Discovery and Structure Elucidation of Natural Products from Underexplored Bacterial Species

Dissertation

zur Erlangung des Grades

der Doktorin der Naturwissenschaften

der Naturwissenschaftlich-Technischen Fakultät

der Universität des Saarlandes

Von

Joy Birkelbach

Saarbrücken

2023

Tag des Kolloquiums:	13.12.2023
Dekan:	Prof. Dr. Ludger Santen
Berichterstatter:	Prof. Dr. Rolf Müller
	Prof. Dr. Anna Hirsch
Vorsitz:	Prof. Dr. Uli Kazmaier
Akad. Mitarbeiter:	Dr. Marc Stierhof

Die vorliegende Arbeit entstand im Zeitraum März 2019 bis September 2023 unter der Anleitung von Herrn Prof. Dr. Rolf Müller am Institut für Pharmazeutische Biotechnologie der Naturwissenschaftlich-Technischen Fakultät der Universität des Saarlandes.

„Das Fragezeichen ist ein Symbol für das Unbekannte. Für unbeantwortete Fragen, ungelöste Rätsel. Unsere Aufgabe ist es, Fragen zu beantworten, Rätsel zu lösen und Geheimnisse jeglicher Art zu lüften.“

Justus Jonas, die drei Fragezeichen

Danksagung

Als erstes möchte ich meinem Doktorvater Prof. Dr. Rolf Müller für die Möglichkeit danken, in seiner Arbeitsgruppe meine Dissertation durchführen zu dürfen. Zum einen habe ich die Freiheit in der Projektgestaltung geschätzt, aber auch die hilfreichen Gespräche und Diskussionen bei Fragen und Unklarheiten. Vielen Dank für das Vertrauen, das du mir entgegengebracht hast!

Zudem bedanke ich mich bei Prof. Dr. Anna Hirsch für die Annahme des Co-Referats und die wissenschaftliche Begleitung. Danke für dein Interesse und die Beiträge während der Promotionskomitees und des Review-scheibens. Sie waren nicht nur sehr hilfreich, sondern auch motivierend.

Bei meinem Betreuer Dr. Daniel Krug möchte ich mich für die stetige Betreuung, viele wissenschaftliche Diskussionen, Motivation, Hilfe und Weitblick bedanken. Danke auch, dass du mir das Vertrauen entgegengebracht hast, dass ich mich mit vielen Analytischen Geräten von innen und außen beschäftigen konnte. Danke Daniel für deine Positive und motivierende Art, für das Korrekturlesen meiner Dissertation und die vielen Tipps beim Schreiben von Manuskripten (und Auffinden von Daten). Auch bei Dr. Judith Hoffmann, Dr. Alexander Popoff und Prof. Dr. Lena Keller, meinen NMR- und Strukturaufklärungs-Betreuer/innen, möchte ich mich bedanken für die vielen Stunden, die ihr mit mir verbracht habt um das Chaos der Moleküle zu bändigen. Danke, dass ihr eure Begeisterung für NMR an mich weitergegeben habt. Ein besonderer Dank dabei geht an dich, Judith, für das unermüdliche Korrekturlesen meiner Strukturaufklärungstexte und -Tabellen und das Einarbeiten in die NMR-Verantwortlichkeiten.

Der gesamten MINS-Gruppe möchte ich für die gute Arbeitsatmosphäre danken. Durch euch war die Promotion nicht nur Arbeit, sondern zusammen forschen, gegenseitiges helfen und miteinander lernen. Besonderer Dank gilt dabei meinen Projektpartnern Hu und Yunsheng, meinen (wechselnden) Büromitbewohnern Maja, Ronald, Christine, Nico, Alex V., Chris und Philipp, sowie der treuen Kollegenfreundin Anna-Lena. Danke, dass ihr meine Projekte bereichert und mich immer wieder motiviert habt. Besonders dankbar bin ich für alle Freundschaften, die sich während der Promotion entwickelt haben. Ein großer Dank geht auch an Bine für die Zeit, die wir zusammen an den HPLC-MSs verbracht haben. Ich habe dabei soo viel gelernt und dabei auch noch Spaß gehabt.

Schließlich möchte ich meiner Familie, besonders meinen Eltern, danken. Vielen Dank, dass ihr mich immer unterstützt und ermutigt das zu tun, was mir Spaß macht und das Neue zu wagen.

Zu guter Letzt danke ich dir, Sebastian. Ich bin dankbar, dass wir dieses Kapitel gemeinsam gemeistert haben und daran wachsen konnten. Danke für die unendlichen Ermutigungen, deine unbezahlbare Unterstützung, und all die schönene Momente dazwischen.

Vorveröffentlichungen aus der Dissertation

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

Publikationen

Hu Zeng†, **Joy Birkelbach**†, Judith Hoffmann, Alexander Popoff, Carsten Volz, and Rolf Müller*: Expanding the ajudazol cytotoxin scaffold: Insights from genome mining, biosynthetic investigations, and novel derivatives. Journal of Natural Products **2022**
DOI: 10.1021/acs.jnatprod.2c00637

Yunsheng Gao†, **Joy Birkelbach**†, Chengzhang Fu, Jennifer Herrmann, Herbert Irschik, Bernd Morgenstern, Kerstin Hirschfelder, Ruijuan Li, Youming Zhang, Rolf Jansen*, and Rolf Müller*: The Disorazole Z Family of Highly Potent Anticancer Natural Products from *Sorangium cellulosum*: Structure, Bioactivity, Biosynthesis and Heterologous Expression. Microbiology spectrum **2023**
DOI: 10.1128/spectrum.00730-23

Publikationen, die nicht Teil dieser Arbeit sind

Sebastian Walesch†, **Joy Birkelbach**†, Gwenaëlle Jézéquel, F P Jake Haeckl, Julian D Hegemann, Thomas Hesterkamp, Anna K H Hirsch, Peter Hammann and Rolf Müller*: Fighting antibiotic resistance—strategies and (pre)clinical developments to find new antibacterials. EMBO reports **2023**
DOI 10.15252/embr.202256033

Julian D Hegemann, **Joy Birkelbach**, Sebastian Walesch and Rolf Müller: Current developments in antibiotic discovery. EMBO reports **2023**
DOI 10.15252/embr.202256184

†These authors contributed equally to this work

Tagungsbeiträge

Joy Birkelbach†, Sebastian Walesch†, Daniel Krug†: Myxobacterial Biodiversity – the Basis for Natural Product Discovery by Screening and Genome/Metabolome-Mining (Oral Communication). A Natural Product Workshop to cover best practice in isolation, separation and characterisation. **2021**, January 18, virtual meeting, Ghana.

Joy Birkelbach†, Yunsheng Gao†, Chengzhang Fu, Jennifer Herrmann, Herbert Irschik, Bernd Morgenstern, Kerstin Hirschfelder, Ruijuan Li, Youming Zhang, Rolf Jansen*, and Rolf Müller*: The Disorazole Z Family of Highly Potent Anticancer Natural Products from *Sorangium cellulosum*: Structure, Bioactivity, Biosynthesis and Heterologous Expression (Poster). Bayreuth Natural Products and Drugs Symposium (BNPDS). **2022**, June 27, Bayreuth, Germany.

Joy Birkelbach†, Yunsheng Gao†, Chengzhang Fu, Jennifer Herrmann, Herbert Irschik, Bernd Morgenstern, Kerstin Hirschfelder, Ruijuan Li, Youming Zhang, Rolf Jansen*, and Rolf Müller*: The Disorazole Z Family of Highly Potent Anticancer Natural Products from *Sorangium cellulosum*: Structure, Bioactivity, Biosynthesis and Heterologous Expression (Poster). VAAM workshop on the biology of microorganisms producing natural products. **2022**, September 7-9, Dortmund, Germany.

†These authors contributed equally to this work

Zusammenfassung

Naturstoffe sind eine wertvolle Quelle für die Entdeckung und Entwicklung von Wirkstoffen mit einzigartigen strukturellen Merkmalen und faszinierenden Bioaktivitäten. In dieser Arbeit wurden verschiedene zeitgemäße Ansätze zur Erforschung unzureichend untersuchter bakterieller *Taxa* verwendet, um sowohl neue chemische Strukturen als auch Bioaktivitäten zu entdecken. Ein bioaktivitätsgeleiteter Ansatz mit dem Stamm *Sorangium cellulosum* So ce1875 führte zur Entdeckung einer neuen Unterfamilie zytotoxischer Disorazole mit der Bezeichnung Disorazol Z. Neben der umfassenden Aufklärung ihrer chemischen Struktur werden auch Studien zur Wirkungsweise, Biosynthese und heterologen Expression vorgestellt. Der Stamm *Cystobacter* sp. SBCb004 wurde nach dem Genom- und Metabolom-geleiteten Ansatz erforscht, der zu neuen Ajudazol-Verbindungen und bisher unbekannten Cystopipecotiden führte, die einen einzigartigen 5-Hydroxyl-6-hydroxymethylpipecolsäure-Baustein enthalten. Beide Verbindungsfamilien wurden hinsichtlich ihrer Struktur und Biosynthese untersucht. Schließlich zeigte die Veränderung der Kultivierungsbedingungen und der Medienzusammensetzung eine differentielle Regulierung der Produktion verschiedener Verbindungen in *Zooshikella marina* sp. Uxx12806. Dabei wurde eine neue Verbindungsfamilie entdeckt, die Zoomarinepane genannt werden und eine faszinierende Imidazo[1,2-a]azepin-amin-Einheit aufweisen. Die chemische Struktur wurde aufgeklärt und es wurde ein Biosyntheseweg vorgeschlagen.

Abstract

Natural products are a valuable source for the discovery and development of therapeutic agents exhibiting unique structural features and intriguing bioactivities. This thesis utilised different contemporary approaches to explore under-investigated bacterial *taxa* to discover both novel chemistry and bioactivities. A bioactivity-guided workflow on the strain *Sorangium cellulosum* So ce1875 led to the discovery of a new sub-family of cytotoxic disorazoles, named disorazole Z. In addition to the comprehensive elucidation of their chemical structure, studies on the mode of action, biosynthesis and heterologous expression are presented. The strain *Cystobacter* sp. SBCb004 was explored following a genome- and metabolome-guided approach leading to new ajudazol congeners and previously unknown cystopipecotides, which harbour an unprecedented 5-hydroxyl-6-hydroxymethylpipecolic acid (HHMPA) building block. Both compound families were investigated regarding their structure and biosynthesis. Finally, alteration of cultivation modalities and media composition revealed differential regulation of the production of several compounds in *Zooshikella marina* sp. Uxx12806. Thereby, a novel compound family was discovered, named zoomarinepanes, exhibiting an intriguing imidazo[1,2-a]azepin amine moiety. The chemical structure was elucidated and a biosynthetic pathway was proposed.

Table of contents

1. Introduction

Introduction

1.1. Natural Products: Prolific Scaffolds for Drugs

Natural products are small molecules produced by living organisms. In a narrower sense, natural products are defined as secondary metabolites, which are produced through energy-expensive processes to fulfil non-essential but advantageous biological functions[1]. Based on their biological activities it is evident that natural products serve valuable purposes in agriculture[2,3], aquaculture[4] or nutrition[5,6]. However, they are probably most known for revolutionising human medicine. Due to their diverse chemical scaffolds, and in many cases because of their biological activities, natural products have played and still play a key role in drug development. Based on their biological targets they often serve as lead structures for anti-cancer[7] or anti-infective agents, like antibacterial[8], antifungal[9], antiviral[10] or antiprotozoal therapeutics[11]. Furthermore, there are natural products whose activity and pharmaceutical application cannot be correlated with a biological function at the time of their discovery. However, in more comprehensive and complex assays some of these natural products were later revealed to possess neurological, immunological or cardiovascular activities, to name a few examples[11].

The percentage of natural products among all small molecule drugs approved by the US food and drug administration (FDA) during the last four decades might seem small (5.1%). However, small molecules derived from natural products (31.7%) and synthetic molecules that mimic natural products (35.1%) add up to 66.8%, which demonstrates the importance of natural products for the drug discovery process (**Figure 1.1**)[11].

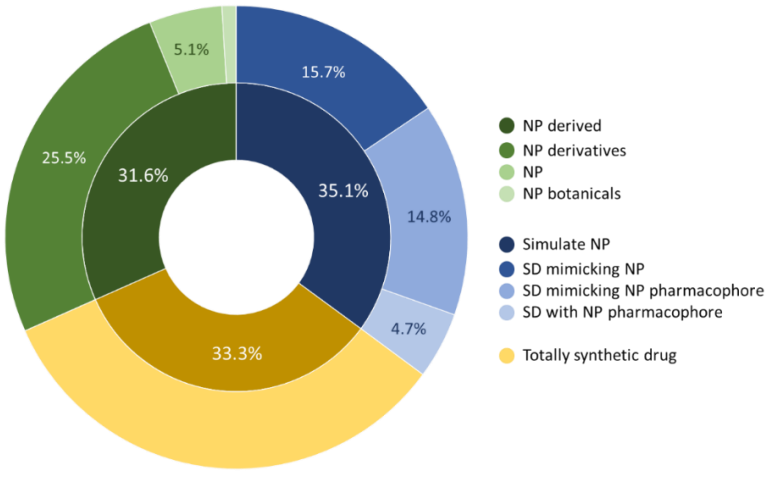


Figure 1.1. Source of small molecule drugs approved by the FDA between 1981 and 2019. *Adapted from Newman & Cragg[11].*

1.2. Natural Products and their Producers – Bacterial Natural Products

In order to improve their fitness and to compete or cooperate with other organisms within the same habitat, many organisms produce natural products that serve a function in inter- and intra-specific communication, symbiosis or competition with other organisms (**Figure 1.2**A)[8,12].

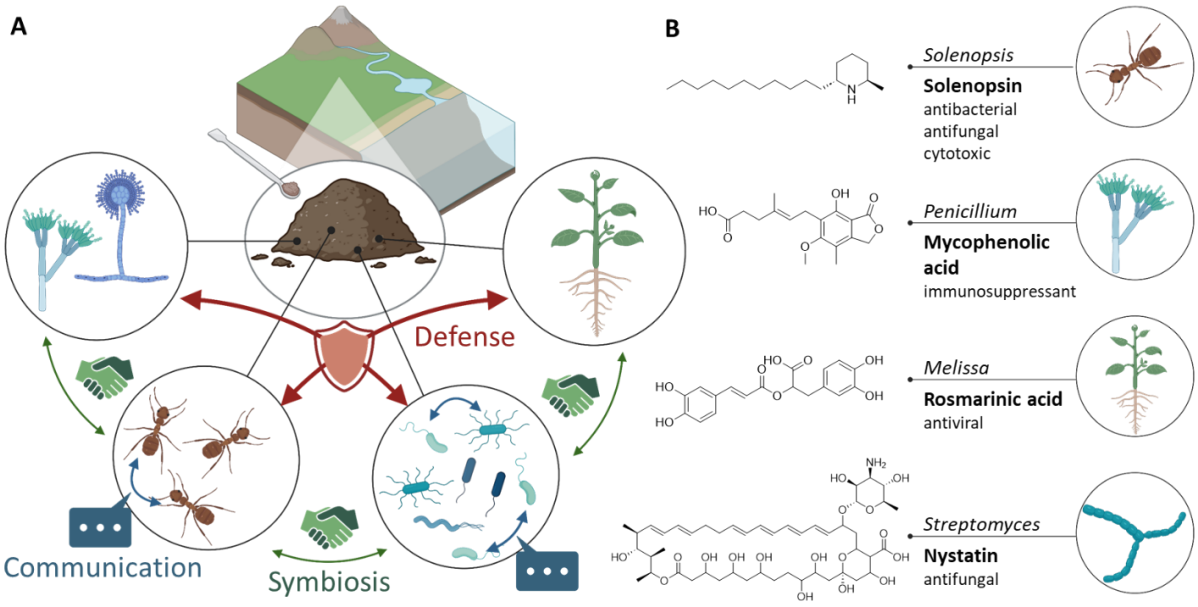


Figure 1.2. Common functions of natural products for organisms of the same habitat (A) and natural products produced by these organisms (B). Examples of natural products: *Antibacterial, antifungal and cytotoxic solenopsins are produced by ants of the* Solenopsis *species*[13]*. Immunosuppressant mycophenolic acid is produced by fungi of the* Penicillium *species*[14]*. Antiviral rosmarinic acid is produced by plants, e.g.* Melissa officinalis[15,16]*. Antifungal nystatin is produced by the bacterium* Streptomyces noursei[17]*.*

Microbes deserve a special place among natural product producers. Generally, they are more easily genetically accessible and easier to cultivate than plants and animals[18]. Moreover, they are oftentimes responsible for - or suspected of being responsible for - the production of natural products through symbiotic relationships especially in animals[19]. About one third of the published microbial natural products originate from bacteria, and two thirds originate from fungi[20]. However, the structural and biosynthetic diversity of bacterial natural products reported until November 2022 in the minimum information about a biosynthetic gene cluster (MIBiG) database is greater than that of natural products from fungi[21]. Additionally, the genes responsible for the biosynthesis of natural products in bacteria are mostly clustered and lack introns, making them easier to genetically manipulate and optimise compared to fungi [18,22]. This in turn is advantageous for large-scale production of agents for clinical trials and application.

Bacteria can be found in diverse habitats ranging from soil to marine environments, living independently or as symbionts. Thus, bacterial natural products evolved to fulfil a broad range of biological purposes in order to communicate and cooperate with each other or with other organisms, or to defend themselves or their host against predators. Therefore, the activity of bacterial natural products covers a very broad spectrum, showing for example anticancer[23] and anti-infective properties[8], which makes them particularly interesting to provide scaffolds for the discovery of novel drugs[1].

In particular, soil-dwelling Gram-positive Actinobacteria played an important role in the discovery of anti-infective and anti-cancer agents[7,8], and still account for the majority of bacterial natural products listed in the NPAtlas database (**Figure 1.3**)[20]. After the discovery and market introduction of streptomycin in 1946[24], the pharmaceutical industry began to screen strain collections of Actinobacteria, specifically prioritising the auspicious genus *Streptomyces*. This approach lead to the development of many antibiotic classes that are still in use today[8]. Further prominent examples of natural product drugs from *Streptomyces* are the anti-cancer drug daunorubicin[25] and the fungicide nystatin (**Figure 1.2**B)[17].

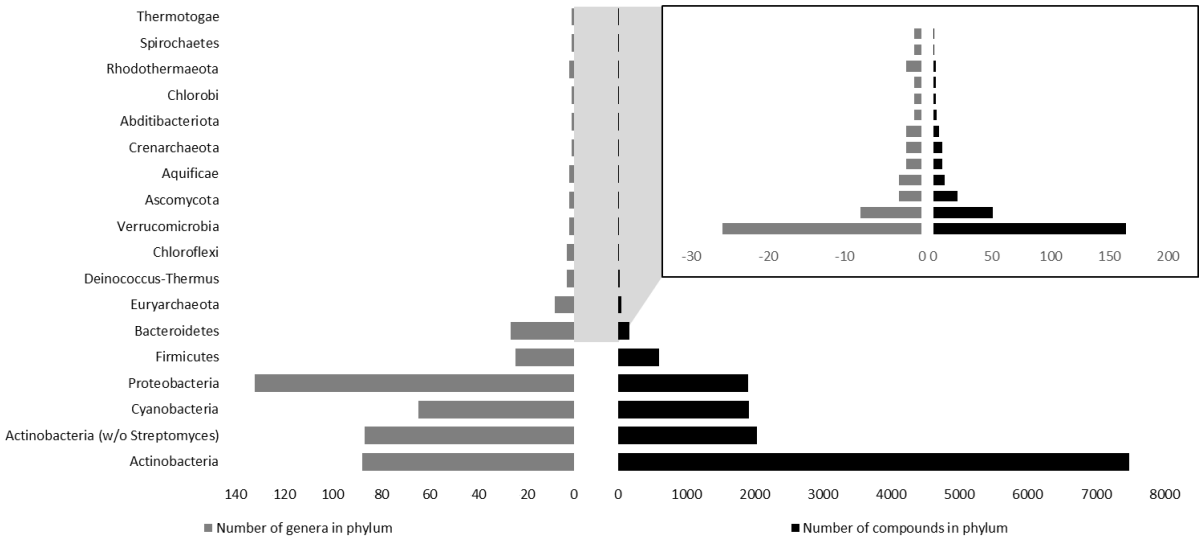


Figure 1.3. The number of compounds per bacterial phylum versus the number of genera per phylum found in the NPAtlas database[20]. *Inset shows phyla with less than 200 compounds. Adapted from Walesch & Birkelbach* et al.*[8].*

Nevertheless, over-mining of phylogenetically limited *phyla* of natural product producing bacteria like Actinobacteria (**Figure 1.3**) led to an increasing compound re-discovery rate[26–28]. Thus, it was assumed that the potential of natural product producers had been exhausted. Three shortcomings of this assumption were elaborated on in our recent review article:

> "First, most screening efforts were built around Actinomycetes, largely ignoring the biosynthetic and antibiotic potential of other microorganisms. Second, the hidden or cryptic biosynthetic potential of microorganisms (including Actinomycetes) was not known. This only changed when complete genome sequences of organisms became available. For example, the genome sequence of the well-studied bacterium *Streptomyces coelicolor* revealed that the number of biosynthetic gene clusters (BGCs) by far exceeded the number of natural products previously connected to the strain (Bentley *et al*, 2002). In routine cultivation settings, most microorganisms produce natural products that account only for fewer than 10% of their BGCs (Katz & Baltz, 2016). Third, with insufficient analytical methods and computational resources it was impossible to quickly de-replicate extracts of newly cultivated microorganisms for known natural products. Without this knowledge, it was difficult to directly de-prioritise extracts featuring known antibiotics or to estimate whether those extracts contained additional antibiotics. Next to Actinobacteria, many well-studied classes, including Myxobacteria, Cyanobacteria, *Pseudomonas*, *Burkholderia*, insect pathogenic bacteria and Firmicutes, have shown potential to produce a wide range of bioactive natural products (Wright, 2017; van Santen *et al*, 2022). Following the observation that chemical diversity correlates with taxonomic distance (Hoffmann *et al*, 2018), there is hope to find new antibiotics in understudied or even new bacterial taxa."[8]

Yet, the genetic potential for secondary metabolite production is not evenly distributed among bacteria – some taxa are more gifted than others[29]. Therefore, it is of utmost importance to assess the amount and nature of unknown BGCs to determine the potential of a bacterium to produce new natural products.

1.2.1. Natural Product Producers from Underexplored Taxa and Habitats

Soils environments are very biodiverse habitats[30] hosting bacteria that have proven to be rich sources for natural products and bioactive compounds[31]. However, under-sampling and a bias towards soil-dwelling Actinobacteria underestimated the potential of other producer-classes and bacteria from other habitats[27]. Compared to soil, marine environments exhibit a rich, albeit less diverse, composition of bacterial *taxa*, which has yielded many novel natural products but has not yet been adequately sampled and explored[30]. Marine natural product producing bacteria are to

date mostly represented by Cyanobacteria (**Figure 1.3**), although there exist halophilic species in many other *taxa* that live autonomously or in symbiosis with marine invertebrates. These *taxa* produce diverse natural products, such as Actinobacteria and Proteobacteria[32]. The structural diversity of marine natural products suggests that they harbour great potential for novel therapeutic agents[32–35].

Analysis of the bacterial natural products reported in the NPAtlas database (**Figure 1.3**) shows that the *phylum* Proteobacteria is less explored than Actinobacteria and harbours more *genera* with less reported natural products. Even though soils are sampled at greater depths than marine habitats, they are still far from being fully exploited to find novel chemistry and potentially new drugs, as exemplified by the enormous potential of Myxobacteria[30,36–38].

Myxobacteria are Gram-negative *δ*-proteobacteria ubiquitous to soils and marine sediments known for their coordinated swarming, predatory behaviour and complex life cycle, as well as the formation of multicellular fruiting bodies under stress[39–41]. Furthermore, reportedly they are among the bacteria with the largest genomes containing up to 16 Mbp[42], and possess a great variety of BGCs encoding for the production of diverse secondary metabolites[37,43]. Currently, the order *Myxococcales* is divided into three suborders *Sorangiineae*, *Cystobacterineae* and *Nannocystineae*. Out of the natural product producers reported in the NPAtlas database[8,20], *Myxococcales* produce 504 compounds from 125 species (**Figure 1.4**). The biggest producer suborder is *Cystobacterineae* with 52% species producing 48% published compounds, closely followed by *Sorangiineae*, which make up 39% of the producers with 45% of myxobacterial compounds. Together they produce more than 90% of all myxobacterial compounds, whereas *Nannocystineae* display 8% of published Myxobactera producing 9% of their compounds (**Figure 1.4**). Analysing the ratio of compounds produced per published *genus*, for Myxobacteria an average of 25 compounds per *genus* were reported, whereas 86 compounds per *genus* were reported for Actinobacteria (the genus *Streptomyces* producing 73% of them). However, on the species level an average of four compounds were reported for Myxobacteria and only three for Actinobacteria. Thus, Myxobacteria have been proposed to be a promising source of novel natural products[44].

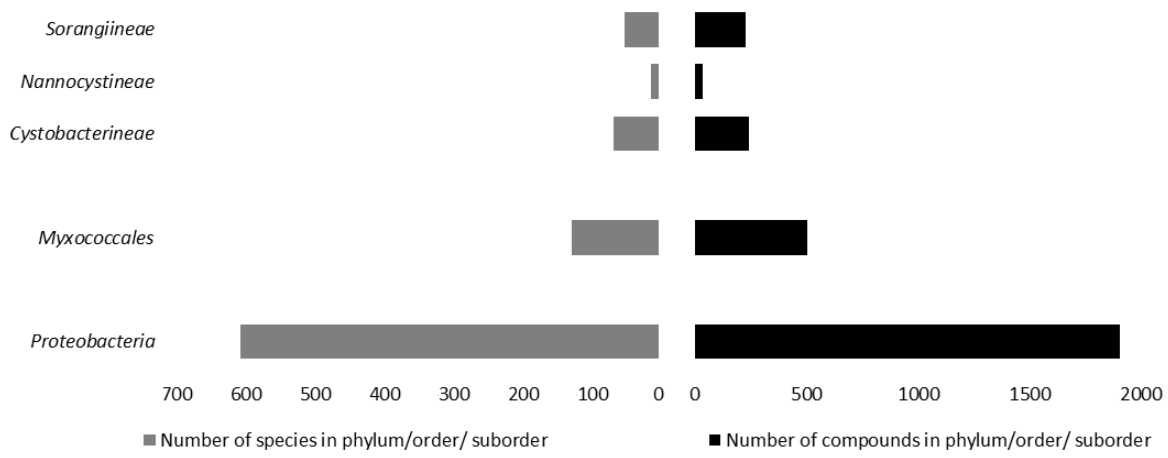


Figure 1.4. The number of compounds per bacterial genus Proteobacteria, the order* Myxococcales *and its suborders* Sorangiineae, Nannocystineae and Cystobacterineae *versus the number of genera species in the NPAtlas database[20]. *Data analysis from Walesch & Birkelbach* et al.*[8].*

1.3. The Chemical Diversity and Biosynthetic Potential of Myxobacterial Natural Products

The physicochemical properties of natural products complement the chemical space of synthetic compound libraries, as they include a vast diversity of structural features optimised through evolution in terms of effectiveness and specificity[8,45,46]. Furthermore, they feature elemental compositions, structural moieties, stereochemistry and an overall complexity that is difficult to achieve with total synthesis[47].

Natural products from Myxobacteria show a variety of activities with astonishing structural diversity that is achieved through complex biosynthetic machineries encoded in BGCs (**Figure 1.5**). Most strains harbour multiple BGCs, which encode for the production of different compound families, as exemplified by the model strain *Myxococcus xanthus* DK1622 that encodes for at least 24 BGCs[48], of which nine compound families have been hitherto isolated and elucidated[38]. Most described natural products from Myxobacteria are non-ribosomal peptides (NRP), polyketides (PK), hybrids thereof (NRP-PK), or ribosomally synthesised and post-translationally modified peptides (RiPP). The immense untapped potential of both known and not-yet cultivated myxobacterial strains has been shown in several studies, both in terms of successful compound discoveries and judged by comprehensive metabolomics-based investigation[38,49–51].

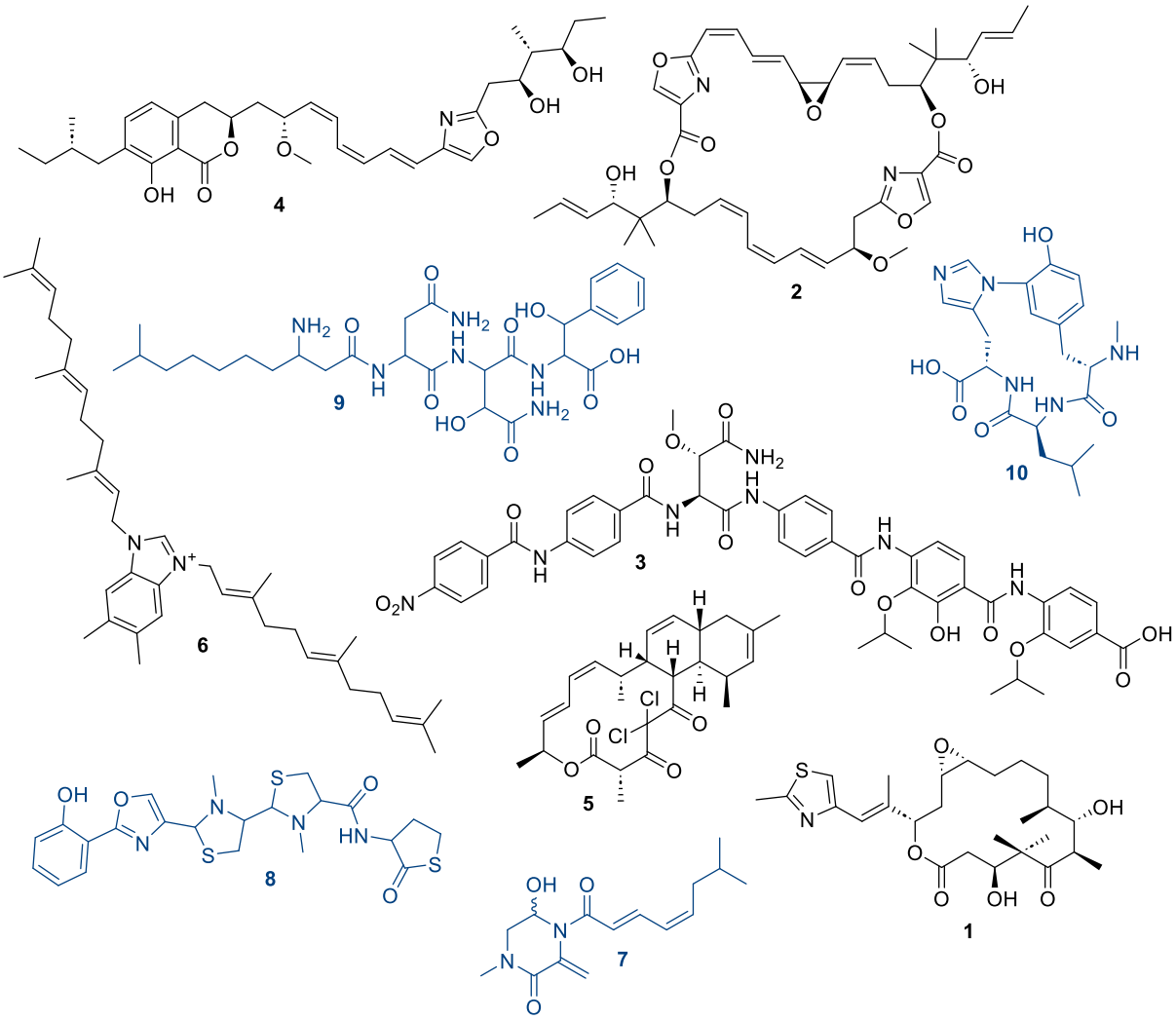


Figure 1.5. Structural, biosynthetic and functional diversity of a selection of myxobacterial natural products. Black*: Compounds with interesting bioactivities:* ***1****: epothilone A[63], a macrocyclic polyketide – non-ribosomal peptide – hybrid (PK-NRP) featuring a thiazole and anti-cancer activity.* ***2****: disorazole A, a* trans*-acyl transferase (*trans*-AT) PK-NRP with a characteristic pseudodimeric structure and anti-cancer activity[52].* ***3****: cystobactamide, a NRP with a striking scaffold of tailored* para*-aminobenzoic acids and broad spectrum antibacterial activity[53].* ***4****: the PK-NRP icumazole with an unique isochromanone-oxazole scaffold and anti-fungal activity[54,55].* ***5****: the chlorinated macrocyclic PK chlorotonil with antiplasmodial activity[56,57].* ***6****: sandacrabin B, a terpene alkaloid with anti-viral activity[58].* ***Blue****: Compounds with striking structural motifs whose bioactivities remain elusive:* ***7****: corallorazine, a piperazine dipeptide[59].* ***8****: the oxazole-thiazolidine-thiolactone-containing NRP sorangibactins[60].* ***9****: lipopeptide cystomanamide A[61].* ***10****: myxaryline, a ribosomally synthesised and post-translationally modified peptide (RiPP) featuring a unique* C-N*-biaryl crosslink[62].*

Prominent examples of natural products from different Myxobacteria (**Figure 1.5**) are the macrocylic PK-NRP epothilone A[63] (**1**) featuring a thiazole and showing anti-cancer activity serving as a model for the FDA approved anticancer drug ixabepilone[64], the *trans*-AT PK-NRP disorazoles

(**2**) with characteristic pseudodimeric structures and anti-cancer activity[52], the NRP cystobactamide (**3**) with a striking scaffold of tailored *para*-aminobenzoic acids and broad spectrum antibacterial activity[53], the PK-NRP icumazole (**4**) with a unique isochromanone-oxazole scaffold and anti-fungal activity[54,55], the chlorinated macrocyclic PK chlorotonil (**5**) with antiplasmodial activity[56,57], and the terpene alkaloid sandacrabin B (**6**) with anti-viral activity[58]. Furthermore, there are many compounds with striking structural motifs whose bioactivities remain to date elusive, such as the piperazine dipeptide corallorazine (**7**)[59], the oxazole-thiazolidine-thiolactone-containing sorangibactins (**8**)[60], the lipopeptides cystomanamides (**9**)[61], or the RiPP myxaryline (**10**)[62] featuring a peculiar *C*-*N*-biaryl crosslink.

NRPS and PKS and their hybrids are among the most prominent biosynthetic machineries in Myxobacteria[49]. Furthermore, most published myxobacterial natural products are biosynthesised by these multimodular megaenzymes, including the natural products studied in this dissertation. Hence, the logic of these assembly lines will be introduced in the following paragraphs.

1.3.1. Non-Ribosomal Peptide Biosynthesis

As the name suggests non-ribosomal peptides (NRP) are not synthesised by ribosomes like most proteins. Instead, the biosynthesis relies on multimodular magasynthetase complexes, the non-ribosomal peptide synthetases (NRPS). Each functional unit, named module, is responsible for the activation of a single amino acid and coupling it to a growing peptide chain (**Figure 1.6**)[65]. However, each module consist of several domains with certain catalytic functions defining the composition of the peptide. Typically, a NRPS assembly line starts with the selection and activation of the carboxylic function of an amino acid by the adenylation domain (A), which loads the substrate to the thiolation domain (often named peptidyl carrier protein, PCP). At this point, if encoded in the BGC or genome, the assembly may be shuttled to tailoring enzymes, which are either located in the same module or are encoded elsewhere in the genome and act in *trans*. A few examples of tailoring enzymes are epimerisation (E), heterocyclisation (Cy), methyl transferase (MT) or oxidation (Ox) domains[66]. After this optional step, the building block is shuttled to the condensation domain (C), which forms the peptide bond to the upstream nascent peptide chain. To terminate the assembly line, the thioesterase domain (TE) disconnects the peptide from the megasynthetase complex by hydrolysis or intramolecular macrocyclisation. In some cases termination domains possess additional functions, such as reducing the resulting carboxylic acid to aldehydes or alcohols[67], and/or mediate macrocyclisation[68]. The resulting peptide can be further modified by tailoring enzymes.

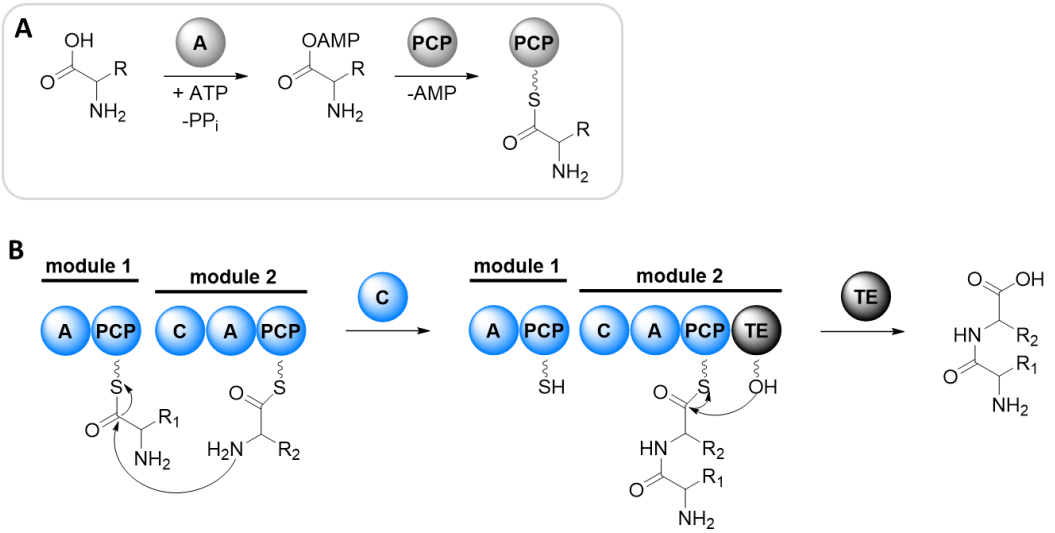


Figure 1.6. Schematic view of the core non-ribosomal peptide assembly line showing the essential domains. A: *Substrate loading by the adenylation domain (A) and the peptidyl carrier protein (PCP).* ***B:*** *Elongation (blue) of the loaded amino acid by the condensation domain (C); and termination (black) via release of the mature peptide by the thioesterase domain (TE).*

As demonstrated in **Figure 1.5**, NRPS assembly lines are not only able to produce "simple" peptides using proteinogenic amino acids but are also able to incorporate highly diverse substrates of which more than 500 are known so far[69], like *para*-aminobenzoic acids (**3**) or β-amino acids (**9**), which creates diverse chemical scaffolds[53,61]. The variability of the incorporation of starter units creates additional diversity in scaffolds. Thus, for example in **9** the assembly line starts with the incorporation of a fatty acid creating a lipopetide scaffold[61].

1.3.2. Polyketide Biosynthesis

Multimodular type I polyketide synthases (type I PKS), like NRPS, are megaenzymes that string together functional units, the modules, with different domains according to the logic of an assembly line. Type II PKS consist of aggregated monofunctional enzymes and type III PKS work independently of acyl carrier proteins, both working iteratively and forming mainly (poly)aromatic structures[70]. As the non-iterative type I PKS machinery is responsible for the biosynthesis of compounds presented in this thesis, it is introduced in more detail below.

Similar to NRPS, the type I PKS machinery consists of three stages: substrate loading, elongation and termination through substrate release (**Figure 1.7**). The assembly line starts with the selection of an acyl-CoA starter unit by an acyl transferase (AT), typically acetyl-CoA. Then, an acyl carrier protein (ACP) transfers the substrate to a ketosynthase (KS) domain of the adjacent elongation

module. Furthermore, the ACP domain of the first elongation module is loaded with a substrate, typically malonyl-CoA, which is also transferred to the KS domain. In this step, the extender-substrate is decarboxylated and attached to the nascent polyketide chain via Claisen condensation. To terminate the assembly line, a thioesterase domain (TE) disconnects the peptide from the megasynthase complex (**Figure 1.7**).[71,72] Similarly to NRPS, some termination domains possess additional functions. For example, macrocyclisation of two monomers mediated by the thioesterase can be observed for disorazole congeners[73].

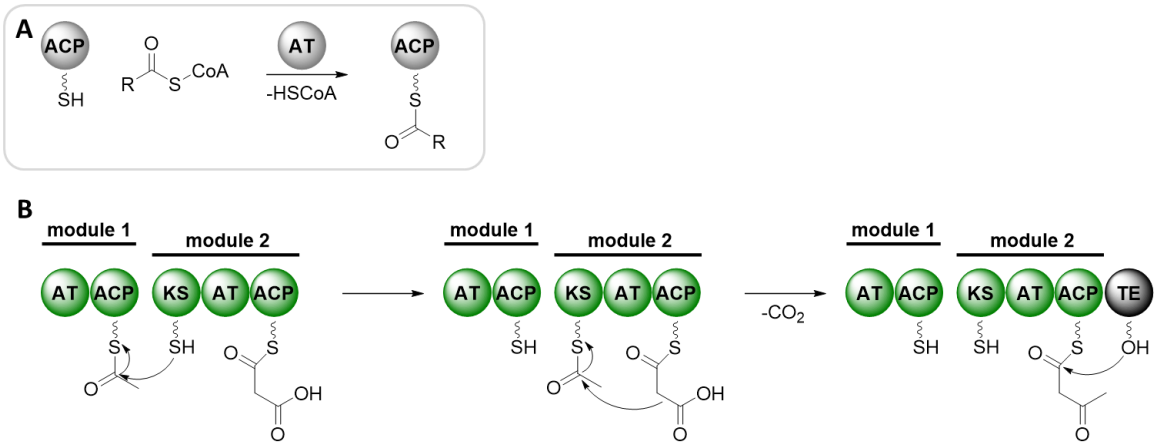


Figure 1.7. Schematic view of the core polyketide assembly line showing the essential domains. ***A:*** *Substrate loading by the acyl transferase (AT) and the acyl carrier protein (ACP).* ***B:*** *Elongation (green) of the loaded building blocks by the ketosynthase (KS) domain and the transfer of the substrate to the thioesterase (TE) domain (black), which releases the mature polyketide of the assembly line.*

The incorporation of different substrates can broaden the structural variety of polyketides, for example propionyl-CoA, malonyl-CoA or benzyl-CoA as starter units[74], or methylmalonyl-CoA, ethylmalonyl-CoA or methoxymalonyl-CoA as extender units[75]. Moreover, tailoring enzymes such as ketoreductases, dehydratases or enoyl-reductases may be implemented in the assembly line and change the chemical scaffold of the produced molecule[71]. A peculiar case of PKS machinery is the *trans*-AT PKS assembly line, which is responsible for the biosynthesis of myxobacterial natural products such as sorangicin[76] and disorazole[77]. In this case, stand-alone AT domains, which are not part of the megaenzyme complex, are recognised by docking-domains, therefore acting in *trans* fashion[78].

1.3.3. Hybrids between NRP and PK Biosynthesis

The structural diversity of NRPS and PKS derived natural products is even more enhanced by the combination of both machineries, which is a common occurrence as PKS-NRPS hybrid BGCs in myxobacterial genomes[49]. PK biosynthesis forms *C-C* bonds, while NRP biosynthesis forms *C-N* bonds. Their common step, the involvement of carboxylic acids in the condensation step during the elongation cycle, facilitates the combination of the two pathways. This allows building blocks to be channelled into the other assembly line and form either *C-N* or *C-C* bonds depending on the order of the modules (**Figure 1.8**)[79].

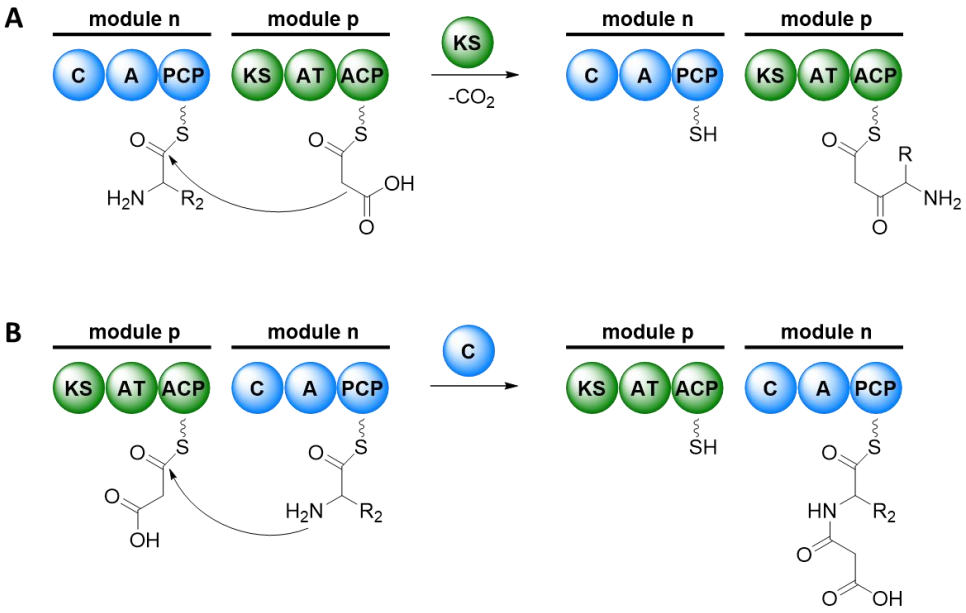


Figure 1.8. Schematic view of the interplay between NRPS and PKS during the elongation cycle. ***A:*** *NRPS-PKS hybrid forming* C-C *bonds between NRP and PK building blocks.* ***B:*** *PKS-NRPS hybrid forming* C-N *bonds between PK and NRP building blocks. Blue: NRPS domains condensation (C), adenylation (A), peptidyl carrier protein (PCP). Green: PKS domains ketosynthase (KS), acyl transferase (AT), acyl carrier protein (ACP).*

PK-NRP hybrids play a big role in providing scaffolds for drugs, as exemplified by prominent examples displayed in **Figure 1.5**, showing diverse structural motifs such as the *trans*-AT PK-NRP disorazole A (**2**) with a characteristic macrocyclic pseudodimeric structure and anti-cancer activity[52] or the linear PK-NRP icumazole (**4**) with a striking isochromanone-oxazole scaffold and anti-fungal activity[54,55].

1.4. How to Discover New Natural Products

In order to discover novel chemical scaffolds and new bioactive natural compounds different strategies were established over time and are still in use today. The oldest approach is the bioactivity-guided compound discovery, whereas contemporary approaches are the metabolome-guided and the genome-guided approaches. Another approach is the OSMAC (one strain many compounds) or cultivation-based approach that feeds into the activity-guided or metabolome-guided workflows (**Figure 1.9**). Common steps in all approaches include the selection of a strain, the cultivation in a small scale in order to analyse the detected metabolome and dereplicate it against known compounds. After the identification of the compound(s) of interest, the strain is cultivated in large scale, target compounds are purified, their structures elucidated and the biological activity profiled. The following five sections will give insights into the different discovery approaches.

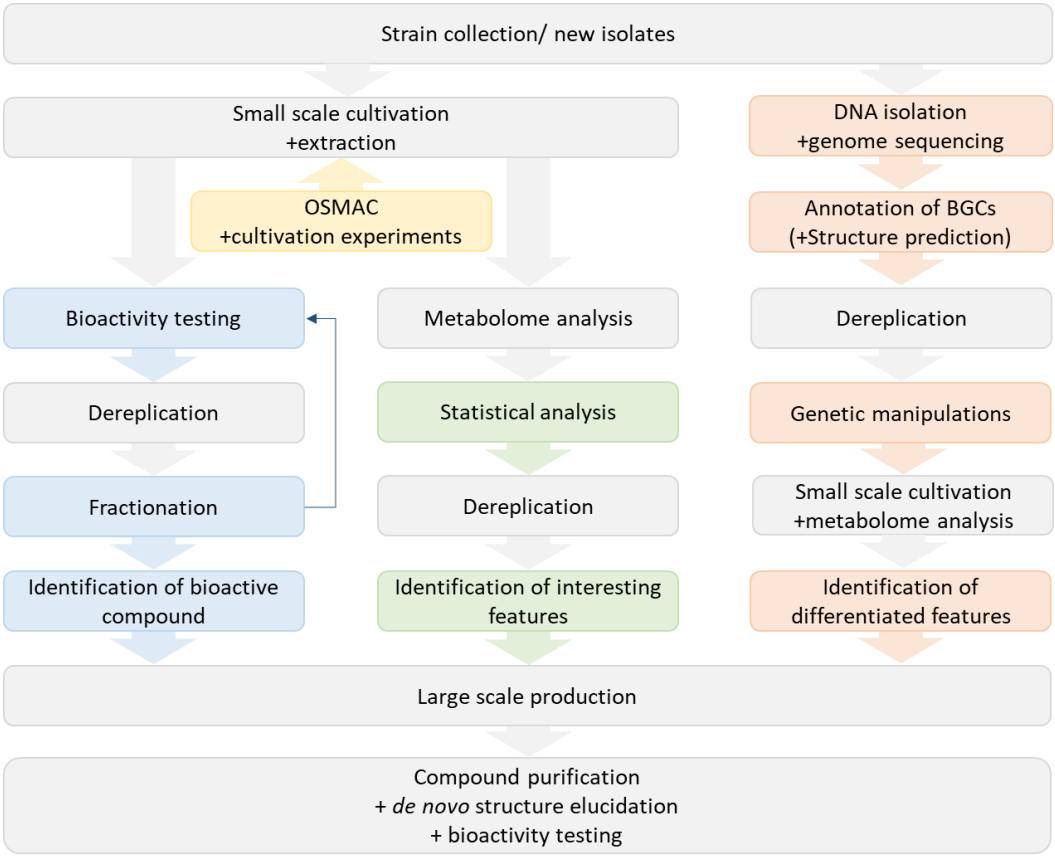


Figure 1.9. Strategies to discover new natural products and the workflow to yield them. *Left, Blue: Bioactivity guided compound discovery. Middle, green: Metabolome-guided compound discovery. Yellow: OSMAC (one strain many compounds) and cultivation-based approaches feed into bioactivity-guided and metabolome-guided workflow. Right, orange: Genome-guided compound discovery. Grey: shared steps in the compound discovery workflow.*

1.4.1. Bioactivity Guided Natural Product Discovery

Bioactivity-guided compound discovery is driven by a straight-forward logical principle. First, the extract of a culture is screened for bioactivity against a selected panel of indicators, for example bacterial or fungal strains, eukaryotic cells or specific proteins. The bioactive extract is then fractionated with selected methods, for example liquid-liquid partitioning, chromatographic separation or crystallisation. The obtained fractions are again tested for bioactivity and further fractionated if applicable. This step may be repeated until a bioactive compound is identified using selected analytical systems. In order to purify sufficient amounts of compound, the strain is cultivated in large scale. Then, the structure of the pure compound is elucidated and further bioactivity tests and mode of action studies may be conducted (**Figure 1.9**).

The bioactivity-guided compound discovery in myxobacteria took off in the 1980's with Reichenbach`s and Höfle's team working mainly with *Sorangium* and *Mycococcus* species[80]. Impressive examples of this work became clear when inspecting the success with the producer strain *S. cellulosum* Soce 12. From a 700 L cultivation they were able to identify five bioactive compounds of which four could be purified: the sorangicins (anti-Gram-positive activity)[81,82], chivosazols (cytotoxic activity)[83], disorazoles (cytotoxic activity)[52], sorangiolids (anti-Gram-positive activity)[84]. The fifth compound, sulfangolids, was identified as early as 1996. However, due to instabilities, it could only be purified and structurally elucidated 16 years later[85,86].

Prominent examples that were discovered during that time are the epothilones[63,87] that served as a model for the clinically approved cancer therapeutic ixabepilone[64], and corallopyronin[88] that is currently in preclinical development for anti-wolbachial treatment of filarial worm infections[89].

The advantage of the bioactivity-guided approach is the relatively straightforward link between a compound and its bioactivity without extensive laboratory work or complete compound purification. However, its drawback is that the re-discovery of known compounds and derivatives thereof without proper de-replication can compromise the work. Furthermore, high production yields of the bioactive compound are often needed in order to achieve a hit during bioactivity screening, while compounds with low production titers may be overlooked. At this point, the genome-guided (Chapter 1.4.2) and metabolome-guided (Chapter 1.4.3) discovery approaches represent complementary alternatives.

1.4.2. Genome Guided Natural Product Discovery

With the first fully sequenced bacterial genome of *Haemophilus influenza* in 1995[90], information on bacterial genomes has expanded rapidly and it became apparent that there is more to the genetic potential of organisms than is translated into the detectable metabolome[18]. One of the first sequenced myxobacterial genomes – and still one of the largest known bacterial genomes – belongs to *Sorangium cellulosum* So ce56 with more than 13 Mbp and 17 putative BGCs[91]. Only recently, however, has more affordable and more reliable next generation sequencing made successful genome-guided natural product discovery possible[92], as this approach requires high quality genome data of larger numbers of promising natural product producers[29]. Today, however, we know several myxobacterial strains that harbour even greater biosynthetic potential, some remarkably devoting 9-13% of their genetic capacity to secondary metabolite biosynthesis[44], which in many cases translates to 20-50 BGCs[93].

After the first steps in the genome-guided compound discovery workflow, cultivation, DNA isolation and sequencing (**Figure 1.9**), the biosynthetic potential of the strain must be evaluated. For this step, the development and availability of annotation tools are essential[94]. The bioinformatics tools antiSMASH[95] and PRISM[96] are commonly used for the identification of BGC, and are even able to provide early insights into the prediction of structural elements using for example the Stachelhaus code for the prediction of NRPS building blocks[97]. Other biosynthetic tools and algorithms, such as BLAST[98] or MiBIG[21], are helpful to dereplicate the BGCs against published genome databases, enabling scientists to identify BGCs of known compound families or similarities to those. As the subsequent steps in the genome-guided workflow are labour extensive and expensive, BGCs for genetic manipulations must be prioritised[99]. This prioritisation can follow several objectives[100], for example: 1.) The investigation of BCS associated with known compounds or unknown compounds. 2.) A selection based on the category or complexity of the BGC. 3.) The presence of certain genes for example resistance genes or genes encoding certain tailoring enzymes. The actual genetic manipulations can be conducted in different ways. BGCs of interest can be inactivated, or cryptic BGC can be either activated or heterologously expressed, in order to identify the respective metabolite(s) using comparative analytical methods and statistical profiling[38,43,101].

Among the early examples of myxobacterial natural products discovered through the genome-guided workflow are the crocapeptins produced by *Chondromyces crocatus* Cm c5. The respective BGC was identified via mutagenesis and the production was increased tenfold by promoter insertion[102]. A recent example of myxobacterial natural products discovered through genome

mining are the sorangibactins, which were targeted because of their similar BGC compared to the not yet fully characterised coelibactin. Intriguing tailoring enzymes and suitability for promoter engineering encouraged Gao *et al*. to heterologously produce, purify and structurally elucidate the compound, giving insights into putative structural elements of coelibactin[60]. A different strategy is to search for antibiotic resistance genes as they may protect antibiotic producer strains from harming themselves. Two examples of such cases are the alkylpyrones[103] and pyxidicyclins[104], the BGCs of which were both targeted because of their close proximity to genes encoding pentapeptide repeat proteins (PRP), which play a role in protecting the producer strains gyrases against gyrase inhibitors. Indeed, after heterologusly expressing the BGCs, novel gyrase inhibitors were produced and Hug *et al*. and Panter *et al*. were able to purify and structurally elucidate the respective compounds[103,104].

The advantage of the genome-guided discovery compared to the bioactivity-guided workflow is that products of cryptic or little expressed BGCs can be identified and analysed. Furthermore, structural feature and BGC homologies to known compounds can be analysed and de-replicated early on in the workflow. Moreover, targeted BGC screening and inactivation/activation can lead to the discovery of natural products with certain properties or scaffolds. However, a prediction about the biological activity of the target compounds is often not possible, and the technique demands genetically accessible producer strains or heterologous hosts that are able to express the target BGC. As the determination of compounds corresponding to the BGC of interest is often challenging, analytical and statistical methods are of great importance to analyse the metabolome[38,99,101].

1.4.3. Metabolome Guided Natural Product Discovery

Crude extracts are complex mixtures of media components, primary and secondary metabolites that were captured at a specific time point after cultivation with a specific method of extraction (**Figure 1.9**)[105,106]. The complexity of the secondary metabolome that is not be accessible through the genome- or bioactivity-guided workflow, however, may still hold untapped potential[43]. This assumption led to another strategy, the metabolome-guided natural product discovery, which is based on the hypothesis that: 1.) The energy-intensive production of secondary metabolites is, however non-essential, advantageous for communication, defence and the establishing of symbiosis[12,38]. 2.) Through evolution, secondary metabolites have been optimised to fit macromolecular targets making them advantageous over chemical libraries[45,46]. 3.) Therefore, a comprehensive overview of all secondary metabolites is of interest in this approach[107]. However,

our view on the metabolome of a strain is always biased on the applied detection methods and analysis tools, which makes the metabolome-guided workflow an analytical-based approach revealing only the detectable metabolome[106]. Due to the availability of analytical systems and the need to progress beyond established systems in order to discover novel natural compounds, the analytical basis of the metabolome-guided workflow has evolved over time.

The history of metabolome-guided discovery story began with TLC- or LC-UV/Vis-based (thin layer chromatography or liquid chromatography) compound identification, as many bioactive compounds known at that time had characteristic UV absorption patterns. For example, DKxanthenes, which are involved in sporulation of Myxobacteria, were discovered as early as 1977 due to their yellow colour[108], and purified and structurally elucidated in 2006 with UV/Vis-guided purification[109]. Nowadays, many easy to detect and easy to purify compounds are already discovered, often referred to as "low hanging fruits"[99], which calls for more sophisticated analytical methods and statistical analysis in order to prioritise compounds for labour-intensive production, purification and structure elucidation (**Figure 1.9**)[107]. Contemporary systems that fulfil these requirements are for example mass spectrometry (MS) and nuclear magnetic resonance (NMR) based methods[110].

Even though NMR applications are most known for *de novo* structure elucidation (see chapter 1.5), it has also been used and is still used for metabolome-guided compound identification. One of the first compounds identified via the application of ^{1}H NMR spectroscopy on crude extracts of *Stigmatella aurantiaca* Sg a15 were the aurachins, exhibiting auspicious signals between 8.6 and 8.8 ppm[111]. Consequently, they were purified, structurally elucidated and profiled for their biological activity. Recent examples of NMR-based compound discovery mostly analyse fractions of crude extracts and compare 1D and 2D NMR spectra to those of published compounds in order to filter out minor or unknown components[112]. A different technique to achieve separation of the crude extract before compound detection is to directly couple chromatographic systems with NMR and MS detectors, for example liquid chromatography-solid phase extraction-NMR-MS (LC-SPE-NMR-MS)[113]. These hyphenate techniques enable relatively low effort in-depth de-replication to prioritise and de-prioritise extracts for compound purification[114]. An example of a compound identified with this workflow are the cystomanamides (**Figure 1.5**, compound **9**).

The most commonly used hyphenated technique for separation and analysis of crude extracts is LC-UV/Vis-electrospray ionisation (ESI)-MS[99,115]. Compared to UV/Vis and NMR, the MS detector is a destructive detector, but it provides valuable information on molecular weight and, in case of high resolution MS, the possibility to predict a molecular formula[115] even of minor components in

crude extracts, which is key information for de-replication. With a directed MS-based metabolomics workflow, Gerth *et al.* were able to identify the double chlorinated compound chlorotonil[56] (**Figure 1.5**, compound **5**), which later proved to have anti-plasmodial activity[57]. Untargeted metabolomics, however, deals with large data sets that require a software-based and statistical analysis in order to evaluate and de-replicate complex natural product samples in an un-biased way[116]. Thereby, for example a subtraction of media components and blanks, the flagging of known metabolites, and the evaluation of up- and down-regulated mass spectrometry features is possible. Often, tandem mass spectrometry (named MS/MS or MS^2) is applied on selected features afterwards[117], which allows for the assignment of compound families using molecular networking for example with the open-source online tool GNPS[118,119]. Furthermore, MS^2 analysis can be used for first predictions of structural moieties, especially amino acid sequences[120]. An example of compounds that were detected using the LC-ESI-MS-guided approach, are the sandacrabins, which showed an intriguing molecular formula prediction[58]. Lacking oxygen, the molecules were supposed to be alkaloids, which is an under-represented microbial natural products class. Therefore, they were prioritised for purification and structure elucidation, sandacrabin B (**Figure 1.5**, compound **6**) showing anti-viral activity after submitting them to biological activity profiling[58].

Nevertheless, the metabolome-guided approach even when analysis includes minor components of natural products mixtures still fails to excessively mine the potential of microbes to produce secondary metabolites. It was shown that using common cultivation techniques usually only less than 10% of biosynthetic gene clusters (BGC) are translated into secondary metabolites[1,121].

1.4.4. Strategies to Optimise or Trigger the Production of Natural Products

The natural products discovery approaches described up to this point deal with either the genetic potential of natural product producers or the metabolomic variety that they produce, ignoring the fact that several factors may regulate or inhibit natural product discovery[122]. In order to unlock the cryptic biosynthetic potential that is not accessible through hitherto described workflows, one can apply various methods[123]. One out of several options, referred to as the one strain many compounds (OSMAC) approach (**Figure 1.9**), shall be introduced in this chapter as it has been explored in this work.

Even though methods belonging to the OSMAC approach have been applied on the cultivation of microbes before[124,125], the term was officially introduced by Schiewe & Zeek in 1999[126], and Bode *et al*. famously reviewed the concept in 2002[122]. In order to encourage a strain with high genetic

potential to produce more than the already known compound(s), altered cultivation approaches can be applied. It is well known that cultivation conditions like the composition of media[127], the pH value of the culture[128], cultivation temperature[129], oxygen supply[130] or the shape and size of the cultivation vessel[131] have great effects on the secondary metabolism of bacteria.

In a recent OSMAC study on *Corallococcus coralloides* different media with or without the addition of chemical elicitors were used to trigger natural product production showing great effects of up- and down regulation of mass spectrometry features when combining several alteration factors[132]. Furthermore, Walt *et al*. were able to enhance the production of several natural products by altering the cultivation conditions of the myxobacterial strain MSr11367, proposing the greatest positive effect by mimicking the natural environment in soil by e.g. the addition of soil extract or lanthanum chloride, and oxygen limitation. Thereby they were able to detect, purify and structurally elucidate three unprecedented secondary metabolites[130].

1.5. Structure Elucidation of Natural Products

Structure elucidation is a key step in the characterisation and quality control of natural products (**Figure 1.9**). It involves several steps, starting with the determination of a molecular formula, followed by the elucidation of the planar structure, and the relative and absolute configuration. Several methods can be applied to solve this workflow, whereas the methods used in this study shall be introduced in the following chapters.

1.5.1. Elemental Composition and Partial Structures: MS and MS/MS

The first step in structure elucidation of natural products is the determination of a molecular formula. First studies reporting the elemental composition of natural products required large amounts of pure compound for a combination of destructive elemental analysis and MS[82]. Nowadays, high-resolution MS (HR-MS), for natural products mostly electrospray ionisation (ESI)-HR-MS, replaced elemental analysis requiring only few micrograms of compound. Furthermore, when applying chromatographic resolution prior to MS, even molecular formulae of compounds in complex mixtures can be predicted[133]. Moreover, the application of MS^2 experiments is able to give insights into partial structures when analysing the fragmentation pattern[133]. Especially for the structure elucidation of peptides, MS^2 fragmentation analysis is often used to establish amino acid sequences. With ion trapping instruments like quadrupole ion traps or Fourier transformation ion cyclotron resonance MS (FT-ICR), MS^n experiments can be applied, which further fragment ions

that resulted from previous fragmentation[134]. These experiments enable the researcher to get insight into the composition of selected fragments, which is very valuable e.g. for the determination of an amino acid sequence of cyclic peptides.

1.5.2. Planar Structure: NMR

30 years ago, the determination of the structural formula of secondary metabolites relied on gram to several hundred-milligram quantities, which were analysed after decomposition or derivatisation reactions, in some cases leading to false structures that were revised by total synthesis[135]. In most cases, an assignment of the configuration of a molecule was unthinkable[135]. Since then, structure elucidation of natural products shifted to using NMR, as the quality of NMR analysis immensely increased due to the optimisation of superconducting magnets, the introduction of cryogenically cooled probeheads and the ability to use low-volume sample tubes, requiring less than one milligram compound[136].

The following two paragraphs will guide the reader through the structure elucidation process to yield a planar structure, which is described in more detail in several textbooks and reviews[137–140]. The experiments used in this study, and typically used in natural products structure elucidation, are listed in **Table 1.1** and a visualisation of the workflow is shown in **Figure 1.10**.

One-dimensional NMR experiments. The determination of a planar structure usually starts with the evaluation of one dimensional (1D) NMR spectra. As NMR is applicable on nuclei with a spin unequal zero, only NMR experiments detecting few isotopes are possible, like ^{1}H, ^{2}H, ^{13}C, ^{15}N, ^{19}F, or ^{31}P. In natural product structure elucidation, most commonly ^{1}H and ^{13}C NMR are used as 1D experiments, since these make up the majority of the molecular backbone. Both experiments are typically used to verify the molecular formula and double bond equivalents (DBE) of the respective compound (**Figure 1.10**), which were proposed by HR-MS beforehand (1.5.1). Integration of the signals equal the amount of magnetically equivalent protons. However, one needs to keep in mind that exchangeable protons, like OH or NH, will not be detected in protic solvents. This effect can be used in comparative experiments, in protic and aprotic solvents, to determine exchangeable protons. Furthermore, the chemical shift of a proton or carbon gives insights into the chemical surroundings of the atom, e.g. a proton with adjacent electron-withdrawing groups is electromagnetically de-shielded, which results in a downfield shift of proton resonances, which can be observed as higher ppm values. A difference of ^{1}H NMR compared to ^{13}C NMR is the higher abundance of isotopes in nature for the latter and their higher gyromagnetic ratio, which results in higher signal-to-noise ratios or shorter measuring times. In order to observe clearer spectra, ^{13}C

spectra are typically de-coupled from proton resonances, yielding line spectra. On the other hand, the multiplicity of proton signals in ^{1}H spectra is highly appreciated as it can be used to determine the amount of homonuclear coupling partners with the n+1 rule, in short: how many protons neighbour the respective proton. Nevertheless, for a detailed *de novo* structural analysis and to discover the connectivity of a molecule, two-dimensional (2D) NMR experiments are necessary.

Table 1.1. The NMR experiments used in this study and commonly used in natural products structure elucidation [140].

Abbreviation	Experiment	Information relevant for structure elucidation	Comment, (mentioned version)
^{1}H	^{1}H NMR	chemical shift, spin-spin coupling, intensity	
^{13}C	^{13}C NMR	chemical shift	Broad-band decoupled
HMQC	Heteronuclear multiple quantum coherence	$^{1}J_{C,H}$ correlation	
HSQC	Heteronuclear single quantum coherence	$^{1}J_{C,H}$ correlation	Phase sensitive
COSY	$^{1}H,^{1}H$ correlation spectroscopy	coupled pairs of protons	
TOCSY	Total correlation spectroscopy	Spin system	(Selective 1D TOCSY)
HMBC	Heteronuclear multiple bond correlation	$^{2\text{-}4}J_{X,H}$ correlation	(N-HMBC)
1,1-ADEQUATE	Adequate sensitivity double quantum spectroscopy	$^{2}J_{C,H}$ correlation	
NOESY	Nuclear Overhauser enhancement spectroscopy	Spatial proximity of protons	Applicable for small and large natural products, (Selective 1D NOESY)
ROESY	Rotating frame Overhauser enhancement spectroscopy	Spatial proximity of protons	Applicable for all sizes of natural products, (Selective 1D ROESY)

Highlighted in grey: The “standard set” of NMR experiments for natural products structure elucidation.

Two-dimensional NMR experiments. In order to determine the connection of protons to a respective carbon, heteronuclear multiple quantum coherence (HMQC) spectroscopy can be used, which uses both proton excitation and proton detection, while the focus lies on the large heteronucelar coupling constant ($^{1}J_{C,H}$). Thus, only cross peaks of direct bonded protons to carbons are visible in the 2D spectrum. Compared to HMQC, heteronuclear single quantum coherence (HSQC) spectroscopy is phase sensitive, which results in different phasing of methyl and methine groups versus methylene groups, in the spectrum visualised with two colour-coding, which gives further insights into structural motifs (**Figure 1.10**). Further, to identify neighbouring protons $^{1}H,^{1}H$ correlation spectroscopy (COSY) can be used. Thereby, geminal, vicinal and long-range couplings are observed, which enable the establishing of a spin system by following one signal to another (**Figure 1.10**). Total correlation spectroscopy (TOCSY), on the other hand, shows the complete spin system in one 2D spectrum, which is especially helpful e.g. for peptides and oligosaccharides to establish structural moieties. The last experiment of a "standard set" for natural product structure elucidation is the heteronuclear multiple bond correlation (HMBC) spectroscopy that displays exactly what the name indicates. Typical are $^{2\text{-}4}J_{C,H}$ correlations, but in some cases $^{2\text{-}6}J_{C,H}$ correlations can be observed. HMBC experiments can be used to link the fragments established through previous experiments, whose spin systems are interrupted by fully substituted heteroatoms (**Figure 1.10**). Furthermore, for example the position of amino or hydroxyl groups can be established when showing correlations through NH or OH protons to adjacent carbons. Structure elucidation of natural products with a high nitrogen content may benefit from N-HMBC experiments, which show mostly $^{2\text{-}4}J_{N,H}$ correlations analogously to standard HMBC experiments. Molecules with a high number of fully substituted carbons, often referred to as quaternary carbons, are often difficult to structurally elucidate. A helpful experiment to selectively establish $^{n}J_{C,H}$ correlations are ADEQUATE experiments, which display long range heteronuclear correlations through proton excitation, carbon-carbon magnetisation transfer ($^{n}J_{C,C}$) and proton detection[141,142]. In this study, 1,1-ADEQUATE experiments were used to exclusively see $^{2}J_{C,H}$ correlations.

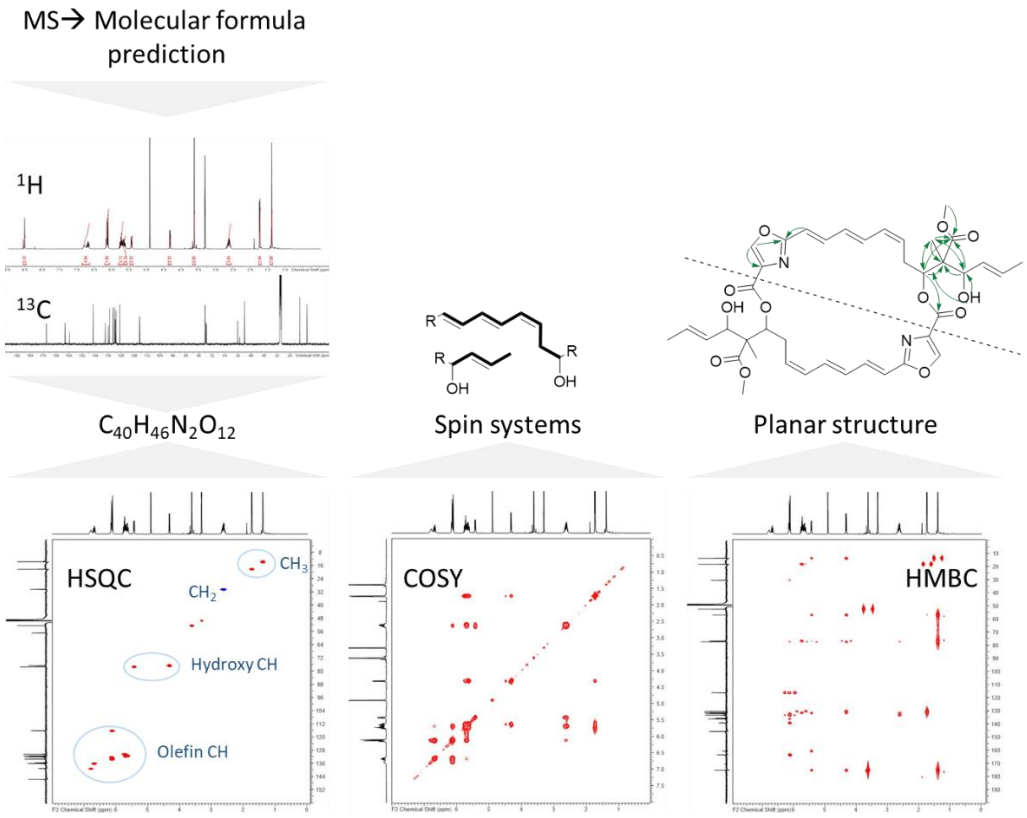


Figure 1.10. Workflow of elucidating the planar structure of natural products using NMR experiments.

1.5.3. Determination of the Relative Configuration

Since most natural products are not planar structures, and bioactivity depends on the configuration of a molecule, complete structure elucidation including the determination of the absolute configuration is of high interest. However, as the structural variety of natural products is vast, there is no general method to solve the configuration – methods are selected on a case-by-case basis. A selection of methods used in this study shall be introduced in the following paragraphs[137–140].

One- and two-dimensional NMR experiments. Non-destructive methods are of great interest, as they do not "waste" precious compound. NMR experiments offer some possibilities for this. Without running any further experiments, some information about the configuration of a molecule can be gathered from the analysis of coupling constants ($J_{H,H}$) observable in ^{1}H spectra. This analysis is especially helpful to determine the configuration of double bonds and six-membered ring systems, e.g. in sugars[139,143]. If spectra are crowded and signals overlap, one can

use experiments that are designed to unravel overlapping signals, like 2D *J*-resolved ^{1}H NMR[140], or selectively excite certain regions, like selective 1D versions of TOCSY, ROESY or NOESY[140]. Furthermore, 2D NOE experiments like the nuclear Overhauser enhancement (NOESY) and the rotating frame Overhauser enhancement (ROESY) spectroscopy are commonly used to get insight into the relative configuration of a molecule. Both experiments, as the name suggests, rely on the Overhauser effect that shows proton-proton correlations through space. However, NOESY is only applicable to smaller and large natural products, as the NOE equals zero at a molecular weight about 1000 Da in this experiment. On the other hand, the NOE in ROESY experiments is always positive, therefore it can be used for molecules of all sizes[140]. Since there is a correlation between the spatial proximity of respective protons and NOE signal intensity, semi-quantitative analysis to determine proton-distances can be applied, which helps to build a spatial representation of a molecule. This is helpful, for example, to establish the relative configuration of sugars[139]. However, in many cases, the presence or absence of a NOE signal already delivers information needed to propose the configuration of certain moieties. For example, this can be used to elucidate the configuration of double bonds as follows: the presence of a signal indicates that the position of both protons are in the same direction (*Z*), whereas no signal means that protons are in the opposite direction (*E*)[144].

Crystallography. X-ray crystallography is the ultimate spectroscopic method to elucidate molecular structures showing the relative and in some cases even the absolute configuration, e.g. in case of crystals with a certain quality and the use of high resolution instruments. Already in 1932 the first crystal structure of a steroid was published, enabling the structure elucidation of compounds that failed to be solved beforehand[145]. Furthermore, X-ray crystallography was able to clarify conflicting NMR data, as exemplified by the structure elucidation of sorangicin A[146] or disorazole Z[144]. However, in order to receive good quality crystallographic data, crystals of sufficient quality are needed, which leads to a few problems often occurring in natural products structure elucidation: 1. not all compounds crystallise or are pure enough to yield good quality crystals. 2. The yield of pure compounds is often not high enough to form crystals in a suitable size from saturated solutions[136].

1.5.4. Determination of the Absolute Configuration

One problem in typical before mentioned NMR and LC-UV/MS analysis is that enantiomers are not distinguishable from each other. Therefore, the determination of the absolute configuration is not possible using methods described in the previous chapter. This chapter will introduce some typical strategies to determine the absolute configuration of natural products[147]. The selection is mainly based on methods used in this study.

ORD and CD. Non-destructive methods traditionally used for the analysis of chiral optically active molecules are optical rotation dispersion (ORD) and circular dichroism (CD)[147]. However, these are comparative analyses, therefore need an authentic standard or computationally generated data sets, which are compared to the natural product of interest[138,148].

Chromatography and derivatisation techniques. Another comparative analysis with authentic standards can be achieved using chiral columns on chromatographic systems, e.g. LC or gas chromatography (GC), coupled with detectors, e.g. UV/Vis or MS[61]. However, respective authentic standards of novel natural products mostly require full chemical synthesis, which is laborious and not feasible in many cases. Thus, the analysis of the absolute configuration of certain building blocks is often achieved through derivatisation reactions with chiral reagents yielding diastereomers that show different physicochemical properties. For example, the determination of the absolute configuration of amino acids in a peptide can be accomplished with Marfey's reaction after hydrolysis (**Figure 1.11**), and the comparative analysis of the derivatisation product with respective derivatised amino acid standards on LC-UV/Vis-MS[149,150]. Furthermore, the configuration of sugar building blocks or fatty acids and respective authentic standards can be analysed by GC-MS after derivatisation and silylation[148,151,152]. Moreover, there are several derivatisation techniques using NMR instead of LC-MS as detection[147]. One example is the Mosher esterification[153]. Typically, after derivatisation of hydroxyl groups with the (*S*)- and (*R*)-Mosher reagent, respectively, ^{1}H and ^{13}C NMR experiments are conducted. The difference of the chemical shifts between the (*S*)- and (*R*)-Mosher derivatives is calculated. Protons and carbons in close proximity to the derivatised stereogenic centre on either side show positive or negative difference values, respectively, indicating their position to be (*R*) or (*S*) (**Figure 1.11**)[153].

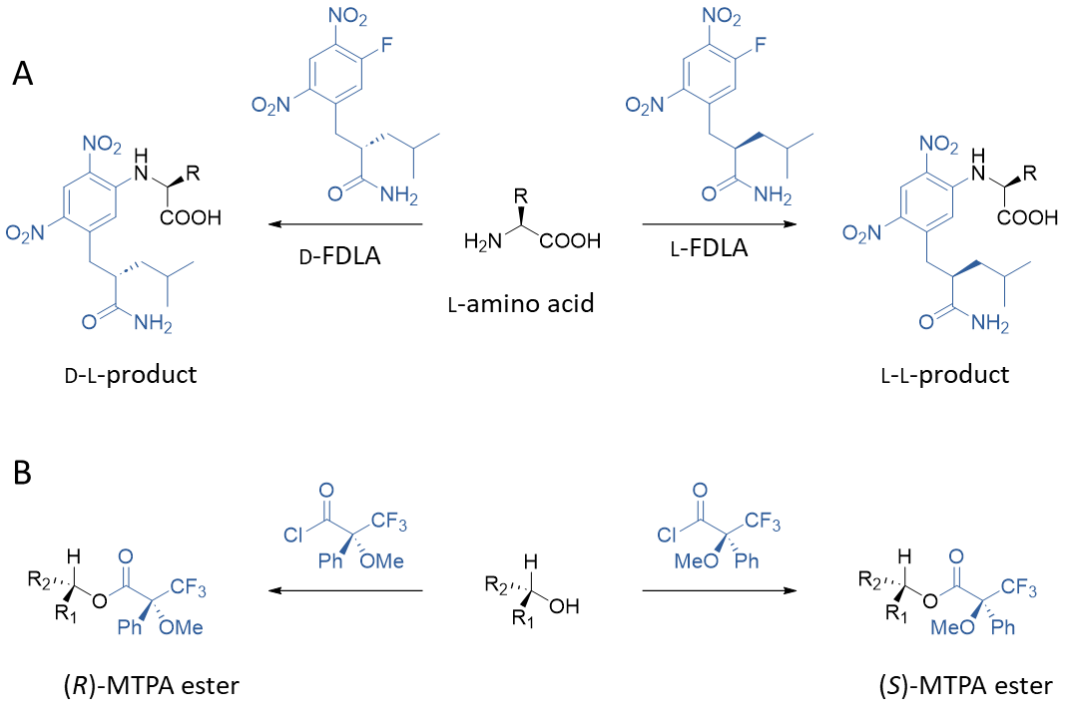


Figure 1.11. Derivatisation reactions A: Mosher-esterification and B: Marfey's reaction.

1.6. Outline of this Study

The overall aim of this study is the discovery and structure elucidation of new natural products. Therefore, two myxobacterial strains and one *Zooshikella* strain, both underexplored bacterial taxa with high biosynthetic potential, were subjected to four different natural product discovery approaches (**Figure 1.12**): The bioactivity-guided approach (Chapter 2), the genome-guided approach (Chapter 3), the metabolome-guided approach (Chapter 4), and cultivation experiments in combination with investigation of the metabolome inspired by the OSMAC approach (Chapter 5). In chapter 2-4, this author's work focused on the characterisation and structure elucidation of identified natural products, whereas in chapter 5 this author took the lead in the natural product discovery process, performed characterisation and structure elucidation, and explored the biosynthetic pathway of discovered compounds.

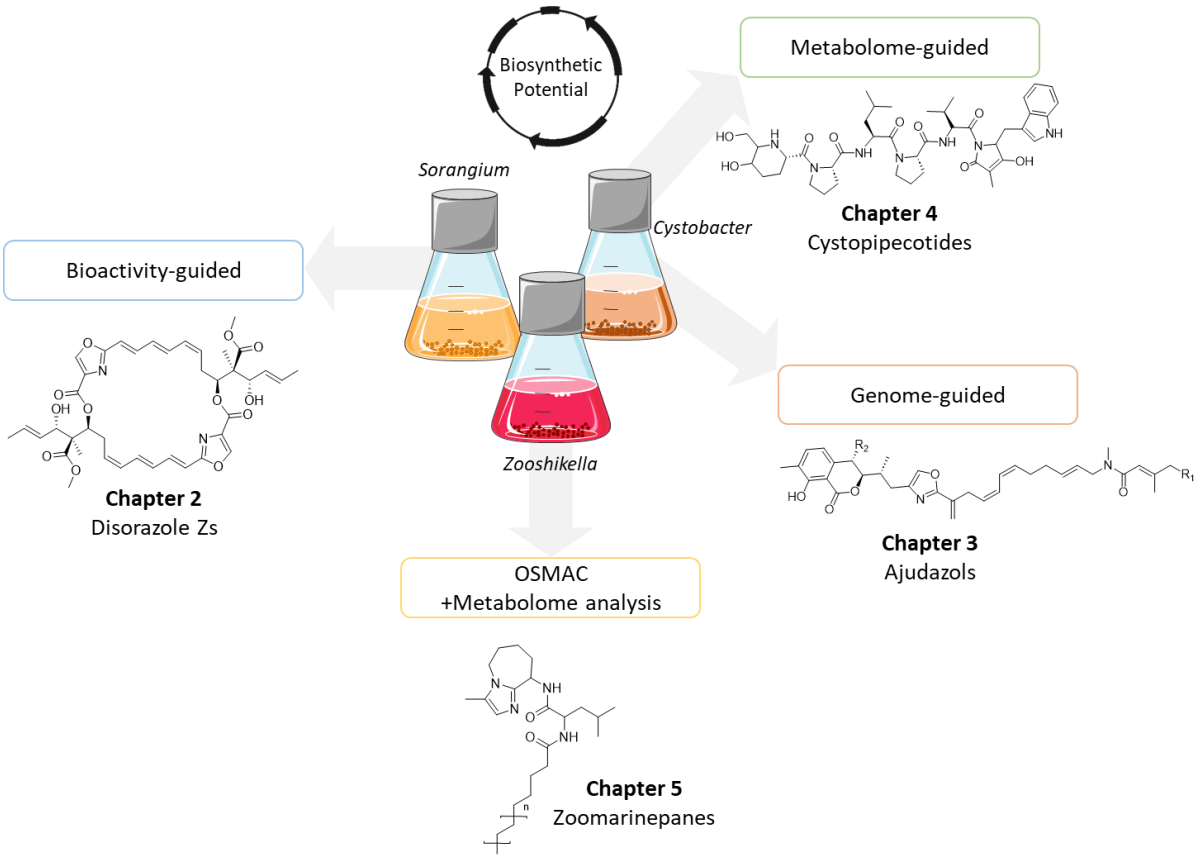


Figure 1.12. Strategies used in this study to discover new natural products. *The bioactivity-guided workflow led to the discovery of a novel subclass of disorazoles (Chapter 2). The genome-guided workflow led to the discovery of new ajudazol congeners (Chapter 3). The metabolome-guided workflow led to the discovery of cystopipecotides (Chapter 4). And an OSMAC inspired workflow led to the discovery of zoomarinepanes (Chapter 5).*

Chapter 2 reports screening of the myxobacterial strain *Sorangium cellulosum* So ce1875 for bioactive natural products. Using the traditional bioactivity-guided natural products discovery approach, a novel subclass of disorazoles was discovered, named disorazole Z. Ten congeners were isolated and subsequently characterised using ESI-HR-MS, the "standard set" of 1D and 2D NMR experiments, Mosher ester analysis and X-ray crystallography. Furthermore, in this study the BGC responsible for the production of disorazole Z was identified in an alternative producer, *S. cellulosum* So ce427, elucidated and heterologously expressed in the host *Myxococcus xanthus* DK1622, which paves the way for biosynthetic engineering and biotechnological optimisation of disorazole production. Finally, the bioactivity of disorazole Z was profiled and the mode of action of the anti-cancer activity characterised.

Chapter 3 and 4 focus on the myxobacterial strain *Cystobacter* sp. SBCb004, which showed high biosynthetic potential with the prediction of 49 secondary metabolite BGCs, whereas 42 of those

had not been characterised previously. Therefore, SBCb004 was subjected to both, the genome-guided workflow discovering new ajudazole congeners by BGC-profiling and metabolome analysis (Chapter 3), and to the metabolome-guided workflow using MS^2 profiling to discover unprecedented cystopipecotides (Chapter 3). Eight ajudazol congeners and two cystopipecotides were isolated, characterised, and structurally elucidated. In these projects, ESI-HR-MS and $–MS^2$ analysis, the "standard set" of NMR experiments plus 1,1-ADEQUATE and N-HMBC were employed for the elucidation of planar structures. Furthermore, coupling constant analysis and ROESY NMR experiments, as well as derivatisation reactions plus chromatographic methods were used to analyse the absolute configuration. Finally, biosynthetic pathways for both ajudazols and cystopipecotides were proposed, and compounds were profiled for bioactivity.

In chapter 5, the under-explored, non-myxobacterial strain *Zooshikella marina* sp. Uxx12806 was submitted to cultivation experiments classifying as OSMAC approach. Changing the salinity of media and cultivation at different temperatures showed the differential production of the known compounds prodigiosin and althiomycin, and led to the discovery of a novel compound family – zoomarinepanes. Four congeners were purified and structurally elucidated using the standard set of NMR experiments. Furthermore, Marfey's derivatisation and FT-ICR-MS and $–MS^2$ experiments led to the elucidation of the absolute configuration. Finally, the biosynthetic pathway was proposed based on *in silico* analysis of the putative BGC, and the compound was submitted to bioactivity profiling.

In summary, this thesis shows the effectiveness of all mentioned discovery approaches to replenish the pool of structural diversity in natural products, and to exploit the biosynthetic potential of natural product producers.

1.7. References

1. Katz, L. & Baltz, R. H. Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **43,** 155–176; 10.1007/s10295-015-1723-5 (2016).

2. Majeed, A., Abbasi, M. K., Hameed, S., Imran, A. & Rahim, N. Isolation and characterization of plant growth-promoting rhizobacteria from wheat rhizosphere and their effect on plant growth promotion. *Front. Microbiol.* **6,** 198; 10.3389/fmicb.2015.00198 (2015).

3. Dayan, F. E., Cantrell, C. L. & Duke, S. O. Natural products in crop protection. *Bioorg. Med. Chem.* **17,** 4022–4034; 10.1016/j.bmc.2009.01.046 (2009).

4. Rather, I. A. *et al.* Diversity of Marine Bacteria and Their Bacteriocins: Applications in Aquaculture. *Reviews in Fisheries Science & Aquaculture* **25,** 257–269; 10.1080/23308249.2017.1282417 (2017).

5. Singh, R., Kumar, M., Mittal, A. & Mehta, P. K. Microbial metabolites in nutrition, healthcare and agriculture. *3 Biotech* **7,** 15; 10.1007/s13205-016-0586-4 (2017).

6. Verpoorte, R. & Memelink, J. Engineering secondary metabolite production in plants. *Curr. Opin. Biotechnol.* **13,** 181–187; 10.1016/S0958-1669(02)00308-7 (2002).

7. Da Rocha, A. B., Lopes, R. M. & Schwartsmann, G. Natural products in anticancer therapy. *Curr. Opin. Pharmacol.* **1,** 364–369; 10.1016/S1471-4892(01)00063-7 (2001).

8. Walesch, S. *et al.* Fighting antibiotic resistance-strategies and (pre)clinical developments to find new antibacterials. *EMBO Rep.,* e56033; 10.15252/embr.202256033 (2022).

9. Heard, S. C., Wu, G. & Winter, J. M. Antifungal natural products. *Current opinion in biotechnology* **69,** 232–241; 10.1016/j.copbio.2021.02.001 (2021).

10. Martinez, J. P., Sasse, F., Brönstrup, M., Diez, J. & Meyerhans, A. Antiviral drug discovery: broad-spectrum drugs from nature. *Nat. Prod. Rep.* **32,** 29–48; 10.1039/C4NP00085D (2015).

11. 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.* **83,** 770–803; 10.1021/acs.jnatprod.9b01285 (2020).

12. Atanasov, A. G., Zotchev, S. B., Dirsch, V. M. & Supuran, C. T. Natural products in drug discovery: advances and opportunities. *Nat. Rev. Drug Discov.* **20,** 200–216; 10.1038/s41573-020-00114-z (2021).

13. Fox, E. G. P. & Adams, R. M. M. On the Biological Diversity of Ant Alkaloids. *Annual review of entomology* **67,** 367–385; 10.1146/annurev-ento-072821-063525 (2022).

14. Bentley, R. Mycophenolic Acid: a one hundred year odyssey from antibiotic to immunosuppressant. *Chem. Rev.* **100,** 3801–3826; 10.1021/cr990097b (2000).

15. *Oriente, G. Isolamento e costituzione dell'acido rosmarinico (dal rosmarinus off.)* (1958).

16. Petersen, M. & Simmonds, M. S. J. Rosmarinic acid. *Phytochemistry* **62,** 121–125; 10.1016/S0031-9422(02)00513-7 (2003).

17. Brown, R. F. & Hazen, E. L. *Nystatin and actidione: Two antifungal agents produced by Streptomyces noursei* (1955).

18. Bode, H. B. & Müller, R. The impact of bacterial genomics on natural product research. *Angewandte Chemie (International ed. in English)* **44,** 6828–6846; 10.1002/anie.200501080 (2005).

19. Mebs, D. Toxicity in animals. Trends in evolution? *Toxicon* **39,** 87–96; 10.1016/S0041-0101(00)00155-0 (2001).

20. van Santen, J. A. *et al.* The Natural Products Atlas 2.0: a database of microbially-derived natural products. *Nucleic Acids Res.* **50,** D1317-D1323; 10.1093/nar/gkab941 (2022).

21. Terlouw, B. R. *et al.* MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res.* **51,** D603-D610; 10.1093/nar/gkac1049 (2023).

22. van der Lee, T. A. J. & Medema, M. H. Computational strategies for genome-based natural product discovery and engineering in fungi. *Fungal genetics and biology : FG & B* **89,** 29–36; 10.1016/j.fgb.2016.01.006 (2016).

23. Sarker, S. D., Nahar, L., Miron, A. & Guo, M. Anticancer natural products. In *Medicinal natural products. A disease-focused approach / editors, Satyajit Dey Sarker, Lutfun Nahar,* edited by S. D. Sarker & L. Nahar (Academic Press, Amsterdam, 2020), Vol. 55, pp. 45–75.

24. Schatz, A., Bugle, E. & Waksman, S. A. Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria.*. *Experimental Biology and Medicine* **55,** 66–69; 10.3181/00379727-55-14461 (1944).

25. Di Marco, A., Gaetani, M., Dorigotti, L., Soldati, M. & Bellini, O. Daunomycin: A New Antibiotic with Antitumor Activity. *Tumori* **49,** 203–217; 10.1177/030089166304900305 (1963).

26. Ribeiro da Cunha, B., Fonseca, L. P. & Calado, C. R. C. Antibiotic Discovery: Where Have We Come from, Where Do We Go? *Antibiotics* **8**; 10.3390/antibiotics8020045 (2019).

27. Hutchings, M. I., Truman, A. W. & Wilkinson, B. Antibiotics: past, present and future. *Curr. Opin. Microbiol.* **51,** 72–80; 10.1016/j.mib.2019.10.008 (2019).

28. Wright, G. D. Opportunities for natural products in 21st century antibiotic discovery. *Nat. Prod. Rep.* **34,** 694–701; 10.1039/c7np00019g (2017).

29. Baltz, R. H. Gifted microbes for genome mining and natural product discovery. *J. Ind. Microbiol. Biotechnol.*; 10.1007/s10295-016-1815-x (2016).

30. Quince, C., Curtis, T. P. & Sloan, W. T. The rational exploration of microbial diversity. *ISME J* **2,** 997–1006; 10.1038/ismej.2008.69 (2008).

31. Findlay, B. L. The chemical ecology of predatory soil bacteria. *ACS Chem. Biol.* **11,** 1502–1510; 10.1021/acschembio.6b00176 (2016).

32. McCauley, E. P. *et al.* Highlights of marine natural products having parallel scaffolds found from marine-derived bacteria, sponges, and tunicates. *J Antibiot (Tokyo)* **73,** 504–525; 10.1038/s41429-020-0330-5 (2020).

33. Demain, A. L. Importance of microbial natural products and the need to revitalize their discovery. *J. Ind. Microbiol. Biotechnol.* **41,** 185–201; 10.1007/s10295-013-1325-z (2014).

34. Montaser, R. & Luesch, H. Marine natural products. A new wave of drugs? *Future Med. Chem.* **3,** 1475–1489; 10.4155/fmc.11.118 (2011).

35. Subramani, R. & Aalbersberg, W. Marine actinomycetes. An ongoing source of novel bioactive metabolites. *Microbiological research* **167,** 571–580; 10.1016/j.micres.2012.06.005 (2012).

36. Vij, R., Hube, B. & Brunke, S. Uncharted territories in the discovery of antifungal and antivirulence natural products from bacteria. *Computational and Structural Biotechnology Journal* **19,** 1244–1252; 10.1016/j.csbj.2021.02.003 (2021).

37. Herrmann, J., Fayad, A. A. & Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **34,** 135–160; 10.1039/C6NP00106H (2017).

38. Bader, C. D., Panter, F. & Müller, R. In depth natural product discovery - Myxobacterial strains that provided multiple secondary metabolites. *Biotechnol. Adv.* **39,** 107480; 10.1016/j.biotechadv.2019.107480 (2020).

39. Velicer, G. J. & Vos, M. Sociobiology of the myxobacteria. *Annu. Rev. Microbiol.* **63,** 599–623; 10.1146/annurev.micro.091208.073158 (2009).

40. Berleman, J. E. & Kirby, J. R. Deciphering the hunting strategy of a bacterial wolfpack. *FEMS Microbiol. Rev.* **33,** 942–957; 10.1111/j.1574-6976.2009.00185.x (2009).

41. Wenzel, S. C. & Müller, R. Myxobacteria - unique microbial secondary metabolite factories. In *Comprehensive Natural Products Chemistry II, Vol 2: Structural Diversity II - Secondary Metabolite Sources, Evolution and Selected Molecular Structures,* edited by B. Moore (Elsevier, Oxford, 2010), pp. 189–222.

42. Garcia, R., Gemperlein, K. & Müller, R. Minicystis rosea gen. nov., sp. nov., a polyunsaturated fatty acid-rich and steroid-producing soil myxobacterium. *International Journal of Systematic and Evolutionary Microbiology* **64,** 3733–3742; 10.1099/ijs.0.068270-0 (2014).

43. Wenzel, S. C. & Müller, R. Myxobacteria—''microbial factories' for the production of bioactive secondary metabolites. *Mol. Biosyst.* **5,** 567–574; 10.1039/b901287g (2009).

44. Baltz, R. H. Molecular beacons to identify gifted microbes for genome mining. *J Antibiot (Tokyo)* **70,** 639–646; 10.1038/ja.2017.1 (2017).

45. Grabowski, K. & Schneider, G. Properties and Architecture of Drugs and Natural Products Revisited. *Current Chemical Biology* **1,** 115–127; 10.2174/187231307779814066 (2007).

46. Bon, R. S. & Waldmann, H. Bioactivity-guided navigation of chemical space. *Acc. Chem. Res.* **43,** 1103–1114; 10.1021/ar100014h (2010).

47. Newman, D. J., Cragg, G. M. & Kingston, D. G. *Natural Products as Pharmaceuticals and Sources for Lead Structures,* 101–139 (Academic Press, 2015).

48. Yang, Y.-J. *et al.* Genome Editing in Model Strain *Myxococcus xanthus* DK1622 by a Site-Specific Cre/loxP Recombination System. *Biomolecules* **8,** 137; 10.3390/biom8040137 (2018).

49. Gregory, K., Salvador, L. A., Akbar, S., Adaikpoh, B. I. & Stevens, D. C. Survey of Biosynthetic Gene Clusters from Sequenced Myxobacteria Reveals Unexplored Biosynthetic Potential. *Microorganisms* **7,** 181; 10.3390/microorganisms7060181 (2019).

50. Gavriilidou, A. *et al.* Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nat. Microbiol.* **7,** 726–735; 10.1038/s41564-022-01110-2 (2022).

51. Hoffmann, T. *et al.* Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. *Nat. Commun.* **9,** 803; 10.1038/s41467-018-03184-1 (2018).

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

53. Baumann, S. *et al.* Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *Angew. Chem. Int. Ed.* **53,** 14605–14609; 10.1002/anie.201409964 (2014).

54. Barbier, J. *et al.* Isolation and total synthesis of icumazoles and noricumazoles—antifungal antibiotics and cation-channel blockers from *Sorangium cellulosum*. *Angew. Chem. Int. Ed. Engl.* **51,** 1256–1260; 10.1002/anie.201106435 (2012).

55. Reichenbach, H., Höfle, G., Böhlendorf, B. & Irschik, H. *Icumazoles and preparation processes* (1994).

56. Gerth, K., Steinmetz, H., Höfle, G. & Jansen, R. Chlorotonil A, a macrolide with a unique gem-dichloro-1,3-dione functionality from *Sorangium cellulosum*, So ce1525. *Angew. Chem. Int. Ed. Engl.* **47,** 600–602; 10.1002/anie.200703993 (2008).

57. Held, J. *et al.* Antimalarial activity of the myxobacterial macrolide chlorotonil a. *Antimicrobial agents and chemotherapy* **58,** 6378–6384; 10.1128/AAC.03326-14 (2014).

58. Bader, C. D. *et al.* Sandacrabins - Structurally Unique Antiviral RNA Polymerase Inhibitors from a Rare Myxobacterium. *Chemistry (Weinheim an der Bergstrasse, Germany)* **28,** e202104484; 10.1002/chem.202104484 (2022).

59. Schmitz, A. *et al.* Corallorazines from the myxobacterium *Corallococcus coralloides*. *J. Nat. Prod.* **77,** 159–163; 10.1021/np400740u (2014).

60. Gao, Y., Walt, C., Bader, C. D. & Müller, R. Genome-Guided Discovery of the Myxobacterial Thiolactone-Containing Sorangibactins. *ACS Chem. Biol.* **18,** 924–932; 10.1021/acschembio.3c00063 (2023).

61. Etzbach, L., Plaza, A., Garcia, R., Baumann, S. & Müller, R. Cystomanamides: structure and biosynthetic pathway of a family of glycosylated lipopeptides from myxobacteria. *Organic letters* **16,** 2414–2417; 10.1021/ol500779s (2014).

62. Hug, J. J. *et al.* Genome-Guided Discovery of the First Myxobacterial Biarylitide Myxarylin Reveals Distinct C–N Biaryl Crosslinking in RiPP Biosynthesis. *Molecules* **26,** 7483; 10.3390/molecules26247483 (2021).

63. Bollag, D. M. *et al.* Epothilones, a new class of microtubule-stabilizing agents with a taxol-like mechanism of action. *Cancer Res.* **55,** 2325–2333 (1995).

64. Rivera, E., Lee, J. & Davies, A. Clinical development of ixabepilone and other epothilones in patients with advanced solid tumors. *Oncologist* **13,** 1207–1223; 10.1634/theoncologist.2008-0143 (2008).

65. Süssmuth, R. D. & Mainz, A. Nonribosomal peptide synthesis - Principles and prospects. *Angew. Chem. Int. Ed.* **56,** 3770–3821; 10.1002/anie.201609079 (2017).

66. Pogorevc, D. *et al.* Biosynthesis and Heterologous Production of Argyrins. *ACS Synth. Biol.* **8,** 1121–1133; 10.1021/acssynbio.9b00023 (2019).

67. Li, Y., Weissman, K. J. & Müller, R. Myxochelin biosynthesis: direct evidence for two- and four-electron reduction of a carrier protein-bound thioester. *Journal of the American Chemical Society* **130,** 7554–7555; 10.1021/ja8025278 (2008).

68. Trauger, J. W., Kohli, R. M. & Walsh, C. T. Cyclization of backbone-substituted peptides catalyzed by the thioesterase domain from the tyrocidine nonribosomal peptide synthetase. *Biochemistry* **40,** 7092–7098; 10.1021/bi010035r (2001).

69. Miller, B. R. & Gulick, A. M. Structural Biology of Nonribosomal Peptide Synthetases. *Methods Mol. Biol.* **1401,** 3–29; 10.1007/978-1-4939-3375-4_1 (2016).

70. Weissman, K. J. Chapter 1 Introduction to Polyketide Biosynthesis. In *Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates* (Elsevier2009), Vol. 459, pp. 3–16.

71. Keatinge-Clay, A. T. The structures of type I polyketide synthases. *Nat. Prod. Rep.* **29,** 1050–1073; 10.1039/c2np20019h (2012).

72. Dutta, S. *et al.* Structure of a modular polyketide synthase. *Nature* **510,** 512–517; 10.1038/nature13423 (2014).

73. Kopp, M., Irschik, H., Pradella, S. & Müller, R. Production of the tubulin destabilizer disorazol in *Sorangium cellulosum*: biosynthetic machinery and regulatory genes. *ChemBioChem* **6,** 1277–1286; 10.1002/cbic.200400459 (2005).

74. Bradley S. Moore & Christian Hertweck. Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Nat. Prod. Rep.* **19,** 70–99; 10.1039/B003939J (2002).

75. Chan, Y. A., Podevels, A. M., Kevany, B. M. & Thomas, M. G. Biosynthesis of polyketide synthase extender units. *Nat. Prod. Rep.* **26,** 90–114; 10.1039/b801658p (2009).

76. Irschik, H. *et al.* 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* **11,** 1840–1849; 10.1002/cbic.201000313 (2010).

77. Carvalho, R., Reid, R., Viswanathan, N., Gramajo, H. & Julien, B. The biosynthetic genes for disorazoles, potent cytotoxic compounds that disrupt microtubule formation. *Gene* **359,** 91–98; 10.1016/j.gene.2005.06.003 (2005).

78. Helfrich, E. J. N. & Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **33,** 231–316; 10.1039/c5np00125k (2016).

79. Du, L., Sánchez, C. & Shen, B. Hybrid Peptide–Polyketide Natural Products: Biosynthesis and Prospects toward Engineering Novel Molecules. *Metabolic engineering* **3,** 78–95; 10.1006/mben.2000.0171 (2001).

80. Reichenbach, H. & Höfle, G. Biologically active secondary metabolites from myxobacteria. *Biotechnol Adv* **11,** 219–277 (1993).

81. 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.* **40,** 7–13 (1987).

82. Jansen, R., Wray, V., Irschik, H., Reichenbach, H. & Höfle, G. Isolation and spectroscopic structure elucidation of sorangicin A, a new type of macrolide-polyether antibiotic from gliding bacteria - XXX. *Tetrahedron Lett.* **26** (1985).

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

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

85. Zander, W. *et al.* Sulfangolids, macrolide sulfate esters from *Sorangium cellulosum*. *Chem. Eur. J.* **18,** 6264–6271; 10.1002/chem.201100851 (2012).

86. Höfle, G. Isolation, structural elucidation and chemistry. In *Scientific Annual Report,* edited by mbH,Gesellschaft für Biotechnologische Forschung (Braunschweig, 1996), pp. 109–113.

87. Gerth, K., Bedorf, N., Hofle, G., Irschik, H. & Reichenbach, H. Epothilons A and B: antifungal and cytotoxic compounds from *Sorangium cellulosum* (Myxobacteria). Production, physico-chemical and biological properties. *J. Antibiot. (Tokyo)* **49,** 560–563 (1996).

88. Irschik, H., Jansen, R., Höfle, G., Gerth, K. & Reichenbach, H. The corallopyronins, new inhibitors of bacterial RNA synthesis from Myxobacteria. *J. Antibiot.* **38,** 145–152 (1985).

89. Schiefer, A. *et al.* Corallopyronin A for short-course anti-wolbachial, macrofilaricidal treatment of filarial infections. *PLoS neglected tropical diseases* **14,** e0008930; 10.1371/journal.pntd.0008930 (2020).

90. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* **269,** 496–512 (1995).

91. Schneiker, S. *et al.* Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat. Biotechnol.* **25,** 1281–1289; 10.1038/nbt1354 (2007).

92. Slatko, B. E., Gardner, A. F. & Ausubel, F. M. Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology* **122**; 10.1002/cpmb.59 (2018).

93. Zaburannyi, N., Bunk, B., Maier, J., Overmann, J. & Müller, R. Genome analysis of the fruiting body forming myxobacterium *Chondromyces crocatus r*eveals high potential for natural product Biosynthesis. *Appl. Environ. Microbiol.* **82,** 1945–1957; 10.1128/AEM.03011-15 (2016).

94. Kenshole, E., Herisse, M., Michael, M. & Pidot, S. J. Natural product discovery through microbial genome mining. *Curr. Opin. Chem. Biol.* **60,** 47–54; 10.1016/j.cbpa.2020.07.010 (2020).

95. Blin, K. *et al.* antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.*; 10.1093/nar/gkad344 (2023).

96. Skinnider, M. A., Merwin, N. J., Johnston, C. W. & Magarvey, N. A. PRISM 3. Expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Res.* **45,** W49-W54; 10.1093/nar/gkx320 (2017).

97. Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6,** 493–505; 10.1016/S1074-5521(99)80082-9 (1999).

98. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410; 10.1016/S0022-2836(05)80360-2 (1990).

99. Panter, F., Bader, C. D. & Müller, R. Synergizing the potential of bacterial genomics and metabolomics to find novel antibiotics. *Chem. Sci.,* 5994–6010; 10.1039/D0SC06919A (2021).

100. Bauman, K. D., Butler, K. S., Moore, B. S. & Chekan, J. R. Genome mining methods to discover bioactive natural products. *Nat. Prod. Rep.* **38,** 2100–2129; 10.1039/D1NP00032B (2021).

101. Zerikly, M. & Challis, G. L. Strategies for the discovery of new natural products by genome mining. *ChemBioChem* **10,** 625–632; 10.1002/cbic.200800389 (2009).

102. Viehrig, K. *et al.* Concerted action of P450 plus helper protein to form the amino-hydroxy-piperidone moiety of the potent protease inhibitor crocapeptin. *J. Am. Chem. Soc.* **135,** 16885–16894; 10.1021/ja4047153 (2013).

103. 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.* **46,** 319–334; 10.1007/s10295-018-2105-6 (2019).

104. Panter, F., Krug, D., Baumann, S. & Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **9,** 4898–4908; 10.1039/C8SC01325J (2018).

105. Sticher, O. Natural product isolation. *Nat. Prod. Rep.* **25,** 517–554; 10.1039/b700306b (2008).

106. 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.*; 10.1021/acs.jnatprod.0c00942 (2021).

107. 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.* **31,** 768–783; 10.1039/c3np70127a (2014).

108. Burchard, R. P., Burchard, A. C. & Parish, J. H. Pigmentation phenotype instability in Myxococcus xanthus. *Can. J. Microbiol.* **23,** 1657–1662; 10.1139/m77-238 (1977).

109. 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.* **103,** 19128–19133; 10.1073/pnas.0606039103 (2006).

110. Hug, J. J., Bader, C. D., Remškar, M., Cirnski, K. & Müller, R. Concepts and Methods to Access Novel Antibiotics from Actinomycetes. *Antibiotics* **7,** 44; 10.3390/antibiotics7020044 (2018).

111. Kunze, B., Höfle, G. & Reichenbach, H. The aurachins, new quinoline antibiotics from myxobacteria. Production, physico-chemical and biological properties. *J Antibiot (Tokyo)* **40,** 258–265; 10.7164/antibiotics.40.258 (1987).

112. Wang, D.-G. *et al.* Constructing a Myxobacterial Natural Product Database to Facilitate NMR-Based Metabolomics Bioprospecting of Myxobacteria. *Analytical chemistry* **95,** 5256–5266; 10.1021/acs.analchem.2c05145 (2023).

113. Schlotterbeck, G. & Ceccarelli, S. M. LC–SPE–NMR–MS: a total analysis system for bioanalysis. *Bioanalysis* **1,** 549–559; 10.4155/bio.09.50 (2009).

114. Johansen, K. T., Wubshet, S. G., Nyberg, N. T. & Jaroszewski, J. W. From retrospective assessment to prospective decisions in natural product isolation: HPLC-SPE-NMR analysis of *Carthamus oxyacantha*. *J. Nat. Prod.* **74,** 2454–2461; 10.1021/np200780m (2011).

115. Chen, F., Wang, M. & Cheng, K.-W. Liquid Chromatography-Mass Spectrometry in Natural Product Research. In *Bioactive natural products. Detection, isolation, and structural determination / edited by Steven M. Colegate, Russell J. Molyneux,* edited by S. M. Colegate & R. J. Molyneux. 2nd ed. (CRC Press, Boca Raton, 2008), pp. 245–2651.

116. Ito, T. & Masubuchi, M. Dereplication of microbial extracts and related analytical technologies. *J Antibiot* **67,** 353–360; 10.1038/ja.2014.12 (2014).

117. Hoffmann, T., Krug, D., Hüttel, S. & Müller, R. Improving natural products identification through targeted LC-MS/MS in an untargeted secondary metabolomics workflow. *Anal. Chem.* **86,** 10780–10788; 10.1021/ac502805w (2014).

118. Nothias, L.-F. *et al.* Feature-based molecular networking in the GNPS analysis environment. *Nat Methods* **17,** 905–908; 10.1038/s41592-020-0933-6 (2020).

119. Phelan, V. V. Feature-Based Molecular Networking for Metabolite Annotation. *Methods Mol. Biol.* **2104,** 227–243; 10.1007/978-1-0716-0239-3_13 (2020).

120. Dancik, V., Addona, T. A., Clauser, K. R., Vath, J. E. & Pevzner, P. A. De novo peptide sequencing via tandem mass spectrometry. *J. Comput. Biol.* **6,** 327–342 (1999).

121. Nett, M., Ikeda, H. & Moore, B. S. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat. Prod. Rep.* **26,** 1362–1384; 10.1039/b817069j (2009).

122. Bode, H. B., Bethe, B., Höfs, R. & Zeeck, A. Big effects from Small Changes: Possible Ways to Explore Nature's Chemical Diversity. *ChemBioChem* **3,** 619–627 (2002).

123. Covington, B. C., Xu, F. & Seyedsayamdost, M. R. A Natural Product Chemist's Guide to Unlocking Silent Biosynthetic Gene Clusters. *Annu. Rev. Biochem.* **90,** 763–788; 10.1146/annurev-biochem-081420-102432 (2021).

124. Schiewe, H. J. Untersuchungen zur Biosynthesesleistung ausgewählter Streptomycetenstämme durch ökomimetische Variation der Fermentationsparameter und Strukturaufklärung der isolierten Sekundärmetabolite. Georg-August-Universität Göttingen, 1997.

125. Bethe, B. Ausarbeitung von Methoden und Konzepten zum Auffinden und Bewerten von Biosyntheseleistungen im Sekundarstoffwechsel sowie Isolierung und Strukturaufklarung der Sekundarstoffe des Streptomyceten Stammes Go 40/14. PhD thesis, 1994.

126. Schiewe, H. J. & Zeeck, A. Cineromycins, *γ*-Butyrolactones and Ansamycins by Analysis of the Secondary Metabolite Pattern Created by a Single Strain of *Streptomyces*. *J Antibiot* **52,** 635–642; 10.7164/antibiotics.52.635 (1999).

127. Machushynets, N. V., Wu, C., Elsayed, S. S., Hankemeier, T. & van Wezel, G. P. Discovery of novel glycerolated quinazolinones from Streptomyces sp. MBT27. *J. Ind. Microbiol. Biotechnol.* **46,** 483–492; 10.1007/s10295-019-02140-2 (2019).

128. Gibson, A. M., Bratchell, N. & Roberts, T. A. Predicting microbial growth: growth responses of salmonellae in a laboratory medium as affected by pH, sodium chloride and storage temperature. *International journal of food microbiology* **6,** 155–178; 10.1016/0168-1605(88)90051-7 (1988).

129. Gallardo, K., Candia, J. E., Remonsellez, F., Escudero, L. V. & Demergasso, C. S. The Ecological Coherence of Temperature and Salinity Tolerance Interaction and Pigmentation in a Non-marine Vibrio Isolated from Salar de Atacama. *Front. Microbiol.* **7,** 1943; 10.3389/fmicb.2016.01943 (2016).

130. Walt, C., Bader, C. D., Krug, D. & Müller, R. Aggravated Cultivation of Myxobacteria Stimulates Secondary Metabolism. *(Manuscript in preparation)*.

131. Schwarz, J., Hubmann, G., Rosenthal, K. & Lütz, S. Triaging of Culture Conditions for Enhanced Secondary Metabolite Diversity from Different Bacteria. *Biomolecules* **11**; 10.3390/biom11020193 (2021).

132. Jenny Schwarz, Georg Hubmann, Ayla Schwarz, Katrin Rosenthal & Stephan Lütz. Bivariate OSMAC Designs Expand the Secondary Metabolite Production Space in Corallococcus Coralloides; 10.20944/preprints202203.0118.v1 (2022).

133. Bouslimani, A., Sanchez, L. M., Garg, N. & Dorrestein, P. C. Mass spectrometry of natural products: current, emerging and future technologies. *Nat. Prod. Rep.* **31,** 718–729; 10.1039/c4np00044g (2014).

134. Kind, T. & Fiehn, O. Advances in structure elucidation of small molecules using mass spectrometry. *Bioanal Rev.* **2,** 23–60; 10.1007/s12566-010-0015-9 (2010).

135. Nicolaou, K. C. & Snyder, S. A. Chasing molecules that were never there: misassigned natural products and the role of chemical synthesis in modern structure elucidation. *Angew. Chem. Int. Ed.* **44,** 1012–1044; 10.1002/anie.200460864 (2005).

136. Molinski, T. F. NMR of natural products at the 'nanomole-scale'. *Nat. Prod. Rep.* **27,** 321–329; 10.1039/b920545b (2010).

137. Holzgrabe, U., Wawer, I. & Diehl, B. *NMR Spectroscopy in Pharmaceutical Analysis* (Elsevier Science, Burlington, 2011).

138. Molinski, T. F. Microscale methodology for structure elucidation of natural products. *Curr. Opin. Biotechnol.* **21,** 819–826; 10.1016/j.copbio.2010.09.003 (2010).

139. Bross-Walch, N., Kühn, T., Moskau, D. & Zerbe, O. Strategies and Tools for Structure Determination of Natural Products Using Modern Methods of NMR Spectroscopy. *Biochemistry and Biodiversity* (2005).

140. Berger, S. & Braun, S. *200 and more NMR experiments* (Wiley-VCH, Weinheim, 2004).

141. Martin, G. E. Chapter 5 - Using 1,1- and 1,n-ADEQUATE 2D NMR Data in Structure Elucidation Protocols. In *Annual Reports on NMR Spectroscopy,* edited by G. A. Webb (Academic Press2011), Vol. 74, pp. 215–291.

142. Reif, B. *et al.* ADEQUATE, a New Set of Experiments to Determine the Constitution of Small Molecules at Natural Abundance. *Journal of Magnetic Resonance, Series A* **118,** 282–285; 10.1006/jmra.1996.0038 (1996).

143. 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. *Carbohydrate research* **343,** 101–112; 10.1016/j.carres.2007.10.008 (2008).

144. Gao, Y. *et al.* The Disorazole Z Family of Highly Potent Anticancer Natural Products from Sorangium cellulosum: Structure, Bioactivity, Biosynthesis, and Heterologous Expression. *Microbiol Spectr,* e0073023; 10.1128/spectrum.00730-23 (2023).

145. BERNAL, J. D. Crystal Structures of Vitamin D and Related Compounds. *Nature* **129,** 277–278; 10.1038/129277a0 (1932).

146. Jansen, R. *et al.* Antibiotika aus Gleitenden Bakterien, XXXVII. Sorangicin A, ein hochwirksames Antibiotikum mit neuartiger Makrolid-Polyether-Struktur aus *Sorangium cellulosum*, So ce12: Spektroskopische Strukturaufklärung, Kristall- und Lösungsstruktur. *Liebigs Ann. Chem.* **1989,** 111–119; 10.1002/jlac.198919890124 (1989).

147. Pescitelli, G., Kurtán, T., Flörke, U. & Krohn, K. Absolute structural elucidation of natural products—a focus on quantum-mechanical calculations of solid-state CD spectra. *Chirality* **21 Suppl 1,** E181-201; 10.1002/chir.20795 (2009).

148. Zeng, H. *et al.* Expanding the Ajudazol Cytotoxin Scaffold: Insights from Genome Mining, Biosynthetic Investigations, and Novel Derivatives. *Journal of natural products*; 10.1021/acs.jnatprod.2c00637 (2022).

149. Marfey, P. Determination of D-amino acids. II. Use of a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzene. *Carlsberg Res. Commun.* **49,** 591–596; 10.1007/BF02908688 (1984).

150. B'Hymer, C., Montes-Bayon, M. & Caruso, J. A. Marfey's reagent: past, present, and future uses of 1-fluoro-2,4-dinitrophenyl-5-L-alanine amide. *J. Sep. Sci.* **26,** 7–19 (2003).

151. Gerwig, G. J., Kamerling, J. P. & Vliegenthart, J. F. Determination of the D and L configuration of neutral monosaccharides by high-resolution capillary G.L.C, 349–357 (1978).

152. Basconcillo, L. S. & McCarry, B. E. Comparison of three GC/MS methodologies for the analysis of fatty acids in Sinorhizobium meliloti. Development of a micro-scale, one-vial method. *J. Chromatogr. B* **871,** 22–31; 10.1016/j.jchromb.2008.06.041 (2008).

153. Hoye, T. R., Jeffrey, C. S. & Shao, F. Mosher ester analysis for the determination of absolute configuration of stereogenic (chiral) carbinol carbons. *Nat. Protoc.* **2,** 2451–2458; 10.1038/nprot.2007.354 (2007).

2. Disorazole Z

The Disorazole Z Family of Highly Potent Anticancer Natural Products from *Sorangium cellulosum*: Structure, Bioactivity, Biosynthesis, and Heterologous Expression

Previously published in:

Microbiology Spectrum **2023** 11 (4), e00730-23

DOI: 10.1128/spectrum.00730-23

Yunsheng Gao†[a,c,d], Joy Birkelbach†[a], Chengzhang Fu[a,d], Jennifer Herrmann[a], Herbert Irschik[b], Bernd Morgenstern[e], Kerstin Hirschfelder[a], Ruijuan Li[c], Youming Zhang[c], Rolf Jansen[b], Rolf Müller[a,d]

[a]Department of Microbial Natural Products, Helmholtz-Institute for Pharmaceutical Research Saarland, Helmholtz Centre for Infection Research and Department of Pharmacy at Saarland University, Saarbrücken, Germany

[b]Department of Microbial Drugs, Helmholtz Centre for Infection Research, Braunschweig, Germany

[c]Helmholtz International Lab for Anti-Infectives, Shandong University-Helmholtz Institute of Biotechnology, State Key Laboratory of Microbial Technology, Shandong University, Qingdao, China

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

[e]Department of Inorganic Chemistry, Saarland University, Saarbrücken, Germany

Contributions and Acknowledgements

The Author's Effort

This author significantly contributed to the concept of this study, designed and performed experiments, and interpreted results. NMR based structure elucidation of disorazole Z1, Z9 and Z10 derived from heterologous expression was conducted by the autor. Furthermore, the author organised and verified the structure elucidation data of all congeners, and wrote the structure elucidation text. Moreover, the author contributed by conception and writing this manuscript.

Contributions by Others

Yunsheng Gao identified, characterised, and heterologously expressed the disorazole Z gene cluster, as well as performed protein expression and *in vitro* reactions. He performed compound purification, metabolome and quantification analysis, and contributed to conceiving and writing this manuscript. Chengzhang Fu contributed to biosynthesis analysis and draft preparation. Jennifer Herrmann and Kerstin Hirschfelder contributed to the biological evaluation section. Herbert Irschik and Rolf Jansen contributed to the production, isolation and characterisation of disorazole Z from *Sorangium cellulosum*, as well as draft preparation. Bernd Morgenstern contributed to crystallographic analysis. Ruijuan Li and Youming Zhang contributed to supervision of the research. Rolf Müller and Rolf Jansen were responsible for the conception and supervision of the project and revised this manuscript.

Acknowledgements

Research in R.M.'s laboratory was partially funded by the Bundesministerium für Bildung und Forschung (BMBF) and Deutsche Forschungsgemeinschaft (DFG). Y.G. thanks China Scholarship Council for a fellowship of the Ph.D. program. The crystallographic analysis, instrumentation and technical assistance for this work were provided by Volker Huch and Bernd Morgenstern at the Service Center X-ray Diffraction, with financial support from Saarland University and German Science Foundation (project number INST 256/506-1). We gratefully acknowledge Marc Stadler, Reinhard Sterlinski, Axel Schulz, Wolfgang Kessler, and Kerstin Schober for their contributions to large-scale fermentation and downstream processing and especially Steffen Bernecker for establishing and monitoring the fermentation process. We declare no conflict of interest.

2.1. Abstract

Myxobacteria serve as a treasure trove of secondary metabolites. During our ongoing search for bioactive natural products, a novel subclass of disorazoles termed disorazole Z was discovered. Ten disorazole Z family members were purified from a large-scale fermentation of the myxobacterium *Sorangium cellulosum* So ce1875 and characterised by electrospray ionisation-high-resolution mass spectrometry (ESI-HR-MS), X-ray, nuclear magnetic resonance (NMR), and Mosher ester analysis. Disorazole Z compounds are characterised by the lack of one polyketide extension cycle, resulting in a shortened monomer in comparison to disorazole A, which finally forms a dimer in the bis-lactone core structure. In addition, an unprecedented modification of a geminal dimethyl group takes place to form a carboxylic acid methyl ester. The main component disorazole Z1 shows comparable activity in effectively killing cancer cells to disorazole A1 via binding to tubulin, which we show induces microtubule depolymerisation, endoplasmic reticulum delocalisation, and eventually apoptosis. The disorazole Z biosynthetic gene cluster (BGC) was identified and characterised from the alternative producer *S. cellulosum* So ce427 and compared to the known disorazole A BGC, followed by heterologous expression in the host *Myxococcus xanthus* DK1622. Pathway engineering by promoter substitution and gene deletion paves the way for detailed biosynthesis studies and efficient heterologous production of disorazole Z congeners.

Importance

Microbial secondary metabolites are a prolific reservoir for the discovery of bioactive compounds, which prove to be privileged scaffolds for the development of new drugs such as antibacterial and small-molecule anticancer drugs. Consequently, the continuous discovery of novel bioactive natural products is of great importance for pharmaceutical research. Myxobacteria, especially *Sorangium* spp., which are known for their large genomes with yet-underexploited biosynthetic potential, are proficient producers of such secondary metabolites. From the fermentation broth of *Sorangium cellulosum* strain So ce1875, we isolated and characterised a family of natural products named disorazole Z, which showed potent anticancer activity. Further, we report on the biosynthesis and heterologous production of disorazole Z. These results can be stepping stones toward pharmaceutical development of the disorazole family of anticancer natural products for (pre)clinical studies.

2.2. Introduction

Myxobacteria are soil-dwelling Gram-negative delta proteobacteria and known as prolific sources of natural products often exhibiting novel chemical scaffolds and unique biological modes of action[1–4]. Disorazoles are a family of macrocyclic dilactones with dimeric or pseudodimeric structures, and they represent one group of myxobacterial secondary metabolites showing promising cytotoxic activity[5,6]. Furthermore, this family of compounds was shown to block the invasion of human epithelial cells by group A streptococci via targeting of the host factor ezrin[7]. Therefore, disorazoles have the potential to be developed as anticancer or anti-infective drugs.

Previously, 29 disorazole A family compounds (disorazole A1 and 28 variants) were isolated from the fermentation culture of the myxobacterium *Sorangium cellulosum* So ce12[8]. The main component, disorazole A1 (compound 1) showed extremely high cytotoxicity, with half-maximal inhibitory concentrations (IC_{50}s) in the range of 2 to 42 pmol/L against a panel of cancer cell lines, including a multidrug-resistant KB line by inhibiting tubulin polymerisation and inducing apoptosis[9]. In 2005, the disorazole A biosynthetic gene cluster from *S. cellulosum* So ce12 (*dis12* gene cluster) was identified by transposon mutagenesis[10,11]. The core biosynthetic gene cluster encodes a multifunctional megasynthetase consisting of polyketide synthases (PKS), a nonribosomal peptide synthetase (NRPS), and a dedicated acyltransferase (AT) and thus belongs to the family of *trans*-AT PKS-NRPS hybrid gene clusters[12,13]. In 2016, the *dis12* gene cluster was subcloned and heterologously expressed in *Myxococcus xanthus* DK1622 leading to the production of disorazole A2 (compound 2) as the major compound[14].

In the course of our ongoing screening for biologically active natural products from myxobacteria, a novel subclass of disorazole, i.e., disorazole Z, was discovered to be produced by several *S. cellulosum* strains, such as So ce1875 and So ce427[15]. Structure elucidation and characterisation of disorazole Z have not been published to date. Nevertheless, and due to its outstanding anticancer activity, disorazole Z was subjected to preclinical evaluation as a cytotoxic component in a drug-targeting approach for the treatment of luteinising hormone-releasing hormone (LHRH) receptor-overexpressing cancers by conjugation with the receptor-targeting moiety D-Lys6-LHRH[16–19]. Compared to doxorubicin-D-Lys6-LHRH, the disorazole Z conjugate demonstrated an increased cytotoxicity *in vitro* in HCC 1806 and MDA-MB-231 triple-negative breast cancer cells[20]. In the meantime, disorazole Z was also identified as a maytansine site ligand, which was confirmed by solving the crystal structure of the tubulin-disorazole Z complex[21].

Due to the novelty of these compounds' structures, their intriguing biological activities, and their promising potential for clinical application, massive efforts were made to develop chemical

synthesis routes. Total synthesis has been achieved for some disorazole congeners and nonnatural analogs, for example, disorazole A1 (compound 1) and also a simplified disorazole Z[22,23]. Nevertheless, biotechnological methods still present significant advantages for obtaining these types of complex natural products[24]. Furthermore, taking advantage of synthetic biotechnology, rational engineering of the biosynthetic pathway in an advantageous heterologous host could also facilitate high-efficiency production of certain target components or generate nonnatural compounds[25,26].

In this work, large-scale fermentation afforded the isolation of 10 disorazole Z family members featuring two-carbon shorter monomers and unprecedented modification of a geminal dimethyl group compared to disorazole A. We report the first full structure elucidation of the disorazole Z congeners using nuclear magnetic resonance (NMR), Mosher ester, and X-ray analysis. Furthermore, we characterized the bioactivity of disorazole Z1 (compound 3). In addition, the disorazole Z biosynthetic pathway was identified by comparative analysis and heterologous expression. Promoter engineering and gene deletion experiments were carried out using the heterologous expression system, which allowed the function of a methyltransferase involved in disorazole Z biosynthesis to be assigned.

2.3. Results and Discussion

Isolation and full structure elucidation of the disorazole Z family of compounds. After large-scale fermentation of *S. cellulosum* So ce1875, we were able to isolate 10 disorazole Z congeners (compounds 3 to 12), achieving a production titer of 60 to 80 mg/L disorazole Z1 (compound 3). An overview of the chemical structures of these compounds is given in **Figure 2.1**. Details of fermentation and purification are provided in Materials and Methods.

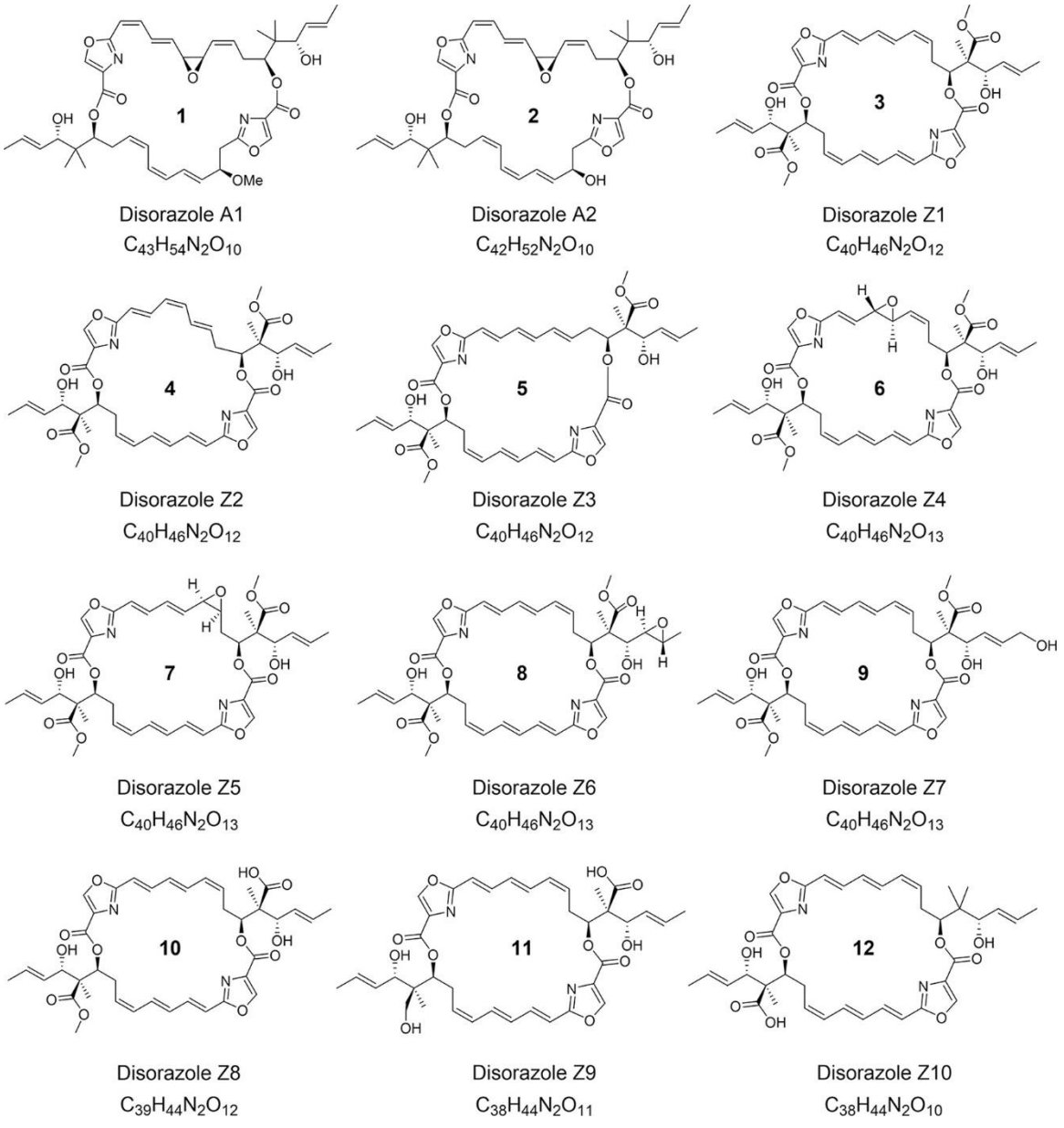


Figure 2.1. Chemical structures of disorazole Z family compounds and disorazole A for comparison.

Molecular ion clusters $[M+H]^+$ (*m/z* 747.3132), $[M+Na]^+$ and $[M-H_2O+H]^+$ in electrospray ionisation-high-resolution mass spectrometry (ESI-HR-MS) revealed the elemental formula $C_{40}H_{46}N_2O_{12}$ of disorazole Z1 (compound 3) from which 19 double-bond equivalents were calculated. Surprisingly, the 13C NMR spectrum showed only half the number of carbon signals solely explicable for a completely symmetrical dimer. The ^1H NMR spectrum in acetone-d_6 presented 23 protons, which were correlated with their corresponding carbons in a $^1H,^{13}C$ heteronuclear single quantum coherence (HSQC) NMR spectrum with the exception of one doublet of a secondary alcohol proton (OH-26) at 4.16 ppm. The $^1H,^1H$ correlation spectroscopy (COSY) NMR spectrum allowed us to assign two main structural parts, A and B (**Figure 2.2**). H-5 of part A and a ^1H singlet (3-H) at 8.53 ppm were connected by a long-range coupling. The corresponding carbon C-3 had no correlation in the heteronuclear multiple-bond correlation (HMBC) NMR spectrum. However, the direct $^1J_{C,H}$ coupling of 213 Hz indicated its role in a hetero-aromatic ring[27]. The further ring carbons C-2 and C-4 were identified by their HMBC correlations with H-3, while the exact connection of the oxazole ring to structural part A was shown by the HMBC correlation between C-4 and H-5. However, H-6 did not show any HMBC correlation, appearing as a flat and broad proton signal. The continuation of structural part A at the opposite oxymethine C-12 (δ_C 76.28) was indicated by an apparent lowfield acylation shift (δ_H 5.45) of the H-12 proton signal that provided the only HMBC correlation of the carboxyl carbon C-13 (δ_C 159.65). By HMBC correlations with H-12, H-26, and the hydroxyl group OH-26, the quaternary sp^3 carbon C-25 (δ_C 56.26) was revealed as the interconnection of parts A and B. Further, the HMBC correlations showed the direct attachment of the methyl group C-30 to C-25. The last open position was filled by a carboxymethyl ester at C-25, as indicated by three HMBC correlations of the carboxyl carbon C-31 (δ_C 173.81) with the methyl group C-30, the alcohol methine H-26, and the acylated oxymethine H-12. The stereochemistry of the *cis* and *trans* double bonds was derived from vicinal coupling constants of about 11 Hz and 15 Hz, respectively. With all NMR structural elements assigned, 18 double-bond equivalents were consumed for two identical structural parts. Using the last equivalent, both halves were connected to give the point symmetrical dilactone ring of compound 3, in which all identical structural elements have the same vicinity.

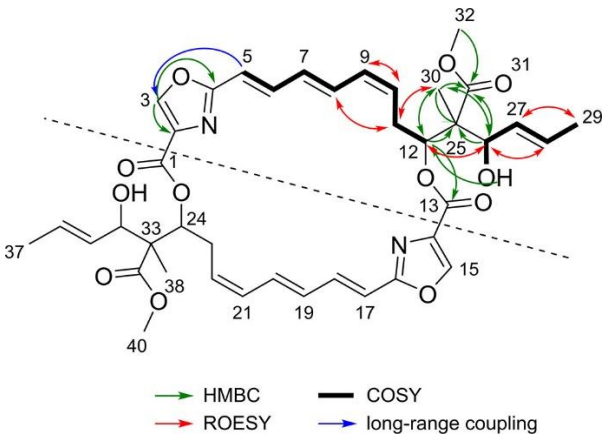


Figure 2.2. NMR correlations of structural parts A and B of disorazole Z1. *^{1}H,^{1}H-COSY correlations (thick bonds) and their interconnection by selected ^{1}H,^{13}C-HMBC correlations (green arrows), ^{1}H,^{1}H-rotating frame Overhauser effect spectroscopy (^{1}H,^{1}H-ROESY) correlations (red arrows) and ^{1}H,^{1}H-long-range coupling (blue arrow). ROESY, rotating-frame nuclear Overhauser effect spectroscopy.*

X-ray analysis revealed the relative configuration of disorazole Z1 (compound 3). In order to elucidate the absolute configuration, Mosher's method based upon ^{1}H and ^{13}C NMR chemical shift differences was applied[28–31]. Consequently, all asymmetric centers (12, 25, 26) of compound 3 are in the *S* configuration, as shown in **Figure 2.1**. A full description of the stereo-chemical analyses is given in the supplemental material. The structural assignment is in full agreement with the structure of the tubulin-disorazole Z complex published recently[21].

Besides disorazole Z1 (compound 3), nine more variants have also been isolated and characterised. Two of them turned out to be isomers of compound 3 with different double bond geometries, i.e., disorazole Z2 (compound 4, $\Delta^{7,8}$-*cis*-disorazole Z) and disorazole Z3 (compound 5, $\Delta^{9,10}$-*trans*-disorazole Z). Four more polar disorazole Z variants were identified as isomers with the ESI-HR-MS-derived elemental formula $C_{40}H_{46}N_2O_{13}$. They contained one additional oxygen, either incorporated as an epoxide group replacing different double bonds, i.e., disorazole Z4 (compound 6, 7,8-epoxy-disorazole Z), disorazole Z5 (compound 7, 9,10-epoxy-disorazole Z), and disorazole Z6 (compound 8, 27,28-epoxy-disorazole Z), or incorporated as primary unsaturated alcohol at the end of a side chain, i.e., disorazole Z7 (compound 9, 29-hydoxy-disorazole Z). Three disorazole Z variants with smaller molecular weight were proposed to be biosynthetic intermediates featuring different degrees of modification on geminal methyl groups, i.e., disorazole Z8 (compound 10, 31-*O*-desmethyl-disorazole Z), disorazole Z9 (compound 11, 31-*O*-desmethyl-39-hydroxy-disorazole Z) and disorazole Z10 (compound 12, 39-*O*-desmethyl-25,25-dimethyl-disorazole Z). Details of the structure elucidation of these analogues are given in the supplemental material.

Discovery and comparative analysis of disorazole Z biosynthetic gene cluster. The significant structural differences between the disorazole Z and A family compounds motivated us to investigate the disorazole Z biosynthetic pathway. The draft genome sequence of the alternative producer strain *S. cellulosum* So ce427 was obtained by Illumina sequencing and subjected to antiSMASH analysis for annotation of biosynthetic gene clusters[32]. A *trans*-AT PKS-NRPS gene cluster (*dis*$_{427}$ gene cluster; GenBank accession number OQ408282) which exhibited significant overall similarity to the known *dis12* gene cluster (GenBank accession number DQ013294 or AJ874112) was speculated to be responsible for the biosynthesis of disorazole Z (**Figure 2.3**). Intriguingly, this gene cluster has two main distinguishing features: on the one hand, one PKS module is lacking in the second polyketide synthase gene, *disB*$_{427}$, which corresponds nicely with the two-carbon shortened monomeric unit of disorazole Z compared to one half-side of the disorazole A congener; on the other hand, the methyltransferase gene *disF*$_{427}$ was found downstream of *disD*$_{427}$, which is not present in the *dis12* gene cluster and could be assigned to the *O*-methylation function needed for the methylation of the carboxyl group found in disorazole Z. The carboxyl group most likely arises from the oxidation of one of the methyl groups at positions C-25 and C-33. However, no gene flanking the corresponding biosynthetic gene cluster (BGC) encoding an expected oxidative function could be assigned for this hypothetical step. The organisation of the remaining part of the disorazole Z pathway is quite similar to that of the disorazole A pathway. The biosynthesis of the half-side of the disorazole Z dilactone begins with condensing one malonyl coenzyme A (malonyl-CoA) with the starter acetate. After five additional rounds of extension with malonyl-CoA, the ketosynthase (KS) domain in module 7 is not expected to extend the polyketide chain, because the catalytic histidine of KS7 is substituted with alanine (Fig. S7). The mutation of the conserved motif C-H-H to C-A-H most likely causes malfunction of this KS domain, as the histidine residue is known to play an essential role in decarboxylation and condensation reactions[33]. However, this type of KS was proven to be a gatekeeper and still capable of transferring the polyketide chain between different domains in FR901464 biosynthesis[34]. The adenylation domain of DisC$_{427}$ activates serine, also found in the disorazole A pathway. Similar tandem heterocyclisation (HC) domains were proven to be essential for vibriobactin and anguibactin biosynthesis[35,36]. It is assumed likely that the HC domains of DisC$_{427}$ work in the same fashion. However, it is also possible that one of the HC domains condenses the serine with the polyketide chain, and the other HC domain might perform the cyclisation of the serine moiety to form the oxazoline ring, which is finally oxidised to its oxazole form by the oxidation domain. The extension of the nascent intermediate bound to the carrier protein of module 9 is supposed to stop here, because the two following PKS modules are most likely nonextending ones, which are

suggested by annotation to be noncondensing KS9 and KS10 domains similar to KS7 (see Fig. S7 in the supplemental material). Finally, the termination of the assembly line and the cyclisation of the two monomeric subunits are likely to be similar to the mechanism described for enterobactin, elaiophylin, and conglobatin biosynthesis[37–39], eventually forming the bis-lactone structure. The carboxylic acid methyl ester in disorazole Z might be introduced before or after release from the assembly line by stereospecific oxidation of a methyl group, giving rise to the free carboxylic acid and subsequent methyl ester formation. Nevertheless, the oxidation process involved remains enigmatic and needs further study. Intriguingly, after expression of the *dis427* gene cluster, disorazole Z1 is also formed, being the major heterologous product in the host *M. xanthus* DK1622 (see below), which indicates that the missing functionalities are encoded within either the BGC or the heterologous host to harbor similar genomic functions capable of oxidising the precursor. The final step of methyl ester formation could be connected to DisF, encoded in the corresponding BGC.

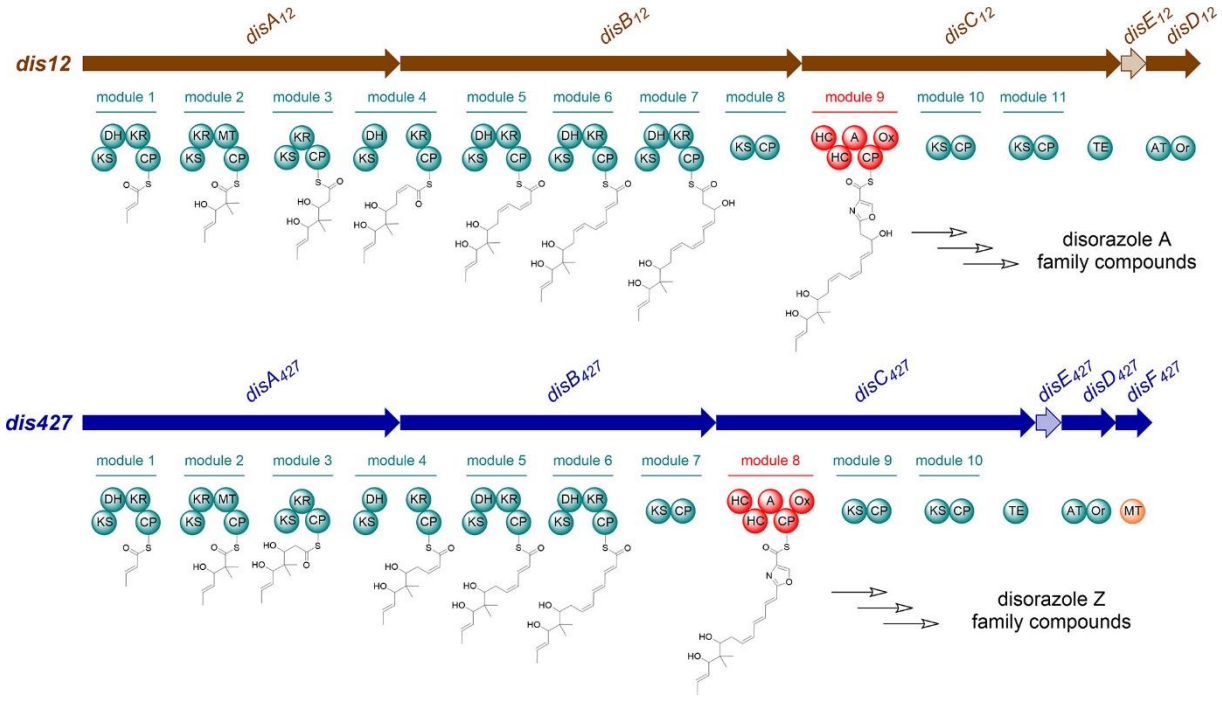


Figure 2.3. Comparison of proposed biosynthetic pathways between disorazole A and Z. dis12*, the disorazole A biosynthetic gene cluster from* S. cellulosum *So ce12;* dis427*, the disorazole Z biosynthetic gene cluster from* S. cellulosum *So ce427; KS, ketosynthase; DH, dehydratase; KR, ketoreductase; CP, carrier protein; MT, methyltransferase; HC, heterocyclisation domain; A, adenylation domain; Ox, oxidation domain; TE, thioesterase; AT, acyl transferase; Or, oxidoreductase. Oxidation of the methyl group and formation of the methyl ester may occur prior to or after release of the bis-lactone.*

Heterologous production and biosynthesis of disorazole Z. Since *Sorangium cellulosum* strains are slow growing and unsuitable for in situ genetic manipulation, we cloned the disorazole Z biosynthetic gene cluster (from $disA_{427}$ to $disF_{427}$) from the genomic DNA of So ce427 for heterologous expression and functional verification. Onestep capture of large gene clusters is usually challenging, especially for myxobacteria, due to their complex genomes. Therefore, the gene cluster was divided into three smaller parts for cloning and then assembled into a p15A-cm vector using RecET-mediated linear-linear homologous recombination (LLHR) (Fig. S9)[40]. After that, the chloramphenicol resistance gene was replaced by a *km-int* cassette coding for phage Mx8 integrase, which could be employed for site-specific integration of the gene cluster into the genome of the heterologous host *M. xanthus* DK1622 (41)[41]. The *dis427* gene cluster under the control of its native promoters was successfully expressed in *M. xanthus* DK1622, leading to the production of disorazole Z1 (compound 3) as the primary compound, which was confirmed by HPLC-MS and NMR analysis (**Figure 2.4**A; Table S3; Fig. S17 and S18). However, the production yield is only about 0.2 mg/L in the standard fermentation procedure (see the supplemental material). Replacement of the native promoter before $disA_{427}$ with a tetracycline-inducible P*tet* promoter resulted in at least a 4-fold increase in heterologous production of compound 3 (Fig. S10). Furthermore, inserting a strong constitutive P*apr* promoter before $disD_{427}$ led to a further improved yield but not by much. When the gene cluster was under the control of a vanillate-inducible Pvan promoter, a nearly 9-fold increase was achieved (Fig. S10). Nevertheless, the production yield of compound 3 (about 1.8 mg/L) is still much lower than that obtained by using the original producer strain under optimised fermentation conditions, which may be due to the low level of phylogenetic similarity between *Sorangium* and *Myxococcus*. Further systematic genetic engineering to improve the yield using heterologous expression is therefore still required to achieve competitiveness. Nevertheless, this system, for the first time, allows for genetic manipulation of the disorazole Z biosynthesis, as the native host was found to be genetically intractable despite significant efforts.

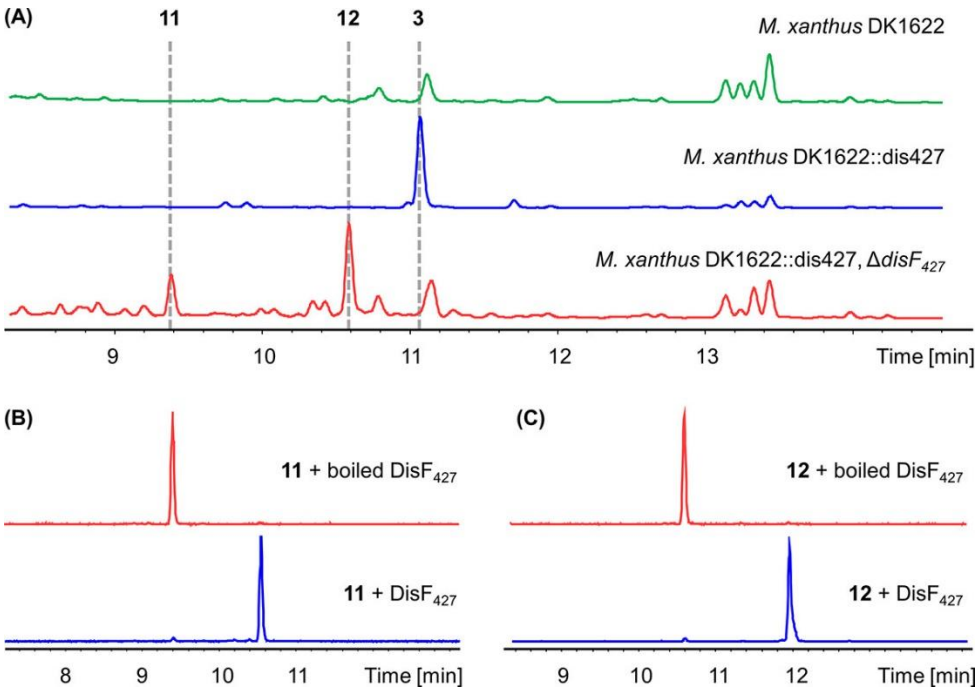


***Figure 2.4. HPLC-MS analysis of disorazole Z formation in M. xanthus and* in vitro.** *(A) HPLC chromatogram analysed by UV detection at 320 nm. Blue trace, heterologous production of disorazole Z1 (compound 3) in M. xanthus DK1622 mutant harboring the* dis427 *gene cluster. Red trace, heterologous production of disorazole Z9 (compound 11) and Z10 (compound 12) after deletion of* $disF_{427}$*. Green trace, M. xanthus DK1622 wild type. (B)* In vitro *reaction of compound 11 with* $DisF_{427}$ *(blue trace) and boiled* $DisF_{427}$ *(red trace), shown as extracted ion chromatograms (EIC) of* m/z *689.31 ± 0.05 and* m/z *703.31 ± 0.05. (C)* In vitro *reaction of 12 with* $DisF_{427}$ *(blue trace) and boiled* $DisF_{427}$ *(red trace), shown as EIC of* m/z *705.30 ± 0.05 and* m/z *719.30 ± 0.05.*

As mentioned above, the methyltransferase gene was found exclusively in type Z disorazole gene clusters and was thus supposed to be involved in methyl ester formation. In order to verify its function in disorazole Z biosynthesis, we replaced $disF_{427}$ with a gentamicin resistance gene using Red*αβ*-mediated linear-circular homologous recombination (LCHR; see the supplemental material)[42]. Deletion of $disF_{427}$ completely abolished disorazole Z1 (compound 3) production in *M. xanthus* DK1622, whereas the compounds disorazole Z9 (compound 11) and disorazole Z10 (compound 12) accumulated in the culture broth, which was confirmed by HPLC-MS and NMR analysis (**Figure 2.4**A; Tables S12 and S14; Fig. S55 and S61). The methyltransferase $DisF_{427}$ was then expressed and purified as N-terminal His-tagged recombinant protein using *E. coli* BL21(DE3). Incubating purified $DisF_{427}$ with compound 11 or compound 12 and *S*-adenosyl methionine resulted in almost complete conversion to the corresponding methylated compounds in vitro at 30°C in 1 h **Figure 2.4**B and C). These results clearly demonstrated the methyltransferase DisF to be responsible for methyl ester formation in disorazole Z biosynthesis. The accumulation of compound 12 as the major component in the absence of $disF_{427}$ also indicated that stereospecific

oxidation and subsequent methylation might occur initially on one side of the symmetric substrate. However, it remains unclear how the geminal methyl group was oxidised to the hydroxy group and furthermore to the carboxyl group, which implies novel tailoring biochemical steps and motivates further investigation, which is ongoing in our laboratory.

Biological activity of disorazole Z. Disorazole Z1 (compound 3) was tested on a small panel of human cancer cell lines and displayed very pronounced cytotoxic activity, with IC_{50}s in the range of 0.07 to 0.43 nM (Table S16). In comparison to the previously described disorazole A1 (compound 1), it showed similar activity, although it tended to be less potent (by a factor of 4 to 5) when tested on hepatocellular carcinoma (HepG2) and osteosarcoma (U-2 OS) cells.

In order to explore the effects of disorazole Z on microtubule dynamics, U-2 OS cells were treated with disorazoles followed by immunostaining of α-tubulin and fluorescence microscopy. After 5 h treatment with compound 1 or compound 3, a slightly higher density of interphase microtubules around the nuclear periphery was observed, which resembles a local destabilisation. The same effect was already described for, e.g., the microtubule-destabilising agent vinblastine[43]. Interestingly, the acetylated microtubule population, which plays an important role in dynamic cellular processes, was much more affected. This might be caused by the ability of disorazoles to preferentially suppress dynamic mechanisms at the binding sites at the end of microtubules. After prolonged treatment, microtubules were completely depolymerised and the low-abundance acetylated tubulin population was no longer detectable by immunostaining (**Figure 2.5**). In particular, endoplasmic reticulum (ER) dynamics are directly associated with acetylated microtubules, an effect termed ER sliding[44]. Thus, we studied whether specific responses of the ER can be observed after treatment of cells with disorazoles. Both compound 1 and compound 3 induced a delocalisation of the ER structure, which coaligns with (depolymerised) microtubules. However, this event does not trigger ER stress, as determined by applying GRP78/BiP immunostaining to disorazole-treated cells (**Figure 2.5**B). The 78-kDa glucose-regulated protein (GRP78) functions as a chaperone and is a master regulator of the unfolded-protein response (UPR)[45], and it is found to be upregulated after treatment of cells with the ER calcium ATPase inhibitor thapsigargin. However, disorazoles did not directly induce ER stress, although dynamics of the ER are probably greatly impaired due to the delocalisation of the complex.

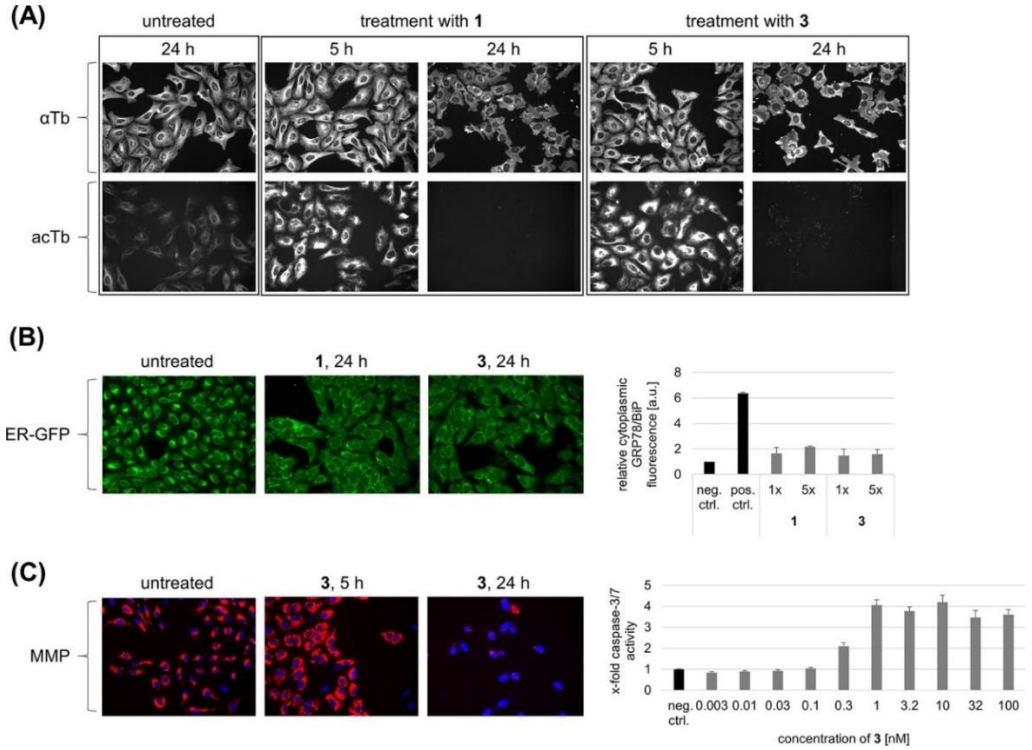


Figure 2.5. Biological characterisation of disorazole Z*. (A) Influence on microtubules. U-2 OS cells were treated with disorazole A1 (compound 1) and disorazole Z1 (compound 3) at their respective IC_{50}s. α-Tubulin (αTb) and acetylated tubulin (acTb) were visualised by immunofluorescence. The appearance of any artifact lines in these images is a result of cropping these images from a larger microscopy image with built-in stitching (see Materials and Methods). (B) Influence on the ER. (Left) U-2 OS cells were transfected with CellLight ER-GFP for live cell imaging of the ER after treatment with compounds 1 and 3 at their respective IC_{50}s. (Right) Relative immunofluorescence quantification of high-content images of the ER stress marker protein GRP78/BiP after 24 h treatment of U-2 OS cells with 1 and 3 at their respective 1-fold and 5-fold IC_{50}s. The positive control was 1 µM thapsigargin, and the negative control was untreated (set to 1). Data are means and standard deviations (SD). (C) Induction of the intrinsic apoptosis pathway. (Left) U-2 OS ells were stained with TMRM as the fluorescence marker of the MMP after treatment of U-2 OS with 3 at its 5-fold IC_{50}. (Right) Quantification of caspase-3/7 activity in HepG2 cells after treatment with compound 3 in serial dilution. The negative control was untreated (set to 1). Data are means and SD.*

To further assess compound 3 and its ability to induce apoptotic processes, the mitochondrial membrane potential (MMP) and caspase activation were determined. Following microtubule depolymerisation and a concurrent cell cycle arrest at the G2/M checkpoint, many tubulin-binding agents are described as inducing a loss of MMP followed by cytochrome c release and activation of the caspase cascade[46]. Here, we could demonstrate that treatment of cells with compound 3 at nanomolar concentrations results in mitochondrial swelling (5 h), followed by a complete loss of MMP (24 h). In line with these findings, we found caspase-3/7 activation upon 24 h treatment of cells with compound 3 at concentrations as low as 0.3 nM (**Figure 2.5**C).

2.4. Experimental Section

General experimental procedures. Melting points were measured on a Büchi-510 melting point apparatus. UV data were recorded on a Shimadzu UV/Vis-2450 spectrophotometer in methanol (UVASOL, Merck). Infrared (IR) data were recorded on a Bruker Tensor 27 IR spectrophotometer. ^{1}H NMR and ^{13}C NMR spectra were recorded on Bruker Avance III 700, DMX 600, UltraShield 500 or DPX 300 NMR spectrometers, locked to the deuterium signal of the solvent. Data acquisition, data processing, and spectral analysis were performed with standard Bruker software and ACD/NMRSpectrus. Chemical shifts are given in parts per million, and coupling constants are in hertz. Analytical reverse-phase high-performance liquid chromatography (RP-HPLC) was carried out with an Agilent 1260 HPLC system equipped with a diodearray UV detector (DAD) and a Corona Ultra detector (Dionex) or a maXis ESI time-of-flight (TOF) mass spectrometer (ESI-HR-MS; Bruker Daltonics). HPLC was carried out with a Waters Acquity C_{18} column, 100 by 2 mm, 1.7 μm; solvent A was H_2O–0.1% formic acid, and solvent B was acetonitrile–0.1% formic acid. The gradient was 5% B for 1 min, increasing to 95% B in 18 min; the flow rate was 0.6 mL/min; and the column temperature was 45°C. All elemental formulae were assigned using the high-resolution data of molecular ion clusters measured with a Bruker maXis ESI-TOF mass spectrometer and calculated with the SmartFormula tool of the Compass DataAnalysis program (Bruker). The myxobacterial strain *Sorangium cellulosum* So ce1875 was isolated in 2001 from soil with plant residues collected near Holbrook, AZ, in 1996 and can be obtained from the DSMZ (German Collection of Microorganisms and Cell Cultures) under the depository number DSM 53600.

Fermentation of *S. cellulosum* So ce1875. A fermentation medium (300 L) was inoculated with 10 L precultured *S. cellulosum* So ce1875 in 2-L Erlenmeyer flasks. The fermentation medium contained 0.8% starch (Cerestar), 0.3% soy meal, 0.05% Casitone, 0.02% soy peptone, 0.1% $MgSO_4$ x $7H_2O$, 0.075% $CaCl_2$ x $2H_2O$, 8 mg/L Na-Fe-EDTA, and 1% Amberlite XAD-16 resin. The pH was 7.3 before autoclaving. Glucose (0.25%) in H_2O was added after autoclaving. The strain was cultivated at 30°C with a pO_2 level at 20% for 14 days. At the end of fermentation, the XAD resin was collected from the culture by sieving. The production of 76 mg/L disorazole Z1 (compound 3) was analysed by HPLC.

Extraction and isolation of disorazole Z from *S. cellulosum* So ce1875. The XAD adsorber resin (3.03 kg) from a large-scale (300-L) fermentation of *S. cellulosum* strain So ce1875 was separated from adhering cells by flotation with water before it was eluted in a glass column with 2 bed volumes of 30% aqueous methanol followed by 3 bed volumes of 100% methanol. The methanol eluate was evaporated to an aqueous mixture, diluted with water, and extracted twice with equal portions of ethyl acetate. After evaporation, the aqueous oil was subjected to a 90% methanol-heptane partitioning removing lipid products with three equal portions of heptane. The aqueous oil was diluted with water and extracted with dichloromethane (DCM) to give 72 g crude extract after evaporation of the solvent. Crystallisation from ethanol provided 36.7 g of raw crystalline disorazole Z1 (compound 3), including several minor structural variants and about 1.7 equivalents of the solvent.

A portion of 10.7 g raw crystals was dissolved in DCM and toluene and evaporated to dryness twice before the material was subjected to Si flash chromatography (Reveleris silica cartridge, 330 g, 61 by 223 mm [Grace], equilibrated with DCM; flow rate, 90 mL/min; solvent A, DCM; solvent B, acetone; gradient, 0% B for 15 min, to 10% B in 15 min, 50 min at 10% B, to 20% B in 5 min, 10 min at 20% B, to 100% B in 35 min, and 10 min at 100% B). A total of 8.2 g of disorazole compound 3 was obtained in the first peak (UV at 280 nm; evaporative light scattering detector) between 40 and 85 min and crystalised from ethanol to give 6.8 g of disorazole compound 3 as crystals containing 1.7 equivalents of ethanol after drying in a vacuum. Further fractions with disorazole variants were collected.

Fraction 3 (91 to 96 min), 760 mg, was separated by reverse-phase medium-pressure liquid chromatography (RP-MPLC; ODS-AQ column, 30 by 480 mm, 25 μm [YMC Ltd.]; solvent A, 50% aqueous methanol; solvent B, 100% methanol; gradient, 20% to 34% B in 120 min, to 45% B in 20 min, to 100% B in 10 min; flow rate, 30 mL/min; UV detection, 310 nm) to give fractions according to UV peaks. The fractions were evaporated, and the remaining water was extracted with ethyl acetate, yielding disorazole Z5 (compound 7) (387 mg), disorazole Z6 (compound 8), and disorazole Z1 (compound 3) (35 mg) after evaporation to dryness.

Fraction 4 (97 to 109 min) (428 mg) was separated similarly with a gradient of 20% B to 32% B in 105 min, to give disorazole Z4 (compound 6) enriched (113 mg, amorphous) and 234 mg which was crystallised from ethanol and furnished compound 6 (195 mg, crystallised, containing 1.2 equivalents of ethanol).

Fraction 6 (628 mg) was separated similarly with a gradient of 10% B to 25% B in 130 min to give a main peak of disorazole Z7 (compound 9) (387 mg), which was crystallised from ethanol, yielding 310 mg of white crystals.

Si flash chromatography of mother liquor (38 g) of a disorazole Z crystallisation provided further fractions, which were separated by RP-MPLC.

Fraction 12 (2.7 g) was separated by RP-MPLC (column, 60 by 500 mm; YMC ODS-AQ, 120 Å and 21 μm; solvent A, 50% methanol; solvent B, methanol; gradient, 15% B to 26% in 225 min, to 50% B in 120 min; flow rate, 65 mL/min; UV detection, 313 nm) to give disorazole Z8 (compound 10) (593 mg) and *O*-desmethyl-dimethyl-disorazole Z10 (compound 12) (503 mg) as main peaks, recovered after evaporation of the methanol and extraction with ethyl acetate.

Fraction 15 (500 mg) was separated be RP-MPLC (column, 30 by 480 mm, ODS-AQ, 120 Å, 16 μm [YMC]; solvent A, 50% aqueous methanol; solvent B, 100% methanol, gradient, 15% for 3 min, 15% B to 30% B in 173 min; flow rate, 30 mL/min; UV detection, 310 nm) to give a main peak at 125 min with 80 mg of disorazole Z9 (compound 11).

A sample of 6 g disorazole Z from Si flash chromatography was further separated by RP-MPLC in two runs (column, 60 by 500 mm, YMC ODS-AQ, 120 Å, 21 μm; solvent A, 50% methanol; solvent B, methanol; gradient, 20% B for 160 min, to 30% B in 240 min; flow rate, 60 mL/min; UV detection, 313 nm) to give a mixture of disorazoles (680 mg) eluting in front of compound 3. This mixture was again separated by RP-MPLC (column, 30 by 480 mm, ODS-AQ, 15 μm [YMC]; solvent A, 50% aqueous methanol; solvent B, 100% methanol; gradient, 30% B for 47 min, 30% B to 40% B in 40 min, to 100% B in 30 min; flow rate, 30 mL/min; UV detection, 310 nm) to give disorazole Z3 (compound 5) (284 mg). Stepwise crystallisation from methanol-water gave about 200 mg of compound 5.

Enriched disorazole Z1 (compound 3) (2.33 g) from Si flash chromatography was separated by RPMPLC (column, 60 by 500 mm, YMC ODS-AQ, 120 Å, 21 μm; solvent A, 50% methanol; solvent B, methanol; gradient, 20% B for 160 min, to 30% B in 240 min; flow rate, 60 mL/min; UV detection, 313 nm) to give compound 3 and a mixture of disorazoles (48 mg) eluting after compound 3. This fraction was separated by preparative thin-layer chromatography (four plates; 20 by 20 cm; silica gel, 0.5 mm; solvent, DCM-methanol [MeOH] [19:1], UV detection, 254 nm) to give compound 3 after elution with the same solvent mixture and compound 5 at a higher R_f value (19 mg).

Disorazole Z1 (compound 3): $C_{40}H_{46}N_2O_{12}$ M = 746.80; t_R = 11.09 min; crystallised from ethanol with 1.7 equivalent ethanol $C_{40}H_{46}N_2O_{12}$–1.7 $C_2H_6O_1$ M = 825.12; mp.128 to 130°C; $[\alpha]^{22}_D$ +169.8 (*c* = 1.01, CH_3OH); UV (MeOH) (amorph): λ_{max} (log ε) 314 (4.948), 340 (sh) nm; UV (chloroform) (crystallised): λ_{max} (log ε) 315 (4.969), 322, 340 (sh) nm; IR (KBr) (amorph): ν_{max} 3486 (m), 2950 (m), 1730 (s), 1614 (m), 1312 (m), 1234 (m), 1153 (m), 1115 (s), 996 (s), 973 (m) cm^{-1}; NMR data, see Table S1; ESI-HR-MS *m/z* 747.3132 $[M+H]^+$ (calculated for $C_{40}H_{47}N_2O_{12}$, 747.3123), *m/z* 769.2949 $[M+Na]^+$ (calculated for $C_{40}H_{46}N_2O_{12}Na$, 769.2943), *m/z* 729.3022 $[M-H_2O+H]^+$ (calculated for $C_{40}H_{45}N_2O_{11}$, 729.3018).

Disorazole Z2 ($\Delta^{7,8}$-*cis*-disorazole Z) (compound 4): $C_{40}H_{46}N_2O_{12}$ M = 746.80; t_R = 11.17 min; crystallised from methanol; mp.133 to 135°C.; $[\alpha]^{22}_D$ +305 (*c* = 0.41, $CHCl_3$); UV (chloroform) (crystallised): λ_{max} (log ε) 313 (5.014), 299, 342 (sh) nm; NMR data, see Table S4; ESI-HR-MS *m/z* 747.3128 $[M+H]^+$ (calculated for $C_{40}H_{47}N_2O_{12}$, 747.3124), *m/z* 769.2944 $[M+Na]^+$ (calculated for $C_{40}H_{46}N_2O_{12}Na$, 769.2943), *m/z* 729.3020 $[M-H_2O+H]^+$ (calculated for $C_{40}H_{45}N_2O_{11}$, 729.3018).

Disorazole Z3 ($\Delta^{9,10}$-*trans*-disorazole Z) (compound 5): $C_{40}H_{46}N_2O_{12}$ M = 746.80; t_R = 10.85 min; $[\alpha]^{22}_D$ –59.6 (*c* = 0.12, $CHCl_3$); UV ($CHCl_3$): λ_{max} (log ε) 314 (4.948), 340 (sh) nm; NMR data, see Table S5; ESI-HR-MS *m/z* 747.3123 $[M+H]^+$ (calculated for $C_{40}H_{47}N_2O_{12}$, 747.3124), *m/z* 769.2935 $[M+Na]^+$ (calculated for $C_{40}H_{46}N_2O_{12}Na$, 769.2943), *m/z* 729.3012 $[M-H_2O+H]^+$ (calculated for $C_{40}H_{45}N_2O_{11}$, 729.3018).

Disorazole Z4 (7,8-epoxy-disorazole Z) (compound 6): $C_{40}H_{46}N_2O_{13}$ M = 762.80; t_R = 10.09 min; m.p.132 to 138°C; $[\alpha]^{22}_D$ +170.9 (*c* = 0.23, $CHCl_3$); UV ($CHCl_3$): λ_{max} (log ε) 266 (4.601), 328 (4.652), 344 (4.558), 276, 314 (sh) nm; NMR data, see Table S6; ESI-HR-MS *m/z* 763.3074 $[M+H]^+$ (calculated for $C_{40}H_{47}N_2O_{13}$ 763.3073), *m/z* 785.2896 $[M+Na]^+$ (calculated for $C_{40}H_{46}N_2O_{13}Na$, 785.2892), *m/z* 745.2975 $[M-H_2O+H]^+$ (calculated for $C_{40}H_{45}N_2O_{12}$, 745.2967), *m/z* 727.2866 $[M-2H_2O+H]^+$ (calculated for $C_{40}H_{43}N_2O_{11}$, 727.2861), *m/z* 807.2991 $[M+HCOOH-H]^-$ (calculated for $C_{41}H_{47}N_2O_{15}$, 807.2982).

Disorazole Z5 (9,10-epoxy-disorazole Z) (compound 7): $C_{40}H_{46}N_2O_{13}$ M = 762.80; t_R = 9.87 min; m.p.137 to 148°C; $[\alpha]^{22}_D$ +135.1 (*c* = 0.28, $CHCl_3$); UV ($CHCl_3$): λ_{max} (log ε) 301 (4.823), 345 (4.416), 291, 323 (sh) nm; NMR data, see Table S7; ESI-HR-MS *m/z* 763.3085 $[M+H]^+$ (calculated for $C_{40}H_{47}N_2O_{13}$ 763.3073), *m/z* 780.3351 $[M+NH_4]^+$ (calculated for $C_{40}H_{50}N_3O_{13}$, 780.3338), *m/z* 785.2903 $[M+Na]^+$ (calculated for $C_{40}H_{46}N_2O_{13}Na$, 785.2892), *m/z* 1542.6339 $[2M+NH_4]^+$ (calculated for $C_{80}H_{96}N_5O_{26}$ 1542.6338), *m/z* 745.2977 $[M+H-H_2O]^+$ (calculated for $C_{40}H_{45}N_2O_{12}$ 745.2967).

Disorazole Z6 (27,28-epoxy-disorazole Z) (compound 8): $C_{40}H_{46}N_2O_{13}$ M = 762.80; t_R = 10.04 min; m.p.121 to 125°C; $[\alpha]^{22}_D$ +20.3 (*c* = 0.272, $CHCl_3$); UV ($CHCl_3$): λ_{max} (log ε) 315 (4.924), 300, 322, 342 (sh) nm; NMR data, see Table S8; ESI-HR-MS *m/z* 763.3079 $[M+H]^+$ (calculated for $C_{40}H_{47}N_2O_{13}$ 763.3073), *m/z* 785.2898 $[M+Na]^+$ (calculated for $C_{40}H_{46}N_2O_{13}Na$, 785.2892), *m/z* 745.2971 $[M-H_2O+H]^+$ (calculated for $C_{40}H_{45}N_2O_{12}$, 745.2967).

Disorazole Z7 (29-hydroxy-disorazole Z) (compound 9): $C_{40}H_{46}N_2O_{13}$ M = 762.80; t_R = 8.96 min; $[\alpha]^{22}_D$ −11.6 (*c* = 0.23, $CHCl_3$); UV ($CHCl_3$): λ_{max} (log ε) 316 (4.893), 300, 342 (sh) nm; NMR data, see Table S9; ESI-HR-MS *m/z* 763.3075 $[M+H]^+$ (calculated for $C_{40}H_{47}N_2O_{13}$ 763.3073), *m/z* 785.2883 $[M+Na]^+$ (calculated for $C_{40}H_{46}N_2O_{13}Na$, 785.2892).

Disorazole Z8 (31-*O*-desmethyl-disorazole Z) (compound 10): $C_{39}H_{44}N_2O_{12}$ M = 732.77; t_R = 9.38 min; $[\alpha]^{22}_D$ +305.2 (*c* = 0.41, $CHCl_3$); UV (MeOH) (amorph): λ_{max} (log ε) 316 (4.954), 300, 343 (sh) nm; NMR data, see Table S10; ESI-HR-MS *m/z* 733.2958 $[M+H]^+$ (calculated for $C_{39}H_{45}N_2O_{12}$, 733.2967), *m/z* 755.2779 $[M+Na]^+$ (calculated for $C_{39}H_{44}N_2O_{12}Na$, 755.2786), *m/z* 715.2866 $[M-H_2O+H]^+$ (calculated for $C_{39}H_{43}N_2O_{11}$, 715.2861).

Disorazole Z9 (31-*O*-desmethyl-39-hydroxy-disorazole Z) (compound 11): $C_{38}H_{44}N_2O_{11}$ M = 704.76; t_R = 9.35 min; $[\alpha]^{22}_D$ +55.8 (*c* = 0.26, $CHCl_3$); UV ($CHCl_3$): λ_{max} (log ε) 317 (4.942), 300, 344 (sh) nm; NMR data, see Table S11; ESI-HR-MS *m/z* 705.3023 $[M+H]^+$ (calculated for $C_{38}H_{45}N_2O_{11}$ 705.3018), *m/z* 687.2920 $[M-H_2O+H]^+$ (calculated for $C_{38}H_{43}N_2O_{10}$, 687.2912), *m/z* 1409.5975 $[2M+H]^+$ (calculated for $C_{76}H_{89}N_4O_{22}$, 1409.5963).

Disorazole Z10 (*O*-desmethyl-dimethyl-disorazole Z) (compound 12): $C_{38}H_{44}N_2O_{10}$ M = 688.76; t_R = 10.59 min; $[\alpha]^{22}_D$ +25.4 (*c* = 0.24, $CHCl_3$); UV ($CHCl_3$): λ_{max} (log ε) 316 (4.825), 300, 342 (sh) nm; NMR data, see Table S13; ESI-HR-MS *m/z* 689.3057 $[M+H]^+$ (calculated for $C_{38}H_{44}N_2O_{10}$ 689.3069), *m/z* 671.2952 $[M-H_2O+H]^+$ (calculated for $C_{38}H_{43}N_2O_9$, 671.2963).

Preparation of disorazole Z Mosher esters. To prepare disorazole Z (*S*)-Mosher ester (compound 13), 20 mg of crystalline disorazole Z1 (compound 3) was twice dissolved in pyridine and toluene and evaporated to dryness. The residue was dissolved in 0.2 mL of dry pyridine. Twenty-five microliters of (*R*)-(–)-*α*-methoxy-*α*-trifluoromethylphenylacetyl chloride was added in three portions to the stirred solution for 24 h. The mixture was dissolved with pyridine and sodium hydrogen carbonate solution (1%), extracted with DCM, washed with water twice, and dried by evaporation with toluene. The residue was purified by Si flash chromatography (12 g silica gel, 40 μm; Reveleris [Grace]; solvent A, petroleum ether; solvent B, ethyl acetate; gradient, 0% B for 1

min, in 1 min to 9% B, 9% B for 1 min, in 4.7 min to 36% B, for 2.5 min 36% B; flow rate, 36 mL/min) The main peak was collected and evaporated to dryness, yielding 39 mg of compound 13. $C_{60}H_{60}F_6N_2O_{16}$ M was 1179.1; for NMR data, see Table S2.2.-2.3.

The disorazole Z (*R*)-Mosher ester (compound 14) was prepared analogously using (*S*)-(+)-*α*-methoxy-*α*-trifluoromethylphenylacetyl chloride, yielding 19 mg of compound 14. $C_{60}H_{60}F_6$ N_2O_{16} M was 1179.1; for NMR data, see Table S2.2.-2.3.

Crystallography and X-ray analysis.

(i) Crystallographic data of compounds 3, 6, and 7. Crystals suitable for single-crystal X-ray analysis were obtained from ethanol. The data set for all structures was collected using a Bruker X8 Apex diffractometer. Graphite-monochromated $Mo_{K\alpha}$ radiation (λ = 0.71073 Å) was used throughout. Data were collected at 133 (2) K (disorazole Z1 [compound 3] = sh3137_a_sq, disorazole Z5 [compound 7] = sh3279) or 152 (2) K (disorazole Z4 [compound 6] = sh3191) and corrected for absorption effects using the multiscan method. The structures were solved by direct methods using SHELXS-97[47] and were refined by full matrix least-squares calculations on F2 (SHELXL2018[48]) in the graphical user interface ShelXle[49].

(ii) Refinement. All non-H atoms were located in the electron density maps and refined anisotropically. C-bound H atoms were placed in positions of optimised geometry and treated as riding atoms. Their isotropic displacement parameters were coupled to the corresponding carrier atoms by a factor of 1.2 (CH, CH_2) or 1.5 (CH_3, OH) for compound 3 (sh3137_a_sq) and compound 6 (sh3191). Restraints of 0.84 (0.01) Å were used for O-H bond lengths. For compound 7 (sh3279), the O-bonded H atoms were located in the electron density maps. Their positional parameters were refined using isotropic displacement parameters, which were set at 1.5 times the Ueq value of the parent atom. Regarding disorder, for compound 7 (sh3279), each oxirane atom is not fully occupied (O4a, 0.76; O4b, 0.24); furthermore, the propylene group of the main compound as well as one solvent ethanol molecule is split over two positions. Their occupancy factors refined to 0.87 and 0.76, respectively, for the major components. For compound 6 (sh3191), two of the solvent ethanol molecules and the propylene residue of the structure were split over two positions. Their occupancy factors refined to 0.55, 0.86, and 0.81 for the major components, respectively. Regarding SQUEEZE, for compound 3 (sh3137_a_sq), the unit cell contains approximately 2 solvent ethanol molecules (occupancy factor less than 1.0 for each ethanol molecule), which were treated

as a diffuse contribution to the overall scattering without specific atom positions by SQUEEZE/PLATON.

Cloning and engineering of the *dis427* gene cluster. The myxobacterium *Sorangium cellulosum* So ce427 was cultivated using CYH medium (1.5 g/L Casitone, 1.5 g/L yeast extract, 4 g/L starch, 1 g/L soy

meal, 1 g/L glucose, 1 g/L calcium chloride dihydrate, 0.5 g/L magnesium sulfate heptahydrate, 5.96 g/L HEPES, 4 mg/L Na-Fe-EDTA; pH 7.3) at 30°C. Clumpy cells were collected by centrifugation and then homogenised to become a suspension for isolation of genomic DNA according to the published protocol[42]. The genomic DNA was treated with RNase A and the appropriate DNA restriction endonuclease (MluCI or BstXI) to eliminate RNA contamination and to release the gene cluster. The digested DNA was recovered by phenolic chloroform extraction and ethanol precipitation and finally dissolved in autoclaved Milli-Q water. The linear cloning vectors (p15A-cm or pBR322-amp) containing homology arms with homology to the end of the released genomic DNA fragment were achieved by PCR. The digested DNA and corresponding cloning vector were then used for electroporation of Escherichia coli GB05-dir expressing RecET recombinase for linear-linear DNA homologous recombination (LLHR)[42]. Recombinant plasmids were isolated from colonies and verified by restriction digestion and Sanger/Illumina sequencing. The correct recombinant plasmids p15A-cm-MluCI-dis427 and pBR322-amp-BstXI-dis427 were then digested with MluCI and BstXI, respectively, to release the cloned fragments. These two fragments were then assembled with a PCR-amplified fragment and a p15A-cm vector by LLHR, leading to the generation of plasmid p15A-cm-dis427. To construct p15A-km-int-dis427 and p15A-km-int-Ptet-dis427, the *km-int* or *km-int*-P*tet* cassette (Table S15) was amplified by PCR and electroporated into *E. coli* GB05-red harboring p15A-cm-dis427 and expressing Redαβ recombinase for linear-circular homologous recombination (LCHR)[42] to replace the *cm* cassette. Similarly, the expression plasmid p15A-km-int-P*van*-dis427 was constructed based on p15A-km-int-dis427 by LCHR using the *apr*-P*van*-disA cassette. p15A-km-int-Ptet-Papr-dis427 and p15A-km-int-Ptet-dis427-gent-delF were constructed based on p15A-km-int-Ptet-dis427 by LCHR using P*apr-disD* and *gent-delF* cassettes, respectively.

Electroporation of *M. xanthus* DK1622. *M. xanthus* DK1622 was inoculated into CTT medium (10 g/L Casitone, 10 mM Tris-HCl, 8 mM magnesium sulfate, 1 mM potassium phosphate; pH 7.6) and incubated at 30°C with shaking until exponential phase. For preparation of electrocompetent cells, 1.8 mL culture was transferred into a 2-mL microcentrifuge tube with a hole punched in the cap. Cells were centrifuged, washed twice, and finally resuspended in 50 μL autoclaved Milli-Q water. After addition of 1 to 3 μg plasmid DNA, the mixture was then transferred into a 1-mm cuvette. Electroporation was performed at 650 V, 400 Ω, and 25 μF using a Bio-Rad Gene Pulser Xcell electroporation system. The pulsed cells were mixed with 1.6 mL fresh CTT medium and transferred back into the 2-mL microcentrifuge tube. After recovery at 30°C for 6 h in an Eppendorf thermomixer, the cells were mixed with 10 mL CTT soft agar (0.6% agar, supplemented with 50 μg/mL kanamycin) and poured onto a CTT agar plate (1.2% agar, supplemented with 50 μg/mL kanamycin). The plate was incubated at 30°C, and colonies appeared after 4 to 7 days. Kanamycin-resistant colonies were picked out and inoculated into a new CTT agar plate supplemented with 50 μg/mL kanamycin. In order to verify intact integration of the biosynthetic gene cluster, three pairs of primers located at different positions of the gene cluster were used for colony PCR. Cells were scraped from the agar plate, suspended in 50 μL autoclaved Milli-Q water, and incubated at 100°C for 20 min. After centrifugation, 1 to 2 μL of the supernatant was used as the PCR template.

Heterologous production of disorazole Z in *M. xanthus* DK1622. *M. xanthus* DK1622 mutants that had the integrated disorazole biosynthetic gene cluster were inoculated in 1.6 mL CTT medium supplemented with 50 μg/mL kanamycin in a 2-mL microcentrifuge tube with a hole punched in the cap and incubated at 30°C in an Eppendorf thermomixer for 1 day. After that, 1 mL of the culture was inoculated into 50 mL CTT medium supplemented with 50 μg/mL kanamycin in a 300-mL baffled flask and incubated at 30°C, 180 rpm for 2 days. After addition of inducer (0.5 μg/mL anhydrotetracycline when the P*tet* promoter was used, 1 mM vanillate when the P*van* promoter was used) and 1 mL XAD-16 resin, incubation was continued for 2 more days. Cells and XAD-16 resin were collected by centrifugation and resuspended in methanol for extraction. After filtration, the extracts were dried by rotary evaporation *in vacuo* and redissolved in 1 mL methanol for HPLC-MS analysis using the method described above.

Isolation of disorazole Z from *M. xanthus* DK1622 mutants. For heterologous production of disorazole Z1 (compound 3), the fully grown culture of *M. xanthus* DK1622::km-int-Ptet-Papr-dis427 or *M. xanthus* DK1622::km-int-Pvan-dis427 was inoculated 1:100 into 2 L CTT medium supplemented with 50 μg/mL kanamycin in 5-L unbaffled flasks and cultivated at 30°C and 180 rpm for 2 days. After addition of inducer (0.5 μg/mL anhydrotetracycline when the P*tet* promoter was used, 1 mM vanillate when the P*van* promoter was used) and 2% XAD-16 resin, incubation was continued for 3 more days. Cells and XAD-16 resin were collected by centrifugation, lyophilised to dryness, and extracted stepwise with methanol. The methanol extract was concentrated and partitioned with *n*-hexane to remove nonpolar impurity. After evaporation *in vacuo*, the methanol extract was then dissolved in Milli-Q water and extracted twice with equal portions of ethyl acetate. The ethyl acetate extract was evaporated to dryness, redissolved in methanol, and fractionated using Si flash chromatography on a Biotage system. The gradient was set as follows: 0 to 20 CV (column volume), hexane to ethyl acetate; 20 to 40 CV, ethyl acetate to MeOH; 40 to 45 CV, MeOH. Fractions containing targeted compounds were combined and further separated by semipreparative HPLC on a Waters XSelect peptide BEH C_{18} column (250 by 10 mm; 5 μm) using A (Milli-Q water plus 0.1% formic acid) and B (acetonitrile plus 0.1% formic acid) as mobile phases at a flow rate of 5 mL/min and a column temperature of 45°C. Gradient conditions were set as follows: 0 to 4 min, 5% B; 4 to 5 min, to 53% B; 5 to 20 min, 53% B; 20 to 21 min, to 95% B; 21 to 25 min, 95% B; 25 to 26 min, to 5% B; 26 to 30 min, 5% B. Fractions were detected by UV detection at 320 nm and were collected using an AFC-3000 fraction collector based on a retention time of 17.7 min. Heterologous production and purification of disorazole Z9 (compound 11) and Z10 (compound 12) were done in a similar manner using the strain *M. xanthus* DK1622::km-int-Ptet-dis427-delF. For purification of compound 11 using semipreparative HPLC, separation was achieved using a Waters XBridge peptide BEH C_{18} column (250 by 10 mm; 5 μm) using the following gradient: 0 to 5 min, 5% B; 5 to 15 min, to 40% B; 15 to 25 min, 40% B; 25 to 26 min, to 95% B; 26 to 31 min, 95% B; 31 to 32 min, to 5% B; 32 to 35 min, 5% B. Fractions were detected by UV detection at 320 nm and were collected using an AFC-3000 fraction collector based on a retention time of 21.8 min. Similarly, compound 12 was purified using a modified gradient: 0 to 5 min, 5% B; 5 to 15 min, to 33% B; 15 to 50 min, 33% B; 50 to 51 min, to 95% B; 51 to 56 min, 95% B; 56 to 57 min, to 5% B; 57 to 60 min, 5% B. Fractions were detected by UV detection at 320 nm and were collected using an AFC-3000 fraction collector based on a retention time of 46.0 min.

Purification and *in vitro* reaction of recombinant protein DisF$_{427}$. The DNA fragment containing *disF*$_{427}$ was amplified by PCR using So ce427 genomic DNA as the template and oligonucleotides HisTEV disF-F and HisTEV-disF-R. The plasmid pHis-TEV[50] was linearised by NcoI/XhoI and assembled with the PCR fragment by Gibson assembly, generating the protein expression construct pHisTEV-disF427. After Sanger sequencing, this construct was electroporated into *E. coli* BL21(DE3). The resulting strain was cultivated in 300-mL flasks at 37°C overnight in 50 mL LB medium (10 g/L tryptone, 5 g/L yeast extract, 5 g/L sodium chloride; pH 7.0) supplemented with 50 μg/mL kanamycin. Twenty milliliters of overnight culture was used to inoculate 2 L fresh LB medium supplemented with 50 μg/mL kanamycin in a 5-L flask. After cultivation at 37°C until the optical density at 600 nm (OD_{600}) was about 0.6, the culture was cooled to 16°C, induced with IPTG (isopropyl-β-D-thiogalactopyranoside) at a final concentration of 0.1 mM and then further cultivated at 16°C overnight. Cells were harvested by centrifugation at 4°C, resuspended in ice-cold lysis buffer (25 mM Tris, 200 mM NaCl, 10% glycerol, 20 mM imidazole; pH 8.0) and lysed using a continuous-flow cell disrupter (Constant Systems) at 25,000 lb/in^2 and 4°C. After centrifugation at 23,500 rpm and 4°C for 30 min, the cell debris was removed and the supernatant was loaded onto a 5-mL HisTrap HP column (GE Healthcare) for nickel affinity chromatography using the ÄKTA protein purification system. Fractions containing recombinant protein in elution buffer (25 mM Tris, 200 mM NaCl, 10% glycerol, 250 mM imidazole; pH 8.0) were collected and loaded onto a HiPrep 26/10 desalting column (GE Healthcare) to remove imidazole using desalting buffer (25 mM Tris, 200 mM NaCl, 10% glycerol; pH 8.0). The eluents were collected, concentrated, and stored at 280°C.

The *in vitro* reaction was carried out at 30°C for 1 h in a 50-μL mixture containing 25 mM Tris, 200 mM NaCl, 2 mM $MgCl_2$, 2mMS-adenosyl methionine, 5 μL (about 1 mg/mL) recombinant protein, 0.5 μL (1 mg/mL in MeOH) disorazole Z9 (compound 11) or Z10 (compound 12). Boiled (100°C for 10 min) recombinant protein was used as a negative control. After addition of 50 μL MeOH, the mixture was vortexed and centrifuged at 15,000 rpm for 15 min. Two microliters of the supernatant was used for HPLC-MS analysis.

Biological characterisation.

(i) IC_{50}. Cell lines were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ) or the American Type Culture Collection (ATCC) and were handled according to standard procedures as recommended by the depositor. Cells were seeded in 96-well plates and treated with disorazole A1 (compound 1) and disorazole Z1 (compound 3) at serial dilutions for 48 h. Viability was determined by adding resazurin sodium salt for 3 h. Fluorescence measurements were performed using a SpectraMax T5 plate reader (Molecular Devices). Readouts were referenced and IC_{50}s were determined by sigmoidal curve fitting using OriginPro software. Data for compound 3 were determined as duplicates in three independent experiments.

(ii) Immunostaining and high-content imaging. U-2 OS cells were seeded in 96-well imaging plates. After overnight equilibration, the cells were treated with compound 1 and compound 3 as assigned and incubated for up to 24 h. Cells were fixed with cold (220°C) acetone-MeOH (1:1) for 10 min. After being washed with phosphate-buffered saline (PBS), the cells were permeabilised with 0.01% Triton X-100 in PBS. The following primary antibodies (Sigma) were used: α-tubulin monoclonal antibody (MAb), acetylated tubulin MAb, and GRP78/BiP MAb. For labeling, cells were incubated with primary antibody for 45 min at 37°C, followed by incubation with the secondary antibody (Alexa Fluor 488 goat anti-mouse or anti-rabbit immunoglobulin; Molecular Probes) under the same conditions. After cells were washed with PBS, the nuclear stain Hoechst 33342 (5 μg/mL) was applied for 10 min. Samples were imaged on an automated microscope (BD Pathway 855) suitable for high-content screening. In order to capture full-width pictures of a larger area, a built-in stitching technique was used to combine multiple frames. A representative part of the larger microscopy image was cropped for illustrations. In case of GRP78 immunostaining, fluorescence intensity was determined in the cytoplasmic segment as defined by a ring around nuclei. The relative intensity of GRP78 fluorescence in this region of interest was used as a measure for ER stress/UPR.

(iii) Live-cell imaging of the ER. U-2 OS cells were seeded in 96-well imaging plates and transfected using CellLight ER-GFP (Thermo Fisher Scientific) according to the manufacturer's protocol. After treatment with compound 1 and compound 3 in triplicate as indicated, live cells were imaged on an automated microscope (BD Pathway 855).

(iv) MMP. U-2 OS cells were seeded at 8 x 10^3 cells/well in 96-well imaging plates and were treated after overnight equilibration with compound 3 as assigned. Following treatment, the cells were washed twice with PBS and 100 μL of a staining solution (50 nM tetramethyl rhodamine methyl ester [TMRM] and 5 μg/mL Hoechst 33342 in assay buffer) was added. The cells were stained for 30 min at 37°C, and after they were washed with assay buffer, the samples were examined on an automated microscope (BD Pathway 855).

(v) Determination of caspase-3/7 activity. HepG2 cells were treated with compound 3 as indicated and lysed in buffer containing 25 mM HEPES, 5 mM $MgCl_2$, 1 mM EGTA and 0.1% (vol/vol) Triton X-100. Protein lysates were mixed with substrate buffer containing 50 mM HEPES, 0.1% (wt/vol) CHAPS {3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate}, 1% (wt/vol) sucrose, 0.2% (wt/vol) dithiothreitol (DTT) and 0.05 mM Ac-DEVD-7-amino-4-trifluoro-methylcoumarin (AFC). Generation of free AFC was determined after 2 h at 37°C by measurement of fluorescence using a SpectraMax T5 plate reader (Molecular Devices). For normalisation, protein concentrations were determined by Bradford assay (absorption, 595 nm) using an Ultra microplate reader EL808 (BioTek Instruments). Data were determined in three independent experiments.

Data availability. Crystallographic data for the structure have been deposited with the Cambridge Crystallographic Data Centre (CCDC), Cambridge, UK. Copies of the data can be obtained free of charge by referencing the depository number: 2236753 (disorazole Z1 [compound 3]; data block data_sh3137_a_sq; unit cell parameters, a 8.5672 [3] b 20.1022 [6] c 25.8711 [9] P212121), 1834868 (disorazole Z4 [compound 6]; data block data_sh3191; unit cell parameters, a 8.6801 [6] b 20.1469 [16] c 26.259 [2] P212121), and 1834869 (disorazole Z5 [compound 7]; data block data_sh3279; unit cell parameters, a 8.7058 [6] b 20.2194 [15] c 26.000 [2] P212121).

2.5. Conclusion and Outlook

In this study, we describe the full structural elucidation of 10 novel disorazole congeners exhibiting a significantly modified basic structure compared to disorazole A and thus grouped as a new subclass of disorazole anticancer drugs termed disorazole Z. This family of compounds possesses a shortened polyketide chain in each half-side of the bis-lactone ring and a carboxymethyl ester at the position where a geminal methyl group is installed in disorazole A. The discovery of the disorazole Z biosynthetic gene cluster and comparison to the disorazole A biosynthetic pathway allowed us to understand the structural differences between these two types of disorazoles. The successful heterologous expression of the disorazole Z gene cluster in an amenable host organism paved the way for detailed biosynthesis studies, e.g., elucidating the intriguing biosynthetic steps involving the oxidation of one methyl centre in the geminal dimethyl group, and rational biosynthetic engineering to further improve the yield of disorazole Z. The system will also allow us to generate nonnatural disorazole family compounds through combinatorial biosynthesis. Activity assays of disorazole Z1 and disorazole A1 revealed similar biological activities in cancer cell lines and thus great potential for this family of compounds to be employed as antitumor drugs, a possibility which is being explored by using a peptide-drug conjugate.

2.6. Reference

1. Herrmann, J., Fayad, A. A. & Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **34,** 135–160; 10.1039/C6NP00106H (2017).
2. Weissman, K. J. & Müller, R. Myxobacterial secondary metabolites: bioactivities and modes-of-action. *Nat. Prod. Rep.* **27,** 1276–1295; 10.1039/c001260m (2010).
3. Wenzel, S. C. & Müller, R. Myxobacteria—''microbial factories' for the production of bioactive secondary metabolites. *Mol. Biosyst.* **5,** 567–574; 10.1039/b901287g (2009).
4. 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.* **106,** 233–253; 10.1016/j.jbiotec.2003.07.015 (2003).
5. Hopkins, C. D. & Wipf, P. Isolation, biology and chemistry of the disorazoles: new anti-cancer macrodiolides. *Nat. Prod. Rep.* **26,** 585–601; 10.1039/b813799b (2009).
6. Irschik, H., Jansen, R., Gerth, K., Höfle, G. & Reichenbach, H. Disorazol A, an efficient inhibitor of eukaryotic organisms isolated from myxobacteria. *J. Antibiot.* **48,** 31–35 (1995).
7. Rox, K., Rohde, M., Chhatwal, G. S. & Müller, R. Disorazoles block group A streptococcal invasion into epithelial cells via interference with the host factor ezrin. *Cell Chem. Biol.* **24,** 159–170; 10.1016/j.chembiol.2016.12.011 (2017).
8. Jansen, R., Irschik, H., Reichenbach, H., Wray, V. & Höfle, G. Disorazoles, highly cytotoxic metabolites from the Sorangicin-producing bacterium *Sorangium cellulosum,* strain So ce12. *Liebigs Ann. Chem.* **1994,** 759–773 (1994).
9. Elnakady, Y. A., Sasse, F., Lünsdorf, H. & Reichenbach, H. Disorazol A1, a highly effective antimitotic agent acting on tubulin polymerization and inducing apoptosis in mammalian cells. *Biochem. Pharmacol.* **67,** 927–935; 10.1016/j.bcp.2003.10.029 (2004).
10. Kopp, M., Irschik, H., Pradella, S. & Müller, R. Production of the tubulin destabilizer disorazol in *Sorangium cellulosum*: biosynthetic machinery and regulatory genes. *ChemBioChem* **6,** 1277–1286; 10.1002/cbic.200400459 (2005).
11. Carvalho, R., Reid, R., Viswanathan, N., Gramajo, H. & Julien, B. The biosynthetic genes for disorazoles, potent cytotoxic compounds that disrupt microtubule formation. *Gene* **359,** 91–98; 10.1016/j.gene.2005.06.003 (2005).

12. Helfrich, E. J. N. & Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **33,** 231–316; 10.1039/c5np00125k (2016).

13. Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **27,** 996–1047; 10.1039/b816430b (2010).

14. Tu, Q. *et al.* Genetic engineering and heterologous expression of the disorazol biosynthetic gene cluster via Red/ET recombineering. *Sci. Rep.* **6,** 21066; 10.1038/srep21066 (2016).

15. Irschik, H., Jansen, R. & Sasse, F. *Biologically active compounds obtainable from sorangium cellulosum. European Patent EP20050106539* (2005).

16. Aicher, B. *et al.* 330 Highly Potent Cytotoxic Conjugates of Disorazol Z Linked to a LHRH Receptor Targeting Peptide Interfere with Cell Cycle Progression in Human Cancer Cell Lines and Suppress Tumor Growth in a LHRH Receptor Positive Ovarian Cancer Xenograft Model. *European Journal of Cancer* **48,** 101; 10.1016/S0959-8049(12)72128-7 (2012).

17. Guenther, E. G. *et al.* Targeting Disorazol Z to LHRH-receptor positive tumors by the cytotoxic conjugate AEZS-125 (2008).

18. Aicher, B. *et al.* Abstract C214: Disorazol Z: A highly cytotoxic natural compound with antitumor properties. *Mol Cancer Ther* **10,** C214-C214; 10.1158/1535-7163.TARG-11-C214 (2011).

19. Guenther, E., Schaefer, O., Teifel, M. & Klaus, P. *Conjugates of Disorazoles and their Derivatives with Cell-Binding Molecules, Novel Disorazole Derivatives, Processes of Manufactoring and Uses Therefof. European Patent EP20070803270* (2007).

20. Seitz, S. *et al.* Triple negative breast cancers express receptors for LHRH and are potential therapeutic targets for cytotoxic LHRH-analogs, AEZS 108 and AEZS 125. *BMC cancer* **14,** 847; 10.1186/1471-2407-14-847 (2014).

21. Menchon, G. *et al.* A fluorescence anisotropy assay to discover and characterize ligands targeting the maytansine site of tubulin. *Nature communications* **9,** 2106; 10.1038/s41467-018-04535-8 (2018).

22. Nicolaou, K. C., Bellavance, G., Buchman, M. & Pulukuri, K. K. Total Syntheses of Disorazoles A_1 and B_1 and Full Structural Elucidation of Disorazole B_1. *J. Am. Chem. Soc.* **139,** 15636–15639; 10.1021/jacs.7b09843 (2017).

23. Schäckel, R., Hinkelmann, B., Sasse, F. & Kalesse, M. The Synthesis of Novel Disorazoles. *Angew. Chem. Int. Ed. Engl.* (2010).

24. Miethke, M. *et al.* Towards the sustainable discovery and development of new antibiotics. *Nat. Rev. Chem.* **5,** 726–749; 10.1038/s41570-021-00313-1 (2021).

25. Huo, L. *et al.* Heterologous expression of bacterial natural product biosynthetic pathways. *Nat. Prod. Rep.* **36,** 1412-1436; 10.1039/C8NP00091C (2019).

26. Ongley, S., Bian, X., Neilan, B. A. & Müller, R. Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat. Prod. Rep.* **30,** 1121–1138; 10.1039/c3np70034h (2013).

27. Hiemstra, H., Houwing, H. A., Possel, O. & van Leusen, A. M. Carbon-13 nuclear magnetic resonance spectra of oxazoles. *Can. J. Chem.* **57,** 3168–3170; 10.1139/v79-517 (1979).

28. Hoye, T. R., Jeffrey, C. S. & Shao, F. Mosher ester analysis for the determination of absolute configuration of stereogenic (chiral) carbinol carbons. *Nat. Protoc.* **2,** 2451–2458; 10.1038/nprot.2007.354 (2007).

29. Pehk, T. *et al.* Determination of the absolute configuration of chiral secondary alcohols; new advances using 13C- and 2D-NMR spectroscopy. *Tetrahedron, asymmetry* **4,** 1527–1532; 10.1016/S0957-4166(00)80354-7 (1993).

30. Lemière, G. L., Willaert, J. J., Dommisse, R. A., Lepoivre, J. A. & Alderweireldt, F. C. Determination of the absolute configuration and enantiomeric purity of alcohols from the 13 C-NMR spectra of the corresponding MTPA esters. *Chirality* **2,** 175–184; 10.1002/chir.530020309 (1990).

31. Dale, J. A., Dull, D. L. & Mosher, H. S. alpha.-Methoxy-.alpha.-trifluoromethylphenylacetic acid, a versatile reagent for the determination of enantiomeric composition of alcohols and amines. *J. Org. Chem.* **34,** 2543–2549; 10.1021/jo01261a013 (1969).

32. Medema, M. H. *et al.* antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39,** W339-46; 10.1093/nar/gkr466 (2011).

33. Robbins, T., Kapilivsky, J., Cane, D. E. & Khosla, C. Roles of Conserved Active Site Residues in the Ketosynthase Domain of an Assembly Line Polyketide Synthase. *Biochemistry* **55,** 4476–4484; 10.1021/acs.biochem.6b00639 (2016).

34. He, H. Y., Tang, M. C., Zhang, F. & Tang, G. L. Cis -double bond formation by thioesterase and transfer by ketosynthase in FR901464 biosynthesis. *J. Am. Chem. Soc.* **136,** 4488–4491; 10.1021/ja500942y (2014).

35. Di Lorenzo, M., Stork, M., Naka, H., Tolmasky, M. E. & Crosa, J. H. Tandem heterocyclization domains in a nonribosomal peptide synthetase essential for siderophore biosynthesis in Vibrio anguillarum. *Biometals* **21,** 635–648; 10.1007/s10534-008-9149-4 (2008).

36. Marshall, C. G., Burkart, M. D., Keating, T. A. & Walsh, C. T. Heterocycle formation in vibriobactin biosynthesis: alternative substrate utilization and identification of a condensed intermediate. *Biochemistry* **40,** 10655–10663; 10.1021/bi010937s (2001).

37. Zhou, Y. *et al.* Iterative Mechanism of Macrodiolide Formation in the Anticancer Compound Conglobatin. *Chem. Biol.* **22,** 745–754; 10.1016/j.chembiol.2015.05.010 (2015).

38. Shaw-Reid, C. A. *et al.* Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of E. coli EntF catalyzes both elongation and cyclolactonization. *Chem. Biol.* **6,** 385–400; 10.1016/S1074-5521(99)80050-7 (1999).

39. Shaw-Reid, C. A. *et al.* Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of E. coli EntF catalyzes both elongation and cyclolactonization. *Chem. Biol.* **6,** 385–400 (1999).

40. Fu, J. *et al.* Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nature biotechnology* **30,** 440–446; 10.1038/nbt.2183 (2012).

41. Magrini, V., Creighton, C. & Youderian, P. Site-specific recombination of temperate *Myxococcus xanthus* phage Mx8: genetic elements required for integration. *J. Bacteriol.* **181,** 4050–4061; 10.1128/JB.181.13.4050-4061.1999 (1999).

42. Wang, H. *et al.* RecET direct cloning and Redαβ recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nat. Protoc.* **11,** 1175–1190; 10.1038/nprot.2016.054 (2016).

43. Jordan, M. A. Mechanism of action of antitumor drugs that interact with microtubules and tubulin. *Current medicinal chemistry. Anti-cancer agents* **2,** 1–17; 10.2174/1568011023354290 (2002).

44. Short, B. Acetylated microtubules let the ER slide. *J.Cell Biol* **190,** 280; 10.1083/jcb.1903iti2 (2010).

45. Lee, A. S. The ER chaperone and signaling regulator GRP78/BiP as a monitor of endoplasmic reticulum stress. *Methods* **35,** 373–381; 10.1016/j.ymeth.2004.10.010 (2005).

46. André, N. *et al.* Paclitaxel Induces Release of Cytochrome c from Mitochondria Isolated from Human Neuroblastoma Cells. *Cancer Res* **60,** 5349–5353 (2000).

47. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. Sect. A* **64,** 112–122; 10.1107/S0108767307043930 (2008).

48. Sheldrick, G. M. Crystal structure refinement with SHELXL. *Acta crystallographica. Section C, Structural chemistry* **71,** 3–8; 10.1107/S2053229614024218 (2015).

49. Hübschle, C. B., Sheldrick, G. M. & Dittrich, B. ShelXle: a Qt graphical user interface for SHELXL. *Journal of applied crystallography* **44,** 1281–1284; 10.1107/S0021889811043202 (2011).

50. Liu, H. & Naismith, J. H. A simple and efficient expression and purification system using two newly constructed vectors. *Protein Expr. Purif.* **63,** 102–111; 10.1016/j.pep.2008.09.008 (2009).

S 2 Supporting Information

Supporting Information

The Disorazole Z Family of Highly Potent Anticancer Natural Products from *Sorangium cellulosum*: Structure, Bioactivity, Biosynthesis, and Heterologous Expression

Previously published in:

Yunsheng Gao†[a,c,d], Joy Birkelbach†[a], Chengzhang Fu[a,d], Jennifer Herrmann[a], Herbert Irschik[b], Bernd Morgenstern[e], Kerstin Hirschfelder[a], Ruijuan Li[c], Youming Zhang[c], Rolf Jansen[b], Rolf Müller[a,d]

Microbiology Spectrum **2023** 11 (4), e00730-23

DOI: 10.1128/spectrum.00730-23

[a]Department of Microbial Natural Products, Helmholtz-Institute for Pharmaceutical Research Saarland, Helmholtz Centre for Infection Research and Department of Pharmacy at Saarland University, Saarbrücken, Germany

[b]Department of Microbial Drugs, Helmholtz Centre for Infection Research, Braunschweig, Germany

[c]Helmholtz International Lab for Anti-Infectives, Shandong University-Helmholtz Institute of Biotechnology, State Key Laboratory of Microbial Technology, Shandong University, Qingdao, China

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

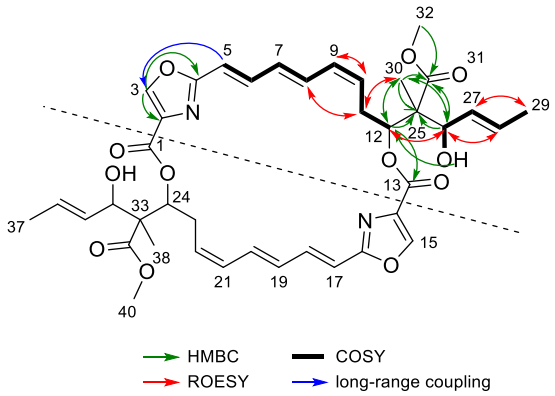
[e]Department of Inorganic Chemistry, Saarland University, Saarbrücken, Germany

S 2.1 Structure Elucidation

STRUCTURE ELUCIDATION OF DISORAZOLE Z1

NMR based structure elucidation of disorazole Z1 is explained in the main text.

Table S2. 1. *NMR data of disorazole Z1 (**3**) in acetone-d_6.*[a]



Pos.	δ_H, m (*J* [Hz])	δ_C, type	COSY[b]	ROESY[b]	HMBC[b,c]
1/13	-	159.76, C	-	-	12
2/14	-	135.84, C	-	-	3
3/15	8.53, s	145.16, CH	5	-	($^1J_{H,C}$ 213 Hz)
4/16	-	162.45, C	-	-	3, 5
5/17	6.19, d (15.5)	116.61, CH	3, 6	overlap with 7	7
6/18	6.81, m br.	137.77, CH	5, 7	overlap with 8	7, 8
7/19	6.17, dd (14.9, 11.1)	131.92, CH	6, 8	-	5, 9
8/20	6.77, dd (14.5, 11.9)	135.10, CH	7, 9	11	9, 10
9/21	6.11, t (11.2)	132.89, CH	8, 10	10	7, 11
10/22	5.69, dt br. (11.2, 9.5, 9.5)	131.19, CH	11a, 11b, 9	9, 11, 12	11, 12
11a/23a	2.71, m	30.31, CH_2	10, 11b, 12	8, 30	9, 10, 12
11b/23b	2.68, m	30.31, CH_2	10, 11a, 12	8, 10, 12, 26, 30	9, 10, 12
12/24	5.45, dd (9.4, 1.5)	76.28, CH	11a, 11b	10, 26, 30	9, 11, 30
25/33	-	56.26, C[d] 56.24	-	-	OH, 12, 26, 30
26/34	4.45, dd br. (4.9, 7.6[e])	75.65, CH[d] 75.52	OH, 27	OH, 11, 12, 28, 30	OH, 27, 28, 29, 30
27/35	5.65, ddquin (15.4, 7.2, 1.3)	131.69, CH[d] 131.66	26, 28, 29	29	OH, 26, 29
28/36	5.74, dqd (15.4, 6.5, 0.9)	128.84, CH	27, 29	26, 29	26, 27, 29
29/37	1.70, dd (6.5, 0.9)	18.06, CH_3	27, 28	27, 28	27, 28
30/38	1.36, s	13.66, CH_3	-	OH, 11, 26	12, 26
31/39	-	173.92, C	-	-	12, 26, 30, 32
32/40	3.60, s	51.97, CH_3	-	30	-
26/34-OH	4.16, d (4.9)	-	26	26, 30	-

[a] $^1H/^{13}C$ at 700/175 MHz; [b] For overview, correlations are given for northern part only, although they are valid for the southern part, too; [c] Carbon showing HMBC correlations to indicated protons; [d] signal doubling ratio 1:1; [e] 7.2 Hz after H/D exchange.

Relative configuration. The crystal structure determination shows the constitution of the cyclic molecule with the relative configuration of the stereo centers C-12, C-25 and C-26 (**Figure S2. 1**. **Crystal structure of 3-(EtOH)$_2$.**, **Table S2. 17**-**Table S2. 23**).

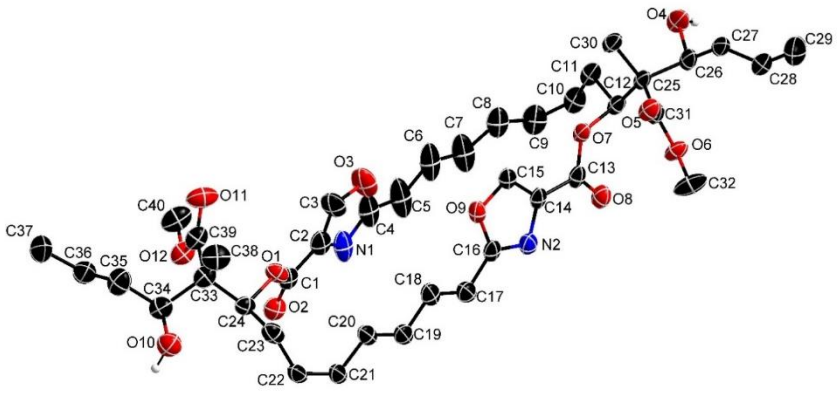


Figure S2. 1. Crystal structure of 3-(EtOH)$_2$. *3 = $C_{40}H_{46}N_2O_{12}$ in the crystal (displacement ellipsoids are drawn at 50% probability level). The C bonded hydrogen atoms are omitted for clarity. Both EtOH solvent molecules are squeezed by the program PLATON.*

In view of the relative configuration in the crystal, the stereo chemical NMR data, i.e. proton coupling constants and NOE correlations, were accessed critically. Indeed, with a dihedral angle of 179° the X-ray structure was not compatible with the strong NOE between methyl group C-30 and H-26 observed in solution and also not with the absence of a strong NOE between the methyl group C-30 and methine H-27 (**Table S2. 1**).

In the ^{13}C NMR spectra in acetone-d_6 a signal doubling of C-25 to C-27 revealed the presence of at least two rotamers in a ratio near 1:1 while the ^{1}H NMR signal of H-26 still appeared as a broad doublet after H/D exchange. Thus, the side chain conformation was explored using the "Conformational Search" module in HyperChem and finally calculated by PM3. The optimized conformation (**Figure S2. 2**. **Side chain of disorazol Z1 (3) after conformational search and PM3 optimisation (right side).**, side chain on the right side) with a dihedral angel of 66° between methyl group C-30 and H-26 now accounted for the observed NOEs (**Table S2. 1**) by distances of 2.4 or 2.6 Å between methyl C-30 and H-26 or H-11, respectively. Contrary to the conformation in the crystal (**Figure S2. 1**. **Crystal structure of 3-(EtOH)$_2$.**), the hydroxyl group is freely exposed in the predominant conformer in solution (**Table S2. 1**, the NOE 30/26 was also the most intensive in DMSO-d_6 and CD_3OD).

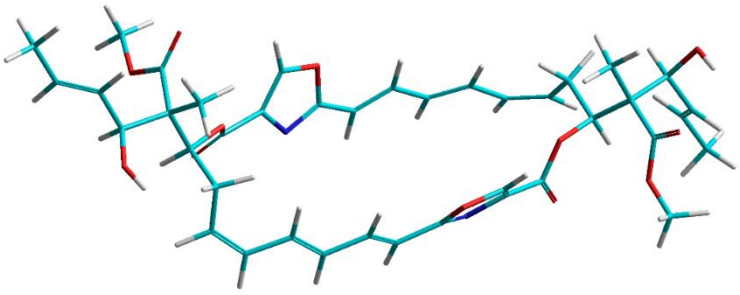


Figure S2. 2. Side chain of disorazol Z1 (3) after conformational search and PM3 optimisation (right side).

Absolute configuration. Since the X-ray analysis provided only the relative configuration of **3**, the absolute configuration was determined by chemical derivatization and NMR comparison of the α-methoxy-α-trifluoromethylphenyl acetic (MTPA) ester diastereomers, i.e. the application of the Mosher method based upon the ^{1}H and ^{13}C NMR chemical shift differences.[1–4] The (*S*)- and (*R*)-bis-methoxy-trifluoromethylphenyl acetic acid (MTPA) esters of **3** were easily prepared using an excess of the (*R*)- and (*S*)-MTPA chlorides in pyridine even without any catalyst.

Defined by convention the signs of the $\Delta\delta^{SR} = \delta_S - \delta_R$ values for protons and carbons residing in R^1 will be positive and those in R^2 will be negative in **Figure S2. 3**. **Mosher ester analysis of disorazole Z1 (3).**A. Both, the ^{1}H- and ^{13}C-NMR shift differences of the disorazole Z (*S*)- and (*R*)-Mosher esters (**Table S2. 2**) unambiguously indicated that the negative $\Delta\delta^{SR}$ values are located at the end of side-chain (C-27 to C-29), which therefore was assigned as R^2. Positive values were observed at positions 10, 11, 12, 30, and 31, all residing in the molecular part R^1 (**Figure S2. 3**. **Mosher ester analysis of disorazole Z1 (3).**B). Consequently, all asymmetric centers (12, 25, 26) of **3** are in the (*S*)-configuration as shown in (**Figure S2. 3**. **Mosher ester analysis of disorazole Z1 (3).**C).

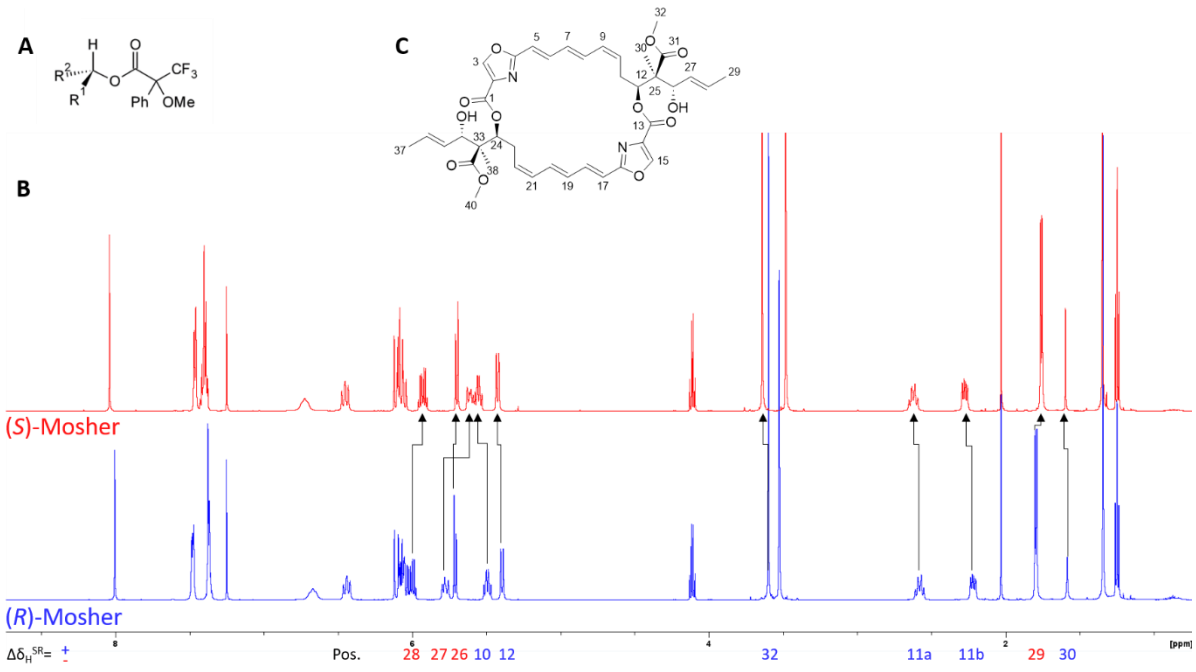


Figure S2. 3. Mosher ester analysis of disorazole Z1 (3). *A: Configuration of Mosher esters. B: ^{1}H NMR spectra and shift differences of (S)- and (R)-Mosher esters of **3**. C: Absolute configuration of **3**.*

Disorazole Z1 (**3**) isolated from the heterologous producer *M. xanthus* DK1622 :: km-int-Ptet-dis427 was characterised by ESI-HR-MS and NMR, accordingly. Comparison of ^{1}H and ^{13}C NMR data of **3** isolated from both producers revealed shift differences <0.1 ppm ($\Delta\delta_H^b$-δ_H^c) and <0.3 ppm ($\Delta\delta_C^b$-δ_C^c). Thus, both compounds are identical (**Table S2. 3**, **Figure S2. 17** and **Figure S2. 18**).

Table S2. 2. *^{1}H- and ^{13}C-NMR data (^{1}H 600 MHz; ^{13}C 150 MHz) of the (S)- and (R)-Mosher esters of **3** in $CDCl_3$.*

Pos.	δ_S -	δ_R =	$\Delta\delta^{SR}$	δ_{CS} -	δ_{CR} =	$\Delta\delta^{SR}$
10	5.57	5.51	+0.06	129.19	129.15	+0.04
11a	2.64	2.60	+0.04	29.47	29.25	+0.22
11b	2.29	2.24	+0.05			
12	5.44	5.41	+0.03	74.19	74.12	+0.07
25	-			53.42	53.54	-0.12
26	5.72	5.73	-0.01	80.02	80.05	-0.03
27	5.63	5.80	-0.17	123.86	123.92	-0.06
28	5.95	6.02	-0.07	135.32	135.98	-0.66
29	1.78	1.81	-0.03	17.95	18.03	-0.08
30	1.37	1.36	+0.01	16.29	15.85	+0.44
31	-			171.93	171.71	+0.22
32	3.66	3.62	+0.04	52.24	52.20	+0.04

Table S2. 3. *Comparison of ^{1}H and ^{13}C NMR data of **3** in acetone-d_6 isolated from* S. cellulosum *and* M. xanthus.*[a]*

Pos.	δ_H^b	δ_C^b	δ_H^c	δ_C^c	$\Delta\delta_H^b$-δ_H^c	$\Delta\delta_C^b$-δ_C^c
1/13	-	159.76	-	159.60	-	0.16
2/14	-	135.84	-	135.69	-	0.15
3/15	8.53	145.16	8.56	145.16	0.03	0.00
4/16	-	162.45	-	162.30	-	0.15
5/17	6.19	116.61	6.17	116.49	0.02	0.12
6/18	6.81	137.77	6.80	137.60	0.01	0.17
7/19	6.17	131.92	6.15	131.78	0.02	0.14
8/20	6.77	135.10	6.78	135.02	0.01	0.08
9/21	6.11	132.89	6.10	132.74	0.01	0.15
10/22	5.69	131.19	5.68	131.17	0.01	0.02
11a/23a	2.71	30.31	2.68	30.17	0.03	0.14
11b/23b	2.68	30.31	-	-	-	-
12/24	5.45	76.28	5.45	76.14	0.00	0.14
25/33	-	56.26/ 56.24[d]	-	56.13	-	0.13/ 0.1[d]

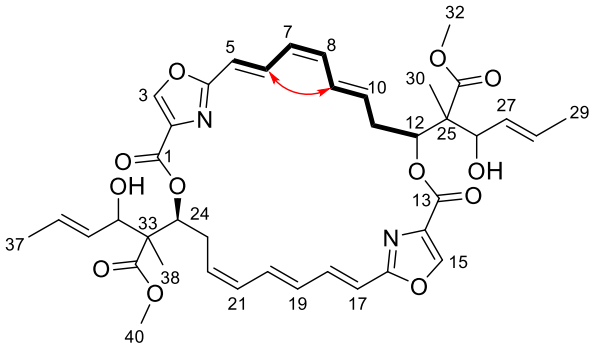
26/34	4.45	75.65/ 75.52[d]	4.43	75.60	0.02	0.05/ 0.08[d]
27/35	5.65	131.69/ 131.66[d]	5.64	131.58	0.01	0.11/ 0.08[d]
28/36	5.74	128.84	5.74	128.77	0.00	0.07
29/37	1.70	18.06	1.70	18.00	0.00	0.06
30/38	1.36	13.66	1.36	13.51	0.00	0.15
31/39	-	173.92	-	173.81	-	0.11
32/40	3.61	51.97	3.61	51.88	0.00	0.09
26/34-OH	4.16	-	4.22	-	0.08	-

[a] $^{1}H/^{13}C$ at 700/175 MHz; [b] isolated from S. cellulosum *So ce1875; [c] isolated from* M. xanthus *DK1622 :: km-int-Ptet-dis427; [d] signal doubling.*

STRUCTURE ELUCIDATION OF DISORAZOLE *CIS*/*TRANS*-ISOMERS Z2 AND Z3

In the analytical RP-HPLC two disorazole Z variants were observed eluting shortly in front and behind disorazole Z1 (**3**). Since ESI-HR-MS provided the elemental formulae $C_{40}H_{46}N_2O_{12}$ similar to the mayor compound **3** for both, they were recognized as isomers. Their ^{1}H NMR spectra showed two well separated pairs of oxazole signals at about 8.5 ppm, which revealed the asymmetry of their structures. In the NMR data of disorazole Z2 (**4**) (**Table S2. 4**) the signals of the side chains were nearly unchanged compared to **3** as well as the signals of the south part of the lactone ring (**Table S2. 1**, **Figure S2. 11**). In the $^{1}H,^{1}H$ COSY spectrum the full correlation sequence of H-5 to H-12 of the north part could be assigned unambiguously which characterised **4** as a $\Delta^{5,6}$-*trans*, $\Delta^{7,8}$-*cis*, $\Delta^{9,10}$-*trans* isomer. This stereo chemistry was indicated by the vicinal coupling constants of about 15 Hz for the *trans* double bonds and 11 Hz for the *cis* double bond. The $\Delta^{7,8}$-*cis* geometry was additionally shown by the strong NOE correlation between H-6 and H-9 in the $^{1}H,^{1}H$ ROESY spectrum.

Table S2. 4. *NMR data of disorazole Z2 (**4**) in acetone-d_6[a].*



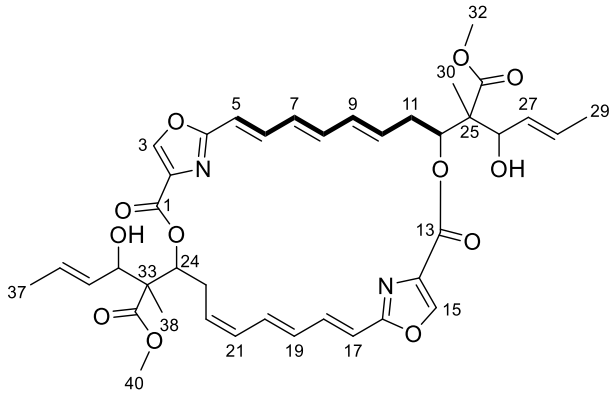
Pos.	δ_H, m (*J* [Hz])	COSY	ROESY	δ_C, type	HMBC[b]
1	-	-	-	159.56, C	24
2[c]	-	-	-	135.59, C	3
3	8.58, s	5	18, 20	145.19, CH	-
4	-	-	-	162.13, C	3, 5, 6
5	6.24, d (15.4)	3, 6	7	117.98, CH	6, 7
6[c]	6.81, dd (15.4, 11.7)	5, 7	9	131.64, CH	5, 8
7	5.93, t (11.4)	6, 8	5, 8	127.42, CH	5, 6, 9
8	6.08, t (11.0)	7, 9	7, 10	135.13, CH	6, 9, 10
9	6.49, dd (14.9, 11.6)	8, 10, 11b	6, 11b	130.44, CH	7, 11ab
10	5.83, ddd (15.2, 9.5, 6.1)	9, 11ab	8, 11a, 12	134.98, CH	8, 11ab, 12
11a	2.79, m	10, 11b	10, 11b, 12, 26	35.06, CH_2	9, 10, 12
11b	2.31, dt (12.8, 10.6)	9, 10, 11a	9, 11a		
12	5.42, dd (11.2, 1.3)	11ab	10, 11a, 26, 30	76.25, CH	11ab, 26, 30
13	-	-	-	159.91, C	12
14[c]	-	-	-	135.86, C	15
15	8.64, s	17	6, 9	145.21, CH	- (213.7 Hz)
16	-	-	-	162.73, C	15, 17, 18
17	6.11, d (15.8)	15, 18	-	115.81, CH	18, 19
18	6.96, dd (15.8, 11.0)	17, 19	20	138.83, CH	17, 19, 20
19	6.13, dd (14.9, 11.4[g])	18	-	130.83, CH	17, 21
20	6.87, dd (14.9, 11.6)	21	18, 23	135.94, CH	18, 22
21	6.13, t (11.4[g])	20, 22	22	132.59, CH	19, 20, 23
22[d]	5.70, m	21, 23	21, 23, 24	131.52, CH	20, 21, 23, 24
23	2.71, m	22, 24	20, 24, 34, 38	30.11, CH_2	21, 22, 24
24	5.46, dd (9.9, 1.5)	23	22, 23, 34, 38	76.31, CH	23, 34, 38
25	-	-	-	55.88, C	11b, 12, 26, 26-OH, 30
26	4.45, m (6.6, 4.4)	27	11a, 12, 28, 30	75.22, CH	12, 27, 28, 30
26-OH	4.13, d (4.8)	-	-	-	-
27[d]	5.61, ddq (15.4, 7.0, 1.5)	26, 28, 36	12, 29, 30	131.63, CH	24, 28
28[e]	5.74, m	27, 29	26, 29	128.77, CH	26, 27, 29
29	2.06, m	28	27	18.04, CH_3	27, 28
30	1.35, m	-	11ab, 12, 26, 27	13.46, CH_3	12, 26
31	-	-	-	173.91, C	12, 26, 30, 32

32[f]	3.64, s	-	12, 27, 28	51.98, CH_3	12, 26
33	-	-	-	56.33, C	24, 34, 34-OH, 38
34	4.43, t (6.2)	34-OH	23, 24, 36, 38	75.64, CH	24, 35, 36, 38
34-OH	4.15, d (4.2)	34	-	-	-
35[d]	5.63, ddq (15.8, 7.0, 1.5)	-	37, 38	131.59, CH	36, 37
36[e]	5.74, m	37	34, 37	128.85, CH	35, 34, 37
37	1.70, d (5.5)	36	35, 36	18.04, CH_3	35, 36
38	1.69, d (5.5)	-	23, 24, 34, 35	13.57, CH_3	24, 34
39	-	-		173.91, C	24, 34, 38, 40
40[f]	3.60, s	-	24, 35	51.96, CH_3	-

[a] 1H/13C at 600/150 MHz; [b] Carbon showing HMBC correlations to indicated protons; [c-f] interchangeable ^{13}C assignments; [g] after H/D exchange.

In the ^1H NMR spectrum of $\Delta^{9,10}$-*trans*-disorazole Z (**5**) the clearly separated signals of 17-, 18-, and 20-H with coupling constants of about 15 Hz as well as the shifts of 23-Ha and -Hb suggested a carbon skeleton similar to disorazole Z1 (**3**) for the southern part (**Table S2. 1**, **Figure S2. 11**). Although the signals of 21- and 22-H of the double bond were hidden in multiplets, the NOE between 20-H and the methylene group protons at C-23 unambiguously indicated the expected $\Delta^{21,22}$-*cis* double bond geometry. In the northern part a $^1H,^1H$ COSY sequence of 5-H to 10-H could be established. While 10-H was completely overlapping with H-12 and H-22, the large coupling constants of about 15 Hz in the signals of H-5, -6, -8 and -9 suggested the all-*trans* geometry for the triene. The assignment of the $\Delta^{9,10}$-*trans* bond was supported by a strong NOE between 9-H and the 11-Hb proton of the methylene group, which was shifted high-field to 2.29 ppm.

Table S2. 5. *NMR data of disorazole Z3 (**5**) in methanol-d_4.[a]*



Pos	δ_C	type	δ_H	m (*J* [Hz])	COSY	ROESY	H in HMBC
1	161.26	C					24
2	135.72	C					3

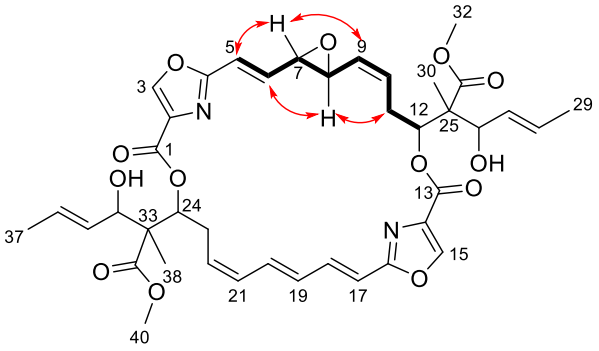
3	146.08	CH	8.32	s	5		-
4	163.00	C					3, 5, 6
5	117.49	CH	6.30	d (15.4)	3, 6	7	7
6	137.86	CH	6.81	dd (15.4, 11.1)	5, 7	8	5, 8
7	133.01	CH	6.03	br dd (14.4, 11.0)	6, 8	5[b]	5, 9[d]
8	137.35	CH	6.16	dd (14.8, 10.8)	7, 9	6, 10	6, 9
9	139.01	CH	6.06	br dd (15.3, 10.3)	8, 10	11b[b]	7, 8, 11ab
10	131.64	CH	5.56	br m	9, 11b	8, 11a	8, 11ab
11a	36.48	CH_2	2.71	m	11b, 12	10, 11b, 12, 26, 30	9, 10
11b			2.29	td (11.9, 11.2)	11a, 12	9, 11a, 30	
12	76.87	CH	5.57	dd (11.6, 3.2)	11ab	11a, 26, 30	11b, 26, 30
13	161.80	C					12
14	134.99	C					15
15	145.86	CH	8.43	s	17		-
16	163.35	C					15, 17, 18
17	116.88	CH	6.14	d (15.3)	15, 18		19
18	139.10	CH	6.96	dd (15.4, 11.3)	17, 19	20	17, 19, 20
19	133.01	CH	6.05	dd (14.3, 11.3)	18, 20	[c]	17, 18[d], 21
20	135.06	CH	6.71	br dd (14.8, 11.4)	19, 21	18, 23b	18, 21, 22
21	133.57	CH	6.03	br t (10.3)	20, 22	22[c]	19, 20, 23b
22	131.78	CH	5.56	t (10.5)	21, 23a	21, 23a	20, 23ab, 24
23a	30.93	CH_2	2.72	br d (3.4)	22, 23b, 24	20, 22, 23b, 38	21, 22, 24
23b			2.59	br dd (13.7, 6.6)	22, 23a, 24	23a, 24, 34, 38	
24	78.01	CH	5.52	dd (10.9, 1.2)	23ab	23a, 34, 38	23ab, 34, 38
25	56.58	C					11b, 12, 26, 30
26	75.91	CH	4.32	d (7.5)	27	11a, 12, 27, 28, 30	27, 28, 29, 30
27	131.27	CH	5.60	ddq (15.3, 7.6, 1.5)	26, 28	26, 29, 30	26, 28, 29
28	130.41	CH	5.72	br m	27, 29	26, 29	27, 29
29	18.17	CH_3	1.71	dd (6.5, 1.3)	28	27, 28	27, 28
30	13.55	CH_3	1.33	s		11ab, 12, 26, 27, 32	12, 26
31	175.21	C					12, 26, 30, 32
32	52.47	OCH_3	3.67	s		30	-
33	56.72	C					24, 34, 38
34	77.32	CH	4.30	d (7.7)	35	23b, 24, 35, 36 38	24, 35, 36, 37, 38
35	131.45	CH	5.68	ddq (15.1, 7.7, 1.2)	34, 36	38, 37, 34	34, 36, 37
36	130.30	CH	5.74	m	37, 35	37, 34	34, 36, 37
37	18.14	CH_3	1.74	dd (6.2, 1.3)	36	35, 36	35, 36
38	13.41	CH_3	1.45	s		23ab, 34, 24, 35, 40	24, 34
39	175.11	C					24, 34, 38, 40
40	52.35	OCH_3	3.59	m		38	-

[a] ^{1}H 700 MHz, ^{13}C 175 MHz; [b] 7-H close to 9-H; [c] H-19 and H-21 overlap; [d] C-7 and C-19 overlap.

STRUCTURE ELUCIDATION OF DISORAZOLE EPOXIDES Z4, Z5 AND Z6

A group of three more polar disorazole Z variants with retention times ~ 15-16 min also were isomers which according to their ESI-HR-MS-derived elemental formulas $C_{40}H_{46}N_2O_{13}$ contained one additional oxygen. While the NMR data of the mayor structural part of the main representative disorazole Z4 (**6**) (**Table S2. 6**) were similar to disorazole Z1 (**3**) (**Table S2. 1**), the COSY NMR spectrum presented a new sequence of correlations for the northern halve which connected completely visible ^{1}H signals between 5-H and 12-H. Thus, instead of a triene the isomer **6** contained a 7,8-epoxide (δ_C 60.4 and 56.5 ppm) flanked by a $\Delta^{5,6}$-*trans* and a $\Delta^{9,10}$-*cis* double bond. The proton signals of H-7 and H-8 (δ_H 3.25 and 3.76) had a small vicinal coupling constant of 1.7 Hz suggesting a *trans* configuration of the epoxide. This was supported by the strong NO effects between 7-H with 5-H and 9-H on one side and between 8-H and 6-H and 11-H_2 on the other side. Assuming the epoxide might be generated by an oxidation of **3** in a late stage of the biosynthesis, the epoxide ring should be directed to the outside of the bislactone ring and the absolute configuration of both epoxide carbons in **6** should be *R.*

Table S2. 6. *NMR data of disorazole Z4 (**6**) in methanol-d_4.*[a]



Pos	δ_C	type	δ_H	m (J [Hz])	COSY	ROESY	H in HMBC
1	160.61	C	-	-	-	-	24
2	135.59	C	-	-	-	-	3
3	145.92	CH	8.50	s	5	18, 20	
4	161.84	C		-	-	-	3, 5, 6
5	121.37	CH	6.46	d (15.8)	3, 6	7	6, 7
6	137.71	CH	6.06	dd (15.9, 9.4)	5, 7	7, 8, 15	5, 7, 8
7	60.42	CH	3.21	dd (9.4, 1.8)	6, 8	5, 6, 8, 9, 10	5, 9
8	56.49	CH	3.72	dd (9.7, 1.8)	7, 9	6, 7, 11, 15	6, 7, 10
9	131.08	CH	5.06	dd (10.7, 9.9)	8, 10	7, 8	7, 8, 11
10	134.04	CH	5.85	ddd (11.0, 10.0, 7.7)	9, 11	7, 11, 12	8, 11, 12
11	30.53	CH_2	2.66	m	10, 12	8, 10, 12, 26, 30	9, 10, 12

12	77.16	CH	5.47	dd (8.4, 4.0)		10, 11, 26, 30	9, 11, 26, 30
13	160.65	C	-	-	-	-	12
14	135.00	C	-	-	-	-	15
15	145.75	CH	8.48	s	-	6, 8, 26, 40	-
16	163.60	C	-	-	-	-	15, 17, 18
17	115.93	CH	6.18	br d (15.4)[c]	18	[c]	18, 19
18	139.36	CH	6.58	dd (15.6, 11.2)	17, 19	3, 20	19, 20
19	131.82[b]	CH	6.19	br dd (14.7, 11.4)[c]	18, 20	[c]	17, 20, 22
20	136.56	CH	6.67	dd (15.0, 11.4)	19	3, 18, 23	18, 19, 22
21	133.29	CH	6.17	br t (11.0)[c]	22	22[c]	19, 23
22	131.82[b]	CH	5.76	br t (9.5)	21, 23	21, 23, 24	23, 24
23	30.27	CH_2	2.63	m	22, 24	20, 22, 24, 34, 38	21, 22
24	77.77	CH	5.42	dd (8.4, 2.2)	23	22, 23, 34, 38	23, 34, 38
25	56.88	C	-	-	-	-	12, 26, 30
33	57.02	C	-	-	-	-	24, 34, 38
26, 34	76.56	CH	4.33	d (7.3)	27/35	11, 12, 23, 24, 27/35, 28/36, 30/38	12, 24, 28/36
27, 35	131.39	CH	5.61	br ddq (15.4, 7.7, 1.60)	26/34, 28/36 29/37	26/34, 29/37, 32/40, 38/35	26/34, 28/36, 29/37
28, 36	130.34	CH	5.72	dq (15.3, 6.6)	27/35, 29/37	26/34, 29/37, 32/40, 38/35	26/34, 27/35, 29/37
29, 37	18.12	CH_3	1.72	dt (6.2, 1.8)	27/35, 28/36	27/35, 28/36	27/35, 28/36
30, 38	13.68	CH_3	1.37	s	-	11/23, 12, 24, 26/34, 27/35, 28/36	12, 24, 26/34,
30, 38	13.61	CH_3	1.35	s	-	-	-
31, 39	175.16	C	-	-	-	-	12, 24, 26/34, 30/38, 32/40
32, 40	52.46	CH_3	3.61	s	-	26/34, 27/35, 28/36	-
32, 40	52.45	CH_3	3.61	s	-	-	-

[a] $^{1}H/^{13}C$ = 600/150 MHZ; [b] double intensity; [c] H-17, -19, -21 overlapping multiplets.

This configuration was proven by X-ray analysis of the 7,8-epoxi-disorazole Z (**6**) after crystallisation from ethanol. The crystals contained about two solvent molecules per one molecule of **6**, which are not presented in **Figure S2. 4**. The X-ray analysis provided a relative configuration shown in **Figure S2. 4** (**Table S2. 24**-**Table S2. 30**).

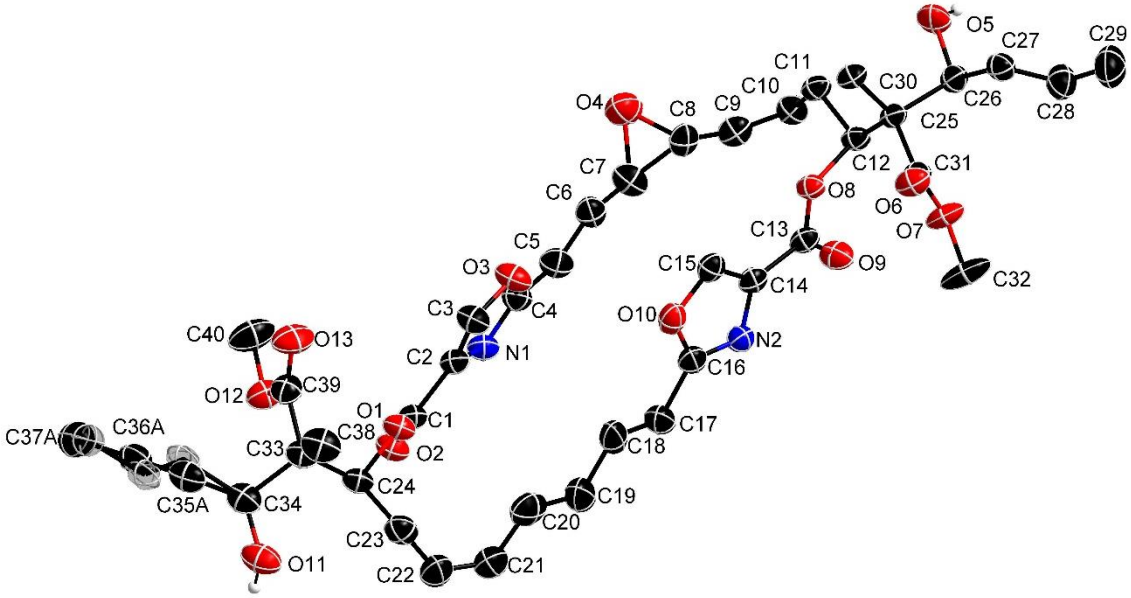
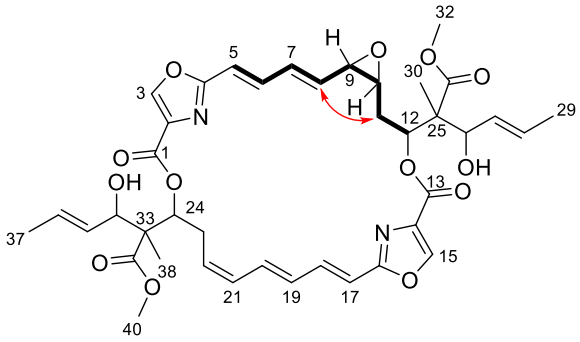


Figure S2. 4. Crystal structure of 6-(EtOH)$_2$. *6 = $C_{40}H_{46}N_2O_{13}$ in the crystal (displacement ellipsoids are drawn at 50 % probability level). The C bonded hydrogen atoms and both solvent ethanol molecules are omitted for clarity. The disordered atoms are highlighted with transparency.*

The second isomer with the elemental formula $C_{40}H_{46}N_2O_{13}$ was easily recognized as 9,10-epoxide disorazole Z (**7**), because the methylene protons H-11a at 2.76 ppm and the conspicuously high-field shifted H-11b at 1.50 ppm showed COSY correlation with the epoxide proton H-10 (3.27 ppm) (**Table S2. 7**). The vicinal coupling $J_{9,10}$ ~ 4 Hz was large for epoxides and indicated a *cis* geometry. As proof, a strong NOE between H-8 and 11-Hb was observed, very similar to the NOE between H-20 and the methylene group protons H-23ab, which characterised the *cis* configuration of the $\Delta^{21.22}$- double bond in the southern part of **7**. On the other side of the *cis*-epoxide three clearly visible dd-signals of H-8, H-7 and H-6, which included a vicinal coupling of ~ 15 Hz, showed the *trans* configuration of the diene in the northern part of **7**.

The relative configuration of disorazole Z5 (**7**) was unambiguously shown by the X-ray structure of **7** after crystallisation from ethanol (**Figure S2. 5**, **Table S2. 31**-**Table S2. 37**).

Table S2. 7. *NMR data of disorazole Z5 (**7**) in methanol-d_4.*[a]



Pos.	δ_H, m (*J* [Hz])	COSY	ROESY	δ_C	H in HMBC
1	-	-	-	160.66, C	24
2	-	-	-	135.58, C	-
3	8.42, br s	-	18, 20, 26/34, 30/38, 32/40	145.52, CH	-
2	-	-	-	163.00, C	3, 5, 10
5	6.14, d (15.8)	10	7	117.87, CH	7, 10
6	6.72, dd (15.4, 11.0)	5, 7	8	136.82, CH	7, 8
7	6.41, dd (15.0, 11.0)	8, 10	5, 9	136.16, CH	9[b], 17/5, 18, 22
8	5.68, dd (15.0, 9.9)	7, 9	10, 11b	135.46, CH	9, 10
9	3.34, dd (9.9, 4.2)	8, 10	7	58.91, CH	7, 8, 9, 11ab, 12
10	3.27, dt (10.5, 3.8)	9, 11a/11b	11a, 12	58.91, CH	7, 8, 9, 11ab, 12
11a	2.76, dd (14.7, 3.7)	10, 11b	10, 11b, 12, 26	32.11, CH_2	10, 12
11b	1.50, dt (14.8, 10.6)	10, 11a, 12	8, 11a		
12	5.57, d (10.6)	11b	10, 11a, 26/34	74.69, CH	12, 26, 30
13	-	-	-	160.98, C	12
14	-		-	135.85, C	-
15	8.37, s		8, 10, 26/34, 30/38, 32/40	145.63, CH	-
16	-	-	-	163.56, C	15, 17, 18
17[c]	6.15, d (15.4)	18	19[c]	116.28, CH	18, 19[c]
18	6.64, dd (15.6, 10.8)	17, 19c	-	139.06, CH	19[c], 20, 21
19[c]	6.16, dd (11.0, 15.0)	20, 22	17[c], 21	132.10, CH	18/20, 21[c]
20	6.61, dd (11.2, 6.4)	18, 19, 21[c]	23ab	136.22, CH	9[b], 17/5, 18, 22
21[c]	6.15, t (11.0)	18, 20	19[c], 22	133.36, CH	19[c], 23ab
22	5.72, m	19, 23ab	21[c], 23ab	131.69, CH	20, 23ab, 24
23	2.62, m	22, 24	20, 22, 24, 26/34	30.37, CH_2	21[c], 22, 24
24	5.41, dd (6.4, 4.6)	23ab	22, 23ab, 26/34, 27/35,	77.52, CH	23ab, 34, 38
25/33	-	-	-	56.90, C	12/24, 26/34, 30/38
25/33	-	-	-	56.86, C	12/24, 26/34, 30/38
26/34	4.32, d (7.3)	27/35	23ab, 24, 28/36, 30/38	76.70, CH	27/35, 28/36, 30/38
26/34	4.33, d (7.3)	27/35	11a, 12, 28/36, 30/38	76.51, CH	27/35, 28/36, 30/38
27/35	5.61, ddt (15.2, 7.9, 1.7)	26/34, 28/36, 29/37	24, 29/37	131.42, CH	28/36, 29/37
27/35	5.61, ddt (15.2, 7.9, 1.7)			131.37, CH	28/36, 29/37

28/36	5.74, m	27/35, 29/37	26/34, 29/37	130.42, CH	10, 26/34, 27/35, 29/37
28/36	5.74, m			130.31, CH	10, 26/34, 27/35
29/37	1.73, br s	27/35, 28/36	27/35	18.12, CH_3	27/35, 28/36
30/38	1.37, s	-	23ab, 24, 26/34, 27/35, 28/36	13.67, CH_3	12/24, 26/34
30/38	1.35, s	-	11ab, 12, 26/34, 27/35, 28/36	13.50, CH_3	12/24, 26/34
31/39	-	-	-	175.18, C	12/24, 26/34, 30/38, 32/40
31/39	-	-	-	175.04, C	12/24, 26/34, 30/38, 32/40
32/40	3.62, s	-	27/35, 28/36	52.49, CH	-
32/40	3.59, s	-	27/35, 28/36	52.44, CH	-

[a] ^1H 600 MHz, ^{13}C 150 MHz; Because the ^1H signals of the side chains closely overlap, they are assigned pairwise, and direct CH assignments may be interchanged; [b] C-20/C-18 overlap; [c] H-17, -19, -21 overlap.

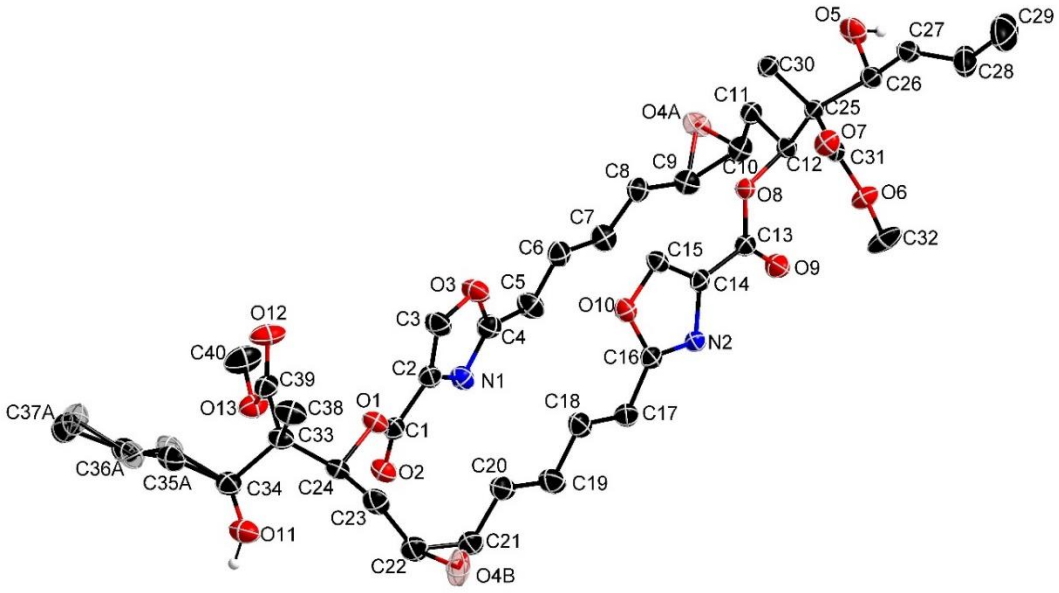


Figure S2. 5. Crystal structure of 7-(EtOH)$_2$. *7 = $C_{40}H_{46}N_2O_{13}$ in the crystal (displacement ellipsoids are drawn at 50 % probability level). The C bonded hydrogen atoms and both solvent ethanol molecules are omitted for clarity. The disordered atoms are highlighted with transparency.*

The NMR data of the third isomer disorazole Z6 (**8**) with the elemental formula $C_{40}H_{46}N_2O_{13}$ were nearly identical to disorazole Z1 (**3**) except those of one side chain. There, the former $\Delta^{27,28}$ double bond was replaced by an epoxide with characteristic ^1H (δ_H 2.87 and 2.92 ppm) and ^{13}C NMR shifts (δ_C 61.34 and 54.54 pppm) (**Table S2. 8**). The vicinal ^1H coupling constant $J_{27,28}$ = 2.1 Hz was small and thus suggested the trans configuration of the epoxide. This finding was supported by the ROESY correlations observed between methyl group (C-29) and H-27 and between H-28 and H-26. Considering, that the oxidation of the former double bond would occur from the least hindered side, i.e. opposite to the hydroxyl group, the absolute configuration of the epoxide **8** was suggested as 27*S*,28*R* (**Figure S2. 6**).

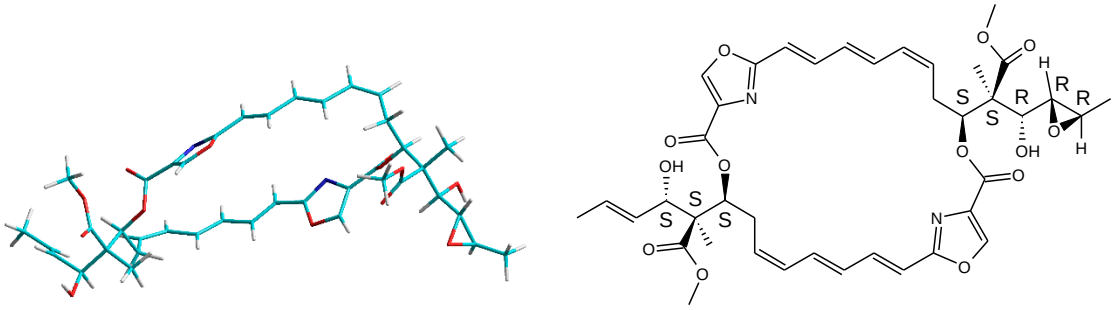
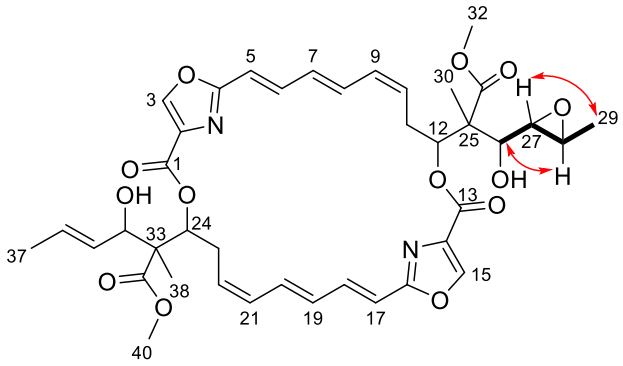


Figure S2. 6. Model and structure of disorazole Z6 (8) with 25R,26R,27R,28R-configuration.

Table S2. 8. *NMR data of disorazole Z6 (**8**) in methanol-d_4.*[a]



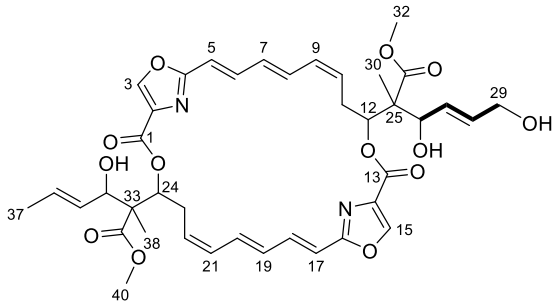
Pos.	δ_H, m (*J* [Hz])	COSY	ROESY	δ_C, type	H in HMBC
1/13	-	-	-	161.03, C	12/24
2/14	-	-	-	136.00, C	3/15, 5/17

2/14	-	-	-	135.92, C	3/15, 5/17
3	8.48, s	-	34, 40, 38	145.94, CH	-[c]
15	8.49, s	-	27, 30, 32	146.05, CH	-[c]
4/16	-	-	-	163.88, C	n.a.
4/16	-	-	-	163.85, C	n.a.
5/17	6.13, m	6/18	n.a.	116.59, CH	n.a.
5/17	-	-	n.a.	116.52, CH	n.a.
6/18	6.79, m	5/17	n.a.	139.41, CH	n.a.
6/18	-	-	n.a.	139.47, CH	n.a.
7/19	6.13, m		n.a.	132.58, CH	n.a.
7/19	-	-	n.a.	132.49, CH	n.a.
8/20	6.68, m	7/19, 9/21	n.a.	136.29, CH	n.a.
8/20	-	-	n.a.	136.34, CH	n.a.
9/21	6.13, m		n.a.	133.79, CH	n.a.
9/21	-	-	n.a.	133.63, CH	n.a.
10/22	5.71, m		n.a.	131.99, CH	12/24
10/22	-	-	-	131.72, CH	12/24
11a	2.70, m		30	31.13, CH_2	7/19, 10/22, 12
11b	2.64, m	10/22, 12	26, 30	-	-
12	5.46, d (10.1)	11	10, 11b, 26[b], 27, 30	77.36, CH	26
25	-	-	-	56.39, C	12, 26, 30
26	3.70, d (5.9)	27	11b[b], 12, 27, 28, 30	75.24, CH	12, 27, 30
27	2.87, dd (5.9, 2.1)	26, 28	15, 26, 29, 30	61.34, CH	26, 29
28	2.92, qd (5.9, 2.1)	27, 29	26[b], 29	54.54, CH	29
29	1.29, d (5.0)	28	26[b], 27, 28	17.86, CH_3	28
30	1.45, s	-	11ab, 12, 26, 27	14.41, CH_3	12, 26
31	-	-	-	175.21, C	12, 26, 30, 32/26
32	3.69, m	-	11b[b], 12, 27, 28, 30	52.99, CH_3	-
23	2.61, m	10/22, 24	24, 34, 38	30.79, CH_2	9/21, 24
24	5.44, d (10.2)	23	22, 23, 34, 38	77.70, CH	23, 34, 38
33	-	-	-	57.18, C	24, 34, 38
34	4.33, d (7.6)	35	3, 23, 24, 35, 36, 38	76.92, CH	24, 30/38
35	5.63, m		34, 37, 38	131.81, CH	36, 37
36	5.74, m	37	34, 37, 38, 40	130.59, CH	34, 35, 37
37	1.73, dd (6.3, 1.3)	10/22, 35, 36	35, 36	18.43, CH_3	35, 36
38	1.39, s	-	3, 23, 24, 34, 35	13.98, CH_3	24, 34
39	-		-	175.46, C	24, 34, 38, 40
40	3.69, s	-	3, 34, 35, 36	52.71, CH_3	-

[a] ^1H 600 MHz, ^{13}C 150 MHz; [b] methoxy-32 and H-26 ^1H signals overlap, however methoxy ROESY correlation signals are expected as very small, like those of methoxy group C-40; [c] carbon signals may be interchanged; n.a. not analyzed due to ^1H signal overlap.

Considerably more polar (t_R = 8.96 min), a fourth isomer disorazole Z7 (**9**) of the elemental formula $C_{40}H_{46}N_2O_{13}$ was identified by HPLC-UV-ESI-HR-MS. The NMR data of the isolated compound again only differed from disorazole Z1 (**3**) (**Table S2. 1**, **Figure S2. 11**) in one side chain, which instead of a methyl group ended in a primary unsaturated alcohol (δ_H 4.11, δ_c 63.0).

Table S2. 9. *NMR data disorazole Z7 (**9**) in methanol-d_4.[a]*



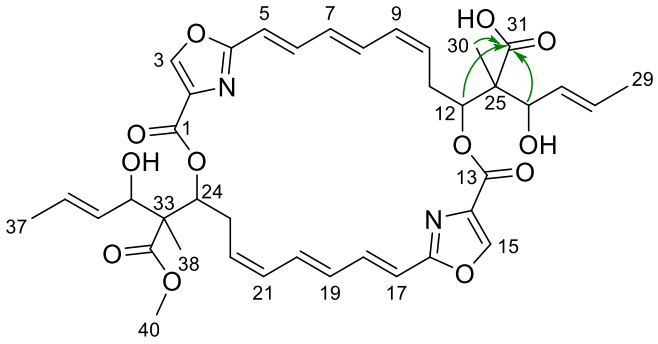
Pos.	δ_H, m (*J* [Hz])	COSY	δ_c, type	H in HMBC
1/13	-	-	160.82, C	12[b]
1/13	-	-	160.75, C	12[b]
2/14	-	-	135.70, C	-
2/14	-	-	135.68, C	-
3/15	8.52, s	-	145.75, CH	-
3/15	8.48, s	-	145.65, CH	-
4/16	-	-	163.57, C	3, 5[b]
5/17	6.12, d (15.0)	6/18	116.25, CH	7[b]
6/18	6.79, t br[d] (12.5)	5/17	139.15, CH	5, 7, 8[b]
7/19	6.13, dd (15.0, 9.5)	8/20	132.20, CH	9[b]
8/20	6.68, t br[e] (12.7)	7/19	136.06, CH	9, 10[b]
9/21	6.13, t (10.6)	10/22	133.33, CH	7, 11[b]
10/22	5.70, m	9/21	131.75, CH	8, 10, 11[b]
10/22	-	-	131.70, CH	8, 10, 11[b]
11/23	2.65, m	10/22, 12/24	30.65, CH_2	9, 10, 12[b]
11/23	-	-	30.51, CH_2	9, 10, 12[b]
12/24	5.45, d br (9.1)	11/23	77.43, CH	9[b], 11, 26, 30, 34
12/24	5.43, d br (9.1)	11/23	77.34, CH	9[b], 11, 26, 30, 34
25/33	-	-	56.89, C	12[b], 25[b], 26, 34
25/33	-	-	56.75, C	12[b], 25[b], 26, 34
26	4.45, d (5.5)	27, 28	75.87, CH	12, 27, 28
27	5.86, m (15.7, 5.4)[c]	26, 28	130.49, CH	26, 28, 29
28	5.86, m (15.3, 3.8)[c]	26, 29	134.18, CH	26, 27, 29
29	4.11, d (3.7)	26, 27	63.00, CH_2	27, 28

30	1.39, s	-	14.01, CH_3	12[b], 26
34	4.33, d (7.7)	35	76.66, CH	24, 35, 36, 38
35	5.63, ddq (15.4, 7.7, 1.5)	34, 36, 37	131.43, CH	34, 36, 37
36	5.74, dq (15.4, 6.2)	35, 37	130.34, CH	34, 35, 37
37	1.73, dd (6.2, 1.1)	35, 36	18.14, CH_3	35, 36
38	1.38, s	-	13.69, CH_3	12[b], 34
31/39	-	-	175.17, C	12[b], 26, 30[b], 32[b], 34
32/40	3.66, s	-	52.55, CH_3	-
32/40	3.62, s	-	52.44, CH_3	-

[a] ^1H 600 MHz, ^{13}C 150 MHz; [b] correlations are given for the northern part only, although they are valid for the southern part too; [c] from J-resolved NMR spectrum; [d] not observed in J-resolved NMR spectrum; [e] from J-resolved NMR spectrum 11.0, 15.6 Hz.

With a RP-HPLC retention time of 9.38 min, disorazole Z8 (**10**) was the second most polar variant. The elemental formula $C_{39}H_{44}N_2O_{12}$ showed a loss of a methylene compared to disorazole Z1 (**3**). The NMR data of **10** (**Table S2. 10**) were most similar to **3** for the bislactone ring and one side chain (C-33 - C-40) (**Table S2. 1**, **Figure S2. 11**). However, the ^1H and ^{13}C spectra contained only one methoxy signal which correlated in the HMBC spectrum with the ester carbon C-39 (δ_C 175.17). The respective carbon C-31 (δ_C 176.66) of the second side chain was assigned as carboxylic acid, which only had HMBC correlations with the oxymethines C-12 and C-26 and methyl group C-30.

Table S2. 10. *NMR data of disorazole Z8 (**10**) in methanol-d_4.[a]*



Pos.	δ_H	m	*J*[Hz]	COSY	δ_C, type	H in HMBC
1/13	-			-	160.85, C	12/24
1/13	-			-	160.72, C	12/24
2/14	-			-	135.90, C	3/15
2/14	-			-	135.68, C	3/15
3/15	8.52	s	-	-	145.62, CH	-
3/15	8.51	s	-	-	145.56, CH	-

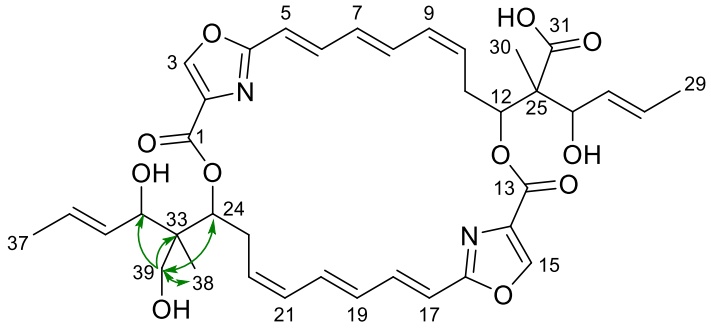
4/16	-			-	163.57, C	3/15, 5/17
4/16	-			-	163.49, C	3/15, 5/17
5/17	6.16	d	15.0	6/18	116.28, CH	7/19
5/17	-			6/18	116.20, CH	7/19
6/18	6.81	t	11.7 br.	5/17, 7/19	139.07, CH	7/19, 8/20
6/18	-			5/17, 7/19	139.14, CH	7/19, 8/20
7/19	6.17	dd	15.0, 11.0	6/18, 8/20	133.34, CH	5/17, 8/20, 9/21
7/19	-			6/18, 8/20	133.29, CH	5/17, 8/20, 9/21
8/20	6.72	dt	14.7, 11.7	7/19, 9/21	136.13, CH	10/22
8/20	-			7/19, 9/21	136.00, CH	10/22
9/21	6.16	t	11.0	8/20, 10/22	132.11, CH	5/17, 7/19, 11b/23b
9/21	-			8/20, 10/22	132.19, CH	5/17, 7/19, 11b/23b
10/22	5.74	m[b]	-	9/21, 11/23	131.95, CH	8/20, 11b/23b, 12/24
10/22	-		-	-	131.67, CH	8/20, 11b/23b, 12/24
11/23	2.73	m	-	10/22, 12/24	30.71, CH_2	7/19, 10/22
11/23	2.66	m	-	10/22, 12/24	30.50, CH_2	7/19, 10/22
12/24	5.47	d	10.3	11/23	77.41, CH	11ab/23ab, 30, 38
12/24	-			-	77.38, CH	11ab/23ab, 30, 38
25	-			-	56.31, C	26, 30
33	-			-	56.89, C	34, 38
26	4.41	d	7.3	27	76.34, CH	28, 30
34	4.37	d	7.7	35	76.64, CH	36, 38
35	5.67	ddq	15.4, 7.7, 1.6	34, 36, 37	131.59, CH	26/34, 28/36, 29/37
27	5.72	m[b]		26, 28	131.43, CH	26/34, 28/36, 29/37
28	5.79	m[b]		27, 29	129.99, CH	26, 29/37
36	5.79	m[b]		35, 37	130.33, CH	29/37, 34
29/37	1.77	d	6.6	28, 35, 36	18.16, CH_3	27/35, 28/36
30	1.38	s	-	-	13.91, CH_3	12/24, 26
31	-			-	176.66, C	12/24, 26, 30
38	1.42	s	-	-	13.69, CH_3	12/24, 34
39	-			-	175.17, C	12/24, 34, 38, 40
40	3.66	s	-	-	52.45, CH_3	-

*[a] ^1H 600 MHz, ^{13}C 150 MHz; [b] with 27, 35, 36, 10/22-H. For comparability reasons side chain numbering starts with position 25 and 33, respectively. Disorazole Z8 (**10**) misses CH_3-32.*

The variant disorazole Z9 (31-*O*-desmethyl-39-hydroxy-disorazole Z) (**11**) with a retention time of 9.35 min had the elemental formula $C_{38}H_{44}N_2O_{11}$, which showed the formal loss of C_2H_2O. The absence of any methoxy group in the NMR data of **11** (**Table S2. 11**) instantly indicated that both side chains were affected. The data of the lactone core were comparable to disorazole Z1 (**3**), although the symmetry of the molecule and the complete overlap of the respective NMR signals were lost (**Table S2. 11**, **Table S2. 1** and **Figure S2. 11**). In the side chain of the northern halve a

free carboxylic acid at C-31 (176.78 ppm) comparable to the mono ester of **10** was assigned from the NMR data. A second carboxyl carbon was also absent in the ^{13}C NMR spectrum of **11**. Instead, a new primary alcohol C-39 ($\delta_{H/C}$ 3.61/66.34 ppm) was assigned to its position from the HMBC correlations observed with lactone C-24, quaternary carbon C-33, secondary alcohol C-34, and methyl group C-38. The exchange of a carboxyl substituent against a primary alcohol group resulted in a high-field shift of 10 ppm for the quaternary carbon C-33 (δ_C 46.91 ppm).

Table S2. 11. *NMR data of disorazole Z9 (**11**) in methanol-d_4.*[a]



Pos	δ_H	m	*J* [Hz]	ROESY	δ_C, type	H in HMBC
1	-		-	-	161.70, C	24
2/14	-		-	-	136.10, C	3/15
2/14	-		-	-	135.95, C	3/15
3	8.53	s	-	34, 39	145.58, CH	-
4/16	-		-	-	163.61, C	3/15, 5/17
4/16	-		-	-	163.52, C	3/15, 5/17
5/17	6.13	m	-	-	116.20, CH	7/19
5/17	6.13	m	-	-	115.99, CH	7/19
6/18	6.84	m	-	-	139.31, CH	7/19, 8/20
6/18	6.84	m	-	-	139.17, CH	7/19, 8/20
7/19	6.13	m	-	-	133.27, CH	[b]
7/19	6.13	m	-	-	132.95, CH	[b]
8/20	6.78	dd	14.5, 11.9	-	136.45, CH	[b]
8/20	6.71	dd	14.3, 12.1	-	136.07, CH	[b]
9/21	6.13	m	-	-	132.78[b], CH	[b]
10/22	5.7	m	-	-	132.12[b], CH	[b]
27/35	5.7	m	-	-	132.05[b], CH 131.95[b], CH 131.88[b], CH 131.63[b], CH	[b] [b] [b] [b]
28/36	5.7	m	-	-	129.62, CH	26/34
28/36	5.7	m	-	-	129.96, CH	26/34

29/37	1.75	br d	6	-	18.18, CH_3	[b]
29/37	1.74	br d	6.6	-	18.15, CH_3	[b]
11a	2.72	br dt	13.7, 10.4	11b, 30	30.76, CH_2	12
11b	2.62	dd	13.9, 7.3	11a, 12, 26, 30		12
12	5.45	d	10.3	11b, 26, 30	77.41, CH	11ab, 26, 30
	-			-	56.29, C	12, 26, 30
26	4.37	d	7.0	11, 12, 30	76.42, CH	12, 30
30	1.35	s	-	11ab, 12, 26	14.06, CH_3	12, 26
31	-		-	-	176.78, C	12, 26, 30
13	-		-	-	161.03, C	12
15	8.47	s	-	26	145.58, CH	-
23a	2.88	br dt	13.4, 10.5	23b	30.40, CH_2	24
23b	2.53	dd	13.9, 7.3	23a, 24, 34, 38, 39		24
24	5.28	d	10.3	23b, 34, 38, 39	77.52, CH	23ab, 34, 38, 39
33	-			-	46.91, C	24, 34, 38, 39
34	4.13	d	6.6	23b, 24, 38, 39	76.90, CH	24, 38, 39
38	1.02	s	-	23b, 24, 34, 39	15.53, CH_3	24, 34, 39
39	3.61	m	-	23b, 24, 34, 38	66.34, CH_2	24, 34, 38

*[a] $^1H/^{13}C$ at 600/150 MHz; [b] overlapping signals. For comparability reasons side chain numbering starts with position 25 and 33, respectively. Disorazole Z9 (**11**) misses CH_3-32 and CH_3-40.*

Comparison of ^1H and ^{13}C NMR data of disorazol Z9 (**11**) isolated from *S. cellulosum* So ce1875 and *M. xanthus* DK1622 :: km-int-Ptet-dis427-gent-delF revealed a shift differences <0.1 ppm ($\Delta\delta_H^b$-δ_H^c). Thus, both compounds are identical (**Table S2. 12**, **Figure S2. 55**). Greater differences of carbon shifts can be explained due to their extraction from HMBC experiments.

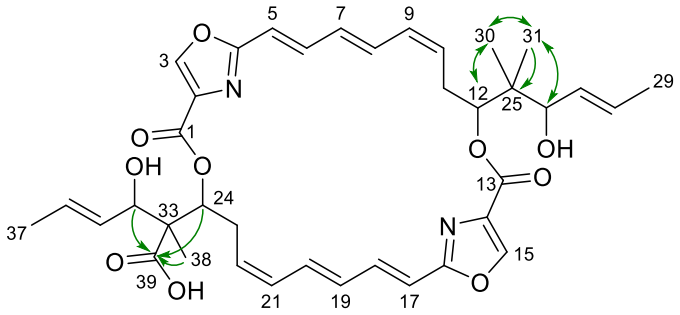
Table S2. 12. *Comparison of ^{1}H and ^{13}C NMR data of* ***11*** *in methanol-d_4 isolated from* S. cellulosum *and* M. xanthus.[a]

Pos.	δ_H[b]	δ_C[b]	δ_H[c]	δ_C[c]	$\Delta\delta_H$[b]-δ_H[c]	$\Delta\delta_C$[b]-δ_C[c]
2/14	-	136.1	-	136.02	-	0.08
2/14	-	135.95	-	136.02	-	0.07
4/16	-	163.61	-	163.53	-	0.08
4/16	-	163.52	-	163.53	-	0.01
5/17	6.13	116.2	6.12	116.02	0.01	0.18
5/17	6.13	115.99	6.12	116.02	0.01	0.03
6/18	6.84	139.31	6.82	139.21	0.02	0.10
6/18	6.84	139.17	6.82	139.21	0.02	0.04
7/19	6.13	133.27	6.12	133.17	0.01	0.10
7/19	6.13	132.95	6.12	133.17	0.01	0.22
8/20	6.78	136.45	6.78	136.48	0.00	0.03
8/20	6.71	136.07	6.73	136.36	0.02	0.29
9/21	6.13	132.78[d]	6.12	132.05	0.01	[d]
10/22	5.7	132.12[d]	5.72	131.97	0.02	[d]
27/35	5.7	132.05[d] 131.95[d] 131.88[d] 131.63[d]	5.72	131.97	0.02	[d] [d] [d] [d]
28/36	5.7	129.62	5.72	129.43	0.02	0.19
28/36	5.7	129.96	5.72	129.43	0.02	0.53
29/37	1.75	18.18	1.74	18.24	0.01	0.06
29/37	1.74	18.15	1.74	18.24	0.00	0.09
11a	2.72	30.76	2.75	31.04	0.03	0.28
11b	2.62	-	2.60	-	0.02	-
12	5.45 -	77.41 56.29	5.40 -	77.53 56.09	0.05 -	0.12 0.20
26	4.37	76.42	4.38	76.69	0.01	0.27
30	1.35	14.06	1.29	15.21	0.06	1.15
31	- -	176.78 161.03	- -	178.19 161.71	- -	1.41 0.68
15	8.47	145.58	8.57	145.74	0.10	0.16
23a	2.88	30.4	2.87	30.37	0.01	0.03
23b	2.53	-	2.53	-	0.00	-
24	5.28 -	77.52 46.91	5.28	77.43 46.89	0.00	0.09 0.02
34	4.13	76.9	4.13	76.77	0.00	0.13
38	1.02	15.53	1.01	15.47	0.01	0.06
39	3.61 -	66.34 161.7	3.61	66.21 161.71	0.00	0.14 0.01
3	8.53	145.58	8.57	145.74	0.04	0.16

[a] $^{1}H/^{13}C$ at 700/175 MHz; [b] isolated from S. cellulosum *So ce1875; [c] isolated from* M. xanthus *DK1622 :: km-int-Ptet-dis427-gent-delF; [d] overlapping signals.*

The elemental formula $C_{38}H_{44}N_2O_{10}$ of disorazole Z10 (39-*O*-desmethyl-25,25-dimethyl-disorazole Z) (**12**) with a retention time of 10.59 min in analytic HPLC showed the formal loss of $C_2H_2O_2$. In the NMR spectra both methoxy signals were absent. Similar to variants **10** and **11** only one carboxyl group C-39 at 176.75 ppm remained as free carboxylic acid on one side chain. Its position was retrieved from HMBC correlations with C-38, -34, and -24. The second side chain featured two geminal methyl groups C-30 and C-31 similar to disorazole A1 (**1**).[5] These were identified from their mutual HMBC correlations and positioned from their HMBC correlations with C-12, C-25 and C-26 together with their NOE interactions.

Table S2. 13. *NMR data of disorazole Z10 (**12**) in methanol-d_4.*[a]



Pos.	δ_H	m (J [Hz])	COSY	ROESY	δ_C	type	H to C HMBC
1	-	-	-	-	160.99	C	24
2	-	-	-	-	135.95	C	3
3	8.47	s	-	-	145.60	CH	
4/16	-	-	-	-	163.59 163.54	C C	3/15, 5/17 3/15, 5/17
5/17	6.13c	m	6/18	-	116.05 116.22	CH CH	7/19 7/19
6/18	6.83	m	5/17	-	139.27 139.15	CH	7/19, 8/20
7/19	6.13[c]	m	-	-	133.28[b] 132.88[b]	CH CH	
8/20	6.72	dd (15.5, 12.3)	9/21	11a/23a	136.26 136.07	CH CH	7/19, 9/21, 10/22 7/19, 9/21, 10/22
9/21	6.13[c]	m	8/20, 10/22	-	132.63[b] 132.13[b]	CH CH	
10	5.69[d]	m	9/21, 23ab	-	131.62[b]	CH	12, 26
11a	2.71	m	10, 12	8/20, 11b, 30, 31	29.81	CH_2	9, 10, 12
11b	2.43	dd (13.9, 7.3)	10, 12	11a, 26, 30, 31			
12	5.13	d (10.3)	11ab	11b, 26, 27/28, 30, 31	79.15	CH	11a/b, 26, 30, 31
13	-	-	-	-	161.75	C	12
14	-	-	-	-	136.07	C	15

15	8.56	s	-	-	145.53	CH	
22			9/21, 23ab	-	131.93[b]	CH	24, 34
23a	2.72	m	22, 24	8/20, 23b, 38	30.76	CH_2	21, 22, 24
23b	2.62	dd (13.6, 7.3)	22, 24	22, 23a, 24, 34, 38			
24	5.45	d (9.9)	23ab	23b, 34, 38	77.41	CH	34, 23a/b, 38
25	-	-	-	-	42.94	C	12, 26, 30, 31
26	3.89	d (7.7)	29	15, 11b, 12, 27/28, 30, 31	77.99	CH	12, 27, 28, 30, 31
27	5.59	ddq (15.4, 8.1, 1.5)	28, 29	11b, 12, 26, 29/37, 30, 31	131.93[b]	CH	
28	5.69[d]	m	27, 29	-	129.78	CH	
29/37	1.72	dd (6.6, 1.5)	26/34, 27/35, 28/36	27/35, 28/36	18.14 18.18	CH_3 CH_3	27/35, 28/36 27/35, 28/36
30	0.98	s	31	11a, 12, 26, 36, 31	19.66	CH_3	12, 26, 31
31	1.05	s	30	11a, 12, 26, 36, 30	19.44	CH_3	12, 26, 30
33	-	-	-	-	56.30	C	23b, 24, 34, 38
34	4.37	d (7.3)	37	23b, 24, 3, 38	76.39	CH	24, 35, 36, 38
35	5.76	m	36, 37	-	129.96	CH	
36	5.69[d]	m	37	-	131.62[b]	CH	
38	1.35	s	-	23ab, 24, 34, 27/35	14.02	CH_3	24, 34
39	-	-	-	-	176.75	C	24, 34, 38

*[a] $^1H/^{13}C$ 600/150 MHz; [b, c, d] interchangeable. For comparability reasons side chain numbering starts with position 25 and 33, respectively. Disorazole Z10 (**12**) misses CH_3-32 and CH_3-40.*

Comparison of ^1H and ^{13}C NMR data of disorazol Z10 (**12**) isolated from *S. cellulosum* So ce1875 and *M. xanthus* DK1622 :: km-int-Ptet-dis427-gent-delF revealed a shift differences <0.2 ppm ($\Delta\delta_H{}^b$-$\delta_H{}^c$). Thus, both compounds are identical (**Table S2. 14** and **Figure S2. 61**). Greater differences of carbon shifts can be explained due to their extraction from HMBC experiments.

Table S2. 14. *Comparison of ^{1}H and ^{13}C NMR data of* ***12*** *in methanol-d_4 isolated from* S. cellulosum *and* M. xanthus.[a]

Pos.	δH^{b}	δC^{b}	δH^{c}	$\delta C^{c, d}$	$\Delta\delta H^{b}$-δH^{c}	$\Delta\delta C^{b}$-$\delta C^{c, d}$
1	-	160.99	-	160.30	-	0.69
2	-	135.95	-	135.26	-	0.69
3	8.47	145.6	8.30	144.68	0.17	0.92
4/16	-	163.59	-	162.19	-	1.40
4/16	-	163.54	-	162.19	-	1.35
5/17	6.13	116.05	6.11	114.31	0.02	1.74
5/17	-	116.22	6.11	114.31	-	1.91
6/18	6.83	139.27	6.81	137.42	0.02	1.85
6/18	-	139.15	6.81	137.42	-	1.73
7/19	6.13	133.28	6.12	131.21	0.01	2.07
7/19	-	132.88	6.12	131.21	-	1.67
8/20	6.72	136.26	6.74	134.66	0.02	1.60
8/20	-	136.07	6.74	134.66	-	1.41
9/21	6.13	132.63	6.12	130.29	0.01	2.34
9/21	-	132.13	6.12	130.29	-	1.84
10	5.69	131.62	5.72	130.91	0.03	0.71
11a	2.71	29.81	2.71	28.48	0.00	1.34
11b	2.43	29.81	2.42	28.48	0.01	1.34
12	5.13	79.15	5.13	77.34	0.00	1.81
13	-	161.75	-	160.30	-	1.45
14	-	136.07	-	135.26	-	0.81
15	8.56	145.53	8.62	144.36	0.06	1.17
22	-	131.93	5.72	130.91	-	1.02
23a	2.72	30.76	2.79	29.92	0.07	0.85
23b	2.62	30.76	2.58	29.92	0.04	0.85
24	5.45	77.41	5.36	76.08	0.09	1.33
25	-	42.94	-	41.54	-	1.40
26	3.89	77.99	3.89	76.42	0.00	1.57
27	5.59	131.93	5.59	130.29	0.00	1.64
28	5.69	129.78	5.73	127.14	0.04	2.64
29/37	-	18.14	1.73	16.89	-	1.25
29/37	1.72	18.18	1.73	16.89	0.01	1.29
30	0.98	19.66	1.05	18.14	0.06	1.52
31	1.05	19.44	0.98	18.30	0.08	1.14
33	-	56.3	-	54.46	-	1.84
34	4.37	76.39	4.39	75.46	0.02	0.93
35	5.76	129.96	5.69	128.08	0.07	1.88
36	5.69	131.62	5.59	130.29	0.10	1.33
38	1.35	14.02	1.24	14.99	0.11	0.97
39	-	176.75	-	178.23	-	1.48

[a] $^{1}H/^{13}C$ at 700/175 MHz; [b] isolated from S. cellulosum *So ce1875; [c] isolated from* M. xanthus *DK1622 :: km-int-Ptet-dis427-gent-delF; [d] δ_C extracted from HMBC.*

S 2.2 Supplemental Tables

Table S2. 15. *DNA oligos used in this study.*

Name	Sequence (5' to 3')	Application
p15A-cm-MluCI-dis427-F	CCTGGCCGAGGATCGCGCCCACCGCCTGATCAGCGGCCCCCGCATGCTCCAGCAGCTCGCCAGGAGATCCCGCCAGCGGCAATTAGATCCGAAAACCCCAAGTTACG	for amplification of the p15A-cm-MluCI vector
p15A-cm-MluCI-dis427-R	AGTCAGGCTATGCCCTCGTTTTATAGACCACAAAATCGCGGTGTCAACGCGGTGCGCGAGAATACAAAGAGTCCGCGCAAAATTAGATCCTTTCTCCTCTTTAGATC	
pBR322-amp-BstXI-dis427-F	GCTGGGGAGCTCGGCCTTCCGGGAGGACTACGGCCTGCGCTACGCCTACGTCGCCGGATCCATGGTCGAGGGCATCTCGTCCAGCGAGCTGGAGATCCGAAAACCCCAAGTTACG	for amplification of the pBR322-amp-BstXI vector
pBR322-amp-BstXI-dis427-R	GGCGAGGTGAATGGCGGTGAGGGACGACGCGCACGCGGTGTCCACCGCCATGCAGGGCCCGTGCAGGTTGAGGAAATACGCCACATGATTGGGATCTAAAGAGGAGAAAGGATCT	
dis427-F3-F	CTGGGGAGCTCGGCCTTCCGGGAGGACTAC	for amplification of a BGC fragment
dis427-F3-R	GCTGTCCGCCGAGAGGTGGTTCCGTGTGGG	
p15A-cm-dis427-F	GAGCGGCTCGTTCGCCCTCGCGGAGCTCCGCGGCCGCGCCCCCGCGACGGACCGGTGGGTCCCACACGGAACCACCTCTCAGATCCGAAAACCCCAAGTTACG	for amplification of the p15A-cm vector
km-int-F	ACTAGTGCTTGGATTCTCACCAA	for amplification of the *km-int* cassette
km-int-R	CAACGCGGTGCGCGAGAATACAAAGAGTCCGCGCAAAATTAAGACGGACGAGGTGCTCACTCA	
Ptet-disA-R	ATGCCGATGATGGCAATGCCGTCCTGCTCTACGTGATTCCCGCTGCTCATAGATCCTTTCTCCTCTTTAGATC	for amplification of the *km-int-Ptet* cassette
Papr-disD-F	ATCACGCAGGTCGCAGGCACGAGCAGGGCTCTCTCATCTAGGAGGCGCTTACGCTCAGTGGAACGAGGTT	for amplification of the *Papr-disD* cassette
Papr-disD-R	CCTCCCATCCCTGGCGACTGCGACCCCTGCCCCGGAAACATGACGGCCTTCATAATCTGTACCTCCTTAAGTCAGCCAATCGACTGGCGAGC	
gent-delF-F	AGGCCTCGGCCGATCTCCTGGACCACCGCCTGCGCGCGCTGTCGCGGTAAGAAGGCACGAACCCAGTTGAC	for amplification of the *gent-delF* cassette
gent-delF-R	CTGAGCCTTTCGTTTTATTTGATGCCTGGAGATCCTTAAGATCCGTAACTTGGGGTTTTCGGATCTTTAGGTGGCGGTACTTGGGTC	
apr-Pvan-disA-F	AATTTTGCGCGGACTCTTTGTATTCTCGCGCACCGCGTTGACACCGCGATTCAGCCAATCGACTGGCGAGCGG	for amplification of the *apr-Pvan-disA* cassette
apr-Pvan-disA-R	GCCATGCCGATGATGGCAATGCCGTCCTGCTCTACGTGATTCCCGCTGCTCATATGCGTTTCCTCGCATCGTG	
dis427-chk01-F	GGATCGTGAGTACCTGGAGAAG	for colony PCR
dis427-chk01-R	GAGCGTCCGGGAGGTCGTGGGC	
dis427-chk02-F	GCAGAAGTACGTGGGCCTCAGC	
dis427-chk02-R	CGACGAGCAGGGTGGCGTATCC	
dis427-chk03-F	CGACCTCTACCTGAAGCACGATG	
dis427-chk03-R	CTGCACCTGATAATCGACCTGGC	
HisTEV-disF-F	TTACGATATCCCAACGACCGAAAACCTGTATTTTCAGGGCCTTGCAAAACTCATGTTGTCTCA	for cloning $disF_{427}$ into pHis-TEV
HisTEV-disF-R	CTTTGTTAGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCTACGCGCTCCGCTGGGCGATGG	

Table S2. 16. *Half-inhibitory concentrations (IC_{50}, mean ± SD) of **3** in comparison of **1** on selected human cancer cell lines.*

Cell line	Origin	disorazole A1 (**1**) IC_{50} [nM]	disorazole Z1 (**3**) IC_{50} [nM]
A549	lung carcinoma	0.35 ± 0.15	0.21 ± 0.11
HCT-116	colon carcinoma	0.24 ± 0.02	0.25
HepG2	hepatocellular carcinoma	0.28 ± 0.07	0.06
HL-60	promyelocytic leukemia	0.07 ± 0.01	nd
KB-3.1	cervix carcinoma	0.15 ± 0.02	0.07 ± 0.02
KB-V1	cervix carcinoma, MDR	0.17 ± 0.01	0.11 ± 0.02
U-2 OS	osteosarcoma	0.43 ± 0.16	0.09

nd: not determined; MDR: multi-drug resistant.

Table S2. 17. *Crystal data and structure refinement for sh3137_a_sq (**3**).*

Identification code	sh3137_a_sq	
Empirical formula	C44 H58 N2 O14	
Formula weight	838.92	
Temperature	133(2) K	
Wavelength	0.71073 Å	
Crystal system	Orthorhombic	
Space group	$P2_12_12_1$	
Unit cell dimensions	a = 8.5672(3) Å	a= 90°.
	b = 20.1022(6) Å	b= 90°.
	c = 25.8711(9) Å	g = 90°.
Volume	4455.5(3) Å3	
Z	4	
Density (calculated)	1.251 Mg/m3	
Absorption coefficient	0.093 mm-1	
F(000)	1792	
Crystal size	0.436 x 0.241 x 0.140 mm3	
Theta range for data collection	1.283 to 27.134°.	
Index ranges	-10<=h<=10, -21<=k<=25, -33<=l<=32	
Reflections collected	30846	
Independent reflections	9363 [R(int) = 0.0376]	
Completeness to theta = 25.242°	94.30%	
Absorption correction	Semi-empirical from equivalents	
Max. and min. transmission	0.7455 and 0.6616	
Refinement method	Full-matrix least-squares on F2	
Data / restraints / parameters	9363 / 2 / 495	
Goodness-of-fit on F2	1.053	

Final R indices [I>2sigma(I)]	R1 = 0.0586, wR2 = 0.1356
R indices (all data)	R1 = 0.0921, wR2 = 0.1606
Absolute structure parameter	-0.2(3)
Extinction coefficient	n/a
Largest diff. peak and hole	0.316 and -0.269 e.Å-3

Table S2. 18. *Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters ($Å^2$ x 10^3) for sh3137_a_sq (**3**). U(eq) is defined as one third of the trace of the orthogonalised U^{ij} tensor.*

	x	y	z	U(eq)		x	y	z	U(eq)
N(1)	5726(5)	4725(2)	2853(1)	54(1)	C(14)	5233(4)	4216(2)	617(1)	32(1)
N(2)	4427(4)	4759(2)	807(1)	39(1)	C(15)	5130(4)	3702(2)	952(1)	33(1)
O(1)	2985(3)	3983(2)	3736(1)	37(1)	C(16)	3856(4)	4559(2)	1244(2)	36(1)
O(2)	3959(4)	5035(2)	3755(1)	43(1)	C(17)	2913(5)	4912(3)	1611(2)	43(1)
O(3)	6092(4)	3700(3)	2539(1)	69(1)	C(18)	2360(5)	4633(3)	2042(2)	43(1)
O(4)	9420(4)	2849(2)	-1292(1)	55(1)	C(19)	1401(5)	4956(3)	2420(2)	43(1)
O(5)	4487(4)	2578(2)	-695(1)	44(1)	C(20)	905(5)	4661(3)	2853(2)	41(1)
O(6)	4898(3)	3611(2)	-975(1)	41(1)	C(21)	28(5)	4982(3)	3254(2)	45(1)
O(7)	6711(3)	3660(2)	8(1)	34(1)	C(22)	-248(5)	4745(3)	3725(2)	46(1)
O(8)	6178(4)	4751(2)	-132(1)	42(1)	C(23)	265(5)	4089(2)	3942(2)	42(1)
O(9)	4243(3)	3913(2)	1360(1)	36(1)	C(24)	1917(5)	4104(2)	4169(2)	34(1)
O(10)	165(4)	3948(2)	5131(1)	53(1)	C(25)	7149(4)	2997(2)	-743(1)	33(1)
O(11)	4675(4)	2947(2)	4444(2)	59(1)	C(26)	7789(5)	2997(3)	-1308(2)	43(1)
O(12)	4803(3)	4002(2)	4712(1)	43(1)	C(27)	6950(6)	2524(3)	-1657(2)	48(1)
C(1)	3910(4)	4474(3)	3575(2)	37(1)	C(28)	6173(8)	2692(3)	-2074(2)	63(2)
C(2)	4872(5)	4262(3)	3139(2)	46(1)	C(29)	5375(10)	2234(4)	-2434(2)	91(2)
C(3)	5084(6)	3652(3)	2948(2)	57(1)	C(30)	7566(5)	2348(2)	-456(2)	39(1)
C(4)	6427(5)	4381(4)	2504(2)	60(2)	C(31)	5371(5)	3016(2)	-795(1)	36(1)
C(5)	7383(6)	4606(4)	2094(2)	79(2)	C(32)	3225(6)	3667(3)	-1039(3)	65(2)
C(6)	7976(5)	4231(4)	1704(2)	75(2)	C(33)	2276(5)	3555(2)	4582(2)	39(1)
C(7)	8863(6)	4502(4)	1283(2)	73(2)	C(34)	1811(5)	3839(3)	5125(2)	45(1)
C(8)	9251(5)	4168(3)	852(2)	53(1)	C(35)	2252(6)	3376(3)	5557(2)	55(1)
C(9)	9922(5)	4492(3)	403(2)	54(1)	C(36)	3303(6)	3493(3)	5902(2)	56(1)
C(10)	9968(5)	4256(3)	-74(2)	46(1)	C(37)	3673(9)	3048(4)	6359(2)	78(2)
C(11)	9395(4)	3587(3)	-254(2)	42(1)	C(38)	1468(6)	2906(3)	4469(2)	54(1)
C(12)	7713(4)	3617(2)	-450(1)	33(1)	C(39)	4029(5)	3455(3)	4572(2)	42(1)
C(13)	6085(4)	4251(2)	126(1)	35(1)	C(40)	6478(6)	3966(3)	4686(2)	59(1)

Table S2. 19. *Bond lengths [Å] and angles [°] for sh3137_a_sq.*

N(1)-C(4)	1.284(7)	C(26)-C(27)	1.494(7)	C(10)-C(9)-H(9)	116.6	C(27)-C(26)-H(26)	108
N(1)-C(2)	1.396(6)	C(26)-H(26)	1	C(8)-C(9)-H(9)	116.6	C(25)-C(26)-H(26)	108
N(2)-C(16)	1.297(5)	C(27)-C(28)	1.312(6)	C(9)-C(10)-C(11)	126.9(4)	C(28)-C(27)-C(26)	125.2(5)
N(2)-C(14)	1.382(6)	C(27)-H(27)	0.95	C(9)-C(10)-H(10)	116.5	C(28)-C(27)-H(27)	117.4
O(1)-C(1)	1.333(5)	C(28)-C(29)	1.478(9)	C(11)-C(10)-H(10)	116.5	C(26)-C(27)-H(27)	117.4
O(1)-C(24)	1.467(5)	C(28)-H(28)	0.95	C(10)-C(11)-C(12)	112.0(4)	C(27)-C(28)-C(29)	126.4(6)
O(2)-C(1)	1.221(6)	C(29)-H(29A)	0.98	C(10)-C(11)-H(11A)	109.2	C(27)-C(28)-H(28)	116.8
O(3)-C(3)	1.369(6)	C(29)-H(29B)	0.98	C(12)-C(11)-H(11A)	109.2	C(29)-C(28)-H(28)	116.8
O(3)-C(4)	1.403(8)	C(29)-H(29C)	0.98	C(10)-C(11)-H(11B)	109.2	C(28)-C(29)-H(29A)	109.5
O(4)-C(26)	1.428(5)	C(30)-H(30A)	0.98	C(12)-C(11)-H(11B)	109.2	C(28)-C(29)-H(29B)	109.5
O(4)-H(14)	0.84	C(30)-H(30B)	0.98	H(11A)-C(11)-H(11B)	107.9	H(29A)-C(29)-H(29B)	109.5
O(5)-C(31)	1.190(5)	C(30)-H(30C)	0.98	O(7)-C(12)-C(11)	106.7(3)	C(28)-C(29)-H(29C)	109.5
O(6)-C(31)	1.346(5)	C(32)-H(32A)	0.98	O(7)-C(12)-C(25)	105.2(3)	H(29A)-C(29)-H(29C)	109.5
O(6)-C(32)	1.447(5)	C(32)-H(32B)	0.98	C(11)-C(12)-C(25)	115.3(4)	H(29B)-C(29)-H(29C)	109.5
O(7)-C(13)	1.340(5)	C(32)-H(32C)	0.98	O(7)-C(12)-H(12)	109.8	C(25)-C(30)-H(30A)	109.5
O(7)-C(12)	1.464(4)	C(33)-C(38)	1.507(7)	C(11)-C(12)-H(12)	109.8	C(25)-C(30)-H(30B)	109.5
O(8)-C(13)	1.208(5)	C(33)-C(39)	1.516(6)	C(25)-C(12)-H(12)	109.8	H(30A)-C(30)-H(30B)	109.5
O(9)-C(15)	1.368(5)	C(33)-C(34)	1.567(7)	O(8)-C(13)-O(7)	125.8(4)	C(25)-C(30)-H(30C)	109.5
O(9)-C(16)	1.375(5)	C(34)-C(35)	1.502(7)	O(8)-C(13)-C(14)	123.5(4)	H(30A)-C(30)-H(30C)	109.5
O(10)-C(34)	1.427(6)	C(34)-H(34)	1	O(7)-C(13)-C(14)	110.8(4)	H(30B)-C(30)-H(30C)	109.5
O(10)-H(29)	0.84	C(35)-C(36)	1.289(8)	C(15)-C(14)-N(2)	110.1(3)	O(5)-C(31)-O(6)	122.8(4)
O(11)-C(39)	1.207(6)	C(35)-H(35)	0.95	C(15)-C(14)-C(13)	128.7(4)	O(5)-C(31)-C(25)	126.7(4)
O(12)-C(39)	1.334(6)	C(36)-C(37)	1.516(7)	N(2)-C(14)-C(13)	121.2(4)	O(6)-C(31)-C(25)	110.6(4)
O(12)-C(40)	1.438(6)	C(36)-H(36)	0.95	C(14)-C(15)-O(9)	107.2(4)	O(6)-C(32)-H(32A)	109.5
C(1)-C(2)	1.459(6)	C(37)-H(37A)	0.98	C(14)-C(15)-H(15)	126.4	O(6)-C(32)-H(32B)	109.5
C(2)-C(3)	1.335(8)	C(37)-H(37B)	0.98	O(9)-C(15)-H(15)	126.4	H(32A)-C(32)-H(32B)	109.5
C(3)-H(3)	0.95	C(37)-H(37C)	0.98	N(2)-C(16)-O(9)	113.1(4)	O(6)-C(32)-H(32C)	109.5
C(4)-C(5)	1.415(7)	C(38)-H(38A)	0.98	N(2)-C(16)-C(17)	129.6(4)	H(32A)-C(32)-H(32C)	109.5
C(5)-C(6)	1.357(9)	C(38)-H(38B)	0.98	O(9)-C(16)-C(17)	117.3(4)	H(32B)-C(32)-H(32C)	109.5
C(5)-H(5)	0.95	C(38)-H(38C)	0.98	C(18)-C(17)-C(16)	123.0(5)	C(38)-C(33)-C(39)	109.6(4)
C(6)-C(7)	1.435(7)	C(40)-H(40A)	0.98	C(18)-C(17)-H(17)	118.5	C(38)-C(33)-C(24)	112.7(4)
C(6)-H(6)	0.95	C(40)-H(40B)	0.98	C(16)-C(17)-H(17)	118.5	C(39)-C(33)-C(24)	106.1(3)
C(7)-C(8)	1.345(8)	C(40)-H(40C)	0.98	C(17)-C(18)-C(19)	125.7(5)	C(38)-C(33)-C(34)	111.9(4)
C(7)-H(7)	0.95			C(17)-C(18)-H(18)	117.1	C(39)-C(33)-C(34)	108.4(4)
C(8)-C(9)	1.452(6)	C(4)-N(1)-C(2)	105.1(5)	C(19)-C(18)-H(18)	117.1	C(24)-C(33)-C(34)	107.8(4)
C(8)-H(8)	0.95	C(16)-N(2)-C(14)	104.6(4)	C(20)-C(19)-C(18)	123.6(5)	O(10)-C(34)-C(35)	109.6(4)
C(9)-C(10)	1.320(7)	C(1)-O(1)-C(24)	119.1(3)	C(20)-C(19)-H(19)	118.2	O(10)-C(34)-C(33)	108.5(4)
C(9)-H(9)	0.95	C(3)-O(3)-C(4)	104.3(4)	C(18)-C(19)-H(19)	118.2	C(35)-C(34)-C(33)	112.2(4)
C(10)-C(11)	1.506(7)	C(26)-O(4)-H(14)	109.5	C(19)-C(20)-C(21)	125.0(5)	O(10)-C(34)-H(34)	108.8
C(10)-H(10)	0.95	C(31)-O(6)-C(32)	114.0(4)	C(19)-C(20)-H(20)	117.5	C(35)-C(34)-H(34)	108.8
C(11)-C(12)	1.528(5)	C(13)-O(7)-C(12)	118.1(3)	C(21)-C(20)-H(20)	117.5	C(33)-C(34)-H(34)	108.8
C(11)-H(11A)	0.99	C(15)-O(9)-C(16)	105.0(3)	C(22)-C(21)-C(20)	126.6(5)	C(36)-C(35)-C(34)	125.3(5)
C(11)-H(11B)	0.99	C(34)-O(10)-H(29)	109.5	C(22)-C(21)-H(21)	116.7	C(36)-C(35)-H(35)	117.4

C(12)-C(25)	1.537(6)	C(39)-O(12)-C(40)	116.3(4)	C(20)-C(21)-H(21)	116.7	C(34)-C(35)-H(35)	117.4
C(12)-H(12)	1	O(2)-C(1)-O(1)	125.8(4)	C(21)-C(22)-C(23)	127.3(4)	C(35)-C(36)-C(37)	125.3(6)
C(13)-C(14)	1.467(5)	O(2)-C(1)-C(2)	123.1(4)	C(21)-C(22)-H(22)	116.4	C(35)-C(36)-H(36)	117.3
C(14)-C(15)	1.351(6)	O(1)-C(1)-C(2)	111.1(4)	C(23)-C(22)-H(22)	116.4	C(37)-C(36)-H(36)	117.3
C(15)-H(15)	0.95	C(3)-C(2)-N(1)	110.1(4)	C(22)-C(23)-C(24)	113.4(4)	C(36)-C(37)-H(37A)	109.5
C(16)-C(17)	1.433(6)	C(3)-C(2)-C(1)	129.1(5)	C(22)-C(23)-H(23A)	108.9	C(36)-C(37)-H(37B)	109.5
C(17)-C(18)	1.334(6)	N(1)-C(2)-C(1)	120.8(5)	C(24)-C(23)-H(23A)	108.9	H(37A)-C(37)-H(37B)	109.5
C(17)-H(17)	0.95	C(2)-C(3)-O(3)	107.9(5)	C(22)-C(23)-H(23B)	108.9	C(36)-C(37)-H(37C)	109.5
C(18)-C(19)	1.434(6)	C(2)-C(3)-H(3)	126.1	C(24)-C(23)-H(23B)	108.9	H(37A)-C(37)-H(37C)	109.5
C(18)-H(18)	0.95	O(3)-C(3)-H(3)	126.1	H(23A)-C(23)-H(23B)	107.7	H(37B)-C(37)-H(37C)	109.5
C(19)-C(20)	1.336(6)	N(1)-C(4)-O(3)	112.6(4)	O(1)-C(24)-C(23)	106.3(3)	C(33)-C(38)-H(38A)	109.5
C(19)-H(19)	0.95	N(1)-C(4)-C(5)	128.8(7)	O(1)-C(24)-C(33)	106.4(3)	C(33)-C(38)-H(38B)	109.5
C(20)-C(21)	1.435(6)	O(3)-C(4)-C(5)	118.5(6)	C(23)-C(24)-C(33)	115.4(4)	H(38A)-C(38)-H(38B)	109.5
C(20)-H(20)	0.95	C(6)-C(5)-C(4)	126.7(7)	O(1)-C(24)-H(24)	109.5	C(33)-C(38)-H(38C)	109.5
C(21)-C(22)	1.328(6)	C(6)-C(5)-H(5)	116.7	C(23)-C(24)-H(24)	109.5	H(38A)-C(38)-H(38C)	109.5
C(21)-H(21)	0.95	C(4)-C(5)-H(5)	116.7	C(33)-C(24)-H(24)	109.5	H(38B)-C(38)-H(38C)	109.5
C(22)-C(23)	1.500(7)	C(5)-C(6)-C(7)	123.5(7)	C(31)-C(25)-C(12)	109.7(3)	O(11)-C(39)-O(12)	122.9(4)
C(22)-H(22)	0.95	C(5)-C(6)-H(6)	118.3	C(31)-C(25)-C(30)	107.0(3)	O(11)-C(39)-C(33)	124.9(4)
C(23)-C(24)	1.532(6)	C(7)-C(6)-H(6)	118.3	C(12)-C(25)-C(30)	112.1(3)	O(12)-C(39)-C(33)	112.2(4)
C(23)-H(23A)	0.99	C(8)-C(7)-C(6)	124.8(6)	C(31)-C(25)-C(26)	105.5(3)	O(12)-C(40)-H(40A)	109.5
C(23)-H(23B)	0.99	C(8)-C(7)-H(7)	117.6	C(12)-C(25)-C(26)	110.6(3)	O(12)-C(40)-H(40B)	109.5
C(24)-C(33)	1.566(6)	C(6)-C(7)-H(7)	117.6	C(30)-C(25)-C(26)	111.6(4)	H(40A)-C(40)-H(40B)	109.5
C(24)-H(24)	1	C(7)-C(8)-C(9)	122.6(6)	O(4)-C(26)-C(27)	110.9(4)	O(12)-C(40)-H(40C)	109.5
C(25)-C(31)	1.530(5)	C(7)-C(8)-H(8)	118.7	O(4)-C(26)-C(25)	108.4(3)	H(40A)-C(40)-H(40C)	109.5
C(25)-C(30)	1.544(6)	C(9)-C(8)-H(8)	118.7	C(27)-C(26)-C(25)	113.3(4)	H(40B)-C(40)-H(40C)	109.5
C(25)-C(26)	1.563(5)	C(10)-C(9)-C(8)	126.7(5)	O(4)-C(26)-H(26)	108		

Table S2. 20. *Anisotropic displacement parameters ($Å^2$ x 10^3) for sh3137_a_sq (**3**). The anisotropic displacement factor exponent takes the form: $-2\pi^2[h^2 a^{*2}U^{11} + ... + 2 h k a^* b^* U^{12}]$.*

	U11	U22	U33	U23	U13	U12
N(1)	36(2)	91(3)	35(2)	-3(2)	-6(2)	-21(2)
N(2)	34(2)	51(2)	32(2)	2(2)	-3(1)	4(2)
O(1)	26(1)	42(2)	44(2)	-8(1)	4(1)	-2(1)
O(2)	34(2)	49(2)	47(2)	1(2)	-4(1)	-5(2)
O(3)	33(2)	118(4)	55(2)	-14(2)	12(2)	0(2)
O(4)	37(2)	86(3)	43(2)	22(2)	19(1)	18(2)
O(5)	35(2)	51(2)	45(2)	17(2)	6(1)	4(2)
O(6)	24(1)	47(2)	51(2)	14(2)	0(1)	7(1)
O(7)	25(1)	47(2)	29(1)	12(1)	6(1)	7(1)
O(8)	43(2)	49(2)	35(2)	11(1)	4(1)	0(2)
O(9)	26(1)	55(2)	27(1)	3(1)	1(1)	-2(1)
O(10)	45(2)	59(2)	54(2)	4(2)	6(2)	1(2)
O(11)	38(2)	45(2)	94(3)	-4(2)	-5(2)	5(2)
O(12)	30(2)	43(2)	55(2)	6(2)	-5(1)	-9(1)

C(1)	18(2)	55(3)	37(2)	1(2)	-7(2)	-2(2)
C(2)	25(2)	71(4)	41(2)	0(2)	-7(2)	-2(2)
C(3)	35(2)	81(4)	53(3)	-9(3)	11(2)	5(3)
C(4)	29(2)	111(5)	41(2)	-7(3)	-7(2)	-25(3)
C(5)	40(3)	155(7)	43(3)	-12(3)	-2(2)	-41(4)
C(6)	27(2)	159(7)	39(2)	1(3)	-5(2)	-30(3)
C(7)	39(3)	134(6)	46(3)	-5(3)	0(2)	-43(3)
C(8)	26(2)	88(4)	45(2)	6(3)	-5(2)	-21(2)
C(9)	30(2)	78(4)	55(3)	10(3)	1(2)	-22(3)
C(10)	28(2)	67(3)	45(2)	19(2)	8(2)	-10(2)
C(11)	21(2)	66(3)	39(2)	22(2)	7(2)	2(2)
C(12)	24(2)	44(3)	32(2)	16(2)	5(2)	5(2)
C(13)	22(2)	54(3)	28(2)	10(2)	-2(1)	-2(2)
C(14)	19(2)	52(3)	27(2)	4(2)	-5(1)	-1(2)
C(15)	21(2)	51(3)	28(2)	2(2)	-1(1)	-2(2)
C(16)	23(2)	56(3)	30(2)	2(2)	-6(2)	1(2)
C(17)	37(2)	54(3)	38(2)	-5(2)	-4(2)	6(2)
C(18)	26(2)	69(3)	36(2)	-9(2)	-6(2)	3(2)
C(19)	31(2)	61(3)	36(2)	-8(2)	-7(2)	11(2)
C(20)	26(2)	61(3)	37(2)	-17(2)	-10(2)	6(2)
C(21)	35(2)	70(3)	32(2)	-14(2)	-9(2)	18(2)
C(22)	35(2)	67(3)	36(2)	-16(2)	-3(2)	18(2)
C(23)	29(2)	53(3)	44(2)	-11(2)	2(2)	0(2)
C(24)	28(2)	34(2)	39(2)	-4(2)	-2(2)	0(2)
C(25)	27(2)	44(3)	27(2)	15(2)	7(1)	7(2)
C(26)	37(2)	64(3)	27(2)	16(2)	10(2)	14(2)
C(27)	52(3)	61(3)	30(2)	14(2)	9(2)	26(3)
C(28)	88(4)	60(4)	41(2)	6(2)	-8(3)	30(3)
C(29)	129(7)	93(5)	53(3)	-9(3)	-28(4)	31(5)
C(30)	33(2)	47(3)	36(2)	16(2)	5(2)	10(2)
C(31)	29(2)	50(3)	27(2)	15(2)	5(2)	8(2)
C(32)	28(2)	59(4)	107(4)	19(3)	-9(3)	10(2)
C(33)	35(2)	34(2)	47(2)	6(2)	-3(2)	-7(2)
C(34)	39(2)	45(3)	51(3)	7(2)	4(2)	-2(2)
C(35)	51(3)	61(4)	53(3)	15(3)	6(2)	-7(3)
C(36)	51(3)	69(4)	49(3)	4(3)	7(2)	14(3)
C(37)	88(5)	101(5)	44(3)	11(3)	6(3)	29(4)
C(38)	42(3)	45(3)	76(3)	-3(3)	0(2)	-16(2)
C(39)	33(2)	46(3)	48(2)	8(2)	-4(2)	-7(2)
C(40)	34(2)	53(3)	89(4)	9(3)	-11(2)	-12(2)

Table S2. 21. *Hydrogen coordinates (x 10^4) and isotropic displacement parameters (Å^2 x 10^3) for sh3137_a_sq (**3**).*

	x	y	z	U(eq)		x	y	z	U(eq)
H(14)	9897	3099	-1499	82	H(27)	6981	2065	-1569	57
H(29)	-38	4281	5315	79	H(28)	6119	3153	-2152	76
H(3)	4618	3255	3074	68	H(29A)	5879	2256	-2774	137
H(5)	7635	5066	2091	95	H(29B)	4277	2363	-2467	137
H(6)	7794	3765	1710	90	H(29C)	5439	1778	-2301	137
H(7)	9201	4951	1312	87	H(30A)	7116	1968	-641	58
H(8)	9078	3701	841	64	H(30B)	7143	2364	-104	58
H(9)	10375	4917	457	65	H(30C)	8703	2299	-441	58
H(10)	10409	4539	-329	56	H(32A)	2858	3318	-1274	97
H(11A)	9461	3266	35	50	H(32B)	2972	4104	-1184	97
H(11B)	10080	3426	-536	50	H(32C)	2713	3616	-703	97
H(12)	7566	4022	-670	40	H(34)	2353	4274	5178	54
H(15)	5587	3275	912	40	H(35)	1712	2964	5578	66
H(17)	2671	5365	1545	52	H(36)	3893	3891	5867	67
H(18)	2625	4181	2103	52	H(37A)	3077	2634	6330	117
H(19)	1102	5404	2360	51	H(37B)	3389	3276	6680	117
H(20)	1152	4204	2898	49	H(37C)	4792	2947	6362	117

H(21)	-397	5406	3176	55	H(38A)	340	2959	4516	81
H(22)	-838	5023	3949	55	H(38B)	1855	2562	4705	81
H(23A)	223	3750	3665	51	H(38C)	1683	2773	4111	81
H(23B)	-476	3953	4216	51	H(40A)	6791	3772	4354	88
H(24)	2129	4553	4319	40	H(40B)	6868	3688	4969	88
H(26)	7662	3455	-1452	51	H(40C)	6917	4415	4717	88

Table S2. 22. *Torsion angles [°] for sh3137_a_sq (3).*

C(24)-O(1)-C(1)-O(2)	-0.3(6)	O(8)-C(13)-C(14)-N(2)	2.3(6)	O(4)-C(26)-C(27)-C(28)	-118.9(5)
C(24)-O(1)-C(1)-C(2)	-179.4(3)	O(7)-C(13)-C(14)-N(2)	-177.0(3)	C(25)-C(26)-C(27)-C(28)	119.0(5)
C(4)-N(1)-C(2)-C(3)	-0.4(5)	N(2)-C(14)-C(15)-O(9)	0.3(4)	C(26)-C(27)-C(28)-C(29)	177.9(6)
C(4)-N(1)-C(2)-C(1)	-179.9(4)	C(13)-C(14)-C(15)-O(9)	178.4(4)	C(32)-O(6)-C(31)-O(5)	-0.3(6)
O(2)-C(1)-C(2)-C(3)	170.4(5)	C(16)-O(9)-C(15)-C(14)	-0.1(4)	C(32)-O(6)-C(31)-C(25)	179.5(4)
O(1)-C(1)-C(2)-C(3)	-10.4(6)	C(14)-N(2)-C(16)-O(9)	0.4(4)	C(12)-C(25)-C(31)-O(5)	-130.7(4)
O(2)-C(1)-C(2)-N(1)	-10.3(6)	C(14)-N(2)-C(16)-C(17)	-179.9(4)	C(30)-C(25)-C(31)-O(5)	-8.9(6)
O(1)-C(1)-C(2)-N(1)	168.9(3)	C(15)-O(9)-C(16)-N(2)	-0.2(4)	C(26)-C(25)-C(31)-O(5)	110.1(5)
N(1)-C(2)-C(3)-O(3)	0.8(5)	C(15)-O(9)-C(16)-C(17)	-179.9(3)	C(12)-C(25)-C(31)-O(6)	49.4(4)
C(1)-C(2)-C(3)-O(3)	-179.9(4)	N(2)-C(16)-C(17)-C(18)	177.7(4)	C(30)-C(25)-C(31)-O(6)	171.3(3)
C(4)-O(3)-C(3)-C(2)	-0.8(5)	O(9)-C(16)-C(17)-C(18)	-2.6(6)	C(26)-C(25)-C(31)-O(6)	-69.7(5)
C(2)-N(1)-C(4)-O(3)	-0.1(5)	C(16)-C(17)-C(18)-C(19)	-179.3(4)	O(1)-C(24)-C(33)-C(38)	-82.5(4)
C(2)-N(1)-C(4)-C(5)	176.3(5)	C(17)-C(18)-C(19)-C(20)	-178.3(4)	C(23)-C(24)-C(33)-C(38)	35.1(5)
C(3)-O(3)-C(4)-N(1)	0.5(5)	C(18)-C(19)-C(20)-C(21)	175.9(4)	O(1)-C(24)-C(33)-C(39)	37.5(4)
C(3)-O(3)-C(4)-C(5)	-176.3(4)	C(19)-C(20)-C(21)-C(22)	-167.3(5)	C(23)-C(24)-C(33)-C(39)	155.1(4)
N(1)-C(4)-C(5)-C(6)	-173.0(5)	C(20)-C(21)-C(22)-C(23)	-1.0(8)	O(1)-C(24)-C(33)-C(34)	153.4(3)
O(3)-C(4)-C(5)-C(6)	3.2(8)	C(21)-C(22)-C(23)-C(24)	85.1(6)	C(23)-C(24)-C(33)-C(34)	-88.9(4)
C(4)-C(5)-C(6)-C(7)	176.2(5)	C(1)-O(1)-C(24)-C(23)	111.5(4)	C(38)-C(33)-C(34)-O(10)	-60.3(5)
C(5)-C(6)-C(7)-C(8)	-168.7(6)	C(1)-O(1)-C(24)-C(33)	-125.0(4)	C(39)-C(33)-C(34)-O(10)	178.7(4)
C(6)-C(7)-C(8)-C(9)	169.6(5)	C(22)-C(23)-C(24)-O(1)	-85.5(4)	C(24)-C(33)-C(34)-O(10)	64.3(4)
C(7)-C(8)-C(9)-C(10)	-161.1(5)	C(22)-C(23)-C(24)-C(33)	156.8(4)	C(38)-C(33)-C(34)-C(35)	61.0(5)
C(8)-C(9)-C(10)-C(11)	-2.7(8)	O(7)-C(12)-C(25)-C(31)	48.5(4)	C(39)-C(33)-C(34)-C(35)	-60.0(5)
C(9)-C(10)-C(11)-C(12)	93.2(5)	C(11)-C(12)-C(25)-C(31)	165.8(3)	C(24)-C(33)-C(34)-C(35)	-174.5(4)
C(13)-O(7)-C(12)-C(11)	103.8(4)	O(7)-C(12)-C(25)-C(30)	-70.2(4)	O(10)-C(34)-C(35)-C(36)	-126.8(6)
C(13)-O(7)-C(12)-C(25)	-133.2(3)	C(11)-C(12)-C(25)-C(30)	47.0(4)	C(33)-C(34)-C(35)-C(36)	112.6(6)
C(10)-C(11)-C(12)-O(7)	-75.6(4)	O(7)-C(12)-C(25)-C(26)	164.5(3)	C(34)-C(35)-C(36)-C(37)	175.6(5)
C(10)-C(11)-C(12)-C(25)	168.0(3)	C(11)-C(12)-C(25)-C(26)	-78.3(4)	C(40)-O(12)-C(39)-O(11)	2.1(7)
C(12)-O(7)-C(13)-O(8)	5.6(6)	C(31)-C(25)-C(26)-O(4)	-168.8(4)	C(40)-O(12)-C(39)-C(33)	-176.5(4)
C(12)-O(7)-C(13)-C(14)	-175.1(3)	C(12)-C(25)-C(26)-O(4)	72.7(5)	C(38)-C(33)-C(39)-O(11)	6.5(7)
C(16)-N(2)-C(14)-C(15)	-0.5(4)	C(30)-C(25)-C(26)-O(4)	-52.9(5)	C(24)-C(33)-C(39)-O(11)	-115.5(5)
C(16)-N(2)-C(14)-C(13)	-178.6(4)	C(31)-C(25)-C(26)-C(27)	-45.3(5)	C(34)-C(33)-C(39)-O(11)	128.9(5)
O(8)-C(13)-C(14)-C(15)	-175.5(4)	C(12)-C(25)-C(26)-C(27)	-163.8(3)	C(38)-C(33)-C(39)-O(12)	-174.8(4)
O(7)-C(13)-C(14)-C(15)	5.2(6)	C(30)-C(25)-C(26)-C(27)	70.7(4)	C(24)-C(33)-C(39)-O(12)	63.1(5)
				C(34)-C(33)-C(39)-O(12)	-52.5(5)

Table S2. 23. *Hydrogen bonds for sh3137_a_sq (**3**) [Å and °].*

D-H...A	d(D-H)	d(H...A)	d(D...A)	<(DHA)
O(10)-H(29)...N(2)#1	0.84	2.37	3.151(5)	155.2

Symmetry transformations used to generate equivalent atoms:
#1 -x+1/2,-y+1,z+1/2

Table S2. 24. *Crystal data and structure refinement for sh3191 (**6**).*

Identification code	sh3191	
Empirical formula	C44 H58 N2 O15	
Formula weight	854.92	
Temperature	152(2) K	
Wavelength	0.71073 Å	
Crystal system	Orthorhombic	
Space group	$P2_12_12_1$	
Unit cell dimensions	a = 8.6801(6) Å	a= 90°.
	b = 20.1469(16) Å	b= 90°.
	c = 26.259(2) Å	g = 90°.
Volume	4592.0(6) Å3	
Z	4	
Density (calculated)	1.237 Mg/m3	
Absorption coefficient	0.093 mm-1	
F(000)	1824	
Crystal size	0.471 x 0.322 x 0.212 mm3	
Theta range for data collection	1.274 to 27.653°.	
Index ranges	-11<=h<=11, -26<=k<=26, -34<=l<=34	
Reflections collected	78838	
Independent reflections	10656 [R(int) = 0.0474]	
Completeness to theta = 25.242°	99.90%	
Absorption correction	Semi-empirical from equivalents	
Max. and min. transmission	0.9805 and 0.9575	
Refinement method	Full-matrix least-squares on F2	
Data / restraints / parameters	10656 / 134 / 639	
Goodness-of-fit on F2	1.055	
Final R indices [I>2sigma(I)]	R1 = 0.0565, wR2 = 0.1400	
R indices (all data)	R1 = 0.0831, wR2 = 0.1570	
Absolute structure parameter	0.1(2)	
Extinction coefficient	n/a	
Largest diff. peak and hole	0.749 and -0.406 e.Å-3	

Table S2. 25. *Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters (Å^2 x 10^3) for sh3191 (**6**). U(eq) is defined as one third of the trace of the orthogonalised Uij tensor.*

	x	y	z	U(eq)		x	y	z	U(eq)
N(1)	5823(4)	4697(2)	2835(1)	38(1)	C(21)	297(6)	5020(3)	3290(2)	65(1)
N(2)	4302(4)	4772(2)	834(1)	37(1)	C(22)	112(5)	4782(3)	3755(2)	59(1)
O(1)	3202(3)	3991(1)	3748(1)	38(1)	C(23)	551(5)	4132(2)	3980(2)	47(1)
O(2)	4211(3)	5027(1)	3749(1)	42(1)	C(24)	2210(4)	4122(2)	4188(2)	36(1)
O(3)	6074(4)	3680(1)	2517(1)	49(1)	C(25)	6747(4)	2976(2)	-739(1)	31(1)
O(4)	10448(4)	4038(2)	1210(1)	64(1)	C(26)	7357(4)	2998(2)	-1300(2)	37(1)
O(5)	8989(3)	2878(2)	-1288(1)	51(1)	C(27)	6615(5)	2495(2)	-1643(2)	39(1)
O(6)	4144(3)	2546(1)	-678(1)	42(1)	C(28)	5978(6)	2649(2)	-2081(2)	54(1)
O(7)	4489(3)	3584(1)	-938(1)	46(1)	C(29)	5328(8)	2157(3)	-2450(2)	74(2)
O(8)	6350(3)	3644(1)	13(1)	33(1)	C(30)	7214(4)	2336(2)	-464(2)	36(1)
O(9)	5858(4)	4734(1)	-115(1)	46(1)	C(31)	4992(4)	2996(2)	-775(1)	33(1)
O(10)	4135(3)	3932(1)	1384(1)	42(1)	C(32)	2832(5)	3653(3)	-954(3)	73(2)
O(11)	546(4)	3975(2)	5141(1)	56(1)	C(33)	2566(5)	3581(2)	4593(2)	40(1)
O(12)	5096(3)	3999(1)	4694(1)	46(1)	C(34)	2169(5)	3857(2)	5128(2)	46(1)
O(13)	4899(4)	2950(2)	4434(1)	59(1)	C(35A)	2494(8)	3380(3)	5553(3)	48(2)
C(1)	4105(4)	4476(2)	3574(1)	34(1)	C(36A)	3519(7)	3488(3)	5896(2)	54(2)
C(2)	4993(4)	4247(2)	3128(2)	36(1)	C(37A)	3856(11)	3054(5)	6338(3)	78(2)
C(3)	5138(5)	3640(2)	2934(2)	44(1)	C(35B)	3120(30)	3460(15)	5532(10)	48(4)
C(4)	6437(5)	4335(2)	2482(2)	40(1)	C(36B)	2510(30)	3139(11)	5899(8)	51(4)
C(5)	7419(5)	4554(2)	2070(2)	45(1)	C(37B)	3220(40)	2689(18)	6285(11)	68(6)
C(6)	8025(5)	4171(2)	1712(2)	43(1)	C(38)	1726(6)	2929(2)	4491(2)	54(1)
C(7)	9006(5)	4430(2)	1305(2)	47(1)	C(39)	4300(5)	3457(2)	4562(2)	42(1)
C(8)	9231(5)	4091(2)	823(2)	41(1)	C(40)	6754(5)	3945(3)	4646(3)	67(2)

C(9)	9655(4)	4470(2)	372(2)	42(1)	O(14)	4013(9)	2231(3)	3524(2)	124(2)
C(10)	9519(4)	4262(2)	-100(2)	40(1)	C(41A)	4067(13)	1530(5)	3585(6)	140(4)
C(11)	8971(4)	3595(2)	-277(2)	36(1)	C(42A)	2977(13)	1237(5)	3851(6)	139(4)
C(12)	7293(4)	3602(2)	-451(1)	33(1)	C(41B)	3300(70)	1960(30)	3073(16)	140(4)
C(13)	5778(4)	4233(2)	140(1)	34(1)	C(42B)	2330(70)	1470(30)	3210(30)	140(4)
C(14)	5029(4)	4218(2)	639(2)	33(1)	O(15A)	4947(16)	6158(6)	3149(5)	59(3)
C(15)	4933(4)	3707(2)	975(2)	37(1)	C(43A)	4990(20)	6308(7)	2601(7)	73(4)
C(16)	3794(4)	4581(2)	1273(2)	40(1)	C(44A)	4200(40)	6947(14)	2540(20)	128(12)
C(17)	2911(5)	4950(3)	1641(2)	51(1)	O(15B)	5414(13)	6188(5)	2923(5)	63(3)
C(18)	2358(5)	4671(3)	2067(2)	58(1)	C(43B)	3913(15)	6222(5)	2743(4)	65(3)
C(19)	1468(6)	4990(3)	2438(2)	68(2)	C(44B)	3530(30)	6922(9)	2652(14)	97(8)
C(20)	1002(6)	4679(3)	2903(2)	62(1)					

Table S2. 26. *Bond lengths [Å] and angles [°] for sh3191 (**6**).*

N(1)-C(4)	1.294(5)	C(43A)-H(43A)	0.99	C(27)-C(28)-H(28)	117.7
N(1)-C(2)	1.390(5)	C(43A)-H(43B)	0.99	C(29)-C(28)-H(28)	117.7
N(2)-C(16)	1.292(5)	C(44A)-H(44A)	0.98	C(28)-C(29)-H(29A)	109.5
N(2)-C(14)	1.382(5)	C(44A)-H(44B)	0.98	C(28)-C(29)-H(29B)	109.5
O(1)-C(1)	1.334(4)	C(44A)-H(44C)	0.98	H(29A)-C(29)-H(29B)	109.5
O(1)-C(24)	1.464(4)	O(15B)-C(43B)	1.388(17)	C(28)-C(29)-H(29C)	109.5
O(2)-C(1)	1.205(4)	O(15B)-H(15B)	0.84	H(29A)-C(29)-H(29C)	109.5
O(3)-C(4)	1.360(5)	C(43B)-C(44B)	1.470(16)	H(29B)-C(29)-H(29C)	109.5
O(3)-C(3)	1.367(5)	C(43B)-H(43C)	0.99	C(25)-C(30)-H(30A)	109.5
O(4)-C(8)	1.469(5)	C(43B)-H(43D)	0.99	C(25)-C(30)-H(30B)	109.5
O(4)-C(7)	1.500(6)	C(44B)-H(44D)	0.98	H(30A)-C(30)-H(30B)	109.5
O(5)-C(26)	1.438(5)	C(44B)-H(44E)	0.98	C(25)-C(30)-H(30C)	109.5
O(5)-H(14)	0.84	C(44B)-H(44F)	0.98	H(30A)-C(30)-H(30C)	109.5
O(6)-C(31)	1.195(4)			H(30B)-C(30)-H(30C)	109.5
O(7)-C(31)	1.331(4)	C(4)-N(1)-C(2)	104.1(3)	O(6)-C(31)-O(7)	122.7(3)
O(7)-C(32)	1.445(5)	C(16)-N(2)-C(14)	104.2(3)	O(6)-C(31)-C(25)	125.4(3)
O(8)-C(13)	1.329(4)	C(1)-O(1)-C(24)	118.9(3)	O(7)-C(31)-C(25)	111.8(3)
O(8)-C(12)	1.471(4)	C(4)-O(3)-C(3)	104.4(3)	O(7)-C(32)-H(32A)	109.5
O(9)-C(13)	1.213(4)	C(8)-O(4)-C(7)	58.5(3)	O(7)-C(32)-H(32B)	109.5
O(10)-C(15)	1.357(5)	C(26)-O(5)-H(14)	109.5	H(32A)-C(32)-H(32B)	109.5
O(10)-C(16)	1.371(5)	C(31)-O(7)-C(32)	114.9(3)	O(7)-C(32)-H(32C)	109.5
O(11)-C(34)	1.429(6)	C(13)-O(8)-C(12)	117.9(3)	H(32A)-C(32)-H(32C)	109.5
O(11)-H(29)	0.84	C(15)-O(10)-C(16)	105.1(3)	H(32B)-C(32)-H(32C)	109.5
O(12)-C(39)	1.338(5)	C(34)-O(11)-H(29)	109.5	C(38)-C(33)-C(39)	108.7(3)
O(12)-C(40)	1.449(5)	C(39)-O(12)-C(40)	115.4(4)	C(38)-C(33)-C(34)	111.2(4)
O(13)-C(39)	1.195(5)	O(2)-C(1)-O(1)	126.0(3)	C(39)-C(33)-C(34)	109.0(4)
C(1)-C(2)	1.476(5)	O(2)-C(1)-C(2)	123.5(3)	C(38)-C(33)-C(24)	112.8(4)
C(2)-C(3)	1.330(6)	O(1)-C(1)-C(2)	110.5(3)	C(39)-C(33)-C(24)	105.9(3)
C(3)-H(3)	0.95	C(3)-C(2)-N(1)	109.7(3)	C(34)-C(33)-C(24)	108.9(3)
C(4)-C(5)	1.446(6)	C(3)-C(2)-C(1)	129.8(4)	O(11)-C(34)-C(35A)	105.8(4)
C(5)-C(6)	1.324(6)	N(1)-C(2)-C(1)	120.5(3)	O(11)-C(34)-C(33)	107.6(4)
C(5)-H(5)	0.95	C(2)-C(3)-O(3)	108.0(4)	C(35A)-C(34)-C(33)	113.7(4)
C(6)-C(7)	1.462(6)	C(2)-C(3)-H(3)	126	O(11)-C(34)-C(35B)	126.1(10)
C(6)-H(6)	0.95	O(3)-C(3)-H(3)	126	C(33)-C(34)-C(35B)	108.2(12)
C(7)-C(8)	1.451(6)	N(1)-C(4)-O(3)	113.7(3)	O(11)-C(34)-H(34A)	109.9
C(7)-H(7)	1	N(1)-C(4)-C(5)	127.4(4)	C(35A)-C(34)-H(34A)	109.9
C(8)-C(9)	1.458(6)	O(3)-C(4)-C(5)	118.9(3)	C(33)-C(34)-H(34A)	109.9
C(8)-H(8)	1	C(6)-C(5)-C(4)	126.0(4)	O(11)-C(34)-H(34B)	104.3
C(9)-C(10)	1.312(6)	C(6)-C(5)-H(5)	117	C(33)-C(34)-H(34B)	104.3
C(9)-H(9)	0.95	C(4)-C(5)-H(5)	117	C(35B)-C(34)-H(34B)	104.3
C(10)-C(11)	1.500(6)	C(5)-C(6)-C(7)	122.9(4)	C(36A)-C(35A)-C(34)	122.9(6)
C(10)-H(10)	0.95	C(5)-C(6)-H(6)	118.6	C(36A)-C(35A)-H(35A)	118.6
C(11)-C(12)	1.527(5)	C(7)-C(6)-H(6)	118.6	C(34)-C(35A)-H(35A)	118.6
C(11)-H(11A)	0.99	C(8)-C(7)-C(6)	123.2(4)	C(35A)-C(36A)-C(37A)	125.8(7)
C(11)-H(11B)	0.99	C(8)-C(7)-O(4)	59.7(3)	C(35A)-C(36A)-H(36A)	117.1
C(12)-C(25)	1.545(5)	C(6)-C(7)-O(4)	114.9(4)	C(37A)-C(36A)-H(36A)	117.1
C(12)-H(12)	1	C(8)-C(7)-H(7)	115.6	C(36A)-C(37A)-H(37A)	109.5
C(13)-C(14)	1.462(5)	C(6)-C(7)-H(7)	115.6	C(36A)-C(37A)-H(37B)	109.5

C(14)-C(15)	1.359(5)	O(4)-C(7)-H(7)	115.6	H(37A)-C(37A)-H(37B)	109.5
C(15)-H(15)	0.95	C(7)-C(8)-C(9)	119.7(4)	C(36A)-C(37A)-H(37C)	109.5
C(16)-C(17)	1.440(6)	C(7)-C(8)-O(4)	61.8(3)	H(37A)-C(37A)-H(37C)	109.5
C(17)-C(18)	1.341(7)	C(9)-C(8)-O(4)	114.7(3)	H(37B)-C(37A)-H(37C)	109.5
C(17)-H(17)	0.95	C(7)-C(8)-H(8)	116.4	C(36B)-C(35B)-C(34)	123(2)
C(18)-C(19)	1.401(7)	C(9)-C(8)-H(8)	116.4	C(36B)-C(35B)-H(35B)	118.3
C(18)-H(18)	0.95	O(4)-C(8)-H(8)	116.4	C(34)-C(35B)-H(35B)	118.3
C(19)-C(20)	1.429(8)	C(10)-C(9)-C(8)	125.2(4)	C(35B)-C(36B)-C(37B)	130(2)
C(19)-H(19)	0.95	C(10)-C(9)-H(9)	117.4	C(35B)-C(36B)-H(36B)	114.8
C(20)-C(21)	1.371(7)	C(8)-C(9)-H(9)	117.4	C(37B)-C(36B)-H(36B)	114.8
C(20)-H(20)	0.95	C(9)-C(10)-C(11)	127.4(4)	C(36B)-C(37B)-H(37D)	109.5
C(21)-C(22)	1.323(7)	C(9)-C(10)-H(10)	116.3	C(36B)-C(37B)-H(37E)	109.5
C(21)-H(21)	0.95	C(11)-C(10)-H(10)	116.3	H(37D)-C(37B)-H(37E)	109.5
C(22)-C(23)	1.487(7)	C(10)-C(11)-C(12)	112.7(3)	C(36B)-C(37B)-H(37F)	109.5
C(22)-H(22)	0.95	C(10)-C(11)-H(11A)	109	H(37D)-C(37B)-H(37F)	109.5
C(23)-C(24)	1.540(6)	C(12)-C(11)-H(11A)	109	H(37E)-C(37B)-H(37F)	109.5
C(23)-H(23A)	0.99	C(10)-C(11)-H(11B)	109	C(33)-C(38)-H(38A)	109.5
C(23)-H(23B)	0.99	C(12)-C(11)-H(11B)	109	C(33)-C(38)-H(38B)	109.5
C(24)-C(33)	1.555(6)	H(11A)-C(11)-H(11B)	107.8	H(38A)-C(38)-H(38B)	109.5
C(24)-H(24)	1	O(8)-C(12)-C(11)	106.4(3)	C(33)-C(38)-H(38C)	109.5
C(25)-C(31)	1.527(5)	O(8)-C(12)-C(25)	106.3(3)	H(38A)-C(38)-H(38C)	109.5
C(25)-C(30)	1.532(5)	C(11)-C(12)-C(25)	115.5(3)	H(38B)-C(38)-H(38C)	109.5
C(25)-C(26)	1.567(5)	O(8)-C(12)-H(12)	109.5	O(13)-C(39)-O(12)	123.0(4)
C(26)-C(27)	1.500(6)	C(11)-C(12)-H(12)	109.5	O(13)-C(39)-C(33)	125.7(4)
C(26)-H(26)	1	C(25)-C(12)-H(12)	109.5	O(12)-C(39)-C(33)	111.2(3)
C(27)-C(28)	1.314(6)	O(9)-C(13)-O(8)	125.6(3)	O(12)-C(40)-H(40A)	109.5
C(27)-H(27A)	0.95	O(9)-C(13)-C(14)	122.5(3)	O(12)-C(40)-H(40B)	109.5
C(28)-C(29)	1.497(7)	O(8)-C(13)-C(14)	111.9(3)	H(40A)-C(40)-H(40B)	109.5
C(28)-H(28)	0.95	C(15)-C(14)-N(2)	110.1(3)	O(12)-C(40)-H(40C)	109.5
C(29)-H(29A)	0.98	C(15)-C(14)-C(13)	128.6(3)	H(40A)-C(40)-H(40C)	109.5
C(29)-H(29B)	0.98	N(2)-C(14)-C(13)	121.3(3)	H(40B)-C(40)-H(40C)	109.5
C(29)-H(29C)	0.98	O(10)-C(15)-C(14)	106.9(3)	C(41A)-O(14)-H(14A)	109.5
C(30)-H(30A)	0.98	O(10)-C(15)-H(15)	126.5	C(42A)-C(41A)-O(14)	118.8(11)
C(30)-H(30B)	0.98	C(14)-C(15)-H(15)	126.5	C(42A)-C(41A)-H(41A)	107.6
C(30)-H(30C)	0.98	N(2)-C(16)-O(10)	113.6(3)	O(14)-C(41A)-H(41A)	107.6
C(32)-H(32A)	0.98	N(2)-C(16)-C(17)	128.7(4)	C(42A)-C(41A)-H(41B)	107.6
C(32)-H(32B)	0.98	O(10)-C(16)-C(17)	117.7(4)	O(14)-C(41A)-H(41B)	107.6
C(32)-H(32C)	0.98	C(18)-C(17)-C(16)	122.2(5)	H(41A)-C(41A)-H(41B)	107
C(33)-C(38)	1.527(6)	C(18)-C(17)-H(17)	118.9	C(41A)-C(42A)-H(42A)	109.5
C(33)-C(39)	1.527(6)	C(16)-C(17)-H(17)	118.9	C(41A)-C(42A)-H(42B)	109.5
C(33)-C(34)	1.549(6)	C(17)-C(18)-C(19)	125.9(6)	H(42A)-C(42A)-H(42B)	109.5
C(34)-C(35A)	1.502(7)	C(17)-C(18)-H(18)	117	C(41A)-C(42A)-H(42C)	109.5
C(34)-C(35B)	1.57(2)	C(19)-C(18)-H(18)	117	H(42A)-C(42A)-H(42C)	109.5
C(34)-H(34A)	1	C(18)-C(19)-C(20)	123.3(6)	H(42B)-C(42A)-H(42C)	109.5
C(34)-H(34B)	1	C(18)-C(19)-H(19)	118.3	C(42B)-C(41B)-O(14)	109(3)
C(35A)-C(36A)	1.284(9)	C(20)-C(19)-H(19)	118.3	C(42B)-C(41B)-H(41C)	109.9
C(35A)-H(35A)	0.95	C(21)-C(20)-C(19)	122.6(5)	O(14)-C(41B)-H(41C)	109.9
C(36A)-C(37A)	1.483(9)	C(21)-C(20)-H(20)	118.7	C(42B)-C(41B)-H(41D)	109.8
C(36A)-H(36A)	0.95	C(19)-C(20)-H(20)	118.7	O(14)-C(41B)-H(41D)	109.9
C(37A)-H(37A)	0.98	C(22)-C(21)-C(20)	124.0(6)	H(41C)-C(41B)-H(41D)	108.3
C(37A)-H(37B)	0.98	C(22)-C(21)-H(21)	118	C(41B)-C(42B)-H(42D)	109.5
C(37A)-H(37C)	0.98	C(20)-C(21)-H(21)	118	C(41B)-C(42B)-H(42E)	109.5
C(35B)-C(36B)	1.28(2)	C(21)-C(22)-C(23)	130.9(5)	H(42D)-C(42B)-H(42E)	109.5
C(35B)-H(35B)	0.95	C(21)-C(22)-H(22)	114.6	C(41B)-C(42B)-H(42F)	109.4
C(36B)-C(37B)	1.49(2)	C(23)-C(22)-H(22)	114.6	H(42D)-C(42B)-H(42F)	109.5
C(36B)-H(36B)	0.95	C(22)-C(23)-C(24)	113.0(4)	H(42E)-C(42B)-H(42F)	109.5
C(37B)-H(37D)	0.98	C(22)-C(23)-H(23A)	109	C(43A)-O(15A)-H(15A)	109.5
C(37B)-H(37E)	0.98	C(24)-C(23)-H(23A)	109	O(15A)-C(43A)-C(44A)	106(2)
C(37B)-H(37F)	0.98	C(22)-C(23)-H(23B)	109	O(15A)-C(43A)-H(43A)	110.5
C(38)-H(38A)	0.98	C(24)-C(23)-H(23B)	109	C(44A)-C(43A)-H(43A)	110.5
C(38)-H(38B)	0.98	H(23A)-C(23)-H(23B)	107.8	O(15A)-C(43A)-H(43B)	110.5
C(38)-H(38C)	0.98	O(1)-C(24)-C(23)	105.9(3)	C(44A)-C(43A)-H(43B)	110.5
C(40)-H(40A)	0.98	O(1)-C(24)-C(33)	107.2(3)	H(43A)-C(43A)-H(43B)	108.7
C(40)-H(40B)	0.98	C(23)-C(24)-C(33)	115.9(3)	C(43A)-C(44A)-H(44A)	109.5
C(40)-H(40C)	0.98	O(1)-C(24)-H(24)	109.2	C(43A)-C(44A)-H(44B)	109.5
O(14)-C(41A)	1.422(10)	C(23)-C(24)-H(24)	109.2	H(44A)-C(44A)-H(44B)	109.5
O(14)-C(41B)	1.44(3)	C(33)-C(24)-H(24)	109.2	C(43A)-C(44A)-H(44C)	109.5

O(14)-H(14A)	0.84	C(31)-C(25)-C(30)	108.5(3)	H(44A)-C(44A)-H(44C)	109.5
C(41A)-C(42A)	1.315(16)	C(31)-C(25)-C(12)	108.3(3)	H(44B)-C(44A)-H(44C)	109.5
C(41A)-H(41A)	0.99	C(30)-C(25)-C(12)	112.1(3)	C(43B)-O(15B)-H(15B)	109.5
C(41A)-H(41B)	0.99	C(31)-C(25)-C(26)	106.1(3)	O(15B)-C(43B)-C(44B)	108.5(13)
C(42A)-H(42A)	0.98	C(30)-C(25)-C(26)	112.2(3)	O(15B)-C(43B)-H(43C)	110
C(42A)-H(42B)	0.98	C(12)-C(25)-C(26)	109.4(3)	C(44B)-C(43B)-H(43C)	110
C(42A)-H(42C)	0.98	O(5)-C(26)-C(27)	108.8(3)	O(15B)-C(43B)-H(43D)	110
C(41B)-C(42B)	1.35(3)	O(5)-C(26)-C(25)	107.9(3)	C(44B)-C(43B)-H(43D)	110
C(41B)-H(41C)	0.99	C(27)-C(26)-C(25)	113.6(3)	H(43C)-C(43B)-H(43D)	108.4
C(41B)-H(41D)	0.99	O(5)-C(26)-H(26)	108.8	C(43B)-C(44B)-H(44D)	109.5
C(42B)-H(42D)	0.98	C(27)-C(26)-H(26)	108.8	C(43B)-C(44B)-H(44E)	109.5
C(42B)-H(42E)	0.98	C(25)-C(26)-H(26)	108.8	H(44D)-C(44B)-H(44E)	109.5
C(42B)-H(42F)	0.98	C(28)-C(27)-C(26)	123.2(4)	C(43B)-C(44B)-H(44F)	109.5
O(15A)-C(43A)	1.47(2)	C(28)-C(27)-H(27A)	118.4	H(44D)-C(44B)-H(44F)	109.5
O(15A)-H(15A)	0.84	C(26)-C(27)-H(27A)	118.4	H(44E)-C(44B)-H(44F)	109.5
C(43A)-C(44A)	1.47(2)	C(27)-C(28)-C(29)	124.7(4)		

Table S2. 27. *Anisotropic displacement parameters ($Å^2$ x 10^3) for sh3191 (**6**). The anisotropic displacement factor exponent takes the form: $-2\pi^2[h^2 a^{*2}U^{11} + ... + 2 h k a^* b^* U^{12}]$.*

	U11	U22	U33	U23	U13	U12
N(1)	32(2)	37(2)	46(2)	-1(1)	4(2)	1(1)
N(2)	34(2)	39(2)	37(2)	2(1)	-1(1)	2(1)
O(1)	35(1)	35(1)	43(2)	-6(1)	9(1)	-1(1)
O(2)	43(2)	36(1)	47(2)	-5(1)	8(1)	0(1)
O(3)	53(2)	40(2)	55(2)	-7(1)	16(2)	3(1)
O(4)	44(2)	78(2)	70(2)	-8(2)	-2(2)	4(2)
O(5)	32(1)	72(2)	48(2)	-2(2)	10(1)	-1(1)
O(6)	32(1)	38(1)	56(2)	11(1)	3(1)	-3(1)
O(7)	25(1)	39(1)	73(2)	16(1)	-3(1)	7(1)
O(8)	30(1)	31(1)	38(1)	6(1)	4(1)	3(1)
O(9)	54(2)	33(1)	51(2)	7(1)	11(1)	6(1)
O(10)	34(1)	53(2)	39(2)	6(1)	-2(1)	-4(1)
O(11)	62(2)	39(2)	65(2)	-1(2)	21(2)	2(1)
O(12)	40(1)	31(1)	66(2)	-2(1)	-4(1)	-2(1)
O(13)	56(2)	33(2)	89(2)	-10(2)	-3(2)	9(1)
C(1)	28(2)	35(2)	39(2)	-2(2)	2(2)	3(2)
C(2)	30(2)	37(2)	41(2)	1(2)	2(2)	3(2)
C(3)	46(2)	40(2)	46(2)	-2(2)	11(2)	2(2)
C(4)	35(2)	43(2)	41(2)	-4(2)	3(2)	2(2)
C(5)	37(2)	45(2)	51(2)	-1(2)	5(2)	-5(2)
C(6)	36(2)	51(2)	42(2)	-1(2)	1(2)	-6(2)
C(7)	43(2)	45(2)	53(2)	-5(2)	9(2)	-12(2)
C(8)	31(2)	50(2)	43(2)	2(2)	-4(2)	-11(2)
C(9)	31(2)	41(2)	54(3)	5(2)	0(2)	-10(2)
C(10)	34(2)	43(2)	44(2)	6(2)	5(2)	-6(2)
C(11)	26(2)	41(2)	42(2)	5(2)	2(2)	2(2)
C(12)	26(2)	34(2)	38(2)	6(2)	5(2)	2(1)
C(13)	27(2)	34(2)	43(2)	4(2)	-1(2)	-1(2)
C(14)	24(2)	33(2)	41(2)	0(2)	-3(2)	-4(1)

C(15)	30(2)	41(2)	40(2)	2(2)	-4(2)	-3(2)
C(16)	30(2)	50(2)	40(2)	-1(2)	-2(2)	1(2)
C(17)	40(2)	70(3)	44(2)	-6(2)	3(2)	5(2)
C(18)	35(2)	100(4)	39(2)	-5(2)	-4(2)	1(2)
C(19)	51(3)	105(4)	47(3)	-11(3)	-6(2)	10(3)
C(20)	43(2)	68(3)	75(3)	-16(3)	-19(3)	4(2)
C(21)	49(3)	82(4)	63(3)	-7(3)	-6(2)	14(3)
C(22)	41(2)	83(3)	53(3)	1(3)	-3(2)	6(2)
C(23)	35(2)	57(3)	49(2)	-6(2)	5(2)	-4(2)
C(24)	32(2)	36(2)	42(2)	-3(2)	7(2)	0(2)
C(25)	26(2)	33(2)	35(2)	6(2)	2(1)	3(1)
C(26)	33(2)	42(2)	36(2)	8(2)	2(2)	2(2)
C(27)	39(2)	40(2)	38(2)	4(2)	5(2)	2(2)
C(28)	72(3)	46(2)	43(2)	2(2)	-4(2)	7(2)
C(29)	94(4)	76(4)	52(3)	-10(3)	-19(3)	6(3)
C(30)	33(2)	33(2)	43(2)	5(2)	-2(2)	4(2)
C(31)	29(2)	34(2)	37(2)	4(2)	4(2)	4(2)
C(32)	27(2)	61(3)	132(5)	26(3)	-8(3)	9(2)
C(33)	44(2)	28(2)	49(2)	-2(2)	4(2)	-1(2)
C(34)	58(3)	31(2)	51(2)	2(2)	10(2)	0(2)
C(35A)	53(4)	36(3)	54(3)	8(2)	19(3)	-2(3)
C(36A)	54(3)	62(3)	47(3)	6(3)	15(3)	8(3)
C(37A)	82(6)	90(6)	61(4)	19(4)	10(4)	29(5)
C(35B)	52(8)	49(8)	43(7)	8(7)	22(8)	4(8)
C(36B)	55(7)	53(7)	44(7)	4(6)	16(7)	5(7)
C(37B)	72(13)	83(14)	49(11)	8(12)	8(11)	31(11)
C(38)	59(3)	34(2)	70(3)	-4(2)	8(2)	-10(2)
C(39)	49(2)	28(2)	48(2)	1(2)	-3(2)	2(2)
C(40)	38(2)	51(3)	111(5)	-3(3)	-11(3)	-1(2)
O(14)	171(6)	89(4)	111(4)	-10(3)	8(4)	-6(4)
C(41A)	110(6)	80(5)	230(11)	-16(6)	-34(7)	2(4)
C(42A)	109(6)	76(5)	231(11)	-11(6)	-37(7)	0(4)
C(41B)	110(7)	80(6)	230(11)	-16(6)	-35(7)	2(5)
C(42B)	110(7)	79(6)	231(11)	-14(6)	-36(7)	1(5)
O(15A)	66(7)	45(4)	68(8)	12(5)	-4(5)	-8(5)
C(43A)	76(8)	65(7)	79(8)	-5(6)	-10(7)	-3(7)
C(44A)	160(30)	115(13)	110(20)	12(12)	-60(20)	15(15)
O(15B)	61(6)	36(4)	92(9)	-9(5)	-18(5)	-8(4)
C(43B)	67(7)	63(6)	66(6)	-10(4)	-13(5)	-2(5)
C(44B)	113(17)	73(8)	106(17)	22(7)	-55(16)	24(9)

Table S2. 28. *Hydrogen coordinates (x 10^4) and isotropic displacement parameters ($Å^2$ x 10^3) for* ***6****.*

	x	y	z	U(eq)		x	y	z	U(eq)
H(14)	9425	3116	-1507	76	H(37A)	4964	2963	6351	117
H(29)	378	4368	5232	83	H(37B)	3291	2636	6302	117
H(3)	4672	3248	3063	53	H(37C)	3537	3277	6652	117
H(5)	7651	5014	2054	54	H(35B)	4214	3457	5502	58
H(6)	7814	3709	1720	52	H(36B)	1429	3196	5932	61
H(7)	9125	4923	1295	56	H(37D)	2778	2784	6621	102
H(8)	8594	3684	765	50	H(37E)	4334	2762	6294	102
H(9)	10063	4902	423	51	H(37F)	3006	2227	6193	102
H(10)	9799	4569	-358	48	H(38A)	2157	2579	4708	81
H(11A)	9086	3271	5	44	H(38B)	1854	2805	4132	81
H(11B)	9630	3444	-562	44	H(38C)	627	2983	4567	81
H(12)	7106	4003	-667	39	H(40A)	7229	4375	4721	100
H(15)	5347	3274	932	45	H(40B)	7018	3811	4298	100
H(17)	2713	5407	1579	61	H(40C)	7138	3612	4887	100
H(18)	2592	4216	2122	69	H(14A)	4366	2415	3786	186
H(19)	1158	5436	2379	81	H(41A)	4071	1329	3241	168
H(20)	1188	4218	2945	75	H(41B)	5066	1419	3746	168
H(21)	-82	5453	3219	78	H(42A)	3168	758	3860	208
H(22)	-385	5077	3985	71	H(42B)	2975	1413	4199	208
H(23A)	445	3783	3717	56	H(42C)	1975	1322	3692	208
H(23B)	-169	4025	4260	56	H(41C)	2730	2311	2891	168
H(24)	2466	4569	4331	44	H(41D)	4104	1782	2842	168
H(26)	7169	3451	-1442	45	H(42D)	1841	1280	2905	210
H(27A)	6605	2044	-1538	47	H(42E)	2902	1118	3388	210
H(28)	5926	3106	-2170	64	H(42F)	1531	1645	3437	210
H(29A)	4289	2294	-2551	111	H(15A)	5201	5761	3196	89
H(29B)	5280	1718	-2288	111	H(43A)	4467	5956	2405	88
H(29C)	5989	2134	-2752	111	H(43B)	6073	6341	2482	88
H(30A)	6778	1953	-643	54	H(44A)	4389	7229	2834	192
H(30B)	6824	2345	-114	54	H(44B)	3089	6870	2500	192
H(30C)	8340	2299	-459	54	H(44C)	4588	7169	2230	192
H(32A)	2564	4085	-1100	110	H(15B)	5890	5885	2770	95
H(32B)	2417	3623	-607	110	H(43C)	3197	6029	2997	78
H(32C)	2392	3298	-1164	110	H(43D)	3819	5966	2423	78
H(34A)	2736	4283	5187	56	H(44D)	4005	7198	2915	146
H(34B)	2635	4311	5131	56	H(44E)	2404	6978	2664	146
H(35A)	1917	2980	5569	57	H(44F)	3907	7055	2316	146
H(36A)	4114	3882	5863	65					

Table S2. 29. *Torsion angles [°] for sh3191 (**6**).*

C(24)-O(1)-C(1)-O(2)	1.2(6)	O(8)-C(13)-C(14)-C(15)	3.6(5)	C(32)-O(7)-C(31)-C(25)	-176.5(4)
C(24)-O(1)-C(1)-C(2)	-179.3(3)	O(9)-C(13)-C(14)-N(2)	3.2(5)	C(30)-C(25)-C(31)-O(6)	-10.5(5)
C(4)-N(1)-C(2)-C(3)	-0.6(5)	O(8)-C(13)-C(14)-N(2)	-177.3(3)	C(12)-C(25)-C(31)-O(6)	-132.4(4)
C(4)-N(1)-C(2)-C(1)	-179.5(3)	C(16)-O(10)-C(15)-C(14)	-0.4(4)	C(26)-C(25)-C(31)-O(6)	110.2(4)
O(2)-C(1)-C(2)-C(3)	170.7(4)	N(2)-C(14)-C(15)-O(10)	0.5(4)	C(30)-C(25)-C(31)-O(7)	170.3(3)
O(1)-C(1)-C(2)-C(3)	-8.9(6)	C(13)-C(14)-C(15)-O(10)	179.6(3)	C(12)-C(25)-C(31)-O(7)	48.3(4)
O(2)-C(1)-C(2)-N(1)	-10.7(6)	C(14)-N(2)-C(16)-O(10)	0.0(4)	C(26)-C(25)-C(31)-O(7)	-69.0(4)
O(1)-C(1)-C(2)-N(1)	169.7(3)	C(14)-N(2)-C(16)-C(17)	-178.4(4)	O(1)-C(24)-C(33)-C(38)	-81.7(4)
N(1)-C(2)-C(3)-O(3)	0.7(5)	C(15)-O(10)-C(16)-N(2)	0.2(4)	C(23)-C(24)-C(33)-C(38)	36.3(5)
C(1)-C(2)-C(3)-O(3)	179.4(4)	C(15)-O(10)-C(16)-C(17)	178.9(3)	O(1)-C(24)-C(33)-C(39)	37.2(4)
C(4)-O(3)-C(3)-C(2)	-0.5(5)	N(2)-C(16)-C(17)-C(18)	174.9(4)	C(23)-C(24)-C(33)-C(39)	155.2(3)
C(2)-N(1)-C(4)-O(3)	0.3(5)	O(10)-C(16)-C(17)-C(18)	-3.5(6)	O(1)-C(24)-C(33)-C(34)	154.3(3)
C(2)-N(1)-C(4)-C(5)	179.7(4)	C(16)-C(17)-C(18)-C(19)	-178.7(5)	C(23)-C(24)-C(33)-C(34)	-87.7(4)
C(3)-O(3)-C(4)-N(1)	0.1(5)	C(17)-C(18)-C(19)-C(20)	-175.5(5)	C(38)-C(33)-C(34)-O(11)	-60.8(4)
C(3)-O(3)-C(4)-C(5)	-179.4(4)	C(18)-C(19)-C(20)-C(21)	171.7(5)	C(39)-C(33)-C(34)-O(11)	179.3(3)
N(1)-C(4)-C(5)-C(6)	-179.6(4)	C(19)-C(20)-C(21)-C(22)	-166.8(5)	C(24)-C(33)-C(34)-O(11)	64.2(4)
O(3)-C(4)-C(5)-C(6)	-0.2(7)	C(20)-C(21)-C(22)-C(23)	-0.8(9)	C(38)-C(33)-C(34)-C(35A)	56.1(6)
C(4)-C(5)-C(6)-C(7)	179.8(4)	C(21)-C(22)-C(23)-C(24)	86.9(6)	C(39)-C(33)-C(34)-C(35A)	-63.8(5)
C(5)-C(6)-C(7)-C(8)	-157.4(4)	C(1)-O(1)-C(24)-C(23)	109.9(4)	C(24)-C(33)-C(34)-C(35A)	-179.0(4)
C(5)-C(6)-C(7)-O(4)	133.8(4)	C(1)-O(1)-C(24)-C(33)	-125.8(3)	C(38)-C(33)-C(34)-C(35B)	78.2(12)
C(8)-O(4)-C(7)-C(6)	115.4(4)	C(22)-C(23)-C(24)-O(1)	-82.9(4)	C(39)-C(33)-C(34)-C(35B)	-41.7(12)
C(6)-C(7)-C(8)-C(9)	154.6(4)	C(22)-C(23)-C(24)-C(33)	158.4(4)	C(24)-C(33)-C(34)-C(35B)	-156.9(12)
O(4)-C(7)-C(8)-C(9)	-103.9(4)	O(8)-C(12)-C(25)-C(31)	49.4(4)	O(11)-C(34)-C(35A)-C(36A)	-127.3(7)
C(6)-C(7)-C(8)-O(4)	-101.5(5)	C(11)-C(12)-C(25)-C(31)	167.2(3)	C(33)-C(34)-C(35A)-C(36A)	114.9(7)
C(7)-O(4)-C(8)-C(9)	111.9(4)	O(8)-C(12)-C(25)-C(30)	-70.3(3)	C(34)-C(35A)-C(36A)-C(37A)	177.2(6)
C(7)-C(8)-C(9)-C(10)	-162.0(4)	C(11)-C(12)-C(25)-C(30)	47.5(4)	O(11)-C(34)-C(35B)-C(36B)	7(4)
O(4)-C(8)-C(9)-C(10)	127.6(4)	O(8)-C(12)-C(25)-C(26)	164.6(3)	C(33)-C(34)-C(35B)-C(36B)	-122(3)
C(8)-C(9)-C(10)-C(11)	-2.6(7)	C(11)-C(12)-C(25)-C(26)	-77.6(4)	C(34)-C(35B)-C(36B)-C(37B)	174(3)
C(9)-C(10)-C(11)-C(12)	100.3(5)	C(31)-C(25)-C(26)-O(5)	-171.4(3)	C(40)-O(12)-C(39)-O(13)	3.1(7)
C(13)-O(8)-C(12)-C(11)	100.7(3)	C(30)-C(25)-C(26)-O(5)	-53.1(4)	C(40)-O(12)-C(39)-C(33)	-176.3(4)
C(13)-O(8)-C(12)-C(25)	-135.7(3)	C(12)-C(25)-C(26)-O(5)	72.0(4)	C(38)-C(33)-C(39)-O(13)	6.3(6)
C(10)-C(11)-C(12)-O(8)	-74.2(4)	C(31)-C(25)-C(26)-C(27)	-50.7(4)	C(34)-C(33)-C(39)-O(13)	127.7(5)
C(10)-C(11)-C(12)-C(25)	168.1(3)	C(30)-C(25)-C(26)-C(27)	67.6(4)	C(24)-C(33)-C(39)-O(13)	-115.2(5)
C(12)-O(8)-C(13)-O(9)	6.5(5)	C(12)-C(25)-C(26)-C(27)	-167.3(3)	C(38)-C(33)-C(39)-O(12)	-174.4(4)
C(12)-O(8)-C(13)-C(14)	-173.0(3)	O(5)-C(26)-C(27)-C(28)	-112.0(5)	C(34)-C(33)-C(39)-O(12)	-53.0(4)
C(16)-N(2)-C(14)-C(15)	-0.3(4)	C(25)-C(26)-C(27)-C(28)	127.8(4)	C(24)-C(33)-C(39)-O(12)	64.1(4)
C(16)-N(2)-C(14)-C(13)	-179.6(3)	C(26)-C(27)-C(28)-C(29)	175.4(5)		
O(9)-C(13)-C(14)-C(15)	-175.9(4)	C(32)-O(7)-C(31)-O(6)	4.3(6)		

Table S2. 30. *Hydrogen bonds for sh3191 (**6**) [Å and °].*

D-H...A	d(D-H)	d(H...A)	d(D...A)	<(DHA)
O(11)-H(29)...N(2)#1	0.84	2.36	3.115(4)	149.4
O(5)-H(14)...O(15A^a)#2	0.84	1.8	2.609(14)	160.6
O(5)-H(14)...O(15B^b)#2	0.84	2.05	2.846(11)	156.9
O(15A^a)-H(15A^a)...N(1)	0.84	2.41	3.151(14)	148.2
O(15B^b)-H(15B^b)...N(1)	0.84	2.4	3.034(10)	132.7
O(14)-H(14A^a)...O(13)	0.84	2.07	2.900(7)	170.7

Symmetry transformations used to generate equivalent atoms:
#1 -x+1/2,-y+1,z+1/2 #2 -x+3/2,-y+1,z-1/2

Table S2. 31. *Crystal data and structure refinement for sh3279 (**7**).*

Identification code	sh3279	
Empirical formula	C44 H58 N2 O15	
Formula weight	854.92	
Temperature	133(2) K	
Wavelength	0.71073 Å	
Crystal system	Orthorhombic	
Space group	$P2_12_12_1$	
Unit cell dimensions	a = 8.7058(6) Å	a= 90°.
	b = 20.2194(15) Å	b= 90°.
	c = 26.000(2) Å	g = 90°.
Volume	4576.7(6) Å3	
Z	4	
Density (calculated)	1.241 Mg/m3	
Absorption coefficient	0.093 mm-1	
F(000)	1824	
Crystal size	0.530 x 0.280 x 0.240 mm3	
Theta range for data collection	1.276 to 27.915°.	
Index ranges	-6<=h<=11, -23<=k<=26, -33<=l<=34	
Reflections collected	24427	
Independent reflections	10915 [R(int) = 0.0400]	
Completeness to theta = 25.242°	99.90%	
Absorption correction	Semi-empirical from equivalents	
Max. and min. transmission	0.9779 and 0.9522	
Refinement method	Full-matrix least-squares on F2	
Data / restraints / parameters	10915 / 221 / 626	
Goodness-of-fit on F2	1.025	
Final R indices [I>2sigma(I)]	R1 = 0.0529, wR2 = 0.1229	
R indices (all data)	R1 = 0.0880, wR2 = 0.1405	
Absolute structure parameter	0.5(5)	
Extinction coefficient	n/a	
Largest diff. peak and hole	0.431 and -0.412 e.Å-3	

Table S2. 32. *Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters (Å^2 x 10^3) for sh3279 (**7**). U(eq) is defined as one third of the trace of the orthogonalised U^{ij} tensor.*

	x	y	z	U(eq)		x	y	z	U(eq)
N(1)	5713(3)	4758(2)	2831(1)	30(1)	C(19)	1355(4)	4953(2)	2429(2)	40(1)
N(2)	4381(3)	4753(2)	831(1)	26(1)	C(20)	926(4)	4669(2)	2878(2)	36(1)
O(1)	3132(3)	4016(1)	3749(1)	28(1)	C(21)	152(4)	5014(2)	3282(2)	40(1)
O(2)	4012(3)	5075(1)	3740(1)	32(1)	C(22)	-14(4)	4787(2)	3779(2)	41(1)
O(3)	6135(3)	3742(1)	2531(1)	37(1)	C(23)	487(4)	4139(2)	3991(2)	36(1)
O(4A)	11256(3)	4306(2)	144(1)	35(1)	C(24)	2136(4)	4138(2)	4195(1)	27(1)
O(4B)	-1380(13)	4723(8)	3491(5)	63(5)	C(25)	6996(4)	2984(2)	-725(1)	22(1)
O(5)	9204(3)	2882(1)	-1294(1)	38(1)	C(26)	7578(4)	3000(2)	-1295(1)	27(1)
O(6)	4725(3)	3589(1)	-907(1)	32(1)	C(27)	6814(4)	2499(2)	-1634(1)	31(1)
O(7)	4422(3)	2529(1)	-683(1)	32(1)	C(28)	6161(6)	2643(2)	-2076(2)	52(1)
O(8)	6599(3)	3650(1)	29(1)	24(1)	C(29)	5496(7)	2151(3)	-2439(2)	75(2)
O(9)	6237(3)	4751(1)	-78(1)	36(1)	C(30)	7477(4)	2350(2)	-440(1)	26(1)
O(10)	4115(3)	3901(1)	1371(1)	29(1)	C(31)	5249(4)	2996(2)	-763(1)	24(1)
O(11)	562(3)	4012(1)	5163(1)	43(1)	C(32)	3069(4)	3645(2)	-930(2)	47(1)
O(12)	4802(3)	2940(1)	4428(1)	45(1)	C(33)	2510(4)	3593(2)	4599(1)	29(1)
O(13)	5065(3)	3992(1)	4682(1)	34(1)	C(34)	2159(4)	3873(2)	5146(1)	33(1)
C(1)	4001(4)	4521(2)	3570(1)	27(1)	C(35A)	2500(6)	3391(3)	5570(2)	36(1)
C(2)	4939(4)	4301(2)	3131(1)	28(1)	C(36A)	3575(6)	3482(3)	5913(2)	42(1)
C(3)	5196(4)	3689(2)	2949(2)	37(1)	C(37A)	3913(8)	3033(4)	6358(2)	63(2)
C(4)	6396(4)	4407(2)	2480(1)	32(1)	C(35B)	3110(40)	3473(19)	5541(13)	36(4)
C(5)	7302(4)	4640(2)	2055(2)	36(1)	C(36B)	2380(40)	3129(16)	5906(11)	44(4)
C(6)	7935(4)	4267(2)	1688(1)	31(1)	C(37B)	3180(60)	2680(30)	6289(17)	70(7)
C(7)	8729(4)	4540(2)	1249(1)	32(1)	C(38)	1638(4)	2946(2)	4500(2)	39(1)

C(8)	9249(4)	4206(2)	846(1)	29(1)	C(39)	4232(4)	3455(2)	4562(1)	31(1)
C(9)	9848(4)	4545(2)	392(1)	31(1)	C(40)	6711(5)	3918(2)	4621(2)	52(1)
C(10)	9806(4)	4283(2)	-123(1)	31(1)	O(14)	3978(9)	2297(3)	3477(2)	122(2)
C(11)	9221(4)	3609(2)	-266(1)	27(1)	C(41)	4158(10)	1618(4)	3475(4)	118(3)
C(12)	7539(4)	3613(2)	-439(1)	24(1)	C(42)	3156(8)	1310(4)	3861(4)	122(3)
C(13)	6044(4)	4246(2)	158(1)	25(1)	O(15)	5121(5)	6215(2)	2943(2)	84(1)
C(14)	5172(4)	4206(2)	644(1)	24(1)	C(43A)	3739(8)	6270(3)	2720(3)	60(2)
C(15)	5012(4)	3697(2)	973(1)	25(1)	C(44A)	3399(12)	6983(4)	2657(3)	97(3)
C(16)	3766(4)	4548(2)	1259(1)	26(1)	C(43B)	4840(30)	6328(12)	2494(10)	70(3)
C(17)	2820(4)	4903(2)	1617(1)	33(1)	C(44B)	4620(50)	6950(16)	2261(13)	116(6)
C(18)	2275(4)	4627(2)	2050(1)	34(1)					

Table S2. 33. *Bond lengths [Å] and angles [°] for sh3279 (**7**).*

N(1)-C(4)	1.301(5)	C(43A)-H(43B)	0.99	C(12)-C(25)-C(26)	109.8(3)
N(1)-C(2)	1.383(5)	C(44A)-H(44A)	0.98	O(5)-C(26)-C(27)	109.1(3)
N(2)-C(16)	1.304(4)	C(44A)-H(44B)	0.98	O(5)-C(26)-C(25)	108.3(3)
N(2)-C(14)	1.390(4)	C(44A)-H(44C)	0.98	C(27)-C(26)-C(25)	113.5(3)
O(1)-C(1)	1.354(4)	C(43B)-C(44B)	1.41(4)	O(5)-C(26)-H(26)	108.6
O(1)-C(24)	1.468(4)	C(43B)-H(43C)	0.99	C(27)-C(26)-H(26)	108.6
O(2)-C(1)	1.204(4)	C(43B)-H(43D)	0.99	C(25)-C(26)-H(26)	108.6
O(3)-C(3)	1.364(4)	C(44B)-H(44D)	0.98	C(28)-C(27)-C(26)	123.8(4)
O(3)-C(4)	1.369(5)	C(44B)-H(44E)	0.98	C(28)-C(27)-H(27)	118.1
O(4A)-C(10)	1.442(5)	C(44B)-H(44F)	0.98	C(26)-C(27)-H(27)	118.1
O(4A)-C(9)	1.467(5)			C(27)-C(28)-C(29)	125.1(4)
O(4B)-C(22)	1.411(12)	C(4)-N(1)-C(2)	104.6(3)	C(27)-C(28)-H(28)	117.4
O(4B)-C(21)	1.557(13)	C(16)-N(2)-C(14)	104.4(3)	C(29)-C(28)-H(28)	117.4
O(5)-C(26)	1.435(4)	C(1)-O(1)-C(24)	118.4(3)	C(28)-C(29)-H(29A)	109.5
O(5)-H(14)	0.843(14)	C(3)-O(3)-C(4)	104.7(3)	C(28)-C(29)-H(29B)	109.5
O(6)-C(31)	1.337(4)	C(10)-O(4A)-C(9)	59.3(2)	H(29A)-C(29)-H(29B)	109.5
O(6)-C(32)	1.448(4)	C(22)-O(4B)-C(21)	55.1(5)	C(28)-C(29)-H(29C)	109.5
O(7)-C(31)	1.205(4)	C(26)-O(5)-H(14)	103(3)	H(29A)-C(29)-H(29C)	109.5
O(8)-C(13)	1.340(4)	C(31)-O(6)-C(32)	114.9(3)	H(29B)-C(29)-H(29C)	109.5
O(8)-C(12)	1.467(4)	C(13)-O(8)-C(12)	117.1(2)	C(25)-C(30)-H(30A)	109.5
O(9)-C(13)	1.204(4)	C(15)-O(10)-C(16)	104.8(3)	C(25)-C(30)-H(30B)	109.5
O(10)-C(15)	1.360(4)	C(34)-O(11)-H(29)	109(3)	H(30A)-C(30)-H(30B)	109.5
O(10)-C(16)	1.375(4)	C(39)-O(13)-C(40)	115.1(3)	C(25)-C(30)-H(30C)	109.5
O(11)-C(34)	1.418(5)	O(2)-C(1)-O(1)	125.4(3)	H(30A)-C(30)-H(30C)	109.5
O(11)-H(29)	0.843(14)	O(2)-C(1)-C(2)	124.2(3)	H(30B)-C(30)-H(30C)	109.5
O(12)-C(39)	1.205(4)	O(1)-C(1)-C(2)	110.4(3)	O(7)-C(31)-O(6)	123.2(3)
O(13)-C(39)	1.343(4)	C(3)-C(2)-N(1)	109.7(3)	O(7)-C(31)-C(25)	124.9(3)
O(13)-C(40)	1.450(5)	C(3)-C(2)-C(1)	130.0(3)	O(6)-C(31)-C(25)	111.9(3)
C(1)-C(2)	1.473(5)	N(1)-C(2)-C(1)	120.3(3)	O(6)-C(32)-H(32A)	109.5
C(2)-C(3)	1.344(5)	C(2)-C(3)-O(3)	107.9(3)	O(6)-C(32)-H(32B)	109.5
C(3)-H(3)	0.95	C(2)-C(3)-H(3)	126.1	H(32A)-C(32)-H(32B)	109.5
C(4)-C(5)	1.438(5)	O(3)-C(3)-H(3)	126.1	O(6)-C(32)-H(32C)	109.5
C(5)-C(6)	1.335(5)	N(1)-C(4)-O(3)	113.2(3)	H(32A)-C(32)-H(32C)	109.5
C(5)-H(5)	0.95	N(1)-C(4)-C(5)	127.7(3)	H(32B)-C(32)-H(32C)	109.5

C(6)-C(7)	1.445(5)	O(3)-C(4)-C(5)	119.1(3)	C(39)-C(33)-C(38)	108.5(3)
C(6)-H(6)	0.95	C(6)-C(5)-C(4)	126.2(4)	C(39)-C(33)-C(24)	107.0(3)
C(7)-C(8)	1.326(5)	C(6)-C(5)-H(5)	116.9	C(38)-C(33)-C(24)	112.8(3)
C(7)-H(7)	0.95	C(4)-C(5)-H(5)	116.9	C(39)-C(33)-C(34)	108.5(3)
C(8)-C(9)	1.462(5)	C(5)-C(6)-C(7)	123.1(4)	C(38)-C(33)-C(34)	111.4(3)
C(8)-H(8)	0.95	C(5)-C(6)-H(6)	118.5	C(24)-C(33)-C(34)	108.5(3)
C(9)-C(10)	1.440(5)	C(7)-C(6)-H(6)	118.5	O(11)-C(34)-C(35A)	107.4(3)
C(9)-H(9A)	1	C(8)-C(7)-C(6)	126.3(4)	O(11)-C(34)-C(35B)	127.4(12)
C(9)-H(9B)	0.95	C(8)-C(7)-H(7)	116.9	O(11)-C(34)-C(33)	107.0(3)
C(10)-C(11)	1.503(5)	C(6)-C(7)-H(7)	116.9	C(35A)-C(34)-C(33)	113.3(3)
C(10)-H(10A)	1	C(7)-C(8)-C(9)	121.3(3)	C(35B)-C(34)-C(33)	108.1(16)
C(10)-H(10B)	0.95	C(7)-C(8)-H(8)	119.3	O(11)-C(34)-H(34A)	109.7
C(11)-C(12)	1.531(4)	C(9)-C(8)-H(8)	119.3	C(35A)-C(34)-H(34A)	109.7
C(11)-H(11A)	0.99	C(10)-C(9)-C(8)	124.7(3)	C(33)-C(34)-H(34A)	109.7
C(11)-H(11B)	0.99	C(10)-C(9)-O(4A)	59.5(2)	O(11)-C(34)-H(34B)	104
C(12)-C(25)	1.548(5)	C(8)-C(9)-O(4A)	119.8(3)	C(35B)-C(34)-H(34B)	104
C(12)-H(12)	1	C(10)-C(9)-H(9A)	114	C(33)-C(34)-H(34B)	104
C(13)-C(14)	1.476(5)	C(8)-C(9)-H(9A)	114	C(36A)-C(35A)-C(34)	123.6(5)
C(14)-C(15)	1.345(5)	O(4A)-C(9)-H(9A)	114	C(36A)-C(35A)-H(35A)	118.2
C(15)-H(15)	0.95	C(10)-C(9)-H(9B)	117.7	C(34)-C(35A)-H(35A)	118.2
C(16)-C(17)	1.435(5)	C(8)-C(9)-H(9B)	117.7	C(35A)-C(36A)-C(37A)	125.7(5)
C(17)-C(18)	1.341(5)	C(9)-C(10)-O(4A)	61.2(2)	C(35A)-C(36A)-H(36A)	117.2
C(17)-H(17)	0.95	C(9)-C(10)-C(11)	124.9(3)	C(37A)-C(36A)-H(36A)	117.2
C(18)-C(19)	1.432(5)	O(4A)-C(10)-C(11)	116.4(3)	C(36A)-C(37A)-H(37A)	109.5
C(18)-H(18)	0.95	C(9)-C(10)-H(10A)	114.5	C(36A)-C(37A)-H(37B)	109.5
C(19)-C(20)	1.354(6)	O(4A)-C(10)-H(10A)	114.5	H(37A)-C(37A)-H(37B)	109.5
C(19)-H(19)	0.95	C(11)-C(10)-H(10A)	114.5	C(36A)-C(37A)-H(37C)	109.5
C(20)-C(21)	1.428(5)	C(9)-C(10)-H(10B)	117.5	H(37A)-C(37A)-H(37C)	109.5
C(20)-H(20)	0.95	C(11)-C(10)-H(10B)	117.5	H(37B)-C(37A)-H(37C)	109.5
C(21)-C(22)	1.379(6)	C(10)-C(11)-C(12)	113.1(3)	C(36B)-C(35B)-C(34)	119(3)
C(21)-H(21A)	0.95	C(10)-C(11)-H(11A)	109	C(36B)-C(35B)-H(35B)	120.4
C(21)-H(21B)	1	C(12)-C(11)-H(11A)	109	C(34)-C(35B)-H(35B)	120.4
C(22)-C(23)	1.488(6)	C(10)-C(11)-H(11B)	109	C(35B)-C(36B)-C(37B)	124(3)
C(22)-H(22A)	0.95	C(12)-C(11)-H(11B)	109	C(35B)-C(36B)-H(36B)	118
C(22)-H(22B)	1	H(11A)-C(11)-H(11B)	107.8	C(37B)-C(36B)-H(36B)	118
C(23)-C(24)	1.531(5)	O(8)-C(12)-C(11)	106.9(3)	C(36B)-C(37B)-H(37D)	109.5
C(23)-H(23A)	0.99	O(8)-C(12)-C(25)	105.7(2)	C(36B)-C(37B)-H(37E)	109.5
C(23)-H(23B)	0.99	C(11)-C(12)-C(25)	115.4(3)	H(37D)-C(37B)-H(37E)	109.5
C(24)-C(33)	1.556(5)	O(8)-C(12)-H(12)	109.5	C(36B)-C(37B)-H(37F)	109.5
C(24)-H(24)	1	C(11)-C(12)-H(12)	109.5	H(37D)-C(37B)-H(37F)	109.5
C(25)-C(31)	1.525(4)	C(25)-C(12)-H(12)	109.5	H(37E)-C(37B)-H(37F)	109.5
C(25)-C(30)	1.538(5)	O(9)-C(13)-O(8)	125.7(3)	C(33)-C(38)-H(38A)	109.5
C(25)-C(26)	1.565(5)	O(9)-C(13)-C(14)	123.7(3)	C(33)-C(38)-H(38B)	109.5
C(26)-C(27)	1.499(5)	O(8)-C(13)-C(14)	110.6(3)	H(38A)-C(38)-H(38B)	109.5
C(26)-H(26)	1	C(15)-C(14)-N(2)	109.7(3)	C(33)-C(38)-H(38C)	109.5
C(27)-C(28)	1.314(5)	C(15)-C(14)-C(13)	129.7(3)	H(38A)-C(38)-H(38C)	109.5
C(27)-H(27)	0.95	N(2)-C(14)-C(13)	120.7(3)	H(38B)-C(38)-H(38C)	109.5
C(28)-C(29)	1.489(6)	C(14)-C(15)-O(10)	108.1(3)	O(12)-C(39)-O(13)	122.9(3)

C(28)-H(28)	0.95	C(14)-C(15)-H(15)	126	O(12)-C(39)-C(33)	125.5(3)
C(29)-H(29A)	0.98	O(10)-C(15)-H(15)	126	O(13)-C(39)-C(33)	111.5(3)
C(29)-H(29B)	0.98	N(2)-C(16)-O(10)	113.1(3)	O(13)-C(40)-H(40A)	109.5
C(29)-H(29C)	0.98	N(2)-C(16)-C(17)	129.2(3)	O(13)-C(40)-H(40B)	109.5
C(30)-H(30A)	0.98	O(10)-C(16)-C(17)	117.7(3)	H(40A)-C(40)-H(40B)	109.5
C(30)-H(30B)	0.98	C(18)-C(17)-C(16)	122.6(4)	O(13)-C(40)-H(40C)	109.5
C(30)-H(30C)	0.98	C(18)-C(17)-H(17)	118.7	H(40A)-C(40)-H(40C)	109.5
C(32)-H(32A)	0.98	C(16)-C(17)-H(17)	118.7	H(40B)-C(40)-H(40C)	109.5
C(32)-H(32B)	0.98	C(17)-C(18)-C(19)	125.7(4)	C(41)-O(14)-H(14A)	112(7)
C(32)-H(32C)	0.98	C(17)-C(18)-H(18)	117.1	O(14)-C(41)-C(42)	110.6(7)
C(33)-C(39)	1.528(5)	C(19)-C(18)-H(18)	117.1	O(14)-C(41)-H(41A)	109.5
C(33)-C(38)	1.536(5)	C(20)-C(19)-C(18)	123.7(4)	C(42)-C(41)-H(41A)	109.5
C(33)-C(34)	1.560(5)	C(20)-C(19)-H(19)	118.2	O(14)-C(41)-H(41B)	109.5
C(34)-C(35A)	1.502(6)	C(18)-C(19)-H(19)	118.2	C(42)-C(41)-H(41B)	109.5
C(34)-C(35B)	1.55(2)	C(19)-C(20)-C(21)	123.9(4)	H(41A)-C(41)-H(41B)	108.1
C(34)-H(34A)	1	C(19)-C(20)-H(20)	118.1	C(41)-C(42)-H(42A)	109.5
C(34)-H(34B)	1	C(21)-C(20)-H(20)	118.1	C(41)-C(42)-H(42B)	109.5
C(35A)-C(36A)	1.305(7)	C(22)-C(21)-C(20)	125.1(4)	H(42A)-C(42)-H(42B)	109.5
C(35A)-H(35A)	0.95	C(22)-C(21)-O(4B)	57.1(5)	C(41)-C(42)-H(42C)	109.5
C(36A)-C(37A)	1.501(7)	C(20)-C(21)-O(4B)	118.4(7)	H(42A)-C(42)-H(42C)	109.5
C(36A)-H(36A)	0.95	C(22)-C(21)-H(21A)	117.5	H(42B)-C(42)-H(42C)	109.5
C(37A)-H(37A)	0.98	C(20)-C(21)-H(21A)	117.5	C(43B)-O(15)-H(15A)	122(6)
C(37A)-H(37B)	0.98	C(22)-C(21)-H(21B)	114.6	C(43A)-O(15)-H(15A)	108(5)
C(37A)-H(37C)	0.98	C(20)-C(21)-H(21B)	114.6	O(15)-C(43A)-C(44A)	108.0(6)
C(35B)-C(36B)	1.34(2)	O(4B)-C(21)-H(21B)	114.6	O(15)-C(43A)-H(43A)	110.1
C(35B)-H(35B)	0.95	C(21)-C(22)-O(4B)	67.8(6)	C(44A)-C(43A)-H(43A)	110.1
C(36B)-C(37B)	1.52(2)	C(21)-C(22)-C(23)	127.6(4)	O(15)-C(43A)-H(43B)	110.1
C(36B)-H(36B)	0.95	O(4B)-C(22)-C(23)	111.3(7)	C(44A)-C(43A)-H(43B)	110.1
C(37B)-H(37D)	0.98	C(21)-C(22)-H(22A)	116.2	H(43A)-C(43A)-H(43B)	108.4
C(37B)-H(37E)	0.98	C(23)-C(22)-H(22A)	116.2	C(43A)-C(44A)-H(44A)	109.5
C(37B)-H(37F)	0.98	C(21)-C(22)-H(22B)	113.6	C(43A)-C(44A)-H(44B)	109.5
C(38)-H(38A)	0.98	O(4B)-C(22)-H(22B)	113.6	H(44A)-C(44A)-H(44B)	109.5
C(38)-H(38B)	0.98	C(23)-C(22)-H(22B)	113.6	C(43A)-C(44A)-H(44C)	109.5
C(38)-H(38C)	0.98	C(22)-C(23)-C(24)	113.9(3)	H(44A)-C(44A)-H(44C)	109.5
C(40)-H(40A)	0.98	C(22)-C(23)-H(23A)	108.8	H(44B)-C(44A)-H(44C)	109.5
C(40)-H(40B)	0.98	C(24)-C(23)-H(23A)	108.8	O(15)-C(43B)-C(44B)	128(2)
C(40)-H(40C)	0.98	C(22)-C(23)-H(23B)	108.8	O(15)-C(43B)-H(43C)	105.4
O(14)-C(41)	1.380(8)	C(24)-C(23)-H(23B)	108.8	C(44B)-C(43B)-H(43C)	105.4
O(14)-H(14A)	0.864(14)	H(23A)-C(23)-H(23B)	107.7	O(15)-C(43B)-H(43D)	105.4
C(41)-C(42)	1.469(12)	O(1)-C(24)-C(23)	106.3(3)	C(44B)-C(43B)-H(43D)	105.4
C(41)-H(41A)	0.99	O(1)-C(24)-C(33)	106.9(3)	H(43C)-C(43B)-H(43D)	106
C(41)-H(41B)	0.99	C(23)-C(24)-C(33)	115.5(3)	C(43B)-C(44B)-H(44D)	109.5
C(42)-H(42A)	0.98	O(1)-C(24)-H(24)	109.3	C(43B)-C(44B)-H(44E)	109.5
C(42)-H(42B)	0.98	C(23)-C(24)-H(24)	109.3	H(44D)-C(44B)-H(44E)	109.5
C(42)-H(42C)	0.98	C(33)-C(24)-H(24)	109.3	C(43B)-C(44B)-H(44F)	109.5
O(15)-C(43B)	1.22(3)	C(31)-C(25)-C(30)	108.4(3)	H(44D)-C(44B)-H(44F)	109.5
O(15)-C(43A)	1.340(7)	C(31)-C(25)-C(12)	108.8(3)	H(44E)-C(44B)-H(44F)	109.5
O(15)-H(15A)	0.841(14)	C(30)-C(25)-C(12)	111.7(3)		

C(43A)-C(44A)	1.480(10)	C(31)-C(25)-C(26)	105.2(3)
C(43A)-H(43A)	0.99	C(30)-C(25)-C(26)	112.6(3)

Table S2. 34. *Anisotropic displacement parameters ($Å^2x\ 10^3$) for sh3279 (**7**). The anisotropic displacement factor exponent takes the form: $-2\pi^2[\ h^2\ a^{*2}U^{11} + ...\ + 2\ h\ k\ a^*\ b^*\ U^{12}\ ]$.*

	U11	U22	U33	U23	U13	U12
N(1)	28(1)	33(2)	28(2)	-3(1)	0(1)	-2(1)
N(2)	23(1)	29(2)	25(1)	-2(1)	-1(1)	3(1)
O(1)	27(1)	28(1)	28(1)	-5(1)	6(1)	-2(1)
O(2)	36(1)	29(1)	33(1)	-5(1)	5(1)	-2(1)
O(3)	43(2)	33(1)	36(1)	-2(1)	15(1)	3(1)
O(4A)	23(2)	45(2)	37(2)	-4(2)	6(1)	-4(2)
O(4B)	30(6)	124(13)	36(7)	17(7)	1(5)	16(7)
O(5)	27(1)	55(2)	32(1)	-1(1)	10(1)	0(1)
O(6)	23(1)	29(1)	44(2)	7(1)	1(1)	4(1)
O(7)	28(1)	31(1)	37(1)	7(1)	2(1)	-3(1)
O(8)	27(1)	24(1)	21(1)	3(1)	3(1)	3(1)
O(9)	50(2)	25(1)	32(1)	6(1)	8(1)	4(1)
O(10)	29(1)	32(1)	26(1)	0(1)	3(1)	-1(1)
O(11)	46(2)	35(2)	47(2)	-3(1)	15(1)	5(1)
O(12)	41(2)	23(1)	71(2)	-7(1)	1(1)	6(1)
O(13)	29(1)	26(1)	49(2)	-4(1)	-4(1)	-2(1)
C(1)	24(2)	31(2)	26(2)	1(2)	-2(1)	2(2)
C(2)	24(2)	33(2)	27(2)	0(2)	0(1)	2(2)
C(3)	39(2)	38(2)	33(2)	2(2)	12(2)	4(2)
C(4)	28(2)	33(2)	34(2)	-2(2)	-1(2)	0(2)
C(5)	32(2)	40(2)	35(2)	-4(2)	6(2)	-4(2)
C(6)	25(2)	37(2)	30(2)	0(2)	-3(2)	-2(2)
C(7)	28(2)	33(2)	34(2)	-1(2)	0(2)	-7(2)
C(8)	28(2)	30(2)	28(2)	0(2)	-2(2)	-2(2)
C(9)	30(2)	30(2)	34(2)	0(2)	3(2)	-3(2)
C(10)	26(2)	33(2)	33(2)	4(2)	-1(2)	-3(2)
C(11)	26(2)	30(2)	26(2)	1(2)	-1(1)	1(2)
C(12)	23(2)	26(2)	22(2)	4(2)	5(1)	2(1)
C(13)	24(2)	27(2)	25(2)	1(2)	-2(1)	2(1)
C(14)	23(2)	25(2)	24(2)	-2(1)	-7(1)	-2(1)
C(15)	23(2)	27(2)	25(2)	-4(2)	0(1)	-2(1)
C(16)	20(2)	33(2)	26(2)	-2(2)	-2(1)	-1(1)
C(17)	27(2)	41(2)	31(2)	0(2)	1(2)	6(2)
C(18)	26(2)	48(3)	30(2)	-6(2)	-4(2)	2(2)
C(19)	32(2)	51(3)	37(2)	-8(2)	1(2)	7(2)
C(20)	31(2)	43(2)	36(2)	-10(2)	-6(2)	6(2)
C(21)	31(2)	52(3)	36(2)	-7(2)	-4(2)	12(2)
C(22)	29(2)	54(3)	39(2)	-6(2)	-1(2)	9(2)

C(23)	26(2)	47(2)	34(2)	-3(2)	4(2)	-1(2)
C(24)	26(2)	28(2)	27(2)	-5(2)	5(1)	1(1)
C(25)	21(1)	23(2)	22(2)	2(1)	1(1)	2(1)
C(26)	26(2)	29(2)	27(2)	4(2)	5(1)	4(2)
C(27)	34(2)	29(2)	28(2)	0(2)	5(2)	5(2)
C(28)	80(3)	42(3)	33(2)	-2(2)	-14(2)	7(2)
C(29)	109(5)	64(3)	52(3)	-14(3)	-34(3)	9(3)
C(30)	26(2)	26(2)	27(2)	4(2)	0(2)	4(1)
C(31)	26(2)	26(2)	20(2)	2(1)	0(1)	3(1)
C(32)	23(2)	47(2)	73(3)	18(2)	1(2)	11(2)
C(33)	33(2)	21(2)	34(2)	-2(2)	3(2)	-2(2)
C(34)	43(2)	23(2)	35(2)	0(2)	4(2)	0(2)
C(35A)	42(3)	32(3)	36(2)	5(2)	15(2)	0(2)
C(36A)	46(3)	47(3)	34(2)	3(2)	11(2)	11(2)
C(37A)	68(4)	78(5)	43(3)	17(3)	10(3)	30(4)
C(35B)	41(7)	33(7)	33(6)	1(6)	13(7)	-2(7)
C(36B)	49(7)	49(7)	35(6)	8(6)	12(6)	9(6)
C(37B)	82(14)	85(14)	44(11)	21(11)	7(12)	35(12)
C(38)	38(2)	30(2)	49(2)	-7(2)	7(2)	-7(2)
C(39)	35(2)	22(2)	35(2)	1(2)	-4(2)	-1(2)
C(40)	31(2)	43(3)	81(3)	-5(2)	-7(2)	-1(2)
O(14)	189(6)	73(3)	103(4)	-10(3)	7(4)	-9(4)
C(41)	121(6)	79(5)	156(8)	-39(5)	-32(6)	1(4)
C(42)	81(5)	68(4)	218(10)	20(5)	-43(5)	-1(4)
O(15)	99(3)	30(2)	122(3)	-11(2)	-61(3)	-1(2)
C(43A)	59(4)	64(4)	57(4)	-10(3)	-14(3)	-10(3)
C(44A)	133(8)	80(5)	78(5)	-6(5)	-62(5)	45(5)
C(43B)	64(7)	65(6)	81(6)	-9(6)	-12(6)	-8(7)
C(44B)	158(15)	89(10)	101(12)	1(9)	-42(12)	41(11)

Table S2. 35. *Hydrogen coordinates (x 10^4) and isotropic displacement parameters (Å^2 x 10^3) for* ***7****.*

	x	y	z	U(eq)		x	y	z	U(eq)
H(14)	9520(50)	3130(20)	-1534(13)	58	H(32B)	2639	3584	-584	71
H(29)	430(60)	4371(14)	5321(18)	64	H(32C)	2658	3305	-1160	71
H(3)	4796	3289	3087	44	H(34A)	2754	4290	5202	40
H(5)	7466	5103	2033	43	H(34B)	2652	4320	5145	40
H(6)	7859	3799	1717	37	H(35A)	1902	3000	5590	44
H(7)	8895	5005	1249	38	H(36A)	4196	3865	5878	51
H(8)	9233	3736	853	34	H(37A)	3752	3273	6681	94
H(9A)	9768	5038	409	38	H(37B)	4982	2883	6339	94
H(9B)	10296	4969	440	38	H(37C)	3226	2649	6346	94
H(10A)	9688	4625	-399	37	H(35B)	4203	3472	5525	43
H(10B)	10177	4557	-393	37	H(36B)	1299	3169	5927	53
H(11A)	9330	3311	33	33	H(37D)	2570	2278	6339	105

H(11B)	9863	3429	-548	33
H(12)	7343	4010	-658	28
H(15)	5448	3269	933	30
H(17)	2570	5350	1546	39
H(18)	2521	4175	2109	41
H(19)	1028	5392	2362	48
H(20)	1152	4214	2928	44
H(21A)	-281	5433	3201	47
H(21B)	163	5507	3247	47
H(22A)	-507	5079	4013	49
H(22B)	-54	5148	4041	49
H(23A)	397	3799	3718	43
H(23B)	-215	4012	4274	43
H(24)	2379	4582	4344	32
H(26)	7383	3450	-1439	33
H(27)	6800	2051	-1523	37
H(28)	6110	3096	-2171	62
H(29A)	5080	2382	-2740	112
H(29B)	6300	1842	-2549	112
H(29C)	4672	1906	-2268	112
H(30A)	7047	1964	-617	40
H(30B)	7089	2364	-87	40
H(30C)	8600	2318	-436	40
H(32A)	2787	4083	-1059	71
H(37E)	3297	2910	6618	105
H(37F)	4201	2559	6157	105
H(38A)	2021	2602	4733	59
H(38B)	1798	2807	4143	59
H(38C)	539	3015	4561	59
H(40A)	7218	4340	4697	78
H(40B)	6942	3786	4267	78
H(40C)	7086	3578	4859	78
H(14A)	4210(120)	2470(50)	3771(19)	183
H(41A)	5242	1507	3551	142
H(41B)	3904	1442	3130	142
H(42A)	3481	852	3918	183
H(42B)	2092	1316	3738	183
H(42C)	3227	1557	4185	183
H(15A)	5210(80)	5828(16)	3060(30)	126
H(43A)	2944	6059	2938	72
H(43B)	3744	6048	2381	72
H(44A)	2392	7036	2494	145
H(44B)	3392	7198	2995	145
H(44C)	4189	7187	2440	145
H(43C)	3898	6073	2416	84
H(43D)	5677	6115	2297	84
H(44D)	4411	6889	1894	174
H(44E)	3751	7175	2423	174
H(44F)	5551	7217	2304	174

Table S2. 36. *Torsion angles [°] for sh3279 (**7**).*

C(24)-O(1)-C(1)-O(2)	-1.5(5)	O(8)-C(13)-C(14)-N(2)	-174.5(3)	C(26)-C(27)-C(28)-C(29)	175.2(5)
C(24)-O(1)-C(1)-C(2)	178.9(3)	N(2)-C(14)-C(15)-O(10)	0.4(4)	C(32)-O(6)-C(31)-O(7)	3.7(5)
C(4)-N(1)-C(2)-C(3)	0.4(4)	C(13)-C(14)-C(15)-O(10)	179.2(3)	C(32)-O(6)-C(31)-C(25)	-177.5(3)
C(4)-N(1)-C(2)-C(1)	-179.2(3)	C(16)-O(10)-C(15)-C(14)	0.0(3)	C(30)-C(25)-C(31)-O(7)	-14.3(5)
O(2)-C(1)-C(2)-C(3)	169.8(4)	C(14)-N(2)-C(16)-O(10)	0.7(4)	C(12)-C(25)-C(31)-O(7)	-136.0(3)
O(1)-C(1)-C(2)-C(3)	-10.6(5)	C(14)-N(2)-C(16)-C(17)	180.0(3)	C(26)-C(25)-C(31)-O(7)	106.4(4)
O(2)-C(1)-C(2)-N(1)	-10.8(5)	C(15)-O(10)-C(16)-N(2)	-0.4(4)	C(30)-C(25)-C(31)-O(6)	166.9(3)
O(1)-C(1)-C(2)-N(1)	168.8(3)	C(15)-O(10)-C(16)-C(17)	-179.8(3)	C(12)-C(25)-C(31)-O(6)	45.2(4)
N(1)-C(2)-C(3)-O(3)	-0.1(4)	N(2)-C(16)-C(17)-C(18)	-179.3(3)	C(26)-C(25)-C(31)-O(6)	-72.4(3)
C(1)-C(2)-C(3)-O(3)	179.4(3)	O(10)-C(16)-C(17)-C(18)	0.0(5)	O(1)-C(24)-C(33)-C(39)	35.5(4)
C(4)-O(3)-C(3)-C(2)	-0.3(4)	C(16)-C(17)-C(18)-C(19)	178.6(3)	C(23)-C(24)-C(33)-C(39)	153.5(3)
C(2)-N(1)-C(4)-O(3)	-0.6(4)	C(17)-C(18)-C(19)-C(20)	-174.9(4)	O(1)-C(24)-C(33)-C(38)	-83.7(3)
C(2)-N(1)-C(4)-C(5)	176.9(4)	C(18)-C(19)-C(20)-C(21)	172.9(4)	C(23)-C(24)-C(33)-C(38)	34.3(4)
C(3)-O(3)-C(4)-N(1)	0.6(4)	C(19)-C(20)-C(21)-C(22)	-165.7(4)	O(1)-C(24)-C(33)-C(34)	152.4(3)
C(3)-O(3)-C(4)-C(5)	-177.2(3)	C(19)-C(20)-C(21)-O(4B)	126.4(7)	C(23)-C(24)-C(33)-C(34)	-89.6(4)
N(1)-C(4)-C(5)-C(6)	-176.9(4)	C(22)-O(4B)-C(21)-C(20)	115.4(5)	C(39)-C(33)-C(34)-O(11)	178.5(3)
O(3)-C(4)-C(5)-C(6)	0.4(6)	C(20)-C(21)-C(22)-O(4B)	-103.7(8)	C(38)-C(33)-C(34)-O(11)	-62.1(4)

C(4)-C(5)-C(6)-C(7)	174.9(3)	C(20)-C(21)-C(22)-C(23)	-3.3(7)	C(24)-C(33)-C(34)-O(11)	62.6(4)
C(5)-C(6)-C(7)-C(8)	-172.8(4)	O(4B)-C(21)-C(22)-C(23)	100.4(8)	C(39)-C(33)-C(34)-C(35A)	-63.2(4)
C(6)-C(7)-C(8)-C(9)	171.4(3)	C(21)-O(4B)-C(22)-C(23)	-123.3(5)	C(38)-C(33)-C(34)-C(35A)	56.2(5)
C(7)-C(8)-C(9)-C(10)	-151.5(4)	C(21)-C(22)-C(23)-C(24)	88.4(5)	C(24)-C(33)-C(34)-C(35A)	-179.1(3)
C(7)-C(8)-C(9)-O(4A)	136.9(4)	O(4B)-C(22)-C(23)-C(24)	166.2(6)	C(39)-C(33)-C(34)-C(35B)	-41.3(14)
C(10)-O(4A)-C(9)-C(8)	115.1(4)	C(1)-O(1)-C(24)-C(23)	108.7(3)	C(38)-C(33)-C(34)-C(35B)	78.1(14)
C(8)-C(9)-C(10)-O(4A)	-107.0(4)	C(1)-O(1)-C(24)-C(33)	-127.4(3)	C(24)-C(33)-C(34)-C(35B)	-157.2(14)
C(8)-C(9)-C(10)-C(11)	-3.4(6)	C(22)-C(23)-C(24)-O(1)	-84.4(4)	O(11)-C(34)-C(35A)-C(36A)	-127.8(5)
O(4A)-C(9)-C(10)-C(11)	103.6(4)	C(22)-C(23)-C(24)-C(33)	157.3(3)	C(33)-C(34)-C(35A)-C(36A)	114.2(5)
C(9)-O(4A)-C(10)-C(11)	-117.2(4)	O(8)-C(12)-C(25)-C(31)	49.6(3)	C(34)-C(35A)-C(36A)-C(37A)	177.0(5)
C(9)-C(10)-C(11)-C(12)	95.3(4)	C(11)-C(12)-C(25)-C(31)	167.5(3)	O(11)-C(34)-C(35B)-C(36B)	11(5)
O(4A)-C(10)-C(11)-C(12)	167.3(3)	O(8)-C(12)-C(25)-C(30)	-70.1(3)	C(33)-C(34)-C(35B)-C(36B)	-119(3)
C(13)-O(8)-C(12)-C(11)	99.7(3)	C(11)-C(12)-C(25)-C(30)	47.8(4)	C(34)-C(35B)-C(36B)-C(37B)	175(4)
C(13)-O(8)-C(12)-C(25)	-136.8(3)	O(8)-C(12)-C(25)-C(26)	164.2(2)	C(40)-O(13)-C(39)-O(12)	2.2(6)
C(10)-C(11)-C(12)-O(8)	-78.4(3)	C(11)-C(12)-C(25)-C(26)	-77.9(3)	C(40)-O(13)-C(39)-C(33)	-175.8(3)
C(10)-C(11)-C(12)-C(25)	164.4(3)	C(31)-C(25)-C(26)-O(5)	-170.8(3)	C(38)-C(33)-C(39)-O(12)	6.6(5)
C(12)-O(8)-C(13)-O(9)	0.7(5)	C(30)-C(25)-C(26)-O(5)	-52.9(4)	C(24)-C(33)-C(39)-O(12)	-115.4(4)
C(12)-O(8)-C(13)-C(14)	-178.1(3)	C(12)-C(25)-C(26)-O(5)	72.2(3)	C(34)-C(33)-C(39)-O(12)	127.8(4)
C(16)-N(2)-C(14)-C(15)	-0.7(4)	C(31)-C(25)-C(26)-C(27)	-49.5(4)	C(38)-C(33)-C(39)-O(13)	-175.5(3)
C(16)-N(2)-C(14)-C(13)	-179.5(3)	C(30)-C(25)-C(26)-C(27)	68.4(4)	C(24)-C(33)-C(39)-O(13)	62.6(4)
O(9)-C(13)-C(14)-C(15)	-171.9(3)	C(12)-C(25)-C(26)-C(27)	-166.5(3)	C(34)-C(33)-C(39)-O(13)	-54.3(4)
O(8)-C(13)-C(14)-C(15)	6.9(5)	O(5)-C(26)-C(27)-C(28)	-111.1(4)		
O(9)-C(13)-C(14)-N(2)	6.7(5)	C(25)-C(26)-C(27)-C(28)	128.0(4)		

Table S2. 37. *Hydrogen bonds for sh3279 (**7**) [Å and °].*

D-H...A	d(D-H)	d(H...A)	d(D...A)	<(DHA)
O(11)-H(29)...N(2)#1	0.843(14)	2.218(19)	3.042(4)	166(5)
O(5)-H(14)...O(15)#2	0.843(14)	1.928(17)	2.760(5)	169(5)
O(14)-H(14A)...O(12)	0.864(14)	2.023(17)	2.885(6)	175(10)
O(15)-H(15A)...N(1)	0.841(14)	2.28(5)	3.004(5)	144(7)

Symmetry transformations used to generate equivalent atoms:
#1 -x+1/2,-y+1,z+1/2 #2 -x+3/2,-y+1,z-1/2

S 2.3 Supplemental Figures

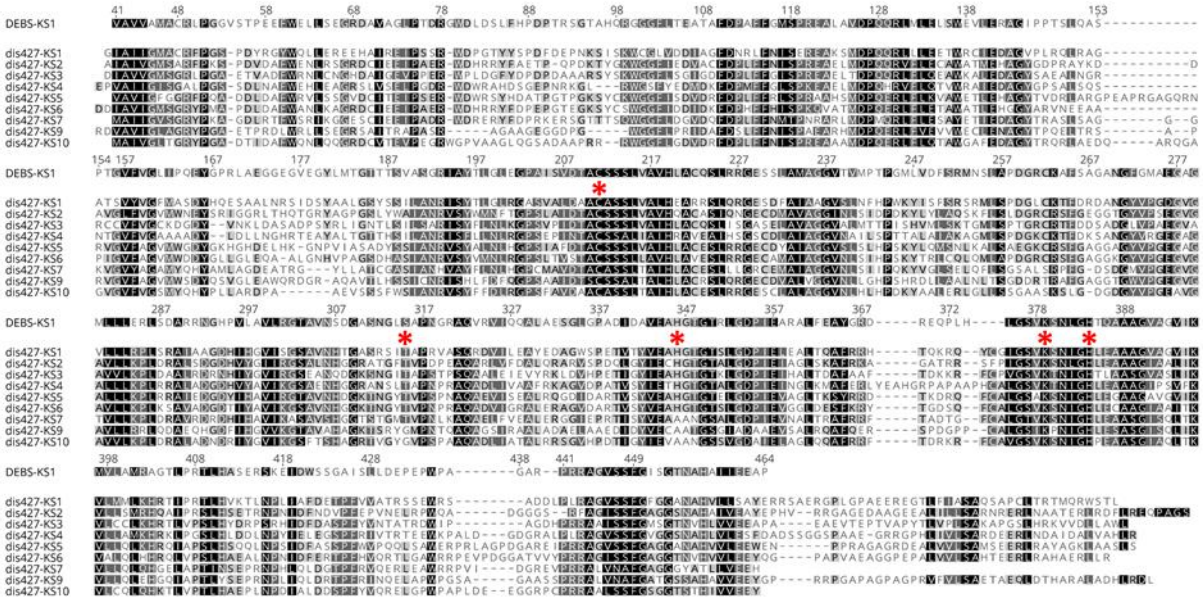


Figure S2. 7. The aligned KS domains. *DEBS-KS1 is the first KS domain of 6-deoxyerythronolide B synthase.[6] dis427-KS with corresponding module numbers are KS domains from the* S. cellulosum *So ce427 disorazole Z biosynthetic pathway. The conserved active site are marked by red asterisks.*

Figure S2. 8. The aligned CP domains. *DEBS-CP2 is the CP domain from the second module of 6-deoxyerythronolide B synthase.[7] dis427-CP with corresponding module numbers are CP domains from the* S. cellulosum *So ce427 disorazole Z biosynthetic pathway. The serine for phosphopantetheinyl arm attachment is marked by red asterisk.*

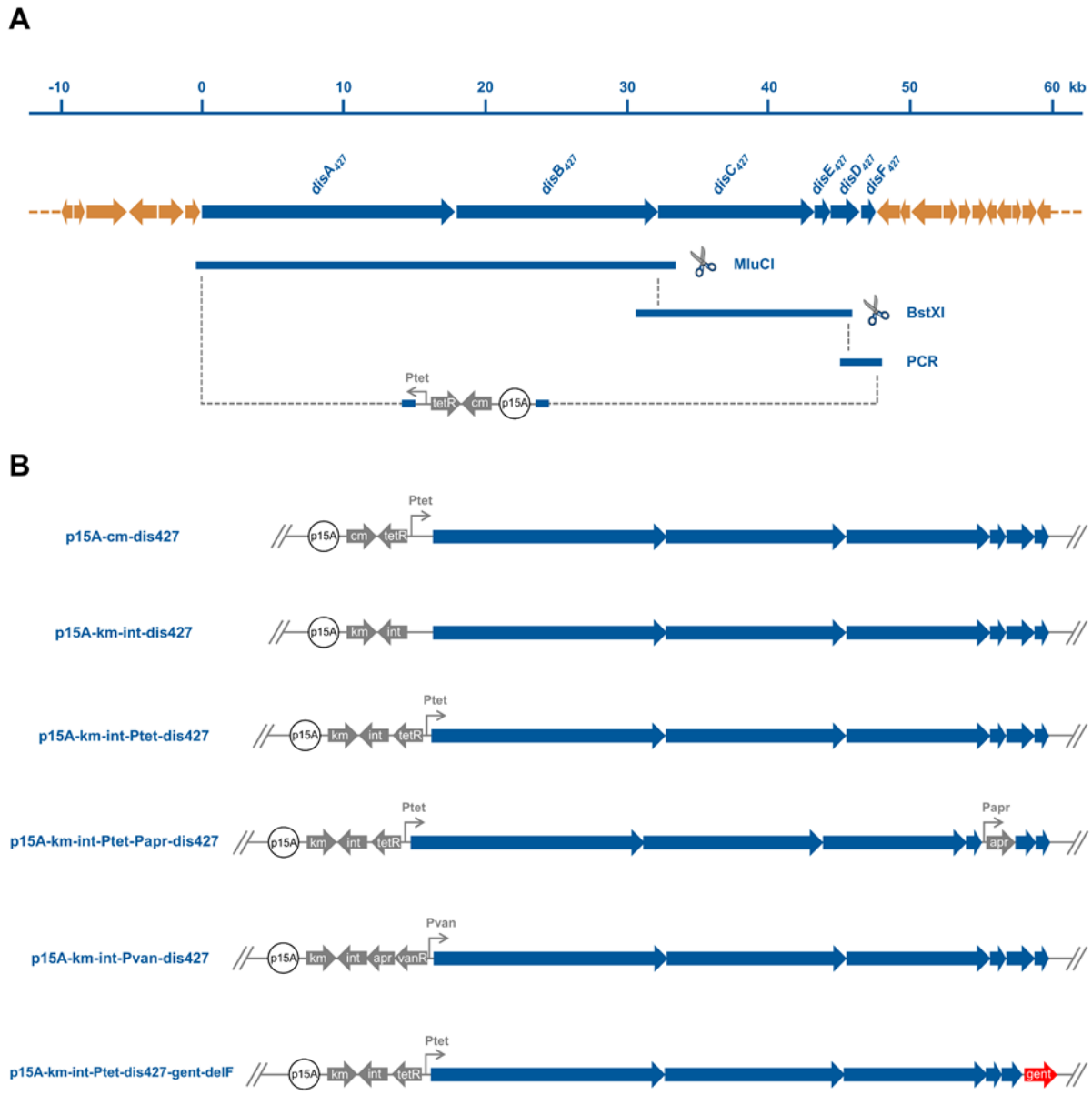
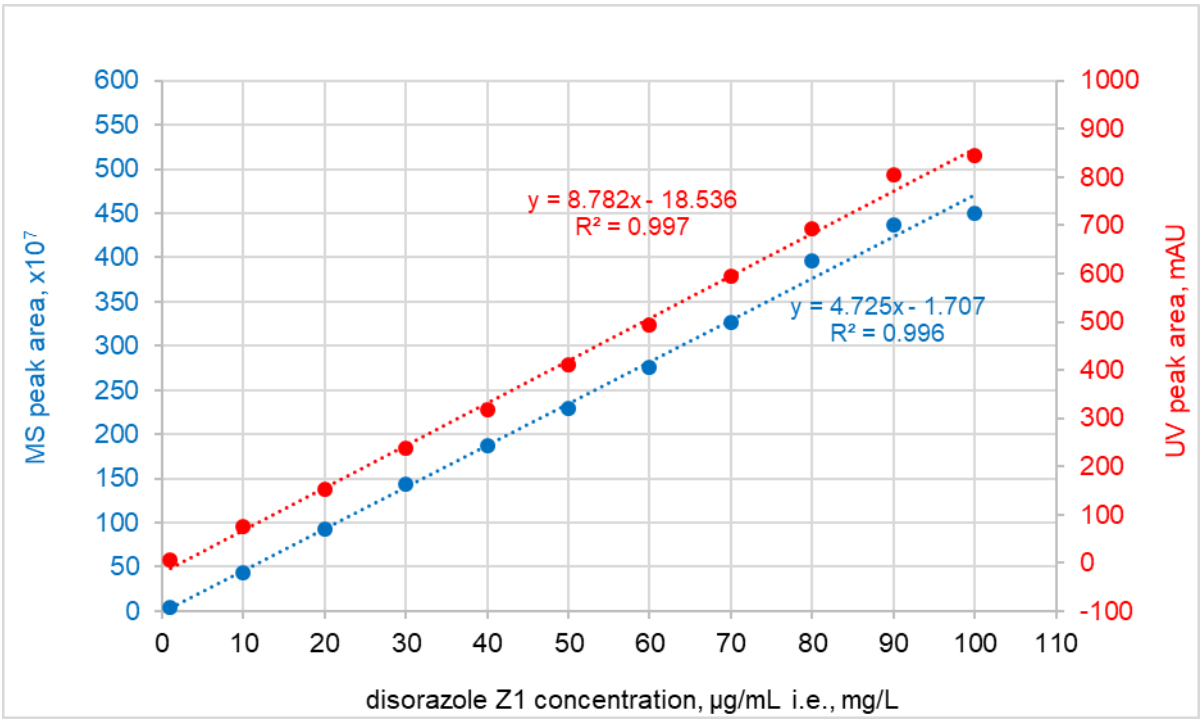


Figure S2. 9. Cloning and engineering of the dis427 gene cluster for heterologous expression. *A: the disorazole Z biosynthesis gene cluster was cloned from the genomic DNA of* S. cellulosum *So ce427 using LLHR as described in methods section. B: scheme of plasmids containing dis427 gene cluster generated in this study, the expression constructs are created using LCHR as described in methods section.*

A



B

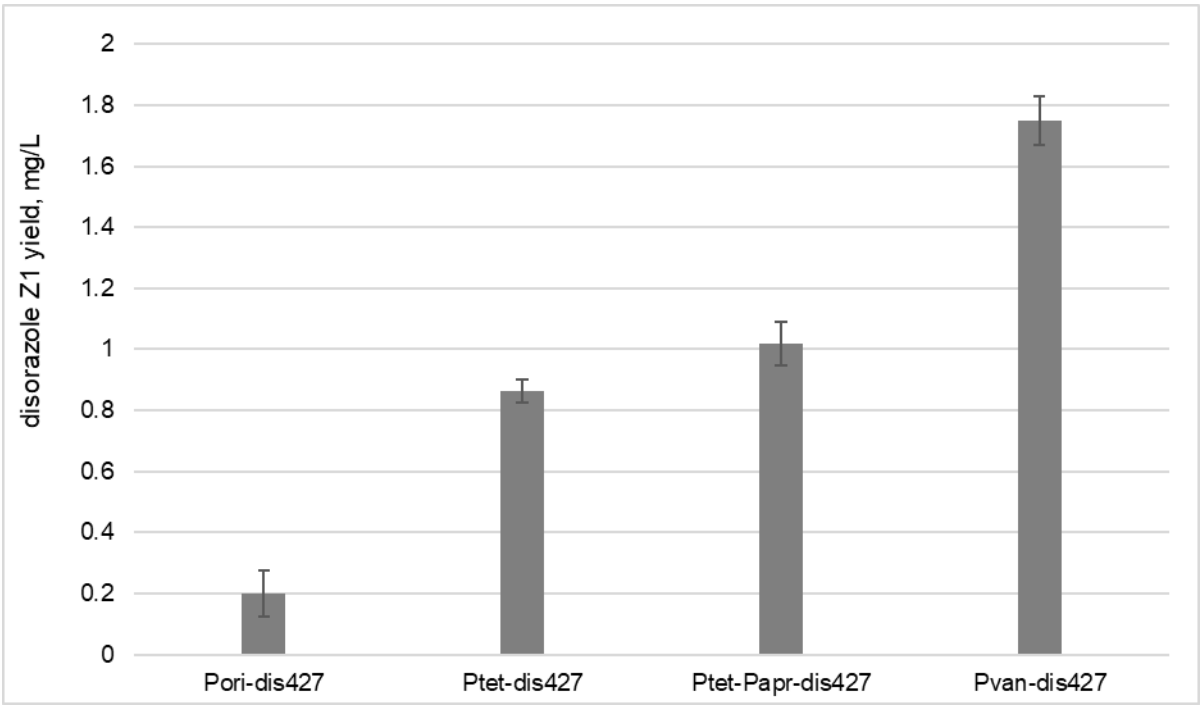


Figure S2. 10. Quantitative analysis of disorazole Z1 (3) using HPLC-UV-MS. *A: the standard curve of* ***3****. B: heterologous production yields of* ***3*** *in* M. xanthus *DK1622 before and after promoter engineering.*

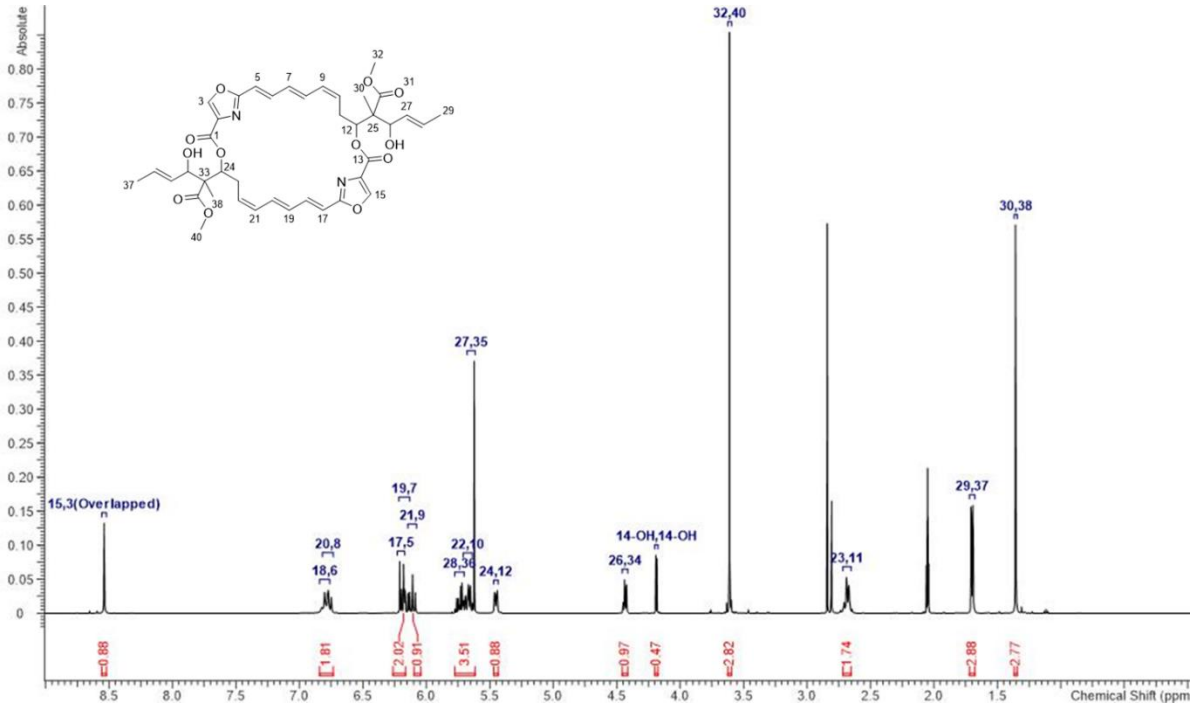


Figure S2. 11. *^{1}H NMR spectrum of disorazole Z1 (**3**) in acetone-d_6 (500 MHz).*

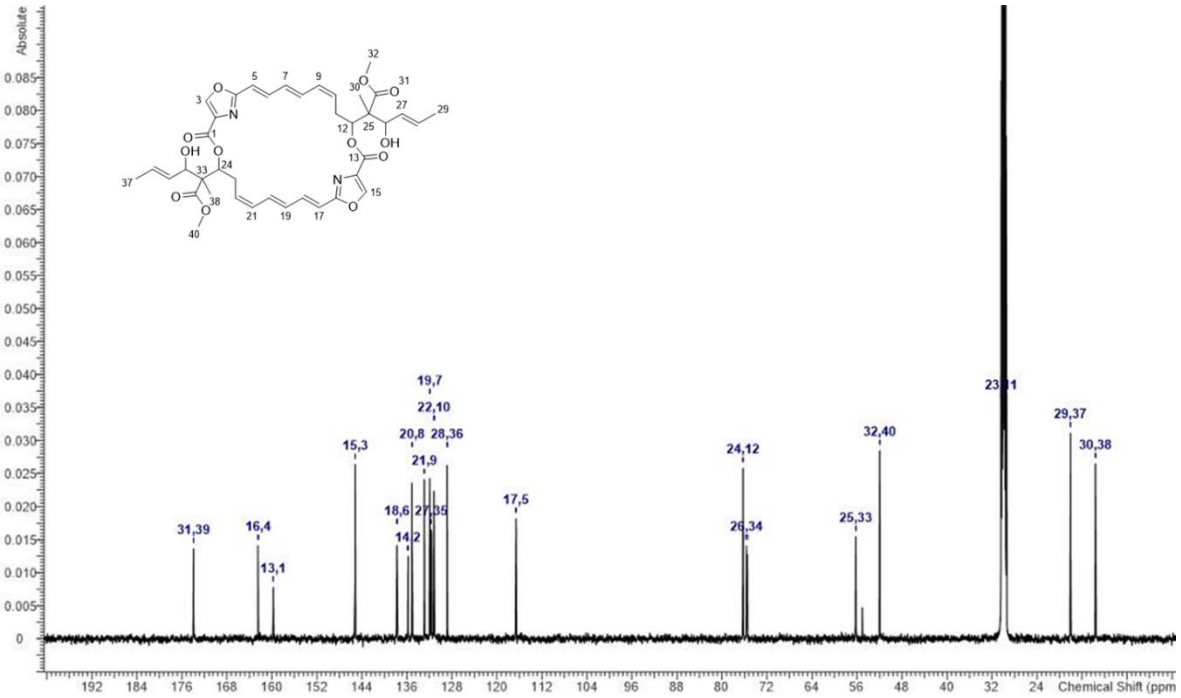


Figure S2. 12. *^{13}C NMR spectrum of disorazole Z1 (**3**) in acetone-d_6 (100 MHz).*

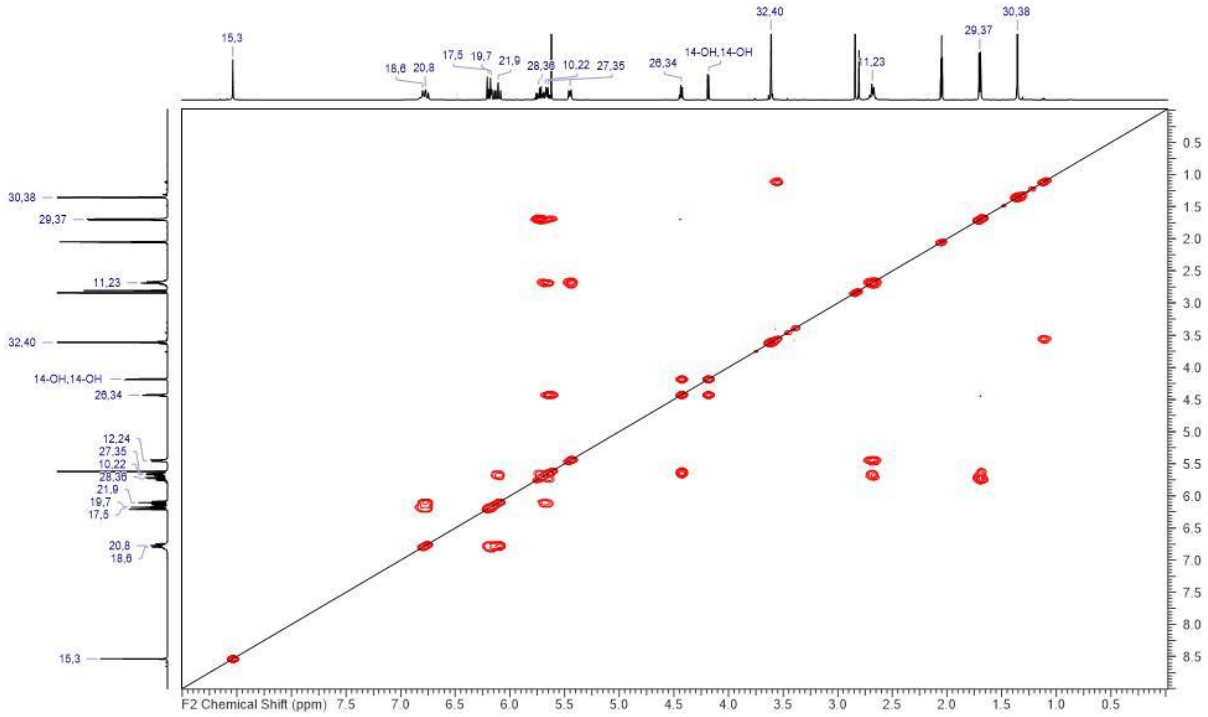


Figure S2. 13. *^{1}H, ^{1}H-COSY NMR spectrum of disorazole Z1 (**3**) in acetone-d_6.*

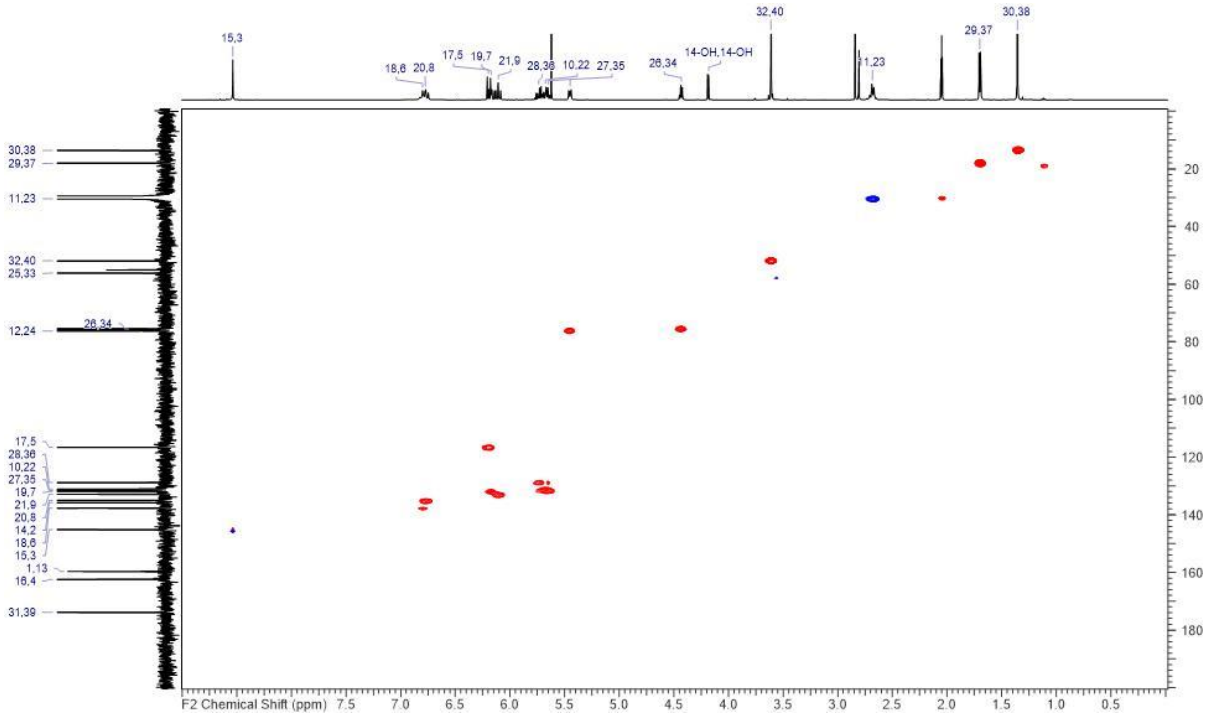


Figure S2. 14. *HSQC-NMR spectrum of disorazole Z1 (**3**) in acetone-d_6.*

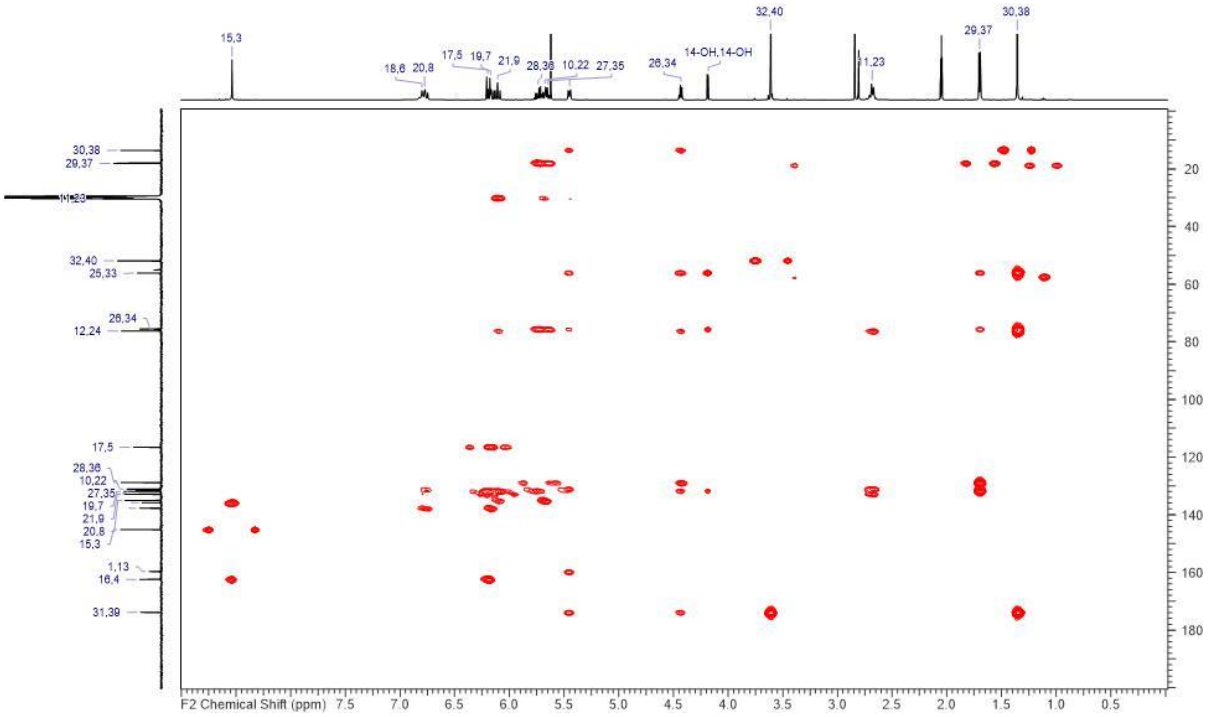


Figure S2. 15. *HMBC-NMR spectrum of disorazole Z1 (**3**) in acetone-d_6.*

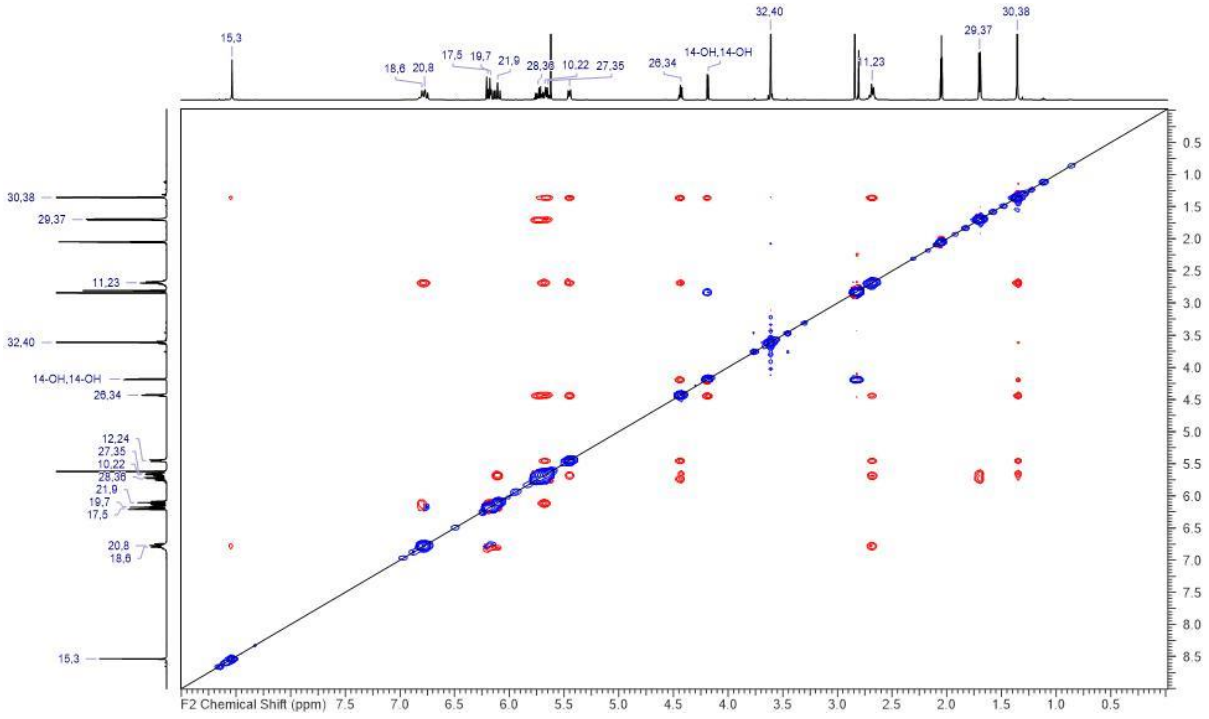


Figure S2. 16. *^{1}H,^{1}H-ROESY NMR spectrum of disorazole Z1 (**3**) in acetone-d_6.*

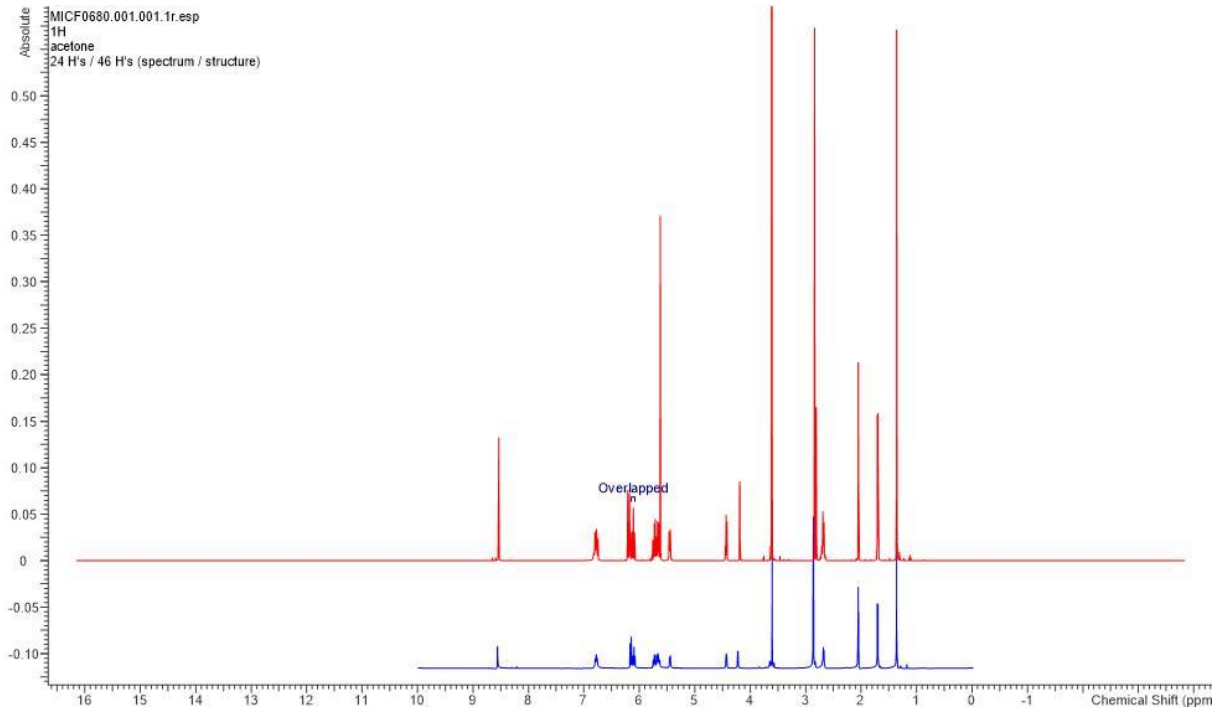


Figure S2. 17. *Comparison of* ^{1}*H NMR spectra of* ***3*** *in acetone-*d_6 *isolated from* S. cellulosum *and* M. xanthus*. Top/red: isolated from* S. cellulosum *So ce1875; Bottom/blue: isolated from* M. xanthus *DK1622 :: km-int-Ptet-dis427.*

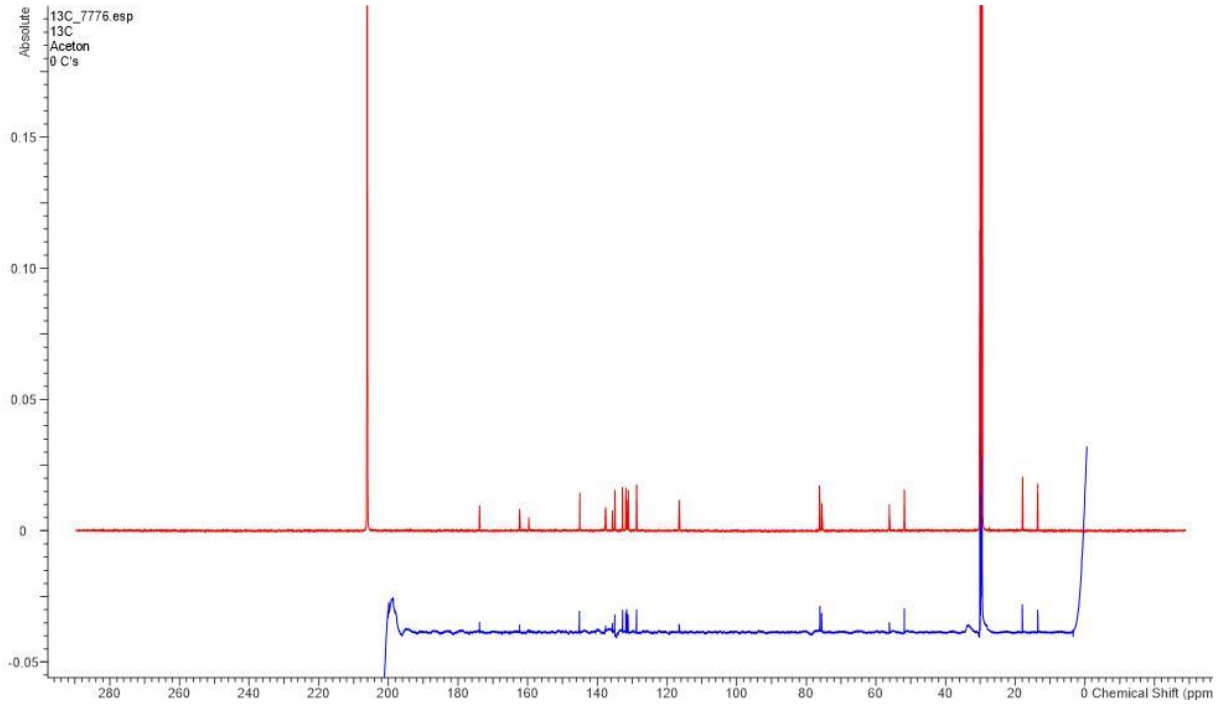


Figure S2. 18. *Comparison of* ^{13}*C NMR spectra of* ***3*** *in acetone-*d_6 *isolated from* S. cellulosum *and* M. xanthus*. Top/red: isolated from* S. cellulosum *So ce1875; Bottom/blue: isolated from* M. xanthus *DK1622 :: km-int-Ptet-dis427.*

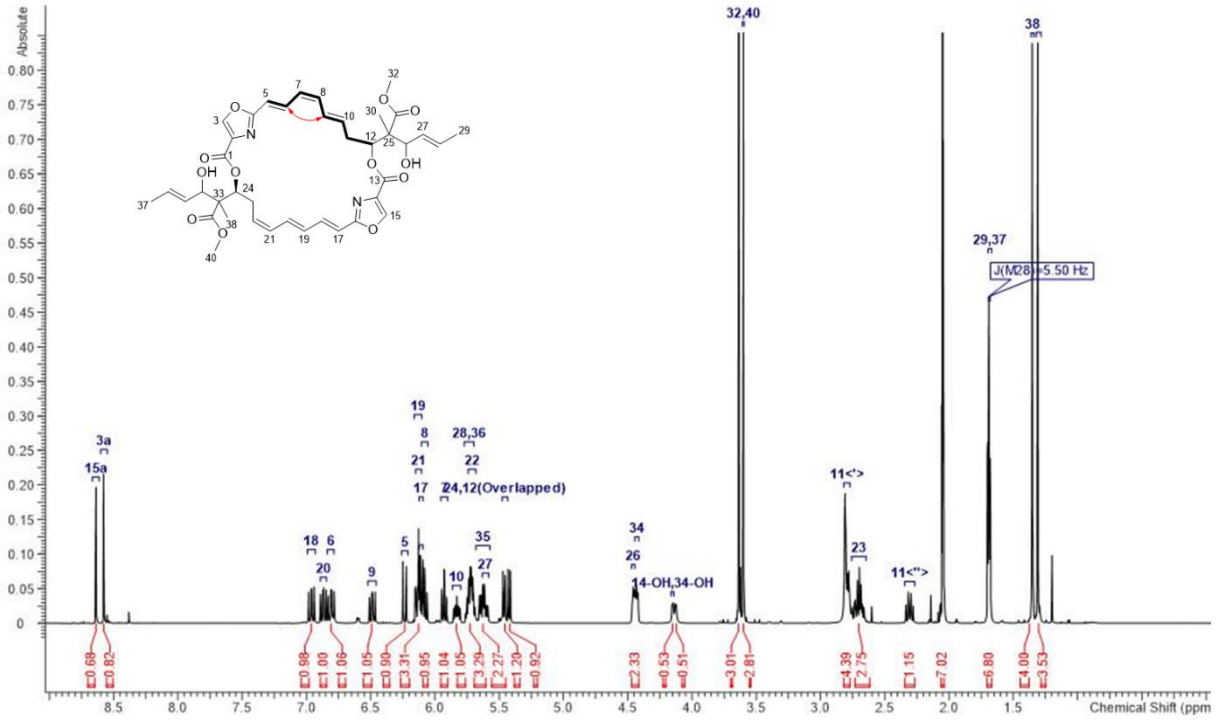


Figure S2. 19. *^{1}H NMR spectrum of $\Delta^{7,8}$-cis-disorazole Z (**4**) in acetone-d_6 (500 MHz).*

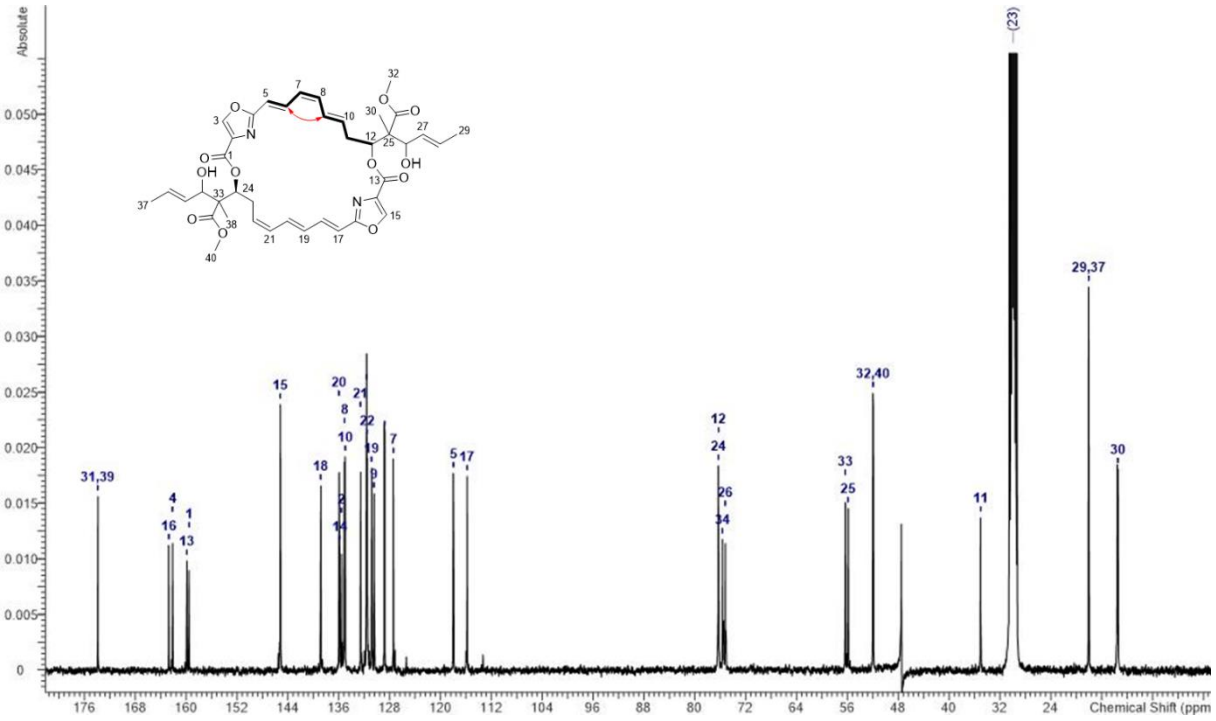


Figure S2. 20. *^{13}C NMR spectrum of $\Delta^{7,8}$-cis-disorazole Z (**4**) in acetone-d_6 (500 MHz).*

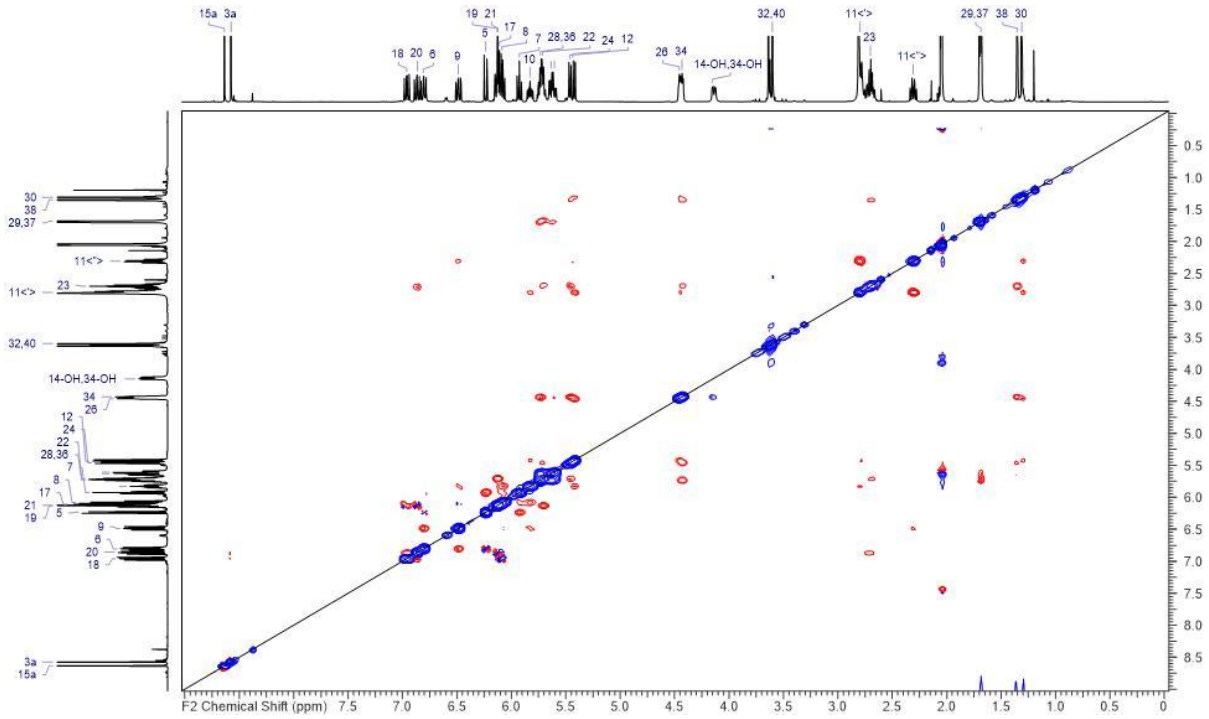


Figure S2. 21. *$^{1}H,^{1}H$-ROESY NMR spectrum of $\Delta^{7,8}$-cis-disorazole Z (**4**) in acetone-d_6.*

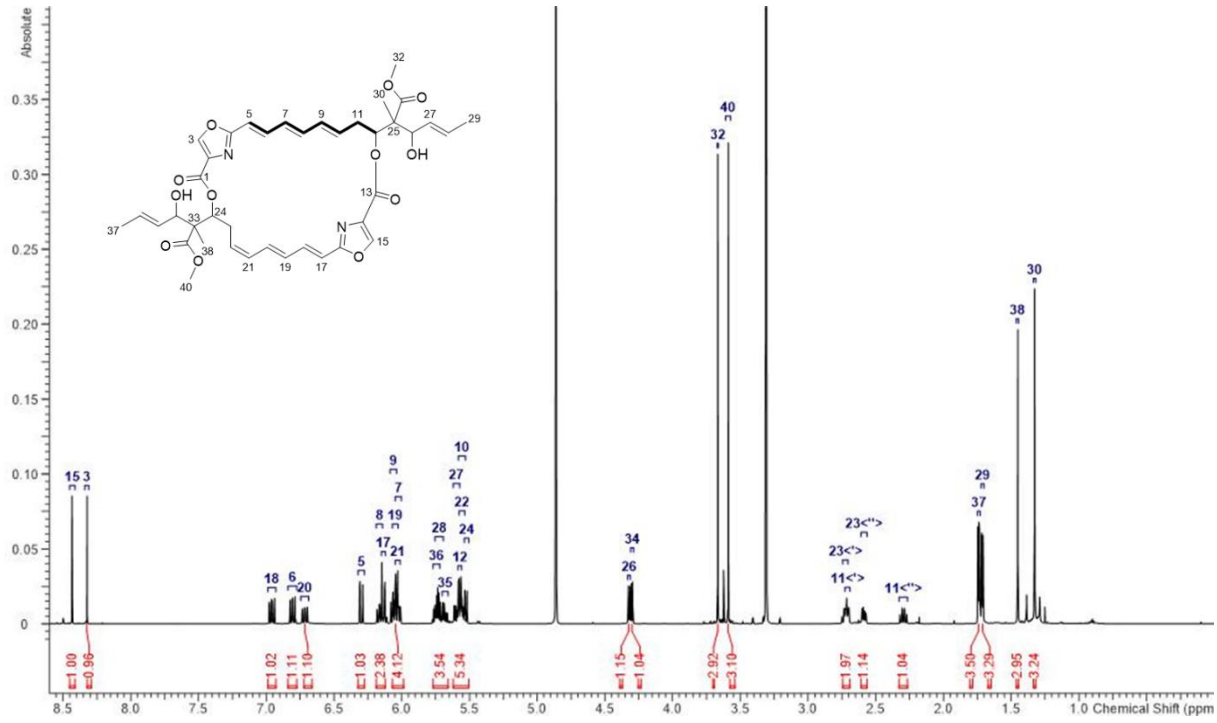


Figure S2. 22. *^{1}H NMR spectrum of $\Delta^{9,10}$-trans-disorazole Z (**5**) in methanol-d_4 (700 MHz).*

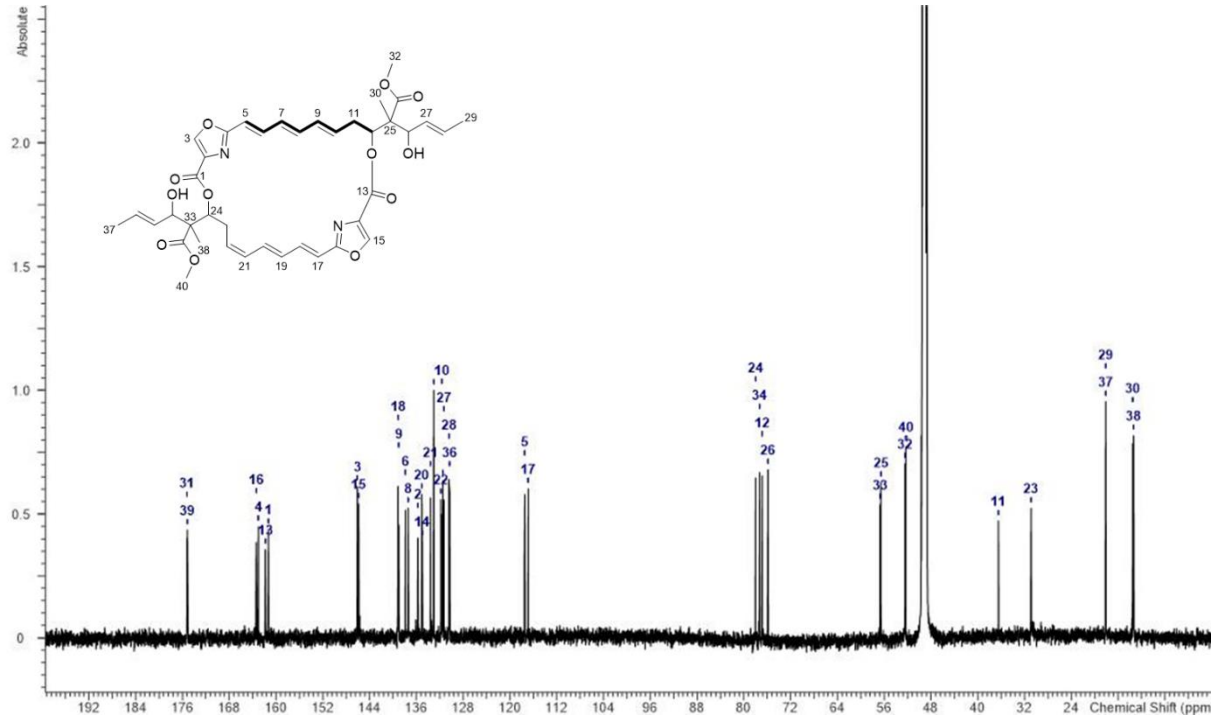


Figure S2. 23. *^{13}C NMR spectrum of $\Delta^{9,10}$-trans-disorazole Z (**5**) in in methanol-d_4 (175 MHz).*

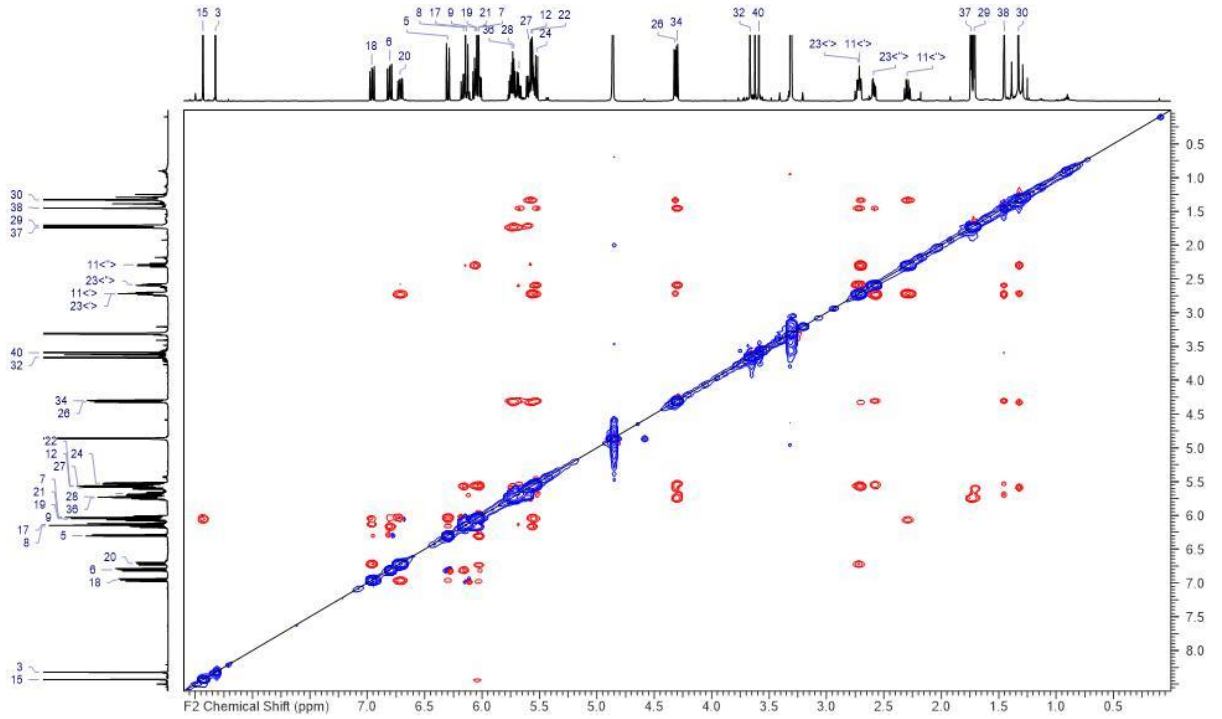


Figure S2. 24. *^{1}H,^{1}H-ROESY NMR spectrum of $\Delta^{9,10}$-trans-disorazole Z (**5**) in methanol-d_4.*

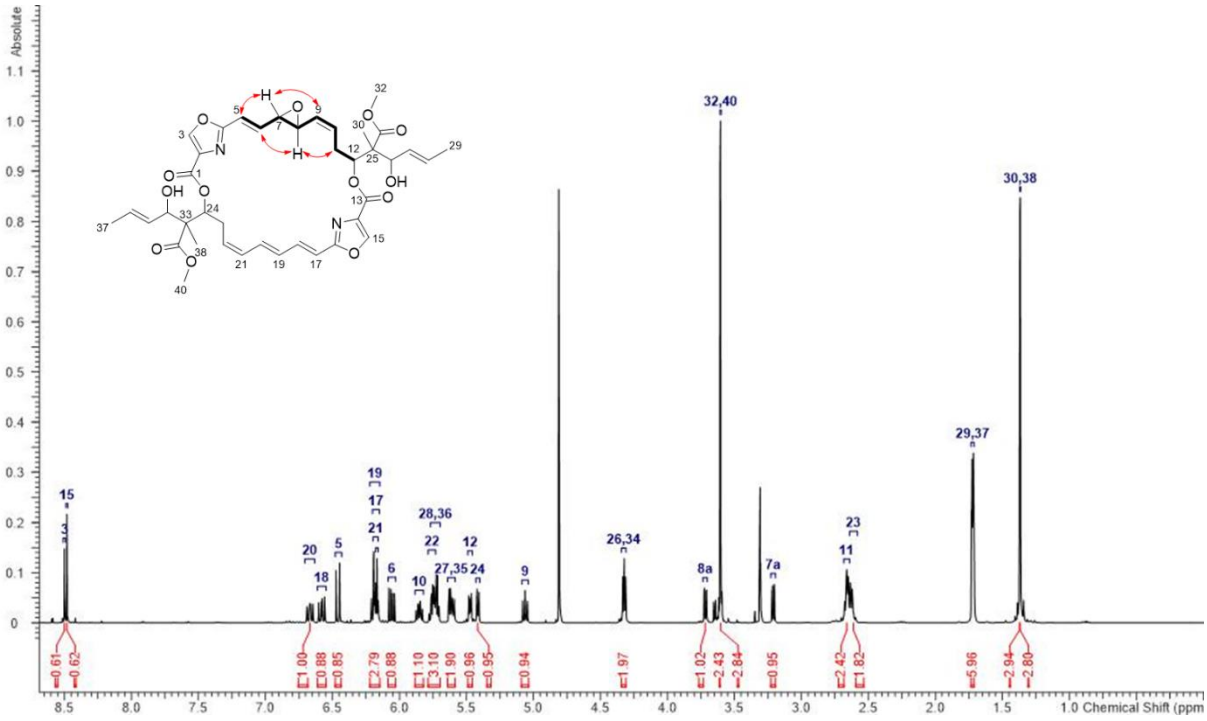


Figure S2. 25. *^{1}H NMR spectrum of 7,8-epoxi-disorazole Z (**6**) in methanol-d_4 (600 MHz).*

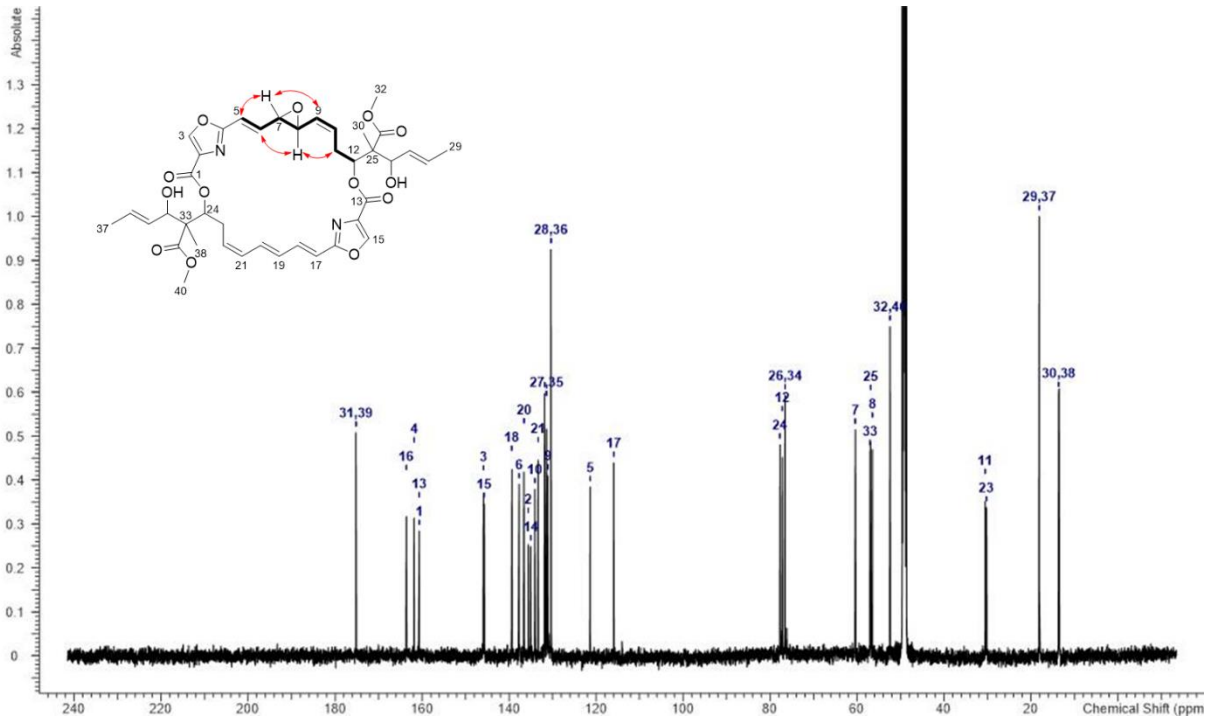


Figure S2. 26. *^{13}C NMR spectrum of 7,8-epoxi-disorazole Z (**6**) in in methanol-d_4 (150 MHz).*

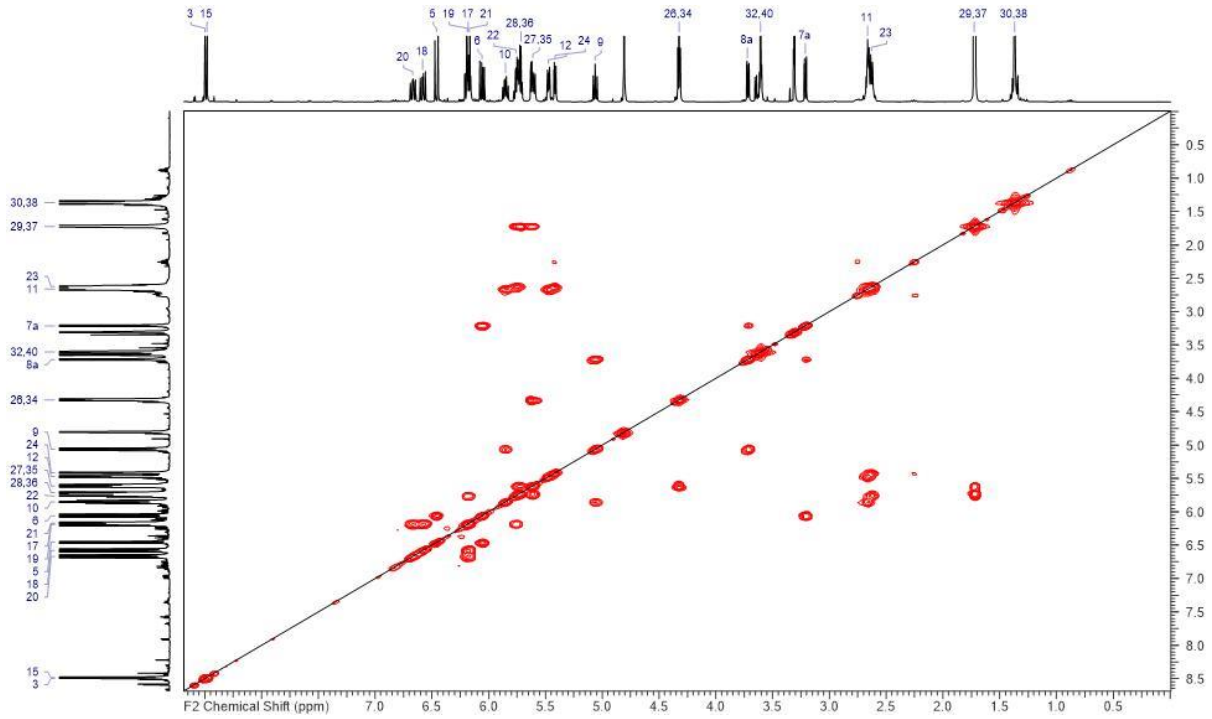


Figure S2. 27. *$^{1}H,^{1}H$-COSY NMR spectrum of 7,8-epoxi-disorazole Z (**6**) in in methanol-d_4.*

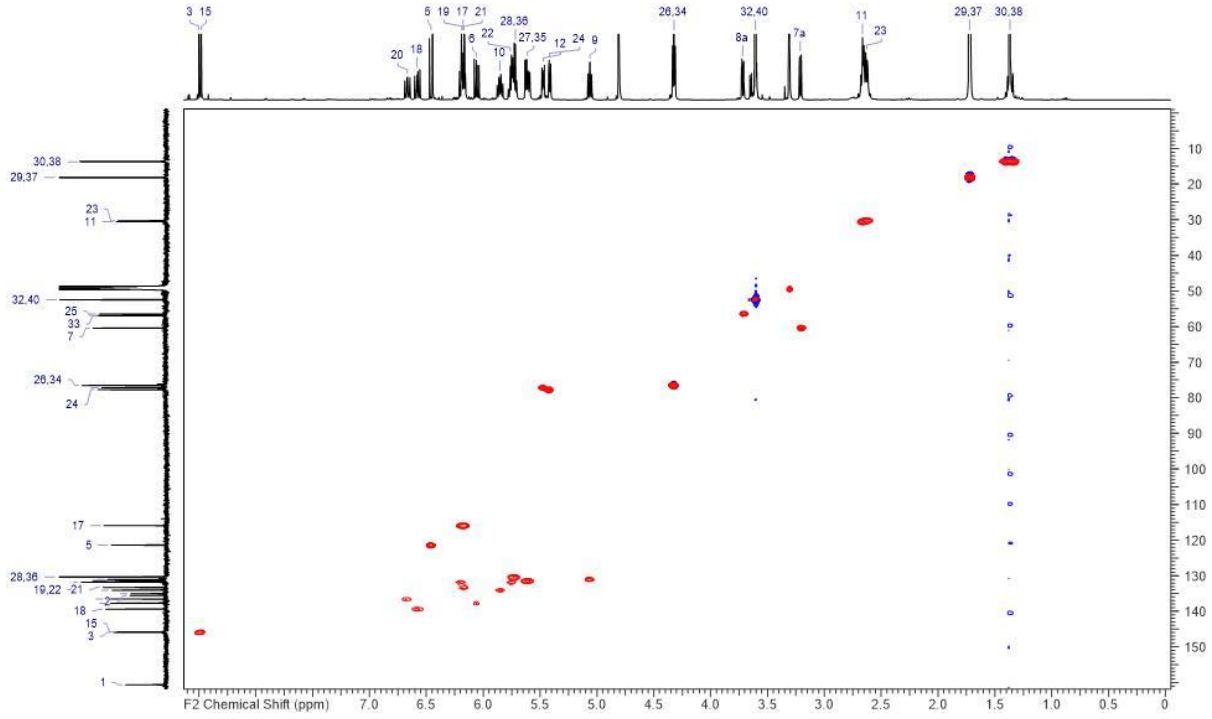


Figure S2. 28. *HMQC-NMR spectrum of 7,8-epoxi-disorazole Z (**6**) in in methanol-d_4.*

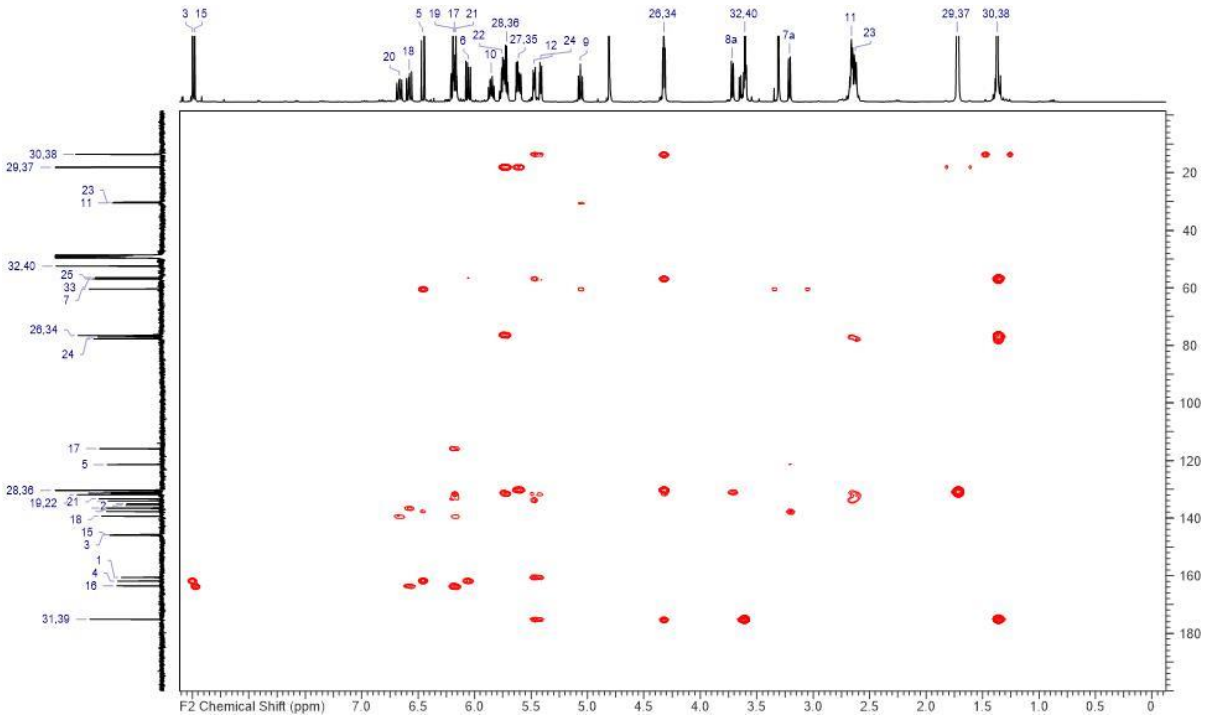


Figure S2. 29. *HMQC-NMR spectrum of 7,8-epoxi-disorazole Z (**6**) in in methanol-d_4.*

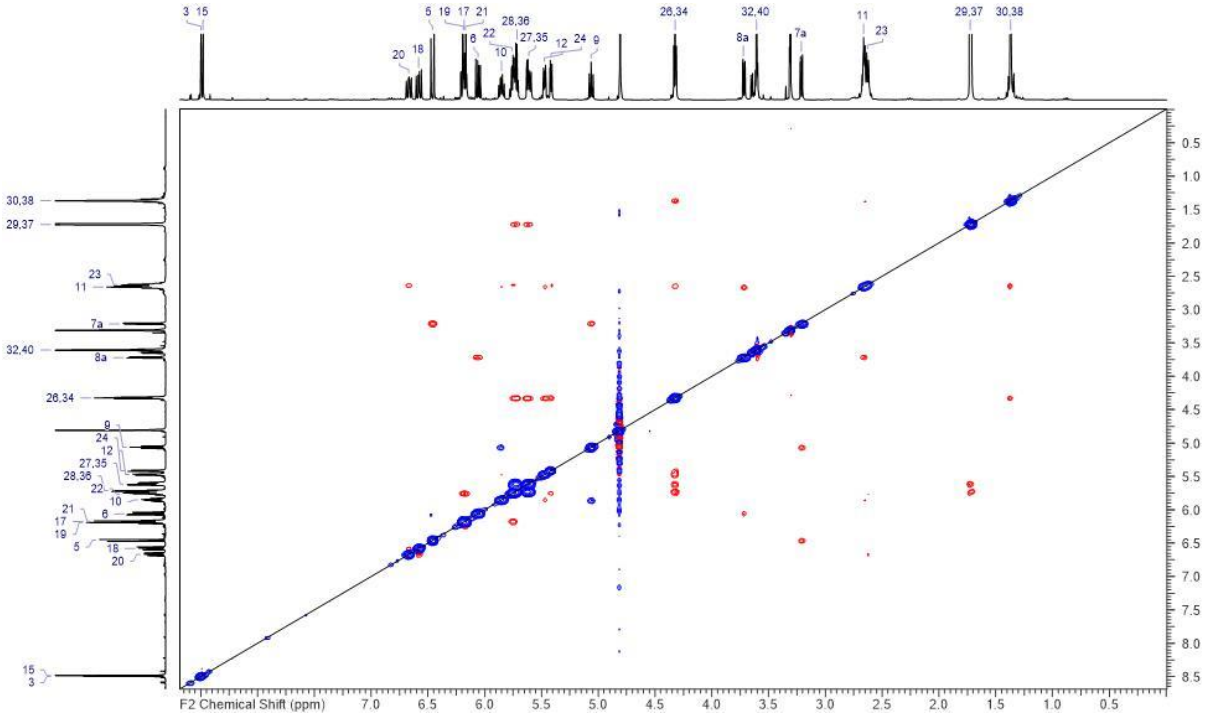


Figure S2. 30. *^{1}H,^{1}H-ROESY NMR spectrum of 7,8-epoxi-disorazole Z (**6**) in methanol-d_4.*

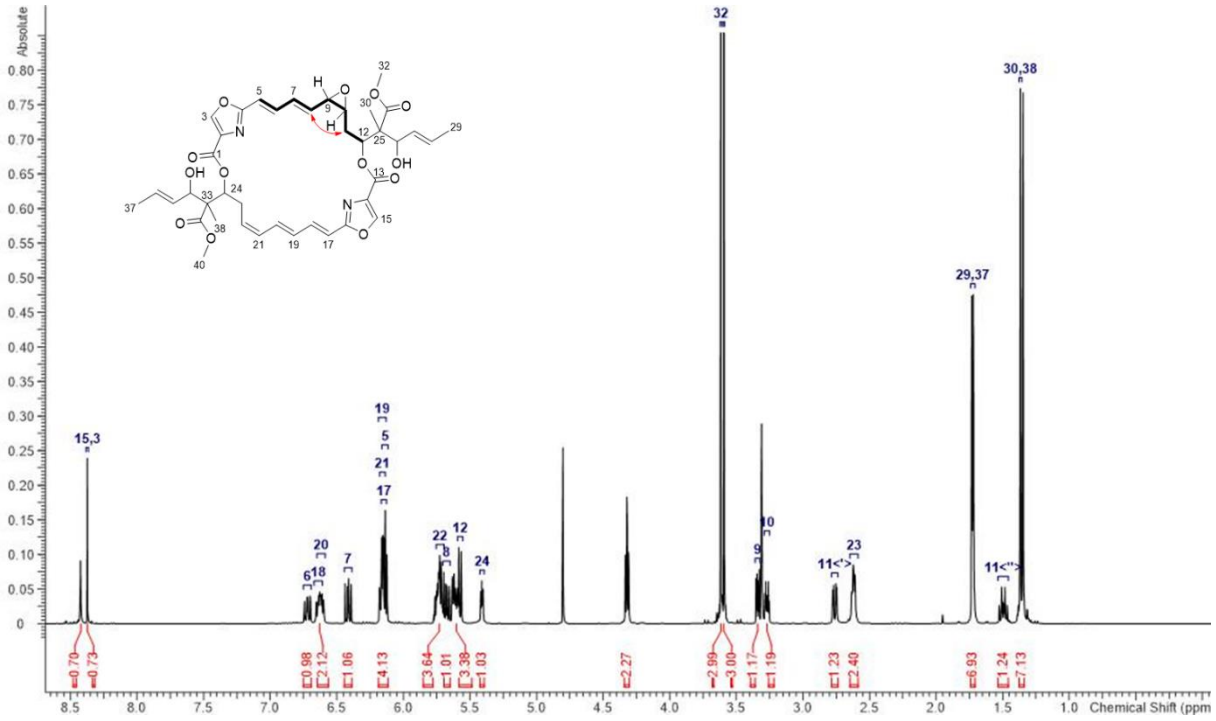


Figure S2. 31. *^{1}H NMR spectrum of 9,10-epoxi-disorazole Z (**7**) in methanol-d_4 (600 MHz).*

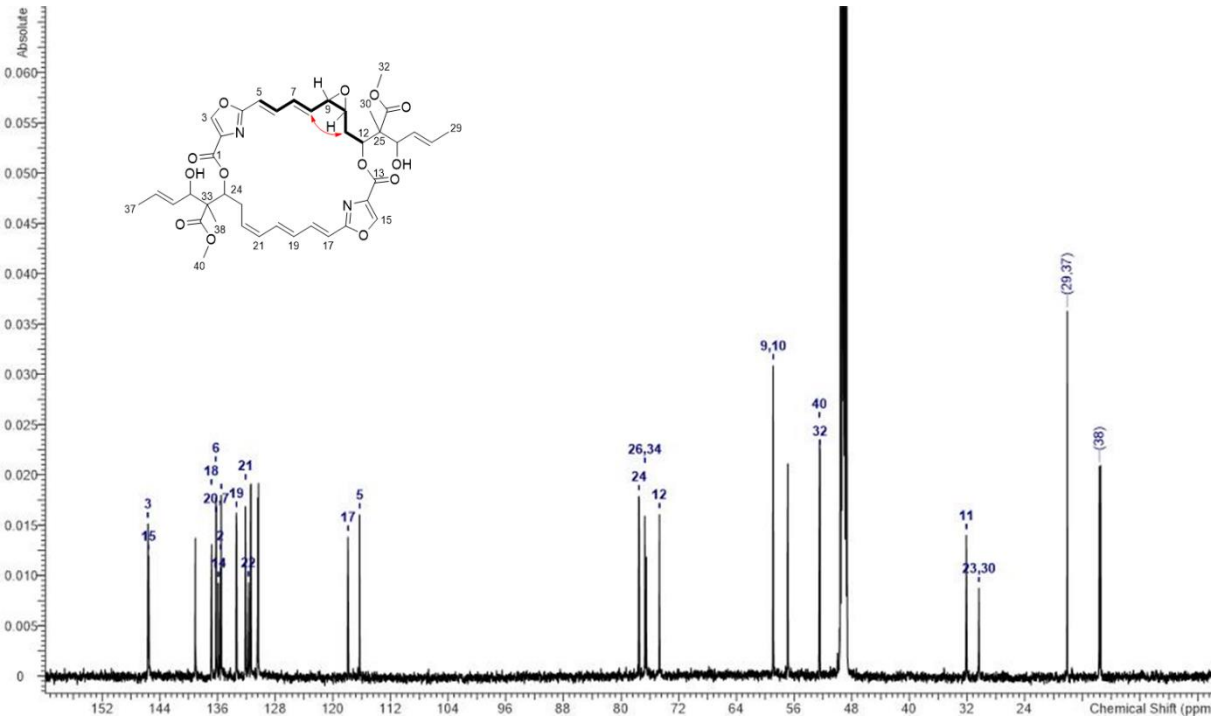


Figure S2. 32. *^{13}C NMR spectrum of 9,10-epoxi-disorazole Z (**7**) in in methanol-d_4 (150 MHz).*

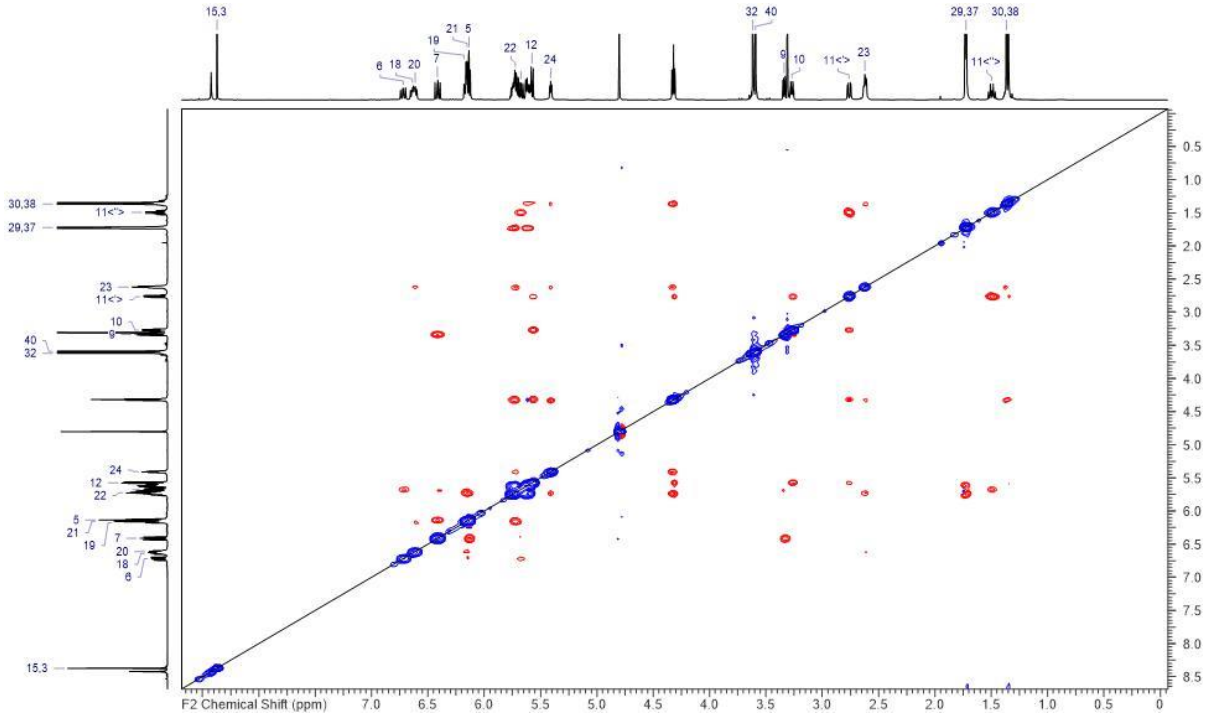


Figure S2. 33. *HMQC-NMR spectrum of 7,8-epoxi-disorazole Z (**6**) in in methanol-d₄.*

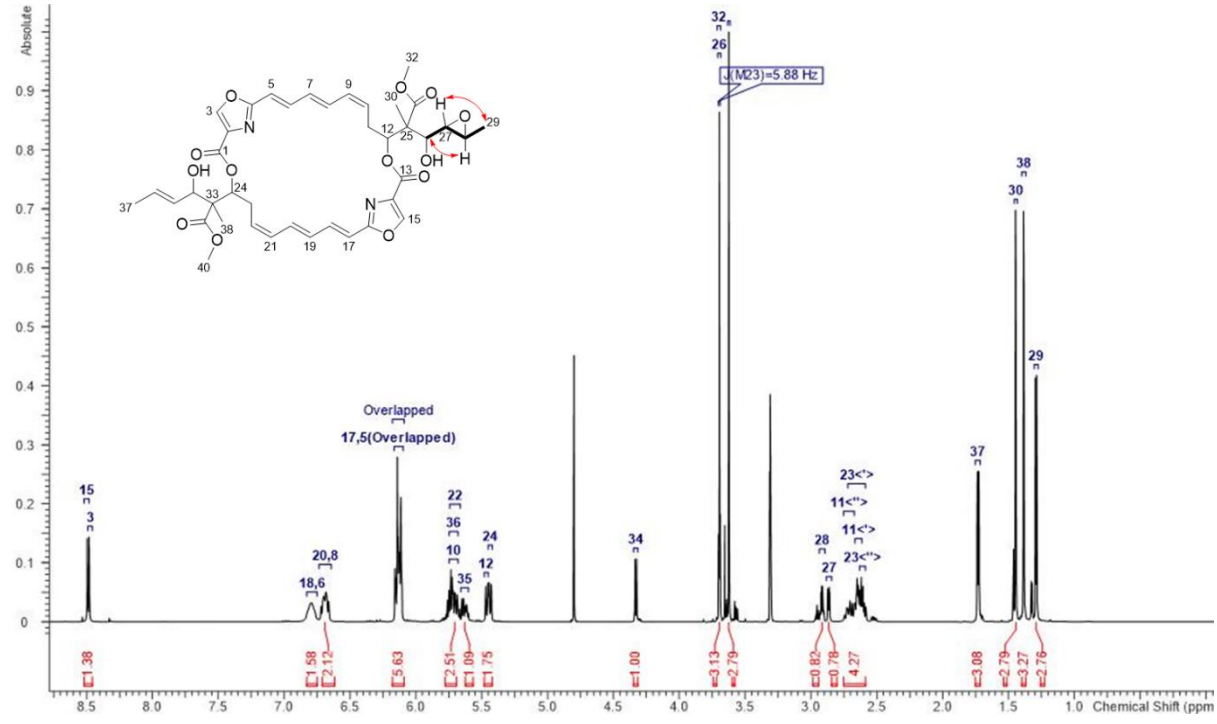


Figure S2. 34. *^{1}H NMR spectrum of 27,28-epoxi-disorazole Z (**8**) in methanol-d₄ (600 MHz).*

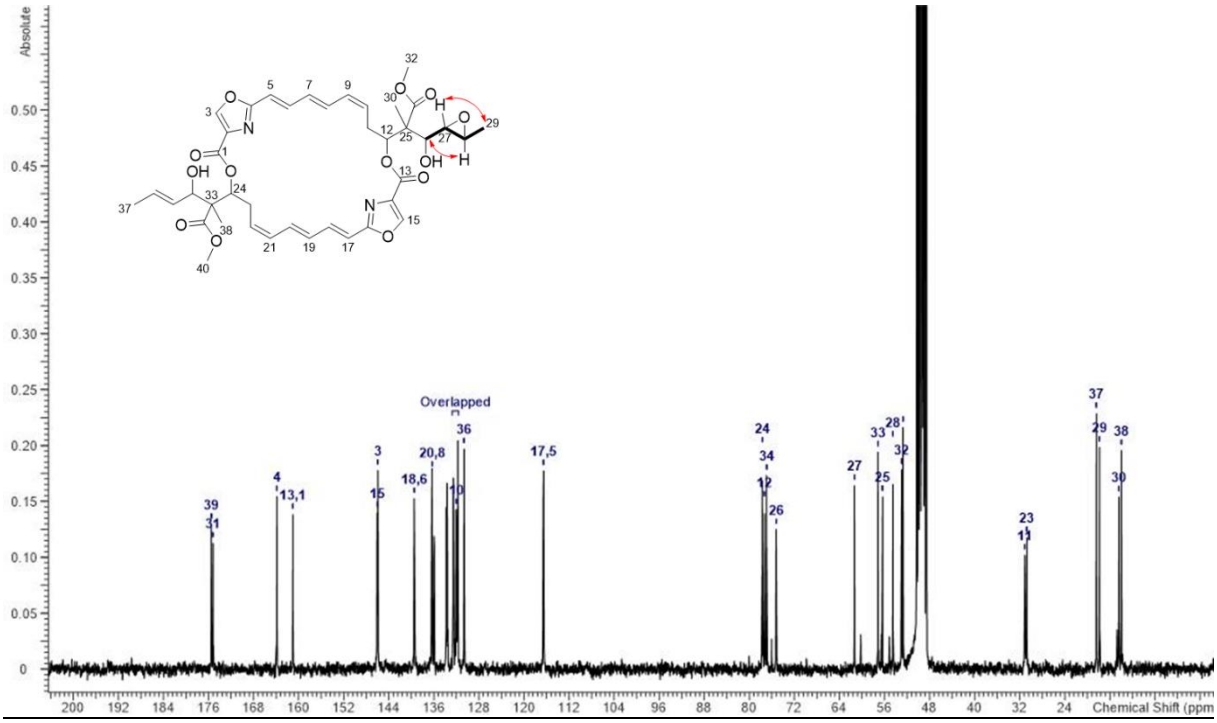


Figure S2. 35. *^{13}C NMR spectrum of 27,28-epoxi-disorazole Z (**8**) in in methanol-d_4 (150 MHz).*

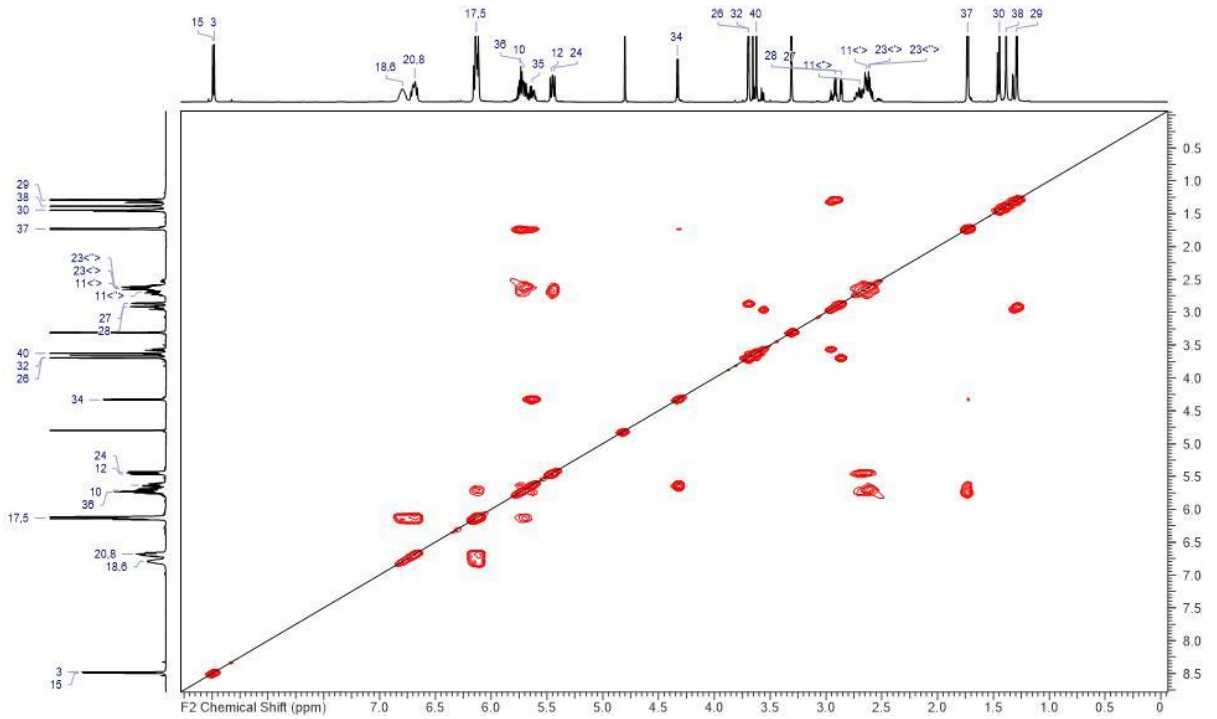


Figure S2. 36. *^{1}H,^{1}H-COSY NMR spectrum of 27,28-epoxi-disorazole Z (**8**) in in methanol-d_4.*

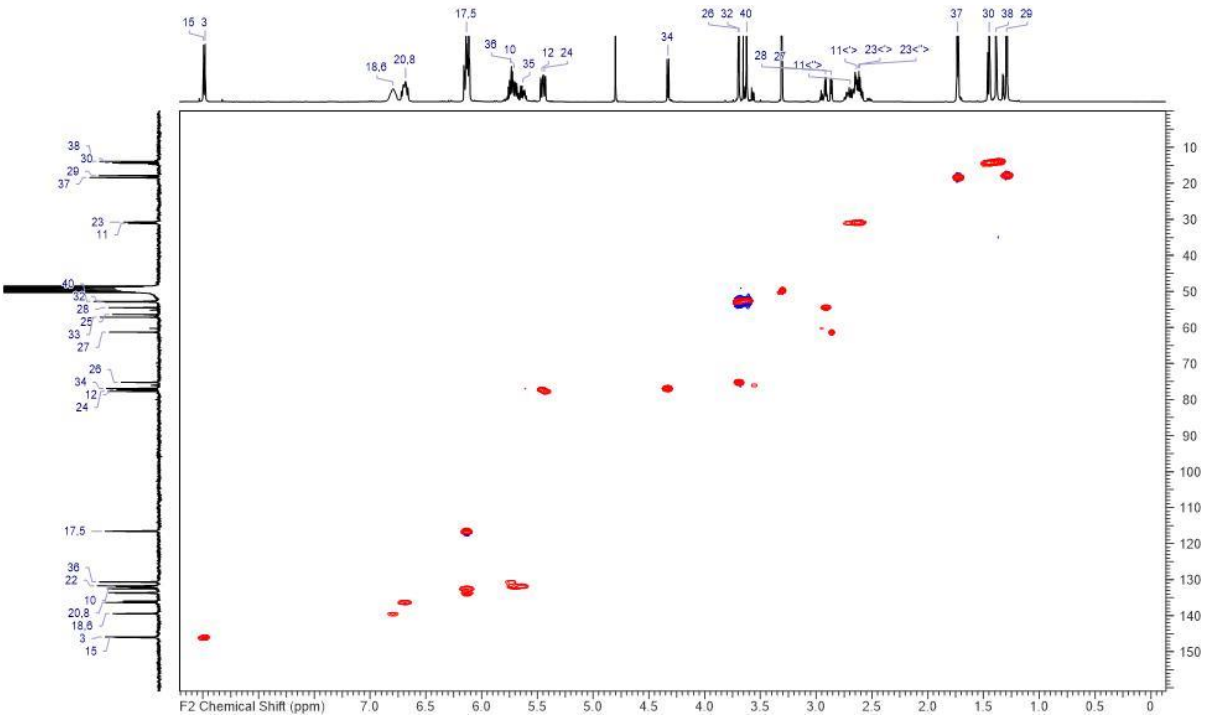


Figure S2. 37. *HMQC-NMR spectrum of 27,28-epoxi-disorazole Z (**8**) in in methanol-d_4.*

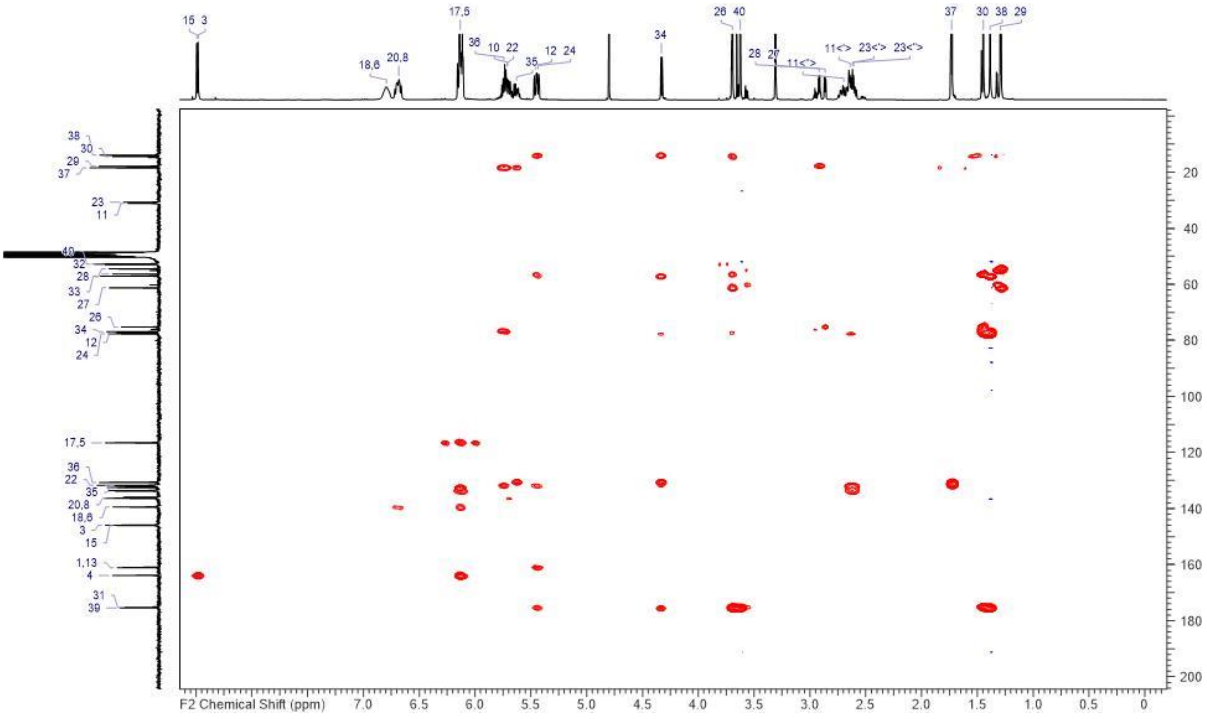


Figure S2. 38. *HMBC NMR spectrum of 27,28-epoxi-disorazole Z (**8**) in in methanol-d_4.*

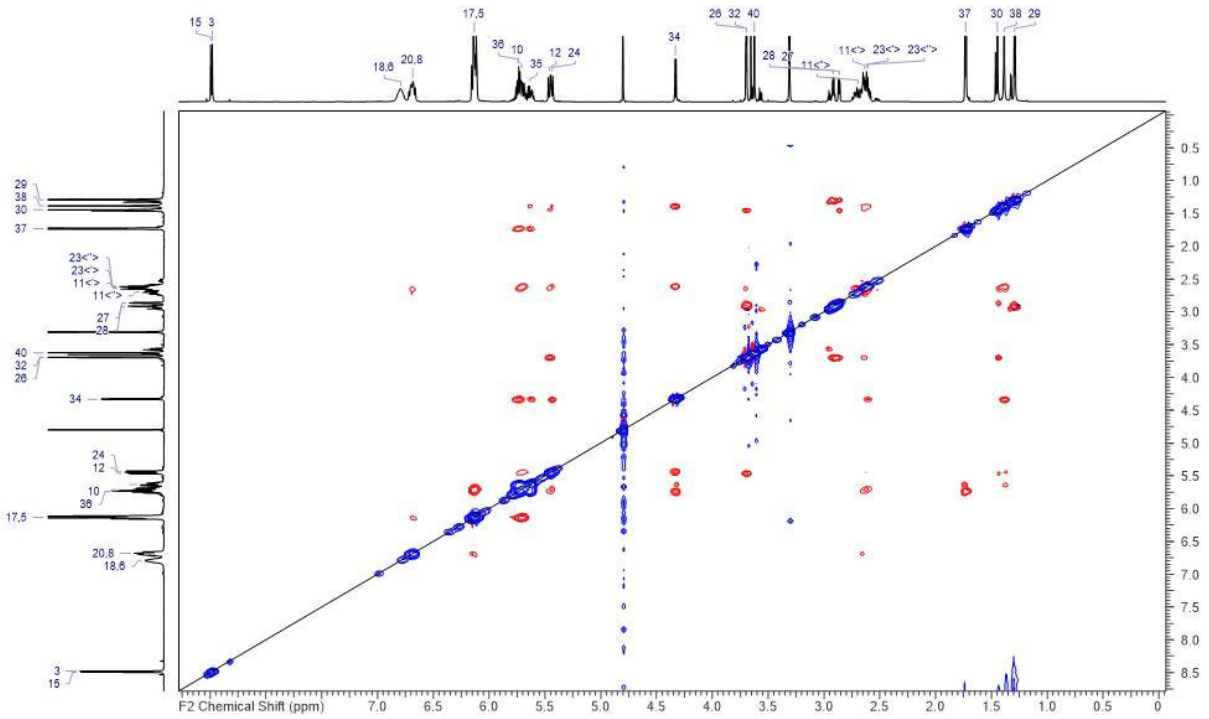


Figure S2. 39. *$^{1}H,^{1}H$-ROESY NMR spectrum of 27,28-epoxi-disorazole Z (**8**) in methanol-d_4.*

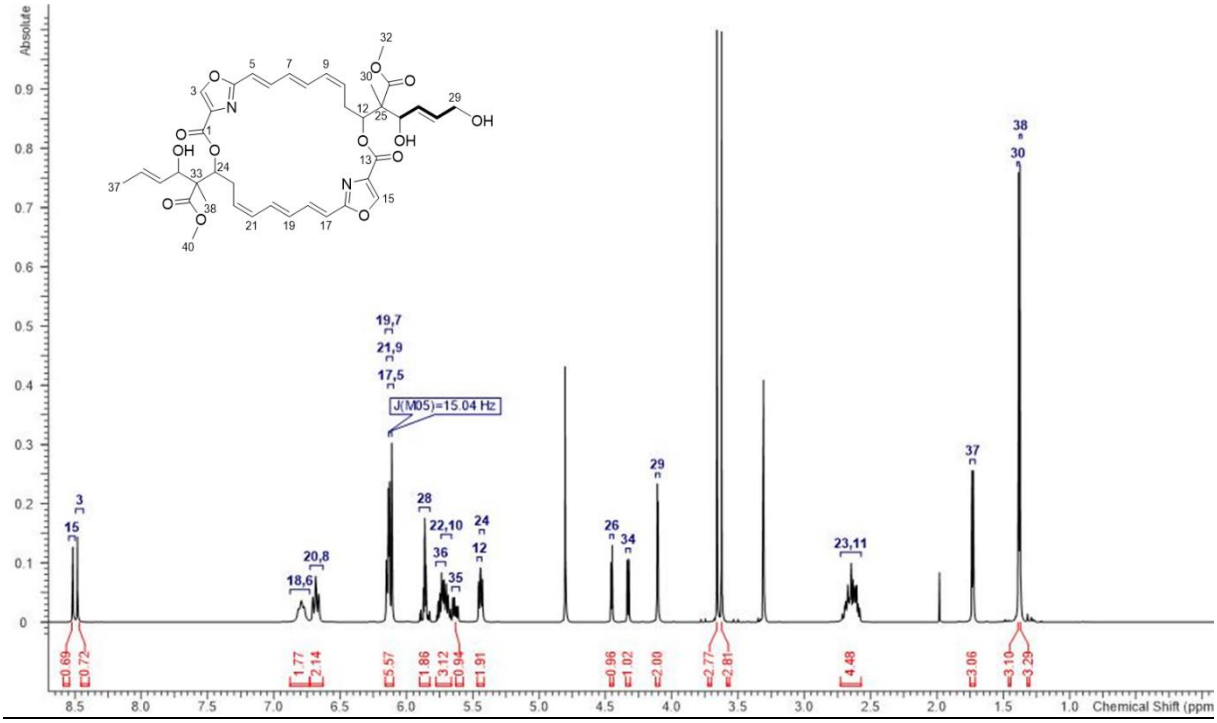


Figure S2. 40. *^{1}H NMR spectrum of 29-hydroxy-disorazole Z (**9**) in methanol-d_4 (600 MHz).*

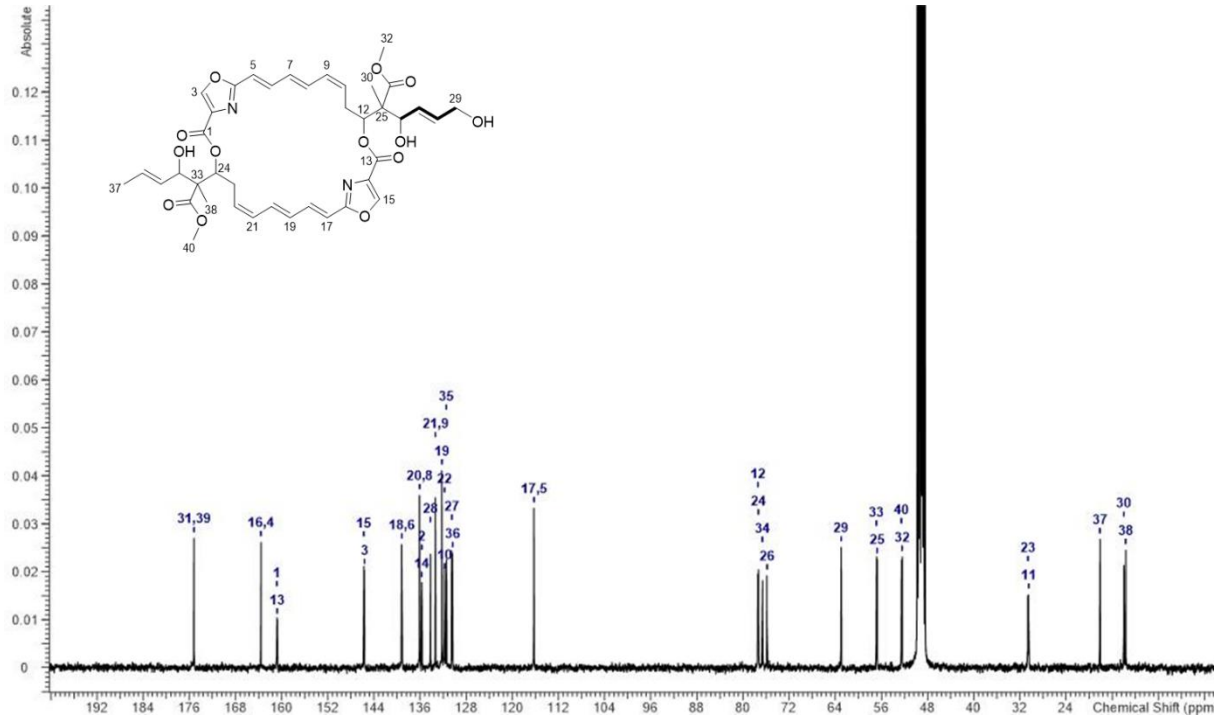


Figure S2. 41. *^{13}C NMR spectrum of 29-hydroxy-disorazole Z (**9**) in in methanol-d$_4$ (150 MHz).*

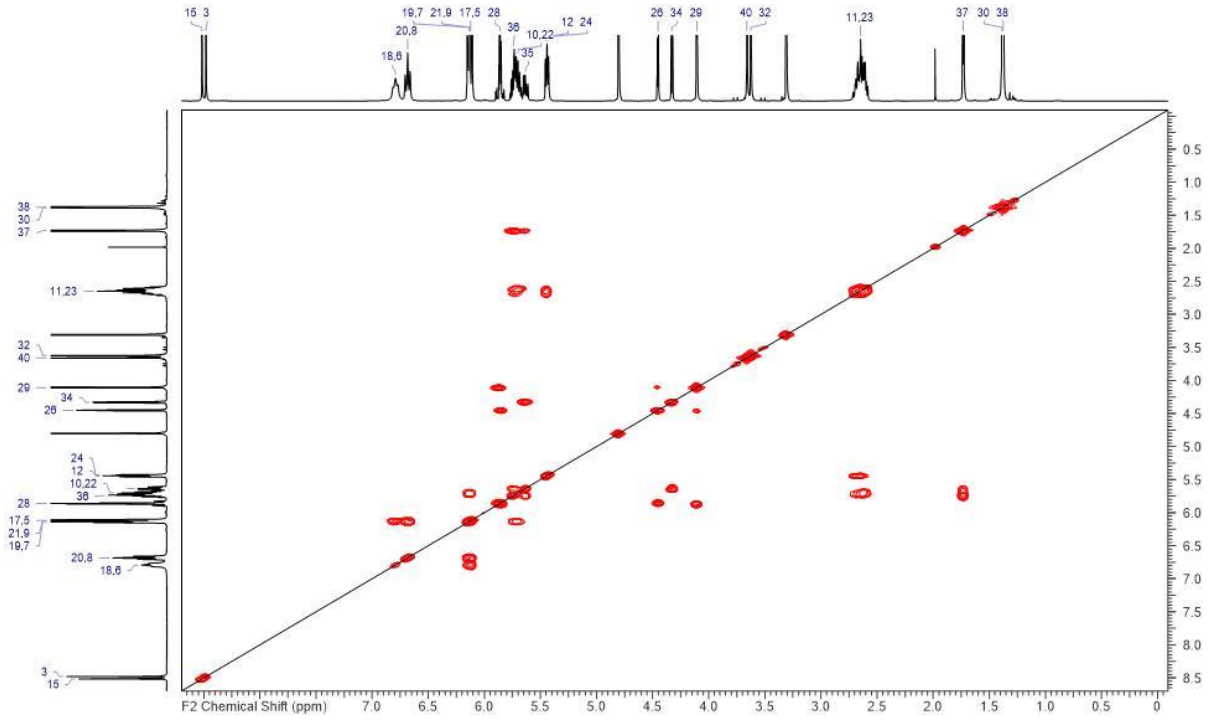


Figure S2. 42. *^{1}H,^{1}H-COSY NMR spectrum of 29-hydroxy-disorazole Z (**9**) in in methanol-d$_4$.*

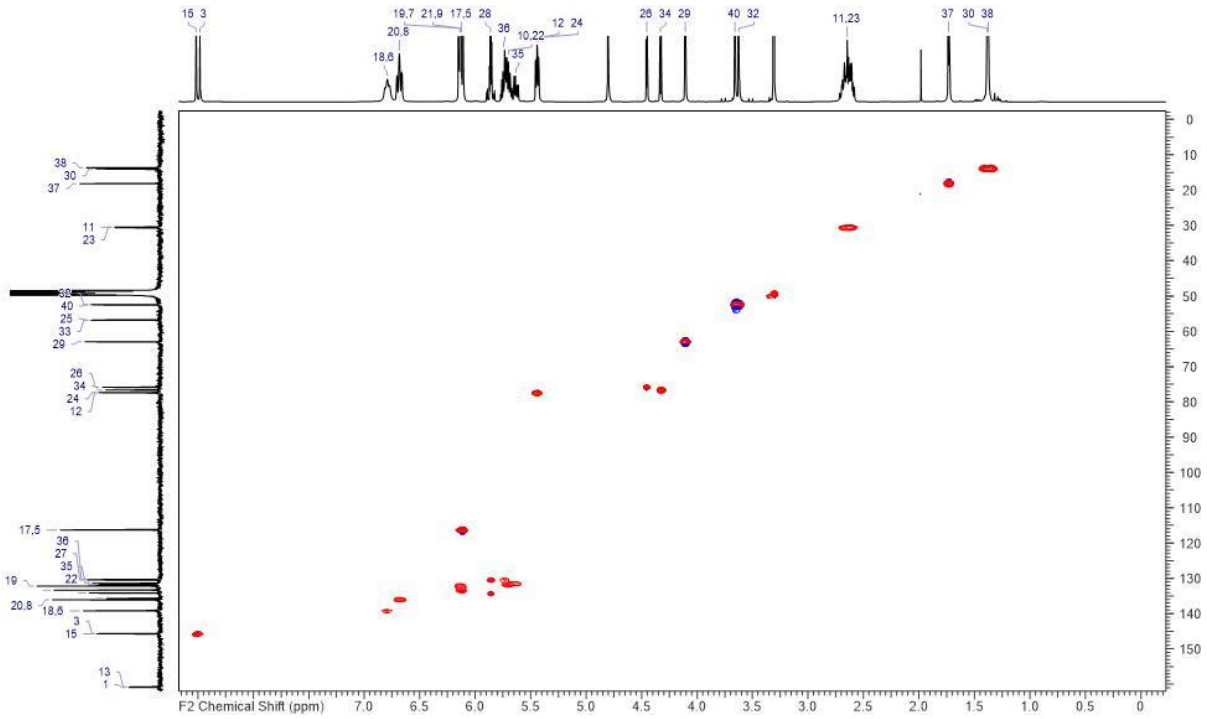


Figure S2. 43. *HMQC-NMR spectrum of 29-hydroxy-disorazole Z (**9**) in in methanol-d_4.*

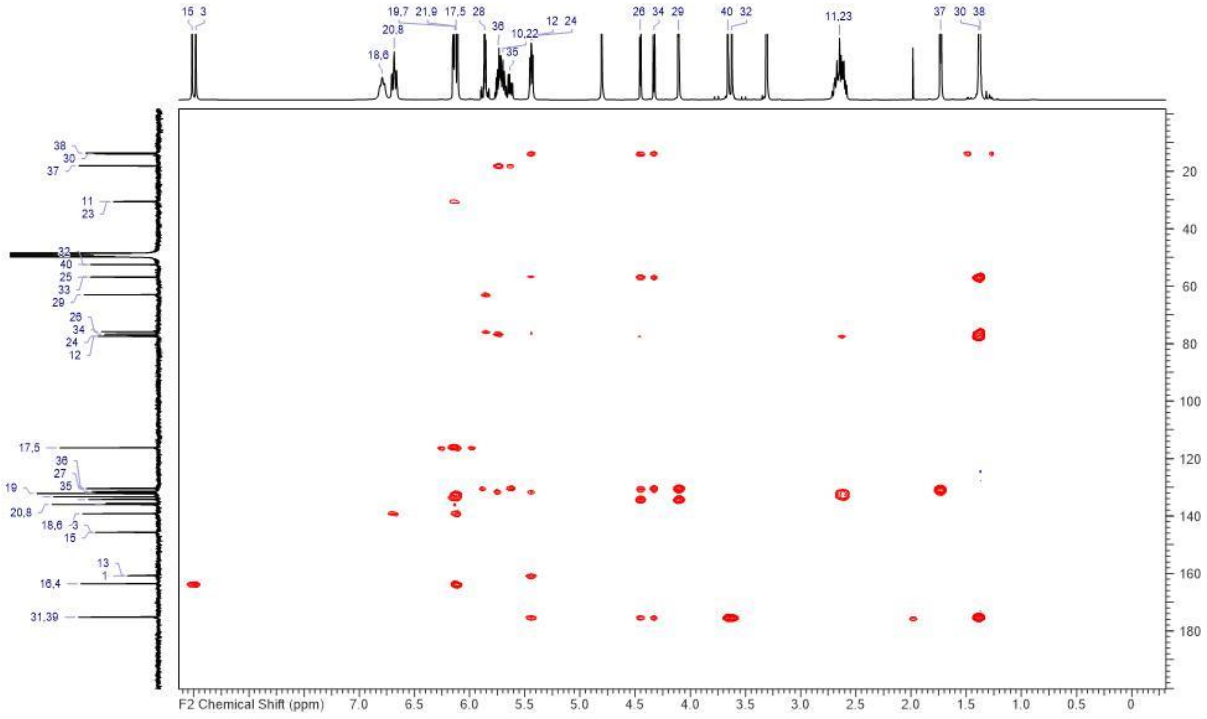


Figure S2. 44. *HMBC NMR spectrum of 29-hydroxy-disorazole Z (**9**) in in methanol-d_4.*

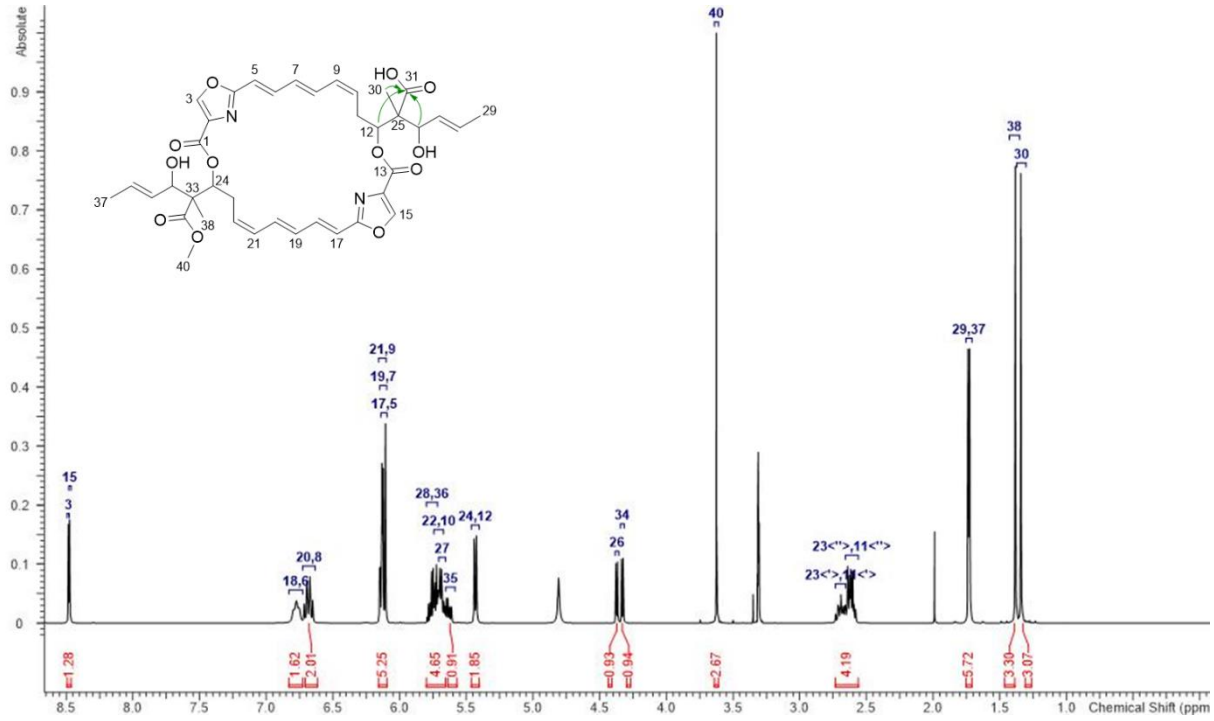


Figure S2. 45. *^{1}H NMR spectrum of 31-O-desmethyl-disorazole Z (**10**) in methanol-d_4 (600 MHz).*

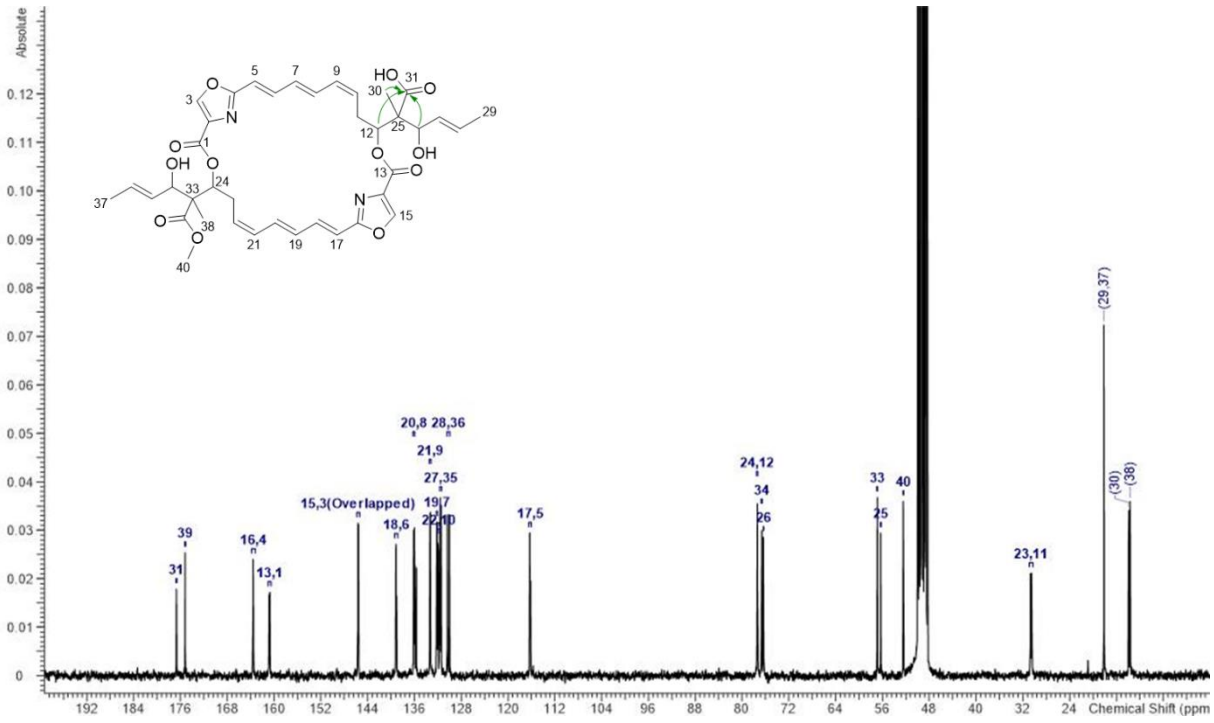


Figure S2. 46. *^{13}C NMR spectrum of 31-O-desmethyl-disorazole Z (**10**) in methanol-d_4 (150 MHz).*

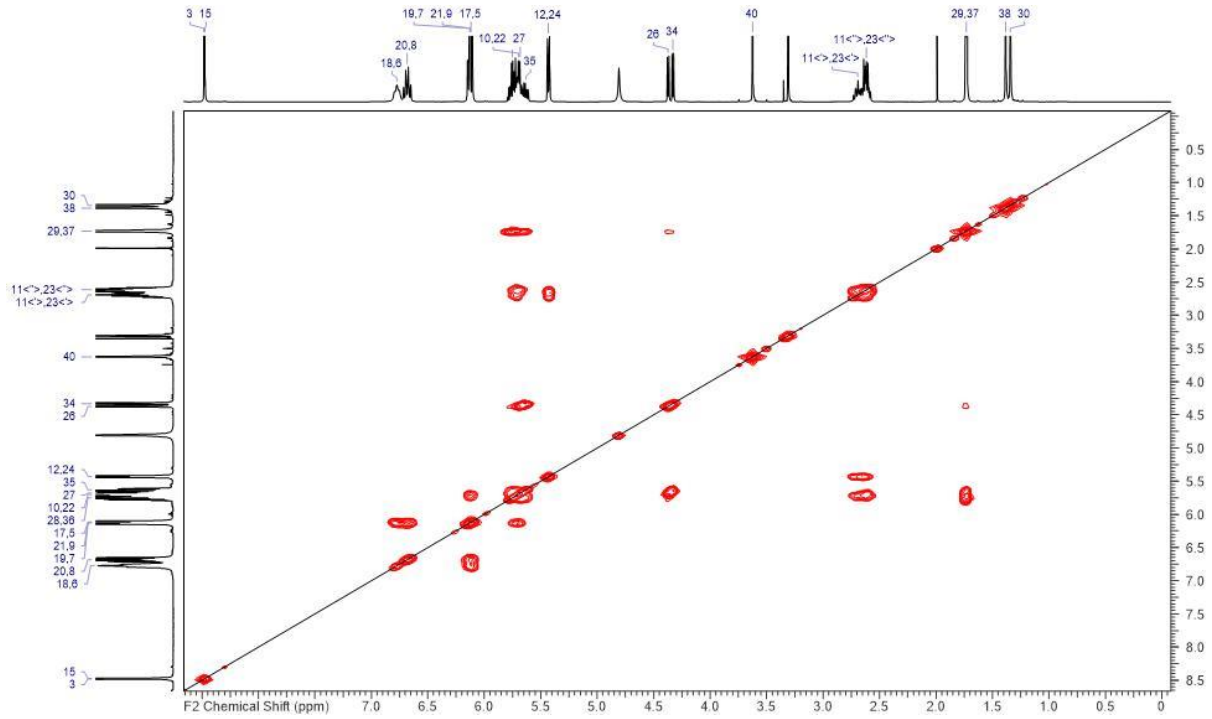


Figure S2. 47. *$^{1}H,^{1}H$-COSY NMR spectrum of 31-O-desmethyl-disorazole Z (**10**) in in methanol-d_4.*

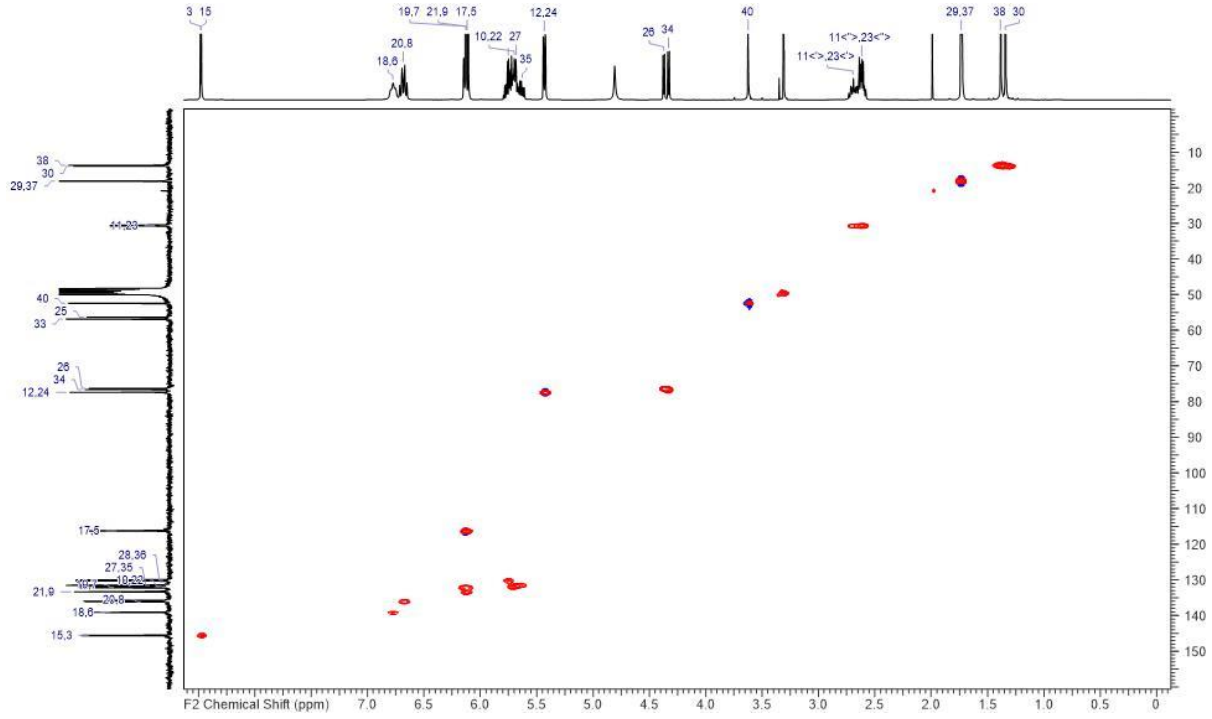


Figure S2. 48. *HMQC-NMR spectrum of 31-O-desmethyl-disorazole Z (**10**) in in methanol-d_4.*

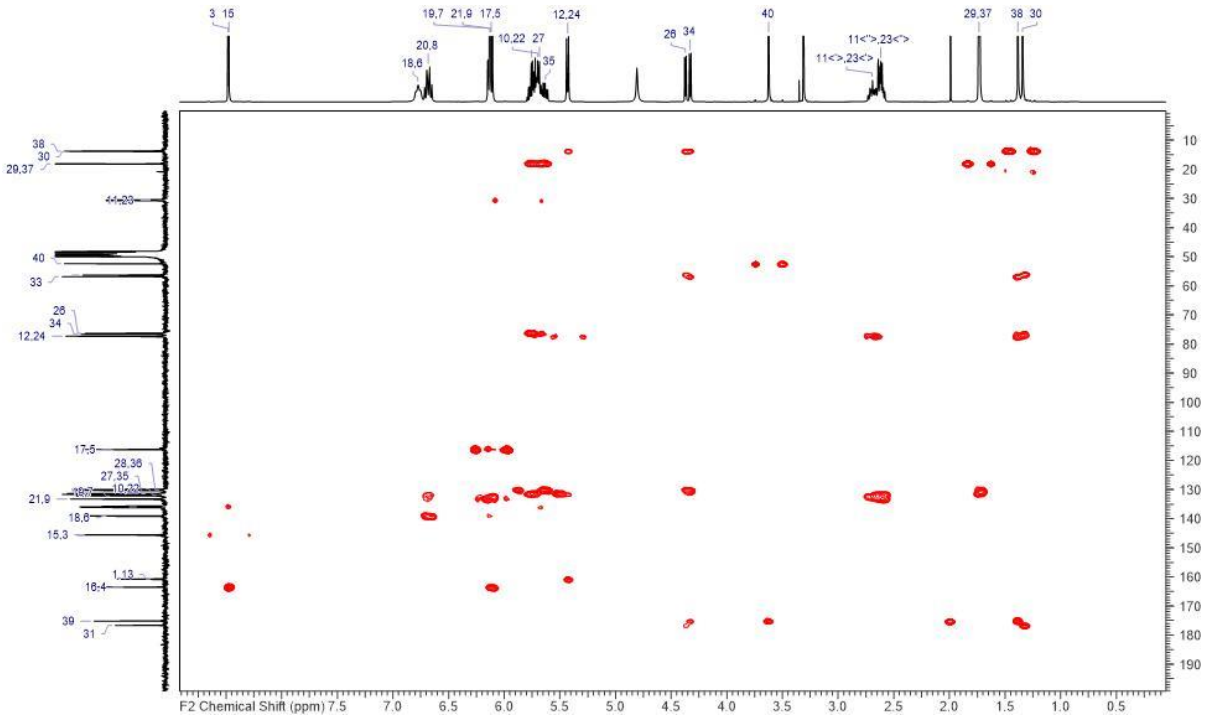


Figure S2. 49. *HMBC NMR spectrum of 31-O-desmethyl-disorazole Z (**10**) in in methanol-d₄.*

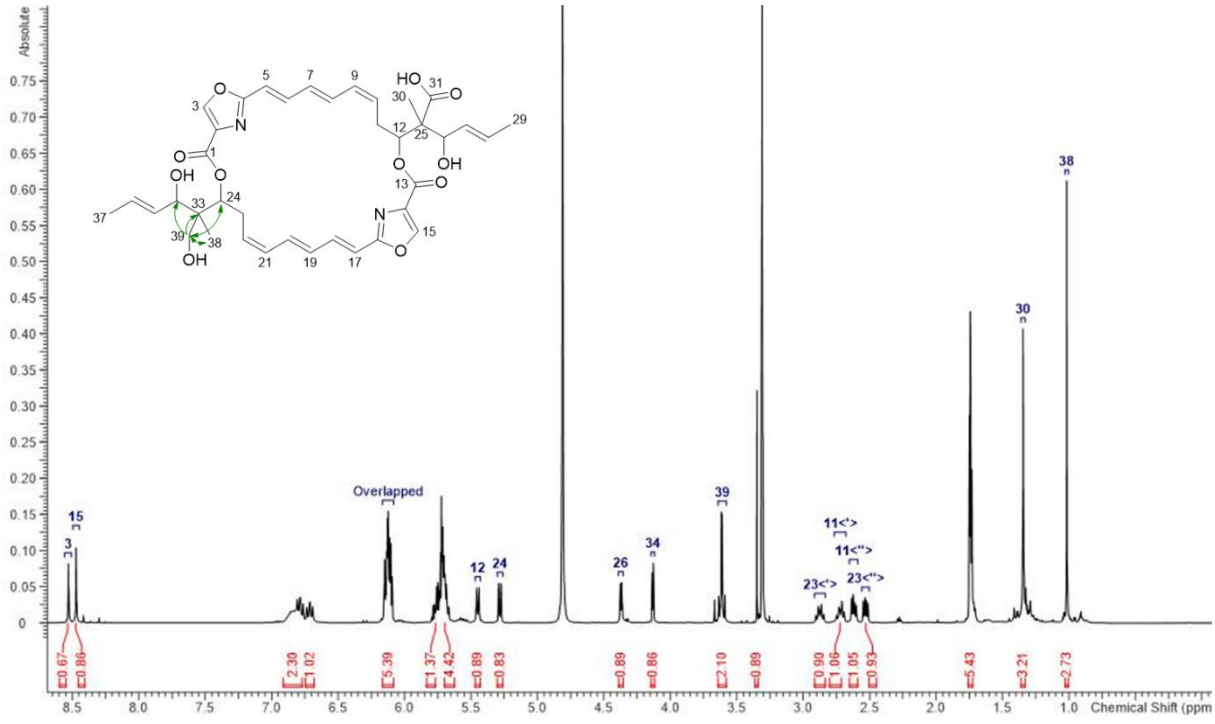


Figure S2. 50. *^{1}H NMR spectrum of 31-O-desmethyl-39-hydroxy-disorazole Z (**11**) in methanol-d₄ (600 MHz).*

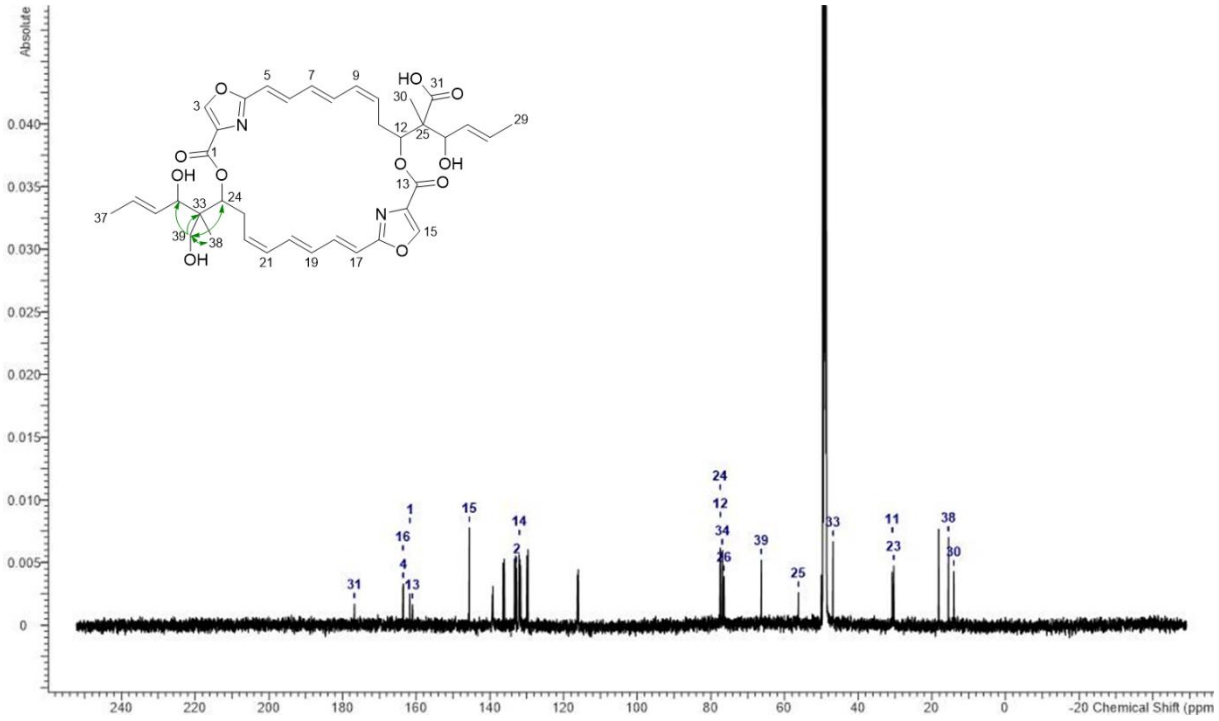


Figure S2. 51. *^{13}C NMR spectrum of 31-O-desmethyl-39-hydroxy-disorazole Z (**11**) in methanol-d_4 (150 MHz).*

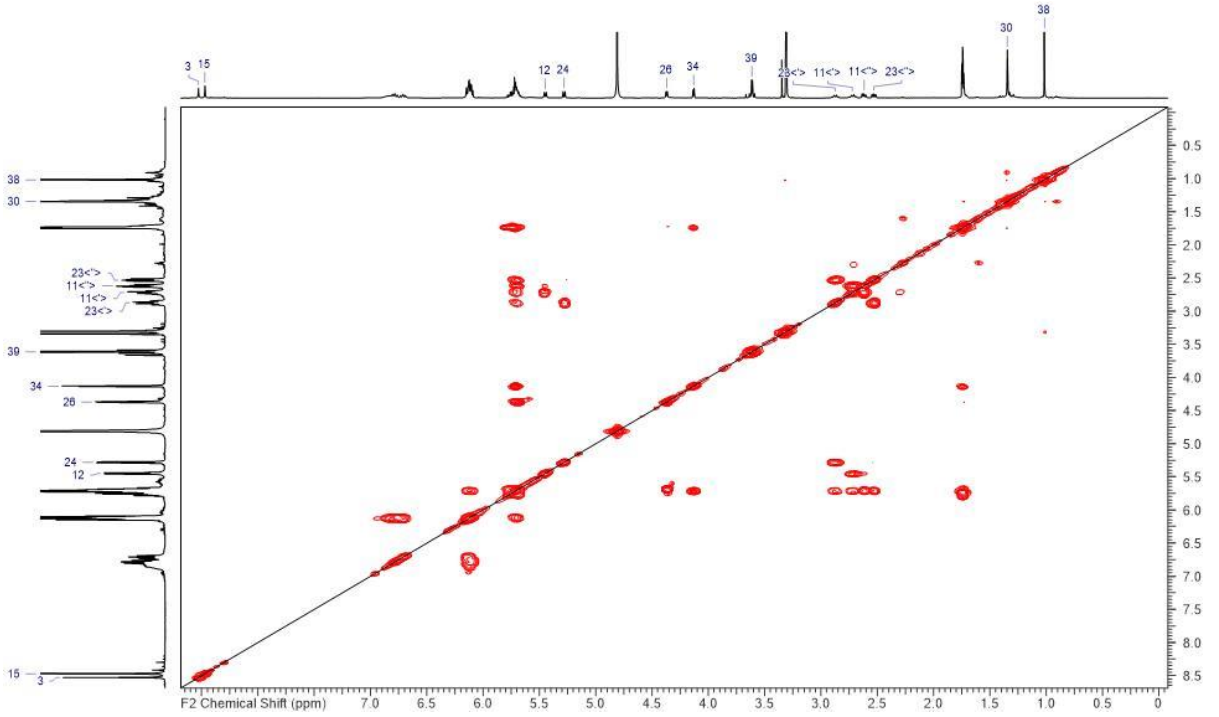


Figure S2. 52. *^{1}H,^{1}H-COSY NMR spectrum of 31-O-desmethyl-39-hydroxy-disorazole Z (**11**) in in methanol-d_4.*

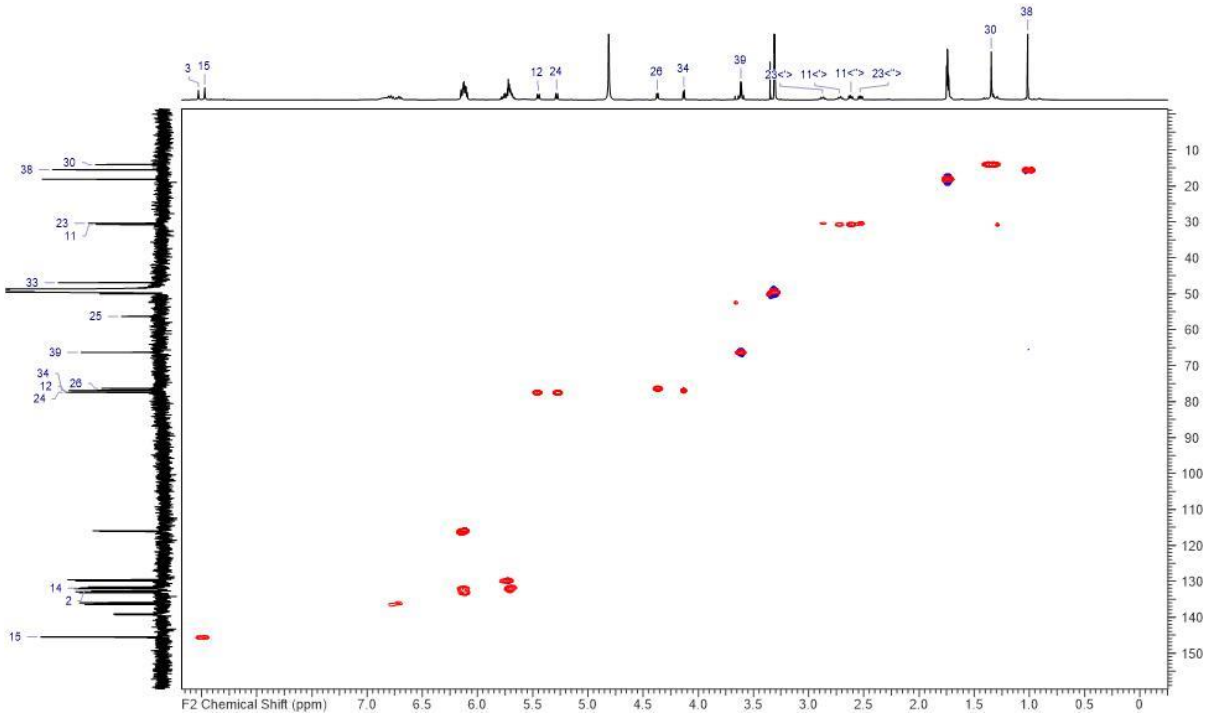


Figure S2. 53. *HMQC-NMR spectrum of 31-O-desmethyl-39-hydroxy-disorazole Z (**11**) in in methanol-d_4.*

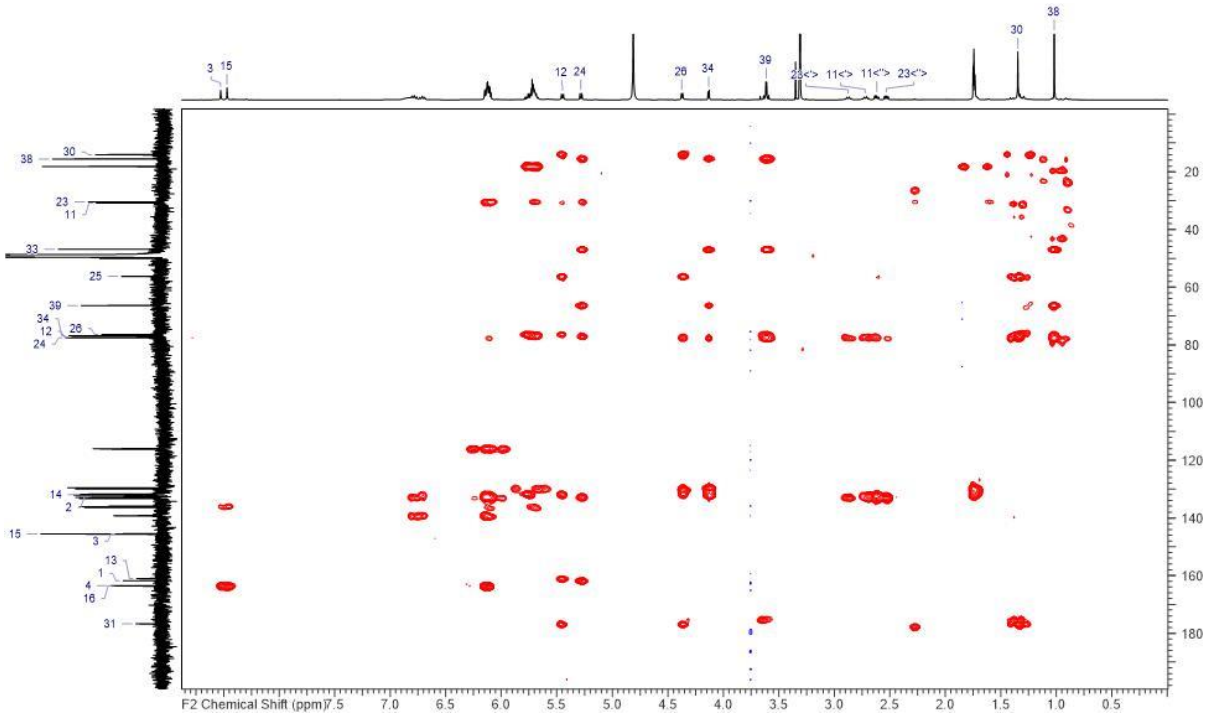


Figure S2. 54. *HMBC NMR spectrum of 31-O-desmethyl-39-hydroxy-disorazole Z (**11**) in in methanol-d_4.*

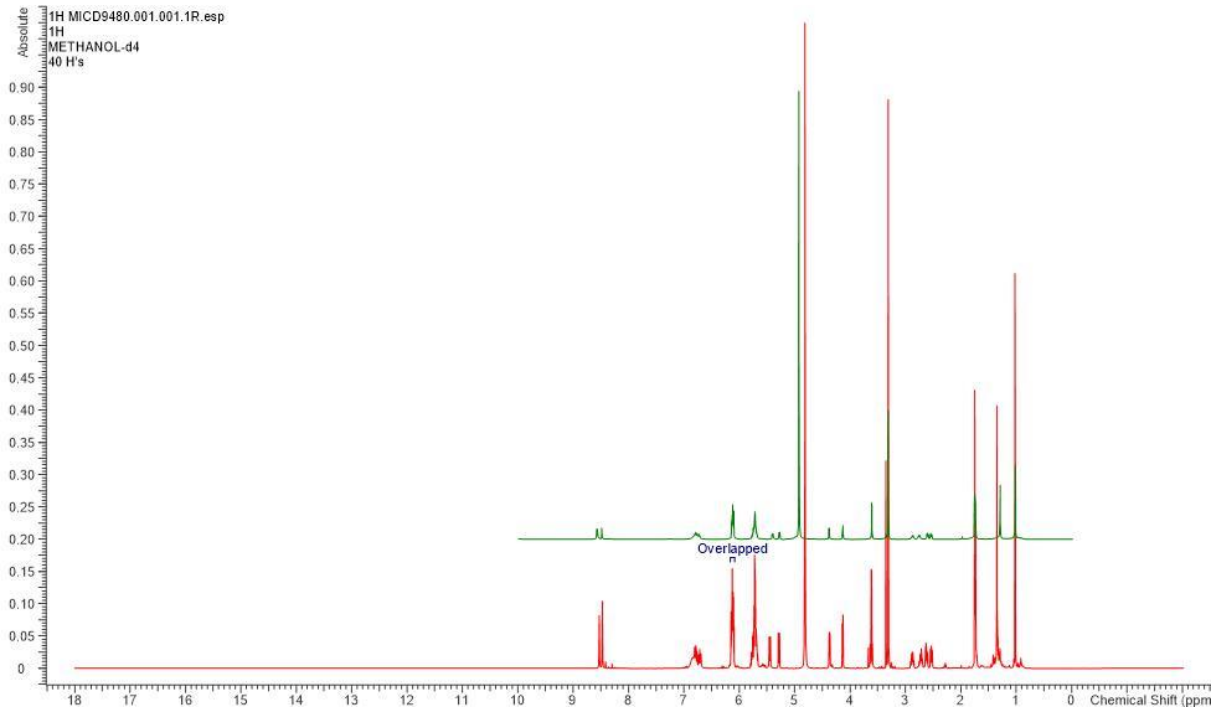


Figure S2. 55. *Comparison of [1]H NMR spectra of* ***11*** *methanol-d₄ isolated from* S. cellulosum *and* M. xanthus*. Bottom/red: isolated from* S. cellulosum *So ce1875; Top/green: isolated from* M. xanthus *DK1622 :: km-int-Ptet-dis427-gent-delF.*

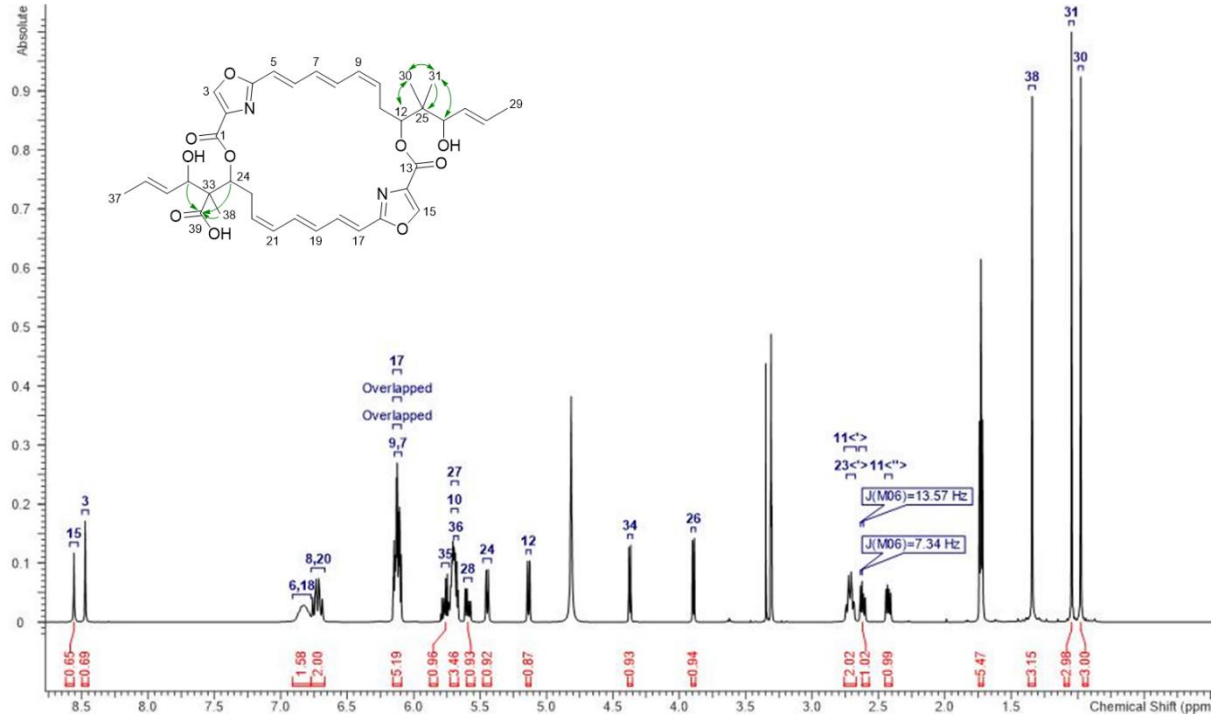


Figure S2. 56. *[1]H NMR spectrum of O-desmethyl-dimethyl-disorazole Z (**12**) in methanol-d₄ (600 MHz).*

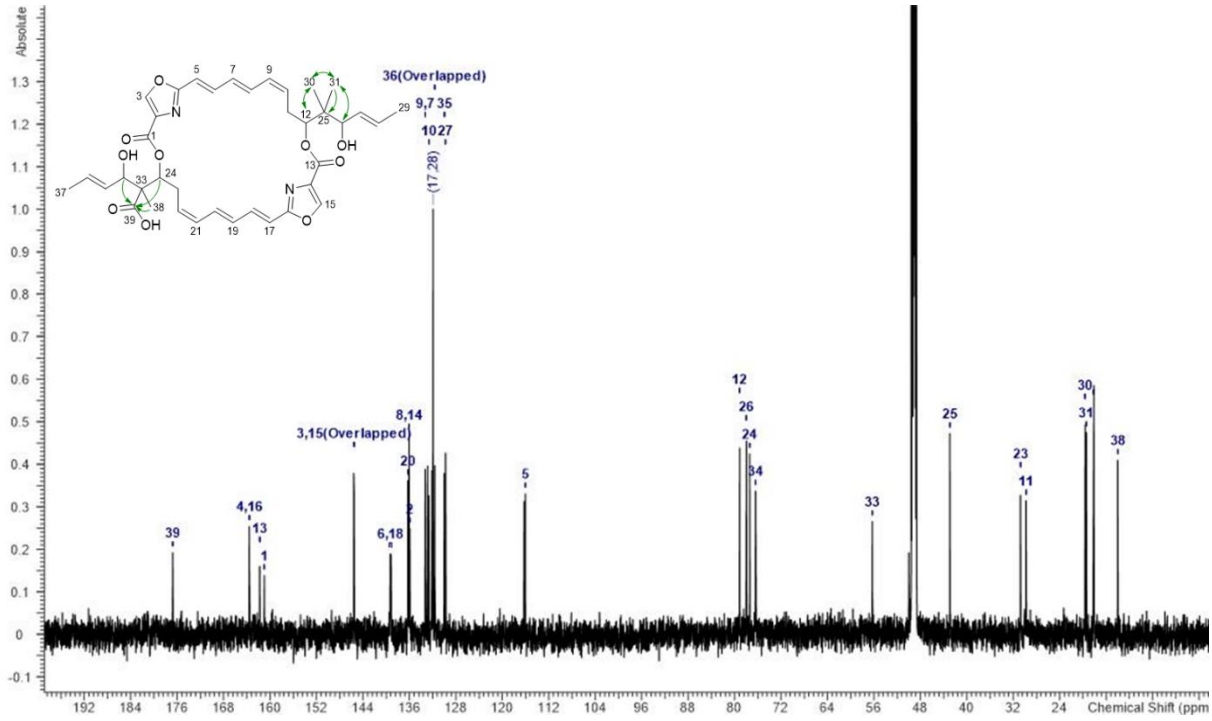


Figure S2. 57. *^{13}C NMR spectrum of O-desmethyl-dimethyl-disorazole Z (**12**) in methanol-d_4 (150 MHz).*

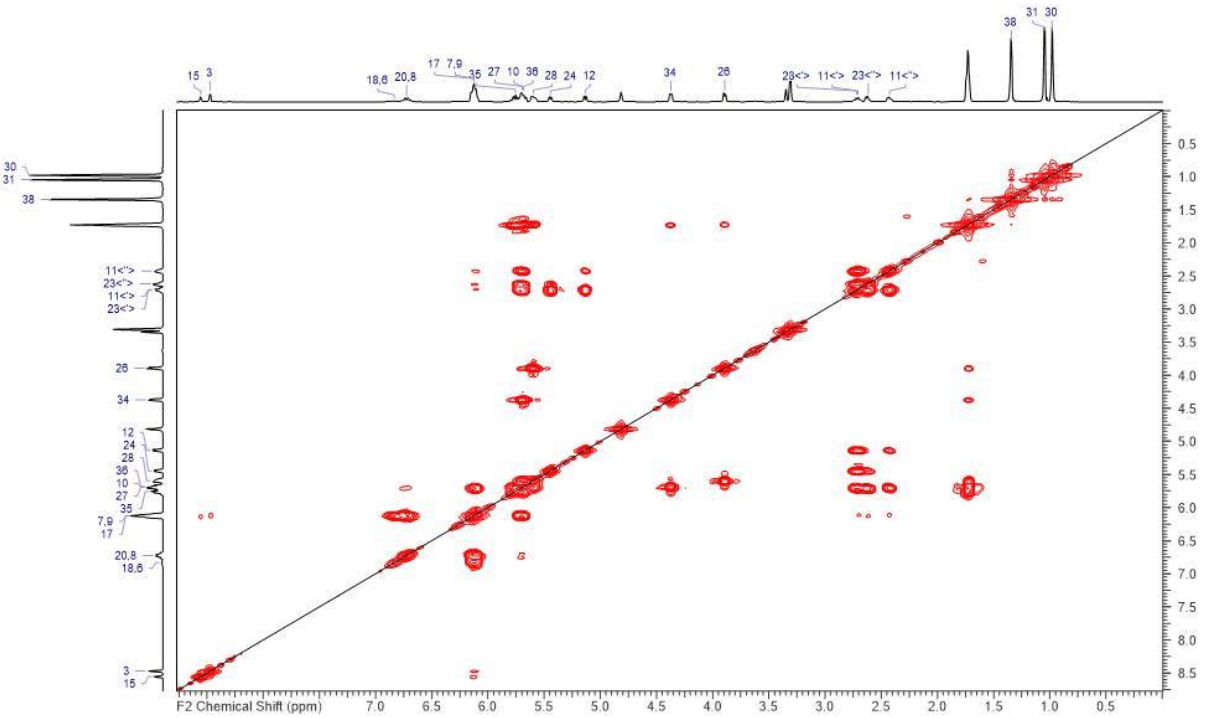


Figure S2. 58. *^{1}H,^{1}H-COSY NMR spectrum of O-desmethyl-dimethyl-disorazole Z (**12**) in in methanol-d_4.*

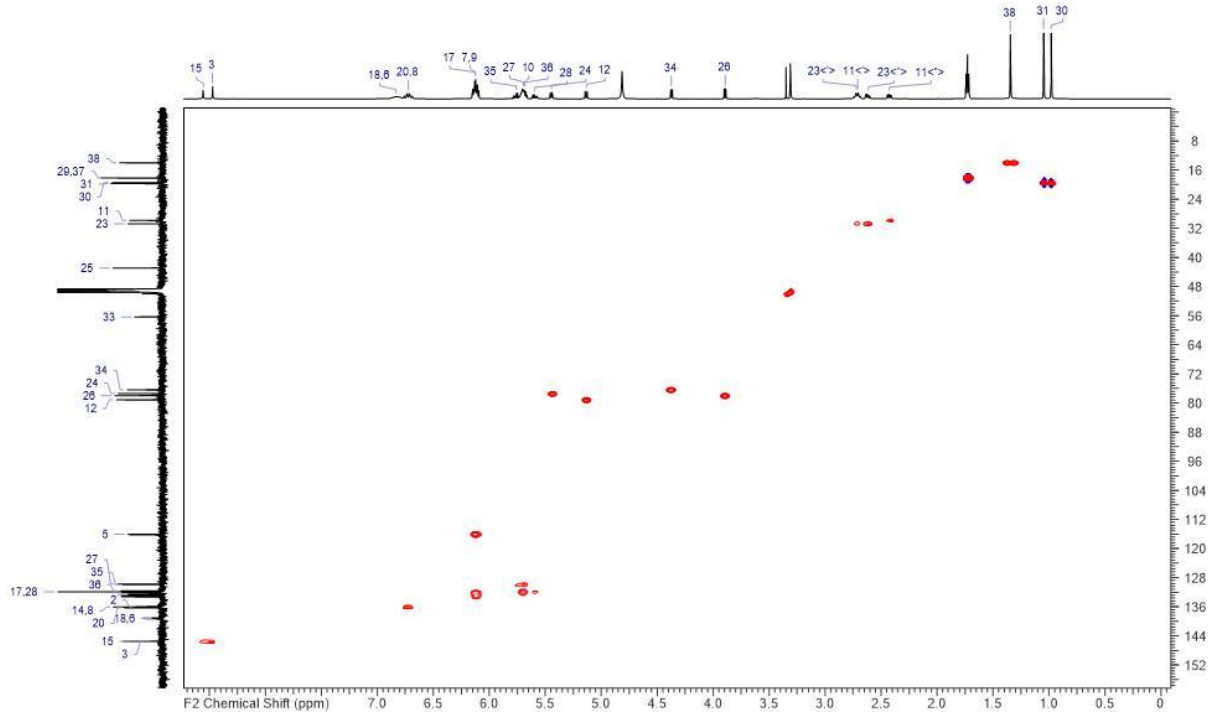


Figure S2. 59. *HMQC-NMR spectrum of O-desmethyl-dimethyl-disorazole Z (**12**) in in methanol-d₄.*

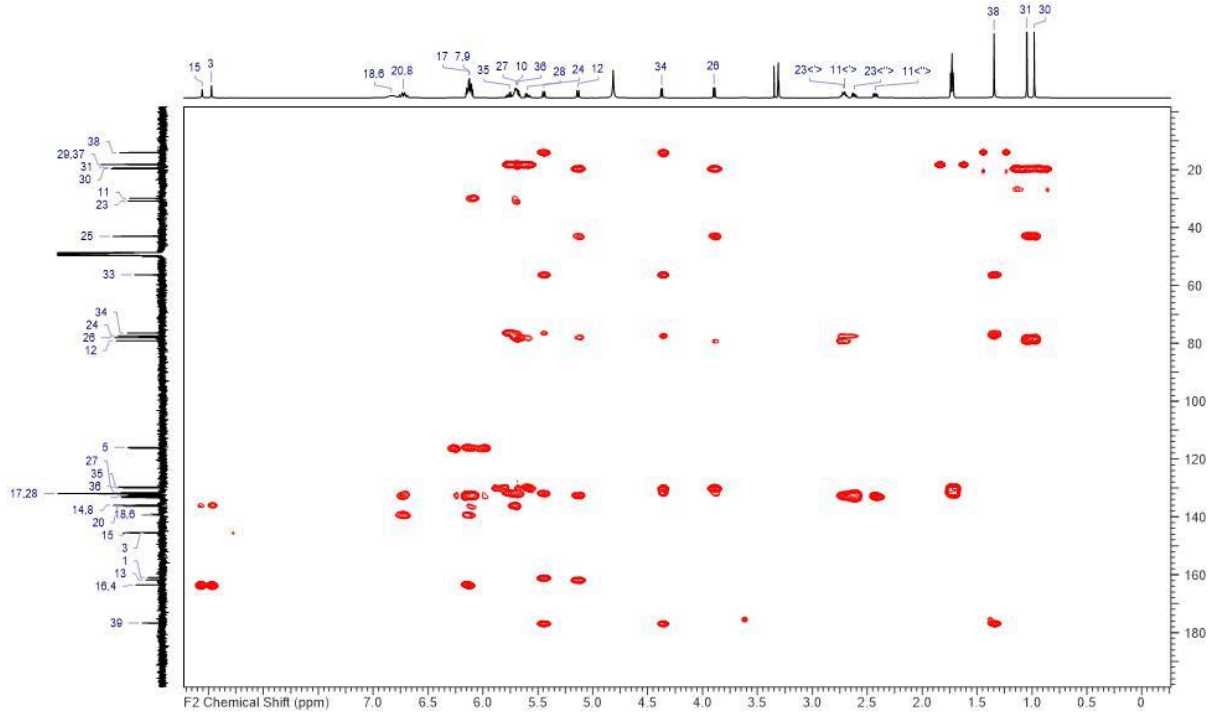


Figure S2. 60. *HMBC NMR spectrum of O-desmethyl-dimethyl-disorazole Z (**12**) in in methanol-d₄.*

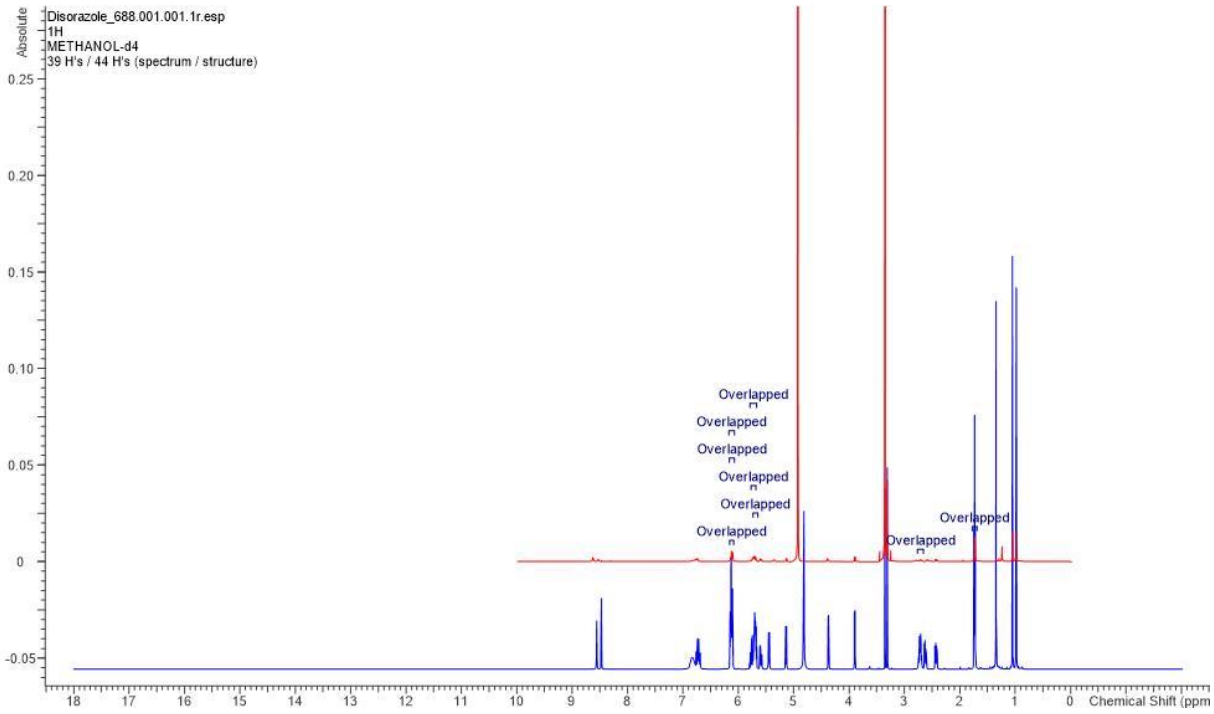


Figure S2. 61. *Comparison of ^{1}H NMR spectra of* ***12*** *in methanol-d_4 isolated from* S. cellulosum *and* M. xanthus*. Bottom/blue: isolated from S. cellulosum So ce1875; Top/red: isolated from M. xanthus DK1622 :: km-int-Ptet-dis427-gent-delF.*

S 2.4 References

1. Dale, J. A., Dull, D. L. & Mosher, H. S. α-Methoxy-α-trifluoromethylphenylacetic acid, a versatile reagent for the determination of enantiomeric composition of alcohols and amines. *J Org Chem* **34,** 2543–2549; 10.1021/jo01261a013 (1969).
2. Lemière, G. L., Willaert, J. J., Dommisse, R. A., Lepoivre, J. A. & Alderweireldt, F. C. Determination of the absolute configuration and enantiomeric purity of alcohols from the 13 C-NMR spectra of the corresponding MTPA esters. *Chirality* **2,** 175–184; 10.1002/chir.530020309 (1990).
3. Pehk, T. *et al.* Determination of the absolute configuration of chiral secondary alcohols; new advances using 13C- and 2D-NMR spectroscopy. *Tetrahedron, asymmetry* **4,** 1527–1532; 10.1016/S0957-4166(00)80354-7 (1993).
4. Hoye, T. R., Jeffrey, C. S. & Shao, F. Mosher ester analysis for the determination of absolute configuration of stereogenic (chiral) carbinol carbons. *Nat. Protoc.* **2,** 2451–2458; 10.1038/nprot.2007.354 (2007).
5. Jansen, R., Irschik, H., Reichenbach, H., Wray, V. & Höfle, G. Disorazoles, Highly Cytotoxic Metabolites from the Sorangicin-Producing BacteriumSorangium Cellulosum, Strain So ce12. *Liebigs Ann Chem* **1994,** 759–773; 10.1002/jlac.199419940802 (1994).
6. Robbins, T., Kapilivsky, J., Cane, D. E. & Khosla, C. Roles of Conserved Active Site Residues in the Ketosynthase Domain of an Assembly Line Polyketide Synthase. *Biochemistry* **55,** 4476–4484; 10.1021/acs.biochem.6b00639 (2016).
7. Alekseyev, V. Y., Liu, C. W., Cane, D. E., Puglisi, J. D. & Khosla, C. Solution structure and proposed domain domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Sci.* **16,** 2093–2107; 10.1110/ps.073011407 (2007).

3. Ajudazols

Expanding the Ajudazol Cytotoxin Scaffold: Insights from Genome Mining, Biosynthetic Investigations, and Novel Derivatives

Previously published in:

Hu Zeng†, Joy Birkelbach†, Judith Hoffmann, Alexander Popoff, Carsten Volz, and Rolf Müller

Journal of Natural Products **2022** 85 (11), 2610-2619

DOI: 10.1021/acs.jnatprod.2c00637

Affiliation

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

Contributions and Acknowledgements

The Author's Effort

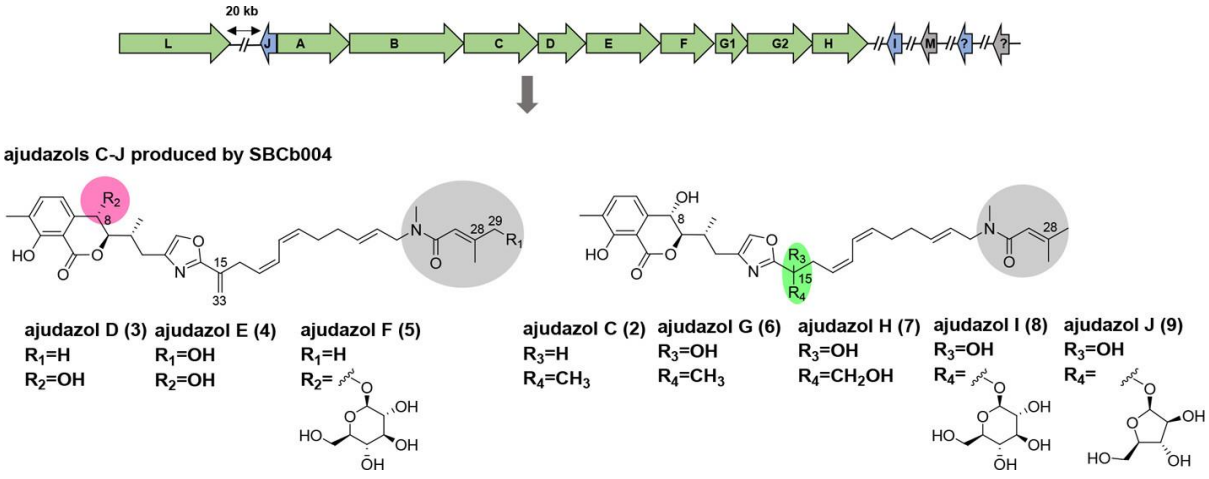
This author significantly contributed to the concept of this study, designed and performed experiments, and interpreted results. NMR based structure elucidation of ajudazol C-E and H was conducted by the autor. Furthermore, the author performed the characterisation, verification and writing of the structure elucidation text of all ajudazol congeners, as well as the sugar analysis of glycosylated ajudazols. Moreover, the author contributed by conception and writing this manuscript.

Contributions by Others

Hu Zeng contributed to the conception of this study, designed, performed and evaluated the *in silico* analysis of the ajudazol gene cluster. He performed compound purification, mutagenesis work and feeding experiments, as well as proposed the biosynthetic pathway of ajudazols. Furthermore, Hu Zeng contributed to conceiving and writing this manuscript. Judith Hoffmann performed the structure elucidation of ajudazol F and J, and Alexander Popoff performed the structure elucidation of ajudazol G. Carsten Volz contributed to supervision, conception, writing and editing of this manuscript. Rolf Müller was responsible for the conception and supervision of the project and performed the proofreading of this manuscript.

3.1. Abstract

Myxobacteria have proven to be a rich source of natural products, but their biosynthetic potential seems to be underexplored given the high number of biosynthetic gene clusters present in their genomes. In this study, a truncated ajudazol biosynthetic gene cluster in *Cystobacter* sp. SBCb004 was identified using mutagenesis and metabolomics analyses and a set of novel ajudazols (named ajudazols C–J, **3–10**, respectively) were detected and subsequently isolated. Their structures were elucidated using comprehensive HR-MS and NMR spectroscopy. Unlike the known ajudazols A (**1**) and B (**2**), which utilise acetyl-CoA as the biosynthetic starter unit, these novel ajudazols were proposed to incorporate 3,3-dimethylacrylyl CoA as the starter. Ajudazols C–J (**3–10**, respectively) are characterised by varying degrees of hydroxylation, desaturation, and different glycosylation patterns. Two P450-dependent enzymes and one glycosyltransferase are shown to be responsible for the hydroxylation at C-8, the desaturation at C-15 and C-33, and the transfer of a D-*β*-glucopyranose, respectively, based on mutagenesis results. One of the cytochrome P450-dependent enzymes and the glycosyltransferase were found to be encoded by genes located outside the biosynthetic gene cluster. Ajudazols C–H (**3–8**, respectively) exhibit cytotoxicity against various cancer cell lines.



Graphical Abstract. *Identification of a truncated ajudazol biosynthetic gene cluster in* Cystobacter *sp. SBCb004 led to a set of novel ajudazols (C-J,* ***3-10****).*

3.2. Introduction

Myxobacteria are Gram-negative *δ*-proteobacteria ubiquitously occurring on earth.[1] They represent an extraordinary order of bacteria due to their ability to swarm, to form fruiting bodies upon starvation and prey on other microorganisms.[2] Myxobacteria have emerged as a rich reservoir of secondary metabolites which exhibit antimicrobial (e.g. cystobactamid), antifungal (e.g. myxalamid) and antitumor (e.g. epothilone) activities.[3] Many myxobacterial secondary metabolites have unique structural features as well as novel modes of action.[4] In addition, myxobacteria harbor the largest genomes among prokaryotes known so far.[5] These genomes possess numerous regions encoding biosynthetic gene clusters (BGCs), and the corresponding secondary metabolites of most of them are still unknown. This observation indicates a large unexplored biosynthetic potential in the order of myxobacteria.

Many of the known secondary metabolites from myxobacteria are synthesised by type I polyketide synthases (PKSs) or non-ribosomal peptide synthetases (NRPSs). Both machineries are multi-modular megaenzymes that are responsible for the biosynthesis of a vast number of highly diverse natural products. This diversity may arise from the incorporation of unusual building blocks, the comprehensive combination of these two types of assembly lines and numerous possible modifications of the core structure introduced by specialised enzymes.[6,7]

Cystobacter sp. SBCb004 possesses a large genome comprising 49 BGCs predicted by antiSMASH 5.0[8] and is able to produce multiple bioactive compounds including tubulysin,[9] stigmatellin[10] and argyrin.[11] However, no natural product could be assigned to 42 of the uncharacterised BGCs we predicted in the *Cystobacter* sp. SBCb004 genome (**Table S3. 1**). Here we report on the functional characterisation of an ajudazol-like BGC from *Cystobacter* sp. SBCb004 and the discovery of a set of novel ajudazol derivatives (**3**-**10**), suggesting that the biosynthetic potential of myxobacteria is underexplored.

3.3. Results and Discussion

Genomic-driven discovery of novel ajudazol derivatives. *Cystobacter* sp. SBCb004 was previously reported to be the producer of diverse bioactive compounds including argyrin and tubulysin.[9,11] However, our *in silico* analysis using antiSMASH 5.0[8] showed that its genome harbors 49 putative biosynthetic gene clusters (BGCs) (**Table S3. 1**). Apart from the BGCs of known compounds including argyrin,[11] tubulysin,[9] stigmatellin,[10] alkylresorcinol,[12] VEPE/AEPE/TG-1,[13] carotenoids[14], and geosmin,[15] the remaining 42 BGCs have no known compound assigned (**Table S3. 1**). Among them, a PKS-NRPS (AjuA-AjuH) encoding region with high similarity to the known BGC of ajudazols A (**1**) and B (**2**) in *Chondromyces crocatus* Cm c5[16] was identified (**Figure 3. 1**A). However, gene *ajuK* and gene *ajuL,* which together encode for the first four modules in the biosynthesis of **1** and **2**, respectively, are missing in this region. By extending the genomic analysis to the upstream and downstream genes of the BGC as defined by antiSMASH, only the homologue of *ajuL* was found 20 kb upstream of *ajuA* (**Figure 3. 1**A, **Table S3. 4**). Furthermore, neither **1** nor **2** could be detected in any of the extracts obtained from *Cystobacter* sp. SBCb004. Owing to this discrepancy, we decided to investigate this further and inactivated the genes homologous to *ajuD* and *ajuL* (**Table S3. 4**) via insertional mutagenesis. Principal component analysis (PCA) of the extracts obtained from the mutants SBCb004 ajuD$^-$, SBCb004 ajuL$^-$ and the wild type strain revealed a set of compounds (**3**-**10**) with *m/z* values ranging from 575 to 771 disappearing in both strains SBCb004 ajuD$^-$ and SBCb004 ajuL$^-$ (**Figure 3. 1**B and **Figure S3. 2**). Compounds correlated to these *m/z* values could not be identified in either the Dictionary of Natural Products[17] or our in-house database Mxbase.[18] These compounds were subsequently isolated and their structures were elucidated using comprehensive HR-MS and NMR spectroscopy. Due to their structural similarities to the known ajudazols A and B, they were considered to be novel ajudazol derivatives and were named ajudazols C-J (**3**-**10**) (**Figure 3. 1**C).

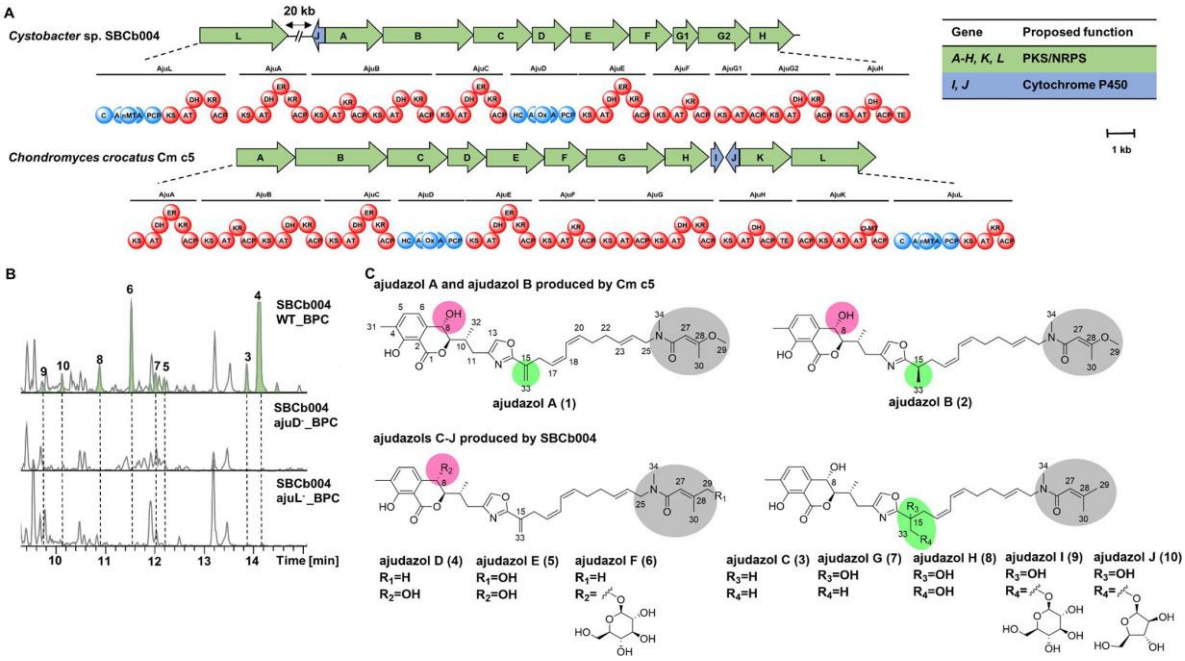


Figure 3. 1. Genomic-driven discovery of novel ajudazol derivatives. *(A) Comparison of the ajudazol biosynthetic gene cluster (BGC) in* Cystobacter *sp. SBCb004 with the ajudazol BGC in* Chondromyces crocatus *Cm c5. The function and organisation of AjuK in* C. crocatus *Cm c5 is shown in* ***Figure 3. 3****A. Gene ajuL in* Cystobacter *sp. SBCb004 is located 20 kb upstream of gene ajuJ. The genes flanking ajuL and found between ajuL and ajuJ are shown in* ***Figure 3. 3*** *and* ***Table S3. 4****. The functions of AjuL and AjuA-AjuH are shown in* ***Figure 3. 4****. (B) Comparative HPLC-MS analysis of SBCb004 ajuD⁻, SBCb004 ajuL, and SBCb004 wild type (WT) extracts. 3-10: Ajudazols C-J. BPC: base peak chromatogram. (C) Chemical structures of ajudazols A-J (1-10). Structural differences are highlighted in different colours.*

HRESIMS analysis of ajudazol C (**3**) revealed a $[M+H]^+$ signal at *m/z* 577.3270 corresponding to the molecular sum formula $C_{34}H_{44}N_2O_6$, which requires 14 double-bond equivalents. Comparison of ^1H and ^{13}C NMR data of ajudazol C (**3**) with those of ajudazol B (**2**)[19] showed identical shifts of the *anti*-configured hydroxyisochromanone core as well as the side chain bearing the oxazole moiety, an (*R*)-configured stereogenic center at position C-15 and the (*E*,*Z*,*Z*)-dodeca-2,6,8-trien-1-yl moiety (for comparison of shifts see, **Table S3. 8** and **Table S3. 9**). Interestingly, one methyl group ($\delta_{C\text{-}34}$ 35.8 and 33.2, $\delta_{H\text{-}34}$ 2.93 and 2.88) showed split ^1H and ^{13}C signals. This phenomenon, named the *E*/*Z*-rotamer effect, can be caused by amide substructures such as those observed in ajudazol B (**2**).[20,21] However, the signals of the remaining $C_6H_{10}NO$ unit of **3** did not match the signals of ajudazol B. Thus, the terminal part of the side chain was elucidated based on HMBC and COSY correlations resulting in a 3,3-dimethylacrylamide moiety (**Figure 3. 2**A and **Table S3. 10**). In detail, the protons of two terminal methyl groups ($\delta_{H\text{-}29}$ 1.84 and $\delta_{H\text{-}30}$ 1.86) showed COSY correlations to a vinylic methine proton ($\delta_{H\text{-}27}$ 5.88) as well as HMBC correlations to C-27 ($\delta_{C\text{-}27}$ 119.4). A sp^2

hybridised carbon ($\delta_{C\text{-}28}$ 148.2) could be identified due to HMBC correlations from H-30 and H-27. Furthermore, the terminal methyl H-30 ($\delta_{H\text{-}30}$ 1.86) and the vinylic methine ($\delta_{H\text{-}27}$ 5.88) protons and the split methyl signals ($\delta_{H\text{-}34}$ 2.93 and 2.88) and methylene protons ($\delta_{H\text{-}25}$ 3.90 and 3.93) showed HMBC correlations to a carbon with a resonance characteristic for a carbonyl carbon ($\delta_{C\text{-}26}$ 170.8), forming a 3,3-dimethylacrylyl moiety. Characteristic proton and carbon shift values as well as the observed split signals of both groups (CH_3-34 and CH_2-25) proved a direct connection to a tertiary amide. Additionally, COSY correlations between H-24 and H-25 link the 3,3-dimethylacrylamide moiety to the (*E*,*Z*,*Z*)-dodeca-2,6,8-trien-1-yl moiety of **3**. Comparative analysis of ^{1}H and ^{13}C NMR data as well as vicinal coupling constants indicated the configuration of the hydroxyisochromanone core (*anti*), the stereogenic center at position C-15 (*R*) and the dodeca-2,6,8-trien-1-yl moiety (*E*,*Z*,*Z*) were identical to those as in ajudazol B (**2**) (for comparison of shifts and coupling constants, see **Table S3. 8** and **Table S3. 9**). Comparing the optical rotation values with the literature known values of ajudazol B confirms these findings.[20] Further evidence of the configuration match of **2** and **3** is provided by the putative biosynthetic pathway and the bioinformatic gene cluster analysis, as their common structural skeleton is assembled by the identical PKS/NRPS processing steps and the configurational assignment based on bioinformatic gene cluster analysis shows that the hydroxyl- and methyl-bearing stereocenters of **3** are the same to those of **2** (**Figure 3. 4** and **Table S3. 5**).

HRESIMS of ajudazol D (**4**) revealed a $[M+H]^+$ signal at *m/z* 575.3128 corresponding to the molecular formula $C_{34}H_{42}N_2O_6$ which requires 15 double-bond equivalents. 2D-NMR analysis of **4** and comparison to **3** revealed many structural similarities. However, the HSQC signal of C-15 in **3** disappeared in **4** in exchange for a new carbon resonance ($\delta_{C\text{-}15}$ 136.0). The protons of a terminal double bond ($\delta_{C\text{-}33}$ 118.8, $\delta_{H\text{-}33}$ 5.97 and 5.42) and the proton of a sp^3 hybridised methylene group ($\delta_{C\text{-}16}$ 31.3, $\delta_{H\text{-}16}$ 3.36) showed HMBC correlations with C-15. Furthermore, comparing **4** to **3**, downfield shifting of H-16 and H-17 ($\delta_{H\text{-}16}$ 3.36, $\delta_{H\text{-}17}$ 5.56) could be observed due to deshielding effects of the exomethylene group, proving its connection to C-15 (**Table S3. 11**), similar to ajudazol A (**1**).[19]

Ajudazol E (**5**) exhibits a $[M+H]^+$ signal at *m/z* 591.3077, which can be assigned to a molecular sum formula of $C_{34}H_{42}N_2O_7$ that requires 15 double-bond equivalents. Analysis of the HSQC spectrum showed its high similarity to **4** (**Table S3. 12**). Nevertheless, instead of two terminal vinylic methyl groups (C-29 and C-30) in **4**, only one could be observed in **5** ($\delta_{C\text{-}30}$ 15.7, $\delta_{H\text{-}30}$ 1.81). Furthermore, an additional methylene group ($\delta_{C\text{-}29}$ 66.9, $\delta_{H\text{-}29}$ 4.03) showed COSY correlations to H-30 and H-27 ($\delta_{H\text{-}30}$ 1.81, $\delta_{H\text{-}27}$ 6.16). Characteristic shift values as well as upfield shifts of H-30 and H-27 indicated the direct proximity of a hydroxy group to C-29. Thus, **5** proved to be the 29-hydroxy derivative of **4**.

The formula of ajudazol F (**6**) $C_{40}H_{52}N_2O_{11}$ was supported by HRESIMS data, requiring 16 double-bond equivalents. 2D-NMR data of **6** highly resembled data of **4** (**Table S3. 13**), whereas close investigation of HSQC spectra showed an additional anomeric signal ($\delta_{C\text{-}1'}$ 103.6, $\delta_{H\text{-}1'}$ 4.57). Furthermore, a group of four methine groups ($\delta_{C\text{-}2'\text{-}5'}$ 71.6-78.2, $\delta_{H\text{-}2'\text{-}5'}$ 3.20-3.37) and one diastereotopic methylene group ($\delta_{C\text{-}6'}$ 62.8, $\delta_{H\text{-}6'}$ 3.78 and 3.62) could be assigned to a pyranose sugar. Based on characteristic carbon shifts, the pyranose moiety could be identified as glucopyranose.[22,23] A large coupling constant between H-1' and H-2' ($^3J_{H,H}$ 7.8 Hz) was observed, indicating a *β*-anomeric configuration of glucopyranose.[22,24] HMBC correlations from H-1' to C-8 proved the presence of an *O*-glycosidic bond between C-1' of the *β*-glucopyranose moiety and C-8 (**Figure 3. 2**B)[22]. GC-MS analysis of the glycon and authentic standards provided the absolute configuration of D-*β*-glucopyranose as the glycon of **6** after derivatisation with (*S*)-2-butanol and trimethylsilylation (**Figure S3. 4**).

HRESIMS analysis of ajudazol G (**7**) revealed a molecular formula of $C_{34}H_{44}N_2O_7$, which requires 14 double-bond equivalents. High similarities of 2D-NMR data with the data of **3** implied identical structural features such as the hydroxyisochromanone core, the oxazole, the (*E*,*Z*,*Z*)-dodeca-2,6,8-trien-1-yl, and the 3,3-dimethylacrylamide moiety. However, the methyl protons H-33 and the methylene protons H-16 showed downfield-shifted proton resonances ($\delta_{H\text{-}33}$ 1.56, $\delta_{H\text{-}16}$ 2.77 and 2.71). Moreover, instead of the methine group C-15 in **2** an additional carbon resonance ($\delta_{C\text{-}15}$ 72.7) could be observed in **6**. H-33 and H-16 showed HMBC correlations to this resonance, confirming **7** was the C-15 hydroxylated derivative of **3** (**Table S3. 14**).

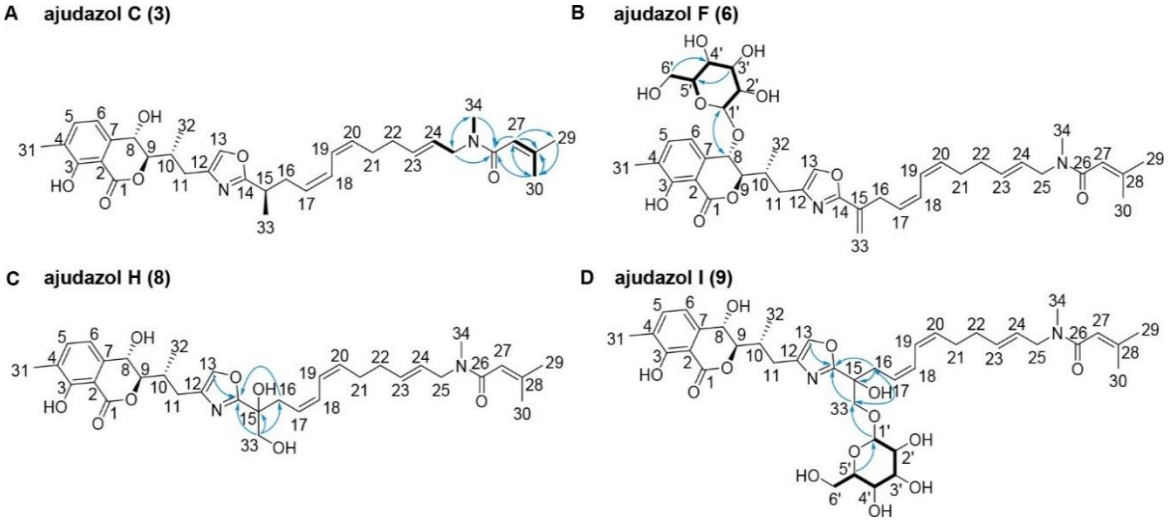


Figure 3. 2. Key HMBC (arrows) and COSY (bold lines) correlations of the moieties in ajudazols C (3), F (6), H (8) and I (9). *(A) Key HMBC and COSY correlations of the 3,3-dimethylacrylamide moiety of **3**. (B) Key HMBC and COSY correlations of the (1'-8)-O-gycosidic bond with D-β-glucopyranose in **6**. (C) Key HMBC correlations of the hydroxymethylene moiety connected to C-15 in **8**. (D) Key HMBC and COSY correlations of the (1'-33)-O-gycosidic bond with D-β-glucopyranose in **9**.*

Ajudazol H (**8**) showed a $[M+H]^+$ signal at *m/z* 609.3171 corresponding to the molecular formula $C_{34}H_{44}N_2O_8$, which requires 14 double-bond equivalents. The NMR data of **8** and **7** revealed similar structural features. Nevertheless, in the HSQC spectra of **8**, a diastereotopic methylene group ($\delta_{C\text{-}33}$ 68.0, $\delta_{H\text{-}33}$ 3.76 and 3.83) with characteristic resonances for a hydroxymethylene group could be identified. Both methylene protons showed HMBC correlations to C-16. Therefore, **8** could be identified as the C-33-hydroxylated derivative of **7** (**Figure 3. 2**C and **Table S3. 15**).

The formulas $C_{40}H_{54}N_2O_{13}$ and $C_{39}H_{52}N_2O_{12}$ for ajudazol I (**9**) and J (**10**), respectively, are supported by HRESIMS analysis and both requiring 15 double bond equivalents. 2D-NMR experiments focused on comparing **9** to **8** showed an additional methine signal ($\delta_{C\text{-}1'}$ 105.0, $\delta_{H\text{-}1'}$ 4.28) in the HSQC spectrum, which is characteristic for an anomeric group of a sugar. Furthermore, a group of four methine groups ($\delta_{C\text{-}2'\text{-}5'}$ 71.7-78.2, $\delta_{H\text{-}2'\text{-}5'}$ 3.13-3.34) and one diastereotopic methylene group ($\delta_{C\text{-}6'}$ 62.9, $\delta_{H\text{-}6'}$ 3.83 and 3.59) could be assigned to a pyranose sugar by COSY and HMBC spectra. The characteristic carbon shifts of the pyranose moiety were used to identify it as glucopyranose.[22,23] Furthermore, the large coupling constant between H-1' and H-2' ($^3J_{H,H}$ 7.7 Hz) proved a β-anomeric configuration of the glucopyranose moiety. The HSQC spectra of **10** and **9** were similar (**Table S3. 16** and **Table S3. 15**). Nevertheless, **10** showed one characteristic sugar methine group than **9**. Due to COSY and HMBC correlations, resonances could be assigned to a pentofuranose moiety. In both **9** and **10**, HMBC correlations proved the presence of an *O*-

glycosidic bond between C-1' and C-33 (**Figure 3. 2**D), respectively. The absolute configuration of the glycons was provided by GC-MS analysis after derivatisation with (*S*)-2-butanol and trimethylsilylation of the sugars and authentic standards. Comparative analysis indicated the glycon of **9** to be D-*β*-glucopyranose and that of **10** to be D-*β*-arabinofuranose (**Figure S3. 4** and **Figure S3. 5**).

In summary, the new ajudazol derivatives (**3-10**) show identical structural features in comparison to ajudazols A and B, such as the isochromanone core, the oxazole and the (*E*,*Z*,*Z*)-dodeca-2,6,8-trien-1-yl moiety. However, **3-10** feature a 3,3-dimethylacrylamide instead of the 3-methoxy-butenamide moiety found in ajudazols A and B. Furthermore, ajudazols A and B only show hydroxylation at C-8, whereas new derivatives feature hydroxylation at additional positions. **5** is hydroxylated at C-29, and **7-10** are hydroxylated at C-15. In addition, **8-10** are hydroxylated at C-33. Moreover, **6**, **9** and **10** are *O*-glycosylated with D-*β*-glucopyranose or D-*β*-arabinofuranose at C-8 or C-33, respectively (**Figure 3. 1**C).

The starter unit of the biosynthesis of 3-10. As shown in **Figure 3. 1**A, all genes encoding the respective PKS and NRPS homologues necessary for the biosynthesis of **1** and **2** are present in *Cystobacter* sp. SBCb004 except gene *ajuK*. In the biosynthetic machinery of **1** and **2** in *C. crocatus* Cm c5, *ajuK* encodes for both the loading module and the first extension module, which are proposed to condense one acetyl-CoA and one malonyl-CoA to form a diketide intermediate. Following the formation of the intermediate, enolisation and *O*-methylation at the distal carbonyl group generate 3-methoxy-butenyl-ACP (**Figure 3. 3**A).[16] Retro-biosynthetic analysis indicated that **3-10** may instead derive from 3,3-dimethylacrylyl CoA (DMA-CoA), which could originate from the catabolism of leucine as well as the 3-methylglutaconyl-CoA decarboxylase-catalysed biosynthesis pathway of isovaleryl CoA in myxobacteria.[25] In order to check if DMA-CoA could be utilised by the assembly line of **3-10**, d_6-3,3-dimethylacrylic acid was fed to *Cystobacter* sp. SBCb004, and the isotopic pattern of **4**, which had the highest intensity in the chromatogram, was inspected. This showed that the isotopic peak m/z 581 of **4** (m/z 575) was enriched (**Figure 3. 3**C and 3D), indicating the incorporation of DMA-CoA. Due to the fact that *ajuK* is not present in *Cystobacter* sp. SBCb004, we propose that DMA-CoA is used as a starter unit of the NRPS-PKS assembly line instead of the commonly incorporated starter units malonyl-CoA and acetyl-CoA. Such an alternative starter unit was also found previously in the biosynthesis of myxothiazol[26] and aurafuron,[27] where isovaleryl-CoA derived from DMA-CoA was shown to be used as the precursor.

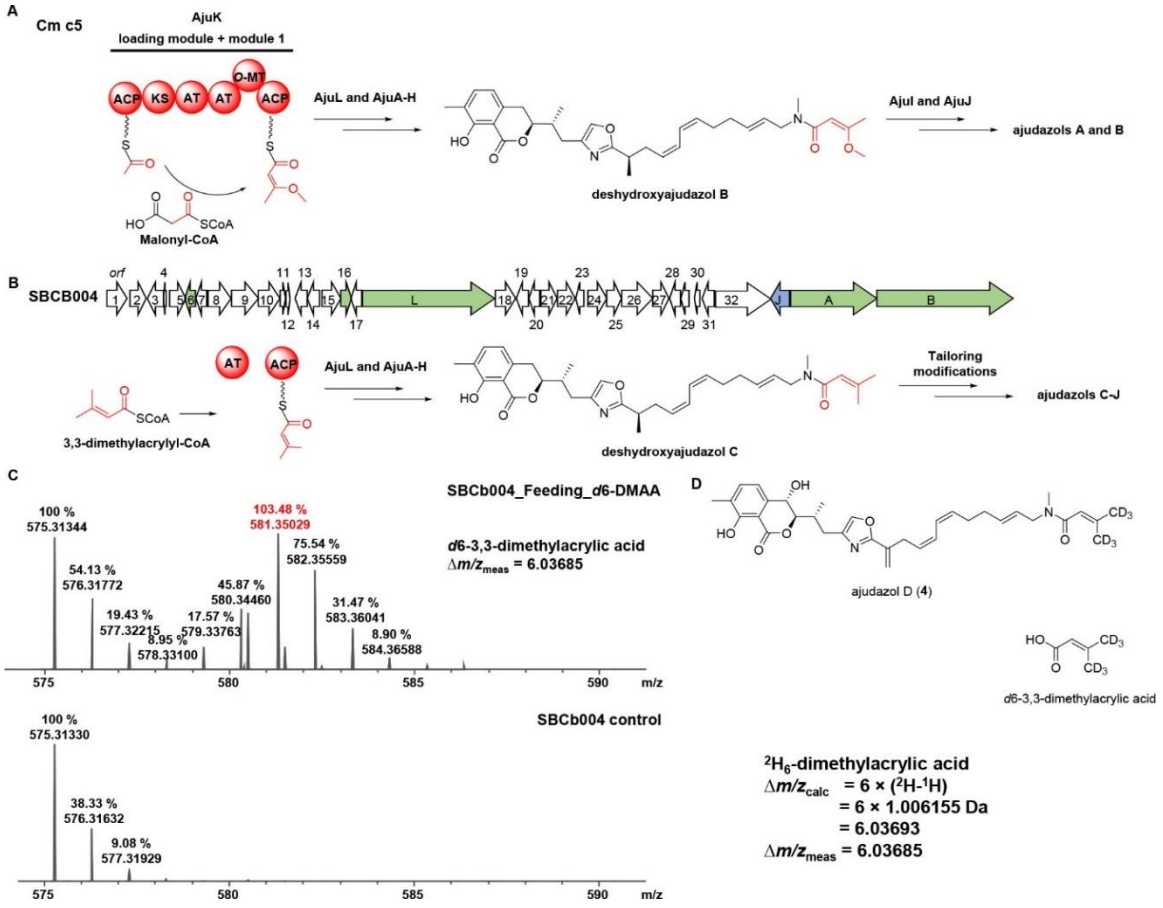


Figure 3. 3. The initiation of ajudazol biosynthesis in Chondromyces crocatus Cm c5 (ajudazols A and B) and Cystobacter sp. SBCb004 (ajudazols C-J). *(A) Proposed generation of 3-methoxy-butenoyl moiety in ajudazols A and B accomplished by AjuK in C. crocatus Cm c5. (B) The putative open reading frames (ORF) adjacent to ajuL in Cystobacter sp. SBCb004 and the proposed generation of 3,3-dimethylacrylyl moiety in ajudazols C-J. The function of AjuL and AjuA-AjuH is shown in* ***Figure 3. 4****. (C) Isotopic peak pattern of the [M+H]+ signals for ajudazol D (575 m/z) prove incorporation of labelled dimethylacrylic acid in Cystobacter sp. SBCb004. The observed mass shifts fit the heavy isotopes that were incorporated. (D) Location of the deuterium label in the fed dimethylacrylic acid and likely incorporation pattern of dimethylacrylic acid skeleton into ajudazol D. ACP: acyl carrier protein. KS: ketosynthase. AT: acyltransferase. O-MT: O-methyltransferase.*

To activate and load DMA-CoA to the assembly line, an AT domain and an ACP domain are needed. Our *in silico* analysis of all the genes 20 kb upstream and downstream of *ajuL* in *Cystobacter* sp. SBCb004 revealed that *orf6* might encode a GCN5-related N-acetyltransferase upstream of *ajuL*. In addition, *orf16*, putatively encoding a 4'-phosphopantetheinyl transferase superfamily protein, was found (**Figure 3. 3**B and **Table S3. 4**). However, as no genetic tools for the construction of in-frame deletion mutants are available for this strain, we had to rely on single crossover-based mutagenesis. Due to the short length of the coding regions, the insertional mutagenesis of *orf6* and *orf16* was not achieved because we found 1kb to be the minimal DNA fragment size required

for homologous recombination. As the hypothetical functions of the other genes among *orf1-orf32* seem not to be involved in the biosynthesis of ajudazols C-J (**3-10**) (**Table S3. 4**), we speculated that the required activation and the loading of DMA-CoA to the assembly line might either be accomplished by ORF6 and ORF16 or by alternative standalone AT domain and ACP domain containing proteins encoded elsewhere in the genome (**Figure 3. 3**B). The precursor would then be subjected to 12 rounds of PKS/NRPS chain extension performed by AjuL and AjuA-H to generate deshydroxyajudazol C (**Figure 3. 4**), which would subsequently be converted into **3-10** in multiple enzymatic reactions (**Figure 3. 5**).

The hydroxylation and desaturation enzymes in the biosynthesis of 3-10. In *C. crocatus* Cm c5, two cytochrome P450-dependent enzymes, AjuI and AjuJ, were described to catalyse the desaturation between C-15 and C-33 and hydroxylation at C-8, respectively.[16] However, only the homologoue of *ajuJ* could be found in the ajudazol BGC in *Cystobacter* sp. SBCb004 (**Figure 3. 5**A). The inactivation of *ajuJ* in *Cystobacter* sp. SBCb004 results in the accumulation of the proposed deshydroxyajudazol C (**11**) (m/z 561) and deshydroxyajudazol D (**12**) (m/z 559) and the abolishment of **3**-**10** (**Figure 3. 5**B and 5C), suggesting that AjuJ is responsible for the hydroxylation at C-8 and the desaturation between C-15 and C-33 is performed independently of this enzyme. In order to identify candidate enzymes responsible for the desaturation between C-15 and C-33, the amino acid sequence of AjuI from *C. crocatus* Cm c5 was used as a query to screen the available genome of *Cystobacter* sp. SBCb004. Consequently, two coding regions, *ctg27_313* (80%, amino acids similarity) and *ctg19_169* (60%, amino acids similarity) were found (**Table S3. 6**) and inactivated via insertional mutagenesis. In the inactivation mutant of *ctg27_313*, from now on referred to as strain SBCb004 ajuI$^-$, only **3** was synthesised (**Figure 3. 5**C), supporting the involvement of AjuI in the desaturation between C-15 and C-33. As the production of **7-10** was also blocked in the SBCb004 ajuI$^-$ mutant (**Figure 3. 5**C), the presence of a double bond between C-15 and C-33 is reasoned to be a prerequisite for the hydroxylations at C-15 and C-33. Compounds **11** and **12** are only present in trace amounts in the extracts of *Cystobacter* sp. SBCb004 wild-type strain (**Figure 3. 5**C), indicating that the reactions catalysed by AjuI and AjuJ are quite efficient. To locate the enzymes that catalyse the hydroxylations at C-15, C-29, and C-33, all the predicted oxidoreductases within the BGC, *orf15*, *orf28* and *orf32* were inactivated (**Figure 3. 3**B and **Table S3. 4**). Against all the expectations, the production of **3-10** was not influenced in any of the mutants (data not shown here), suggesting the involvement of additional enzymes encoded elsewhere in the genome or some redundancy in the functionality of single enzymes, as they may complement

each other. It remains elusive whether the hydroxylations at C-15, C-29, and C-33 result from the activity of three different enzymes or from a broad substrate tolerance of one enzyme.

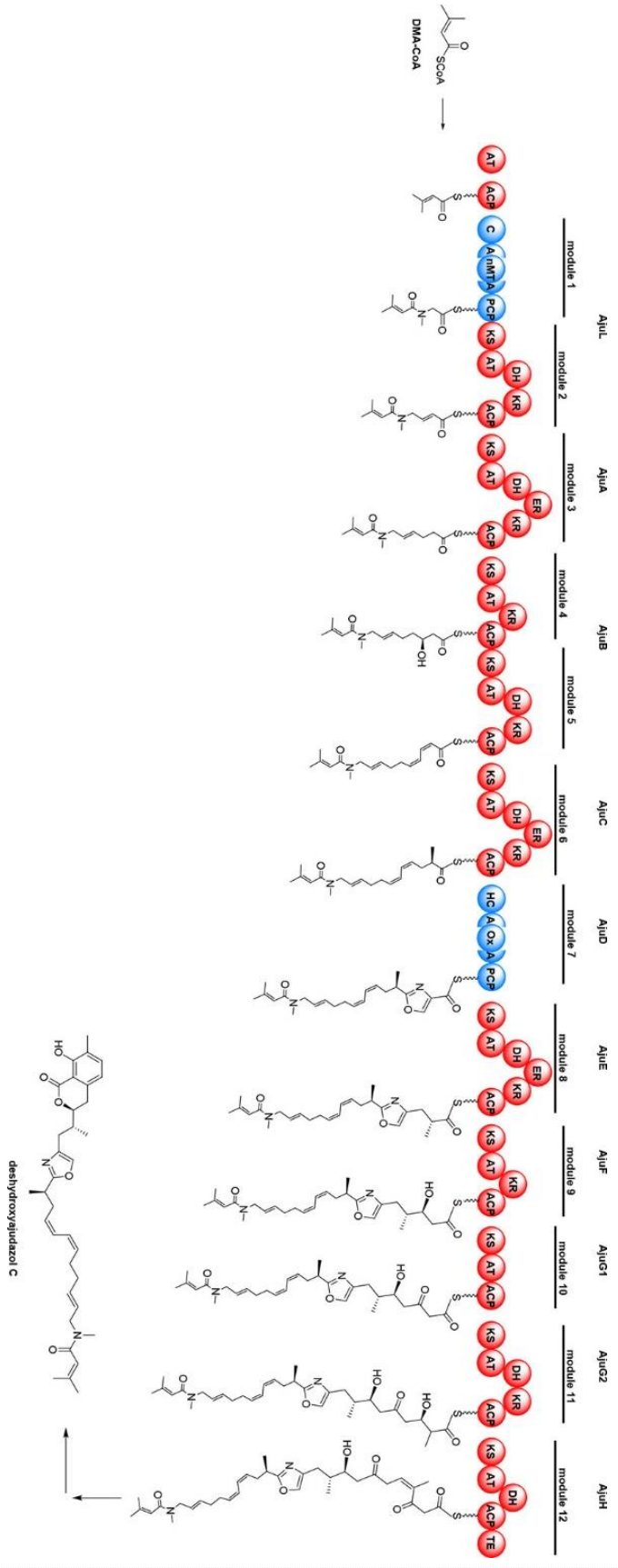


Figure 3. 4. Model for the biosynthesis of deshydroxyajudazol C under the action of the mixed PKS–NRPS enzymes in Cystobacter sp. SBCb004. *DMA-CoA: 3,3-dimethylacrylyl CoA. ACP: acyl carrier protein. KS: ketosynthase. AT: acyltransferase. DH: dehydratase. ER: enoyl reductase. KR: ketoreductase. TE: thioesterase. C: condensation. A: adenylation. PCP: peptidyl carrier peptide. nMT: N-methyltransferase. HC: heterocyclisation. Ox: oxidation.*

The glycosylation in the biosynthesis of 3-10. Glycosylation is one of the most common modification of natural products after the synthesis of the core structure. Such glycosylation may have profound effects on bioactivities, selectivity, and pharmacokinetic properties of the mature products. This kind of modification has been observed in the biosynthesis of numerous myxobacterial compounds, including sorangicins,[28] disciformycins[29], and cystomanamides.[30] However, to our knowledge, only the glycosyltransferase involved in sorangicin biosynthesis, SorF,[28] was characterised from myxobacteria, while those involved in other compounds were not encoded in the clusters and therefore not identified yet. Flanking the core biosynthetic genes *ajuA-H* and *ajuL*, only one putative glycosyltransferase encoded by *orf2* could be identified (**Figure 3. 3**B and **Table S3. 4**). However, a respective inactivation mutant constructed during our investigations did not show any influence on the biosynthesis of ajudazol derivatives (data not shown here), implying that the involved proteins are not encoded within this region or may be complemented by other enzymes encoded in the genome. Considering the fact that SorF is able to utilise activated hexose as a natural donor substrate *in vivo* and has a relaxed donor substrate tolerance *in vitro*,[28] it is reasonable to conclude that the glycosyl groups in **6**, **9**, and **10** may be transferred by similar enzymes. Thus, SorF was used as a query to identify such candidate enzymes in the available genome of *Cystobacter* sp. SBCb004. Three hits were obtained in our search, namely *ctg1_130* (41%, amino acids similarity), *ctg10_386* (44%, amino acids similarity), and *ctg19_179* (50%, amino acids similarity) (**Table S3. 7**) and corresponding gene inactivation mutants were generated. In the inactivation mutant of *ctg19_179* (now referred to as *ajuM*; SBCb004 ajuM$^-$), the production of **6** and **9** was abolished while the production of ajudazol J (**10**) was not influenced (**Figure 3. 5**C). We reasoned that AjuM is able to perform the transfer of D-*β*-glucopyranose to *O*-8 in **4** and *O*-33 in **8**, yielding **6** and **9**, respectively (**Figure 3. 5**B). However, the enzyme that catalyses the transfer of the D-*β*-arabinofuranose remains elusive.

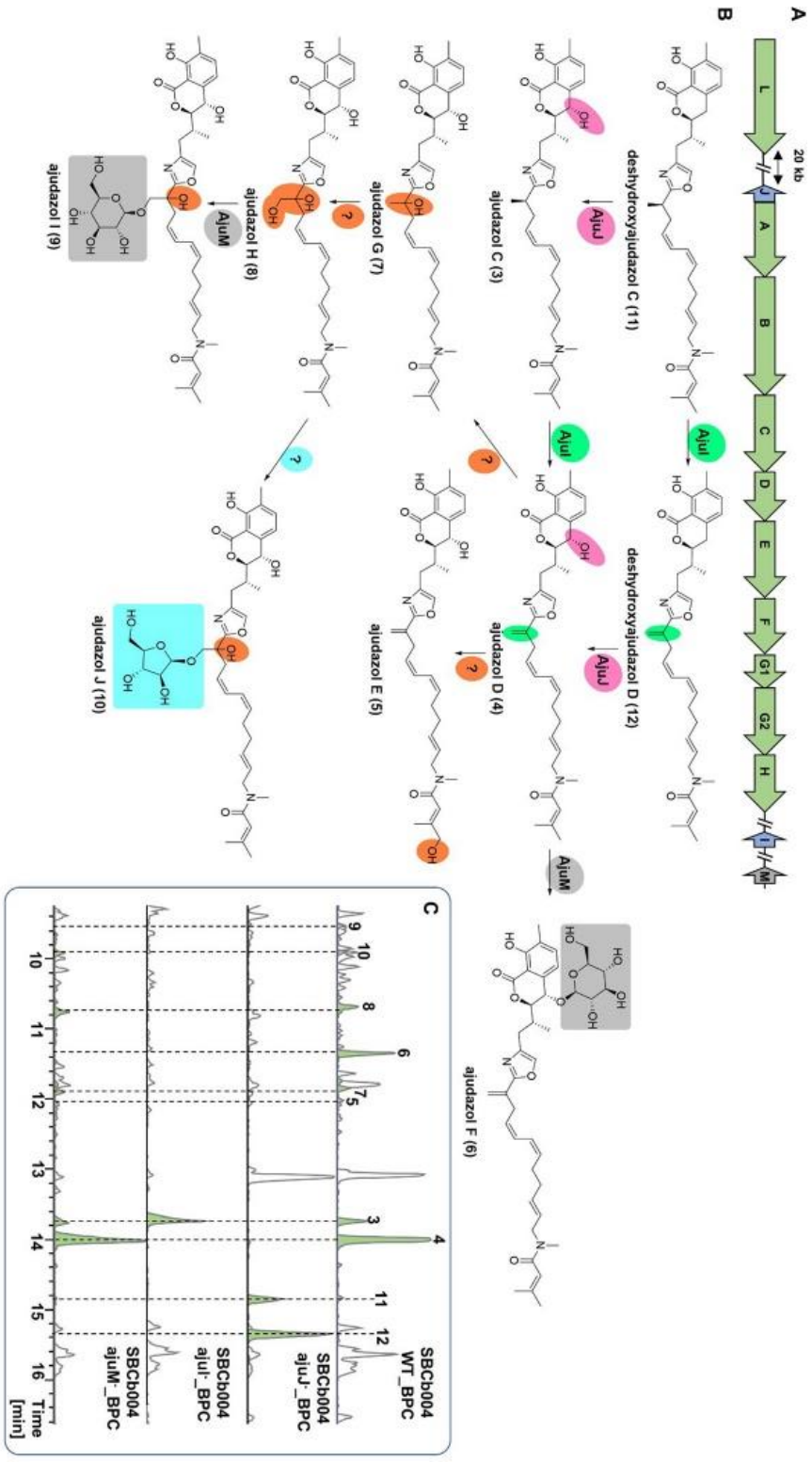


Figure 3. 5. Tailoring modifications of ajudazols C-J (3-10, respectively) in **Cystobacter** ***sp. SBCb004.*** *(A) Putative biosynthetic genes found in the biosynthesis of ajudazols C-J (**3-10**, respectively). (B) Proposed modification steps to generate ajudazols C-J (**3-10**, respectively) in* Cystobacter *sp. SBCb004. Modifications in each step of the biosynthesis are highlighted in colour. (C) Comparative HPLC-MS analysis of SBCb004 ajuJ$^-$, SBCb004 ajuI$^-$, SBCb004 ajuM$^-$ and SBCb004 wild-type (WT) extracts. **3-10**: ajudazols C-J. **11** and **12**: deshydroxyajudazols C and D.*

Biosynthetic overview of 3-10. Based on the above results, a proposal for the biosynthesis of **3-10** was devised. In *Cystobacter* sp. SBCb004, DMA-CoA is used as the starter unit. Thus, the enzymatic functions encoded by *orf6* and *orf16* or other standalone enzymes are proposed to be involved (**Figure 3. 3**B). The subsequent steps of the biosynthesis are accomplished by NRPS-PKS extensions similar to the described ajudazol biosynthesis pathway in *C. crocatus* Cm c5 (AjuL and AjuA-H) (**Figure 3. 4**). The resulting product deshydroxyajudazol C (**11**) is either hydroxylated by AjuJ at C-8 to yield ajudazol C (**3**) or desaturated by AjuI between C-15 and C-33 to yield deshydroxyajudazol D (**12**). The latter product is subjected to further hydroxylation at C-8 to form ajudazol D (**4**). Compound **3** could also be transformed into **4** after an initial desaturation conducted by AjuI, while the double bond between C-15 and C-33 in **4** could be hydroxylated to yield **7** and **8**, successively. Compound **4** could be modified to yield **5** with a hydroxylation at C-29. The hydroxylated products **4** and **8** would subsequently be glycosylated with a D-*β*-glucopyranose or D-*β*-arabinofuranose, yielding **6**, **9**, and **10** (**Figure 3. 5**B). In theory, **3**, **5**, and **7** could also be glycosylated and **4** could accept a D-*β*-arabinofuranose. Their putative products could be detected by LC-MS based on their expected mass. However, their production yields were not sufficient for isolation, which could be explained by the specificity and selectivity of the enzymes responsible for the subsequent modifications. Considering the structural similarity of AjuL substrates utilised in strains *C. crocatus* Cm c5 and *Cystobacter* sp. SBCb004, their ajudazol BGCs are hypothesised to be evolutionarily closely related. It may be either that either the gene *ajuK* was lost in *Cystobacter* sp. SBCb004 during evolution and the production of **3-10** was enabled by a certain substrate plasticity of the residual enzymes of the assembly line encoded in the BGC or *ajuK* was involved later, leading to the production of ajudazols A and B in *C. crocatus* Cm c5. The distribution of genes encoding glycosyltransferases and oxidoreductases in the genome of *Cystobacter* sp. SBCb004 involved in the biosynthesis of **3-10** supports the hypothesis that genes work synergistically regardless of the linear distance to each other. It is still not clear if the genome-wide distribution of these tailoring genes involved in the biosynthesis of **3-10** is caused by the insertion of other DNA fragments or the rearrangement of the genome. A more appealing hypothesis is that **3-10** are the accidental byproducts of some discrete enzymes with a broad substrate tolerance.

Bioactivity. The biological activities of natural products may be severely influenced by both the nature and degree of modifications of the primary core structure. Ajudazol A was reported to only show a minor activity against a few fungi and Gram-positive bacteria.[31] As the novel derivatives **3-10** reported herein exhibit different modifications when compared to ajudazols A and B, **3-10** were

tested for their activity against several pathogens, including bacteria and fungi. In addition, cytotoxicity was determined using various cancer cell lines. However, none of the novel ajudazols exhibited antimicrobial activity against the tested strains (**Table S3. 18**). In contrast, **3**-**8** exhibited cytotoxicity against HepG2, HCT-116, and KB3.1 cells, with half-maximal inhibitory concentrations (IC_{50}) in the sub- to low-micromolar range (**Table S3. 19**). The glycosylated derivatives **9** and **10** exhibited only marginal cytotoxicity (**Table S3. 19**). It thus seems that a glycosylation at position *O*-33 decreases cytotoxicity. Eventually, such glycosylation may act as a mechanism of self-resistance to convert an active compound into inactive form, which can be released by a hydrolase. Similar observations could be made in the case of sorangicin, where glycosylated sorangiosides did not show biological activity.[28]

3.4. Experimental Section

General experimental procedures. UV-vis data were obtained using the DAD of the Dionex Ultimate 3000 SL system linked to a Bruker maXis 4G UHRqTOF for mass spectrometric analysis. Specific optical rotations were measured on a JASCO P-2000 polarimeter. NMR data were recorded on an UltraShield 500 MHz (^{1}H at 500 MHz, ^{13}C at 125 MHz) or an AVANCE III 700 MHz NMR (^{1}H at 700 MHz, ^{13}C at 175 MHz) equipped with a 5 mm inverse TCI cryoprobe (Bruker, Billerica, MA, USA). Shift values (δ) were calculated in ppm, and coupling constants (J) were calculated in Hz. Spectra were recorded in methanol-d_4, acetone-d_6, or chloroform-d and adjusted to the solvent signals (methanol-d_4 δ_H 3.31, δ_C 49.2, acetone-d_6 δ_H 2.05, δ_C 29.9, and chloroform-d δ_H 7.27, δ_C 77.0). Measurements were conducted in 5 mm Shigemi tubes (Shigemi Inc., Allison Park, PA 15101, USA). For the two-dimensional experiments HMBC, HSQC, ROESY, and gCOSY standard pulse programs were used. HMBC experiments were optimised for $^{2,3}J_{C\text{-}H}$ = 6 Hz, and HSQC were optimised for $^{1}J_{C\text{-}H}$ = 145 Hz. UPLC-MS measurement were performed on a Dionex Ultimate 3000 RSLC system equipped with Waters BEH C_{18} column (100 × 2.1 mm, 1.7 µm). The flow rate was set to 600 µL/min, and the column temperature was 45 °C. The gradient was as follows: 0-0.5 min, 95% water with 0.1% formic acid (A) and 5% acetonitrile with 0.1% formic acid (B); 0.5-18.5 min, 5-95% B; 18.5-20.5 min, 95% B; 20.5-21 min, 95-5% B; 21-22.5 min, 5% B. UV spectra 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 Bruker Daltonics maXis 4G hr-qToF mass spectrometer using the Apollo II ESI source. Mass spectra were acquired in centroid mode ranging from 150-2500 m/z at a 2 Hz full scan rate. Full scan and MS spectra were acquired at 2 Hz.

Bacterial strains, culture conditions and reagents. *Escherichia coli* DH10B was used for cloning and was cultivated at 37 °C in LB medium (1% trypton, 0.5% yeast extract, 0.5% NaCl). *Cystobacter* sp. SBCb004 was cultivated at 30 °C and 180 rpm in CYH medium (0.3% casitone, 0.3% yeast extract, 0.2% $CaCl_2 \bullet 2H_2O$, 0.2% soy meal, 0.2% glucose, 0.8% soluble starch, 0.1% $MgSO_4 \bullet 7H_2O$, 1.19% HEPES, 8 mg/L Fe-EDTA, adjusted to pH 7.3 with 10 N KOH) for principal component analysis. M medium (1% phytone, 1% maltose, 0.1% $CaCl_2 \bullet 2H_2O$, 0.1% $MgSO_4 \bullet 7H_2O$, 8 mg/L Fe-EDTA, 1.19% HEPES, adjusted to pH 7.2 with 10 N KOH) was used for transformation and feeding. VY2M medium (0.5% Baker's Yeast, 0.05% $CaCl_2 \bullet 2H_2O$, 0.2% maltose, 0.119% HEPES, 0.5 mg/L Vitamin B12, adjusted to pH 7.0 with 10 N KOH) was used for large-scale fermentation and compound isolation. When necessary, 50 µg/mL Kanamycin was added to the culture.

Restriction endonucleases, Taq DNA polymerase, Phusion DNA polymerase, FastAP, and T4 DNA ligase was purchased from Thermo Scientific. NEBuilder HiFi DNA assembly Master Mix was purchased from New England Biolabs. Oligonucleotide primers were synthesised by Sigma-Aldrich and DNA sequencing of PCR products was performed by LGC Biosearch Technologies.

Gene inactivation in *Cystobacter* sp. SBCb004. Gene inactivation in *Cystobacter* sp. SBCb004 was accomplished by the homologous recombination of a suitable plasmid backbone in the encoding region. Internal DNA fragments of the target genes served as homologous regions to mediate the insertion of the constructed plasmids into the target gene via homologous recombination. Internal DNA regions of the target gene with sizes of 800-1200 bp were amplified from *Cystobacter* sp. SBCb004 genomic DNA by using the corresponding primers 9101-sc-gene-F and 9101-sc-gene-R (**Table S3. 2**). Subsequently, these DNA fragments were cloned into plasmid pbluekan kindly provided by Dr. Carsten Volz (unpublished). pbluekan encodes a aminoglycoside phosphotransferase (*aph*(3')-II) that confers kanamycin resistance[32] as well as a *β*-lactam resistance cassette (*bla*) and a *lacZ* gene under the control of the T7A1 promoter (**Figure 3. 1**).[33] Stop codons were introduced at both ends of the cloned homologous regions upon PCR amplification using the mutated oligonucleotide primer. The constructed plasmids (**Table S3. 3**) were confirmed by Sanger sequencing using primer seq-pblueCN-R (**Table S3. 2**). Correct plasmids were introduced into *Cystobacter* sp. SBCb004 by electroporation following a previously published protocol.[9] The correct transformants were confirmed by PCR reactions with primer CN-check-R, which binds the plasmid backbone, and primer gene_CK_F (where "gene" stands for the designation of the target gene, **Table S3. 2**), which binds the upstream of the target gene.

Principal Component Analysis. *Cystobacter* sp. SBCb004 wild-type strain and inactivation mutants were cultivated in 50 mL CYH medium supplemented with 2% Amberlite XAD-16 resin. The cultivations were performed in triplicate. A CYH medium blank was treated and extracted in parallel as a negative control. After 10 days, the cells and the Amberlite XAD-16 resin were harvested by centrifugation and extracted with 40 mL acetone. The solvent was dried by rotary evaporation, and the extract was dissolved in MeOH, followed by UPLC-MS analysis. The molecular feature finder implemented in Bruker Compass Data Analysis 4.2 was used to process the MS data with the compound detection parameters: SN threshold of 1, correlation coefficient of 0.9; minimum compound length of 10 spectra, and smoothing width of 3 spectra. Bucketing was performed using Bruker Compass Profile Analysis 2.1 with advanced bucketing and window parameters of 30 s and 15 ppm, respectively. The PCA t-Test function was then used to distinguish the features derived from each group. The features not derived from the medium and disappearing in a mutant were proposed to be related to the inactivated genes.

Compound isolation. *Cystobacter* sp. SBCb004 was grown in 20 L VY2M medium supplemented with 2% Amberlite XAD-16 resin at 30 °C for 10 days. Cells and XAD-16 resin were harvested by centrifugation and extracted with 3 × 1.5 L acetone. The extract was partitioned between hexane and methanol (3:1, v/v), and the methanol extract was subsequently partitioned between ethyl acetate and deionised water (3:1, v/v). The ethyl acetate extract was fractionated on a Sephadex LH-20 column (GE healthcare) eluted with methanol to afford two fractions.

The fraction containing ajudazol C-E (**5**), G, and H (**3-5**, **7**, and **8**, respectively) was separated by semi-preparative HPLC (Waters Xbridge Peptide BEH C_{18}, 5 μm, 250 × 10 mm; 5 mL/min; UV 220 nm) using a gradient as follows: 0-5 min, 5% H_2O with 0.1% formic acid (A) and 5% acetonitrile with 0.1% formic acid (B); 5-8 min, 5% B to 58% B; 8-31 min, 58% to 75% B; 31-33 min, 75% to 95% B; 33-34 min, 95% to 5% B; 34-36 min, 5% B. 2.04 mg **3** (t_R=23.2 min) and 11.91 mg **4** (t_R=25.4 min) were obtained. The impure fraction collected at t_R=13.7 min was further purified on a Phenomenex Kinetex Biphenyl HPLC column (10 × 250 mm, 5 µm; 5 mL/min; UV 220 nm) using the following gradient: 0-5 min, 5% H_2O with 0.1% formic acid (A) and 5% acetonitrile with 0.1% formic acid (B); 5-8 min, 5% B to 43% B; 8-31 min, 43% to 44% B; 31-33 min, 44% to 95% B; 33-34 min, 95% to 5% B; 34-36 min, 5% B. 1.21 mg of **8** (t_R=22.0 min) was obtained. The impure fraction collected from the C_{18} column at t_R=16.8 min was further purified on a Waters XSelect CSH Fluro-Phenyl HPLC column (10 × 250 mm, 5 µm; 5 mL/min; UV 220 nm) using the following gradient: 0-5 min, 5% H_2O with 0.1% formic acid (A) and 5% acetonitrile with 0.1% formic acid (B); 5-8 min, 5% B to 45% B; 8-

31 min, 45% to 45.8% B; 31-33 min, 45.8% to 95% B; 33-34 min, 95% to 5% B; 34-36 min, 5% B. 0.4 mg of **7** (t_R=21.0 min) was obtained. The impure fraction collected from the C_{18} column at t_R=17.2 min was further purified on a Phenomenex Kinetex Biphenyl HPLC column (10 × 250 mm, 5 µm; 5 mL/min; UV 220 nm) using the following gradient: 0-5 min, 5% H_2O with 0.1% formic acid (A) and 5% acetonitrile with 0.1% formic acid (B); 5-8 min, 5% B to 50% B; 8-31 min, 50% to 50.4% B; 31-33 min, 50.4% to 95% B; 33-34 min, 95% to 5% B; 34-36 min, 5% B. 0.7 mg of **5** (t_R=23.6 min) was obtained.

The fraction containing ajudazol F (**6**), I (**9**), and J (**10**) was separated by semi-preparative HPLC (Waters Xbridge Peptide BEH C_{18}, 5 µm, 250 × 10 mm; 5 mL/min; UV 220 nm). The following gradient was used: 0-5 min, 5% H_2O with 0.1% formic acid (A) and 5% acetonitrile with 0.1% formic acid (B); 5-8 min, 5% B to 42% B; 8-31 min, 42% to 52% B; 31-33 min, 52% to 95% B; 33-34 min, 95% to 5% B; 34-36 min, 5% B. The **9**-containing fraction was further purified on a Phenomenex Synergi™ Fusion-RP (10 × 250 mm, 4 µm; 5 mL/min; UV 220 nm) and eluted with the following gradient to afford **9** (0.8 mg, t_R=23.0 min): 0-5 min, 5% H_2O with 0.1% formic acid (A) and 5% acetonitrile with 0.1% formic acid (B); 5-8 min, 5% B to 40% B; 8-31 min, 40% to 40.4% B; 31-33 min, 40.4% to 95% B; 33-34 min, 95% to 5% B; 34-36 min, 5% B. The fraction containing **10** was further purified using a Phenomenex Kinetex Biphenyl HPLC column (10 × 250 mm, 5 µm; 5 mL/min; UV 220 nm) and eluted with the following gradient to afford **10** (0.9 mg, t_R=20.9 min): 0-5 min, 5% H_2O with 0.1% formic acid (A) and 5% acetonitrile with 0.1% formic acid (B); 5-8 min, 5% B to 40% B; 8-31 min, 40% to 40.7% B; 31-33 min, 46.4% to 95% B; 33-34 min, 95% to 5% B; 34-36 min, 5% B. The **6**-containing fraction was further purified on a Phenomenex Synergi™ Fusion-RP (10 × 250 mm, 4 µm; 5 mL/min; UV 220 nm) and eluted with the following gradient to afford **6** (2.4 mg, t_R=20.4 min): 0-5 min, 5% H_2O with 0.1% formic acid (A) and 5% acetonitrile with 0.1% formic acid (B); 5-8 min, 5% B to 45% B; 8-31 min, 45% to 46.4% B; 31-33 min, 46.4% to 95% B; 33-34 min, 95% to 5% B; 34-36 min, 5% B.

Structure elucidation. *Ajudazol C (**3**)*. Colourless solid; *M* = 576.32; UV (82.2% ACN) λ_{max} 224 nm; $[\alpha]_D^{23} = +0.2$ (*c* = 0.007, in methanol); for ^{1}H and ^{13}C NMR data see **Table S3. 8-Table S3. 10**; HR-ESIMS *m/z* 577.3270 $[M+H]^+$ (calc. for $C_{34}H_{45}N_2O_6$ 577.3272).

*Ajudazol D (**4**)*. Colourless solid; *M* = 574.30; UV (83.3% ACN) λ_{max} 224 nm; $[\alpha]_D^{23} = +2.4$ (*c* = 0.018, in methanol); for ^{1}H and ^{13}C NMR data see **Table S3. 11**; HR-ESIMS *m/z* 575.3128 $[M+H]^+$ (calc. for $C_{34}H_{43}N_2O_6$ 575.3116).

*Ajudazol E (**5**)*. Colourless solid; *M* = 590.30; UV (63.5% ACN) λ_{max} 222 nm; for ^{1}H and ^{13}C NMR data see **Table S3. 12**; HR-ESIMS *m/z* 591.3077 $[M+H]^+$ (calc. for $C_{34}H_{42}N_2O_7$ 591.3065).

*Ajudazol F (**6**)*. Colourless solid; *M* = 736.36; UV (60.2% ACN) λ_{max} 220 322 nm; $[\alpha]_D^{23} = +58.6$ (**c** = 0.006, in methanol); for ^{1}H and ^{13}C NMR data see **Table S3. 13**; HR-ESIMS *m/z* 737.3648 $[M+H]^+$ (calc. for $C_{40}H_{52}N_2O_{11}$ 737.3644).

*Ajudazol G (**7**)*. Colourless solid; *M* = 592.31; UV (62.4% ACN) λ_{max} 218 322 nm; for ^{1}H and ^{13}C NMR data see **Table S3. 14**; HR-ESIMS *m/z* 593.3228 $[M+H]^+$ (calc. for $C_{34}H_{44}N_2O_7$ 593.3221).

*Ajudazol H (**8**)*. Colourless solid; *M* = 608.31; UV (56.8% ACN) λ_{max} 220 nm; $[\alpha]_D^{23} = +158.2$ (*c* = 0.005, in methanol); for ^{1}H and ^{13}C NMR data see **Table S3. 15**; HR-ESIMS *m/z* 609.3171 $[M+H]^+$ (calc. for $C_{34}H_{44}N_2O_8$ 609.3170).

*Ajudazol I (**9**)*. Colourless solid; *M* = 770.36; UV (51.4% ACN) λ_{max} 216 322 nm; $[\alpha]_D^{23} = +4.1$ (*c* = 0.007, in methanol); for ^{1}H and ^{13}C NMR data see **Table S3. 16**; HR-ESIMS *m/z* 771.3699 $[M+H]^+$ (calc. for $C_{40}H_{54}N_2O_{13}$ 771.3699).

*Ajudazol J (**10**)*. Colourless solid; *M* = 740.35; UV (53.3% ACN) λ_{max} 216 322 nm; for ^{1}H and ^{13}C NMR data see **Table S3. 17**; HR-ESIMS *m/z* 741.3593 $[M+H]^+$ (calc. for $C_{39}H_{52}N_2O_{12}$ 741.3593).

Feeding experiment. The feeding studies of *Cystobacter* sp. SBCb004 were performed in 50 mL M medium. In 1 mL water was dissolved 5.3 mg d_6-3,3-dimethylacrylic acid were, and the solution was sterilised by filtration. Equal aliquots (250 μL) were added to the culture after 24, 36, 48, and 60 h to a final concentration of 1 mM. A control experiment without isotopically labelled precursor was done in parallel. To the culture was added 2% Amberlite XAD-16 resin after 72 h. The culture was harvested after 90 h by centrifugation. Cells including XAD were extracted with 50 mL acetone. The solvents were evaporated, and the extract was resolved in MeOH and analysed by UPLC-MS.

Assignment of the absolute configuration of glucose and arabinose. Hydrolysis of ca. 0.3 mg **6**, **9,** or **10** was conducted with 0.25 mL 1 M HCl at 60 °C for 4 hours, respectively. The reaction mixture volume was reduced using a Speedvac concentrator (Eppendorf), followed by *in vacuo* drying. The residue was dissolved in 0.3 mL (*S*)-2-butanol and 5 μL acetyl chloride, and stirred at 80 °C for 8 hours. After the reaction mixture was quenched with 0.5 mL saturated aqueous $NaHCO_3$, the solution was extracted two times with 0.5 mL ethyl acetate. After ethyl acetate layer was dried first by a Speedvac concentrator and then overnight *in vacuo*, silylation was conducted by

dissolving the residue in 40 μL 1-(trimethylsilyl)imidazole and 160 μL pyridine and stirring the solution at 60 °C for 15 min. The mixture was dried by a Speedvac concentrator and extracted two times with 0.4 mL ethyl acetate and water (1:1 *v*/*v*). The organic layer was dried over $MgSO_4$, spun down, and analysed by GC-MS (Agilent) using a 6890N gas chromatograph equipped with a 5973 mass detector, a 7683B injector, a HP-1ms column (Agilent, 100% dimethylpolysiloxane, 0.1 µm film, 30 m length, 0.25 mm ID), and helium as the carrier gas. General settings were as follows: inlet temperature of 275 °C; GC-MS transfer line of 280 °C, ion source of 230 °C, and quadrupole of 150 °C; MS detector operated in the scan mode (40-700 *m*/*z*). GC was performed with 1 mL/min carrier gas flow, and 1 μL of the sample was injected in the split mode (1:10). The column oven temperature protocol was as follows: held at 50 °C for 2.5 min, increased from 50 to 180 °C with a 5 °C/min rate, held at 180 °C for 10 min, and decreased to 50 °C with a rate of 30 °C/min, which was held for 5 min. Authentic standards of D- and L-arabinose with (*S*)-2-butanol as well as *D*-glucose with (*S*)-2-butanol and (*R*)-2-butanol were treated in the same manner. Analytes were detected as trimetylsilylated (*S*)- or (*R*)-2-butyl glycosides. Since the retention times of enantiomers (*S*)-2-butyl L-glucoside and (*R*)-2-butyl D-glucoside are the same, the authentic D-glucose standard was derivatised with (*S*)- and (*R*)-2-butanol. The glycosides of **6** and **9** exhibited the same retention time as the derivative of D-glucose (t_R=37.50 min), and the glycoside of **10** exhibited the same retention time as the derivative of D-arabinose (t_R=28.25 min).

Antimicrobial activity test. All microorganisms were obtained from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung für Mikroorganismen und Zellkulturen, DSMZ). For microdilution assays, overnight cultures of each strain were diluted in the growth medium to approximately 10^6 cfu/mL and added to sterile 96-well plates. The cell suspension was treated with the respective compounds in a serial dilution. After cells were incubated on a microplate shaker (750 rpm, 37 °C) for 16 h, the growth inhibition was assessed by visual inspection.

Cytotoxicity assay. Cell lines were obtained from the DSMZ and were cultured under conditions as recommended by the depositor. Cells were seeded at 6×10^3 cells per well of the 96-well plates in 180 μL complete medium and treated with compounds in serial dilution after 2 h of equilibration. Both compounds and the internal solvent control was tested in duplicate. After 5 d of incubation, 20 μL of 5 mg/mL MTT (thiazolyl blue tetrazolium bromide) in PBS (phosphate-buffered saline) was added per well, and samples were further incubated for 2 h at 37 °C. The

medium was then discarded, and cells were washed with 100 μL PBS before 100 μL 2-propanol/10 N HCl (250:1) was added in order to dissolve formazan granules. The absorbance at 570 nm was measured using a microplate reader, and cell viability was expressed as percentage relative to the respective methanol control. The half-maximal inhibitory concentration (IC_{50}) values were determined by sigmoidal curve fitting by plotting % relative growth versus applied concentrations using 0% (full inhibition) and 100% (no inhibition) as borders (**Figure S3. 3**).

***In silico* analysis of the genome of *Cystobacter* sp. SBCb004 and the sequence of ajudazol biosynthetic genes in *Cystobacter* sp. SBCb004.** The genome of *Cystobacter* sp. SBCb004 was provided by 454 technology.[34,35] The prediction of all the BGCs were done in antiSMASH5.0.[8] The gaps in genes *ajuA*, *ajuC*, *ajuE*, *ajuG1* and *ajuL,* and the gap between *orf31* and *orf32* were refilled by PCR and Sanger sequencing with the primers 9104-gene-gap-F and 9104-gene-gap-R, as shown in **Table S3. 2** ("gene" refers to the gene with a gap). The sequence of ajudazols C-J BGC (ON420332) and the list of AjuI and SorF homologs (ON420327-ON420331) (**Table S3. 6** and **Table S3. 7**) in *Cystobacter* sp. SBCb004 genome have been deposited in GenBank.

Acknowledgements

Research in R.M.'s laboratory was partially funded by the Bundesministerium für Bildung und Forschung and Deutsche Forschungsgemeinschaft. H.Z. is grateful to the China Scholarship Council for a Ph.D. scholarship. We thank our colleagues Jennifer Hermann, Stefanie Neuber, and Alexandra Amann for performing and interpreting the bioactivity evaluation tests.

3.5. References

(1) Dawid, W. Biology and global distribution of myxobacteria in soils. *FEMS Microbiol. Rev.* **2000**, *24* (4), 403–427. DOI: 10.1111/j.1574-6976.2000.tb00548.x.

(2) Kaplan, H.B.; Whitworth, D., Eds. *Multicellularity and differentiation among the myxobacteria and their neighbors*, 2007.

(3) Herrmann, J.; Fayad, A. A.; Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **2017**, *34* (2), 135–160. DOI: 10.1039/C6NP00106H.

(4) Weissman, K. J.; Müller, R. Myxobacterial secondary metabolites: bioactivities and modes-of-action. *Nat. Prod. Rep.* **2010**, *27* (9), 1276–1295. DOI: 10.1039/c001260m.

(5) Wenzel, S. C.; Müller, R. The impact of genomics on the exploitation of the myxobacterial secondary metabolome. *Nat. Prod. Rep.* **2009**, *26* (11), 1385–1407. DOI: 10.1039/b817073h.

(6) Hertweck, C. The Biosynthetic Logic of Polyketide Diversity. *Angew. Chem. Int. Ed. Engl.* **2009**, *48* (26), 4688–4716. DOI: 10.1002/anie.200806121.

(7) Süssmuth, R. D.; Mainz, A. Nonribosomal peptide synthesis - Principles and prospects. *Angew. Chem. Int. Ed.* **2017**, *56* (14), 3770–3821. DOI: 10.1002/anie.201609079.

(8) Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **2019** (47), W81-W87. DOI: 10.1093/nar/gkz310.

(9) Chai, Y.; Pistorius, D.; Ullrich, A.; Weissman, K. J.; Kazmaier, U.; Müller, R. Discovery of 23 natural tubulysins from *Angiococcus disciformis* An d48 and *Cystobacter* SBCb004. *Chem. Biol.* **2010**, *17* (3), 296–309. DOI: 10.1016/j.chembiol.2010.01.016.

(10) 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* (2), 185–195. DOI: 10.1016/s0167-4781(99)00041-x.

(11) 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* (5), 1121–1133. DOI: 10.1021/acssynbio.9b00023.

(12) Ohnishi, Y.; Ishikawa, J.; Hara, H.; Suzuki, H.; Ikenoya, M.; Ikeda, H.; Yamashita, A.; Hattori, M.; Horinouchi, S. Genome sequence of the streptomycin-producing microorganism

Streptomyces griseus IFO 13350. *Journal of bacteriology* **2008**, *190* (11), 4050–4060. DOI: 10.1128/JB.00204-08.

(13) Lorenzen, W.; Ahrendt, T.; Bozhüyük, K. A. J.; Bode, H. B. A multifunctional enzyme is involved in bacterial ether lipid biosynthesis. *Nat Chem Biol* **2014**, *10* (6), 425–427. DOI: 10.1038/nchembio.1526. Published Online: May. 11, 2014.

(14) Botella, J. A.; Murillo, F. J.; Ruiz-Vázquez, R. A cluster of structural and regulatory genes for light-induced carotenogenesis in *Myxococcus xanthus*. *Eur. J. Biochem.* **1995**, *233* (1), 238–248. DOI: 10.1111/j.1432-1033.1995.238_1.x.

(15) Giglio, S.; Jiang, J.; Saint, C. P.; Cane, D. E.; Monis, P. T. Isolation and characterization of the gene associated with geosmin production in cyanobacteria. *Environ. Sci. Technol.* **2008**, *42* (21), 8027–8032. DOI: 10.1021/es801465w.

(16) Buntin, K.; Rachid, S.; Scharfe, M.; Blöcker, H.; Weissman, K. J.; Müller, R. Production of the antifungal isochromanone ajudazols A and B in *Chondromyces crocatus* Cm c5: biosynthetic machinery and cytochrome P450 modifications. *Angew. Chem. Int. Ed. Engl.* **2008**, *47* (24), 4595–4599. DOI: 10.1002/anie.200705569.

(17) *Dictionary of Natural Products*. http://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml.

(18) 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* (6), 768–783. DOI: 10.1039/c3np70127a.

(19) Jansen, R.; Kunze, B.; Reichenbach, H.; Höfle, G. The Ajudazols A and B, Novel Isochromanones from Chondromyces crocatus (Myxobacteria): Isolation and Structure Elucidation. *Eur. J. Org. Chem.* **2002**, *2002* (5), 917–921. DOI: 10.1002/1099-0690(200203)2002:5<917:AID-EJOC917>3.0.CO;2-Z.

(20) Essig, S.; Schmalzbauer, B.; Bretzke, S.; Scherer, O.; Koeberle, A.; Werz, O.; Müller, R.; Menche, D. Predictive bioinformatic assignment of methyl-bearing stereocenters, total synthesis and an additional molecular target of ajudazol B. *J. Org. Chem.* **2016**, *81* (4), 1333–1357. DOI: 10.1021/acs.joc.5b02844.

(21) Fischer, G. Chemical aspects of peptide bond isomerisation. *Chem Soc.Rev.* **2000**, *29* (2), 119–127. DOI: 10.1039/A803742F.

(22) Popoff, A. Exploiting the biosynthetic potential of myxobacteria for natural product discovery. Doctoral Thesis, Saarland University, Saarbrücken, Germany, 2020.

(23) Bock, K.; Pedersen, C. Carbon-13 Nuclear Magnetic Resonance Spectroscopy of Monosaccharides. In *Advances in Carbohydrate Chemistry and Biochemistry*; pp 27–66. DOI: 10.1016/S0065-2318(08)60055-4.

(24) Brito-Arias, M. Nuclear Magnetic Resonance of Glycosides. In *Synthesis and Characterization of Glycosides,* 2nd ed.; Brito-Arias, M., Ed.; Springer International Publishing, 2016; pp 369–389. DOI: 10.1007/978-3-319-32310-7_8.

(25) Li, Y.; Luxenburger, E.; Müller, R. An alternative isovaleryl CoA biosynthetic pathway involving a previously unknown 3-methylglutaconyl CoA decarboxylase. *Angew. Chem. Int. Ed. Engl.* **2012**, *52* (4), 1304–1308. DOI: 10.1002/anie.201207984.

(26) Silakowski, B.; Schairer, H. U.; Ehret, H.; Kunze, B.; Weinig, S.; Nordsiek, G.; Brandt, P.; Blöcker, H.; Höfle, G.; Beyer, S.; Müller, R. New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. *J. Biol. Chem.* **1999**, *274* (52), 37391–37399. DOI: 10.1074/jbc.274.52.37391.

(27) Frank, B.; Wenzel, S. C.; Bode, H. B.; Scharfe, M.; Blöcker, H.; Müller, R. From genetic diversity to metabolic unity: studies on the biosynthesis of aurafurones and aurafuron-like structures in myxobacteria and streptomycetes. *J. Mol. Biol.* **2007**, *374* (1), 24–38. DOI: 10.1016/j.jmb.2007.09.015.

(28) Kopp, M.; Rupprath, C.; Irschik, H.; Bechthold, A.; Elling, L.; Müller, R. SorF, a glycosyltransferase with promiscuous donor substrate specificity in vitro. *ChemBioChem* **2007**, *8* (7), 813–819. DOI: 10.1002/cbic.200700024.

(29) Surup, F.; Viehrig, K.; Mohr, K. I.; Herrmann, J.; Jansen, R.; Müller, R. Disciformycins A and B: 12-membered macrolide glycoside antibiotics from the myxobacterium *Pyxidicoccus fallax* active against multiresistant staphylococci. *Angew. Chem. Int. Ed. Engl.* **2014**, *49* (53), 13588–13591. DOI: 10.1002/anie.201406973.

(30) 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* (9), 2414–2417. DOI: 10.1021/ol500779s.

(31) Kunze, B.; Jansen, R.; Hofle, G.; Reichenbach, H. Ajudazols, new inhibitors of the mitochondrial electron transport from Chondromyces crocatus. Production, antimicrobial activity and mechanism of action. *J. Antibiot.* **2004**, *57* (2), 151–155.

(32) Wright, G. D.; Thompson, P. R. *Aminoglycoside phosphotransferases: proteins, structure, and mechanism*; Front Biosci, 1999. https://pubmed.ncbi.nlm.nih.gov/9872733/.

(33) Deuschle, U.; Kammerer, W.; Gentz, R.; Bujard, H. *Promoters of Escherichia coli: a hierarchy of in vivo strength indicates alternate structures*, 1986.

(34) Margulies, M.; Egholm, M.; Altman, W. E.; Attiya, S.; Bader, J. S.; Bemben, L. A.; Berka, J.; Braverman, M. S.; Chen, Y. J.; Chen, Z.; Dewell, S. B.; Du, L.; Fierro, J. M.; Gomes, X. V.; Godwin, B. C.; He, W.; Helgesen, S.; Ho, C. H.; Irzyk, G. P.; Jando, S. C.; Alenquer, M. L.; Jarvie, T. P.; Jirage, K. B.; Kim, J. B.; Knight, J. R.; Lanza, J. R.; Leamon, J. H.; Lefkowitz, S. M.; Lei, M.; Li, J.; Lohman, K. L.; Lu, H.; Makhijani, V. B.; McDade, K. E.; McKenna, M. P.; Myers, E. W.; Nickerson, E.; Nobile, J. R.; Plant, R.; Puc, B. P.; Ronan, M. T.; Roth, G. T.; Sarkis, G. J.; Simons, J. F.; Simpson, J. W.; Srinivasan, M.; Tartaro, K. R.; Tomasz, A.; Vogt, K. A.; Volkmer, G. A.; Wang, S. H.; Wang, Y.; Weiner, M. P.; Yu, P.; Begley, R. F.; Rothberg, J. M. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **2005.** DOI: 10.1038/nature03959.

(35) Rothberg, J. M.; Leamon, J. H. The development and impact of 454 sequencing. *Nat. Biotechnol.* **2008**, *26* (10), 1117–1124. DOI: 10.1038/nbt1485.

Supporting Information

Expanding the Ajudazol Cytotoxin Scaffold: Insights from Genome Mining, Biosynthetic Investigations, and Novel Derivatives

Previously published in:

Hu Zeng†, Joy Birkelbach†, Judith Hoffmann, Alexander Popoff, Carsten Volz, and Rolf Müller

Journal of Natural Products **2022** 85 (11), 2610-2619

DOI: 10.1021/acs.jnatprod.2c00637

Affiliation

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

S 3.1 Analysis of the Biosynthetic Gene Clusters in SBCb004

Table S3. 1. *Predicted Biosynthetic Gene Clusters in* Cystobacter *sp. SBCb004.*

BGC region	Type	Known compound
1	PKS	
2	PKS	Alkylresorcinol
3	RiPPs	
4	terpene	
5	terpene	
6	NRPS	
7	terpene	
8	terpene	
9	PKS	
10	PKS	
11	NRPS	
12	PKS	
13	NRPS	
14	amglyccycl	
15	PKS	
16	RiPPs	
17	NRPS	
18	phenazine	
19	NRPS	Argyrin
20	RiPPs	
21	PKS-NRPS	
22	RiPPs	
23	RiPPs	
24	RiPPs	
25	NRPS	
26	PKS	Stigmatellin
27	RiPPs	
28	RiPPs	
29	other	VEPE/AEPE/TG-1
30	NRPS	
31	PKS	

32	terpene	Carotenoids
33	terpene	
34	PKS-NRPS	Tubulysin
35	terpene	
36	PKS	
37	PKS	
38	other	
39	PKS	
40	RiPPs	
41	RiPPs	
42	PKS-NRPS	ajudazols C-J
43	PKS-NRPS	
44	PKS-NRPS	
45	RiPPs	
46	other	
47	RiPPs	
48	NRPS	
49	terpene	Geosmin

PKS: polyketide synthase. NRPS: non-ribosomal peptide synthetase. PKS-NRPS: hybrid polyketide synthase-non-ribosomal peptide synthetase. RiPPs: Ribosomally synthesized and post-translationally modified peptides.

Table S3. 2. *Oligonucleotide Primers Used in this Study.*

Primer	Oligonucleotide in 5'-3' direction (introduced restriction site)
9101-sc-ajuD-F	GCGCCTAGGGTAGGGCTGGGAGCAGGTGG (XmaJI)
9101-sc-ajuD-R	CCTCAATTGGGCCCAGGTCACCGCTCTTG (MfeI)
9101-sc-ajuL-F	GCGCCTAGGCCCTTCTAGGAGCTGGACGC (XmaJI)
9101-sc- ajuL-R	CCTCAATTGGCTGGGCTAGTTCCTCCTCG (MfeI)
9101-sc-orf43-F	CCAGTCTAGCTATTAATAGGCCTAGGGTCTCTAGCAGCTCCACTC
9101-sc-orf43-R	GGGCTGGCTTAAAGTCGACTCAATTGCTTAGGCATCCTTCCCCAAT
9101-sc-orf42-F	CCAGTCTAGCTATTAATAGGCCTAGGTGACATCACCGGCGCCTCC
9101-sc-orf42-R	GGGCTGGCTTAAAGTCGACTCAATTGCTACACCACGTCGTTGAGGG
9101-sc-ajuJ-F	CCAGTCTAGCTATTAATAGGCCTAGGTAGTTCGTCTTCGTGAGCGC
9101-sc-ajuJ-R	GGGCTGGCTTAAAGTCGACTCAATTGTCCGGCTAGATGTCCTCGC

9101-sc-orf32-F	CCAGTCTAGCTATTAATAGGCCTAGGAGGCCACTAATCCTCTGTGA
9101-sc-orf32-R	GGGCTGGCTTAAAGTCGACTCAATTGCTACGCGGGACGGAAGATGT
9101-sc-orf28-F1	CCAGTCTAGCTATTAATAGGCCTAGGGCCTGGTATGATGCAGGTGA
9101-sc-orf28-R1	GGGCTGGCTTAAAGTCGACTCAATTGAACCTAAGCTGGGAATGTCC
9101-sc- orf28-F2	TATTCCTCCGTAGCAGAAAG
9101-sc- orf28-R2	CTTTCTGCTACGGAGGAATA
9101-sc-orf15-F	CCAGTCTAGCTATTAATAGGCCTAGGCTAAGCGGGACTTCACCGTC
9101-sc-orf15-R	GGGCTGGCTTAAAGTCGACTCAATTGGGAAGTCCTACCAGTTCAGC
9101-sc-orf2-F1	CCAGTCTAGCTATTAATAGGCCTAGGGCATCAGTCGCCTCAAGCT
9101-sc-orf2-R1	GGGCTGGCTTAAAGTCGACTCAATTGGGCCTACGTCGGGTTGTAGA
9101-sc- orf2-F2	GTGATGGCGTAACTGGTGTT
9101-sc- orf2-R2	AACACCAGTTACGCCATCAC
9101-sc-ctg19_169-F	CCAGTCTAGCTATTAATAGGCCTAGGCGGTAGCGGATGGATTCCTT
9101-sc-ctg19_169-R	GGGCTGGCTTAAAGTCGACTCAATTGGGAGACGTCTTCACCATCCG
9101-sc-ajuI-F	CCAGTCTAGCTATTAATAGGCCTAGGCAACGTGTAGTCCGTCAGCT
9101-sc-ajuI-R	GGGCTGGCTTAAAGTCGACTCAATTGATGGTCTTCATCAGCTCGCC
9101-sc-ctg1_130-F	CCAGTCTAGCTATTAATAGGCCTAGGCAGATCGAGCAGATGACCGT
9101-sc-ctg1_130-R	GGGCTGGCTTAAAGTCGACTCAATTGTCGGCGGATTATGAGGAGTG
9101-sc-ctg10_386-F	CCAGTCTAGCTATTAATAGGCCTAGGGTGACGTCCAACCCTTCGT
9101-sc-ctg10_386-R	GGGCTGGCTTAAAGTCGACTCAATTGGTTGTCATCCACGTACACGC
9101-sc-ajuM-F	CCAGTCTAGCTATTAATAGGCCTAGGAATTGATCGAACGCGTGTGG
9101-sc-ajuM-R	GGGCTGGCTTAAAGTCGACTCAATTGCGTTCGGTATCCCCTTCGG
ajuD-CK-F	CATCCCGGACAAGTCACTGC
ajuL-CK-F	CAGGAACCCGTGTCCCTTC
orf43-CK-F	CACCCTCTTCTCCTATGCGC
orf42-CK-F	GATTGGGGAAGGATGCCGAA
ajuJ-CK-F	AGATGTCGCTGTACATGGCC
orf32-CK-F	GAAACTGTAACGAGCTGCGC
orf28-CK-F	TACATGCAGCAGCCCAATGA
orf15-CK-F	GTTCGATTCAGCGTGGATGC
orf2-CK-F	GTGGCTCACGAAATCCAACG
ctg19_169-CK-F	GGCCAACACCACCTTCATCT
ajuI-CK-F	CGAGCATGATGTGGCGATTG
ctg1_130-CK-F	TCAGCTCACCCAACCTGTTC

ctg10_386-CK-F	GTCGGTGTTCTGGATGGAGC
ajuM-CK-F	CCTTCTCTCGCATCGGTCC
CN-check-R	ACCTCTGACACATGCAGCTC
seq-pblueCN-R	ACCTCTGACACATGCAGCTC
9101-ajuG1-gap-F	TCCATGATCCGGGAGATGGA
9101-ajuG1-gap-R	GGCGAGACAGTCATCTCCG
9101-ajuE-gap-F	CGTTCTTGAGGATGGGGTGG
9101-ajuE-gap-R	TTTCGACAACAGCCCCTTCT
9101-ajuC-gap-F	GTGACAGTCACCGTCGAGTA
9101-ajuC-gap-R	ATATCGATTGGGACCGGATGC
9101-ajuA-gap-F	ATCGAAACGTGACTCGCTCC
9101-ajuA-gap-R	CCAGGAGGCGAAGGTGTC
9101-ajuL-gap-F	GTGCGATACCTGCAACGAAC
9101-ajuL-gap-R	GCGGAGTACCTCGCGAAG
9101-orf31-32-gap-F	GCAGCCCATCACAGAGGATT
9101-orf31-32-gap-R	GCACTGTCCTTGCTCGTCTA

The bold letters indicate the overlap sequence to the vector pbluekan.

Table S3. 3. *Plasmids Used in this Study.*

Plasmids	Description/derivation	Reference
pbluekan	ori (pMB1), bla, lacZ, neo	Unpublished
pbluekan-ajuD-HR	pbluekan derivative bearing internal homologous region of gene ajuD	This study
pbluekan-ajuL-HR	pbluekan derivative with gene ajuL homologous region	This study
pbluekan-orf42-HR	pbluekan derivative with gene orf42 homologous region	This study
pbluekan-orf43-HR	pbluekan derivative with gene orf41 homologous region	This study
pbluekan-ajuJ-HR	pbluekan derivative with gene ajuJ homologous region	This study
pbluekan-orf32-HR	pbluekan derivative with gene orf31 homologous region	This study

pbluekan-orf28-HR	pbluekan derivative with gene orf27 homologous region	This study
pbluekan-orf15-HR	pbluekan derivative with gene orf15 homologous region	This study
pbluekan-orf2-HR	pbluekan derivative with gene orf2 homologous region	This study
pbluekan-ctg19_169-HR	pbluekan derivative with gene ctg19_169 homologous region	This study
pbluekan-ajuI-HR	pbluekan derivative with gene ajuI homologous region	This study
pbluekan-ctg1_130-HR	pbluekan derivative with gene ctg1_130 homologous region	This study
pbluekan-ctg10_386-HR	pbluekan derivative with gene ctg10_386 homologous region	This study
pbluekan-ajuM-HR	pbluekan derivative with gene ajuM homologous region	This study

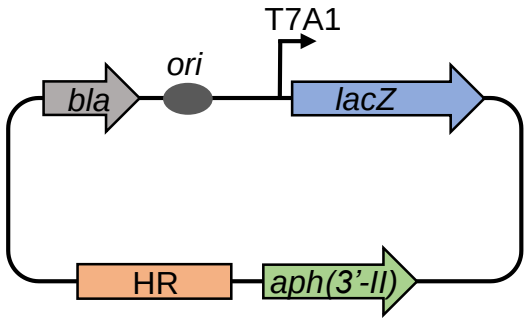


***Figure S3. 1*. Schematic representation of plasmid pbluekan-HR used to perform insertional gene inactivation.** aph(3'-II)*: kanamycin resistance gene.* bla*: β-lactam resistance cassette. HR: homologous region amplified from target gene.* ori*: origin of replication of* E. coli *plasmid pMB1. The angled arrow indicates the T7A1 transcriptional promoter.* lacZ*: β-galactosidase gene.*

Table S3. 4. *Predicted Proteins Encoded in the Ajudazol BGC of Strain* Cystobacter *sp. SBCb004.*

Protein	Size aa*	Proposed function of the similar protein	Sequence similarity to source	Similarity / identity aa* level	Accession number of the similar protein
Orf1	515	zinc-ribbon domain-containing protein	*Vitiosangium* sp. GDMCC 1.1324	77 / 67 %	WP_108075970.1
Orf2	394	glycosyltransferase family 2 protein	*Archangium* sp. Cb G35	96 / 94 %	WP_073567095.1
Orf3	410	GTP-binding protein	*Pyxidicoccus fallax*	85 / 79 %	WP_169349557.1
Orf4	55	MULTISPECIES: 50S ribosomal protein L33	*Archangium*	98 / 87 %	WP_043409257.1
Orf5	386	hypothetical protein	*Cystobacter ferrugineus*	71 / 60 %	WP_143177574.1
Orf6	273	GNAT family N-acetyltransferase	*Cystobacter ferrugineus*	88 / 82 %	WP_071905383.1
Orf7	251	C40 family peptidase	*Vitiosangium* sp. GDMCC 1.1324	94 / 86 %	WP_108075973.1
Orf8	563	HEAT repeat domain-containing protein	*Vitiosangium* sp. GDMCC 1.1324	95 / 89 %	WP_108075974.1
Orf9	676	ATP-dependent helicase	*Vitiosangium* sp. GDMCC 1.1324	96 / 93 %	WP_108075975.1
Orf10	544	Hsp70 family protein	*Vitiosangium* sp. GDMCC 1.1324	99 / 96 %	WP_108075976.1
Orf11	142	hypothetical protein	*Vitiosangium* sp. GDMCC 1.1324	92 / 89 %	WP_108075977.1
Orf12	93	hypothetical protein	*Archangium gephyra*	96 / 93 %	WP_047857416.1
Orf13	328	TerC family protein	*Archangium* sp. Cb G35	92 / 89 %	WP_073567110.1
Orf14	299	zinc metalloprotease HtpX	*Archangium gephyra*	97 / 91 %	WP_047857418.1
Orf15	381	MBL fold metallo-hydrolase	*Vitiosangium* sp. GDMCC 1.1324	92 / 85 %	WP_108075982.1

Orf16	263	4'-phosphopantetheinyl transferase superfamily protein	*Archangium gephyra*	92 / 87 %	WP_047857428.1
Orf17	255	thioesterase	*Vitiosangium* sp. GDMCC 1.1324	91 / 87 %	WP_108075988.1
AjuL	3307	hybrid NRPS/type I PKS ($C_{7\text{-}454}$, $A_{461\text{-}980}$, $nMT_{927\text{-}1338}$, $PCP_{1403\text{-}1467}$, $KS_{1493\text{-}1918}$, $AT_{2023\text{-}2317}$, $DH_{2383\text{-}2542}$, $KR_{2858\text{-}3036}$, $ACP_{3137\text{-}3204}$)	*Chondromyces crocatus*	73 / 60 %	WP_050433124.1
Orf18	491	glycoside hydrolase family 6 protein	*Vitiosangium* sp. GDMCC 1.1324	88 / 80 %	WP_108069950.1
Orf19	303	DUF3396 domain-containing protein	*Archangium violaceum*	85 / 72 %	WP_152622499.1
Orf20	283	hypothetical protein	*Vitiosangium* sp. GDMCC 1.1324	79 / 69 %	WP_108072391.1
Orf21	401	hypothetical protein DAT35_32570	*Vitiosangium* sp. GDMCC 1.1324	73 / 62 %	PTL79551.1
Orf22	452	SWIM zinc finger family protein	*Vitiosangium* sp. GDMCC 1.1324	92 / 87 %	WP_108075989.1
Orf23	181	MarR family transcriptional regulator	*Vitiosangium* sp. GDMCC 1.1324	82 / 70 %	WP_108069654.1
Orf24	480	hypothetical protein	*Vitiosangium* sp. GDMCC 1.1324	88 / 81 %	WP_108075990.1
Orf25	362	AAA family ATPase	*Vitiosangium* sp. GDMCC 1.1324	98 / 97 %	WP_108075991.1
Orf26	754	hypothetical protein	*Vitiosangium* sp. GDMCC 1.1324	96 / 93 %	WP_108075992.1
Orf27	384	VWA domain-containing protein	*Vitiosangium* sp. GDMCC 1.1324	97 / 93 %	WP_108075993.1

Orf28	254	alpha/beta fold hydrolase	*Chondromyces fuscus*	92 / 87 %	WP_002628336.1
Orf29	198	hypothetical protein BON30_35460	*Cystobacter ferrugineus*	84 / 76 %	OJH35915.1
Orf30	117	serine/threonine protein kinase	*Archangium* sp. Cb G35	95 / 95 %	WP_073562591.1
Orf31	311	hypothetical protein BO221_22375	*Archangium* sp. Cb G35	95 / 86 %	OJT22519.1
Orf32	1482	SIR2 family protein	*Archangium* sp. Cb G35	79 / 70 %	WP_073567252.1
AjuJ	456	cytochrome P450	*Chondromyces crocatus*	77 / 59 %	WP_169796669.1
AjuA	2147	type I PKS (KS_{30-458}, $AT_{564-861}$, $DH_{923-1088}$, $ER_{1435-1741}$, $KR_{1753-1931}$, $ACP_{2031-2096}$)	*Chondromyces crocatus*	71 / 60 %	WP_050433113.1
AjuB	3380	type I PKS (KS_{37-463}, $AT_{568-862}$, $KR_{1146-1324}$, $ACP_{1430-1499}$, $KS_{1523-1951}$, $AT_{2059-2352}$, $DH_{2422-2597}$, $KR_{2970-3148}$, $ACP_{3264-3321}$)	*Chondromyces crocatus*	73 / 61 %	WP_050433114.1
AjuC	2190	type I PKS (KS_{36-465}, $AT_{570-860}$, $DH_{924-1094}$, $ER_{1475-1787}$, $KR_{1808-1987}$, $ACP_{2090-2156}$)	*Chondromyces crocatus*	73 / 61 %	WP_050433115.1
AjuD	1404	NRPS (HC_{70-503}, $A_{514-1044}$, $PCP_{1310-1374}$)	*Chondromyces crocatus*	76 / 63 %	WP_050433116.1
AjuE	2167	type I PKS (KS_{7-433}, $AT_{538-828}$, $DH_{891-1057}$, $ER_{1437-1749}$, $KR_{1771-1950}$, $ACP_{2051-2117}$)	*Chondromyces crocatus*	70 / 57 %	WP_050433117.1
AjuF	1552	type I PKS (KS_{36-462}, $AT_{568-862}$, $KR_{1154-1332}$, $ACP_{1434-1501}$)	*Chondromyces crocatus*	67 / 55 %	WP_050433118.1

AjuG1	968	type I PKS (KS_{37-464}, $AT_{569-866}$)	*Chondromyces crocatus*	74 / 62 %	WP_050433119.1
AjuG2	1936	type I PKS (ACP_{2-69}, KS_{95-524}, $AT_{625-920}$, $DH_{988-1153}$, $KR_{1516-1705}$, $ACP_{1810-1877}$)	*Chondromyces crocatus*	66 / 53 %	WP_050433119.1
AjuH	1635	type I PKS (KS_{37-464}, $AT_{569-868}$, $DH_{934-1110}$, $ACP_{1276-1342}$, $TE_{1387-1622}$)	*Chondromyces crocatus*	68 / 56 %	WP_050433120.1

*aa: amino acids

Table S3. 5. *Configurational Assignment by Bioinformatic Gene Cluster Analysis.*

Location of KR domain	Assignment of hydroxyl-bearing stereocenter	DH in module of KR	Double bond geometry	Assignment of methyl-bearing stereocenter
Module 2	D	+	*E*-configuration	
Module 3	D	+	obscured by ER	
Module 4	L	-	*Z*-configuration	
Module 5	L	+	*Z*-configuration	
Module 6	D	+	obscured by ER	(*R*)-configuration
Module 8	D	+	obscured by ER	(*R*)-configuration
Module 9	D	-		
Module 11	L	+	*Z*-configuration	

Assignments were retrieved from outcomes of the antiSMASH 5.0[1] gene cluster analysis.

Table S3. 6. *List of AjuI (Cm c5) Homologs in* Cystobacter *sp. SBCb004.*

Protein	Size aa*	Similarity / identity to AjuI (Cm c5) aa* level	Accession number
Ctg27_313 (AjuI)	459	80 / 63 %	ON420331
Ctg19_169	476	60 / 41 %	ON420327

*aa: amino acids.

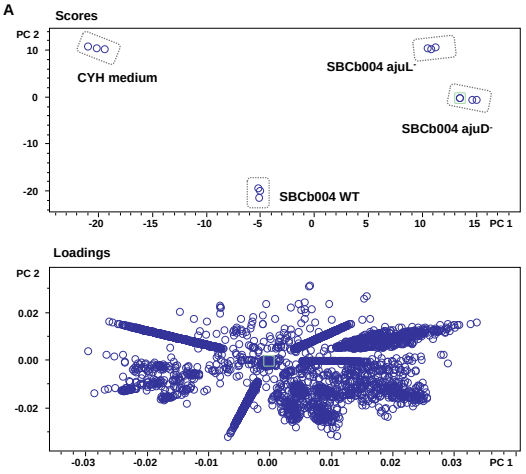
Table S3. 7. *List of SorF Homologs in* Cystobacter *sp. SBCb004.*

Protein	Size aa*	Similarity / identity to SorF aa* level	Accession number
Ctg19_179 (AjuM)	419	50 / 36 %	ON420328
Ctg1_130	433	41 / 29 %	ON420329
Ctg10_386	438	44 / 30 %	ON420330

*aa: amino acids.

S 3.2 Statistical Analysis of Extracts from SBCb004 WT and Mutants

A



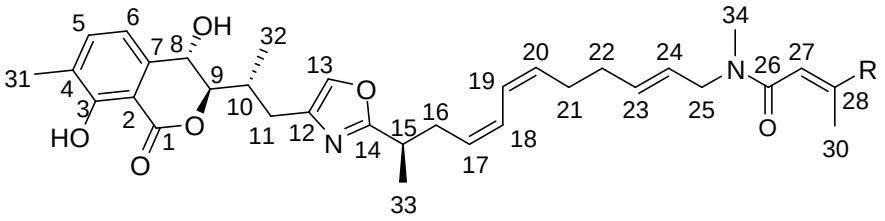
B

Molecular features	P-value	Max. intensity	WT count (3)	Media count (3)	delta ajuD count (3)	delta ajuL count (3)
10.32 min :755.374m/z	n.a.	5871898	3	0	0	0
14.11 min : 575.315m/z	n.a.	5201352	3	0	0	0
10.14 min : 741.359m/z	n.a.	4471474	3	0	0	0
12.25min : 707.354m/z	n.a.	1894118	3	0	0	0
12.02 min : 593.322m/z	n.a.	1757720	3	0	0	0
9.77 min : 771.370m/z	n.a.	1637718	3	0	0	0
13.86 min : 577.326m/z	n.a.	1046710	3	0	0	0
11.45min : 737.364m/z	n.a.	913102	3	0	0	0
12.57min : 591.306m/z	n.a.	882834	3	0	0	0
11.19min : 723.348m/z	n.a.	786580	2	0	0	0
10.84min : 609.315m/z	n.a.	267294	3	0	0	0
10.20min : 753.357m/z	n.a.	196414	2	0	0	0
5.05min : 269.211m/z	n.a.	524536	3	0	3	3
7.42min : 249.184m/z	n.a.	160746	3	0	3	3
3.19min : 513.414m/z	n.a.	141596	3	0	3	3
9.04min : 603.333m/z	n.a.	136404	3	0	3	3
8.41min : 235.205m/z	n.a.	130962	3	0	3	3
6.04min : 515.372m/z	n.a.	124768	3	0	3	3

Figure S3. 2. PCA and T-test models. *(A) PCA for SBCb004 ajuL- vs. SBCb004 ajuD- vs. SBCb004 WT vs. CYH medium. (B) T-test for datasets from A. n.a.: not available.*

S 3.3 Structure Elucidation

Table S3. 8. *^{1}H-NMR Spectroscopic Data for Ajudazol B[2] (**2**) vs C (**3**).*



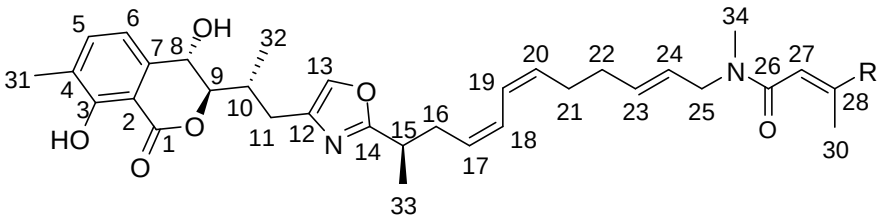
ajudazol B (**2**) $R=OCH_3$ ajudazol C (**3**) $R=CH_3$

Position	Chemical shift (δ_H, in ppm), Ajudazol B[2] (acetone-d_6, 600 MHz)[a]	Chemical shift (δ_H, in ppm), Ajudazol C (acetone-d_6, 700 MHz)[a]	Δ ($\delta_{Ajudazol\ B}$ - $\delta_{Ajudazol\ C}$, in ppm)
1	-	-	-
2	-	-	-
3	-	-	-
3-OH	11.28, s	11.29, s	-0.01
4	-	-	-
5	7.47, dd (7.6, 1.3)	7.47, dd (7.6, 0.7)	0.00
6	7.08, dd (7.4, 1.0)	7.07, d (7.6)	0.01
7	-	-	-
8	4.96, d (8.5)	4.97, d (8.3)	0.00
9	4.43, dd (8.3, 4.1)	4.43, dd (8.3, 4.3)	0.00
10	2.47, m	2.45, m	0.02
11a	2.89, ddd (14.5, 4.4, 1.2)	2.89, ddd (14.8, 4.9, 1.3)	0.00
11b	2.49, m	2.50, m	-0.01
12	-	-	-
13	7.61, d (1.3)	7.61, s	0.00
14	-	-	-
15	3.03, tq (7.0, 7.0)	3.03, m	0.00
16a	2.62, dtd (14.4, 7.3, 1.2)	2.61, dtd (14.5, 7.3, 1.4)	0.01
16b	2.49, m	2.49, m	0.00
17	5.40, m	5.40, m	0.00
18	6.29, ddd (11.9, 10.7, 1.3)	6.29, m	0.00
19	6.24, ddd (11.6, 10.4, 1.5)	6.24, m	0.00

20	5.45, m	5.44, m	0.01
21	2.24, m	2.24, m	0.00
22	2.11, m	2.12, m	-0.01
23	5.60, dtd (15.2, 6.4, 1.5)	5.59, m	0.01
24	5.47, m	5.43, m	0.04
25	3.92, d (5.8)	3.90/3.88[b],d (6.03)	0.02/-
26	-	-	-
27	5.33, br s	5.89, br d (12.7)	-0.56[c]
28	-	-	-
29	3.60, br s	1.81, br d (11.4)	1.79[c]
30	2.13, s	1.89, br d (5.9)	0.24[c]
31	2.22, s	2.22, s	0.00
32	1.06, d (6.6)	1.06, d (6.0)	0.00
33	1.29, d (6.9)	1.292, br d (6.9)	0.00
34	2.84/2.94[b]	2.81/2.91[b], s	0.03/0.03

[a]Adjusted to the solvent signal of acetone-d_6 (δ_H 2.05 ppm). [b]Due to the amide substructure signal doubling is observed.[2] [c]larger Δ ($\delta_{\text{Ajudazol B}}$ - $\delta_{\text{Ajudazol C}}$) due to difference in structure. Positions with a stereogenic center are marked with a grey background.

Table S3. 9. *^{13}C-NMR Spectroscopic Data for Ajudazol B[2] (**2**) vs C (**3**).*



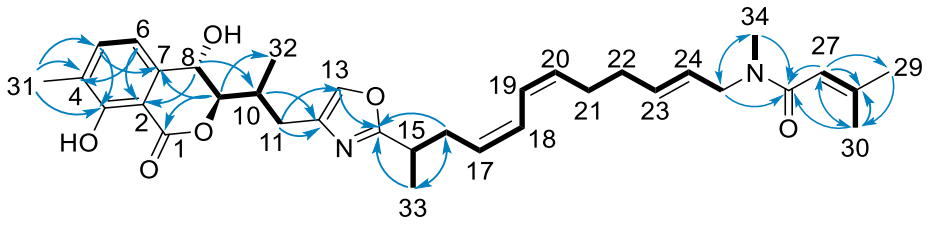
ajudazol B (**2**) $R=OCH_3$ ajudazol C (**3**) $R=CH_3$

Position	Chemical shift (δ_C, in ppm), Ajudazol B[2] (acetone-d_6, 150 MHz)[a]	Chemical shift (δ_C, in ppm), Ajudazol C (acetone-d_6, 175 MHz)[a]	Δ ($\delta_{Ajudazol\ B}$ - $\delta_{Ajudazol\ C}$, in ppm)
1	170.3	170.3	0
2	107.2	107.3	-0.1
3	160.6	160.7	-0.1
4	126.1	126.2	-0.1
5	137.9	138.1	-0.3
6	116.5	116.6	-0.1
7	142.1	142.2	-0.1
8	65.3	65.4	-0.1
9	88.1	88.2	-0.1
10	33.8	33.9	-0.1
11	27.8	27.9	-0.1
12	139.4	139.5	-0.1
13	135.8	135.9	-0.1
14	168.2	168.3	-0.1
15	34.6	34.7	-0.1
16	33.5	33.6	-0.1
17	129.0	129.1	-0.1
18	126.3	126.4	-0.1
19	124.6	124.8	-0.2
20	132.4	132.4	0
21	27.8	27.9	-0.1
22	32.8	32.8	0
23	132.7	133.1	-0.4

24	127.0	126.8	0.2
25a[b]	52.5	52.4	0.1
25b[b]	49.2	48.9	0.3
26	167.9	167.9	0
27	92.3	119.1	-26.8[c]
28	168.7	146.7	22.0[c]
29	55.3	26.4	-28.9[c]
30	18.7	20.2	-1.5[c]
31	15.4	15.5	-0.1
32	16.9	17.0	-0.1
33	18.3	18.4	-0.1
34a[b]	34.3	35.0	-0.7
34b[b]	33.2	32.5	0.7

[a]Adjusted to the solvent signal of acetone-d_6 (δ_C 29.9 ppm). [b]Due to the amide substructure signal doubling is observed.[2] [c]larger Δ ($\delta_{\text{Ajudazol B}}$ - $\delta_{\text{Ajudazol C}}$) due to difference in structure.

Table S3. 10. *NMR Spectroscopic Data for Ajudazol C (**3**)*[a].

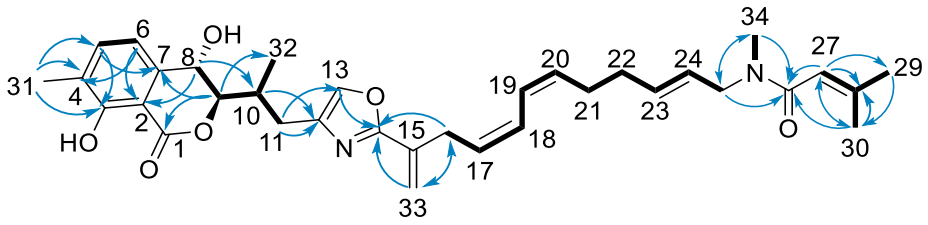


Position	δ_C[b], type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]
1/26[f]	170.8, C	-	-	-
2	107.5, C	-	-	-
3	161.2, C	-	-	-
4	127.3, C	-	-	-
5	138.6, CH	7.46, dd (7.6, 0.7)	6, 31	2, 3, 6, 7, 8, 31
6	117.4, CH	7.01, d (7.9)	5, 8, 31	1, 2, 3, 4, 5, 8
7	141.4, C	-	-	-
8	65.8, CH	4.89, m	6, 9	2, 4, 5, 6, 7, 9, 10
9	89.0, CH	4.35, dd (7.5, 5.0)	8, 10	1, 7, 8, 10, 11, 32
10	34.5, CH	2.29, m	9, 11a, 11b, 32	9, 11, 12, 32
11a	28.4, CH2	2.45, m	10, 11b, 13, 32	9, 10, 12, 13, 32
11b		2.87, m	10, 11a, 13, 32	9, 10, 12, 13, 32
12	139.3, C	-	-	-
13	136.9, CH	7.58, s	11a, 11b	10, 14
14	169.8, C	-	-	-
15	35.5, CH	3.05, m	16a, 16b, 33	14, 16, 17, 33
16a	34.1, CH2	2.60, m	15, 16b, 17, 18	14, 15, 17, 18, 33
16b		2.49, m	15, 16a, 17, 18	14, 15, 17, 18, 33
17	129.0, CH	5.35, m	16a, 16b, 18	15, 16, 19
18	126.9, CH	6.28, m	16a, 16b, 17, 19	16, 20
19	124.9, CH	6.20, m	18, 20, 22b	17, 22
20	132.9, CH	5.41, m	19, 22a, 22b	18, 21, 22, 23
21	33.3, CH2	2.13, m	22a, 22b, 23a, 23b, 25a	20, 22, 23, 24
22a	28.2, CH2	2.11, m	20, 21, 22b, 23a, 23b, 24, 25b	20, 23b, 24
22b		2.22, m	19, 20, 21, 22a, 24	19, 20, 21, 23a, 23b
23a[g]	134.4, CH	5.56, m	21, 22a, 24, 25a	20, 21, 22, 24, 25a
23b[g]	135.0, CH	5.62, m	21, 22a, 24, 25b	21, 22, 25b

24	126.1, CH	5.42, m	22a, 22b, 23a, 23b, 25a, 25b	21, 22, 23a, 23b, 25a, 25b
25a[g]	53.4, CH2	3.90, dd (5.7, 0.9)	22a, 23a, 24, 34a	23a, 24, 26, 34a
25b[g]	50.0, CH2	3.93, d (6.3)	22a, 23b, 24, 34b	23b, 24, 26, 34b
27a[g]	119.1, CH	5.86, m	29a	26, 28, 29a, 29b, 30
27a[g]	119.4, CH	5.88, m	29b, 30	26, 28, 29a, 29b, 30
28	148.2, C	-	-	-
29a[g]	26.3, CH3[f]	1.84, d (1.3)	27a	27a, 28, 30
29a[g]	26.3, CH3[f]	1.86, m	27b	27b, 28, 30
30	20.5, CH3	1.86, s	27b	26, 27a, 27a, 28, 29a, 29b
31	15.7, CH3	2.25, m	5, 6	2, 3, 4, 5, 6
32	16.6, CH3	1.00, d (7.0)	10, 11	9, 10, 11
33	18.5, CH3	1.32, d (7.0)	15	14, 15, 16
34a[g]	33.2, CH3	2.88, m	25a, 27a	25a, 26
34b[g]	35.8, CH	2.93, m	25b, 27b	25b, 26

[a]Recorded in methanol-d_4. [b]Acquired at 125 MHz, adjusted to the solvent signal of methanol-d_4 (δ_C 49.15 ppm). [c]Acquired at 500 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 3.31 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Protons showing HMBC correlations to indicated carbons. [f]Overlapping carbons signals. [g]Due to the amide substructure signal doubling is observed (rotamer effect).[2]

Table S3. 11. *NMR Spectroscopic Data for Ajudazol D (**4**)*[a].

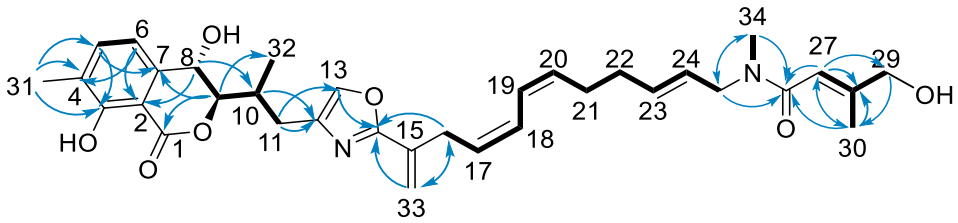


Position	δ_C[b], type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]
1	170.8, C	-	-	-
2	107.5, C	-	-	-
3	161.2, C	-	-	-
4	127.4, C	-	-	-
5	138.6, CH	7.45, dd (7.6, 0.8)	6, 31	2, 3, 6, 7, 31
6	117.5, CH	7.01, d (7.5)	5, 8, 31	1, 2, 3, 4, 5, 8
7	141.4, C	-	-	-
8	65.8, CH	4.91, m	6, 9, 31	2, 6, 7, 9, 10
9	89.0, CH	4.38, dd (7.4, 5.1)	8, 10, 32	1, 7, 8, 10, 11, 32
10	34.6, CH	2.34, m	9, 11a, 11b, 32	8, 9, 11, 32
11a	28.6, CH2	2.53, dd (14.7, 9.4)	10, 11b, 13, 32	9, 10, 12, 13, 32
11b	28.6, CH2	2.91, m	10, 11a, 13, 32	9, 10, 12, 13, 32
12	140.9, C	-	-	-
13	137.4, CH	7.66, s	11a, 11b	12, 14
14	163.6, C	-	-	-
15	136.0, C	-	-	-
16	31.3, CH2	3.36, m	17, 18, 33a, 33b	14, 15, 17, 18, 33
17	128.4, CH	5.56, m	16, 18	16
18	126.9, CH	6.41, m	16, 17, 19	16, 23
19	125.0, CH	6.35, m	18, 22b, 23	17, 22
20	131.0, CH	5.34, m	22a	22
21	33.3, CH2	2.16, m	22a, 25a, 25b	22, 23, 24
22a	28.3, CH2	2.03, br d (6.2)	20	20, 21
22b	28.3, CH2	2.30, m	19, 21, 23	19, 20, 21, 23
23	133.2, CH	5.49, m	19, 22b	21, 22
24	126.1, CH	5.43, m	21, 25a, 25b	21, 25a, 25b
25a[f]	53.4, CH2	3.90, dd (5.7, 1.1)	21, 24	24, 26, 34a

25b[f]	49.8, CH2	3.93, dd (6.2, 0.5)	21, 24	24, 26, 34b
26	171.1, C	-	-	-
27	119.2, CH	5.87, m	29, 30, 34a	26, 28, 29, 30
28	148.2, C	-	-	-
29	26.3, CH3	1.84, d (1.3)	27	27, 28, 30
30	20.6, CH3	1.86, d (1.2)	27	26, 27, 28, 29
31	15.7, CH3	2.24, s	5, 6, 8	2, 3, 4, 5, 6
32	16.7, CH3	1.04, d (7.0)	9, 10, 11a, 11b	9, 10, 11
33a	118.8, CH2	5.97, s	16, 33b	14, 15, 16
33b	118.8, CH2	5.42, m	16, 33a	14, 15, 16
34a[f]	33.2, CH3	2.87, m	27	25a, 26
34b[f]	35.8, CH3	2.93, s	-	25b, 26

[a]Recorded in methanol-d_4. [b]Acquired at 125 MHz, adjusted to the solvent signal of methanol-d_4 (δ_C 49.15 ppm). [c]Acquired at 500 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 3.31 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Protons showing HMBC correlations to indicated carbons. [f]Due to the amide substructure signal doubling is observed.[2]

Table S3. 12. *NMR Spectroscopic Data for Ajudazol E (**5**)*[a].

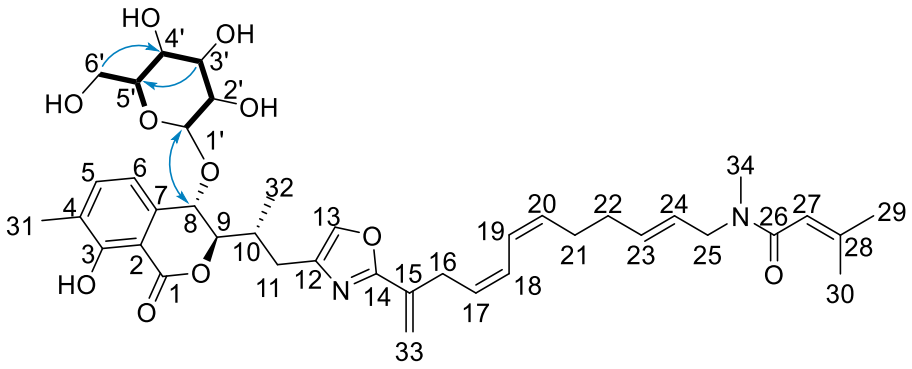


Position	δ_C[b], type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]
1, 26[f]	170.8, C	-	-	-
2	107.5, C	-	-	-
3	161.2, C	-	-	-
4	127.4, C	-	-	-
5	138.6, CH	7.45, dd (7.5, 0.7)	6, 31	2, 3, 6, 7, 31
6	117.5, CH	7.01, d (7.5)	5, 8, 31	1, 2, 3, 4, 5, 8
7	141.2, C	-	-	-
8	65.9, CH	4.93, m	6, 9, 31	2, 6, 7, 9, 10
9	89.1, CH	4.39, dd (5.4, 7.3)	8, 10	1, 8, 10, 11, 12, 32

10	34.6, CH	2.35, m	9, 11a, 11b, 32	8, 9, 11, 32
11a	28.6, CH2	2.53, m	10, 11b, 13	9, 10, 12, 13, 32
11b	28.6, CH2	2.91, m	10, 11a, 13	9, 10, 12, 13, 32
12	140.9, C	-	-	-
13	137.3, CH	7.65, s	11a, 11b	10, 12, 14
14	163.4, C	-	-	-
15	136.0, C	-	-	-
16	31.3, CH2	3.36, m	17, 18, 33a, 33b	14, 15, 17, 18, 33
17	128.4, CH	5.56, m	16, 18	15, 16, 19
18	126.9, CH	6.41, m	16, 17, 19	15, 16, 19, 20
19	125.0, CH	6.35, m	18, 20, 22	17, 22, 24
20	133.3, CH	5.50, m	19, 22	18, 21, 22
21	33.3, CH2	2.17, m	22, 23	20, 22, 23, 24
22	28.3, CH2	2.30, q (7.1, 7.1, 7.1)	19, 20, 21	19, 20, 21, 23
23	134.6, CH	5.64, m	21, 24, 25a, 25b	21, 22, 25a, 25b
24	126.2, CH	5.44, m	23, 25a, 25b	21, 25a, 25b
25a[g]	53.4, CH2	3.92, dd (7.5, 13.3)	23, 24, 34a	23, 24, 26, 34a
25b[g]	49.9, CH2	3.96, d (6.2)	23, 24, 34b	23, 24, 26, 34b
27	116.8, CH	6.16, m	29, 30	26, 28, 29, 30
28	148.2, C	-	-	-
29	66.9, CH2	4.03, br d (10.0)	27, 30	27, 28, 30
30	15.7, CH3	1.81, br s	27, 29	26, 27, 28, 29
31	15.7, CH3	2.24, s	5, 6	2, 3, 4, 5
32	16.7, CH3	1.04, m	10	9, 10, 11
33a	118.8, CH2	5.97, m	16, 33b	14, 15, 16
33b	118.8, CH2	5.42, m	16, 33a	14, 15, 16
34a[g]	33.2, CH3	2.90, s	25a	25a, 26
34b[g]	36.0, CH3	2.95, m	25b	25b, 26

[a]Recorded in methanol-d_4. [b]Acquired at 175 MHz, adjusted to the solvent signal of methanol-d_4 (δ_C 49.15 ppm). [c]Acquired at 700 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 3.31 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Protons showing HMBC correlations to indicated carbons. [f]Overlapping carbons signals. [g]Due to the amide substructure signal doubling is observed.[2]

Table S3. 13. *NMR Spectroscopic Data for Ajudazol F (**6**)*[a].

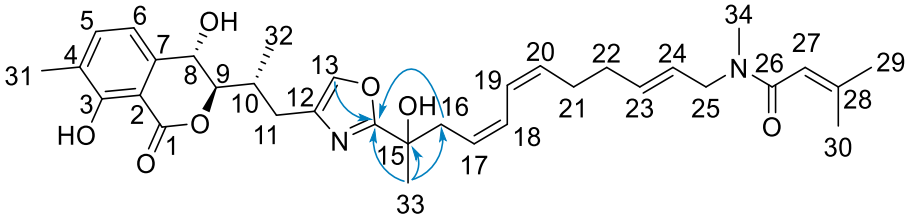


Position	δ_C[b], type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]
1	170.3, C	-	-	-
2	108.3, C	-	-	-
3	161.0, C	-	-	-
4	128.3, C	-	-	-
5	138.5, CH	7.44, dd (7.6, 0.8)	6, 31	2, 3, 6, 7, 8, 31
6	120.6, CH	7.10, d (7.6)	5, 8, 31	1, 2, 3, 4, 5, 8, 31
7	137.1, C	-	-	-
8	73.4, CH	5.10, d (3.9)	6, 9, 31	1‘, 1, 2, 5, 6, 7, 9, 10
9	87.6, CH	4.71, dd (7.7, 3.9)	8, 31	1, 7, 8, 10, 11, 32
10	35.0, CH	2.23, m	11a, 11b, 32	8, 9, 11, 12, 13, 32
11a	29.7, CH2	2.49, dd (14.8, 8.7)	10, 11b, 13, 32	9, 10, 12, 13, 32
11b	29.7, CH2	2.91, m	10, 11a, 13, 32	9, 10, 12, 13, 32
12	140.8, C	-	-	-
13	137.3, CH	7.64, s	11a, 11b, 33	10, 11, 12, 14
14	163.6, C	-	-	-
15	136.0, C	-	-	-
16	31.3, CH2	3.33, m	17, 18, 33	14, 15, 17, 18, 33
17	128.5, CH	5.55, m	16, 18	16, 19
18	126.9, CH	6.39, d (10.1)	16, 17	16, 19, 20
19	125.1, CH	6.35, br d (11.6)	20, 22	17, 18, 22
20	133.2, CH	5.47, m	19, 22	18, 21, 22, 23
21	33.3, CH2	2.16, m	22, 23, 24	20, 22, 23, 24
22	28.3, CH2	2.29, m	19, 20, 21	19, 20, 21, 23
23	134.4, CH	5.59, m	21, 24, 25a, 25b	20, 21, 22, 24

24	126.2, CH	5.43, m	21, 23, 25	21, 23, 25
25a[f]	53.4, CH2	3.91, m	23, 24	24, 26
25b[f]	49.8, CH2	3.93, m	21, 23, 24	23, 24, 34b
26	170.6, C	-	-	-
27	119.4, CH	5.87, dt (2.8, 1.3, 1.3)	30, 34	26, 28, 29, 30
28	147.8, C	-	-	-
29	26.3, CH3	1.84, m	-	27, 30
30	20.6, CH3	1.86, m	27	26, 27, 28, 29
31	15.8, CH3	2.25, m	5, 6, 8, 9	1, 2, 3, 4, 5, 6, 7
32	16.7, CH3	0.98, d (6.9)	10, 11a, 11b	9, 10, 11
33a	118.7, CH2	5.96, s	16, 33	14, 15, 16
33b	118.7, CH2	5.39, m	13, 16, 33	14, 15, 16
34a[f]	33.2, CH	2.87, m	25a	25a, 26
34b[f]	35.8, CH3	2.93, m	27	26
1'	103.6, CH	4.57, d (7.8)	2'	2', 8
2'	75.2, CH	3.20, dd (9.2, 7.8)	1', 3'	3', 1', 4', 5'
3'	78.2, CH	3.37, m	2', 4'	2', 4', 5'
4'	71.6, CH	3.28, m	3', 5'	2', 3', 6'
5'	78.2, CH	3.26, m	4', 6'	2', 3', 6'
6'a	62.8, CH2	3.78, dd (11.9, 1.9)	5', 6'	4', 5', 6'
6'b	62.8, CH2	3.62, dd (11.8, 5.4)	5', 6'	4', 5', 6'

[a]Recorded in methanol-d_4. [b]Acquired at 125 MHz, adjusted to the solvent signal of methanol-d_4 (δ_C 49.15 ppm). [c]Acquired at 500 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 3.31 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Protons showing HMBC correlations to indicated carbons. [f]Due to the amide substructure signal doubling is observed.[2]

Table S3. 14. *NMR Spectroscopic Data for Ajudazol G (**7**)*[a].

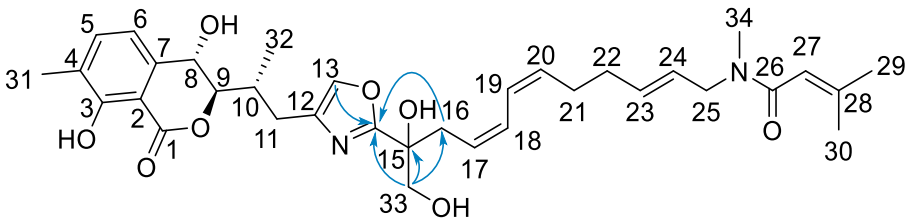


Position	δ_C[b], type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]
1	170.7, C	-	-	-
2	107.3, C	-	-	-
3	161.1, C	-	-	-
4	127.2, C	-	-	-
5	138.6, CH	7.46, dd (7.5, 0.8)	6, 31	3, 6, 7, 31
6	117.3, CH	7.02, dd (7.5, 0.6)	5	2, 4, 5, 8
7	141.6, C	-	-	-
8	65.7, CH	4.9, br d (7.8)	9	2, 6, 7, 9, 10
9	89.0, CH	4.35, dd (7.7, 4.7)	8, 10	1, 7, 8, 10, 11, 32
10	34.5, CH	2.31, m	9, 11a, 11b, 32	8, 9, 11, 32
11a	28.3, CH2	2.48, dd (14.6, 9.7)	11b, 13	9, 10, 12, 13, 32
11b	28.3, CH2	2.88, m	10, 11a	9, 10, 12, 13, 32
12	139.4, C	-	-	-
13	137.3, CH	7.63, s	11a, 11b	11, 12, 14
14	169.2, C	-	-	-
15	72.7, C	-	-	-
16a	40.7, CH2	2.77, m	16, 17	14, 15, 17, 18, 33
16b	40.7, CH2	2.71, m	16, 17	14, 15, 17, 18, 33
17	126.2, CH	5.37, m	16, 18	16
18	127.7, CH	6.31, m	17, 19	16, 19, 20
19	125.1, CH	6.19, m	18, 20	18, 24
20	133.0, CH	5.41, m	19, 22, 23	18, 21, 22
21	33.3, CH2	2.12, m	22, 23	20, 22, 23, 24
22	28.2, CH2	2.22, m	20, 21	20, 21, 23
23	134.4, CH	5.57, m	20, 21, 25	21, 22, 25
24	126.2, CH	5.39, m	25	19, 21, 25
25a[f]	53.4, CH2	3.90, br dd (5.8, 1.1)	24, 25	23, 24

25b[f]	49.8, CH2	3.93, br d (6.4)	23, 24	23, 24, 26, 34b
26	170.9, C	-	-	-
27	119.4, CH	5.88, m	29, 30, 34	26, 29, 30
28	148.2, C	-	-	-
29	26.3, CH3	1.84, m	27	27, 28, 30
30	20.5, CH3	1.86, m	27	26, 27, 28, 29
31	15.7, CH3	2.25, s	5	2, 3, 4, 5
32	16.6, CH3	1.00, d (6.9)	10	9, 10, 11
33	26.4, CH3	1.56, s	-	-
34a[f]	33.3, CH3	2.87, m	-	-
34b[f]	35.8, CH3	2.93, s	27	25b, 26

[a]Recorded in methanol-d_4. [b]Acquired at 175 MHz, adjusted to the solvent signal of methanol-d_4 (δ_C 49.15 ppm). [c]Acquired at 700 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 3.31 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Protons showing HMBC correlations to indicated carbons. [f]Due to the amide substructure signal doubling is observed.[2]

Table S3. 15. *NMR Spectroscopic Data for Ajudazol H (**8**)*[a].

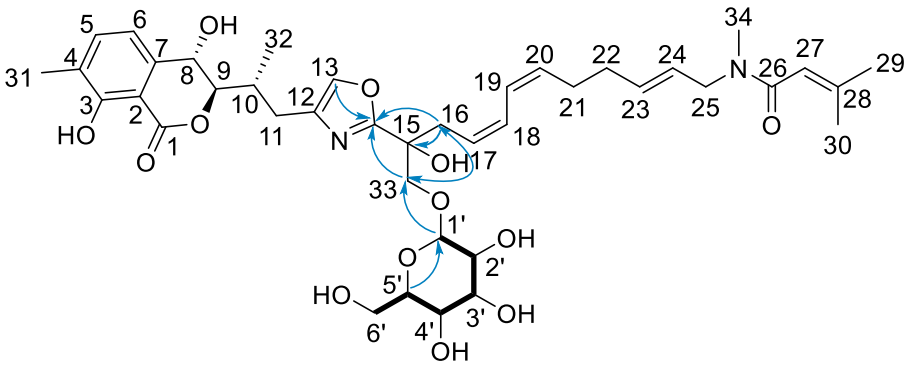


Position	δ_C[b], type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]
1	170.9, C	-	-	-
2	107.5, C	-	-	-
3	161.2, C	-	-	-
4	127.3, C	-	-	-
5	138.6, CH	7.46, d (7.7)	6, 31	2, 3, 6, 7, 8, 31
6	117.4, CH	7.02, d (7.6)	5, 8, 31	1, 2, 3, 4, 5, 8, 31
7	141.6, C	-	-	-
8	65.7, CH	4.89, m	6, 9, 31	1, 2, 5, 6, 7, 9, 10
9	88.9, CH	4.35, dd (7.7, 4.6)	8, 10, 32	1, 7, 8, 10, 11, 32
10	34.6, CH	2.32, m	9, 11a, 11b, 32	8, 9, 11, 12, 13, 32

11a	28.4, CH2	2.48, dd (9.6, 14.6)	10, 11b, 13, 32	9, 10, 12, 13, 32
11b	28.4, CH2	2.88, m	10, 11a, 13, 32	9, 10, 12, 13, 32
12	139.7, C	-	-	-
13	137.4, CH	7.64, s	11a, 11b	10, 11, 12, 14
14	167.6, C	-	-	-
15	76.3, C	-	-	-
16	35.6, CH2	2.76, m	17, 18	14, 15, 17, 18, 33
17	132.8, CH	5.42, m	16, 18, 19	15, 16, 18, 19, 20, 21, 22
18	127.7, CH	6.30, m	16, 17, 19	16, 17
19	125.2, CH	6.21, m	17, 18, 22	17, 21, 22, 24
20	134.4, CH	5.56, m	21	17, 21, 22, 24
21	33.3, CH2	2.11, m	20, 22, 23, 25	17, 19, 20, 22, 23, 24
22	28.2, CH2	2.21, m	19, 21, 24	17, 19, 20, 21, 23, 24
23	135.0, CH	5.62, m	21, 24, 25	21, 22, 25
24	125.9, CH	5.41, m	22, 23, 25	19, 20, 21, 22, 25
25a[f]	53.4, CH2	3.90, br d (5.6)	21, 23, 24, 34a	23, 24, 26, 34
25b[f]	49.8, CH2	3.93, d (6.2)	21, 23, 24, 34b	23, 24, 26, 34
26	171.1, C	-	-	-
27	119.1, CH	5.87, br d (8.0)	29, 30	26, 28, 29, 30
28	148.2, C	-	-	-
29	26.3, CH3	1.84, d (0.8)	27, 30	27, 28, 30
30	20.6, CH3	1.86, s	27, 29	26, 27, 28, 29
31	15.7, CH3	2.25, s	5, 6	2, 3, 4, 5, 6
32	16.7, CH3	1.01, m	9, 10, 11a, 11b	9, 10, 11
33a	68.0, CH2	3.70, d (11.3)	33b	14, 15, 16
33b		3.83, d (11.3)	33a	14, 15, 16
34a[f]	33.2, CH3	2.87, m	25a	25a, 26
34b[f]	35.8, CH3	2.93, s	25b	25b, 26

[a]Recorded in methanol-d_4. [b]Acquired at 125 MHz, adjusted to the solvent signal of methanol-d_4 (δ_C 49.15 ppm). [c]Acquired at 500 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 3.31 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Protons showing HMBC correlations to indicated carbons. [f]Due to the amide substructure signal doubling is observed.[2]

Table S3. 16. *NMR Spectroscopic Data for Ajudazol I (**9**)*[a].

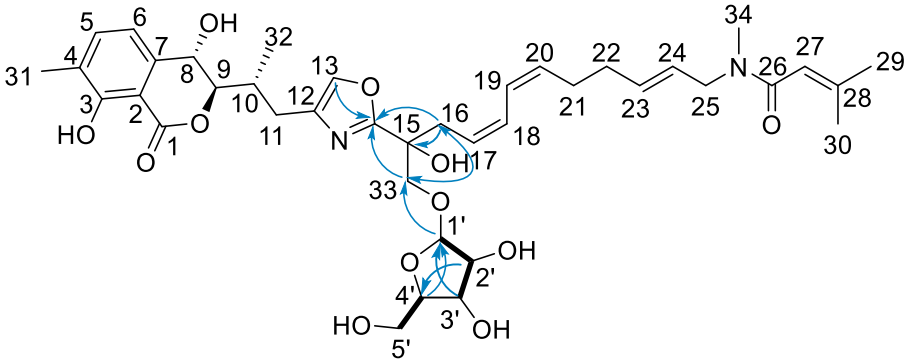


Position	δ_C[b], type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]
1, 26[f]	170.9, C	-	-	-
2	107.5, C	-	-	-
3	161.3, C	-	-	-
4	127.4, C	-	-	-
5	138.7, CH	7.46, br d (7.6)	6, 31	2, 3, 6, 7, 8, 31
6	117.4, CH	7.02, d (7.5)	5, 8, 31	1, 2, 3, 4, 5, 8, 31
7	141.5, C	-	-	-
8	65.8, CH	4.89, m	6, 9, 31	1, 2, 4, 5, 6, 7, 9, 10
9	89.0, CH	4.35, m	8, 31	1, 7, 8, 10, 11, 32
10	34.5, CH	2.32, m	11a, 11b, 32	8, 9, 11, 12, 13, 32
11a	28.3, CH2	2.48, m	10, 11b, 13, 32	9, 10, 12, 13, 32
11b	28.3, CH2	2.87, br d (4.4)	10, 11a, 13, 32	9, 10, 12, 13, 32
12	139.8, C	-	-	-
13	137.4, CH	7.64, s	11a, 11b	10, 11, 12, 14
14	167.2, C	-	-	-
15	87.1, C	-	-	-
16	36.4, CH2	2.76, m	17, 19	14, 15, 17, 19, 33
17	126.2, CH	5.42, m	16, 18, 19	16, 20
18	127.9, CH	6.29, m	17, 21	21
19	125.1, CH	6.17, m	16, 17, 20	16, 22
20	133.1, CH	5.41, m	19, 22, 23	17, 19, 22
21	33.3, CH2	2.12, m	18, 22, 23, 24	20, 22, 23
22	28.2, CH2	2.20, m	19, 20, 21, 24	20, 21, 23
23	135.0, CH	5.59, m	19, 20, 21, 22, 24, 25	21, 22, 25a, 25b

24	125.4, CH	5.40, m	21, 22, 23, 25	21, 22, 25a, 25b
25a[g]	53.4, CH2	3.90, d[h]	22, 23, 24	22, 23, 24, 26, 34a
25b[g]	49.9, CH2	3.93, dd[h]	22, 23, 24	18, 22, 23, 24, 26, 34b
27	119.1, CH	5.87, br d (8.7)	30, 34a, 34b	26, 28, 29, 30
28	148.2, C	-	-	-
29	26.3, CH3	1.84, d (1.0)	-	26, 27, 28, 30
30	20.6, CH3	1.85, d (1.2)	-	26, 27, 28, 29
31	15.7, CH3	2.24, br s	5, 6, 8	2, 3, 4, 5, 6, 7
32	16.7, CH3	1.00, m	10, 11a, 11b	9, 10, 11
33a	75.7, CH2	4.33, m	33b	1‘, 14, 33
33b	75.7, CH2	3.75, d (10.3)	33a	1‘, 14, 33
34a[g]	33.3, CH3	2.87, s	27	25, 26, 27
34b[g]	35.8, CH3	2.94, m	27	25, 26
1'	105.0, CH	4.28, m	2'	2', 3', 5', 52
2'	75.2, CH	3.13, dd (9.2, 7.8)	1', 3'	1', 3'
3'	77.9, CH	3.34, m	2', 4'	1', 2', 4', 5'
4'	71.7, CH	3.21, m	3', 5'	1, 3', 5', 6'
5'	78.2, CH	3.24, m	4', 6'a, 6'b	1', 3', 4', 6'
6'a	62.9, CH2	3.59, dd (12.0, 5.8)	5', 6'b	4', 5'
6'b	62.9, CH2	3.83, br dd (11.9, 2.0)	5', 6'a	4', 5'

[a]Recorded in methanol-d_4. [b]Acquired at 125 MHz, adjusted to the solvent signal of methanol-d_4 (δ_C 49.15 ppm). [c]Acquired at 500 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 3.31 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Protons showing HMBC correlations to indicated carbons. [f]Overlapping carbons signals. [g]Due to the amide substructure signal doubling is observed.[2]

Table S3. 17. *NMR Spectroscopic Data for Ajudazol J (**10**)*[a].



Position	δ_C[b], type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]
1	169.9, C	-	-	-
2	105.8, C	-	-	-
3	160.1, C	-	-	-
4	125.5, C	-	-	-
5	137.5, CH	7.42, br d (7.5)	6, 31	2, 3, 6, 7, 8, 31
6	114.4, CH	7.13, br d (7.7)	5, 8	1, 2, 3, 4, 5, 8, 31
7	140.8, C	-	-	5, 8, 9
8	63.8, CH	4.97, br d (10.8)	6, 9	5, 6, 7, 9, 10, 32
9	86.7, CH	4.26, br d (10.9)	8	7, 8, 10, 11, 32
10	31.2, CH	2.55, br s	11a, 11b, 32	8, 9, 11, 13, 32
11a	25.4, CH2	2.80, br d (14.1)	11b	9, 10, 12, 13, 32
11b	25.4, CH2	2.48, br d (13.7)	11a, 13	9, 10, 12, 13, 32
12	138.1, C	-	-	-
13	135.5, CH	7.45, s	11b	10, 11, 12, 14
14	164.8, C	-	-	-
15	72.5, C	-	-	-
16	34.2, CH2	2.89, s	17, 18	14, 15, 18, 33
17	125.1, CH	5.38, m	16, 18, 19, 20	-
18	127.6, CH	6.41, br d (11.7)	16, 17, 19	15, 16, 19, 20
19	123.4, CH	6.25, br d (5.0)	16, 17, 18	16, 20
20	133.1, CH	5.53, m	17, 21, 23	17, 19, 21, 22
21	31.9, CH2	2.18, m	20, 22	20, 22, 23, 24
22	26.9, CH2	2.28, m	19, 20, 21	21, 23, 24
23	133.4, CH	5.57, br d (11.6)	21, 22	21, 22, 25a, 25b

24	123.2, CH	5.34, br dd (15.5, 5.8)	21, 22	21, 22, 25a, 25b
25a	48.7, CH2	3.91, m	22, 24	21, 24, 26, 34a
25b	48.7, CH2	4.02, br dd (14.3, 6.5)	24, 25c	23, 24, 26, 34b
25c	52.1, CH2	3.85, m	24, 25b	23, 24, 26, 34b
26	168.3, C	-	-	-
27	117.8, CH	5.78, m	29, 30	26, 28, 29, 30
28	147.0, C	-	-	-
29	26.3, CH3	1.84, m	27	26, 27, 28, 30
30	20.2, CH3	1.92, s	27	26, 27, 28, 29
31	15.5, CH3	2.27, s	5	2, 3, 4, 5, 6, 8, 9
32	16.3, CH3	1.07, br d (6.8)	10	8, 9, 10, 11
33a	69.2, CH2	3.48, br dd (18.8, 9.8)	33b	1', 14, 15, 16
33b	69.2, CH2	4.13, br d (10.0)	33a	1', 14, 16
34a	34.9, CH	2.92, s	-	25a, 26, 27, 28
34b	32.5, CH3	2.90, m	-	25b, 26
1'	107.2, CH	5.10, s	-	2', 3', 4', 33
2'	78.5, CH	4.18, br s	3', 4'	3', 4', 5'
3', 4'	87.9, CH	4.10, m	3', 4', 5', OH	1', 3', 4', 5'
5'a	61.8, CH2	3.82, m	5'b, OH	2', 3', 4'
5'b	61.8, CH2	3.88, m	5'a, OH	2', 3', 4'
38	OH	6.06, m	-	-
3‘-OH	OH	4.46, br d (8.4)	3'	-
5‘-OH	OH	3.38, m	5'	-
3-OH	OH	11.08, s	-	1, 2, 3, 4, 5, 31
8-OH	OH	5.84, br s	8	-

[a]Recorded in methanol-d_4. [b]Acquired at 125 MHz, adjusted to the solvent signal of methanol-d_4 (δ_C 49.15 ppm). [c]Acquired at 500 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 3.31 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Protons showing HMBC correlations to indicated carbons. [f]Due to the amide substructure signal doubling is observed.[2]

S 3.4 Bioactivity Data

Table S3. 18. *Minimum-Inhibitory Concentration (MIC, µg/mL) of Ajudazols C-J against Common Microbial Pathogens.*

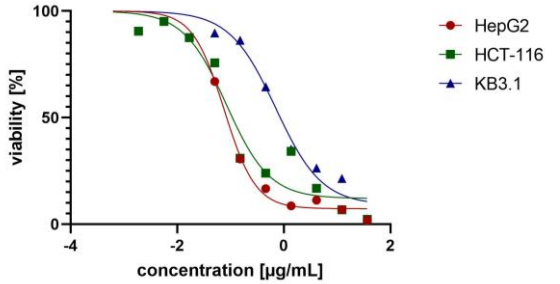
Test organisms	C	D	E	F	G	H	I	J
Staphylococcus aureus Newman	> 64	> 64	> 64	> 64	> 64	> 64	> 64	> 64
Mycobacterium smegmatis mc2-155	> 64	> 64	> 64	64	> 64	> 64	> 64	64
Bacillus subtilis DSM10	> 64	> 64	> 64	> 64	> 64	> 64	> 64	> 64
Citrobacter freundii DSM30039	> 64	> 64	> 64	> 64	> 64	> 64	> 64	> 64
Acinetobacter baumannii DSM30007	> 64	> 64	> 64	> 64	> 64	> 64	> 64	> 64
Candida albicans DSM1665	> 64	> 64	> 64	> 64	> 64	> 64	> 64	> 64
Escherichia coli Δ*acrB* JW0451-2	> 64	> 64	> 64	> 64	> 64	> 64	> 64	> 64
Escherichia coli DSM1116	> 64	> 64	> 64	> 64	> 64	> 64	> 64	> 64
Cryptococcus neoformans DSM11959	> 64	> 64	> 64	> 64	> 64	> 64	> 64	> 64
Pseudomonas aeruginosa PA14 (DSM19882)	> 64	> 64	> 64	> 64	> 64	> 64	> 64	> 64
Pichia anomala DSM6766	> 64	> 64	> 64	> 64	> 64	> 64	> 64	> 64

Table S3. 19. *Half-Inhibitory Concentration (IC_{50}, µM) of Ajudazols C-J against Cancer Cell Lines.*

Test cancer cell line	C	D	E	F	G	H	I	J
HepG2	0.9	4.1	0.17	n.d.	0.02	5.72	48	n.d.
HCT-116	0.5	1.7	0.22	1.68	0.03	3.17	60	25
KB3.1	2.2	4.8	1.75	10.5	2.4	6.76	48	69

n.d.: not determined

IC_{50} plots Ajudazol E



IC_{50} plots Ajudazol G

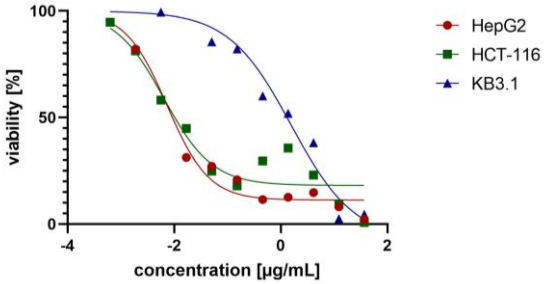


Figure S3. 3. Example of IC_{50} values obtained from sigmoidal curve fitting.

S 3.5 GC-MS Analysis

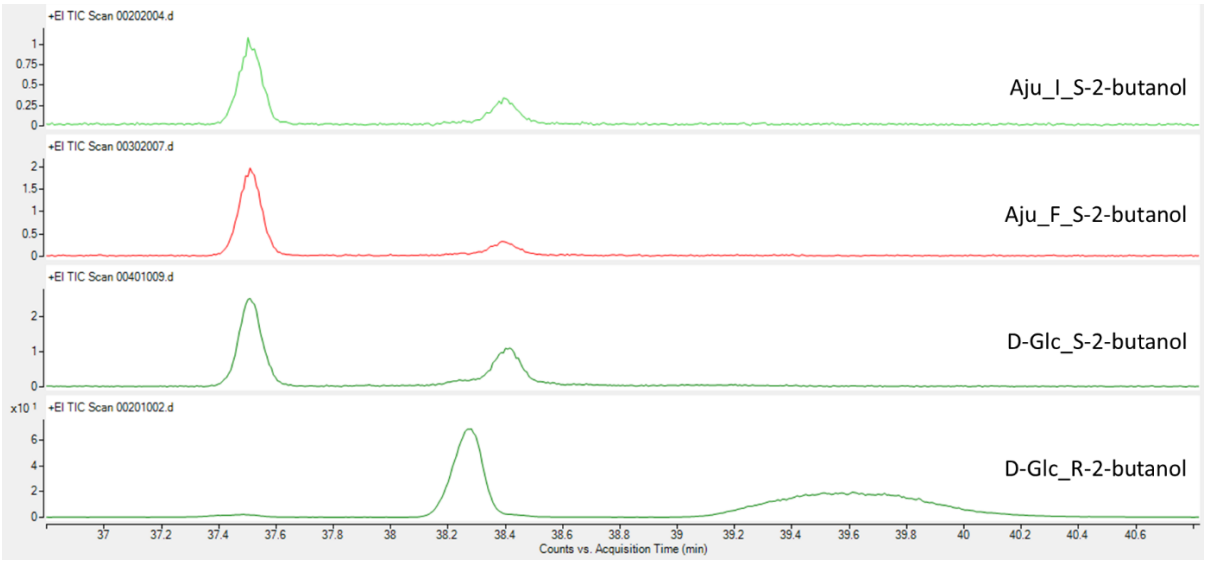


Figure S3. 4. GC-MS chromatograms of trimetylsilylated (S)- and (R)-2-butyl glycosides, comparison with glucose. *TIC of (S)-2-butyl glycosides, 4 TMS derivative of ajudazol I (**9**), ajudazol F (**6**) and of the 4 TMS derivatives of (S)-2-butyl D-glucoside and (R)-2-butyl D-glucoside. Each glycoside shows open- and pyranose form (two peaks).*

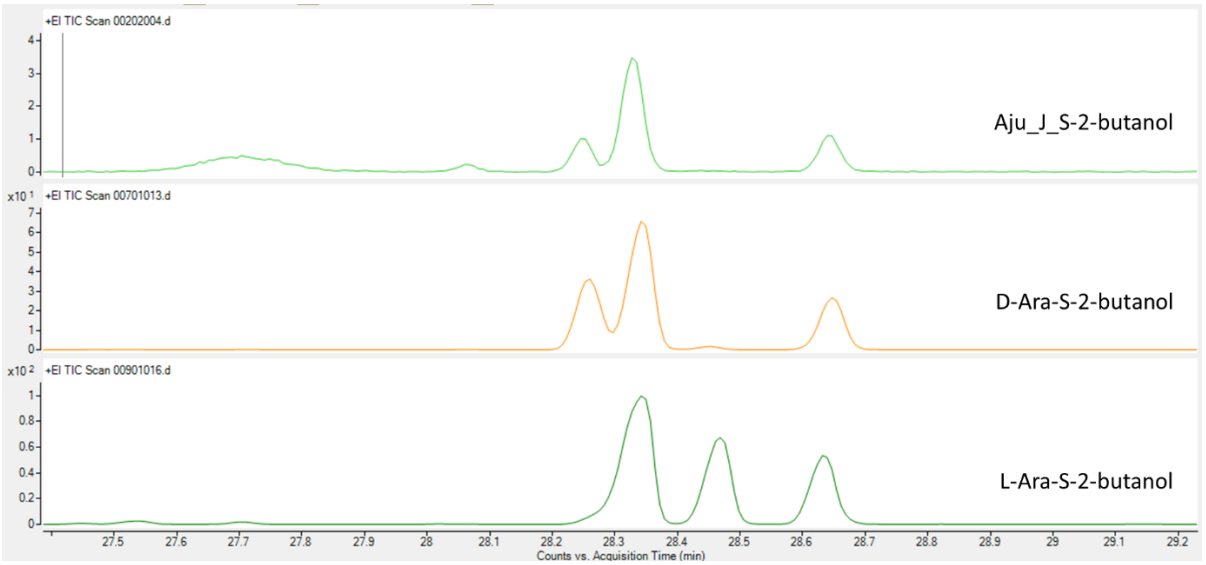


Figure S3. 5. GC-MS chromatograms of trimetylsilylated (S)- and (R)-2-butyl glycosides, comparison with arabinose. *TIC of (S)-2-butyl glycosides, 4 TMS derivative of ajudazol J (**10**) and of the 4 TMS derivatives of (S)-2-butyl D-arabinoside and (S)-2-butyl L-arabinoside. Each glycoside shows open-, furanose- and pyranose form (three peaks).*

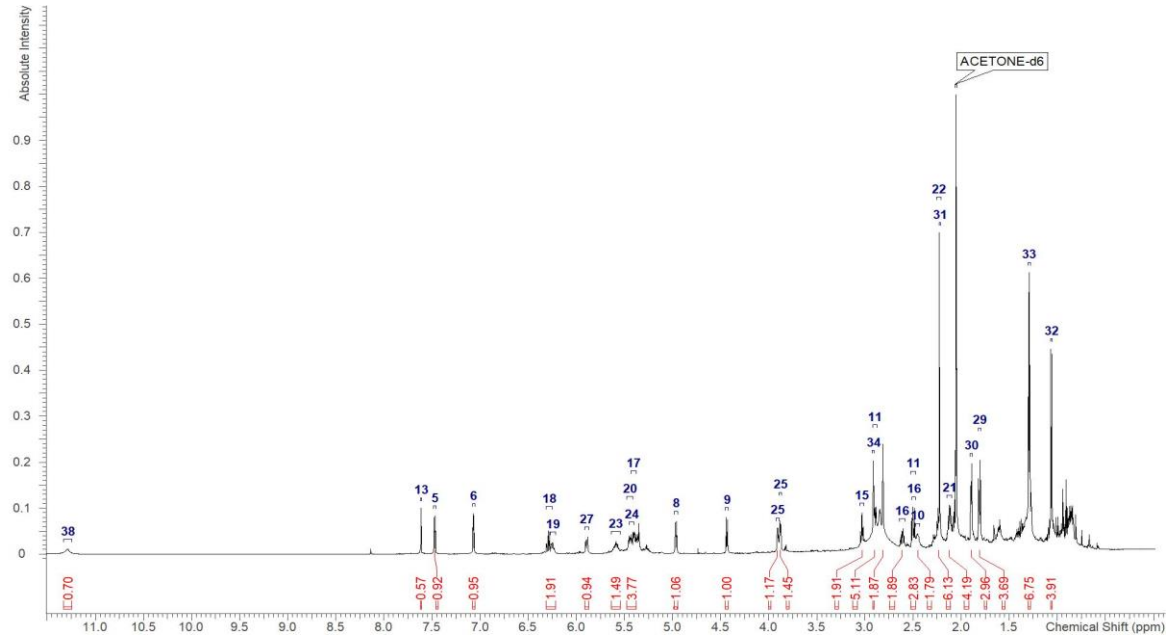


Figure S3. 6. *Ajudazol C (**3**) - ^{1}H NMR recorded in acetone-d$_6$ at 700 MHz*

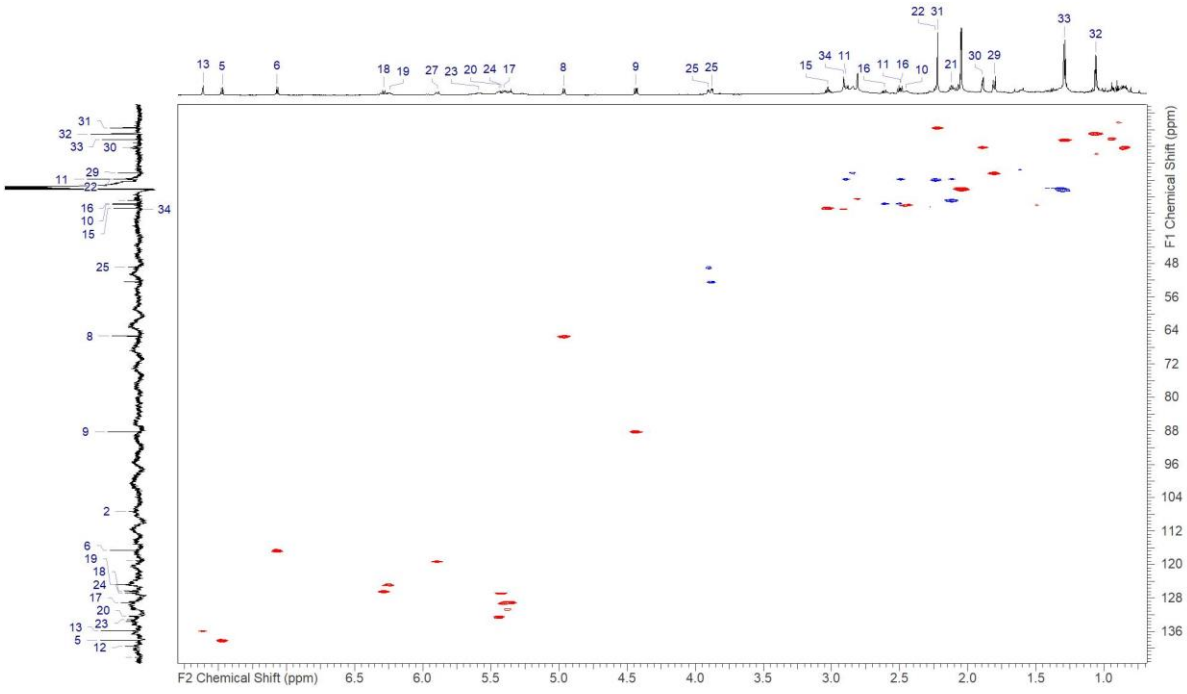


Figure S3. 7. *Ajudazol C (**3**) - HSQC NMR recorded in acetone-d$_6$ at 175/700 (F1/F2) MHz*

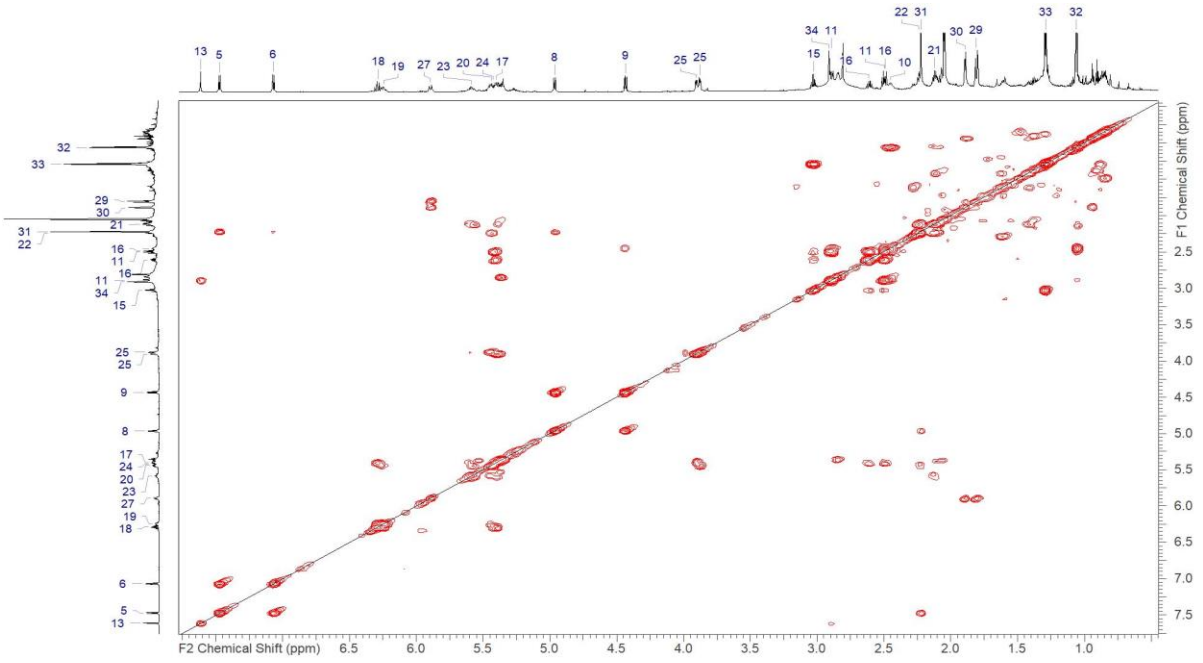


Figure S3. 8. *Ajudazol C (**3**) - COSY NMR recorded in acetone-*d_6 *at 700 MHz*

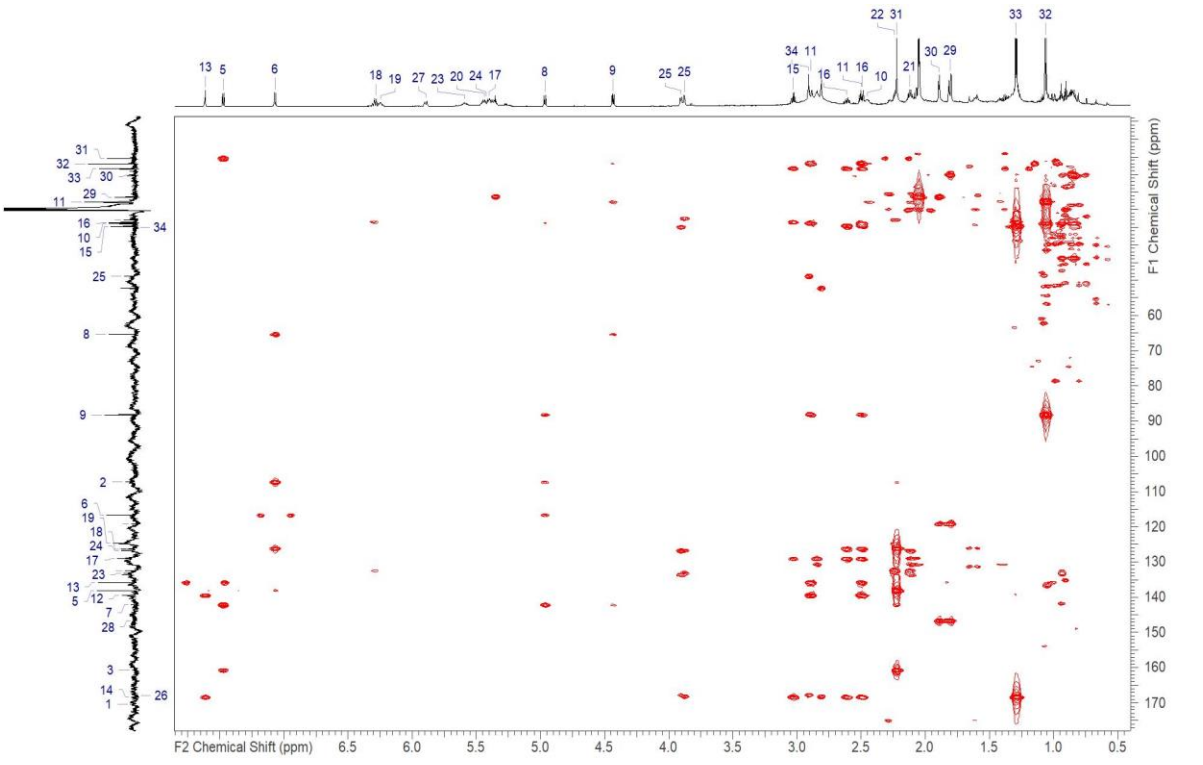


Figure S3. 9. *Ajudazol C (**3**) - HMBC NMR recorded in acetone-*d_6 *at 700/175 (F1/F2) MHz*

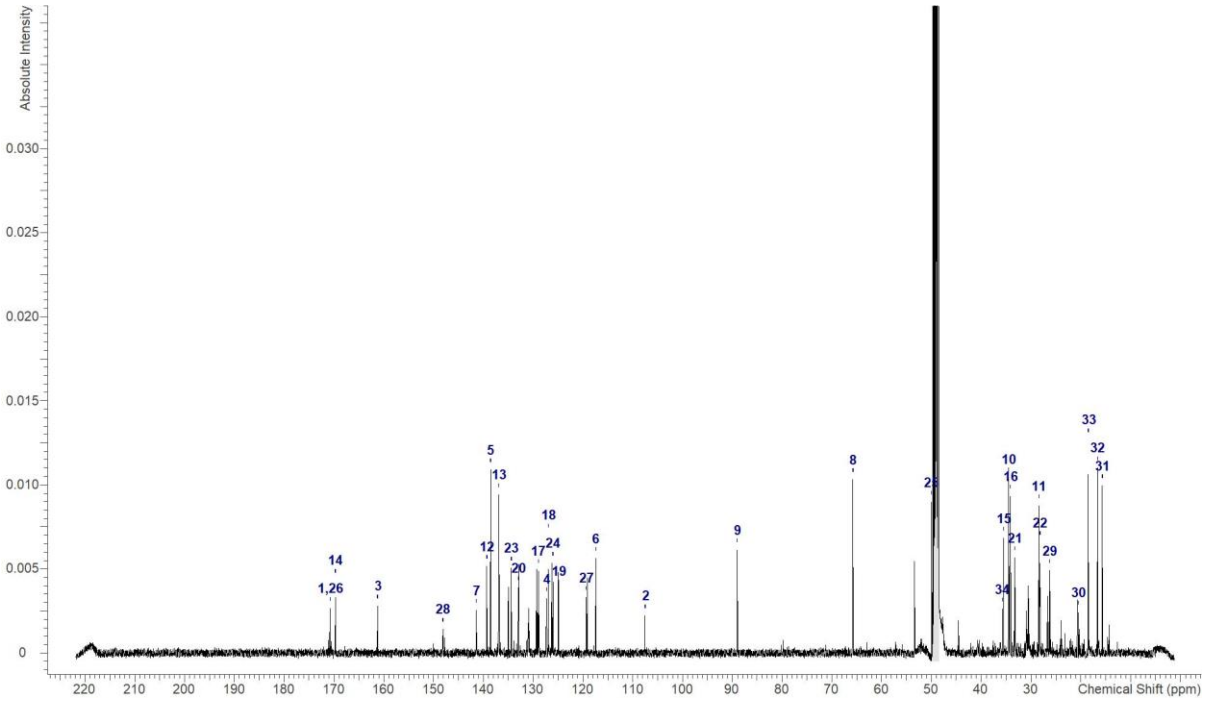


Figure S3. 10. *Ajudazol C (**3**) - ^{13}C NMR recorded in methanol-*d_4 *at 125 MHz*

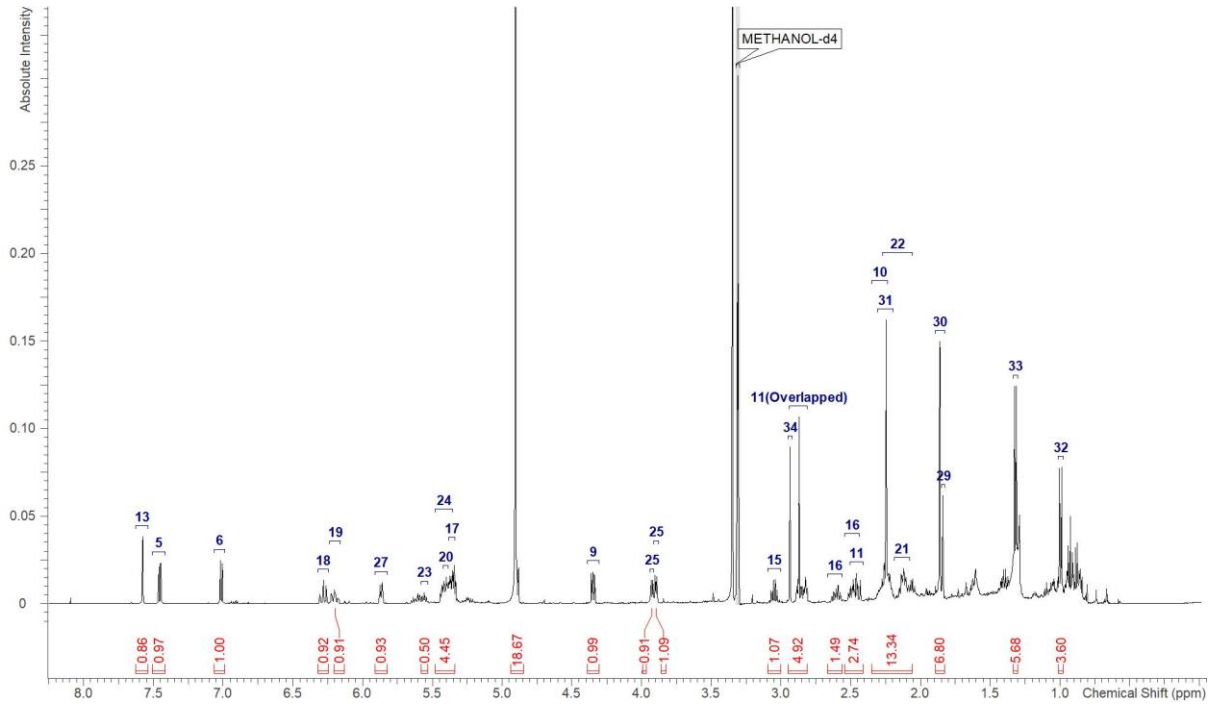


Figure S3. 11. *Ajudazol C (**3**) - ^{1}H NMR recorded in methanol-*d_4 *at 500 MHz*

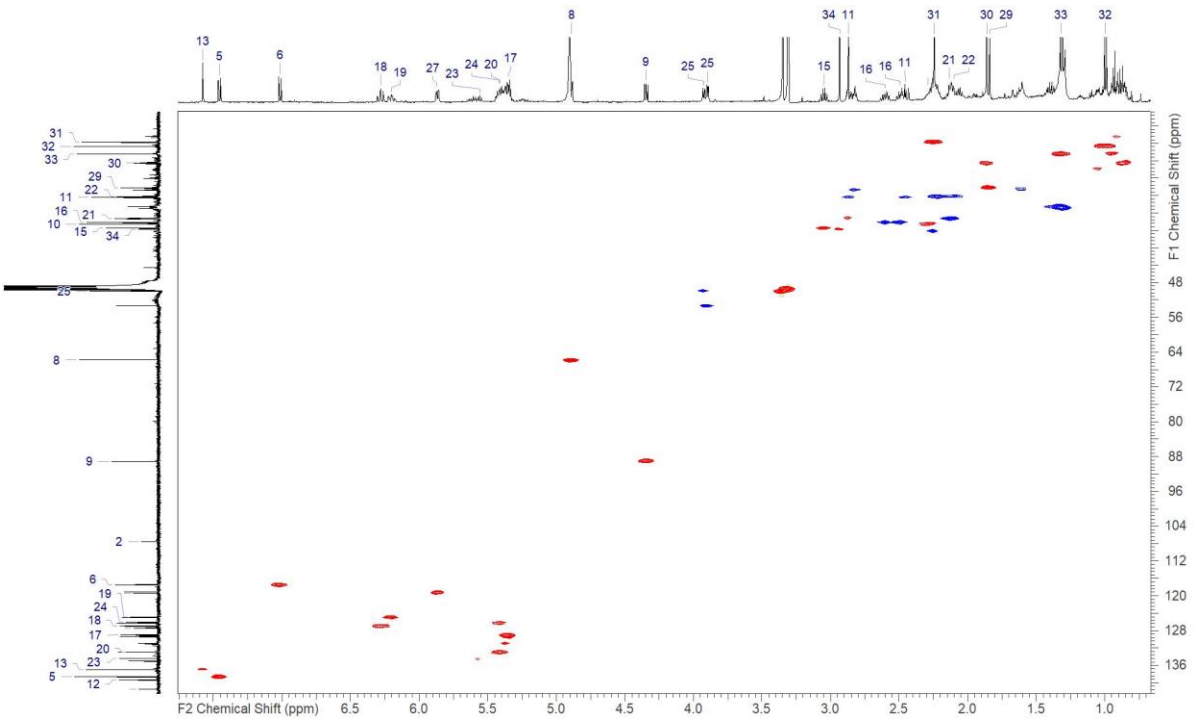


Figure S3. 12. *Ajudazol C (**3**) - HSQC NMR recorded in methanol-*d_4 *at 125/500 (F1/F2) MHz*

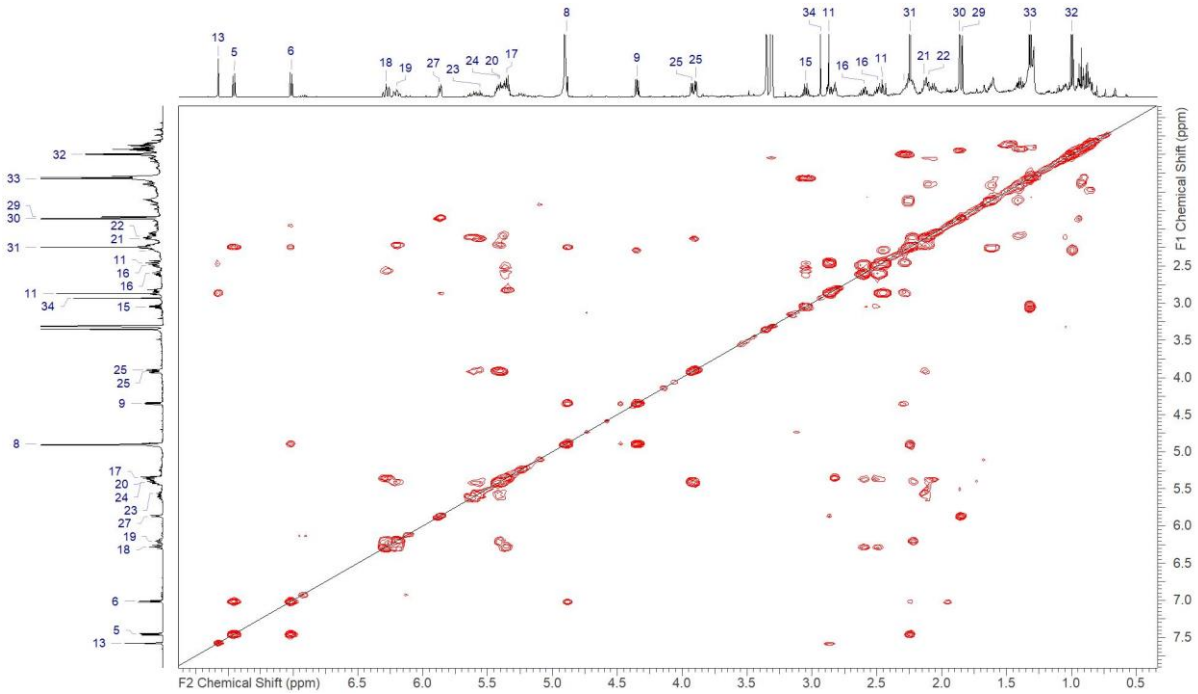


Figure S3. 13. *Ajudazol C (**3**) - COSY NMR recorded in methanol-*d_4 *at 500 MHz*

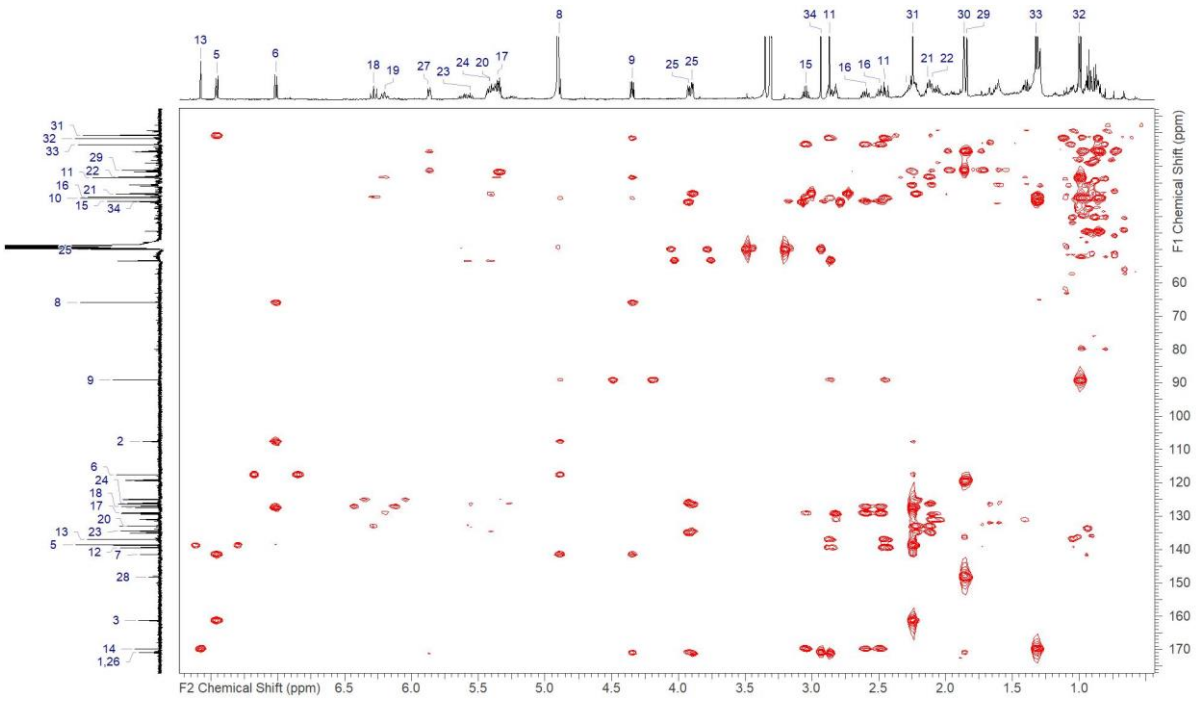


Figure S3. 14. *Ajudazol C (**3**) - HMBC NMR recorded in methanol-*d_4 *at 125/500 (F1/F2) MHz*

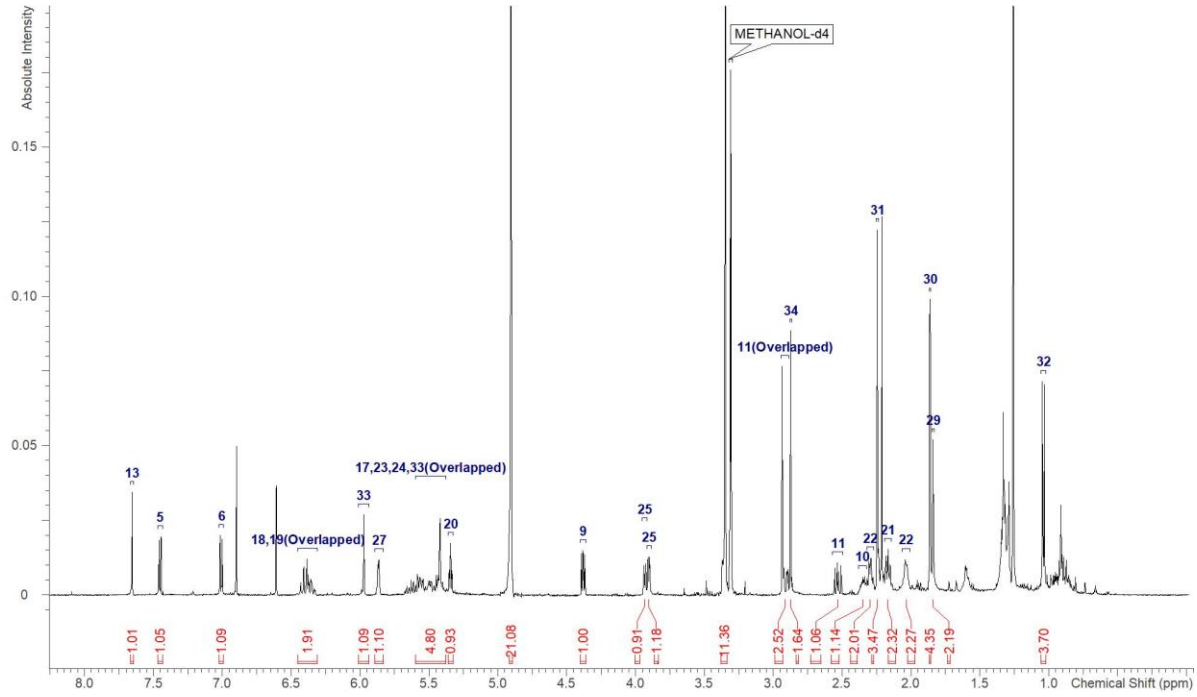


Figure S3. 15. *Ajudazol D (**4**) – ^{1}H NMR recorded in methanol-*d_4 *at 500 MHz*

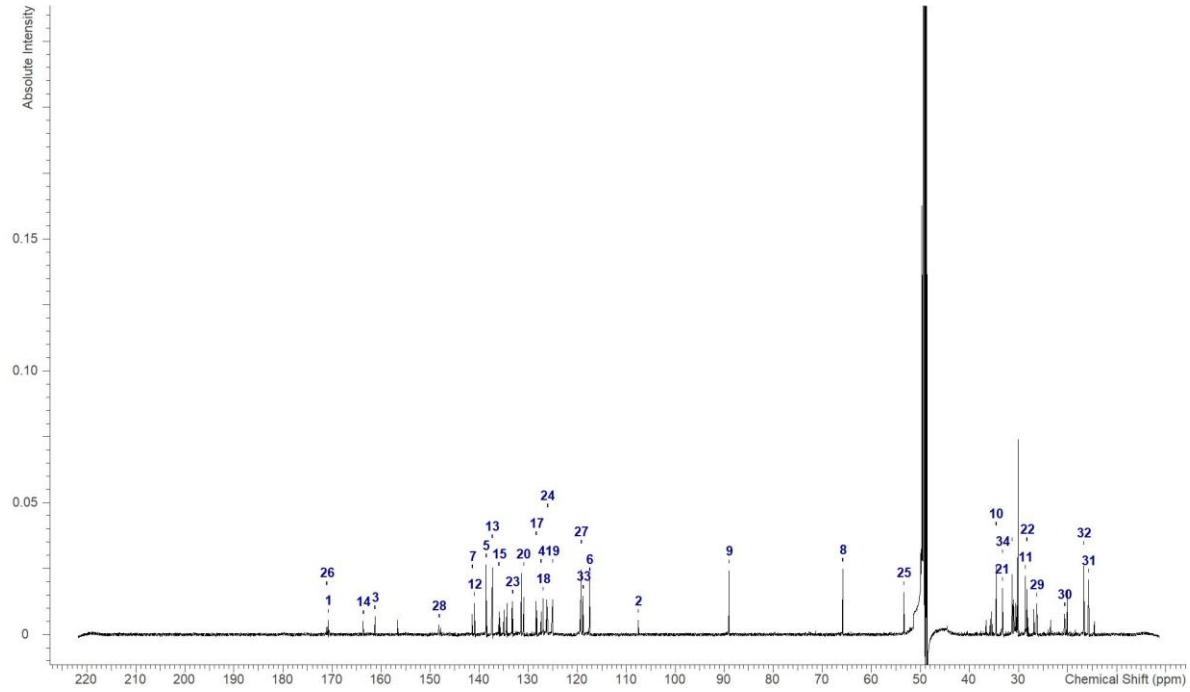


Figure S3. 16. *Ajudazol D (**4**) – ^{13}C NMR recorded in methanol-d$_4$ at 125 MHz*

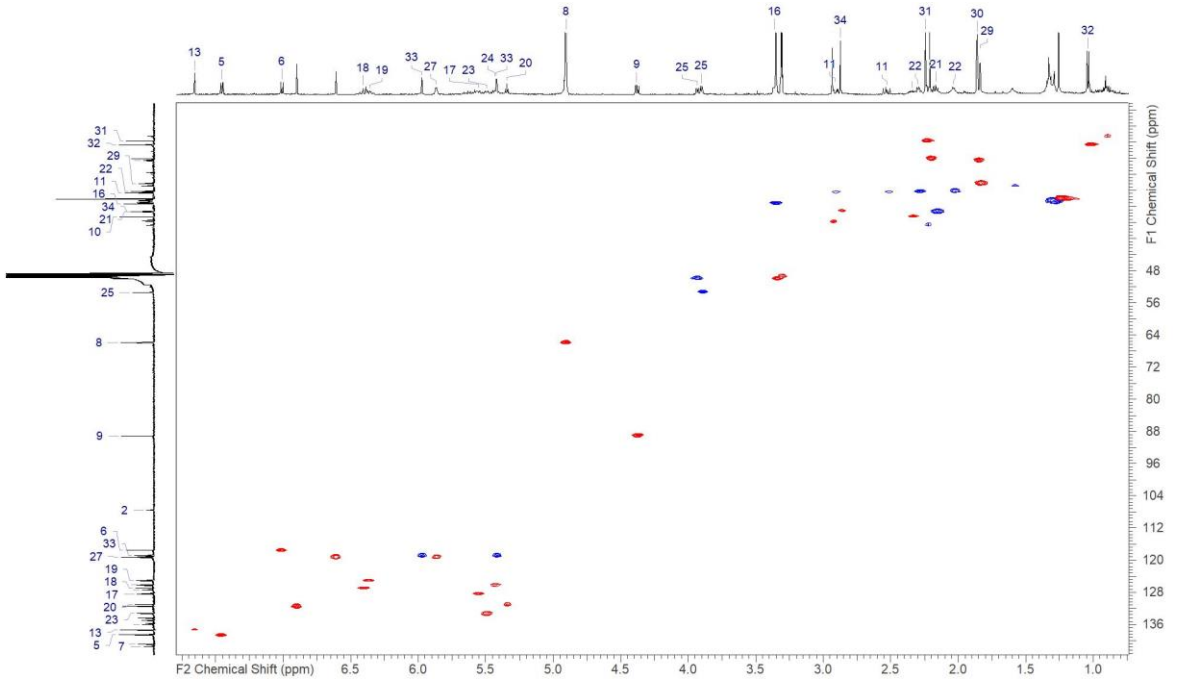


Figure S3. 17. *Ajudazol D (**4**) – HSQC NMR recorded in methanol-d$_4$ at 125/500 (F1/F2) MHz*

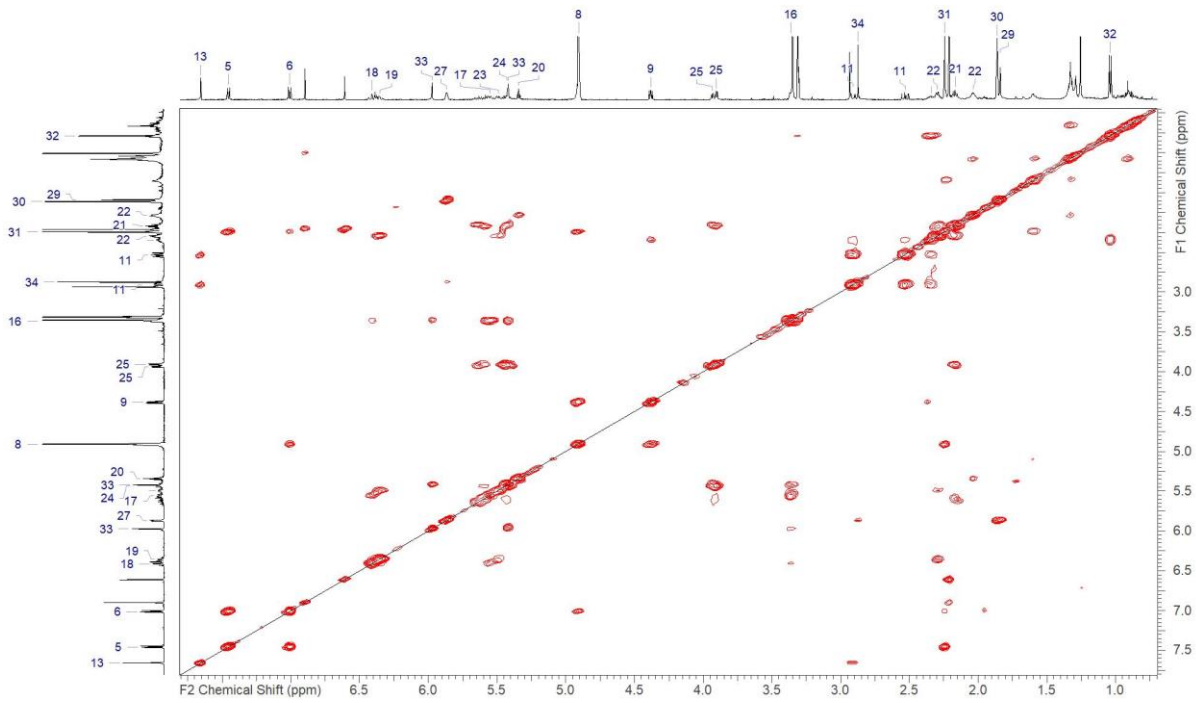


Figure S3. 18. *Ajudazol D (**4**) – COSY NMR recorded in methanol-*d_4 *at 500 MHz*

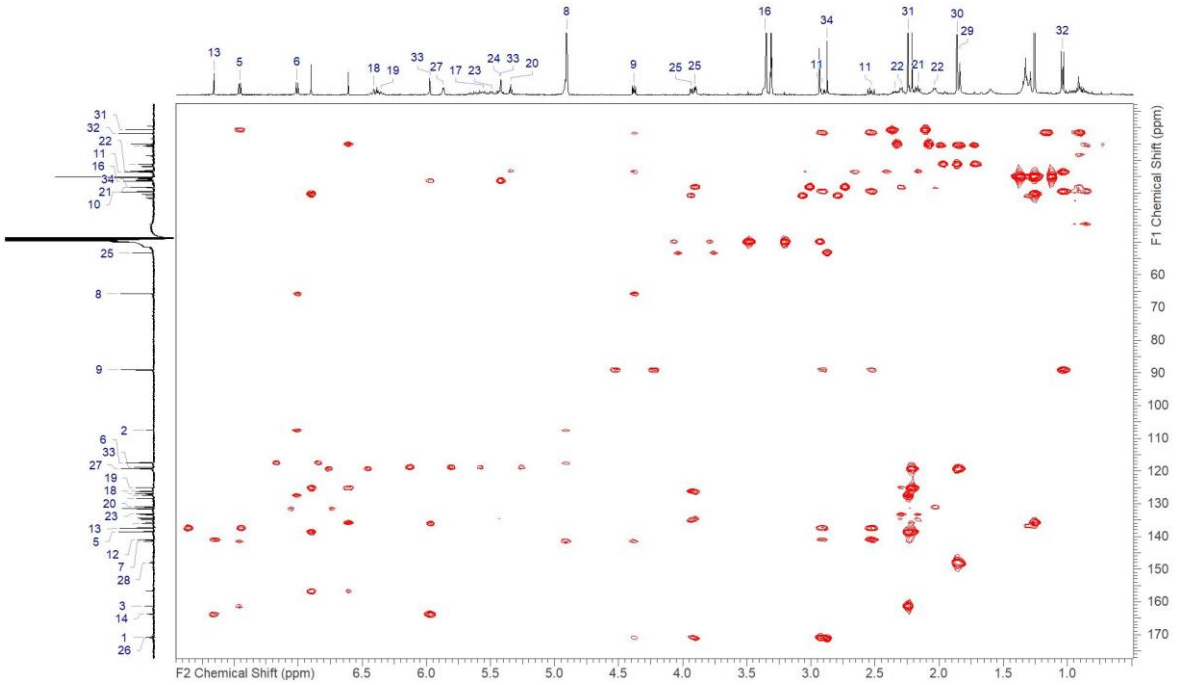


Figure S3. 19. *Ajudazol D (**4**) – HMBC NMR recorded in methanol-*d_4 *at 125/500 (F1/F2) MHz*

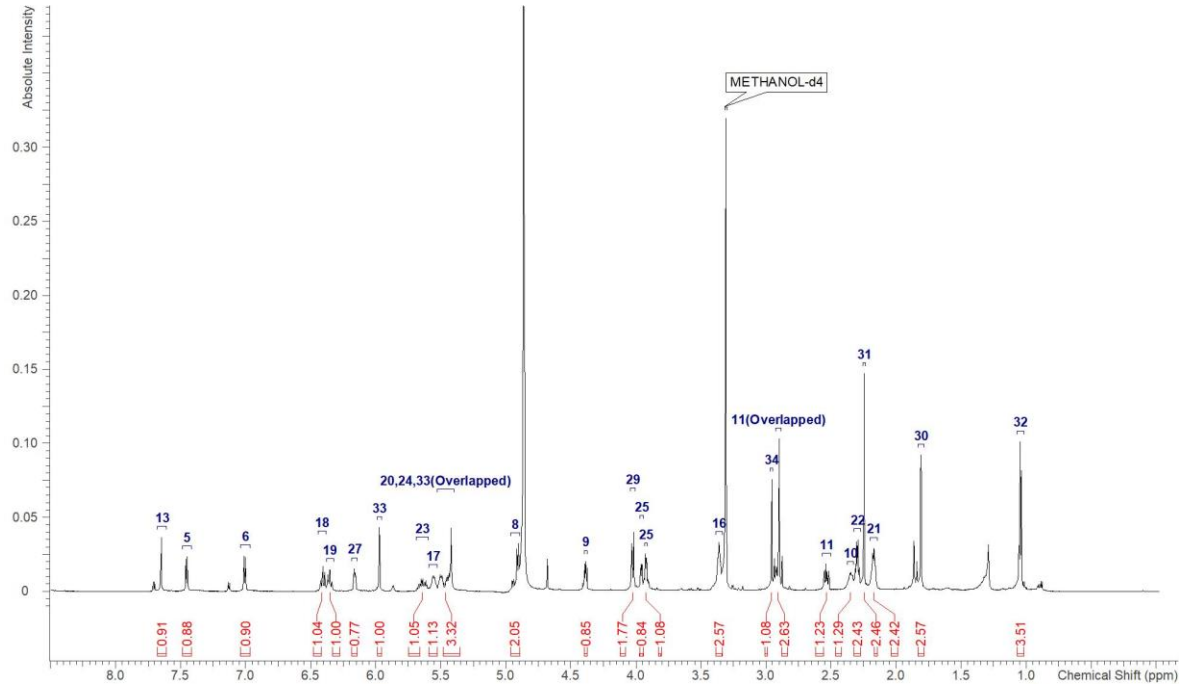


Figure S3. 20. *Ajudazol E (**5**) - ^{1}H NMR recorded in methanol-*d_4 *at 700 MHz*

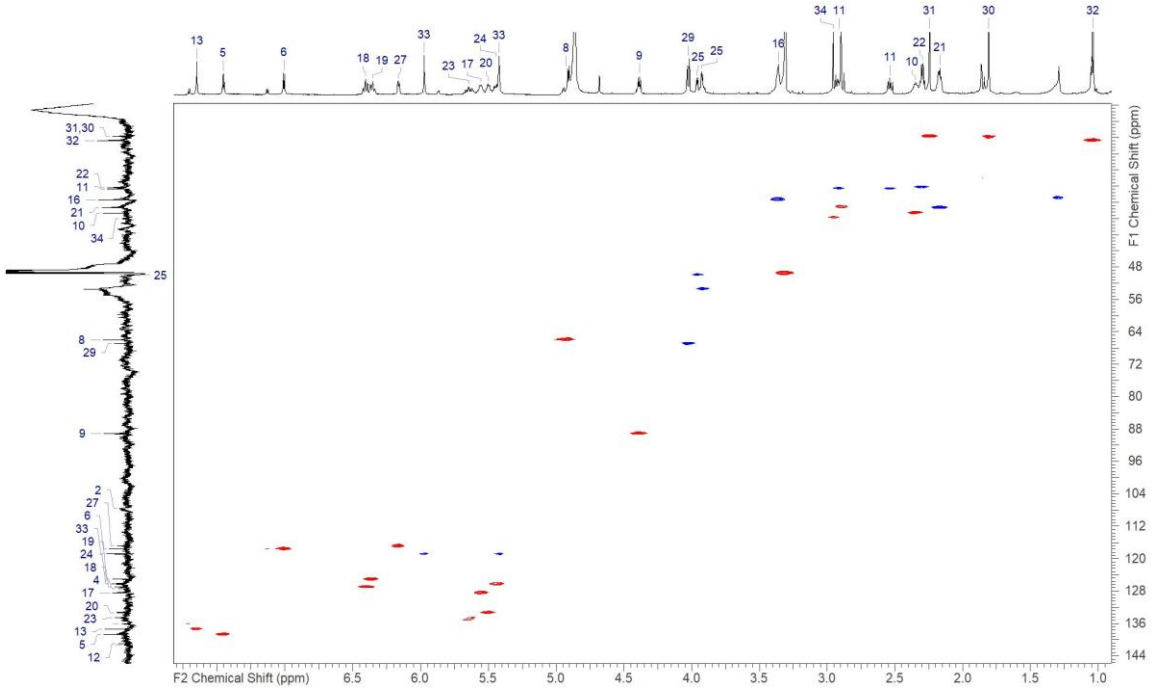


Figure S3. 21. *Ajudazol E (**5**) – HSQC NMR recorded in methanol-*d_4 *at 175/700 (F1/F2) MHz*

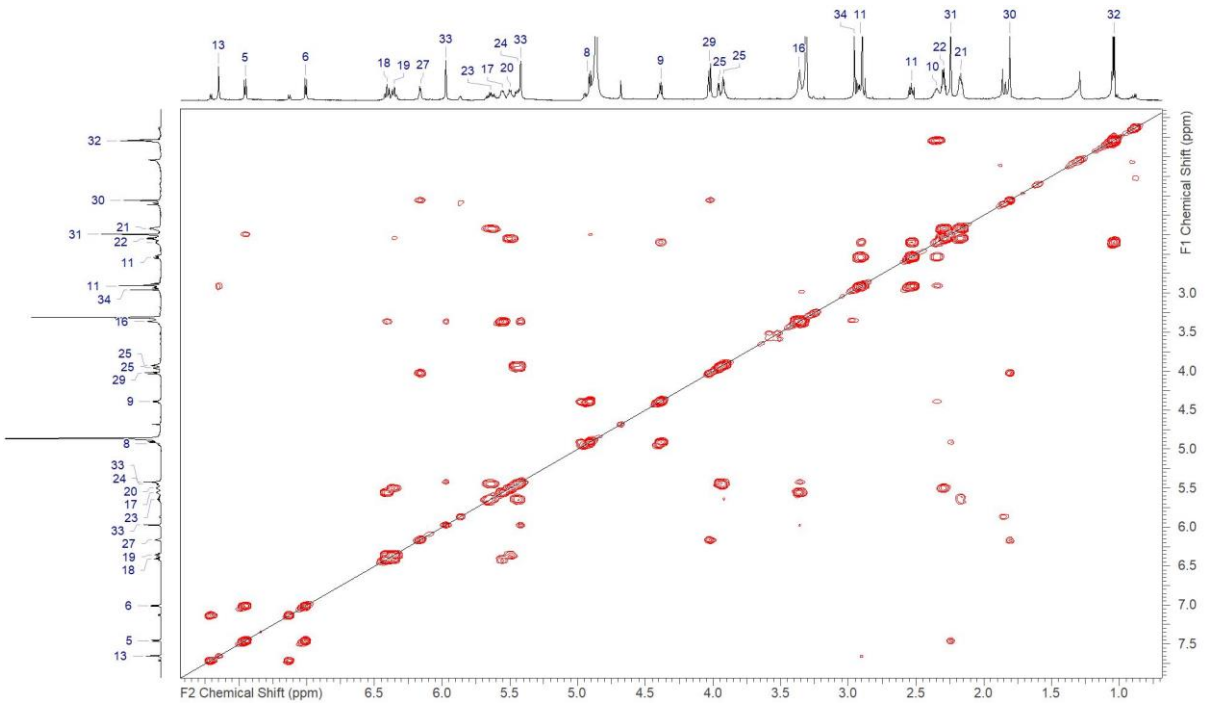


Figure S3. 22. *Ajudazol E (**5**) – COSY NMR recorded in methanol-*d_4 *at 700 MHz*

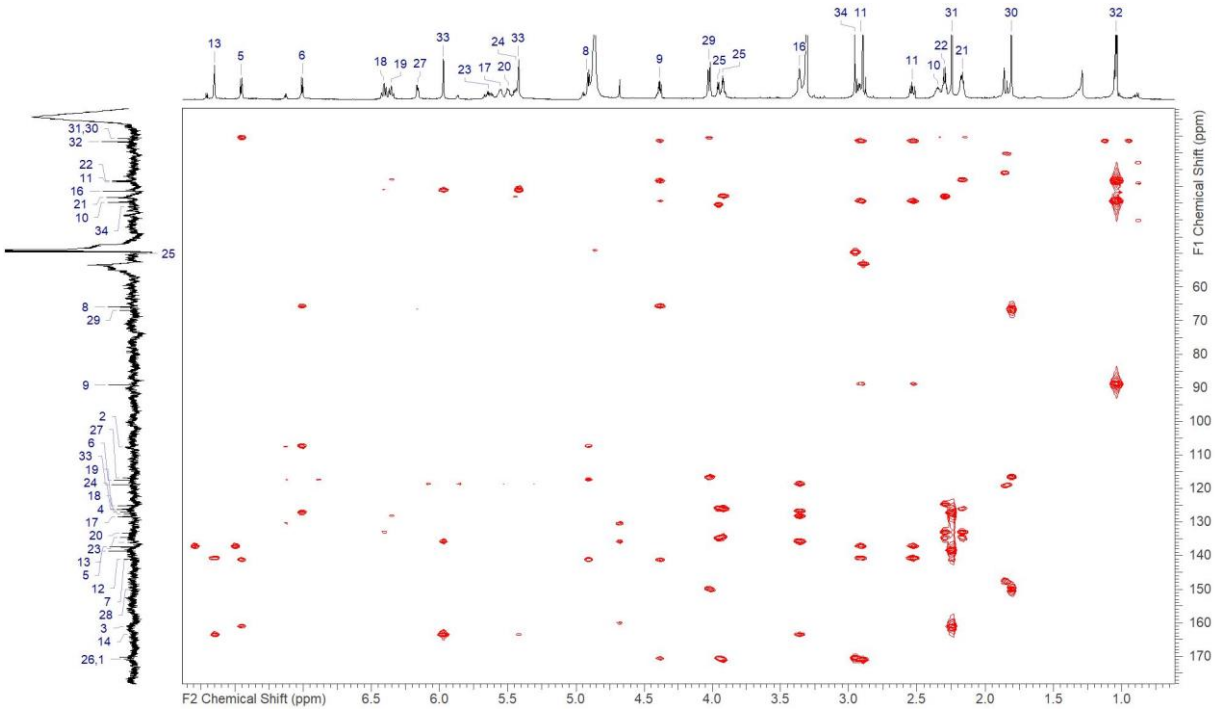


Figure S3. 23. *Ajudazol E (**5**) – HMBC NMR recorded in methanol-*d_4 *at 175/700 (F1/F2) MHz*

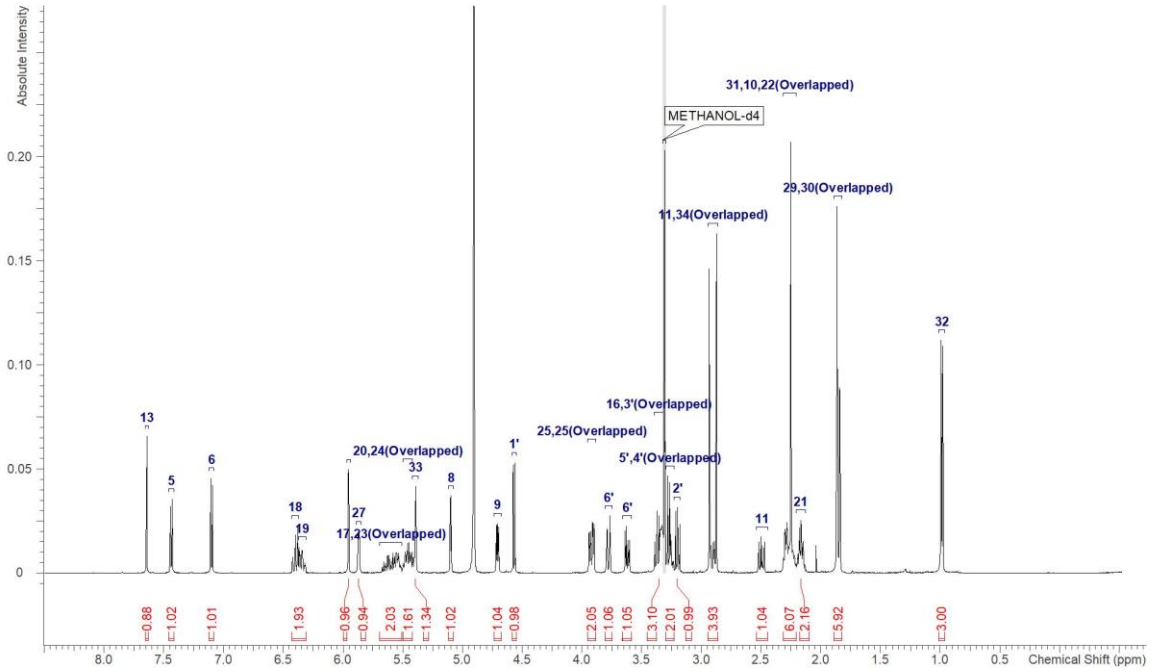


Figure S3. 24. *Ajudazol F (**6**) - ^{1}H NMR recorded in methanol-d_4 at 500 MHz*

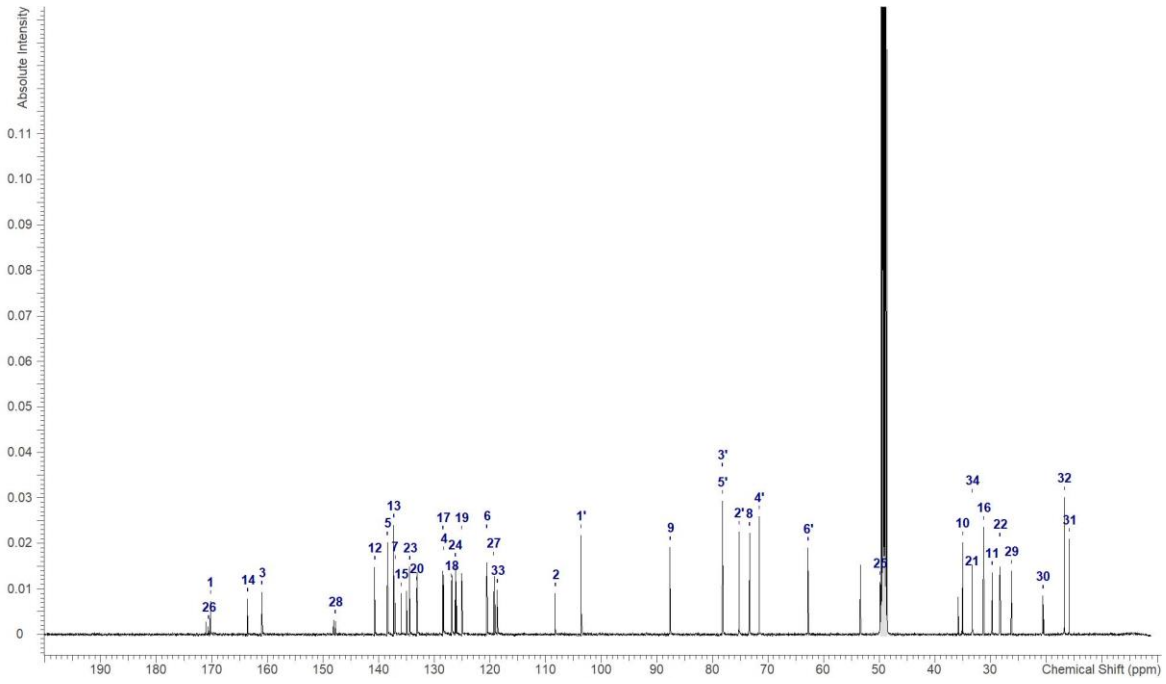


Figure S3. 25. *Ajudazol F (**6**) – ^{13}C NMR recorded in methanol-d_4 at 125 MHz*

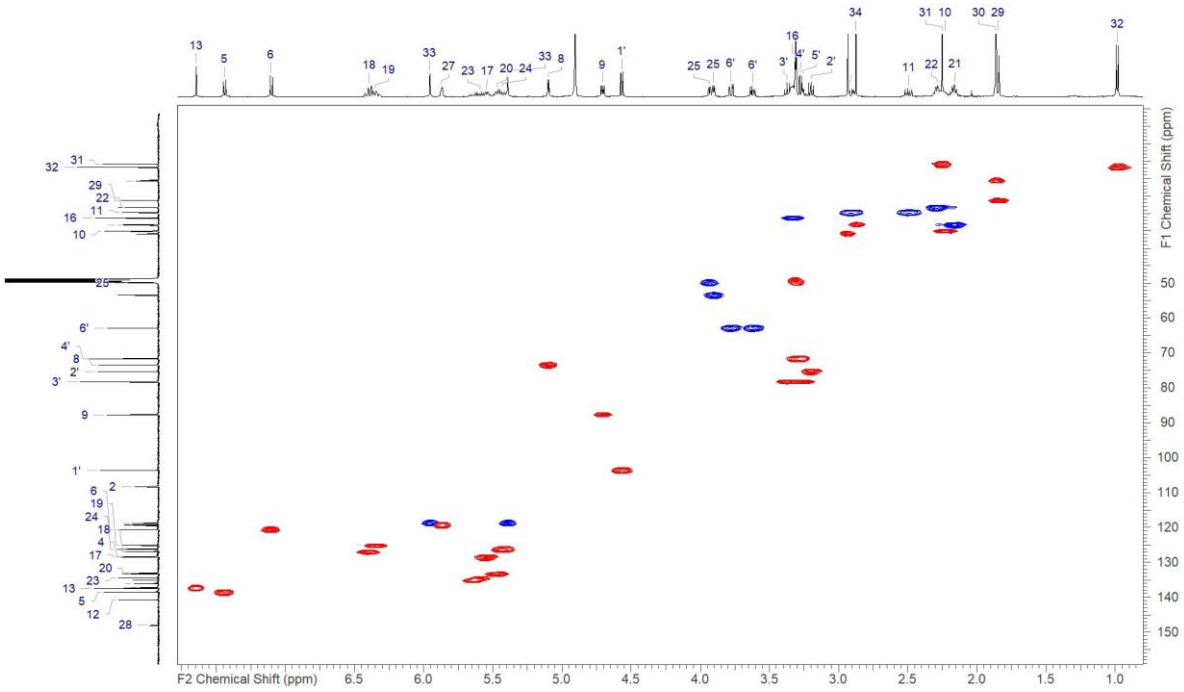


Figure S3. 26. *Ajudazol F (**6**) – HSQC NMR recorded in methanol-*d_4 *at 125/500 (F1/F2) MHz*

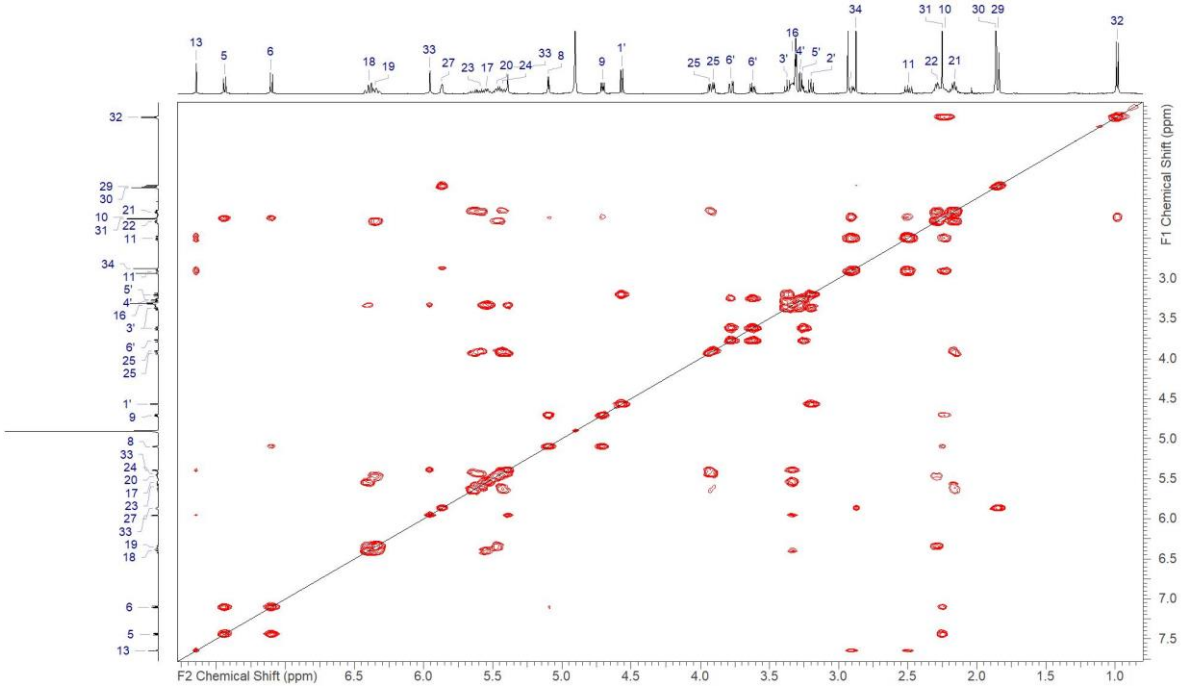


Figure S3. 27. *Ajudazol F (**6**) – COSY NMR recorded in methanol-*d_4 *at 500 MHz*

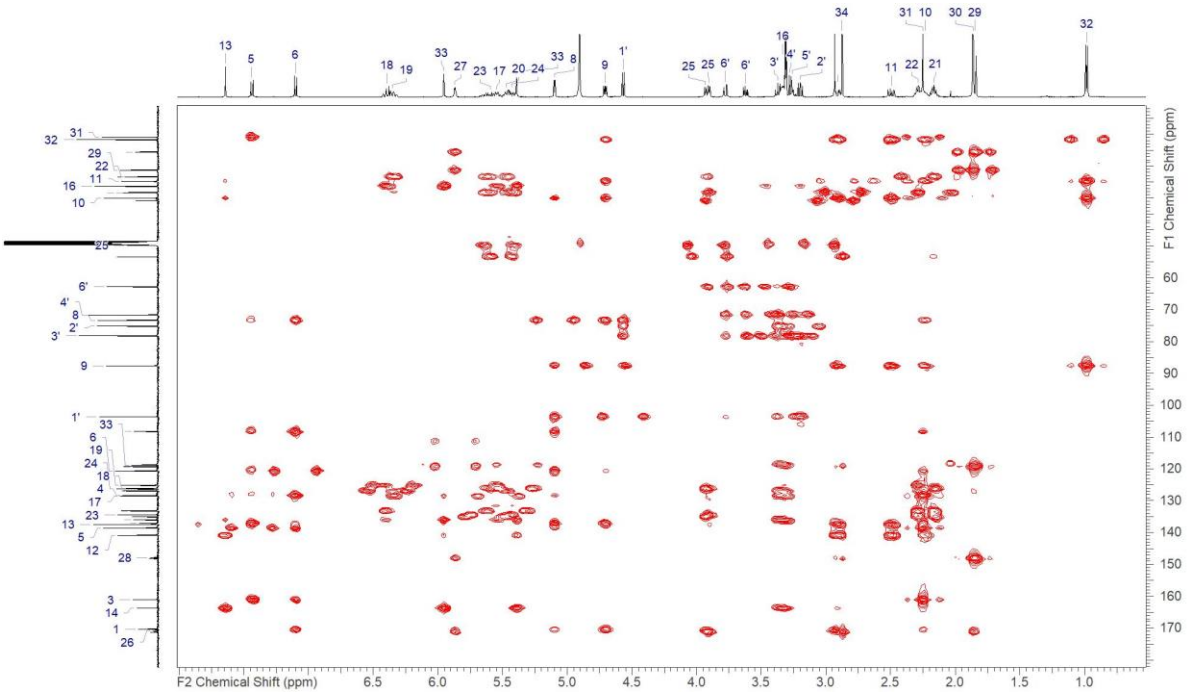


Figure S3. 28. *Ajudazol F (**6**) – HMBC NMR recorded in methanol-*d_4 *at 125/500 (F1/F2) MHz*

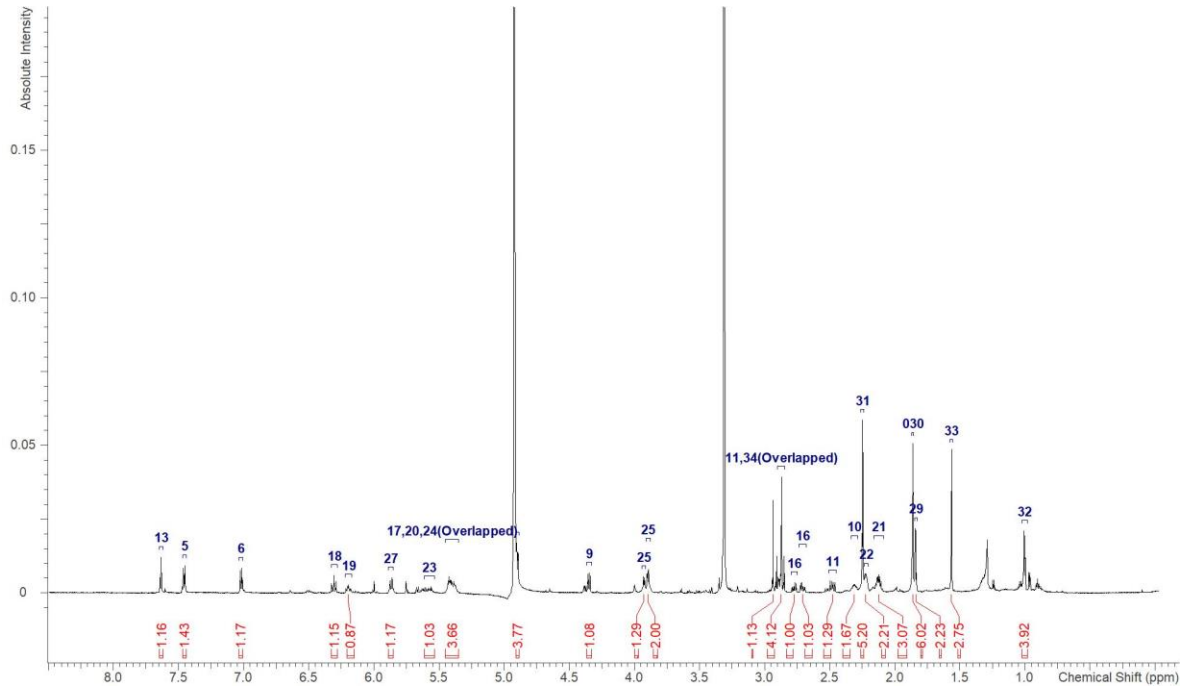


Figure S3. 29. *Ajudazol G (**7**) - ^{1}H NMR recorded in methanol-*d_4 *at 700 MHz*

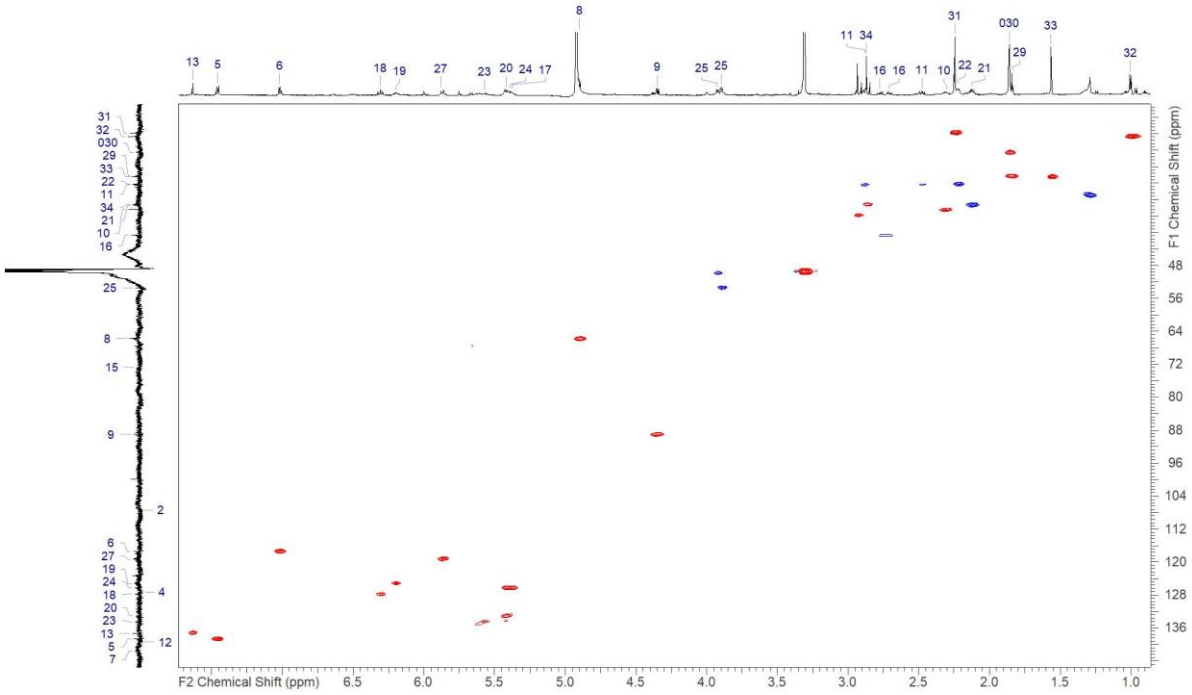


Figure S3. 30. *Ajudazol G (**7**) – HSQC recorded in methanol-*d_4 *at 175/700 (F1/F2) MHz*

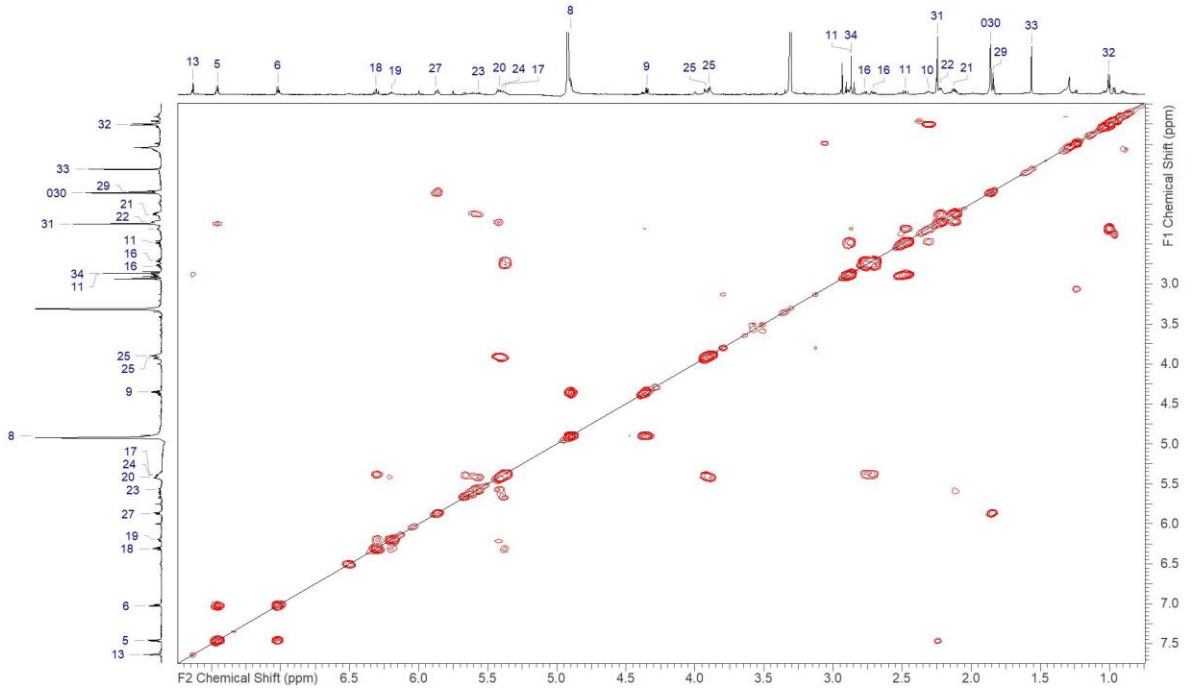


Figure S3. 31. *Ajudazol G (**7**) – COSY recorded in methanol-*d_4 *at 700 MHz*

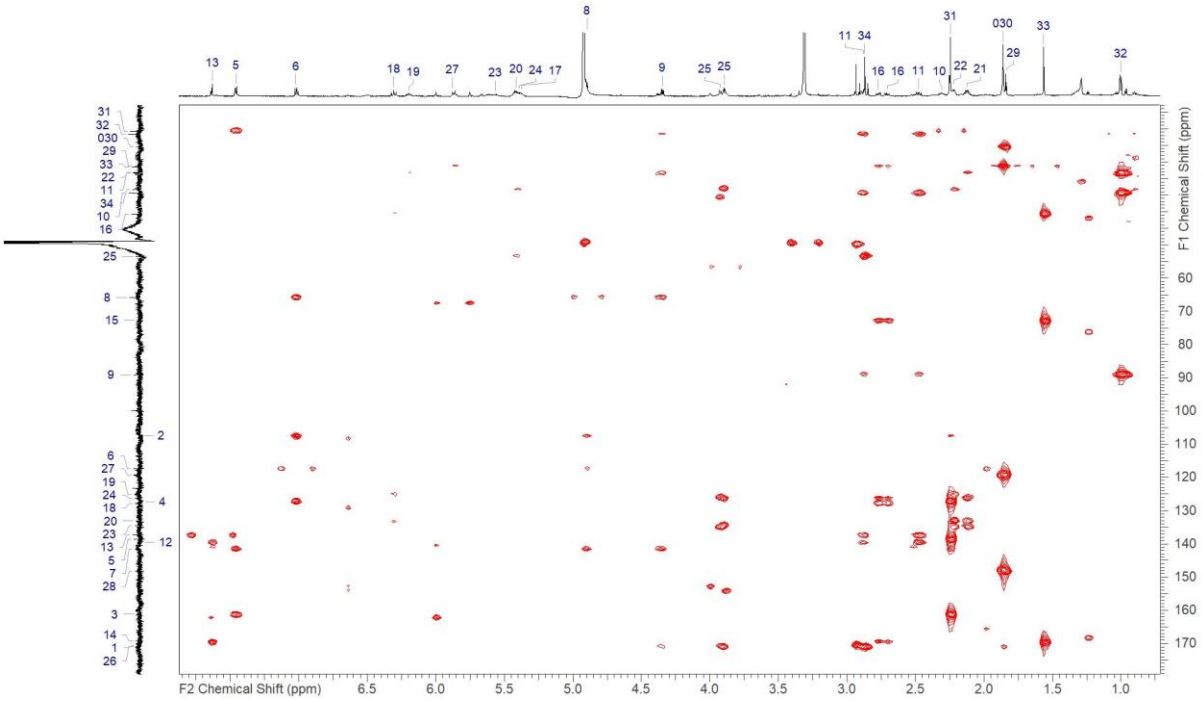


Figure S3. 32. *Ajudazol G (**7**) - HMBC recorded in methanol-*d_4 *at 175/700 (F1/F2) MHz*

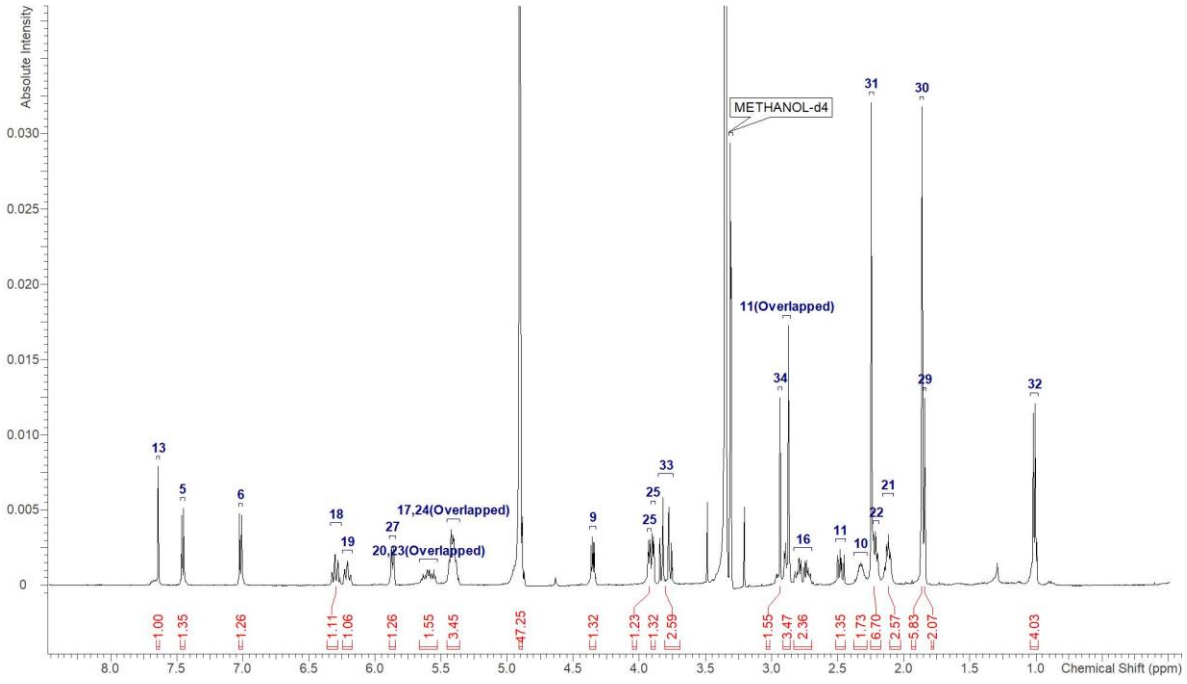


Figure S3. 33. *Ajudazol H (**8**) - ^{1}H NMR recorded in methanol-*d_4 *at 500 MHz*

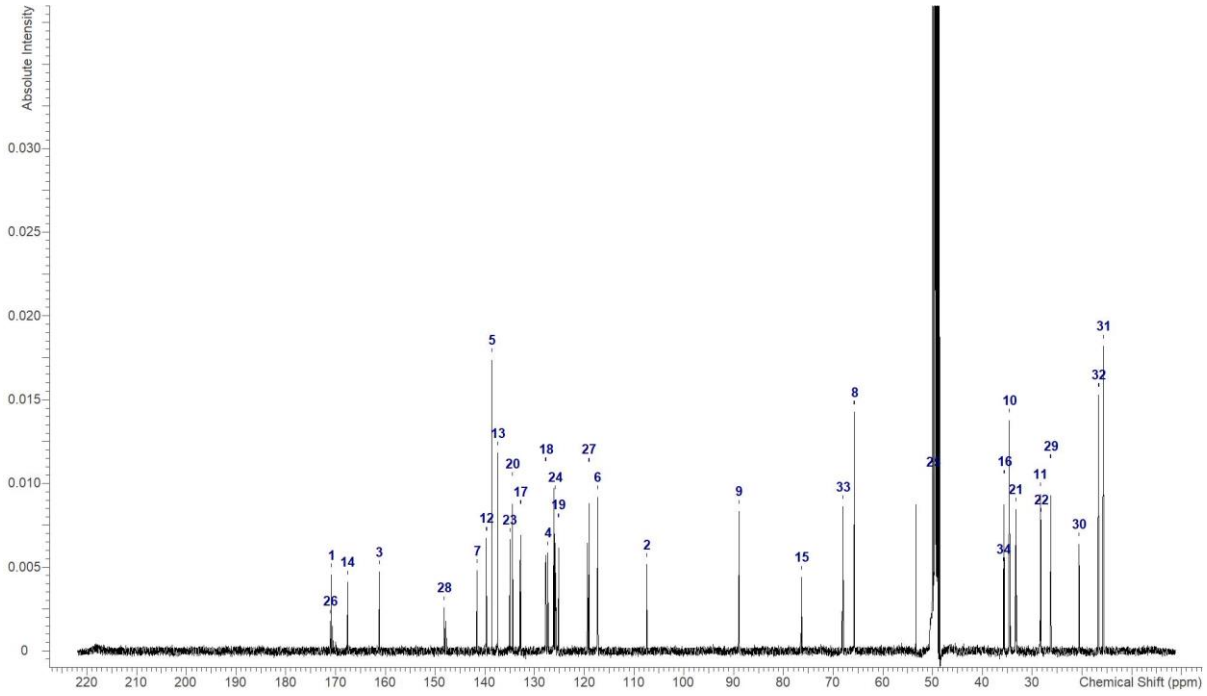


Figure S3. 34. *Ajudazol H (**8**) - ^{13}C NMR recorded in methanol-d_4 at 125 MHz*

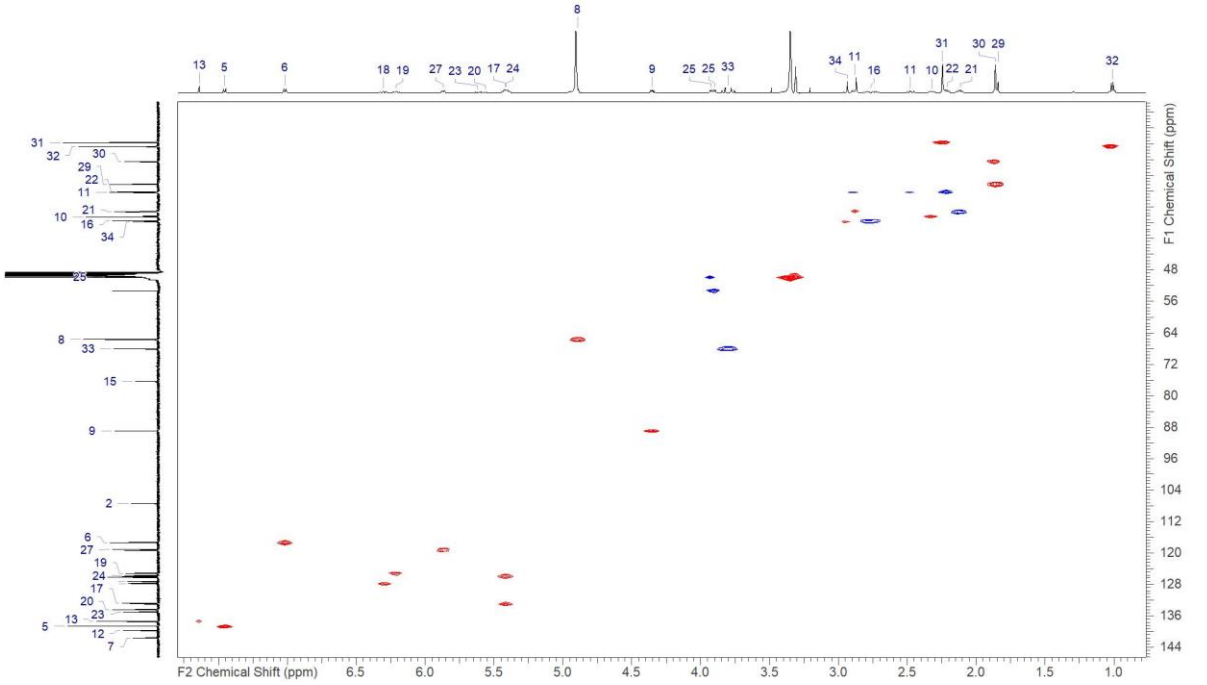


Figure S3. 35. *Ajudazol H (**8**) - HSQC NMR recorded in methanol-d_4 at 125/500 (F1/F2) MHz*

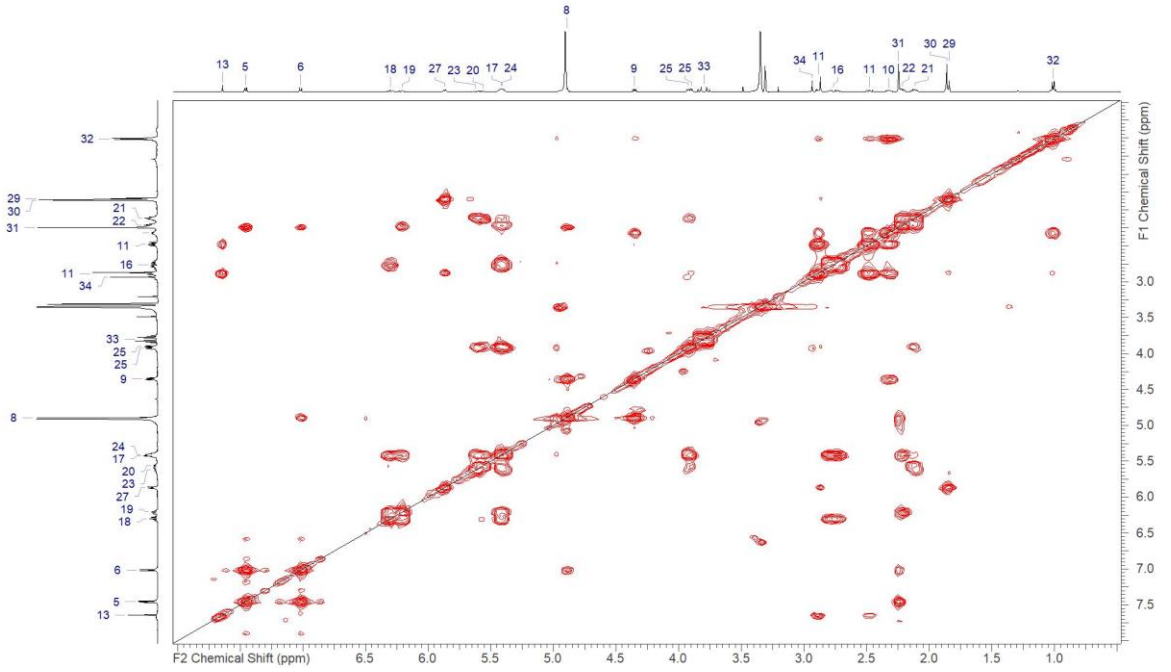


Figure S3. 36. *Ajudazol H (**8**) - COSY NMR recorded in methanol-*d_4 *at 500 MHz*

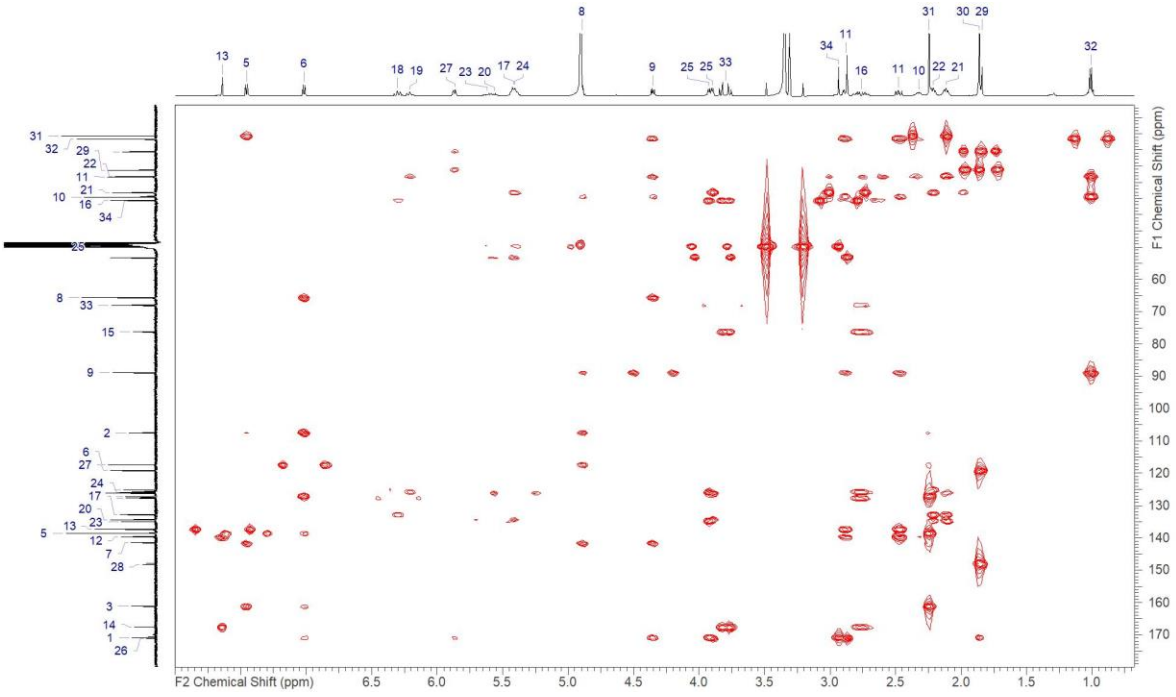


Figure S3. 37. *Ajudazol H (**8**) - HMBC NMR recorded in methanol-*d_4 *at 125/500 (F1/F2) MHz*

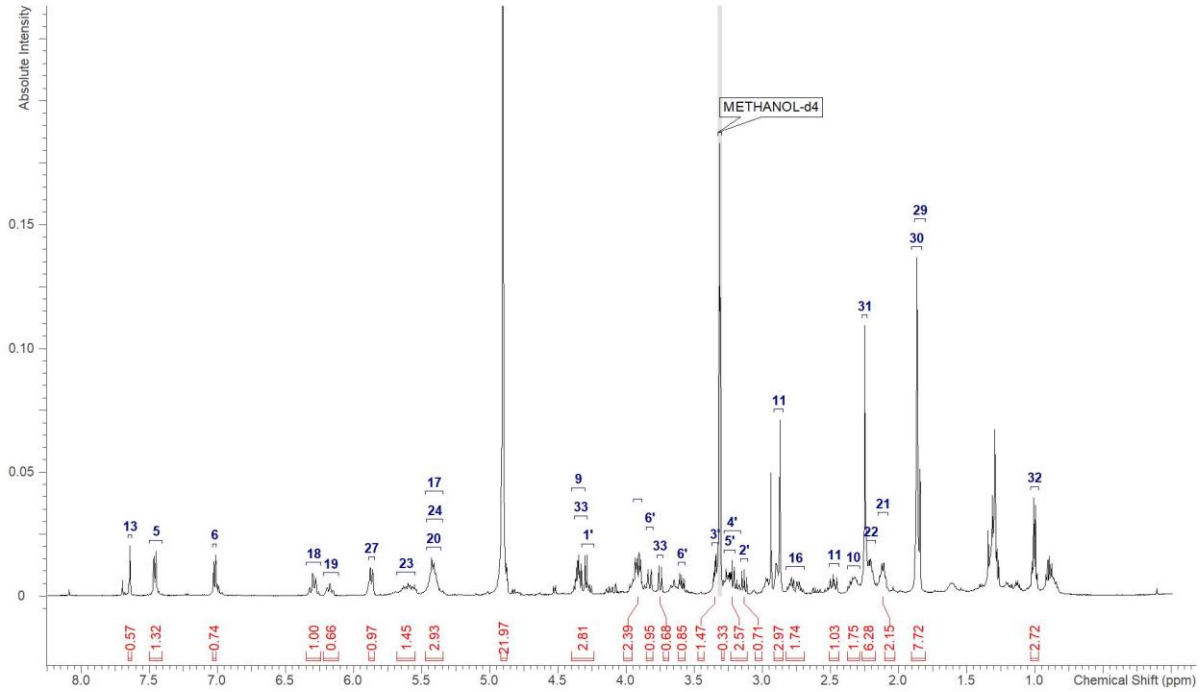


Figure S3. 38. *Ajudazol I (**9**) - ^{1}H NMR recorded in methanol-d_4 at 500 MHz*

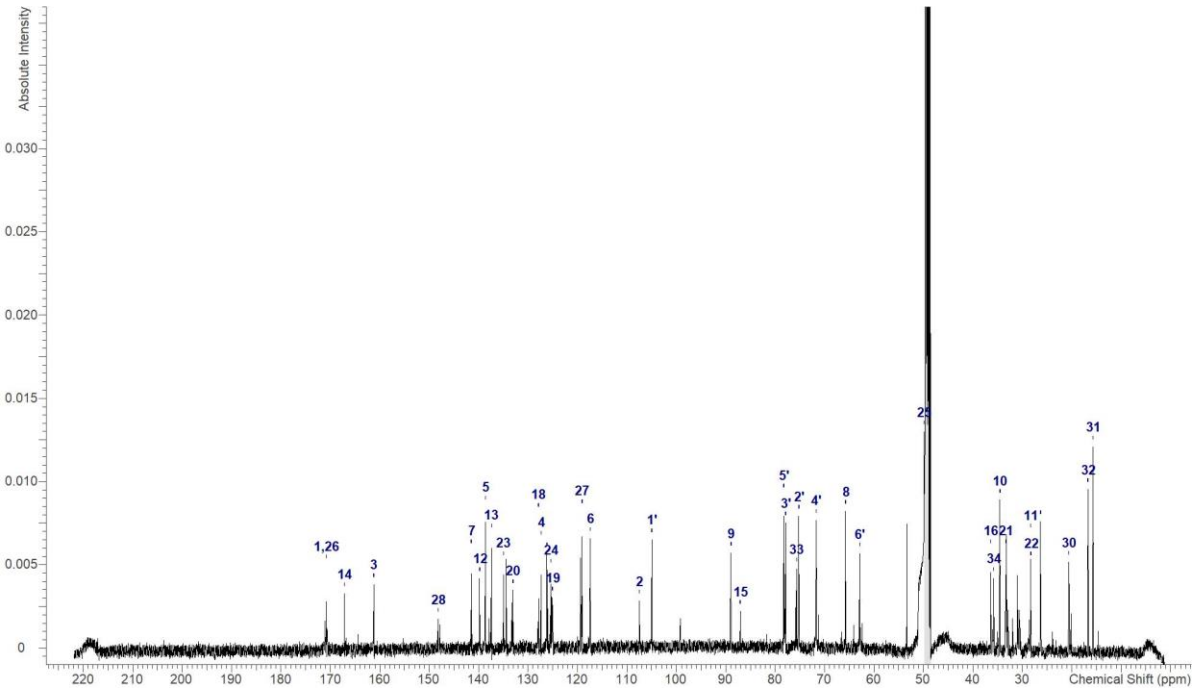


Figure S3. 39. *Ajudazol I (**9**) - ^{13}C NMR recorded in methanol-d_4 at 125 MHz*

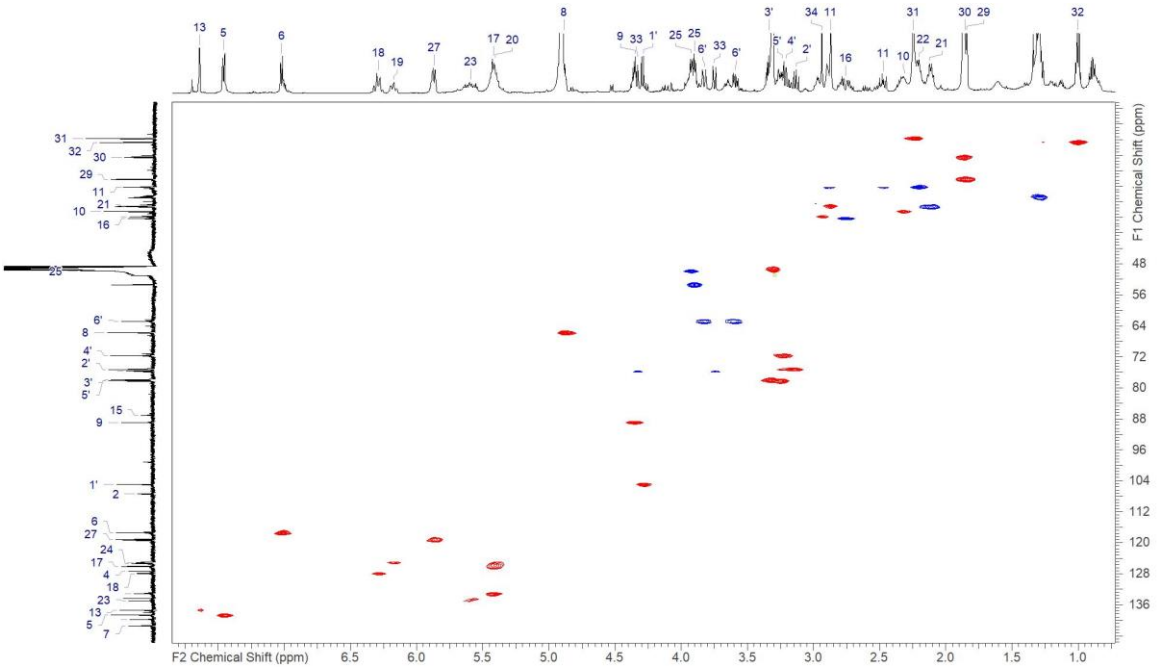


Figure S3. 40. *Ajudazol I (**9**) – HSQC NMR recorded in methanol-*d_4 *at 125/500 (F1/F2) MHz*

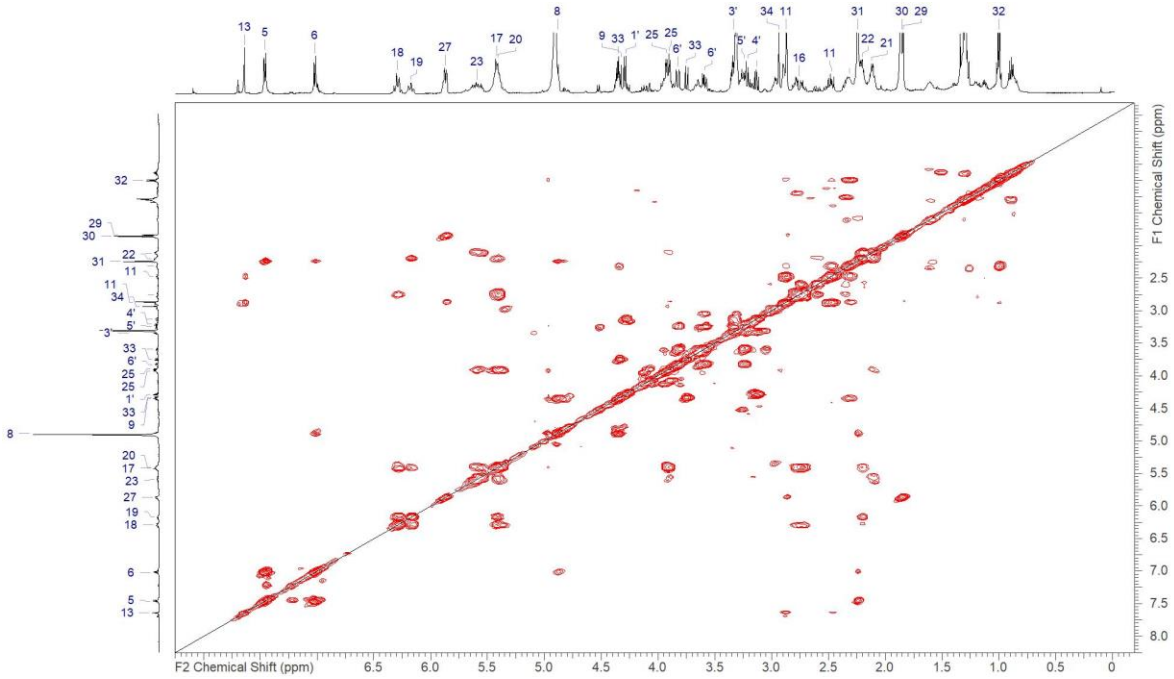


Figure S3. 41. *Ajudazol I (**9**) – COSY NMR recorded in methanol-*d_4 *at 500 MHz*

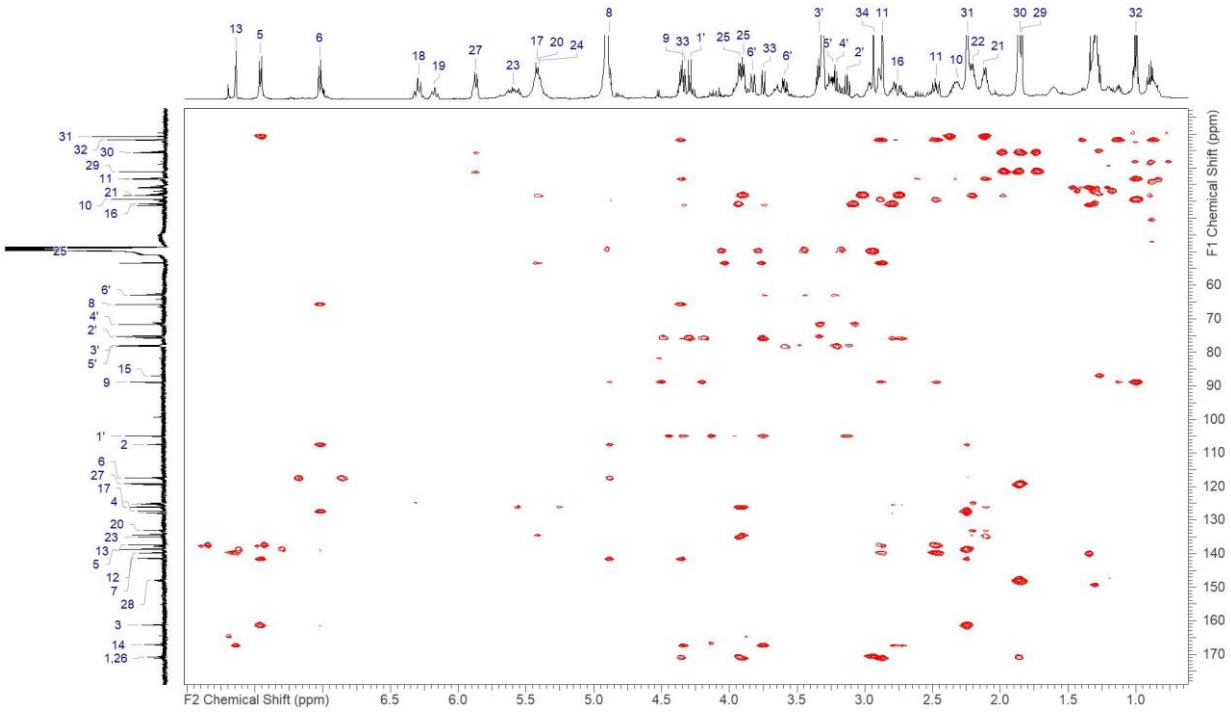


Figure S3. 42. *Ajudazol I (**9**) – HMBC NMR recorded in methanol-*d_4 *at 125/500 (F1/F2) MHz*

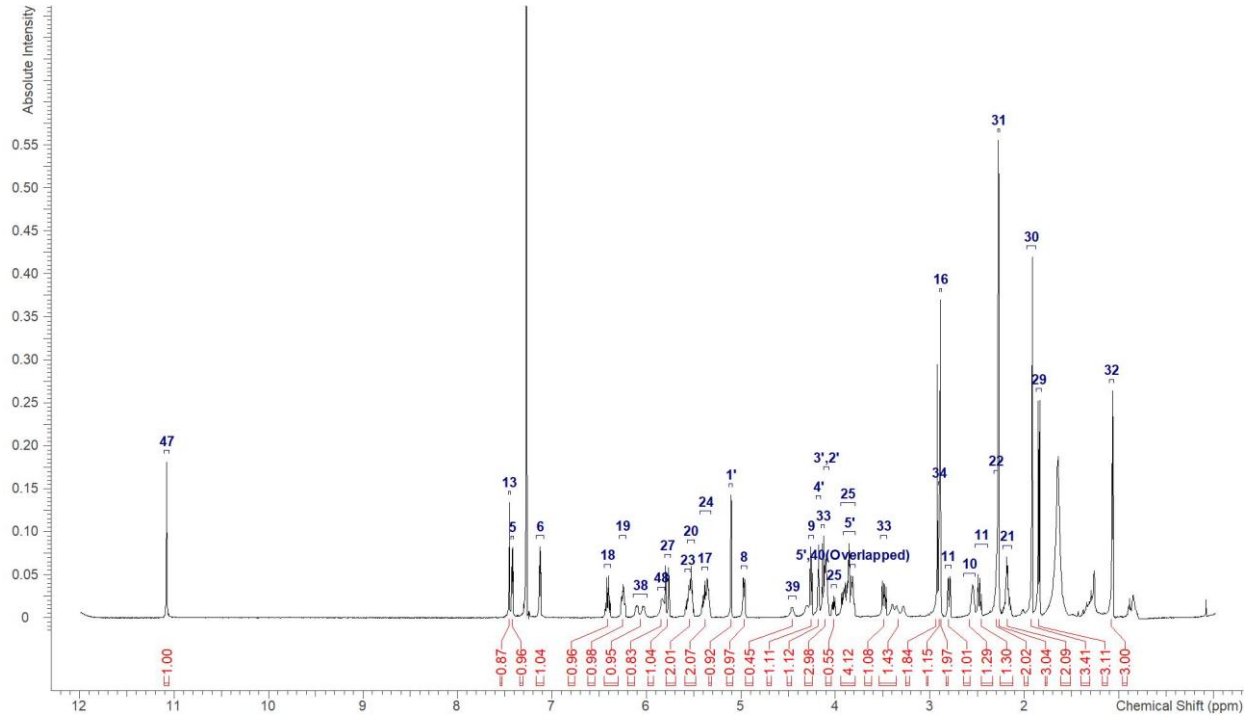


Figure S3. 43. *Ajudazol J (**10**) –* ^{1}*H NMR recorded in chloroform-*d *at 700 MHz*

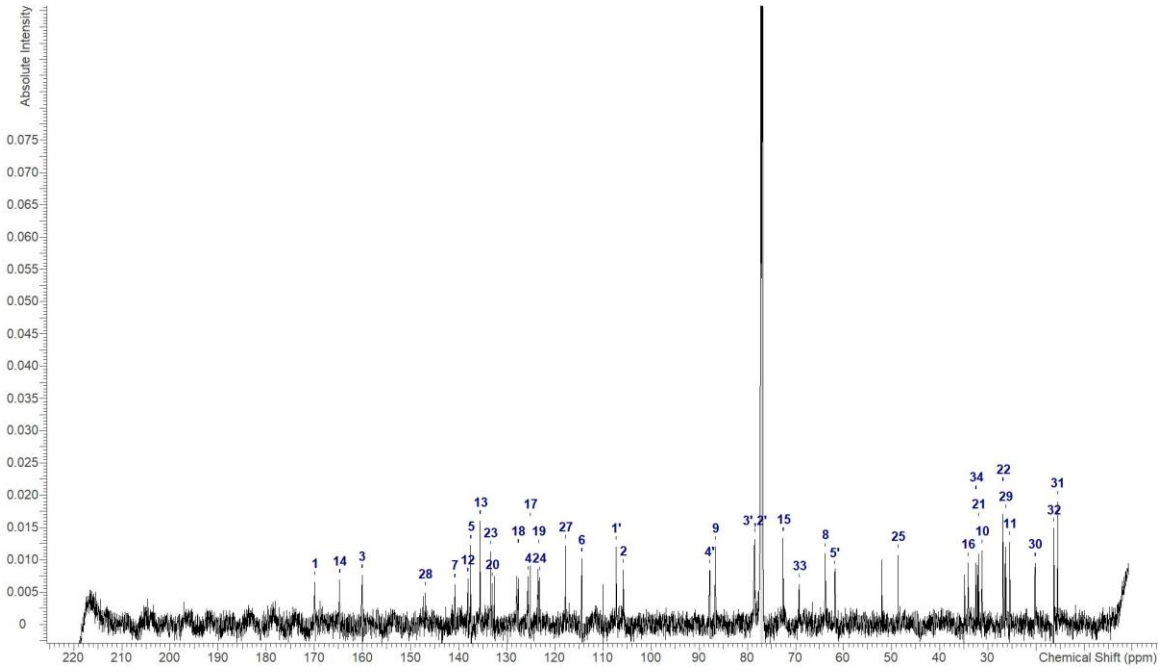


Figure S3. 44. *Ajudazol J (**10**) – ^{13}C NMR recorded in chloroform-*d *at 175 MHz*

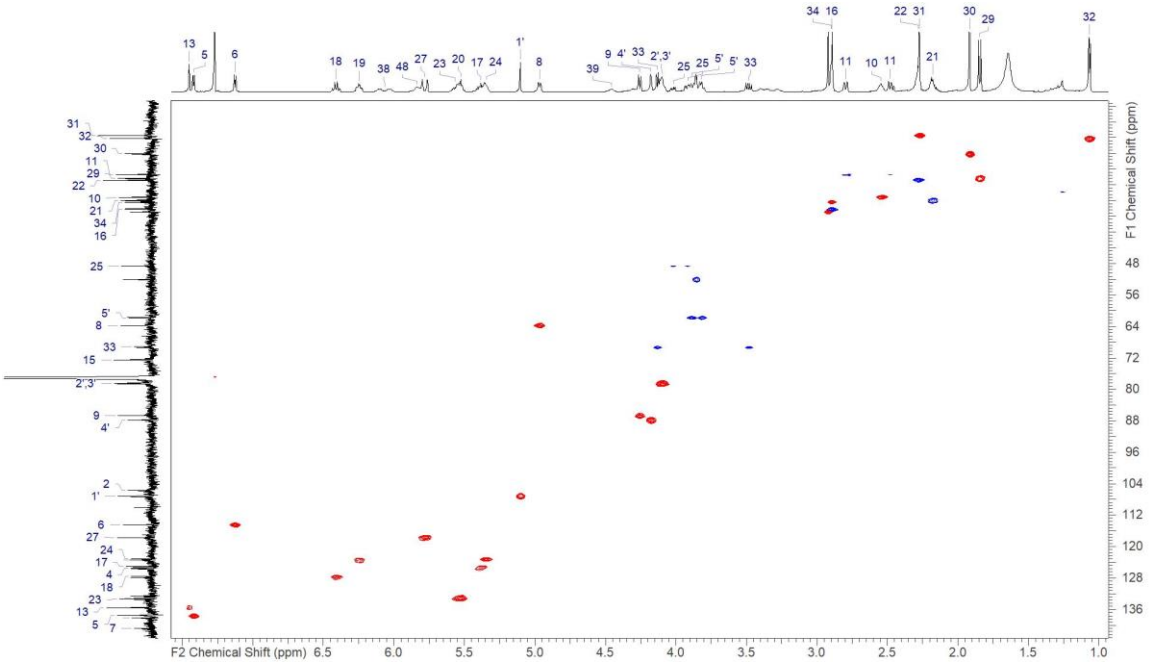


Figure S3. 45. *Ajudazol J (**10**) – HSQC recorded in chloroform-*d *at 175/700 (F1/F2) MHz*

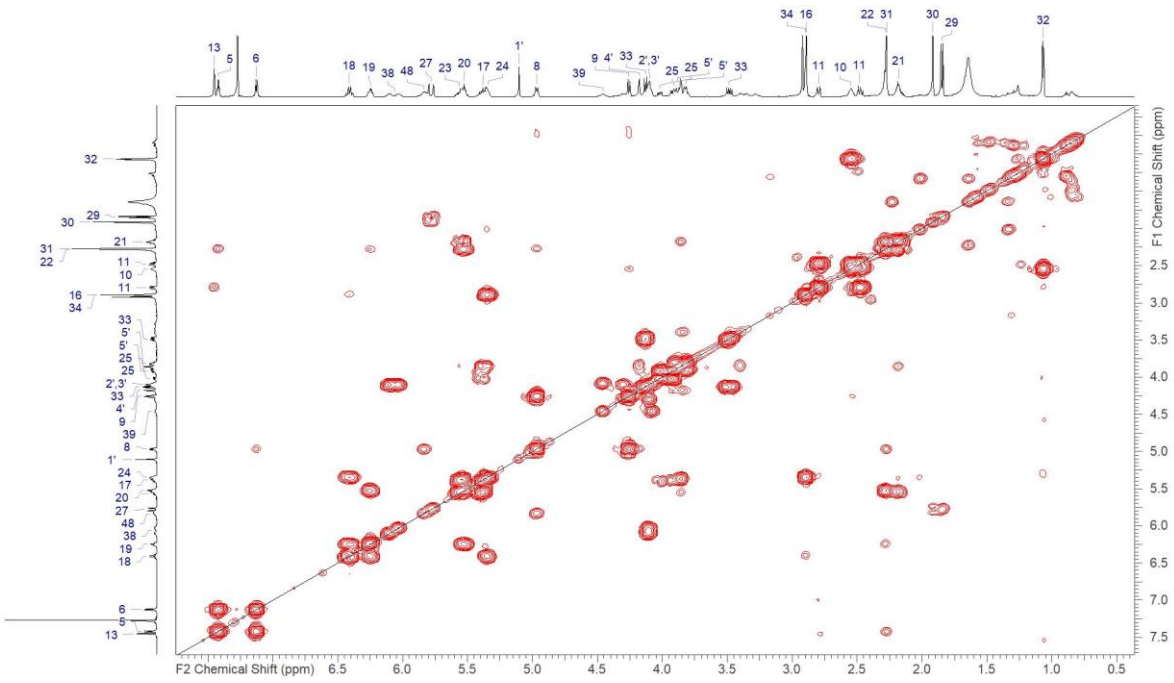


Figure S3. 46. *Ajudazol J (**10**) – COSY recorded in chloroform-*d *at 700 MHz*

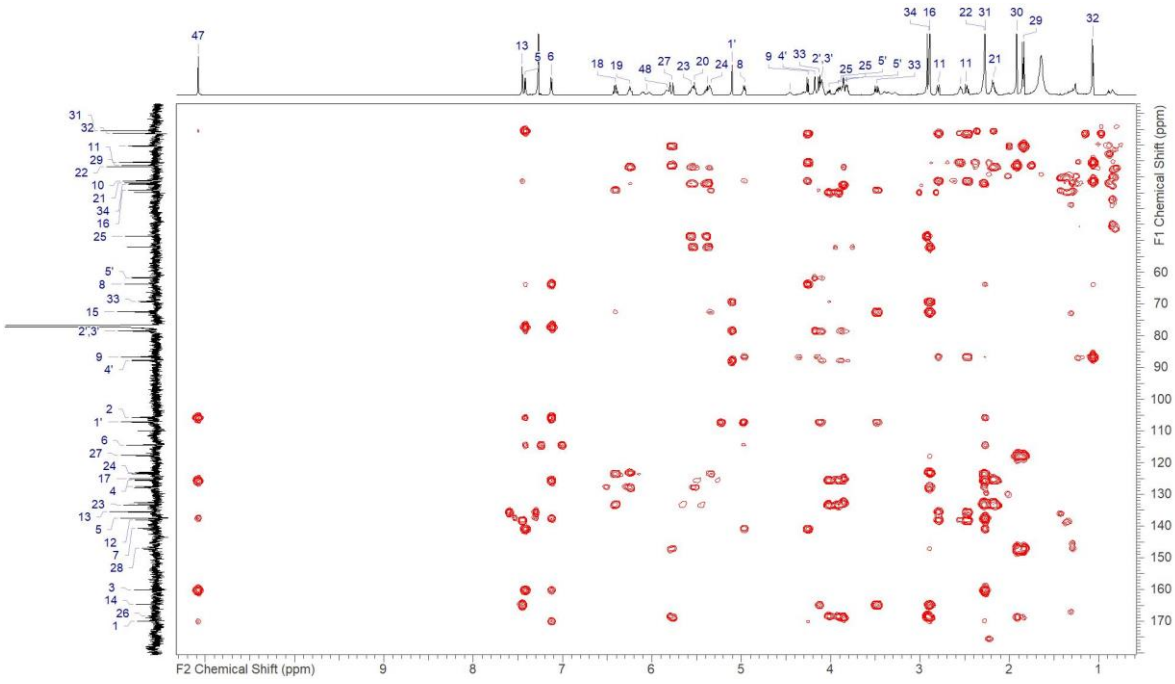


Figure S3. 47. *Ajudazol J (**10**) – HMBC recorded in chloroform-*d *at 175/700 (F1/F2) MHz*

S 3.7 References

(1) Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **2019** (47), W81-W87. DOI: 10.1093/nar/gkz310.

(2) Essig, S.; Schmalzbauer, B.; Bretzke, S.; Scherer, O.; Koeberle, A.; Werz, O.; Müller, R.; Menche, D. Predictive bioinformatic assignment of methyl-bearing stereocenters, total synthesis and an additional molecular target of ajudazol B. *J. Org. Chem.* **2016**, *81* (4), 1333–1357. DOI: 10.1021/acs.joc.5b02844.

4. Cystopipecotides

Cystopipecotides: Discovery, Structure Elucidation and Biosynthesis of Myxobacterial Peptides Featuring a 5-hydroxyl-6-hydroxymethylpipecolic acid Building Block

Hu Zeng†, Joy Birkelbach†, Lena Keller, Carsten Volz, Chengzhang Fu, and Rolf Müller

Affiliation

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

Contributions and Acknowledgements

The Author's Effort

This author significantly contributed to the concept of this study, designed and performed experiments, and interpreted results. *De novo* structure elucidation of cystopipecotides and verification of the *in* vivo generated pipecolic acid was conducted by the author. Furthermore, the author performed the characterisation and Marfey's analysis. Moreover, the author contributed by conception and writing this manuscript.

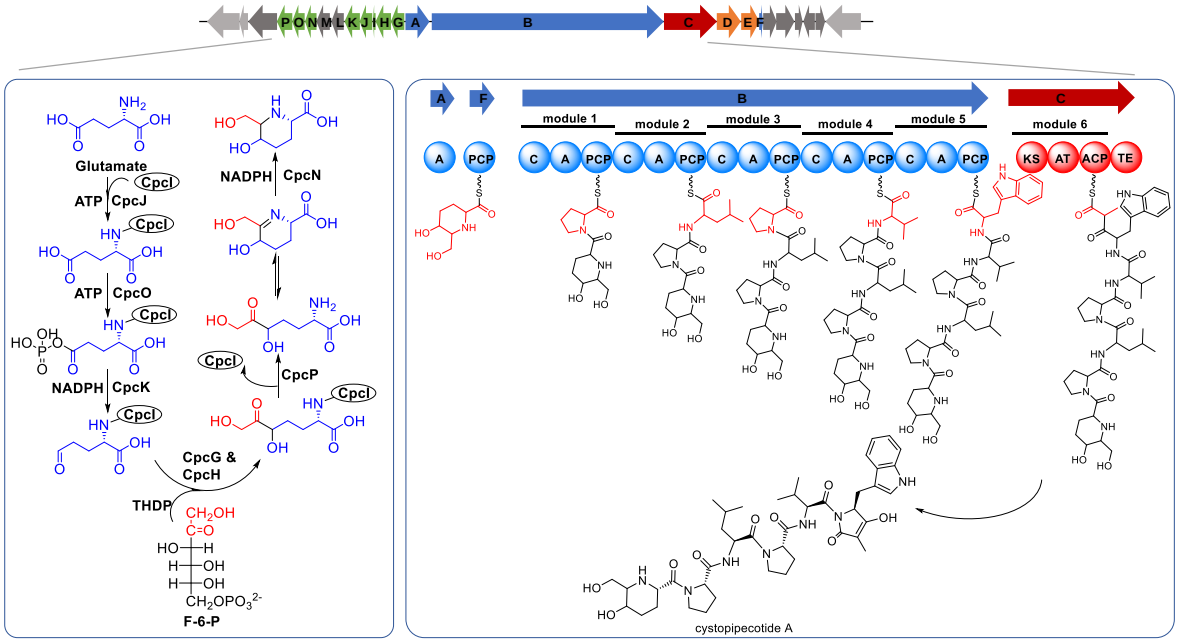
Contributions by Others

Hu Zeng contributed to the conception of this study, designed, performed and evaluated the analysis of the cystopipecotide gene cluster. He performed compound purification, mutagenesis work and feeding experiments, as well as proposed the biosynthetic pathway of cystopipecotides and reconstructed the pathway to the substituted pipecolic acid *in vitro*. Furthermore, Hu Zeng contributed to conceiving and writing this manuscript. Lena Keller was involved in the conception of NMR experiments and contributed through proofreading. We are thankful for Alexander Horn to assist with the ozonolysis, and Judith Hoffmann for scientific discussions and proofreading regarding the NMR analysis. Carsten Volz and Chengzhang Fu contributed to supervision, conception, writing and editing of this manuscript. Rolf Müller was responsible for the conception and supervision of the project and involved in writing of this manuscript.

Work presented in this chapter builds upon result contained in the doctoral thesis of Hu Zeng[1]. Additionally, it presents the full structure elucidation of cystopipecotides, and biochemical work.

4.1. Abstract

Myxobacteria are an abundant source of natural products featuring unique structures and biosynthetic machineries. The biosynthetic potential of myxobacteria, even in some well-explored species, is still far from being exhausted. In this work, a novel family of linear peptides, named cystopipecotides, were discovered from *Cystobacter* sp. SBCb004. These peptides were found to contain an unprecedented 5-hydroxyl-6-hydroxymethylpipecolic acid (HHMPA) building block. A NRPS-PKS hybrid gene cluster responsible for the biosynthesis of cystopipecotides was identified via retro-biosynthetic analysis and confirmed by insertional mutagenesis. The biosynthesis of HHMPA was shown to be conducted by a concerted action of eight biosynthetic proteins including enzymes exhibiting homology to transketolases and several lysine biosynthetic enzymes involved in the α-aminoadipate pathway based on heterologous expression. The HHMPA biosynthesis pathway was reconstructed *in vitro* based on protein expression and purification in *Escherichia coli*, followed by stepwise biochemical assays and LC-MS analysis of intermediates: Starting from the activation and linkage of L-Glutamic acid to the LysW-like carrier protein CpcI, biosynthesis follows with sequential phosphorylation, reduction, formal addition of a dihydroxyethyl group, cyclisation and imine reduction. This work thus not only revealed a novel peptide natural product exhibiting unprecedented building blocks, but also accomplished *in vitro* reconstruction of the unprecedented pathway to the substituted pipecolic acid.



Graphical Abstract.

4.2. Introduction

Besides actinomycetes and fungi, myxobacteria have proven to be producers of a large variety of natural products with unique structural features and biological activities[2–5]. Additionally, representatives of myxobacteria possess some of the largest genomes known from prokaryotes[6,7]. In these large genomes, numerous putative biosynthetic gene clusters (BGCs) encoding for the assembly lines of yet unknown natural products can be found, implying that the biosynthetic potential of myxobacteria is far from exhausted[2,8–11].

The secondary metabolites discovered in myxobacteria are mainly synthesised by polyketide synthases (PKS), non-ribosomal peptide synthetases (NRPS), and by the action of NRPS-PKS hybrid assembly lines[12]. The non-iterative type I PKS and NRPS are generally multimodular megaenzymes that can use various amino acids and carboxylic acids as building blocks, with each module usually responsible for only one extension cycle[13–15]. Thus, the number of modules usually agrees with the number of extension cycles. This one-to-one correspondence of biosynthetic assembly line and metabolic structure is termed the collinearity rule, and facilitates the prediction of the product's structure based on enzymatic domain composition, and vice versa[16–18].

Our previous *in silico* analysis of the *Cystobacter* sp. SBCb004 genome showed that it harbors 24 PKS and NRPS BGCs in its genome[19]. Besides the BGCs encoding the biosynthetic machineries for stigmatellin[20], tubulysin[21], alkylresorcinol[22], ajudazol[19] and argyrin[23], no natural product could be assigned to the other 19 BGCs. These existing known metabolites exhibit antimicrobial, antifungal and antitumor activities as well as interesting chemical scaffolds, which have piqued our interest in the yet unknown natural products and their potential.

In this work, we found a novel family of linear peptides named cystopipecotides by metabolomic profiling of extracts obtained from *Cystobacter* sp. SBCb004. These peptides bear a 5-hydroxyl-6-hydroxymethylpipecolic acid moiety (HHMPA) at their *N*-terminus and a γ-lactam at the C-terminus. The corresponding BGC was identified as an NRPS-PKS hybrid gene cluster via insertional mutagenesis. The biosynthetic pathway of HHMPA, which involves an unprecedented and concerted action of transketolases and a set of lysine biosynthetic enzymes via fragments of the *α*-aminoadipate pathway, was proposed and reconstructed *in vivo* and *in vitro*. The biosynthesis of HHMPA reveals a novel route to generate substituted pipecolates in bacteria.

4.3. Results and Discussion

Discovery of cystopipecotides based on specific MS^2 features. In order to explore the potential of the secondary metabolome of *Cystobacter* sp. SBCb004, UHPLC-MS and MS^2 data of the crude extract was analysed and dereplicated against our in house database Myxobase[24], the Npatlas database[25], and the dictionary of Natural Products[26]. Apart from representatives of the known metabolite families stigmatellin[20], tubulysin[21] , ajudazol[19] and argyrin[23], two hitherto unknown molecular features, later designated as cystopipecotide A (*m/z* 806.4450 $[M+H]^+$, t_R = 6.89 min) and B (*m/z* 792.4298 $[M+H]^+$, t_R = 6.37 min), stood out due to their specific fragmentation pattern (**Figure 4. 1**A and B). MS^2 fragmentation indicated the presence of at least four amino acids in both molecules, with a difference between the two molecules in the third amino acid, which was later identified as leucine and valine, respectively (**Figure 4. 1**B). Furthermore, both N- and C-termini contained the same fragments that could not be identified as proteinogenic amino acids, respectively, which sparked our interest to investigate their structure and biosynthetic origin. As no compound matching these molecules and their fragmentation pattern could be found in neither the Dictionary of Natural Products[26], the Npatlas database[25], nor our in house database Myxobase[24], these amino acids-containing-metabolites were potentially novel compounds and therefore subsequently isolated. Their structures were elucidated by a combination of HR-MS^2 and NMR experiments.

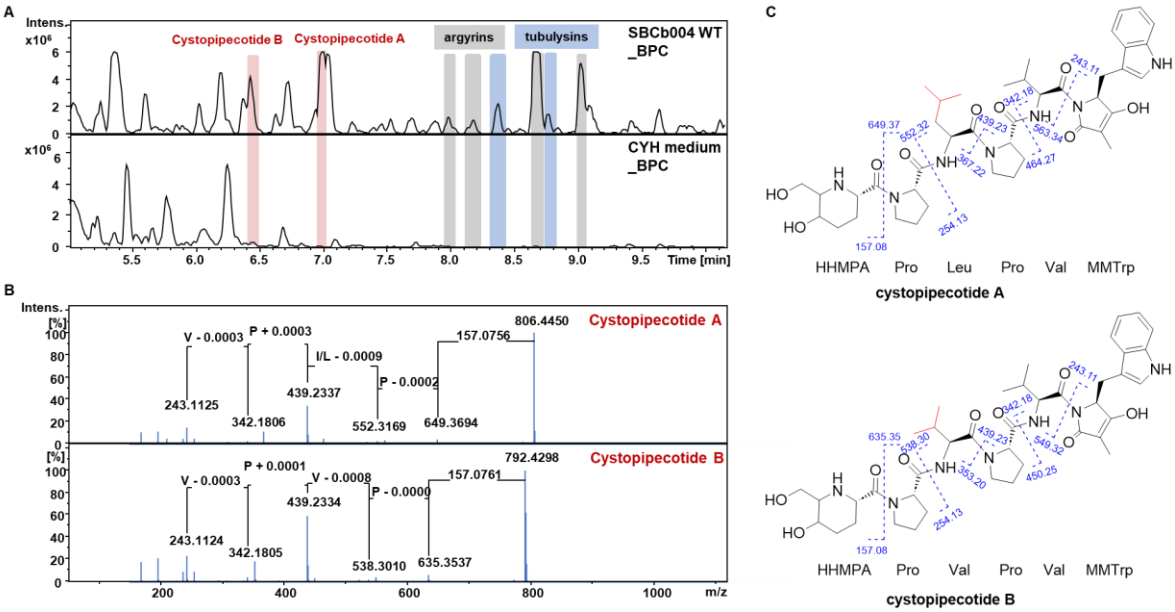


Figure 4. 1. Discovery, fragmentation pattern and chemical structures of cystopipecotides A and B. *(A) Comparative LC-MS analysis of* Cystobacter *sp. SBCb004 extracts and CYH medium. WT: wild type. BPC: base peak chromatogram. (B) MS^2 spectra of cystopipecotides A and B. (C) Structure and fragmentation scheme of cystopipecotides A and B in the MS^2 experiments. HHMPA: 5-hydroxyl-6-hydroxymethylpipecolic acid moiety. MMTrp: Methylmalonyl-CoA condensed Tryptophan.*

Structure elucidation. ESI-HR-MS analysis of cystopipecotide A (**1**) revealed an $[M+H]^+$ ion with *m/z* 806.4450 (calcd for $C_{42}H_{60}N_7O_9$, 806.44470), consistent with the molecular formula $C_{42}H_{59}N_7O_9$ containing 17 double bond equivalents. The ^{1}H NMR spectrum of **1** showed signals characteristic for a peptide, including five typical α-protons (δ 3.8-4.6), and one with a downfield shift of $\delta_{H\text{-}Val\text{-}2}$ 5.66. Furthermore, characteristic shifts of five aromatic/olefinic protons (δ 6.8-7.5), including an aromatic amine proton (δ 10.66), are to be emphasised. In alignment with these findings, the ^{13}C NMR spectrum showed characteristic signals for six amide carbonyl carbons (δ 169.2-172.4), which fits the assumption from MS^2 that **1** is a hexapeptide. Furthermore, the ^{13}C NMR spectrum showed signals for five quaternary carbons (δ 94.1-135.5), including a downfield shifted carboxyl carbon ($\delta_{C\text{-}MMTrp\text{-}3}$ 182.1). Additionally, the HSQC spectrum revealed characteristic signals for two hydroxylated methines, several diastereotopic methylene groups and five methyl groups, whereas one methyl group showed a ^{13}C upfield and a ^{1}H downfield shift characteristic for a methyl substituent at or on an aromatic system ($\delta_{H\text{-}MMTrp\text{-}2Me}$ 1.20, $\delta_{C\text{-}MMTrp\text{-}2Me}$ 6.3) (**Table S4. 1**, **Figure S4. 1**-**Figure S4. 9**). The 2D NMR spectra COSY, HMBC and N-HMBC showed the presence of the amino acids leucine (Leu), valine (Val), and two prolines (Pro1 and Pro2), as shown in **Figure 4. 2**. Leucine and valine were revealed by the presence of a spin system starting at geminal methyl groups Leu-5/6 (δ_H 0.89 and 0.90) and Val-4/5 (δ_H 0.75 and 0.78), respectively, in both cases leading to a methine (Leu-4: δ_H 1.72, Val-3: δ_H 1.98), in case of leucine further to a diastereotopic methylene Leu-3 (δ_H 1.45 and 1.52), and finally to the α-protons Leu-2 (δ_H 4.53) and Val-2 (δ_H 5.66), respectively. Both prolines were elucidated based on their spin systems from α-protons Pro1-2 (δ_H 4.37) and Pro-2-2 (δ_H 4.55) to three consecutive methylenes (Pro1 and 2 position 3-5 δ_H 1.76-3.62), and the ring closure was shown by HMBC correlations from α-protons to carbons Pro1-5 and Pro-2-5 (both δ_C 46.7), respectively. Furthermore, a 5-hydroxy-6-hydroxymethyl pipecolic acid (HHMPA) moiety could be deduced from a spin system starting with the α-proton HHMPA-2 (δ_H 3.86) to two diastereotopic methylenes (HHMPA-3: δ_H 1.65 and 1.87; and HHMPA-4: δ_H 1.52 and 1.74), a hydroxylated methine HHMPA-5 (δ_H 3.40), another methine HHMPA-6 (δ_H 3.19) and a hydroxylated diastereotopic methylene group HHMPA-7 (δ_H 3.49 and 3.64). The pipecolic acid ring closure was elucidated based on the HMBC correlation from H-HHMPA-2 to C-HHMPA-6 (δ_C 59.2) (**Figure 4. 2**). Both, chemical shifts and correlations of the HHMPA moiety showed high resemblance to those of the 6-chloromethyl-5-methoxypipecolic acid (CMPA) unit in chloromyxamides[27]. Additionally, a moiety that was later shown to be derived from methylmalonyl-CoA and tryptophan, therefore named methylmalonyl tryptophane (MMTrp), could be elucidated as follows. COSY and HMBC correlations as well as characteristic shifts revealed an indole moiety[28], which was connected via the methylene HHMPA-5 to the alpha-

proton-like position MMTrp-4 (δ_H 4.20) of a 5-yl-4-hydroxy-3-methyl-1,5-dihydro-2H-pyrrol-2-one moiety, also called a methylated tetramic acid moiety. The tetramic acid derivative was elucidated as follows. MMTrp-4 showed HMBC correlations to a fully substituted olefinic carbon MMTrp-2 (δ_C 94.1). This carbon was also observed due to HMBC correlations from an upfield shifted methyl group MMTrp-2Me (δ_H 1.20, δ_C 6.3). Furthermore, the protons of MMTrp-2Me showed HMBC correlations to the carbonyl carbon MMTrp-1 (δ_C 172.4). The connectivity of this moiety was confirmed by 1,1-ADEQUATE correlations between the methylene group MMTrp-5 and the *α*-like carbon MMTrp-4 (**Figure 4. 2**), and showed some resemblance in chemical shifts and correlations to those of the tetramic acid moiety of griseofamine B[29].

Finally, the amino acid sequence was determined by HMBC and N-HMBC correlations from *α*-protons to carbonyl carbons and *β*-protons to amide nitrogens, respectively (**Figure 4. 2**). However, due to the absence of a HMBC correlation between MMTrp-4 and Val-1, the connectivity of the MMTrp moiety to valine could only established through ESI-HR-MS^2 experiments (**Figure S4. 25**) Thereby, the number of double bonds equivalents fits the calculated number, completing the structure of **1** as a linear hexapeptide with the sequence from *N*-to *C*-terminus of HHMPA-Pro1-Leu-Pro2-Val-MMTrp.

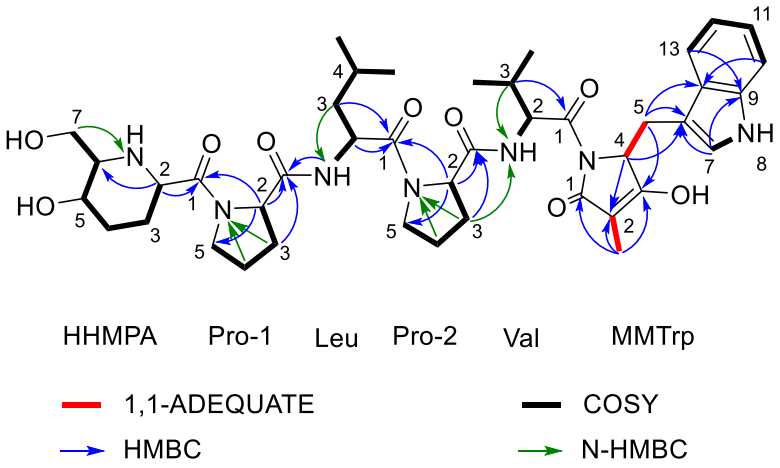


Figure 4. 2. Key NMR correlations of cystopipecotide A (1). *1,1-ADEQUATE correlations shown in bold red, COSY in bold black, HMBC with a blue arrow and N-HMBC correlations with a green arrow.*

ESI-HR-MS analysis of cystopipecotide B (**2**) displayed an $[M+H]^+$ signal at *m/z* 792.4298 (calcd for $C_{41}H_{58}N_7O_9$, 792 .42905), consistent with the molecular formula $C_{41}H_{57}N_7O_9$ thus containing one carbon and two protons less than **1**. Consistent with this finding, ^{1}H, ^{13}C and HSQC spectra of **2** showed similar signals compared to **1**, however, one less diastereotopic methylene group (**Table S4. 2**, **Figure S4. 10**-**Figure S4. 14**). However, compared to **1**, COSY and HMBC experiments

revealed the presence of a second valine in **2** instead of leucine in **1**, resulting in the hexapeptide sequence HHMPA-Pro1-Val1-Pro2-Val2-MMTrp that is consistent with the ESI-HR-MS^2 analysis.

The absolute configuration of the proteinogenic amino acid residues was elucidated by UHPLC-ESI-HR-MS analysis of D- and L-FDLA (1-fluoro-2,4-dinitrophenyl-5-L/D-leucinamide) derivatives after hydrolysis of **1** and **2**, respectively, in comparison to authentic standards[30,31]. The amino acids proline, leucine and valine were assigned as L-configured (**Figure S4. 21**).

The absolute configuration of MMTrp was elucidated by ozonolysis followed by Marfey's analysis. After hydrolysis of **1**, the product was submitted to ozonolysis known to lead to the degradation of tryptophan to yield aspartate[28,32,33]. The ozonolysis product and authentic amino acid standards were derivatised with D- and L-FDLA, respectively. By comparison of the ozonolysis product to authentic amino acid standards L-aspartate was identified, thereby showing MMTrp to be L-configured (**Figure S4. 22**). In addition, feeding isotopically labelled L- and D-tryptophane-d_8 showed the incorporation of L-tryptophane.

The *in vitro* reconstitution of the HHMPA biosynthesis showed the incorporation of L-glutamic acid in HHMPA, thereby suggesting that HHMPA is L-configured. In order to elucidate the configuration of position HHMPA-5 and HHMPA-6, S^3HMBC experiments were employed showing a $^3J_{HH}$ coupling of 2.9 Hz, which suggests an equatorial-equatorial orientation of H-HHMPA-5 and H-HHMPA-6 (**Figure S4. 10**)[27,34]. However, as we did not observe a reference signal to validate this finding, further proof in form of a comparative Marfey's analysis with authentic HHMPA standards is needed to determine its absolute configuration.

Identification of the cystopipecotide biosynthetic gene cluster. The structural similarity between cystopipecotides A and B suggested that they should originate from the same biosynthetic machinery followed by alternative subsequent modifications or using different extender units. To shed light on the origin of the building blocks, L-Leucine-5,5,5-d_3 and L-Valine-d_8 were fed to the *Cystobacter* sp. SBCb004 wild type strain. Consequently, one leucine and one valine were observed to be incorporated into cystopipecotide A, while no leucine but two valines were incorporated into cystopipecotide B (**Figure S4. 23**). The results clearly show that the structural difference between cystopipecotides A and B originates from the incorporation of different building blocks rather than from alternative modification steps after the common core structure is synthesised.

The C-terminus of cystopipecotides A and B comprises the backbone of a tryptophan and a γ-lactam, which is postulated to be formed by a carboxyl group provided by a C_3 unit and the amine

group of tryptophan. Feeding of sodium propionate-1-^{13}C and L-Tryptophan-d_8 led to the enrichment of cystopipecotide A (*m/z* 806) isotopic peaks *m/z* 807 and *m/z* 814, respectively (**Figure S4. 24**), supporting the proposal of the incorporation of methylmalonyl-CoA and L-Tryptophan. Therefore, the C-terminus is proposed to derive from tryptophan and one methylmalonyl-CoA is assembled to the carbon chain through a Claisen-like condensation. The product is offloaded from the assembly line by lactamisation between the terminal carbonyl group and the amine group of tryptophan. The nascent carbonyl group resulting from the Claisen condensation then spontaneously tautomerises to an enolate (**Figure 3. 3**B).

As no putative precursor peptide in the genome of *Cystobacter* sp. SBCb004 matches the oligopeptide sequence of P-L/V-P-V-W in cystopipecotide A and B, we proposed that cystopipecotides do not belong to ribosomally synthesised and post-translationally modified peptides, but are generated by a machinery comprising at least five NRPS modules and one PKS module. Thus, a NRPS-PKS hybrid gene cluster encoding one AMP-binding domain (CpcA), five entire NRPS modules (CpcB) and one PKS module (CpcC), was identified in the genome of strain *Cystobacter* sp. SBCb004 (**Figure 3. 3**, **Figure S4. 23**). According to our *in silico* analysis, the predicted substrate specificity of A domains in NRPS module 1, 3 and 4 matches amino acids (P, P and V) in the corresponding positions of cystopipecotides (**Figure S4. 24**). The AT domain in PKS module (CpcC) may utilise methylmalonyl-CoA (**Figure S4. 25**), which is also consistent to our proposal. *cpcD* and *cpcE* downstream to the NRPS-PKS genes putatively encode a transporter and a secretion protein, respectively, which may act during transportation and resistance.

The inactivation of *cpcB* via insertional mutagenesis abolished the production of both, cystopipecotides A and B. In addition, three peaks exhibiting *m/z* values of 667.381 (t_R = 6.26 min), *m/z* of 681.398 (t_R = 6.68 min) and *m/z* of 766.451 (t_R = 6.73 min) vanished in the HPLC/MS chromatogram of extracts obtained from mutant *Cystobacter* sp. SBCb004 cpcB$^-$ (**Figure S4. 25**A), suggesting that cystopipecotides A and B may not be the only compounds produced by the inactivated biosynthetic machinery. These molecules were postulated to be the derivatives of cystopipecotides A and B and were designated as cystopipecotides C-E (**Figure S4. 25**B and C). According to the MS/MS analyses, the fifth amino acid valine in cystopipecotides A and B is missing in cystopipecotides C and D (**Figure S4. 25**B). The fragment of the C-terminus in cystopipecotides C-E is 26 Dalton lighter than that in A and B (**Figure S4. 25**B), which may result from a decarboxylation after the product is liberated from the NRPS-PKS assembly line. Therefore, the C-terminus of cystopipecotides C-E may be a propionyl residue rather than the *γ*-lactam (**Figure S4. 25**C). Based on the analysis of all possible combinations of amino acids and the varying C-terminus,

we proposed the existence of cystopipecotides F-H (**Figure S4. 25**C). However, the amount of the predicted compound with respective *m*/*z* values is too low to be detected with confidence in the chromatogram.

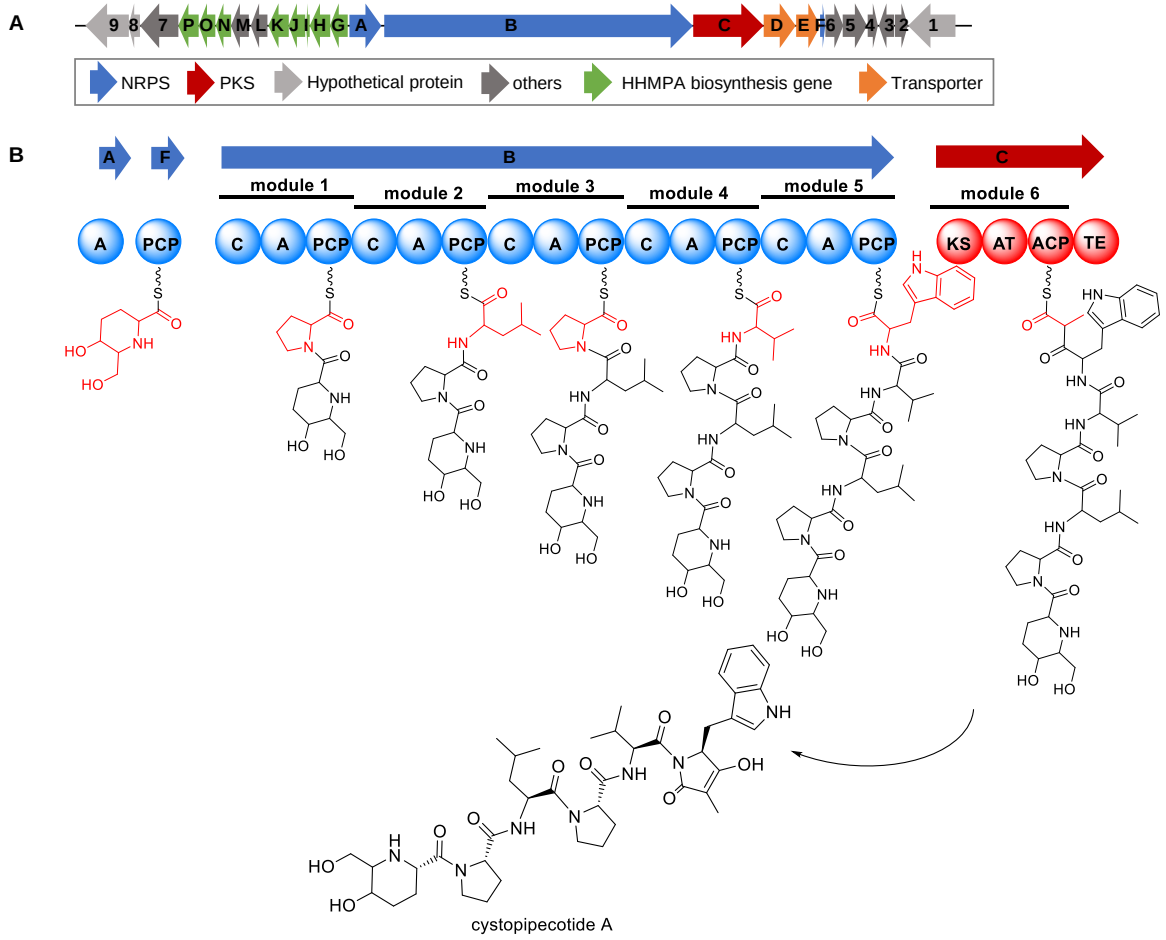


Figure 4. 3. Biosynthetic gene cluster and proposed biosynthetic pathway of the cystopipecotides. *(A) Genetic organisation of the putative biosynthetic gene cluster of cystopipecotides. (B) Putative biosynthetic pathway of cystopipecotide A. A: adenylation. C: condensation. PCP: peptidyl carrier protein. AT: acyl transferase. KS: ketosynthase. ACP: acyl carrier protein. TE: thioesterase.*

Taken together, we propose that the biosynthesis of cystopipecotides A and B is initiated with the formation of HHMPA (see below) which is adenylated by an AMP-binding protein encoded by *cpcA*. The activated HHMPA is then loaded to a dissociated PCP domain, which is probably encoded by the standalone *cpcF* found downstream (**Figure 3. 3**A). Subsequently, the C domain of NRPS module 1 within CpcB recognises the HHMPA-tethered PCP domain and condenses HHMPA to the amino group of PCP-tethered proline, which is delivered by the A domain of NRPS module 1. Then

one leucine/valine, one proline, one valine and one tryptophan are successively added by NRPS modules 2-5 within CpcB, followed by the PKS module (CpcC) mediated incorporation of one methylmalonyl-CoA to the peptide chain via a Claisen condensation. Ultimately, the product is offloaded from the assembly line by γ-lactam formation in cystopipecotides A and B including the spontaneous tautomerisation resulting in the enolate (**Figure 3. 3**B). In the biosynthesis of cystopipecotides C and D, module 4 of CpcB incorporating valine is presumably skipped. Moreover, in the biosynthesis of cystopipecotides C-E, the carboxyl group resulting from the release from the TE using water instead of the intermolecular amide group is decarboxylated after release of the free acid.

Biosynthesis of the HHMPA building block. In microbes, pipecolic acid is usually synthesised from lysine[35] (**Figure S4. 26**A). It was also reported that the γ-substituted pipecolate could be biosynthesised from *O*-acetyl-L-homoserine and acetoacetate in fungus[36] (**Figure S4. 26**B). Besides, 5-substituted L-pipecolic acid could be chemoenzymatically synthesised from aldehyde and acetoacetate[37] (**Figure S4. 26**C). One important myxobacterial metabolite, tubulysin, harbors a D-*N*-methyl pipecolic acid building block (D-Mep), which also originates from lysine while *N*-methylation and epimerisation occur during the NPRS assembly[21] (**Figure S4. 26**D). Another important case in myxobacteria is chloromyxamide which contains a 6-chloromethyl-5-methoxypipecolic acid (CMPA) building block. Its biosynthesis was proposed to originate from L-Glutamic acid[27] (**Figure S4. 26**E). However, the pivotal genes involved in the above-mentioned biosynthetic routes could not be found in the cystopipecotide BGC, suggesting an alternative biosynthetic pathway of HHMPA.

In the upstream region of the NRPS and PKS genes *cpcABC*, a conserved operon (*cpcG* to *cpcP*) which we found widely distributed in myxobacteria can be identified (**Figure 3. 3**A, **Table S4. 6**). CpcG and CpcH share high similarity (88 % and 99 %, respectively, amino acid level) to two-subunit transketolases which can transfer a C-2 keto-containing unit from a 2-ketose to the first carbon atom of an aldose[38]. Proteins CpcI, CpcJ, CpcO, CpcK and CpcP are homologous to the enzymes involved in lysine biosynthesis through the α-aminoadipate (AAA) pathway and arginine biosynthesis (LysW, LysX, LysZ/ArgB, LysY/ArgC and LysK/ArgE) (**Figure S4. 27**A and B, **Table S4. 6**). What is missing in AAA pathway is the homolog of LsyJ/ArgD, which can transfer the amino group from glutamine to the semialdehyde[39]. Thus, the hypothetical product of CpcI, CpcJ, CpcO and CpcK may stay as semialdehyde instead of being further transformed to lysine or arginine. AAA pathway enzymes coupled with transketolases were also reported to synthesise the precursor

(2*S*,6*R*)-diamino-(5*R*,7)-dihydroxy-heptanoic acid (DADH) in streptomycetes[40,41]. In this case, the whole AAA pathway cassette besides transketolases are present. The transketolases interfere the process of AAA enzymes by formally transferring a dihydroxyethyl group to the semialdehyde. With the presence of the LysJ homologue, transamination occurs at the nascent keto group introduced by the transketolases. The linear intermediate DADH is then released from LysW by hydrolysation (**Figure S4. 27**C).

Therefore, we proposed that the biosynthesis of HHMPA starts from the binding of glutamate to the putative LysW like protein CpcI, catalysed by the putative RimK family alpha-L-glutamate ligase CpcJ. CpcI may act as a carrier protein and in parallel protects the amino group of glutamate[39]. The γ-carboxyl group of glutamate is then phosphorylated by the putative LysW-aminoadipate kinase CpcO, followed by a reduction to the semialdehyde catalysed by the putative N-acetyl-γ-glutamyl-phosphate reductase CpcK. The putative transketolases CpcG and CpcH, in the presence of the essential cofactor thiamin diphosphate (ThDP), might transfer a dihydroxyethyl group from a 2-ketose to the semialdehyde via a nucleophile attack, generating a carbon-carbon bond accompanied by the reduction of the aldehyde to the keto group. Then, LysW-like CpcI releases the amino group hydrolysed by CpcP. In contrast to the biosynthesis of lysine, arginine and DADH, no transaminase exists in the HHMPA biosynthetic cassette. Thus, the keto group introduced by the transketolases CpcG and CpcH might next attack the free amino group and form the tetrahydropicolinate, which is further saturated by the putative short-chain dehydrogenase/reductase family oxidoreductase CpcN thus yielding HHMPA (**Figure 4. 4**A).

To verify our hypothesis, D-Glucose-1-^{13}C, D-Glucose-2-^{13}C, and L-Glutamic-2,3,3,4,4-d_5-acid were fed to *Cystobacter* sp. SBCb004. Intriguingly, only glucoses could be incorporated into cystopipecotides (**Figure S4. 24**), indicating that the putative transketolases may be involved. Nevertheless, both glucose and glutamate could participate in various metabolic pathways in myxobacteria. Therefore, the source of HHMPA cannot be evaluated based on feeding experiments alone.

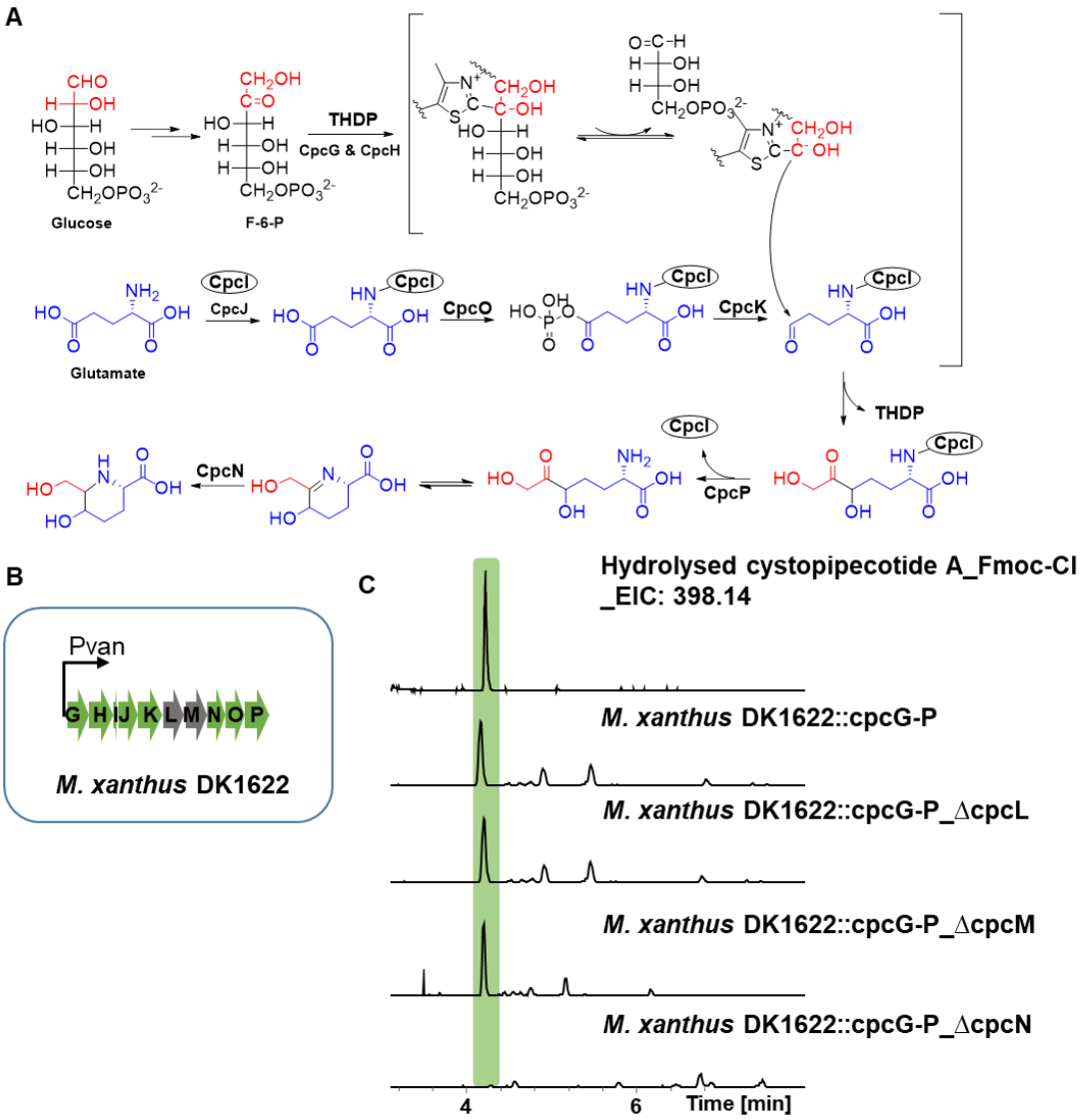


Figure 4. 4. Biosynthesis of 5-hydroxyl-6-hydroxymethylpipecolic acid (HHMPA). *(A) Proposed biosynthesis of HHMPA. F-6-P: Fructose 6-phosphate. THDP: Thiamin diphosphate. (B) Schematic representation of heterologous expression of cpcG-P in* M. xanthus *DK1622. (C) Extracted ion chromatogram of Fmoc-HHMPA (m/z 398.14) from standard HHMPA,* M. xanthus DK1622*::cpcG-P,* M. xanthus *DK1622::cpcG-P_ΔcpcL,* M. xanthus *DK1622::cpcG-P_ΔcpcM and* M. xanthus *DK1622::cpcG-P_ΔcpcN.*

In order to investigate whether genes flanking the NRPS-PKS encoding genes *cpcA-C* are involved in the biosynthesis of cystopipecotides as postulated, the insertional mutants of *cpcG*, *cpcK*, *cpcM*, *cpcN*, *cpcP* and *orf6* were obtained, respectively, and the mutant metabolite profiles were analysed. However, the production of cystopipecotide A in these mutants were not totally abolished like in SBCb004 cpcB^{-}, but were only reduced compared to *Cystobacter* sp. SBCb004 wild type strain (**Figure S4. 28**). After interrogation of the *Cystobacter* sp. SBCb004 genome, we found

another BGC, designated as BGC13, containing an additional set of homologs of *cpcA*, *cpcD-F*, *cpcG-P* and *orf4-6*. The difference between BGC13 and BGC of cystopipecotide is that BGC13 does not contain the homolog of the PKS encoding gene *cpcC* besides the NRPS encoding gene in BGC13 may only encode three NRPS modules (**Table S4. 9**). To investigate whether BGC13 participates in the biosynthesis of cystopipecotide or other metabolites, the insertional mutants of *BGC13_B* and *BGC13_K* were created. However, no metabolic change were found in the insertional mutant SBCb004 BGC13_B$^-$ and SBCb004 BGC13_K$^-$ (**Figure S4. 28**), indicating that the NRPS modules in BGC13 may be silent under the applied laboratory condition. As no resistance marker except kanamycin could be used in *Cystobacter* sp. SBCb004, insertional mutagenesis of *BGC13 G-P* and *orf4-6* in the mutants of their corresponding congeners within cystopipecotide BGC is not achievable. Given in-frame gene deletion in *Cystobacter* sp. SBCb004 could not be established, it is not possible to investigate whether *cpcG-P* or *BGC13 G-P* are involved in the biosynthesis of HHMPA via *in vivo* knockout.

To confirm the correlation of HHMPA and *cpcG-P* operon, the *cpcG-P* operon was cloned into pFPvan vector[42] under the control of a vanillate inducible promoter (**Figure 4. 4**B). The *cpcG-P* operon containing construct was introduced in *Myxococcus xanthus* DK1622, leading to the production of HHMPA. HHMPA was purified and its structure was verified by NMR showing a 2(*S*),5(*S*),6(*R*)-configuration, which matches that of a pipecolic acid unit of a different myxobacterial compound, CMPA in chloromyxamides (**Table S4. 2**, **Figure S4. 15**-**Figure S4. 20**)[27]. Nevertheless, as the structure elucidation of cystopipecotide A and B hints towards a different configuration, further investigations should be conducted to secure the configuration of the natural product, such as Marfey's analysis in comparison with authentic synthetic standards of HHMPS conformers. At presence, without the synthetic reference compound at hand, this discrepancy cannot be resolved. Gene deletion of *cpcL* and *cpcM* showed that HHMPA production is not influenced while deletion of *cpcN* abolished its production (**Figure 4. 4**C), which is consistent with our proposal.

Reconstitution of the HHMPA biosynthesis *in vitro*. To elucidate the biosynthesis of HHMPA, we expressed and purified the recombinant proteins CpcI, CpcJ, CpcO, CpcK, CpcG/CpcH, CpcP and CpcN (**Figure S4. 26**) and reconstructed the proposed biosynthetic pathway stepwise.

The biosynthesis is supposed to be initiated by the linkage of L-Glutamic acid to LysW like protein CpcI via the amine group of L-Glutamic acid and the γ-carboxyl group of the C-terminal Glutamic acid in CpcI, catalysed by CpcJ[43] (**Figure 4. 5**A). As our feeding experiments did not provide any

information about the origin of HHMPA, different amino acids including L-Glutamic acid, L-Aspartic acid, L-Arginine and L-Lysine were used to react with CpcI and CpcJ, respectively, to test their substrate specificity. In the intact protein analysis performed by LC-MS, only CpcI tethered Glutamic acid (CpcI-*γ*-Glu) could be detected (**Figure 4. 5**B), suggesting that L-Glutamic acid is the specific substrate of CpcI and CpcJ.

Next, we added CpcO to potentially phosphorylate the γ-carboxyl group of CpcI-γ-Glu in the presence of ATP (**Figure 4. 5**C). However, the CpcO product CpcI-γ-Glu-phosphate was found to be unstable and easily converted to CpcI-γ-Glu during MS analysis (**Figure 4. 5**D), which is consistent with the report about LysW-γ-AAA-phosphate in lysine biosynthesis[44]. Hence, we added *N*-Methylhydroxylamine to convert the unstable carbonyl phosphate group to the hydroxamate group and observed the expected product CpcI-γ-Glu-hydroxamate (**Figure 4. 5**C and D).

CpcK was then added to the reaction and CpcI-γ-Glu-phosphate was observed to convert to CpcI-γ-Glu-semialdehyde (**Figure 4. 5**E and F). Adjacent to the product (molecular weight: 10271.65 Da), we identified an unexpected peak with the molecular weight of 10253.6372 Da (**Figure 4. 5**F), which may be caused by the instability of the aldehyde group.

To investigate the functions of transketolases CpcG/CpcH, a reaction containing CpcI, CpcJ, CpcO, CpcK and CpcG/CpcH was set up. In this reaction, fructose-6-phosphate was used as the donor of C_2 unit (**Figure 4. 5**E). A peak with the molecular weight of 10332, corresponding to the CpcI linked 2-amino-5,7-dihydroxy-6-oxoheptanoic acid (ADHOHA), was detected (**Figure 4. 5**F), which is consistent with our assumption.

When CpcP was added to the above reaction, the peak corresponding to the CpcI-ADOHOHA decreased while the peaks corresponding to CpcI and CpcI-γ-Glu were found enriched (**Figure 4. 5**F). This result suggests the release of ADHOHA from CpcI, which in turn increased the amount of upstream intermediate.

The released ADHOHA equilibrates with tetrahydropicolinate, which might subsequently be reduced by CpcN (**Figure 4. 5**G). To confirm this assumption, a one pot reaction with all the above-mentioned enzymes was performed. As the final product HHMPA is a small hydrophilic molecule, which is hard to be detected in reverse-phase LC-MS, Fmoc-Cl was used to derivatise it and facilitate the detection[45]. Cystopipecotide A was hydrolysed by HCl to release the HHMPA unit as a positive control. In the reaction, a signal with the *m*/*z* of 398.14, which corresponds to the derivative Fmoc-HHMPA, was detected in LC-MS (**Figure 4. 5**H). Thus, the complete biosynthetic pathway of HHMPA was successfully reconstituted *in vitro*.

It is noteworthy that CMPA unit from another myxobacterial compound, chloromyxamide[27], is quite similar to HHMPA and could in theory be generated from the latter by *O*-methylation at C-5 position and chlorination at C-7 position. The biosynthesis of CMPA was proposed to start from glutamate and the C_2 unit was presumed to be provided by malonyl-CoA through Claisen condensation[27]. Intriguingly, the proposed operon of HHMPA could be found in the genome of chloromyxamide producing strains *Myxococcus* sp. MCy10608 and *Pyxidicoccus* sp. MCy10649 (**Figure 4. 6**, **Table S4. 10**), while no proposed CMPA producing operon exists in the available genome of *Cystobacter* sp. SBCb004, implying the possibility of multiple pathways to form similar structures in myxobacteria.

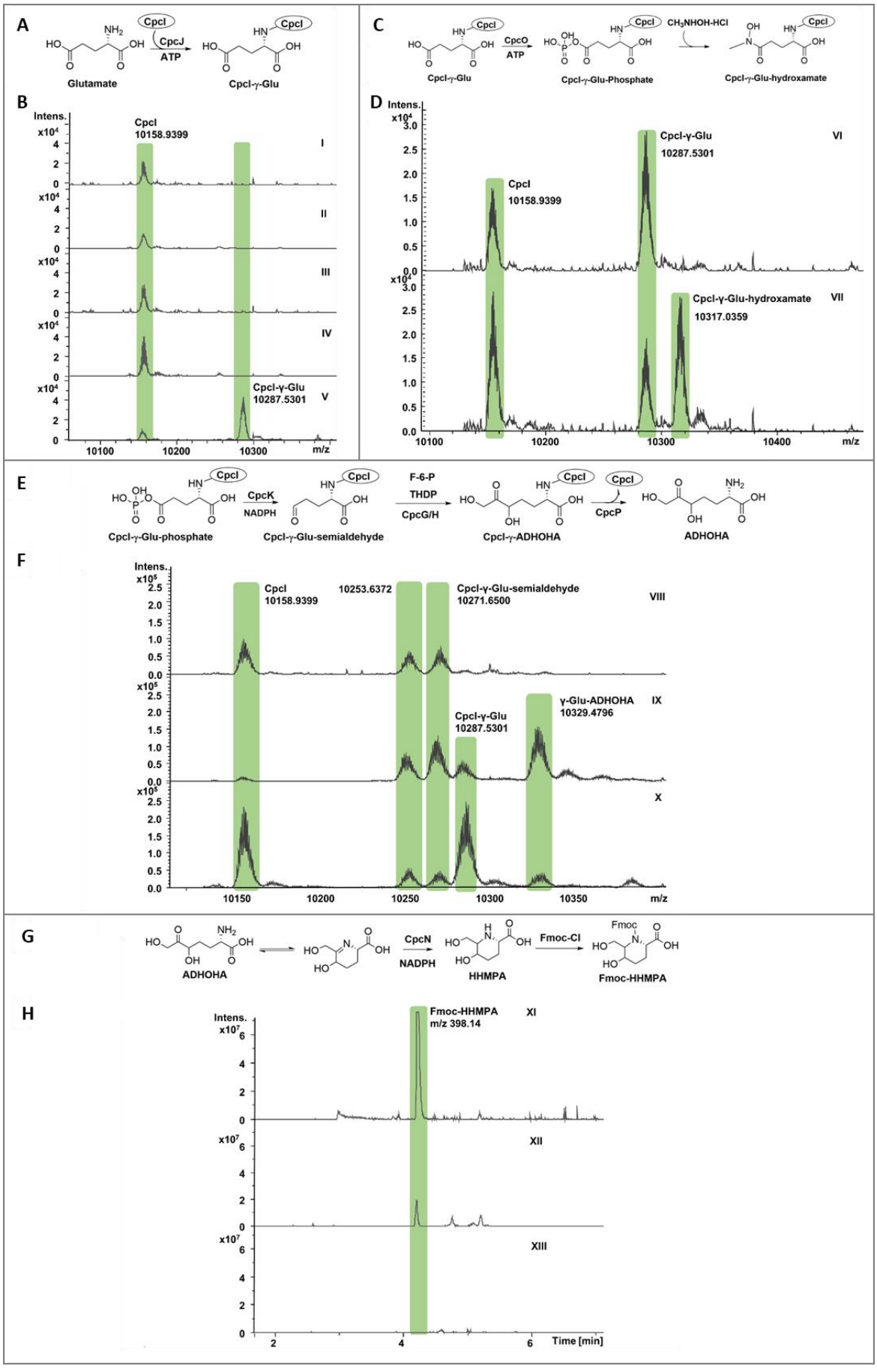

A
CpcI
CpcJ
ATP
Glutamate
CpcI-γ-Glu
B
Intens.
CpcI
10158.9399
I
II
III
IV
CpcI-γ-Glu
10287.5301
V
m/z
C
CpcO
ATP
CpcI-γ-Glu
CpcI-γ-Glu-Phosphate
CH3NHOH-HCl
CpcI-γ-Glu-hydroxamate
D
Intens.
CpcI-γ-Glu
10287.5301
VI
CpcI
10158.9399
CpcI-γ-Glu-hydroxamate
10317.0359
VII
m/z
E
CpcK
NADPH
F-6-P
THDP
CpcG/H
CpcP
CpcI-γ-Glu-phosphate
CpcI-γ-Glu-semialdehyde
CpcI-γ-ADHOHA
ADHOHA
F
Intens.
CpcI
10158.9399
10253.6372
CpcI-γ-Glu-semialdehyde
10271.6500
VIII
γ-Glu-ADHOHA
10329.4796
IX
CpcI-γ-Glu
10287.5301
X
m/z
G
ADHOHA
CpcN
NADPH
HHMPA
Fmoc-Cl
Fmoc-HHMPA
H
Intens.
Fmoc-HHMPA
m/z 398.14
XI
XII
XIII
Time [min]

Figure 4. 5. Reconstitution of HHMPA biosynthesis in vitro. A *Scheme of linkage of Glutamate to CpcI, catalyzed by CpcJ.* ***B*** *Deconvoluted mass spectra of CpcI reacted with: Glu (**I**), Arg and CpcJ (**II**), Asp and CpcJ (**III**), Lys and CpcJ (**IV**) and Glu and CpcJ (**V**). The calculated molecular weight (MW) of CpcI, CpcI-γ-Arg, CpcI-γ-Asp, CpcI-γ-Lys and CpcI-γ-Glu are 10159.06 Da, 10315.25 Da, 10274.15 Da, 10287.23 Da and 10288.18 Da, respectively.* ***C*** *Scheme of CpcO-mediated reaction and derivatisation by N-Methylhydroxylamine.* ***D*** *Deconvoluted mass spectra of CpcI reacted with Glu, CpcJ and CpcO.* ***VI****: without N-Methylhydroxylamine.* ***VII****: with N-Methylhydroxylamine. The calculated MW of CpcI-γ-Glu-phosphate and CpcI-γ-Glu-hydroxamate are 10368.16 Da and 10317.13 Da.* ***E*** *Scheme of reactions successively mediated by CpcK, CpcG/H and CpcP.* ***F*** *Deconvoluted mass spectra of CpcI reacted with Glu, CpcJ, CpcO and CpcK (**VIII**), CpcG/H (**IX**), CpcP (**X**). The calculated MW of CpcI-γ-Glu-semialdehyde and CpcI-γ-Glu-ADHOHA (2-amino-5,7-dihydroxy-6-oxoheptanoic acid) are 10272.18 Da and 10332.13 Da.* ***G*** *Scheme of CpcN-mediated reaction and derivatisation by Fmoc-Cl.* ***H*** *Extracted ion chromatogram of Fmoc-HHMPA (*m/z *398.14) from: **XI** cystopipecotide A hydrolysis product, **XII** reaction with CpcI, CpcJ, CpcO, CpcK, CpcG/H, CpcP and CpcN, **XIII** reaction withCpcI, CpcJ, CpcO, CpcK, CpcG/H and CpcP.*

Distribution and diversity of the putative HHMPA biosynthetic operons in myxobacteria. Inspired by the existence of the proposed HHMPA biosynthetic operon in chloromyxamide producing strains, we analysed all myxobacterial genomes in our in-house database with the amino acid sequence of CpcG-P and revealed that there are 32 myxobacteria containing the putative biosynthetic HHMPA operon (**Table S4. 10**). Intriguingly, all these 32 species belong to the suborder of *Cystobacterineae*, whereas 60 *Cystobacterineae* strains, 120 *Sorangiineae* strains and 20 *Nannocystineae* strains in database were checked. This indicates that the putative HHMPA operon may be exclusively distributed in *Cystobacterineae*.

According to the organisation of the NRPS-PKS genes adjacent to the HHMPA biosynthetic operon, these BGCs which may generate HHMPA containing compounds could be classified into two major types. Type A refers to the BGCs with a PKS module downstream of the NRPS modules, suggesting that the *C*-terminus of the product may form the γ-lactam as in cystopipecotides A and B. Type B depicts BGCs containing only NRPS modules. Each type could be further classified into two subtypes based on the position of the loading PCP domain, which is putatively responsible for the delivery of HHMPA at the beginning of the biosynthesis. Subtype 1 refers to the case that the loading PCP domain is located adjacent to the loading A domain. While in subtype 2, the loading PCP domain is at the downstream of all the NRPS-PKS genes as in the BGC of cystopipecotide (**Figure 4. 6**, **Table S4. 10**). Among all the 32 species, *Cystobacter* sp. SBCb004 and *Archangium minus* MCy10943 have type A2 and type B2 BGCs simultaneously. All other strains only exhibit one HHMPA biosynthetic operon containing BGC. The number of NRPS modules in each BGC is not fixed (**Figure 4. 6**, **Table S4. 10**), thus the size of their products would be variable as well.

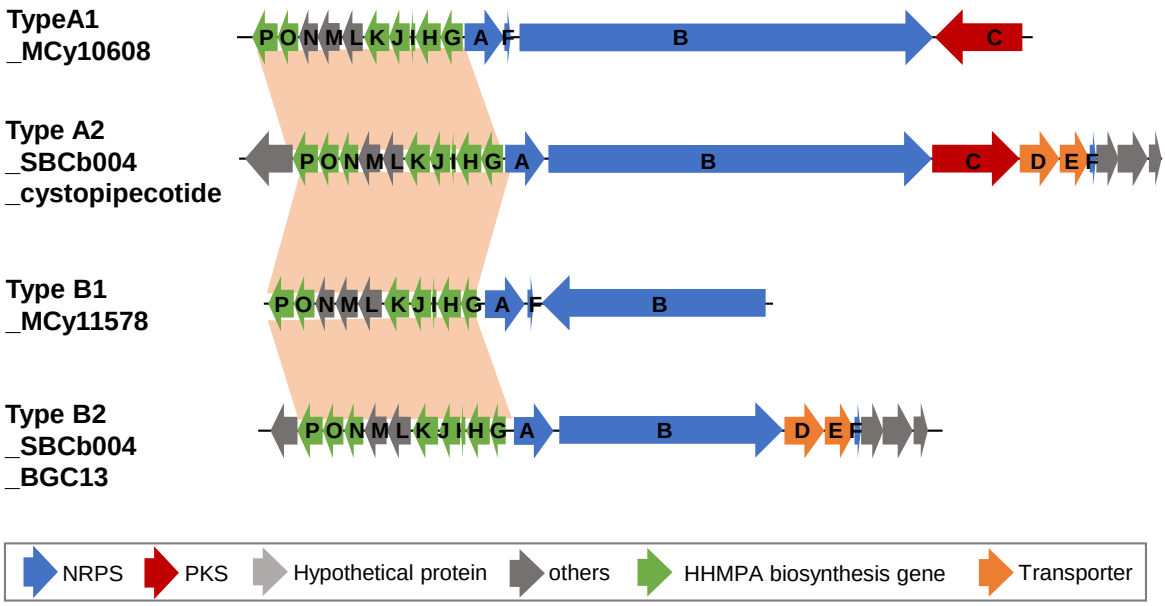


Figure 4. 6. Organisation of representative BGCs containing the putative HHMPA biosynthetic operons.

Bioassay. Cystopipecotide A and B were tested against several bacterial pathogen strains and cancer cell lines. However, neither of them exhibited a significant antimicrobial activity against the tested strains even at the concentration of 64 µg/ml (**Table S4. 3**). Only cystopipecotide B was weakly cytotoxic (**Table S4. 11**). Thus, cystopipecotides may be generated to perform some unknown function, which requires further exploration.

Conclusion. In this study, we uncovered a group of novel peptides containing the non-proteinogenic amino acid HHMPA through metabolomics analysis. The biosynthetic gene cluster of the cystopipecotides was identified by an insertional mutagenesis experiment and a putative biosynthetic pathway was proposed. The biosynthesis of HHMPA, which partially mimics the lysine biosynthesis via the AAA pathway and assembles a C_2 unit by transketolases, was reconstituted *in vitro* and could exhibit an alternative pathway to form substituted pipecolate in microbial natural products.

Aknowledgement. Research in R.M.'s laboratory was partially funded by the Bundesministerium für Bildung und Forschung and Deutsche Forschungsgemeinschaft. H.Z. is grateful to China Scholarship Council for a Ph.D. scholarship. We thank our colleagues Jennifer Hermann and Stefanie Neuber for performing and interpreting the bioactivity evaluation tests.

4.4. Experimental Section

Bacterial strains, culture conditions and reagents. *Escherichia coli* DH10B was used for cloning. *E. coli* BL21 (DE3) was used as a host to express *cpc* single genes. *E. coli* GB05-red was used for gene deletion of HHMPA biosynthesis operon[46]. *E. coli* cells were cultivated at 37 °C in LB medium (1 % trypton, 0.5 % yeast extract, 0.5 % NaCl). *Cystobacter* sp. SBCb004 was cultivated at 30 °C and 180 rpm in CYH medium (0.3 % casitone, 0.3 % yeast extract, 0.2 % $CaCl_2 \bullet 2H_2O$, 0.2 % soy meal, 0.2 % glucose, 0.8 % soluble starch, 0.1 % $MgSO_4 \bullet 7H_2O$, 1.19 % HEPES, 8 mg/L Fe-EDTA, adjusted to pH 7.3 with 10 N KOH) for cystopipecotide production analysis and compound isolation while in M medium (1 % phytone, 1 % maltose, 0.1 % $CaCl_2 \bullet 2H_2O$, 0.1 % $MgSO_4 \bullet 7H_2O$, 8 mg/L Fe-EDTA, 1.19 % HEPES, adjusted to pH 7.2 with 10 N KOH) for transformation and feeding. *Myxococcus xanthus* DK1622 was cultivated in CTT medium (10 g/L casitone; 10 mM Tris-HCl, pH 8.0; 8 mM $MgSO_4$; 10 mM potassium phosphate, pH 7.6) for seed culture, and in CYH medium for HHMPA production analysis and compound isolation. 50 µg/mL kanamycin, 50 µg/mL spectinomycin and 35 µg/mL chloramphenicol were added to the culture when necessary.

Restriction endonucleases, Taq DNA polymerase, Phusion DNA polymerase, FastAP, T4 DNA ligase were purchased from Thermo scientific. NEBuilder HiFi DNA assembly Master Mix was purchased from NEB. All the isotope precursors, unlabeled amino acids, Isopropyl β-D-1-thiogalactopyranoside (IPTG), *N*-methylhydroxylamine hydrochloride, fluorenylmethyloxycarbonyl chloride (Fmoc-Cl), 1,4-Dithiothreitol (DTT), Thiamine pyrophosphate (ThDP) and β-Nicotinamide adenine dinucleotide 2'-phosphate reduced tetrasodium salt hydrate (NADPH) were purchased from Sigma-Aldrich. D-fructose-6-phosphate (F-6-P) was purchased from Biomol. Oligonucleotide primers were synthesised by Sigma-Aldrich and DNA sequencing of PCR products were performed by LGC, Biosearch Technologies.

Gene inactivation in *Cystobacter* sp. SBCb004. Gene inactivation in *Cystobacter* sp. SBCb004 was accomplished by homologous recombination of a suitable plasmid backbone in the encoding region. Internal DNA fragments of the target genes served as homologous regions to mediate the insertion of the constructed plasmids into the target gene via homologous recombination. Internal DNA regions of the target gene with a size of 800 bp to 1200 bp were amplified from *Cystobacter* sp. SBCb004 genomic DNA by using the corresponding primers 9101-sc-gene-F and 9101-sc-gene-R (**Table S4. 4**). Subsequently, these DNA fragments were cloned into the plasmid pbluekan kindly provided by Dr. Carsten Volz (unpublished). pbluekan encodes a aminoglycoside phosphotransferase (*aph*(3')-II) conferring kanamycin resistance[47] as well as a β-lactam resistance cassette (*bla*) and a *lacZ* gene under the control of the T7A1 promoter[48]. Stop codons were introduced at both ends of the cloned homologous regions upon PCR amplification by using mutated oligonucleotide primer. The constructed plasmids (**Table S4. 5**) were confirmed by Sanger sequencing using the primer seq-pblueCN-R. Correct plasmids were introduced into *Cystobacter* sp. SBCb004 by electroporation following a previously published protocol[21]. The correct transformants were confirmed by PCR reactions with the primer CN-check-R, which binds the plasmid backbone, and primer gene_CK_F (where "gene" stands for the designation of the target gene), which binds the upstream of the target gene.

UPLC-MS/MS Measurement. The measurements of crude extracts and pure compounds were performed on a Dionex UltiMate 3000 RSLC system equipped with a Waters BEH C18 column (100 × 2.1 mm, 1.7 µm). The flow rate was 0.6 mL/min and the column temperature was 45 °C. The gradient was as follows: 0 - 0.5 min, 95 % water with 0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 0.5 - 18.5 min, 5 - 95 % B; 18.5 - 20.5 min, 95 % B; 20.5 - 21 min, 95 - 5 % B; 21 - 22.5 min, 5 % B. The measurements of Fmoc-Cl derivatised HHMPA were performed using a Waters BEH C18 column (50 × 2.1 mm, 1.7 µm). The gradient was as follows: 0 - 0.5 min, 5 % B; 0.5 - 9.5 min, 5 - 95 % B; 9.5 - 11.5 min, 95 % B; 11.5 - 12 min, 95 - 5 % B; 12 - 13.5 min, 5 % B. UV spectra were recorded by a DAD ranging from 200 to 600 nm. The LC flow was split to 75 µL/min before entering the Bruker Daltonics maXis 4G hr-qToF mass spectrometer using the Apollo II ESI source. Mass spectra were acquired in centroid mode ranging from 150 - 2500 *m/z* at a 2 Hz full scan rate. Full scan and MS/MS spectra were acquired at 2 Hz.

Intact protein measurements. Intact protein measurements were performed on a Dionex UltiMate 3000 RSLC system using an Aeris Widepore XB-C8 column (150 × 2.1 mm, 3.6 µm, Phenomenex). The flow rate was 0.3 mL/min and the column temperature was 45 °C. The separation of 1 µL sample was achieved using the following gradient: 0 - 0.5 min, 98 % water with 0.1 % formic acid (A) and 2 % acetonitrile with 0.1 % formic acid (B); 0.5 - 10.5 min, 2 - 75 % B; 10.5 - 13.5 min, 75 % B; 13.5 - 14 min, 75 - 2 % B; 14-15.5 min, 2 % B. UV spectra were recorded by a DAD ranging from 200 to 600 nm. The LC flow was split to 75 µL/min before entering the maXis-4 hr-qToF mass spectrometer (Bruker Daltonics) using the Apollo II ESI source. In the source region, the temperature was set to 200 °C, the capillary voltage was 4,000 V, the dry-gas flow was 5.0 L/min and the nebuliser was set to 1.0 bar. Mass spectra were acquired in positive ionisation mode in the range from 600 - 1800 *m/z* at a scan rate of 2.5 Hz. Protein masses were deconvoluted using the Maximum Entropy algorithm (Copyright 1991-2004 Spectrum Square Associates, Inc.).

Isolation and Purification of Cystopipecotides A and B. *Cystobacter* sp. SBCb004 was grown in 20 L CYH medium supplemented with 2 % Amberlite XAD-16 resin at 30 °C for 10 days. Cells and XAD-16 resin were harvested by centrifugation and extracted with 3 × 1.5 L acetone. The combined extracts were partitioned between hexane and methanol (3:1, v/v) and the methanol extract was subsequently partitioned between ethyl acetate and deionised water (3:1, v/v). The water phase extract was separated by semi-preparative HPLC (Waters XSelect CSH Phenyl-Hexyl, 5 µm, 250 × 10 mm; 5 mL/min; UV 220 nm) using a gradient as follows: 0 - 5 min, 95 % H_2O with 0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 5 - 8 min, 5 % B to 23 % B; 8 - 31 min, 23 % to 23.7 % B; 31 - 33 min, 23.7 % to 95 % B; 33 - 34 min, 95 % to 5 % B; 34 - 36 min, 5 % B. Two impure fractions collected at 16.2 min and 21.9 min were obtained.

The impure fraction collected at t_R = 21.9 min was further purified on a Waters Xbridge Peptide BEH C18 HPLC column (10 × 250 mm, 5 µm; 5 mL/min; UV 220 nm) using the following gradient: 0 - 5 min, 95 % H_2O with 0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 5 - 8 min, 5 % B to 30 % B; 8 - 31 min, 30 % to 30.4 % B; 31 - 33 min, 30.4 % to 95 % B; 33 - 34 min, 95 % to 5 % B; 34 - 36 min, 5 % B. The resulting impure fraction was further purified on a Phenomenex Kinetex Biphenyl HPLC column (10 × 250 mm, 5 µm; 5 mL/min; UV 220 nm) using the following gradient: 0 - 5 min, 95 % H_2O with 0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 5 - 8 min, 5 % B to 30 % B; 8 - 31 min, 30 % to 30.3 % B; 31 - 33 min, 30.3 % to 95 % B; 33 -34 min, 95 % to 5 % B; 34 - 36 min, 5 % B, affording cystopipecotide A (9.8 mg, t_R = 16.5 min).

The impure fraction collected at t_R = 16.2 min was further purified on a Waters Xbridge Peptide BEH C18 HPLC column (10 × 250 mm, 5 µm; 5 mL/min; UV 220 nm) using the following gradient: 0 - 5 min, 95 % H_2O with 0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 5 - 8 min, 5 % B to 30 % B; 8 - 31 min, 30 % to 30.4 % B; 31 - 33 min, 30.4 % to 95 % B; 33 - 34 min, 95 % to 5 % B; 34 - 36 min, 5 % B. The resulting impure fraction was further purified on a Phenomenex Kinetex Biphenyl HPLC column (10 × 250 mm, 5 µm; 5 mL/min; UV 220 nm) using the following gradient: 0 - 5 min, 95 % H_2O with 0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 5 - 8 min, 5 % B to 25 % B; 8 - 31 min, 25 % to 26.5 % B; 31 - 33 min, 26.5 % to 95 % B; 33 - 34 min, 95 % to 5 % B; 34 - 36 min, 5 % B, affording cystopipecotide B (4.2 mg, t_R = 17.0 min).

Structure elucidation. UV/Vis data were obtained using the DAD of the Dionex Ultimate 3000 SL system linked to a Bruker maXis 4G UHRqTOF for mass spectrometric analysis.

NMR data were recorded on an AVANCE III 700 MHz NMR (1H at 700 MHz, 13C at 175 MHz) equipped with a 5 mm inverse TCI cryoprobe (Bruker, Billerica, MA, USA). Shift values (δ) are given in ppm and coupling constants (J) in Hz. Measurements were conducted in 5 mm Shigemi tubes (Shigemi Inc., Allison Park, PA 15101, USA). For the two-dimensional experiments, standard pulse programs were used. HMBC experiments were optimised for $^{2,3}J_{C\text{-}H}$ = 6 Hz, and HSQC were optimised for $^{1}J_{C\text{-}H}$ = 145 Hz.

Ozonolysis. 75 µg of **1** was hydrolysed in 50 µL 6 M HCl at 100 °C for 30 min. The mixture was diluted with 1 mL methanol and ozone was bubbled through the solution at -70 °C until blue colour occurred (about 10 min). After flushing the system with oxygen, oxidative workup was conducted at 100 °C for 5 min after adding 200 µL 1 M NaOH and 20 µL 30 % H_2O_2[49,50]. Solvents were evaporated using a Speedvac concentrator (Eppendorf) and the sample was dried *in vacuo* before dissolving it in 50 µL H_2O and submitting it to Marfey's analysis.

Marfey's analysis. 75 µg of **1** was hydrolysed in 50 µL 6 M HCl at 100 °C for 30 min. The hydrolysate was reduced using a Speedvac concentrator (Eppendorf), dried *in vacuo* and redissolved in 50 µL H_2O. After the addition of 60 µL 1 M $NaHCO_3$ and 20 µL 1 % L- or D-FDLA reagent in acetone, respectively, the mixture was incubated for two hours at 40 °C and shaken with 700 rpm. In order to stop the reaction, 30 µL 2 M HCl was added and diluted with 150 µL acetonitrile and 150 µL

methanol. Derivatisation of authentic L-amino acid standards was performed analogously. LC-MS analysis was conducted on a Waters BEH C18 column (2.1 x 100 mm, 1.7 µm) with a gradient using H_2O +0.1 % formic acid and acetonitrile +0.1 % formic acid at a flow rate of 0.55 mL/min at 45 °C. The gradient for the separation of proline, leucine, valine and aspartate derivatives was as follows: from 5-10 % acetonitrile in 1 min, to 35 % in 14 min, to 50 % in 7 min and to 80 % in 3 min. UV spectra were recorded by a DAD ranging from 200 to 600 nm. The LC flow was split to 75 µL/min before entering the Bruker Daltonics maXis 4G hr-qToF mass spectrometer using the Apollo II ESI source. Mass spectra were acquired in centroid mode ranging from 150 - 2500 *m/z* at a 2 Hz full scan rate[30,31].

L-FDLA derivatives eluted at 12.39 min (L-Asp/ L-Trp after ozonolysis), 14.26 min (L-Pro), 15.95 min (L-Val), and 17.58 min (L-Leu). D-FDLA derivatives eluted at 13.17 min (L-Asp/ L-Trp after ozonolysis), 15.90 min (L-Pro), 19.43 min (L-Val), and 20.91 min (L-Leu).

Feeding experiments. The feeding studies of *Cystobacter* sp. SBCb004 were performed in 50 mL M medium. 11.34 mg D-Glucose-1-^{13}C, 11.30 mg D-Glucose-2-^{13}C, 3.87 mg L-Glutamic-2,3,3,4,4-d_5-acid, 5.10 mg L-Valine-d_8, 22.50 mg Sodium propionate-1-^{13}C and 5.20 mg L-Tryptophan-d_8 were resolved in 1 mL water, respectively, while 1.6 mg L-Leucine-5, 5, 5-d_3 were resolved in 1 mL 1 N HCl and sterilised by filtration. Equal aliquots (250 µL) were added to the culture after 24 h, 36 h, 48 h and 60 h. A control experiment without isotopically labelled precursor was done in parallel. 2 % Amberlite XAD-16 resin was added after 72 h. The culture was harvested after 90 h by centrifugation. The cells/XAD were extracted with 50 mL acetone. The solvents were evaporated and the residue was resolved in MeOH and analysed by UPLC-MS.

Heterologous Expression of HHMPA in *M. xanthus DK1622*. Two DNA fragments which cover *cpcG-cpcL* and *cpcL-cpcP*, respectively, were amplified from *Cystobacter* sp. SBCb004 genomic DNA by using the primers listed in **Table S4. 4** and cloned into pFPVan[42] via Gibson Assembly. The resulting plasmid pFPVan-cpcG-P was verified by endonuclease digestion and Sanger sequencing before being transformed into *M. xanthus* DK1622. The transformant was pre-cultured in CTT medium supplemented with kanamycin (50 µg/mL) at 30 °C for two days and then transferred (1 to 100) into 12 L fresh CYH medium containing 50 µg/mL kanamycin and 1 mM vanilate. The cells were grown at 30 °C, 180 rpm for ten days before harvest by centrifugation.

The supernatant was acidified to pH 3.0 by HCl, followed by being applied to a cation exchange column packed with 200 g Dowex 50W X8 (hydrogen form, 16-50 mesh) resin. The resin was then washed with 3 column volumes (CV) H_2O and 5 CV 1 N ammonium hydroxide. HHMPA was eluted with ammonium hydroxide. Fractions with HHMPA were pooled together and were dried by lyophiliser. The residues were then fractionated with Biotage with a 25 g Snapfit column, Versa Flash Spherical C_{18} silica 45-75 µm 70 A. LC conditions were as follows: mobile phase A, H_2O; mobile phase B, methanol; 0 % B, 10 CV; 0-100 % B, 3CV; at a flow rate of 50 mL/min. The HHMPA containing fractions were pooled and the solvents were evaporated. The residues were further purified on a Waters XBridge BEH Amide column (10 × 250 mm, 5 µm; 5 mL/min; UV 220 nm) using the following gradient: 0 - 5 min, 5 % H_2O with 10 mM ammonium formate (A) and 95 % acetonitrile with 10 mM ammonium formate (B); 5 - 20 min, 95 % B to 40 % B; 20 - 23 min, 40 % B; 23 - 24 min, 40 % to 95 % B; 24 - 30 min, 95 % B, affording a impure HHMPA containing fraction (t_R =12.0 min), which was subsequently purified on a semi-preparative HPLC (Waters XSelect CSH Fluro-Phenyl, 5 µm, 250 × 10 mm; 5 mL/min; UV 220 nm) using a gradient as follows: 0 - 10 min, 5 % H_2O with 0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 10 - 12 min, 5 % B to 95 % B; 12 - 15 min, 95 % B; 15-17 min, 95 % B to 5 % B; 17 - 20 min, 5 % B, affording HHMPA (t_R = 2.8 min, 1.6 mg).

Gene Deletion of the HHMPA Biosynthetic Operon in *M. xanthus* DK1622. Gene deletions of *cpcL*, *cpcM* and *cpcN* were achieved by linear-circular recombination conducted by λ-redαβ recombineering system[46]. The plasmid pFPVan-cpcG-P was transformed to *E. coli* GB05-red firstly. The linear gene deletion cassette contains a chloramphenicol resistance gene (cat) and two 39 bp homologs flanking each targeted gene and was amplified by using primers in **Table S4. 4**. The introduction of the gene deletion cassette to *E. coli* GB05-red/pFPVan-cpcG-P and the induction of λ-redαβ recombineering system followed a previously published protocol[51]. The targeted gene deleted plasmids (pFPVan-cpcG-P_del-cpcL, pFPVan-cpcG-P_del-cpcM and pFPVan-cpcG-P_del-cpcN) were verified by digestion and were introduced to *M. xanthus* DK1622, respectively. The strains were cultivated in 50 mL CYH medium following the above-mentioned procedure. The production of HHMPA was checked by derivatisation of the extracts with Fmoc-Cl as described below.

Protein Expression and Purification. The coding sequences of CpcG, CpcH, CpcI, CpcJ, CpcK, CpcN, CpcO and CpcP were separately amplified from *Cystobacter* sp. SBCb004 genomic DNA. All the PCR primers and plasmids used are listed in **Table S4. 4** and **Table S4. 5**. CpcI, CpcJ, CpcO and CpcK were cloned into pHisTEV, while CpcG, CpcP and CpcN were cloned into pHisSUMOTEV (a gift from Dr. David Owen, DIAMOND Light Source) and CpcH was cloned into pCDF-1b (Novagen). The resulting expression plasmids were verified by endonuclease digestion and Sanger sequencing before being transformed into *E. coli* BL21 (DE3). CpcG and CpcH were coexpressed. Each transformant was pre-cultured in LB medium supplemented with kanamycin (50 μg/mL, for vecto pHisTEV and pHisSUMOTEV) or spectinomycin (50 μg/mL, for vector pCDF-1b) at 37 °C overnight and then transferred (1 to 100) into fresh LB medium containing appropriate antibiotics. The *E. coli* cells were grown at 37 °C till the optical density (OD_{600}) reached 0.6. The cultures were then transferred to 16 °C followed by adding 0.1 mM IPTG to induce the protein expression. The cells were continued at 16 °C overnight before being harvested by centrifugation.

The cell pellets except those used to co-express CpcG and CpcH, were resuspended in lysis buffer (20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 20 mM imidazole, 1 mM DTT, 10 % glycerol). For co-purification of CpcG and CpcH, lysis buffer (20 mM Bis-Tris, pH 6.8, 200 mM NaCl, 20 mM imidazole, 1 mM DTT, 10 % glycerol) was used. The cell suspension was lysed via passing through a cell disrupter (Constant Systems) at 30,000 psi, and cell debris was removed by centrifugation (23,500 rpm, 4 °C, 30 min). The supernatant was collected and directly loaded onto a 5-mL Ni-NTA Superflow Cartridge (QIAGEN) pre-equilibrated with lysis buffer. The column was washed extensively with 30 column volumes lysis buffer before the protein was eluted with lysis buffer supplemented with 250 mM imidazole. Fractions containing the targeted protein were directly loaded onto a gel filtration column (HiPrep™ 26/100, GE healthcare) pre-equilibrated in gel filtration buffer (20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM DTT). The fractions of the highest purity indicated by SDS-PAGE were pooled and concentrated by Amicon Ultra centrifugal filters (Merck Millipore). Concentrated proteins were flash frozen by liquid nitrogen and stored at -80 °C. Protein concentrations were measured by Nanodrop UV-Vis spectrometer.

Enzymatic Reactions *in vitro*

Substrate specificity of CpcI and CpcJ. To determine the native amino acid substrate of CpcI and CpcJ, 20 mM L-Glutamic acid, L-Aspartic acid, L-Arginine or L-Lysine was mixed with 10 mM ATP, 10 mM $MgCl_2$, 10 μM CpcI, 1 μM CpcJ and 100 mM Tris-HCl (pH 8.0) and incubated at 30 °C for 3 h. The reaction product was applied to intact protein measurement.

Idenfication of CpcI-γ-Glu-phosphate generated by CpcI, CpcJ and CpcO *in vitro*. To determine CpcI-γ-Glu-phosphate, 20 mM L-Glutamic acid, 10 mM ATP, 10 mM $MgCl_2$, 20 μM CpcI, 1 μM CpcJ, 1 μM CpcO, 100 mM Tris-HCl (pH 8.0) and 160 mM *N*-Methylhydroxylamine hydrochloride were mixed and incubated at 30 °C for 3 h. The reaction product was applied to intact protein measurement.

Detection of CpcI-γ-Glu-semialdehyde produced by CpcI, CpcJ, CpcO and CpcK *in vitro*. CpcI-γ-Glu-semialdehyde was produced *in vitro* reactions as follows. 20 mM L-Glutamic acid, 10 mM ATP, 10 mM $MgCl_2$, 20 μM CpcI, 1 μM CpcJ, 1 μM CpcO, 1 μM CpcK, 1mM NADPH and 100 mM Tris-HCl (pH 8.0) were mixed and incubated at 30 °C for 3 h. The reaction product was applied to intact protein measurement.

Analysis of reactions mediated by CpcI, CpcJ, CpcO, CpcK and CpcG/CpcH. The reactions were conducted with 20 mM L-Glutamic acid, 10 mM ATP, 10 mM $MgCl_2$, 20 μM CpcI, 1 μM CpcJ, 1 μM CpcO, 1 μM CpcK, 0.3 μM CpcG/CpcH, 50 mM D-fructose-6-phosphate (F-6-P), 0.1 mM thiamine pyrophosphate (ThDP), 1mM NADPH and 100 mM Tris-HCl (pH 8.0) at 30 °C for 3 h. The reaction product was applied to intact protein measurement.

Reactions mediated by CpcI, CpcJ, CpcO, CpcK, CpcG/CpcH and CpcP. The reactions were conducted with 20 mM L-Glutamic acid, 10 mM ATP, 10 mM $MgCl_2$, 20 μM CpcI, 1 μM CpcJ, 1 μM CpcO, 1 μM CpcK, 0.3 μM CpcG/CpcH, 0.8 μM CpcP, 50 mM D-fructose-6-phosphate (F-6-P), 0.1 mM thiamine pyrophosphate (ThDP), 1mM NADPH and 100 mM Tris-HCl (pH 8.0) and incubated at 30 °C for 3 h. The reaction product was applied to intact protein measurement.

***In vitro* synthesis of HHMPA.** HHMPA was prepared by *in vitro* reactions as follows. 20 mM L-Glutamic acid, 10 mM ATP, 10 mM $MgCl_2$, 170 μM CpcI, 60 μM CpcJ, 3 μM CpcO, 3 μM CpcK, 0.3 μM CpcG/CpcH, 1 μM CpcP, 1 μM CpcN, 50 mM D-fructose-6-phosphate (F-6-P), 0.1 mM thiamine pyrophosphate (ThDP), 1mM NADPH and 100 mM PBS (pH 6.5) were mixed and incubated at 30 °C overnight.

Fmoc-Cl derivatisation was used to determine the production of HHMPA as follows. The reaction was dried *in vacuo*, and was resuspended in 120 μL 83 % acetonitrile. After centrifugation, 10 μL 0.2 M sodium borate (pH 10.5) and 10 μL 100 mM Fmoc-Cl were added to the supernatant and the mixture was incubated at 50 °C for 30 min followed by LC-MS analysis.

Antimicrobial Activity Test. All tested pathogens were obtained from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung für Mikroorganismen und Zellkulturen, DSMZ). For micro dilution assays, overnight cultures of each strain were diluted in the growth medium to approximately 10^6 cfu/mL and added to sterile 96-well plates. The cell suspension was treated with the respective compounds in serial dilution. After incubation on a microplate shaker (750 rpm, 37 °C) for 16 h, the growth inhibition was assessed by visual inspection.

Cytotoxicity assay. Cell lines were obtained from the DSMZ and were cultured under conditions as recommended by the depositor. 6×10^3 cells were seeded in 180 µL complete medium in each well of the 96-well plates and treated with compounds in serial dilution after 2 h of equilibration. Both compounds and the internal solvent control was tested in duplicates. After 5 d of incubation, 20 µL of 5 mg/mL MTT (Thiazolyl blue tetrazolium bromide) in PBS was added per well and it was continued to be at 37 °C for 2 h. The medium was then discarded and cells were washed with 100 µL PBS before adding 100 µL 2-propanol/10 N HCl (250:1) in order to dissolve formazan granules. The absorbance at 570 nm was measured using a microplate reader and cell viability was expressed as percentage relative to the respective methanol control. The half-maximal inhibitory concentration (IC_{50}) values were determined by sigmoidal curve fitting by plotting the relative growth (%) versus applied concentrations using 0 % (full inhibition) and 100 % (no inhibition) as borders.

Sequence of cystopipecotide biosynthetic gene cluster in *Cystobacter* sp. SBCb004. The major sequence of the cystopipecotide BGC and BGC13 in *Cystobacter* sp. SBCb004 were provided by 454 technology[52,53]. The gaps in genes *cpcB*, *cpcC, cpcL*, *orf7* and *BGC13_L* and the gap between *BGC13_B* and *BGC13_C* were refilled by PCR and Sanger sequencing with the primers 9104-gene-gap-F and 9104-gene-gap-R in **Table S4. 4** ("gene" refers to the gene with a gap). The complete sequence of the cystopipecotide BGC (ON668024) and BGC13 (ON668025) have been deposited in GenBank.

4.5. References

1. Zeng, H. Exploring the biosynthetic potential of *Cystobacter* sp. SBCb004. Dissertation. Saarland University, 2021.
2. Wenzel, S. C. & Müller, R. Myxobacteria—''microbial factories' for the production of bioactive secondary metabolites. *Mol. Biosyst.* **5,** 567–574; 10.1039/b901287g (2009).
3. Wenzel, S. C. & Müller, R. The biosynthetic potential of myxobacteria and their impact on drug discovery. *Curr. Opin. Drug Discov. Devel.* **12,** 220–230 (2009).
4. Bader, C. D., Panter, F. & Müller, R. In depth natural product discovery - Myxobacterial strains that provided multiple secondary metabolites. *Biotechnol. Adv.* **39,** 107480; 10.1016/j.biotechadv.2019.107480 (2020).
5. Mohr, K. I. Diversity of Myxobacteria-We Only See the Tip of the Iceberg. *Microorganisms* **6**; 10.3390/microorganisms6030084 (2018).
6. Garcia, R., Gemperlein, K. & Müller, R. Minicystis rosea gen. nov., sp. nov., a polyunsaturated fatty acid-rich and steroid-producing soil myxobacterium. *International Journal of Systematic and Evolutionary Microbiology* **64,** 3733–3742; 10.1099/ijs.0.068270-0 (2014).
7. Han, K. *et al.* Extraordinary expansion of a Sorangium cellulosum genome from an alkaline milieu. *Scientific Reports* **3,** 2101; 10.1038/srep02101 (2013).
8. Wenzel, S. C. & Müller, R. The impact of genomics on the exploitation of the myxobacterial secondary metabolome. *Nat. Prod. Rep.* **26,** 1385–1407; 10.1039/b817073h (2009).
9. Herrmann, J., Fayad, A. A. & Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **34,** 135–160; 10.1039/C6NP00106H (2017).
10. Gavriilidou, A. *et al.* Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nat. Microbiol.* **7,** 726–735; 10.1038/s41564-022-01110-2 (2022).
11. Zaburannyi, N., Bunk, B., Maier, J., Overmann, J. & Müller, R. Genome analysis of the fruiting body forming myxobacterium *Chondromyces crocatus r*eveals high potential for natural product Biosynthesis. *Appl. Environ. Microbiol.* **82,** 1945–1957; 10.1128/AEM.03011-15 (2016).
12. Gregory, K., Salvador, L. A., Akbar, S., Adaikpoh, B. I. & Stevens, D. C. Survey of Biosynthetic Gene Clusters from Sequenced Myxobacteria Reveals Unexplored Biosynthetic Potential. *Microorganisms* **7,** 181; 10.3390/microorganisms7060181 (2019).

13. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Natural product reports* **18,** 380–416 (2001).

14. Sieber, S. A. & Marahiel, M. A. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem Rev* **105,** 715–738; 10.1021/cr0301191 (2005).

15. Wenzel, S. C. & Müller, R. Myxobacterial natural product assembly lines: fascinating examples of curious biochemistry. *Natural product reports* **24,** 1211–1224; 10.1039/b706416k (2007).

16. Süssmuth, R. D. & Mainz, A. Nonribosomal peptide synthesis - Principles and prospects. *Angew. Chem. Int. Ed.* **56,** 3770–3821; 10.1002/anie.201609079 (2017).

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

18. Hwang, S., Lee, N., Cho, S., Palsson, B. & Cho, B.-K. Repurposing Modular Polyketide Synthases and Non-ribosomal Peptide Synthetases for Novel Chemical Biosynthesis. *Front. Mol. Biosci.* **7,** 87; 10.3389/fmolb.2020.00087 (2020).

19. Zeng, H. *et al.* Expanding the Ajudazol Cytotoxin Scaffold: Insights from Genome Mining, Biosynthetic Investigations, and Novel Derivatives. *Journal of natural products* **85,** 2610–2619; 10.1021/acs.jnatprod.2c00637 (2022).

20. 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* **1445,** 185–195 (1999).

21. Chai, Y. *et al.* Discovery of 23 natural tubulysins from *Angiococcus disciformis* An d48 and *Cystobacter* SBCb004. *Chem. Biol.* **17,** 296–309; 10.1016/j.chembiol.2010.01.016 (2010).

22. Ohnishi, Y. *et al.* Genome sequence of the streptomycin-producing microorganism Streptomyces griseus IFO 13350. *Journal of bacteriology* **190,** 4050–4060; 10.1128/JB.00204-08 (2008).

23. Pogorevc, D., Popoff, A., Fayad, A. A., Wenzel, S. C. & Müller, R. Production profile engineering and precursor directed biosynthesis approaches for production of novel argyrin derivatives **unpublished results**.

24. 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.* **31,** 768–783; 10.1039/c3np70127a (2014).

25. Linington, R. G. Npatlas - The Natural Products Atlas. Available at https://www.npatlas.org (2018).

26. Dictionary of Natural Products. Available at http://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml (2019).

27. Gorges, J. *et al.* Structure, Total Synthesis, and Biosynthesis of Chloromyxamides: Myxobacterial Tetrapeptides Featuring an Uncommon 6-Chloromethyl-5-methoxypipecolic Acid Building Block. *Angewandte Chemie (International ed. in English)* **57,** 14270–14275; 10.1002/anie.201808028 (2018).

28. Hoffmann, T., Müller, S., Nadmid, S., Garcia, R. & Müller, R. Microsclerodermins from terrestrial myxobacteria: An intriguing biosynthesis likely connected to a sponge symbiont. *J. Am. Chem. Soc.* **45,** 16904–16911; 10.1021/ja4054509 (2013).

29. Zang, Y. *et al.* Griseofamines A and B: Two Indole-Tetramic Acid Alkaloids with 6/5/6/5 and 6/5/7/5 Ring Systems from Penicillium griseofulvum. *Org. Lett.* **20,** 2046–2050; 10.1021/acs.orglett.8b00584 (2018).

30. Marfey, P. Determination of D-amino acids. II. Use of a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzene. *Carlsberg Res. Commun.* **49,** 591–596; 10.1007/BF02908688 (1984).

31. B'Hymer, C., Montes-Bayon, M. & Caruso, J. A. Marfey's reagent: past, present, and future uses of 1-fluoro-2,4-dinitrophenyl-5-L-alanine amide. *J. Sep. Sci.* **26,** 7–19 (2003).

32. Bewley, C. A., Debitus, C. & Faulkner, D. J. Microsclerodermins A and B. Antifungal cyclic peptides from the lithistid sponge *Microscleroderma* sp. *J. Am. Chem. Soc.* **116,** 7631–7636; 10.1021/ja00096a020 (1994).

33. Kobayashi, J. *et al.* Keramamides B .apprx. D, novel peptides from the Okinawan marine sponge Theonella sp. *J. Am. Chem. Soc.* **113,** 7812–7813; 10.1021/ja00020a081 (1991).

34. Kjaerulff, L., Benie, A. J., Hoeck, C., Gotfredsen, C. H. & Sørensen, O. W. S(3) HMBC: Spin-State-Selective HMBC for accurate measurement of homonuclear coupling constants. Application to strychnine yielding thirteen hitherto unreported J(HH). *Journal of magnetic resonance (San Diego, Calif. : 1997)* **263,** 101–107; 10.1016/j.jmr.2015.11.007 (2016).

35. He, M. Pipecolic acid in microbes: Biosynthetic routes and enzymes. *J. Ind. Microbiol. Biotechnol.,* 1–7; 10.1007/s10295-006-0078-3 (2006).

36. Chen, M., Liu, C.-T. & Tang, Y. Discovery and Biocatalytic Application of a PLP-Dependent Amino Acid *γ*-Substitution Enzyme That Catalyzes C-C Bond Formation. *J. Am. Chem. Soc.* **142,** 10506–10515; 10.1021/jacs.0c03535 (2020).

37. Hai, Y., Chen, M., Huang, A. & Tang, Y. Biosynthesis of Mycotoxin Fusaric Acid and Application of a PLP-Dependent Enzyme for Chemoenzymatic Synthesis of Substituted l-Pipecolic Acids. *J. Am. Chem. Soc.* **142,** 19668–19677; 10.1021/jacs.0c09352 (2020).

38. Kochetov, G. A. & Solovjeva, O. N. Structure and functioning mechanism of transketolase. *Biochimica et biophysica acta* **1844**; 10.1016/j.bbapap.2014.06.003 (2014).

39. Ouchi, T. *et al.* Lysine and arginine biosyntheses mediated by a common carrier protein in Sulfolobus. *Nat. Chem. Biol.* **9,** 277–283; 10.1038/nchembio.120 (2013).

40. Matsuda, K. *et al.* Genome Mining of Amino Group Carrier Protein-Mediated Machinery: Discovery and Biosynthetic Characterization of a Natural Product with Unique Hydrazone Unit. *ACS Chem. Biol.* **12**; 10.1021/acschembio.6b00818 (2017).

41. Hasebe, F. *et al.* Amino-group carrier-protein-mediated secondary metabolite biosynthesis in Streptomyces. *Nature chemical biology* **12**; 10.1038/nchembio.2181 (2016).

42. Panter, F., Krug, D., Baumann, S. & Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **9,** 4898–4908; 10.1039/C8SC01325J (2018).

43. Horie, A. *et al.* Discovery of proteinaceous N-modification in lysine biosynthesis of Thermus thermophilus. *Nature chemical biology* **5**; 10.1038/nchembio.198 (2009).

44. Yoshida, A. *et al.* Structural insight into amino group-carrier protein-mediated lysine biosynthesis: crystal structure of the LysZ·LysW complex from Thermus thermophilus. *The Journal of biological chemistry* **290,** 435–447; 10.1074/jbc.M114.595983 (2015).

45. FMOC-Cl as derivatizing agent for the analysis of amino acids and dipeptides by the absolute analysis method.

46. Fu, J. *et al.* Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* **30,** 440–446; 10.1038/nbt.2183 (2012).

47. Wright, G. D. & Thompson, P. R. Aminoglycoside phosphotransferases: proteins, structure, and mechanism. *Frontiers in bioscience: a journal and virtual library* **4,** D9-21 (1999).

48. Deuschle, U., Kammerer, W., Gentz, R. & Bujard, H. Promoters of Escherichia coli: a hierarchy of in vivo strength indicates alternate structures. *EMBO J.* **5,** 2987–2994 (1986).

49. Hoffmann, T., Müller, S., Nadmid, S., Garcia, R. & Müller, R. Microsclerodermins from terrestrial myxobacteria: an intriguing biosynthesis likely connected to a sponge symbiont. *Journal of the American Chemical Society* **135,** 16904–16911; 10.1021/ja4054509 (2013).

50. Vollbrecht, L. *et al.* Argyrins, immunosuppressive cyclic peptides from myxobacteria. II. Structure elucidation and stereochemistry. *J. Antibiot.* **55,** 715–721; 10.7164/antibiotics.55.715 (2002).

51. Wang, H. *et al.* RecET direct cloning and Redαβ recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nat. Protoc.* **11,** 1175–1190; 10.1038/nprot.2016.054 (2016).

52. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature*; 10.1038/nature03959 (2005).

53. Rothberg, J. M. & Leamon, J. H. The development and impact of 454 sequencing. *Nat. Biotechnol.* **26,** 1117–1124; 10.1038/nbt1485 (2008).

Supporting Information

Cystopipecotides: Discovery, structure elucidation and biosynthesis of myxobacterial peptides featuring a 5-hydroxyl-6-hydroxymethylpipecolic acid building block

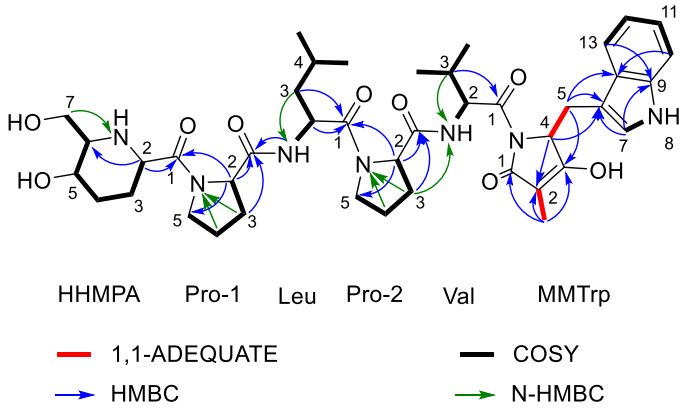
Hu Zeng†, Joy Birkelbach†, Lena Keller, Carsten Volz, Chengzhang Fu and Rolf Müller

Affiliation

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

S 4.1 Structure Elucidation

Table S4. 1. *NMR Spectroscopic Data for Cystopipecotide A (**1**)*[a].



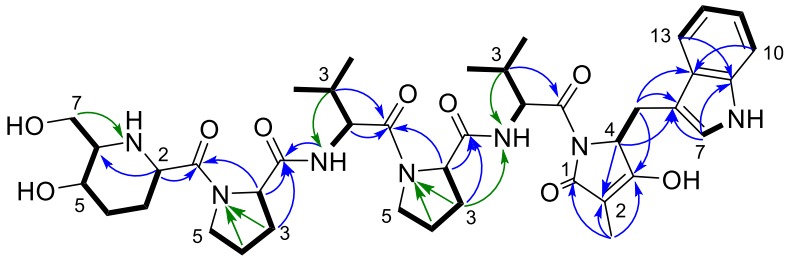
	Pos.	δ_X[b], X type	δ_H[c], (J in Hz)	COSY[d]	HMBC[e]	N-HMBC[f]	1,1-ADEQUATE
HHMPA	1	169.2, C	-	-	-	-	-
	2	51.5, CH	3.86, m	3a, 3b	1, 3, 4, 6, 2(Pro-1)	-	1, 3
	3a	23.3, CH_2	1.87, m	2, 3b, 4a	1, 2, 4, 5	8	-
	3b		1.52, m	2, 3a, 4b	2, 5	8	-
	4a	27.7, CH_2	1.65, m	3a, 4b, 5	2, 3, 5, 6	-	-
	4b		1.74, m	3b, 4a, 5	2, 3, 5, 6	-	-
	5	64.8, CH	3.40, m	4a, 4b, 6	- [g]	-	4, 6
	6	59.2, CH	3.19, m	5, 7b	4, 5, 7	-	5
	7a	60.3, CH_2	3.49, m	7b	5, 6	8	-
	7b		3.64, m	6, 7a	5, 6	8	-
	8	40.3, NH	n.a.	-	-	-	-
Pro-1	1	171.3, C	-	-	-	-	-
	2	58.9, CH	4.37, m	3b, 4	1, 3, 4, 5, 1(HHMPA)	6	1, 3
	3a	28.9, CH_2	1.86, m	3b, 4	2, 4, 5	6	-
	3b		2.03, m	2, 3a, 4	1, 2, 4, 5	6	-
	4	24.6, CH_2	1.76, m	2, 3a, 3b, 5a, 5b	1, 2, 3, 5	6	-
	5a [h]	46.7, CH_2	3.46, m	4, 5b	3, 4	-	4
	5b [h]		3.62, m	4, 5a	3, 4	-	4
	6 [i]	129.2, N	-	-	-	-	-
Leu	1 [j]	170.8, C	-	-	-	-	-

	2	48.6, CH	4.53, m	3b, 7	1, 3, 1(Pro-2)	7	3
	3a	40.4, CH_2	1.45, m	3b, 4	1, 2, 4, 5, 6	7	2, 4
	3b		1.52, m	2, 3a, 4	1, 2, 4, 5, 6	7	2, 4
	4	24.0, CH	1.72, m	3a, 3b, 5, 6	2, 3, 5, 6	-	-
	5	23.3, CH_2	0.90, m	4	3, 4, 6	-	-
	6	21.6, CH_3	0.89, m	4	3, 4, 5	-	-
	7	117.5, NH	8.09, br d (7.7)	2	2, 3, 1(Pro-1)	-	-
Pro-2	1[j]	170.8, C	-	-	-	-	-
	2	59.2, CH	4.55, m	3, 4	1, 3, 4, 5, 1(Leu)	-	1, 3
	3	28.6, CH_2	2.01, m	2, 4	1, 2, 4	6[i]	-
	4	24.6, CH_2	1.91, m	2, 3, 5a, 5b	1, 3, 2, 5	6[i]	3, 5
	5a[h]	46.7, CH_2	3.46, m	4, 5b	3, 4	-	4
	5b[h]		3.62, m	4, 5a	3, 4	-	4
	6[i]	129.2, N	-	-	-	-	-
Val	1	169.7, C	-	-	-	-	-
	2	55.1, CH	5.66, br s	3, 6	-[g]	-	-
	3	30.8, CH	1.98, m	2, 4, 5	1, 2, 4, 5	6[i]	4, 5
	4	16.7, CH_3	0.75, m	3	2, 3, 5	-	3
	5	19.7, CH_3	0.78, m	3	2, 3, 4	-	3
	6	110.4, NH	7.61, br s	2	-[g]	-	-
MMTrp	1	172.4, C	-	-	-	-	1
	2	94.1, C	-	-	-	-	1
	2-Me	6.3, CH_3	1.20, s	-	1, 2, 3	-	2
	3	182.1, C	-	-	-	-	-
	4	60.0, CH	4.20, br s	5a, 5b	1, 3, 5, 6	-	5
	5a	24.4, CH_2	3.22, m	4, 5b, 7	3, 4, 6, 7, 14	15	-
	5b		3.50, m	4, 5a, 7	3, 4, 6, 7, 14	15	-
	6	108.2, C	-	-	-	-	-
	7	123.8, CH	6.80, m	5a, 5b, 8	4, 5, 6, 9, 14	-	6
	8	n.a., NH	10.66, s	7	6, 7, 9, 14	-	-
	9	135.5, C	-	-	7, 8, 12, 13	-	-
	10	110.9, CH	7.20, m	12	11, 13, 14	-	9, 12
	11	117.9, CH	6.82, m	12, 13	9, 10, 12, 13, 14	-	12
	12	120.2, CH	6.94, t (7.5)	10, 11	9, 11, 13, 14	-	10, 11
	13	118.8, CH	7.44, br d (8.0)	11	6, 9, 10, 12, 14	-	14

	14	128.4, C	-	-	-	-	-
	15	163.7, N	-	-	-	-	-

[a]Recorded in DMSO-d_6. [b]Acquired at 175 MHz, adjusted to the solvent signal of DMSO-d_6 (δ_C 39.51 ppm). [c]Acquired at 700 MHz, adjusted to the solvent signal of DMSO-d_6 (δ_H 2.50 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Proton showing HMBC correlation to indicated carbon. [f]Proton showing HMBC correlation to indicated nitrogen. [g]Signals not observed. [h-i]Overlapping signals.

Table S4. 2. *NMR Spectroscopic Data for Cystopipecotide B (**2**)*[a].

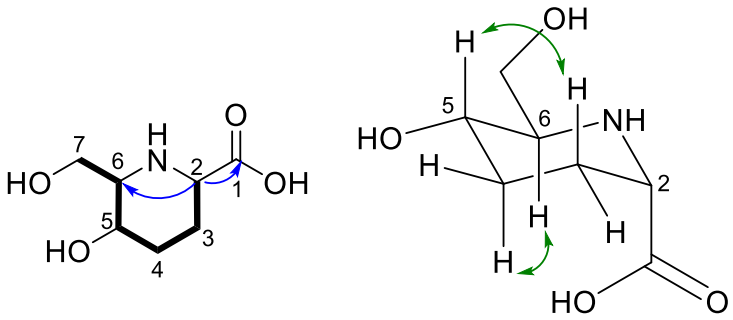


	Position	δ_C[b], type	δ_H[c], (J in Hz)	COSY[d]	HMBC[e]
HHMPA	1	171.9, C	-	-	-
	2	51.1, CH	3.61, m	3a, 3b	1, 3, 4, 6
	3a	24.2, CH_2	1.48, m	2, 3b, 4a, 4b	2, 4
	3b		1.75, m	2, 3a, 4b	2, 4
	4a	28.6, CH_2	1.23, m	3a	-[f]
	4b		1.64, m	3a, 3b, 5	2, 5, 6
	5	65.5, CH	3.28, m	4b	6, 7
	6	58.3, CH	2.84, m	7a, 7b	2, 4, 6, 7
	7a	61.8, CH_2	3.35, m	7b	5, 6
	7b		3.56, m	7a	5, 6
Pro-1	1	171.0, C	-	-	-
	2	59.3, CH	4.50, m	3a, 3b, 4	-[f]
	3a	28.6, CH_2	1.88, m	2, 4	4, 5
	3b		1.99, m	2, 4	1
	4[g]	24.3, CH_2	1.87, m	2, 3a, 3b, 5	3, 5
	5	46.4, CH_2	3.48, m	4	-[f]
Val-1	1[h]	169.3, C	-	-	-
	2	55.2, CH	4.35, m	3, 6	1, 3, 4, 5, 1(Pro-2)

	3[i]	30.1, CH_2	2.00, m	2, 4, 5	1, 2, 4, 5
	4	17.9, CH_3	0.90, m	3	2, 3, 5
	5	19.2, CH_3	0.94, m	3	2, 3, 4
	6	n.a., NH	7.81, br d (8.3)	2	1(Pro-1)
Pro-2	1[h]	169.3, C	-	-	-
	2	58.9, CH	4.41, m	3a, 3b	-[f]
	3a	28.3, CH_2	1.81, m	2, 3b, 4	4, 5
	3b		1.99, m	2, 3a, 4	4, 5
	4[g]	24.3, CH_2	1.87, m	3a, 3b, 5	2, 3, 5
	5	46.8, CH_2	3.62, m	4	3, 4
Val-2	1	169.7, C	-	-	-
	2	57.4, CH	5.54, m	6	-[f]
	3[i]	31.1, CH	2.00, m	4, 5	1, 4, 5
	4	19.5, CH_3	0.78, br d (6.4)	3	3, 5
	5	16.7, CH_3	0.75, br d (6.9)	3	3, 4
	6	n.a., NH	8.11, m	2, 2(Pro-2)	-[f]
MMTrp	1	172.3, C	-	-	-
	2	89.8, C	-	-	-
	2-Me	6.4, CH_3	1.17, m	4	1, 2, 3
	3	186.9, C	-	-	-
	4	60.8, CH	3.92, m	2-Me, 5a, 5b	1, 3, 5, 6
	5a	24.5, CH_2	3.15, m	4, 5b, 7	4, 6, 7, 14
	5b		3.39, d (13.5)	4, 5a, 7	3, 4, 6, 7, 14
	6	108.8, C	-	-	-
	7[j]	123.2, CH	6.80, m	5a, 5b, 8	5, 6, 9, 14
	8	n.a., NH	10.53, s	7, 13	6, 7, 9, 14
	9	135.2, C	-	-	-
	10	110.3, CH	7.19, d (8.1)	12	11, 13, 14
	11[j]	117.2, CH	6.80, m	12, 13	10, 12
	12	119.7, CH	6.92, t (7.5, 7.5)	10, 11	9, 11, 13
	13	118.8, CH	7.48, d (7.8)	8, 11	6, 9, 10, 12, 14
	14	127.9, C	-	-	-

[a]Recorded in DMSO-d_6. [b]Acquired at 175 MHz, adjusted to the solvent signal of DMSO-d_6 (δ_C 39.51 ppm). [c]Acquired at 700 MHz, adjusted to the solvent signal of DMSO-d_6 (δ_H 2.45 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Protons showing HMBC correlations to indicated carbons. [f]Signals not observed. [f-j]Overlapping signals.

Table S4. 3. *NMR Spectroscopic Data for HHMPA derived from the expression of the* cpcG-P *operon in* Myxococcus xanthus *DK1622 (**3**)*[a].



2(*S*),5(*S*),6(*R*)

	Position	δ_C[b], type	δ_H[c], (J in Hz)	COSY[d]	HMBC[e]	NOESY[f]
HHMPA	1	172.6, C	-	-	-	-
	2	55.3, CH	3.94, t (4.2)[g]	3a, 3b	1, 3, 4, 6	3ab
	3a	22.5, CH_2	1.90, m	3b, 4ab, 2	1, 2, 4, 5	2, 3b
	3b		2.28, ddt (15.0, 5.2, 3.8)	3a, 4ab, 2	1, 2, 4, 5	2, 3a, 5
	4a	27.7, CH_2	1.43, m	3ab, 4b, 5	2, 3, 5, 6	4b, 6
	4b		2.00, ddt (15.0, 4.6, 4.1)	3ab, 4a, 5	2, 3, 5, 6	4a, 5
	5	63.9, CH	3.74, td (9.5, 9.6, 4.1)	4ab, 6	3, 6, 7	3a, 4b, 6
	6	59.0, CH	3.39, ddd (9.5, 6.0, 3.5)[g]	5, 7ab	4, 5, 7	4a, 7ab
	7a	57.6, CH_2	3.86, dd (12.6, 6.0)[g]	6, 7b	5	6
	7b		3.91, dd (12.6, 3.6)[g]	6, 7a	5, 6	6

[a]Recorded in D_2O. [b]Acquired at 175 MHz, shifts calculated based on HMBC. [c]Acquired at 700 MHz, adjusted to the solvent signal of D_2O (δ_H 4.75 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Protons showing HMBC correlations to indicated carbons. [f]Using a dwell time of 1.2 s. [g]Coupling constants derived from 1D selective NOESY irrigating the respective proton and using a dwell time of 0.3 s.

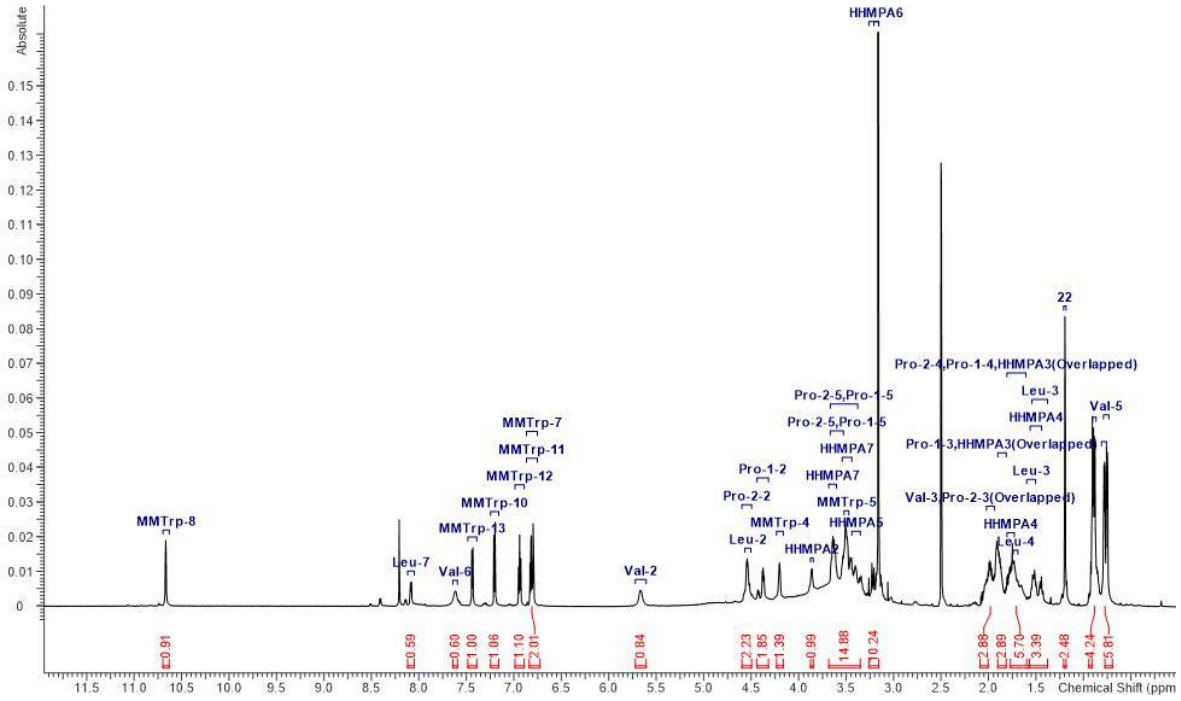


Figure S4. 1. *^{1}H NMR of cystopipecotide A recorded in DMSO-*d_6 *at 700 MHz.*

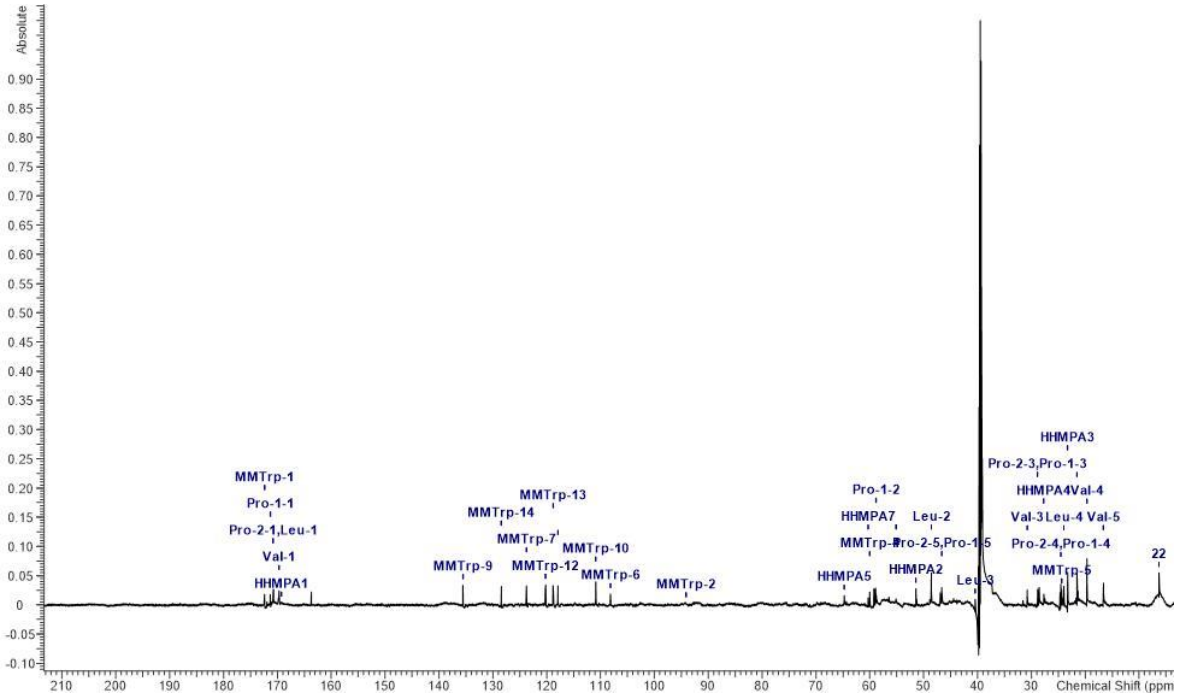


Figure S4. 2. *^{13}C NMR of cystopipecotide A recorded in DMSO-*d_6 *at 175 MHz.*

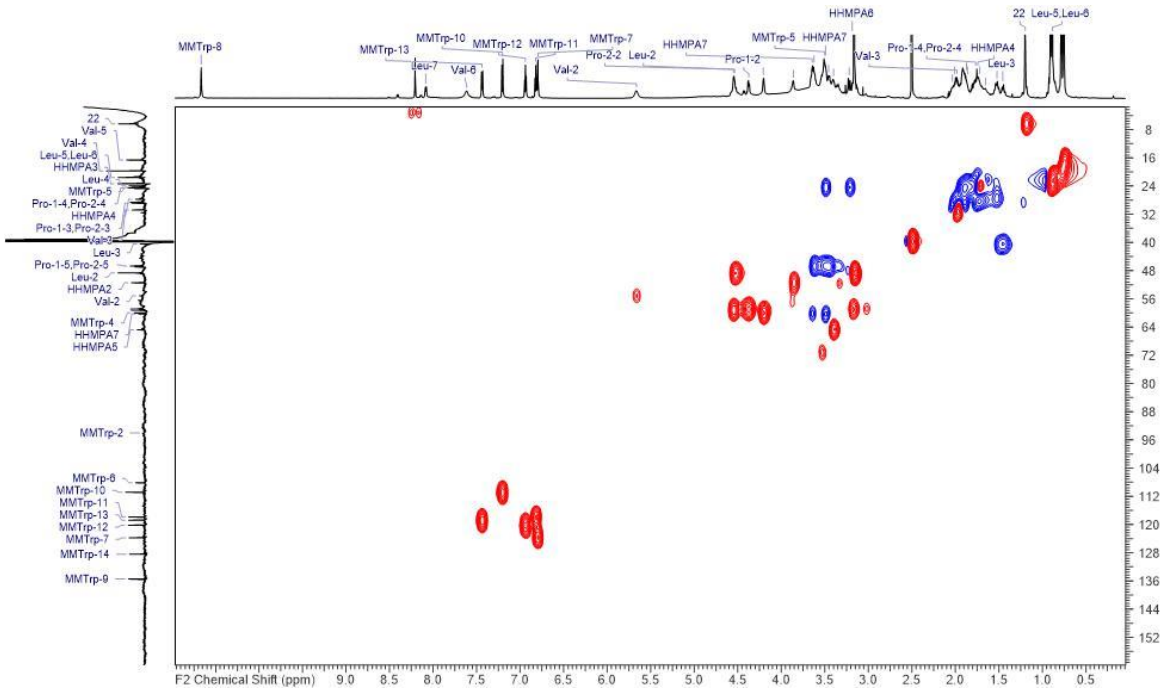


Figure S4. 3. *HSQC NMR of cystopipecotide A recorded in DMSO-*d_6*.*

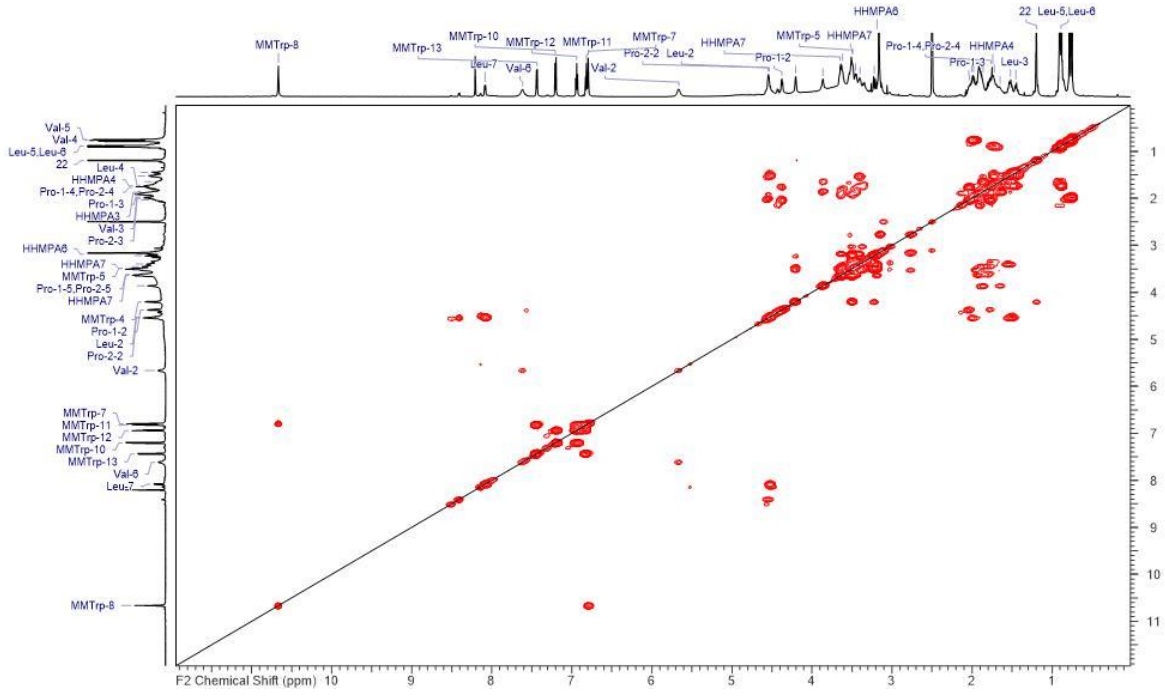


Figure S4. 4. *COSY NMR of cystopipecotide A recorded in DMSO-*d_6*.*

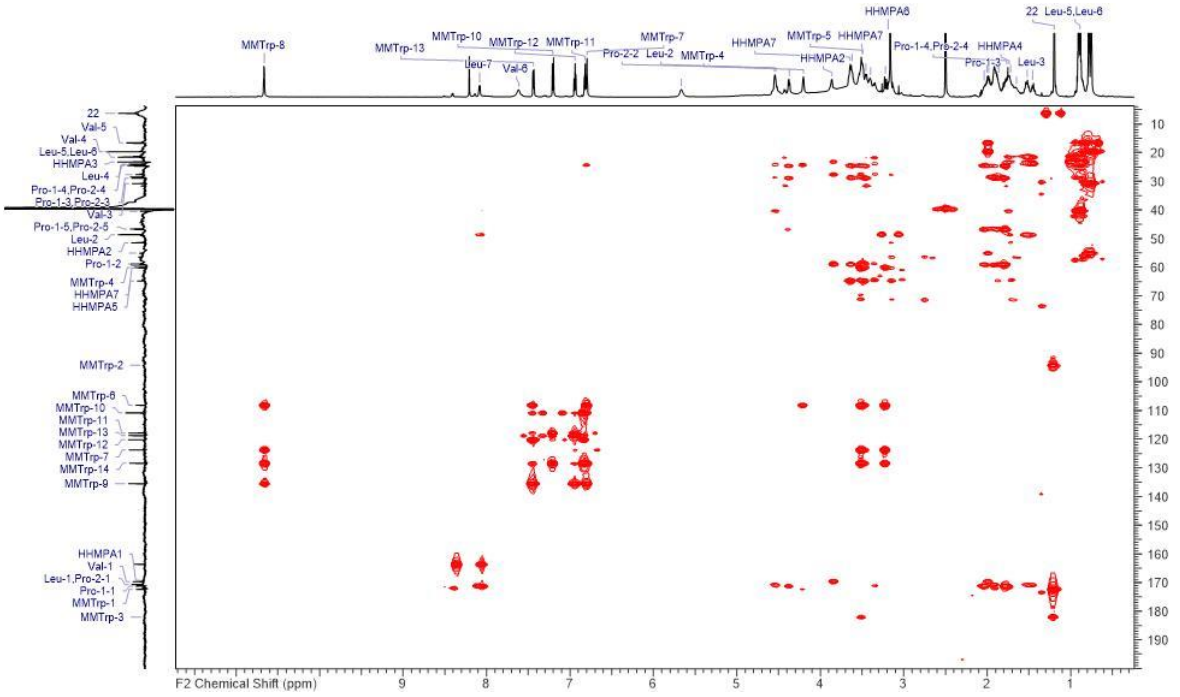


Figure S4. 5. *HMBC NMR of cystopipecotide A recorded in DMSO-*d_6.

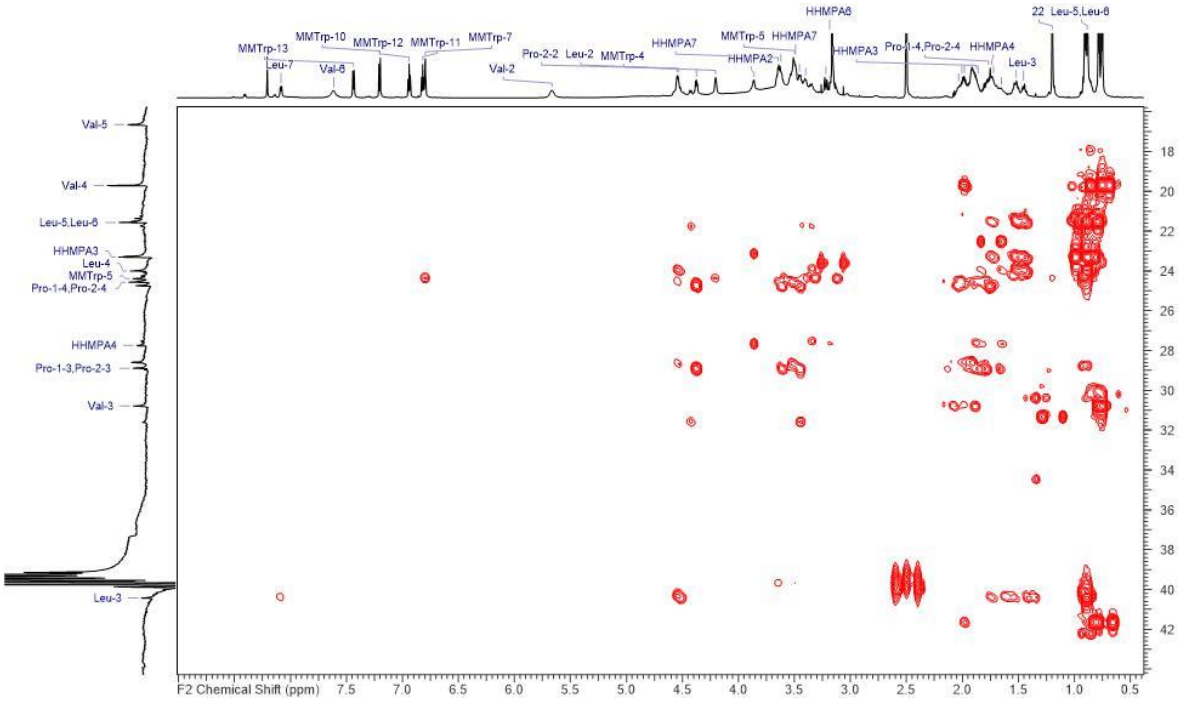


Figure S4. 6. *Selective HMBC NMR (F1 0-42 ppm) of cystopipecotide A recorded in DMSO-*d_6.

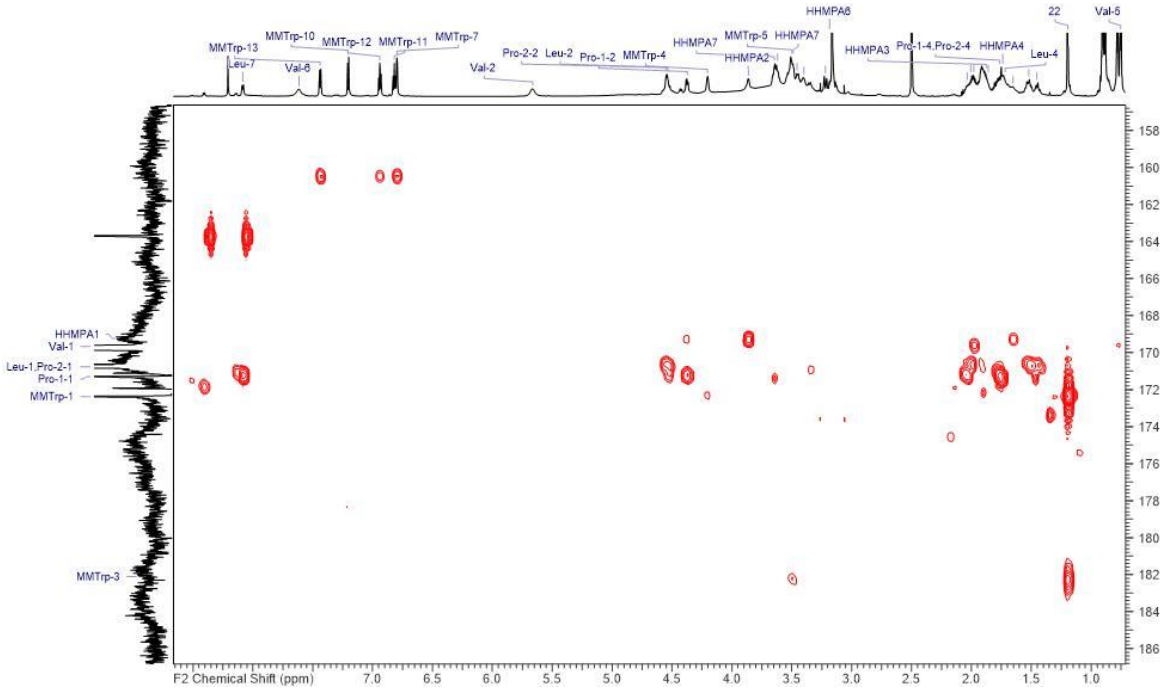


Figure S4. 7. *Selective HMBC NMR (F1 160-190 ppm) of cystopipecotide A recorded in DMSO-*d_6*.*

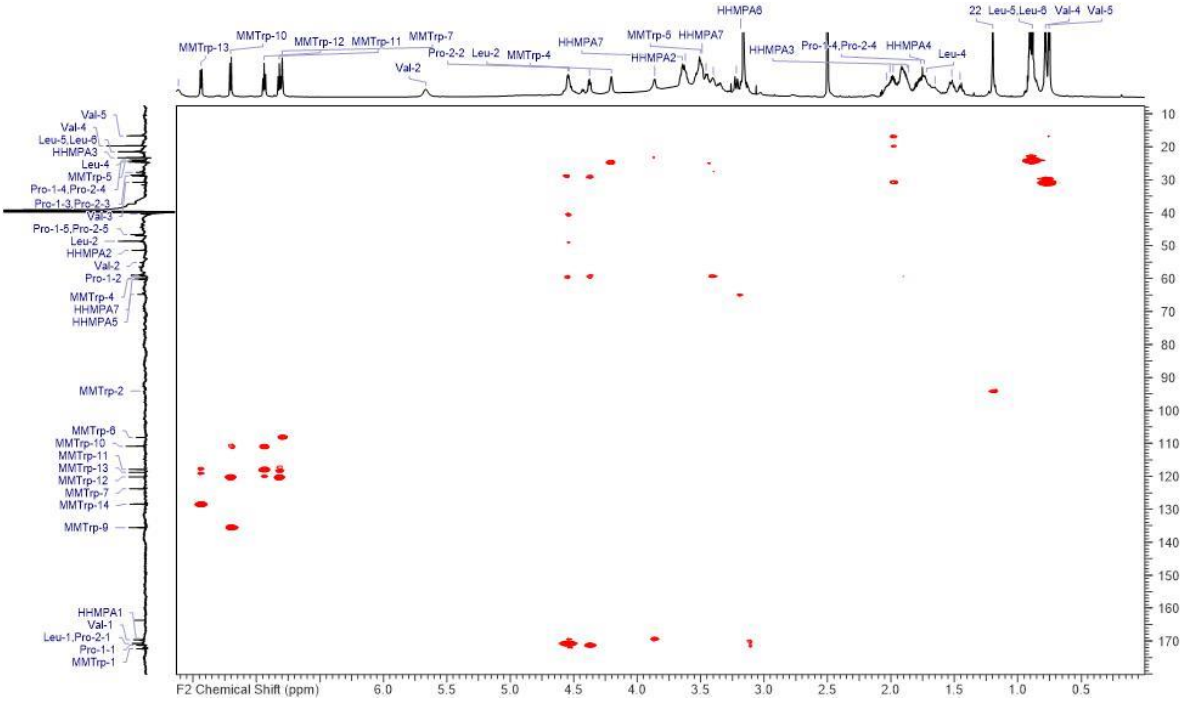


Figure S4. 8. *1,1-ADEQUATE NMR of cystopipecotide A recorded in DMSO-*d_6*.*

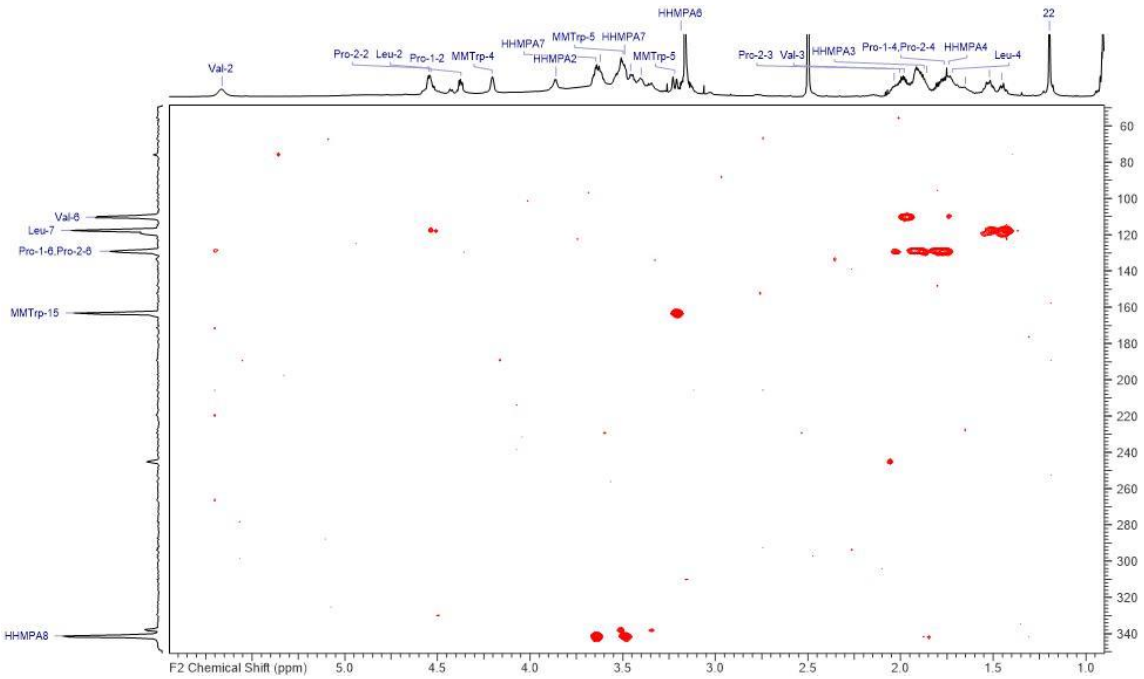


Figure S4. 9. *N-HMBC NMR (F1 50-350 ppm) of cystopipecotide A recorded in DMSO-*d_6*.*

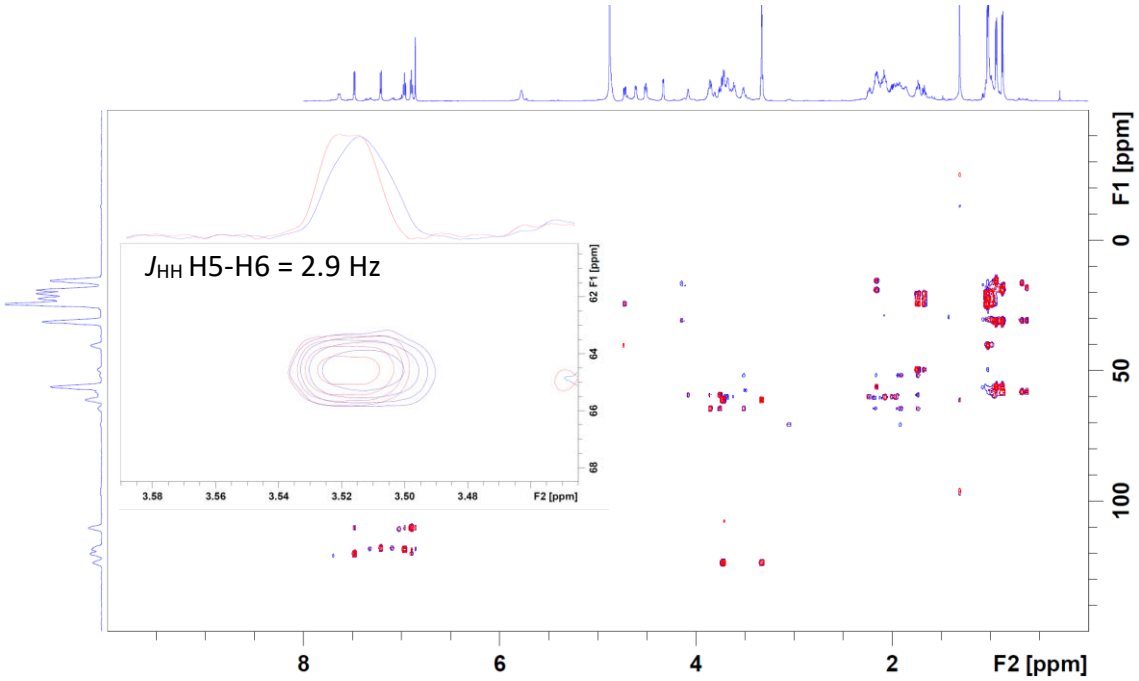


Figure S4. 10. *S^3HMBC of cystopipecotide A; add/substract CH/CH_3 spectra; recorded in methanol-*d_4*.*

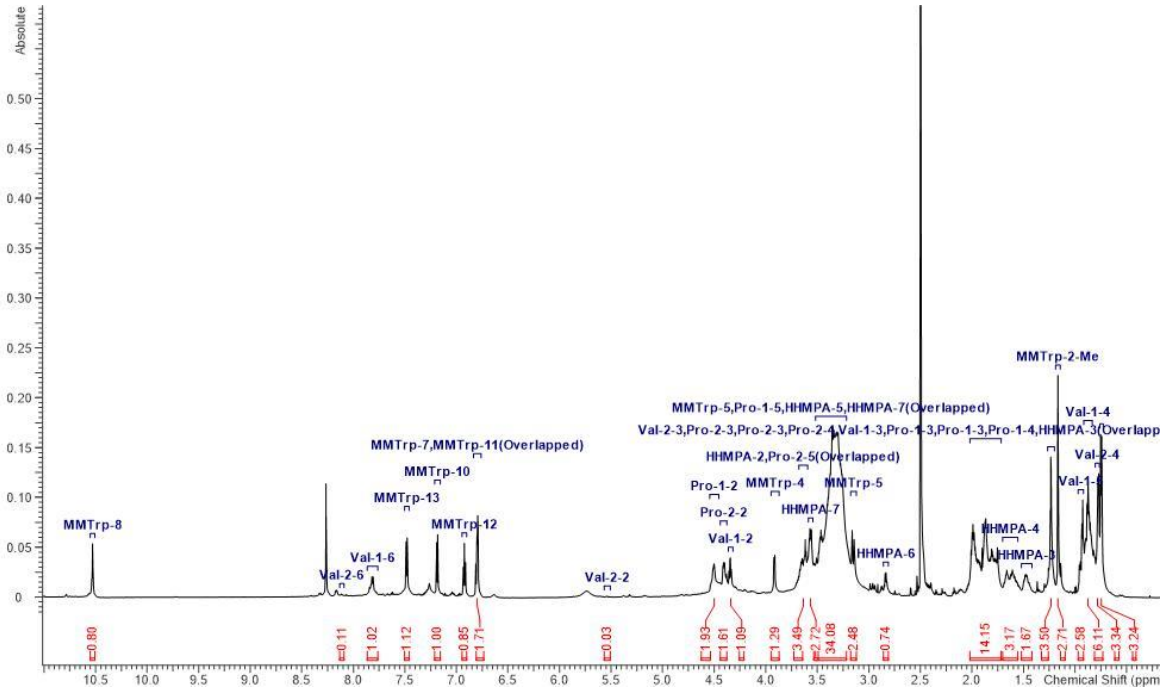


Figure S4. 11. *^{1}H NMR of cystopipecotide B recorded in DMSO-d_6 at 700 MHz.*

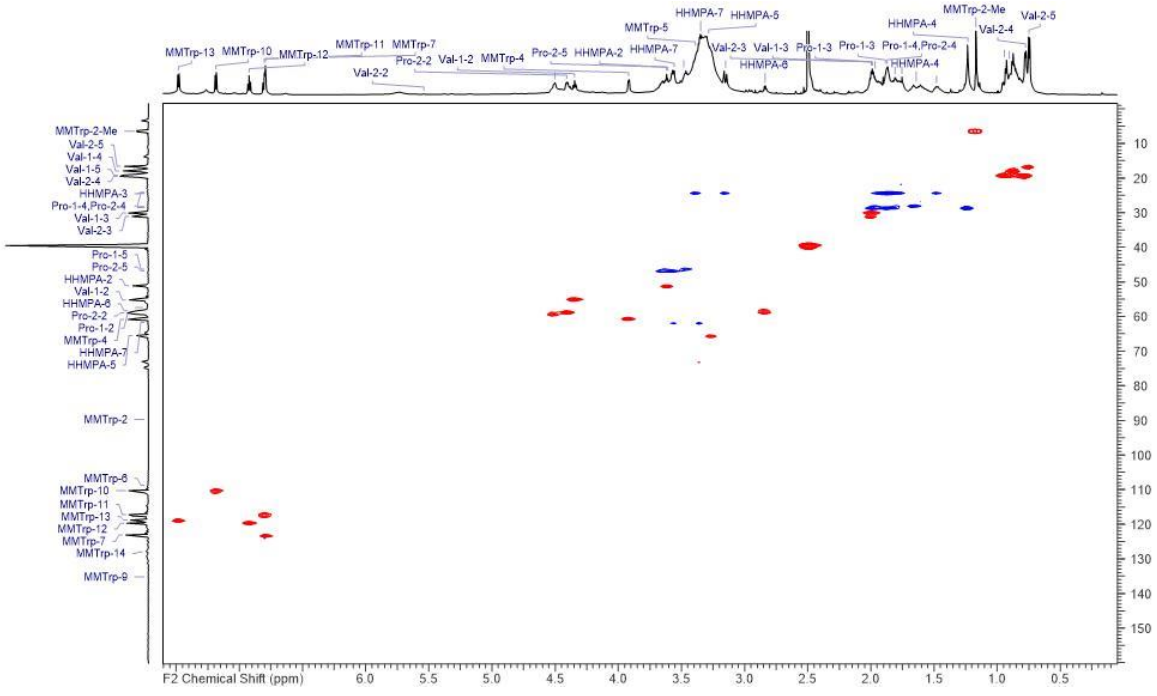


Figure S4. 12. *HSQC NMR of cystopipecotide B recorded in DMSO-d_6 at 175 MHz.*

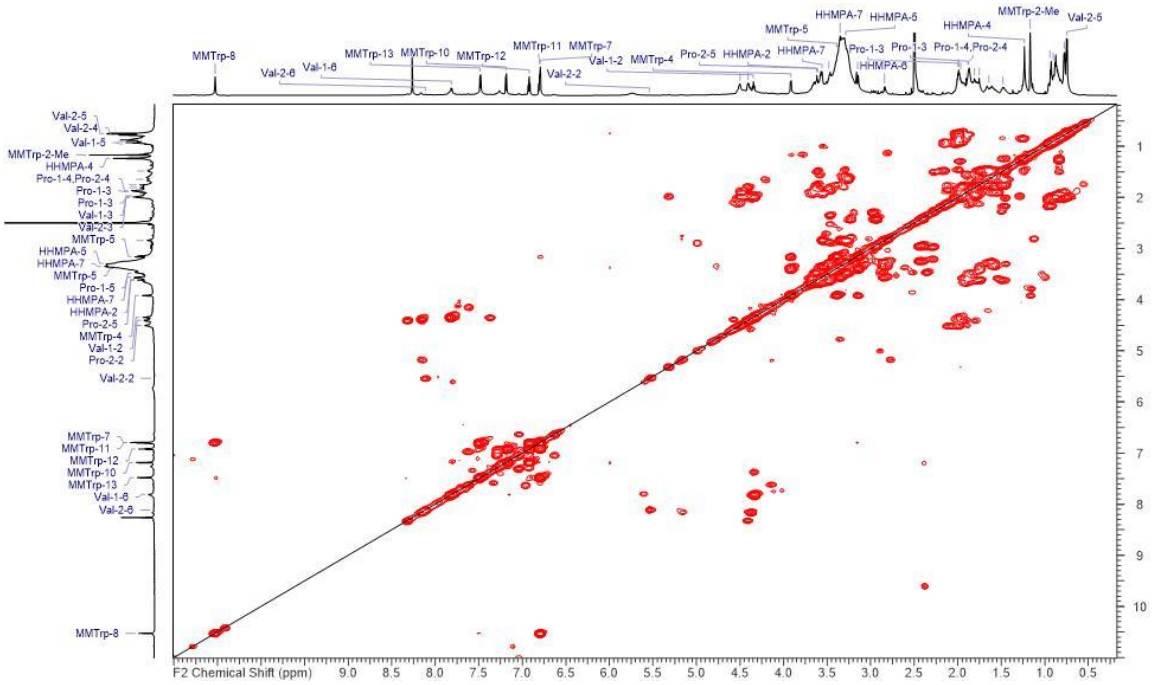


Figure S4. 13. *COSY NMR of cystopipecotide B recorded in DMSO-d_6.*

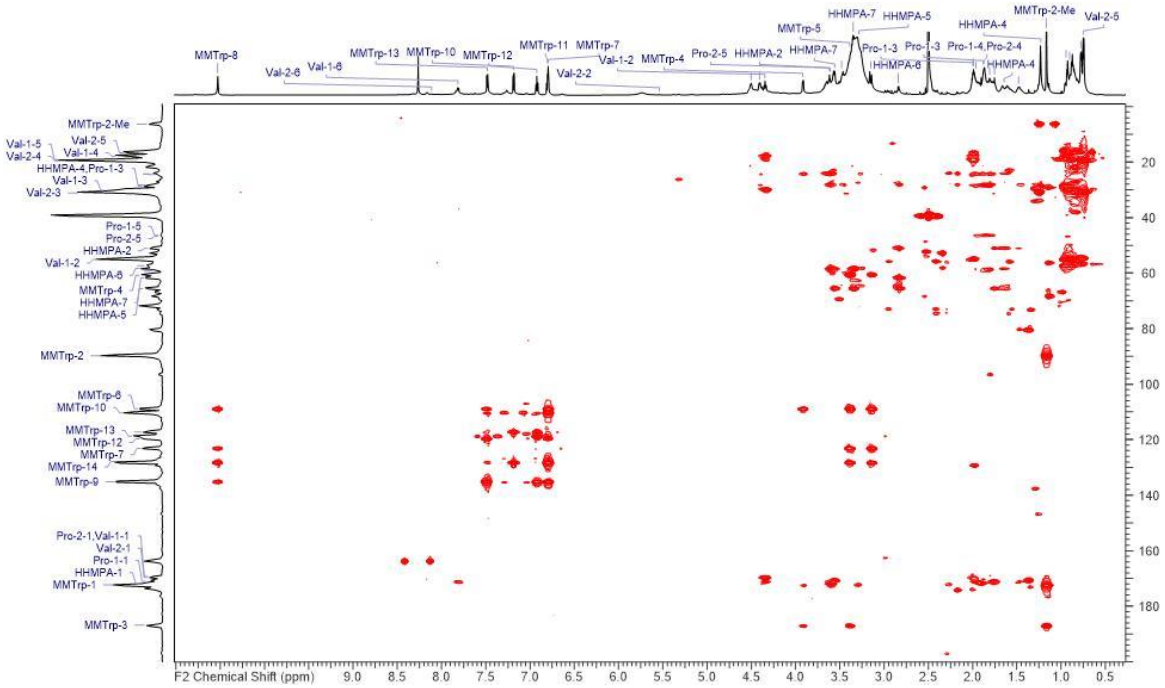


Figure S4. 14. *HMBC NMR of cystopipecotide B recorded in DMSO-d_6.*

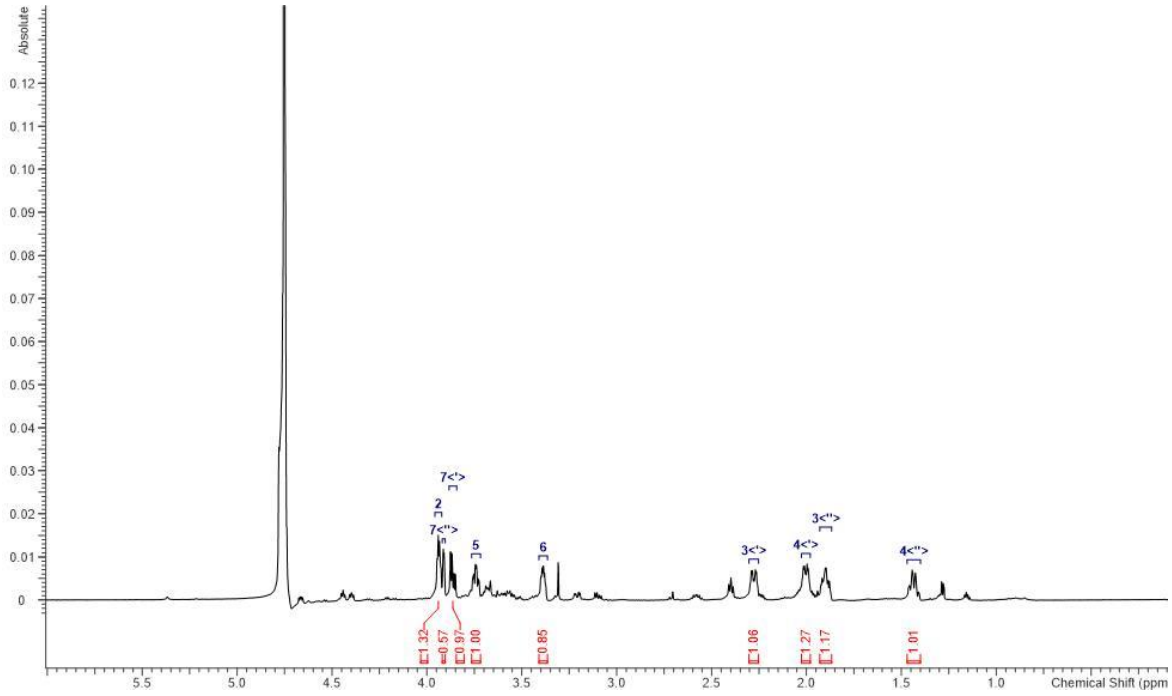


Figure S4. 15. *^{1}H NMR of HHMPA, derived from heterologous expression, recorded in D_2O at 700 MHz.*

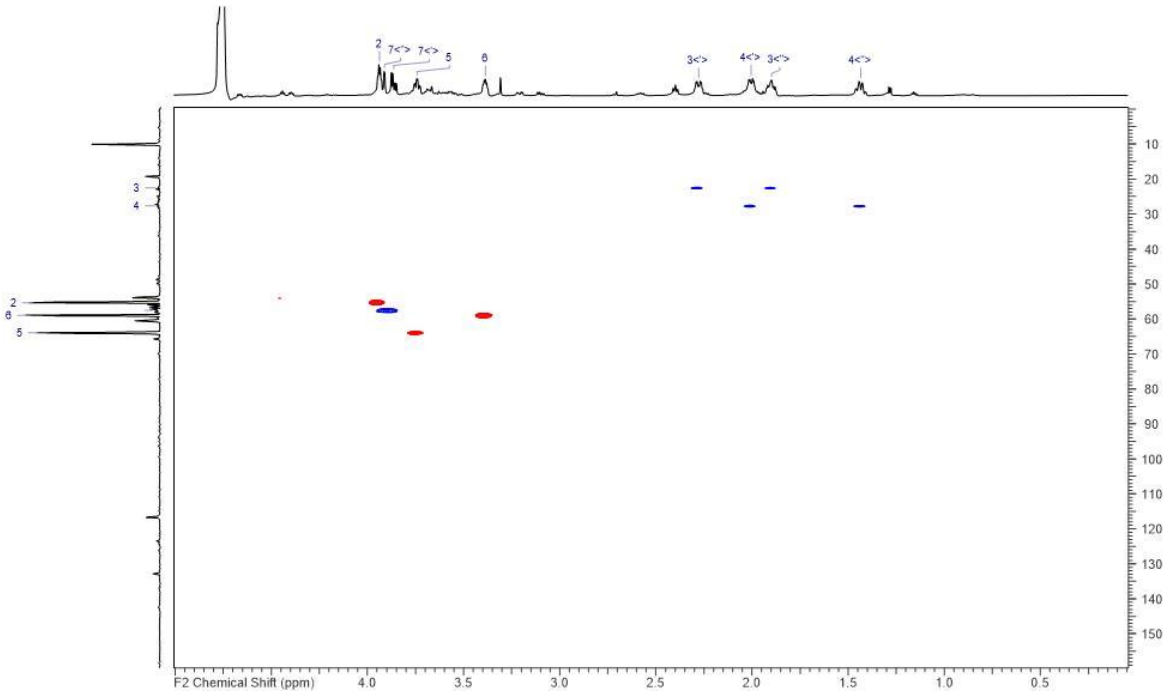


Figure S4. 16. *HSQC NMR of HHMPA, derived from heterologous expression, recorded in D_2O at 175 MHz.*

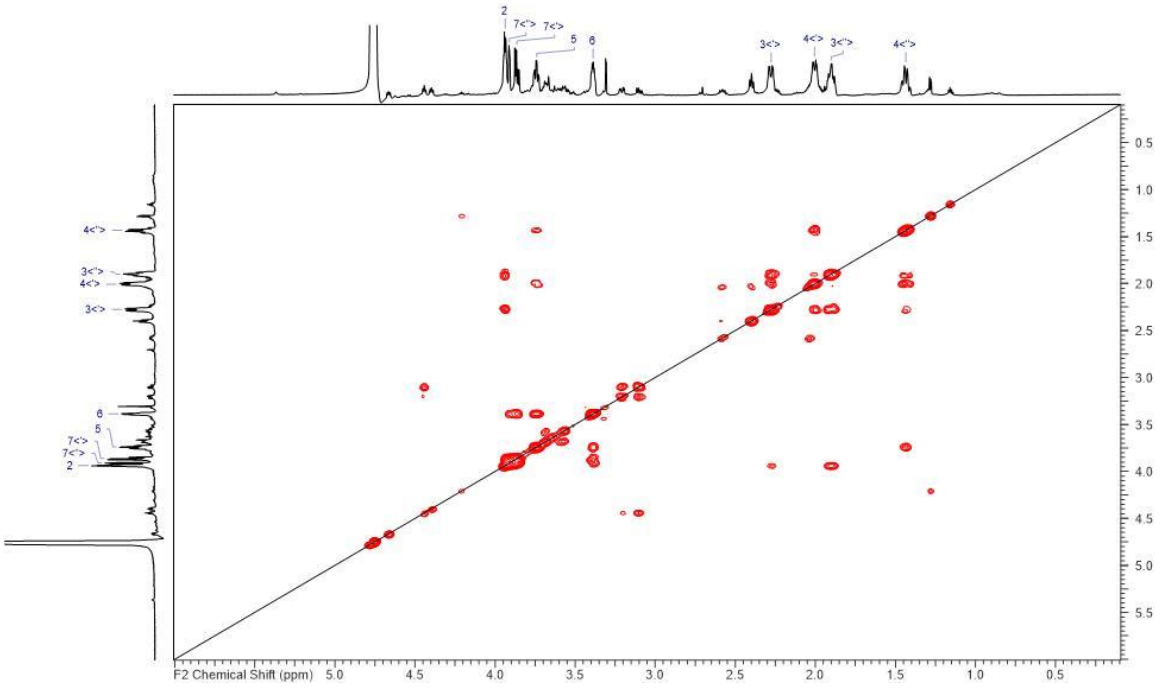


Figure S4. 17. *COSY NMR of HHMPA, derived from heterologous expression, recorded in D_2O.*

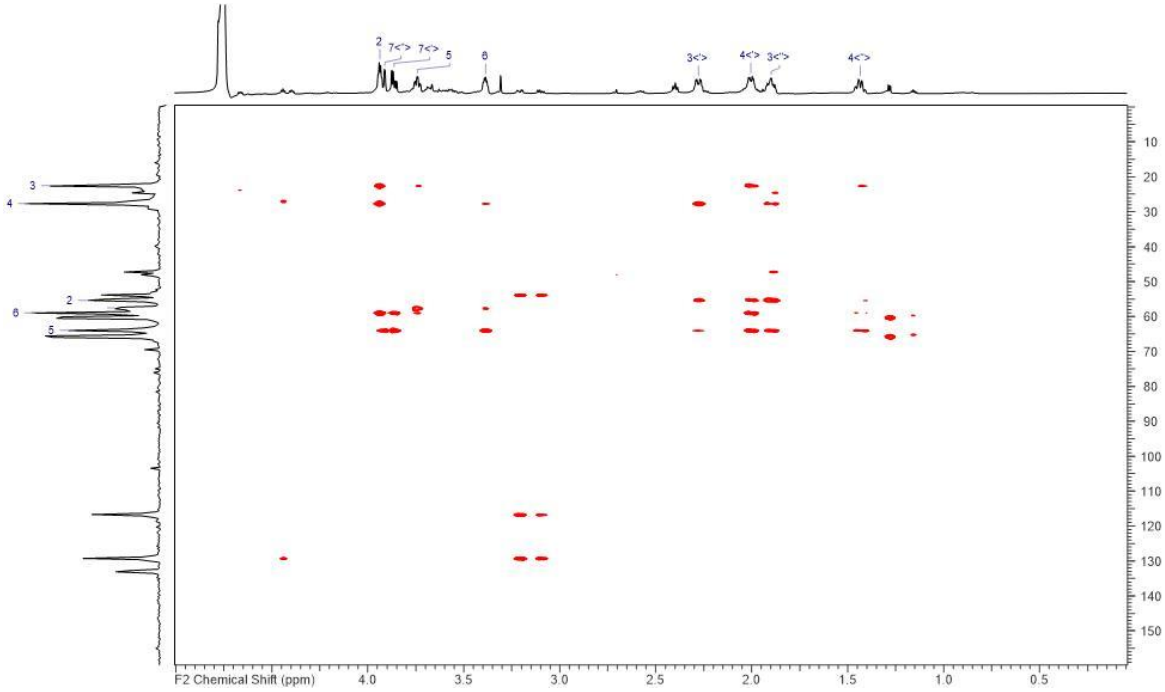


Figure S4. 18. *HMBC NMR of HHMPA, derived from heterologous expression, recorded in D_2O.*

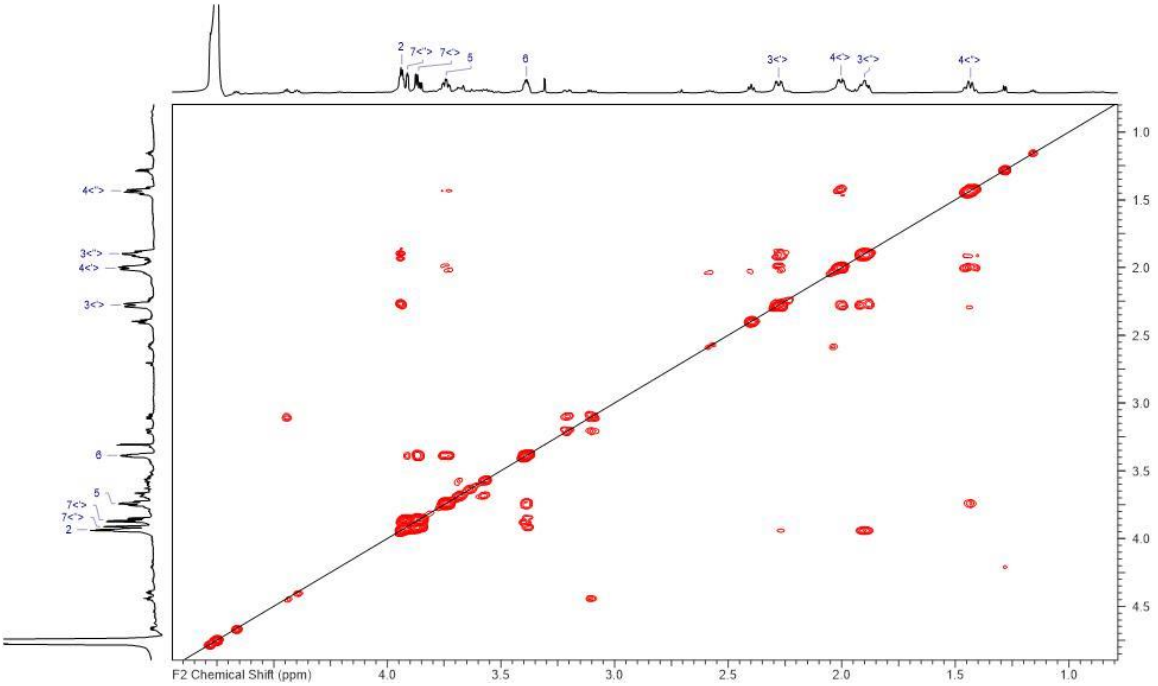


Figure S4. 19. *NOESY NMR of HHMPA, derived from heterologous expression, recorded in D_2O.*

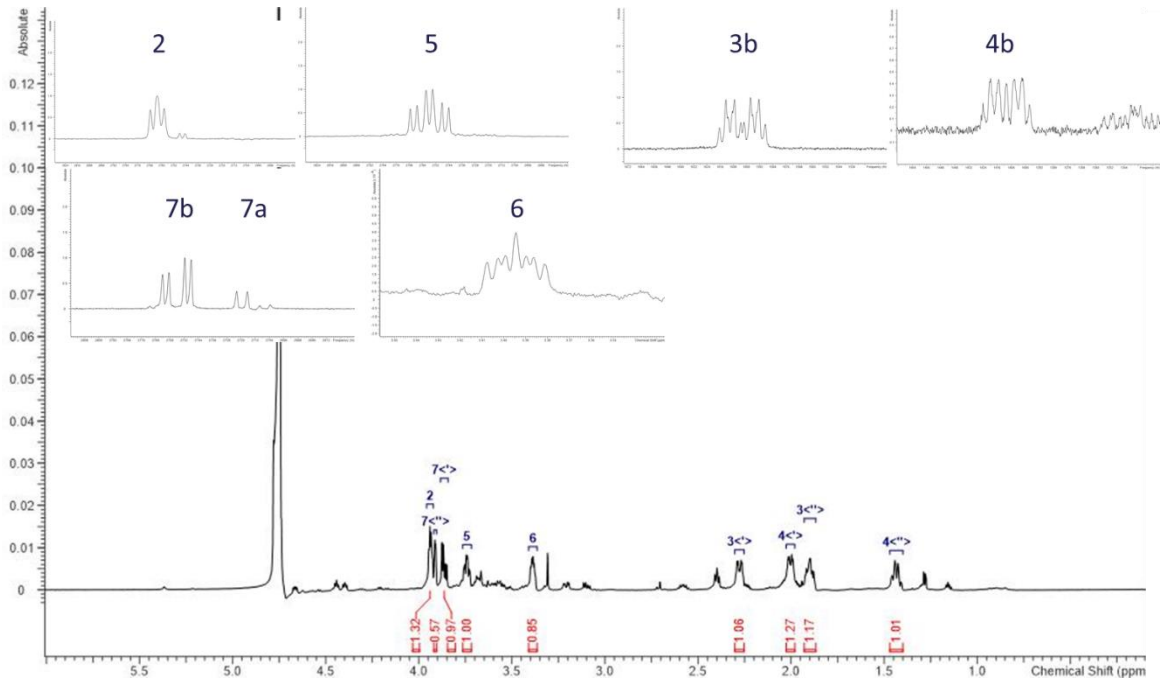


Figure S4. 20. *^1H NMR of HHMPA, derived from heterologous expression, including multiplets derived from 1D selective NOESY.*

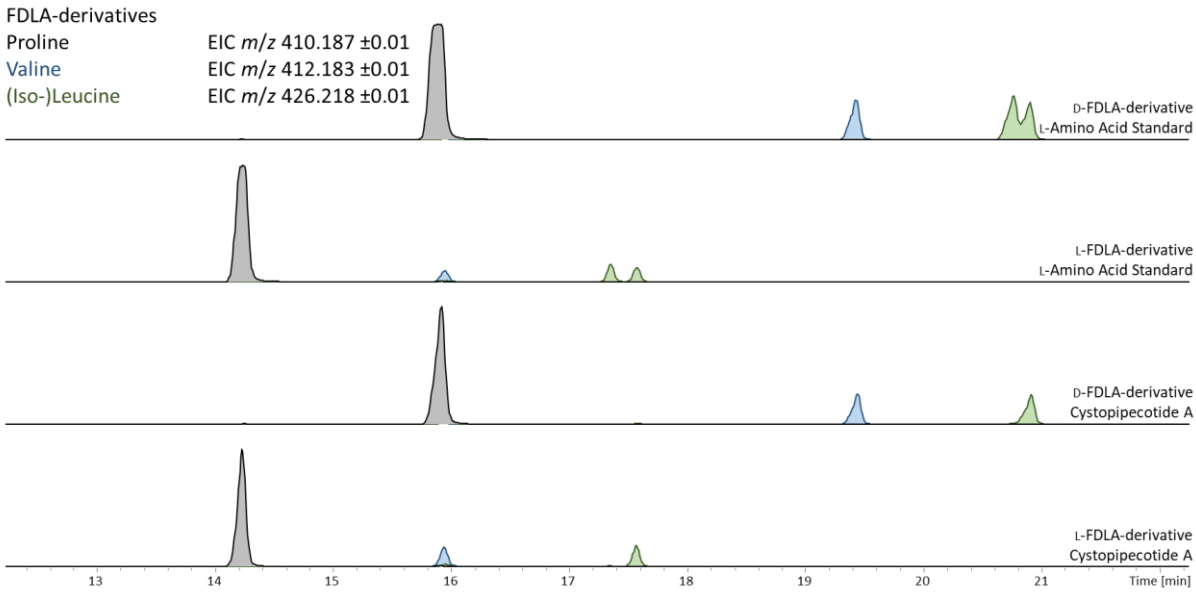


Figure S4. 21. Overlay of the LC-MS chromatograms after Marfey's derivatisation for the characterisation of amino acids using FDLA. *The stereogenic centers of proline (grey, left), valine (blue, middle) and leucine (green, right) are L-configured.*

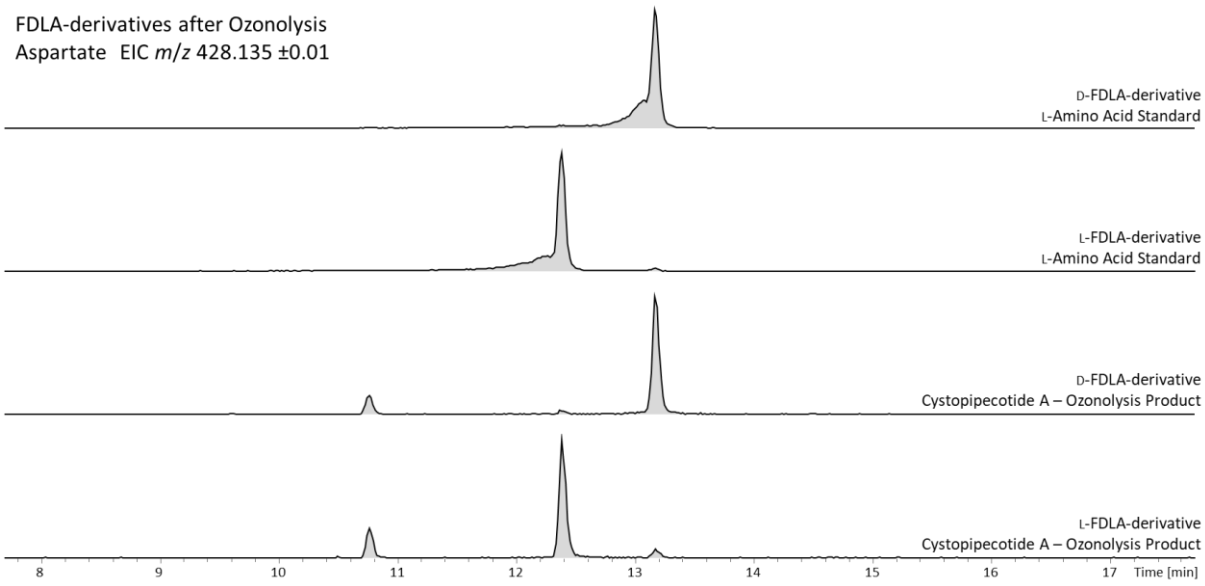


Figure S4. 22. Overlay of the LC-MS chromatograms after ozonolysis and Marfey's derivatisation for the characterisation MMTrp using FDLA. *During the reaction MMTrp is converted to aspartate, which stereogenic center is L-configured.*

S 4.2 Biosynthesis

Table S4. 4. *Oligonucleotide Primers Used in this Study.*

Primer	Oligonucleotide in 5'-3' direction (introduced restriction site)
9101-sc-cpcB-F	**CCAGTCTAGCTATTAATAGGCCTAGG**TCAACCCACCACGATGTCCA
9101-sc-cpcB-R	**GGGCTGGCTTAAAGTCGACTCAATTG**GGTGATGTAAGAGGTGGTGG
9101-sc-cpcG-F	**CCAGTCTAGCTATTAATAGGCCTAGG**CTAAGTCACGAATTGCCTCC
9101-sc-cpcG-R	**GGGCTGGCTTAAAGTCGACTCAATTG**CTAGAACATGGGAACGCCCG
9101-sc-cpcK-F	**CCAGTCTAGCTATTAATAGGCCTAGG**CTAGACGGGCGGAGAGGTG
9101-sc-cpcK -R	**GGGCTGGCTTAAAGTCGACTCAATTG**CCAGTTCCTGTCGGGATCCA
9101-sc-cpcM-F	**CCAGTCTAGCTATTAATAGGCCTAGG**GTAGGGTTTCTTTCACGCCG
9101-sc-cpcM-R	**GGGCTGGCTTAAAGTCGACTCAATTG**GAGAGCTACTTCTTCCACCC
9101-sc-cpcN-F	**CCAGTCTAGCTATTAATAGGCCTAGG**TAACTGGGACTGGCAGGTAG
9101-sc-cpcN-R	**GGGCTGGCTTAAAGTCGACTCAATTG**TCAGCCATCGACGAGCTG
9101-sc-cpcP-F	**CCAGTCTAGCTATTAATAGGCCTAGG**CTAGTGGATGGTGGAGCAGT
9101-sc-cpcP-R	**GGGCTGGCTTAAAGTCGACTCAATTG**CTAGGCGAGGGTGGGCACC
9101-sc-orf6-F	**CCAGTCTAGCTATTAATAGGCCTAGG**TCACGTCATGTCGCCCAC
9101-sc-orf6-R	**GGGCTGGCTTAAAGTCGACTCAATTG**TAGGAGTTGTCCCATCTCCAG
9101-sc- BGC13_B-F	GCGCCTAGGGCGGACGGGAGCATCTAGTT (MfeI)
9101-sc- BGC13_B-R	CCTCAATTGAGCTGTCAATGAGCGCCAGG (XmaJI)
9101-sc- BGC13_K-F	**CCAGTCTAGCTATTAATAGGCCTAGG**ACCTAGGATCAATTCGCGG
9101-sc- BGC13_K-R	**GGGCTGGCTTAAAGTCGACTCAATTG**ATTGACGACGATCCGGTTTC
cpcB-CK-F	CTCCTGTAGCGCGAACATCA
cpcG-CK-F	ACTTCTCCGCCATCTCGTTC
cpcK -CK-F	GAGGTCAACCACTGTGTCGA
cpcM-CK-F	CACTTCTACGTGGCTGCACT
cpcN-CK-F	CGACCTGGGAAAGAACCACA
cpcP-CK-F	AATGATCCGTTGACGCTCGT
orf6-CK-F	AATATCCCTCGCGCTTACGG
BGC13_B-CK-F	CATCGAGGACAGGTTCACGG
BGC13_K-CK-F	GACAAGCCCGGCTATGATCT
CN-check-R	ACCTCTGACACATGCAGCTC
seq-pblueCN-R	ACCTCTGACACATGCAGCTC
9101-orf4 gap-F	AACAGCTCACCAACCCGTC
9101-orf4 gap-R	GTGGAGCAGGTGATTCAGGA

9101-cpcC gap-F	GTGGATGCCAATCTCCTCGA
9101-cpcC gap-R	CTGTCTTCCTGTTTCCCGGA
9101-cpcB gap1-F	GCTGGGGATCCATCTGCTC
9101-cpcB gap1-R	ACTACGTCGCTCCGGAGT
9101-cpcB gap2-F	GAAGAAGTTGTCGTGCGCAC
9101-cpcB gap2-R	CACGAACTCTTTGAGGCGTG
9101-cpcB gap3-F	CCACCACCTCTTCCATCACC
9101-cpcB gap3-R	GCTGGCGTTGATGCTCAATG
9101-cpcL gap-F	CCCCAGACAGAAATCACCGA
9101-cpcL gap-R	CGCATGTGATTCGGGAGTCT
9101-cpcQ gap-F	GGTTCGTGGAGAAGGAGCTG
9101-cpcQ gap-R	GTAGCAGTCCTGACACTGGC
9101-BGC13_B_C gap-F	CGACTTCCTGTCCACCTTCC
9101 BGC13_B_C gap-R	CACGCTTCGACATCACGTTG
9101-BGC13_L gap-F2	TCGCGACGTTCATCGACTAC
9101- BGC13_L gap-R2	TGAAGCCGACCCAATAGCTG
CpcG-sumohistev-F	**AAAACCTGTATTTTCAGGGC**ACGAGCCATATCCTGATTTC
CpcG-sumohistev-R	**TCGAGTGCGGCCGCAAGCTT**TTATGGGACCTCTCCCATGG
CpcH-CDF-F	**TAATAAGGAGATATACCATG**AGCGAGCCCCAGACGGCGGG
CpcH-CDF-R	**CGAGCTCCCAATTGGGATCC**TTAGTTCCTCCTCGTTGTGA
CpcI-histev-F	**AAAACCTGTATTTTCAGGGC**CAGGTGCAAAAGCAGCTGGA
CpcI-histev-R	**TCGAGTGCGGCCGCAAGCTT**TTATTCGCCCCAGTCTTCCT
CpcJ-histev-F	**AAAACCTGTATTTTCAGGGC**AAGAAGATCGATCTGATCTA
CpcJ-histev-R	**TCGAGTGCGGCCGCAAGCTT**TTATGGATGCACCCCGGCG
CpcK-histev-F	**AAAACCTGTATTTTCAGGGC**ACGCGCGTGGCCGTGCTGGG
CpcK-histev-R	**TCGAGTGCGGCCGCAAGCTT**TTAGATGGGATGCAGGCCGC
CpcN-sumohistev-F	**AAAACCTGTATTTTCAGGGC**GAACTGGGACTGGCAGGTAG
CpcN-sumohistev-R	**TCGAGTGCGGCCGCAAGCTT**TTATCCGAGCACACTCGCTC
CpcO-histev-F	**AAAACCTGTATTTTCAGGGC**AGCGCGTCCTCCATCGTGGT
CpcO-histev-R	**TCGAGTGCGGCCGCAAGCTT**TTACGTGCCCGCATGCGTCG
CpcP-sumohistev-F	**AAAACCTGTATTTTCAGGGC**CGGGCACGTGAGAAGACGAT
CpcP-sumohistev-R	**TCGAGTGCGGCCGCAAGCTT**TTACCCTCCCGACTTCGCGG
T7_cpcG-L_F	**AAAACCTGTATTTTCAGGGC**ACGAGCCATATCCTGATTTC
T7_cpcG-L_R	TGACGTGCTCGGTGATTTCT

T7_cpcL-P_F	CCCCAGACAGAAATCACCGA
T7_cpcL-P_R	**TCGAGTGCGGCCGCAAGCTT**TCACCCTCCCGACTTCGCGG
del_cpcL_cat-F	**AGCATCCCGCGGTCTGGAATATCCAAACAGGGAGTAGAC**CATACCAACAT GGTCAAATA
del_cpcL_cat-R	**GGTTGGATGGAGACACGCGGGCCCCGCTCGGATTGGACG**GAGCTCATCG CTAATAACTT
del_cpcM_cat-F	**ACCGGCTGGTGCCGCAACGGCAGCACGATCATCGGAAAC**CATACCAACAT GGTCAAATA
del_cpcM_cat-R	**CAGGAGCGTGGGCCCGCTGCGGAAGCGATCCGGATTTCC**GAGCTCATCG CTAATAACTT
del_cpcN_cat-F	**ACCCGGGAGGATGCCCGTGGAACTGGGACTGGCAGGTAG**CATACCAACA TGGTCAAATA
del_cpcN_cat-R	**CCGGCGAAGCGGGATGTGGGCCGCCATGGCGTGGAGCCC**GAGCTCATCG CTAATAACTT
del-cpcL_cat-CK-up-F	GATCGCTCGTCCCATCACC
del_cpcL_cat-CK-down-R	GAAGCCAACCCAGTAGACGG
del-cpcM_cat-CK-up-F	CTCGCGACGTTCATCGACTA
del_cpcM_cat-CK-down-R	CAATGCCAAACTTCGTCGCC
del-cpcN_cat-CK-up-F	GAAGCAATACGTCGTCCCGA
del_cpcN_cat-CK-down-R	CGATGCGTTTGTTCACCTCG
cat-CK-up-R	GCTCACCGTCTTTCATTGCC
cat-CK-down-F	CGGTGAGCTGGTGATATGGG

The bold letters indicate the overlap sequence to the vector.

Table S4. 5. *Plasmids Used in this Study.*

Plasmids	Description/derivation	Reference
pbluekan	*ori (pMB1)*, *bla*, *lacZ*, *neo*	C. Volz, unpublished
pbluekan-cpcB-HR	pbluekan derivative with *cpcB* homologous region	This study
pbluekan-cpcG-HR	pbluekan derivative with *cpcG* homologous region	This study
pbluekan-cpcK-HR	pbluekan derivative with *cpcK* homologous region	This study
pbluekan-cpcM-HR	pbluekan derivative with *cpcM* homologous region	This study
pbluekan-cpcN-HR	pbluekan derivative with *cpcN* homologous region	This study
pbluekan-cpcP-HR	pbluekan derivative with *cpcP* homologous region	This study
pbluekan-orf6-HR	pbluekan derivative with *orf6* homologous region	This study
pbluekan-BGC13_B-HR	pbluekan derivative with *BGC13_B* homologous region	This study
pbluekan-BGC13_K-HR	pbluekan derivative with *BGC13_K* homologous region	This study
pHisTEV	*ori (pBR322)*, *lac* operator, *neo*, T7 promoter, ribosome binding site, N-HisTag	Novagen
pHisSUMOTEV	*ori (pBR322)*, *lac* operator, *neo*, T7 promoter, ribosome binding site, N-HisTag, SUMO-Tag	Dr. David Owen, unpublished
pCDF-1b	*ori (CloDF13)*, *lac* operator, *spect(variant)*, T7 promoter, ribosome binding site, HisTag	Novagen
pHisTEV-cpcG-P	pHisTEV derivative with CpcG-CpcP coding sequence	This study
pHisTEV-cpcG-P_del-cpcL	pHisTEV-cpcG-P derivative with *cpcL* deleted	This study
pHisTEV-cpcG-P_del-cpcM	pHisTEV-cpcG-P derivative with *cpcM* deleted	This study
pHisTEV-cpcG-P_del-cpcN	pHisTEV-cpcG-P derivative with *cpcN* deleted	This study
pHisSUMOTEV-CpcG	pHisSUMOTEV derivative with CpcG coding sequence	This study
pCDF-CpcH	pCDF-1b derivative with CpcH coding sequence	This study
pHisTEV-CpcI	pHisTEV derivative with CpcI coding sequence	This study
pHisTEV-CpcJ	pHisTEV derivative with CpcJ coding sequence	This study
pHisTEV-CpcK	pHisTEV derivative with CpcK coding sequence	This study
pHisSUMOTEV-CpcN	pHisSUMOTEV derivative with CpcN coding sequence	This study
pHisTEV-CpcO	pHisTEV derivative with CpcO coding sequence	This study
pHisSUMOTEV-CpcP	pHisSUMOTEV derivative with CpcP coding sequence	This study

Table S4. 6. *Putative proteins encoded in the cystopipecotide biosynthetic gene cluster.*

Protein	Size aa*	Proposed function of the similar protein	Sequence similarity to source	Similarity/ identity aa* level	Accession number of the closest homologue
Orf1	818	penicillin acylase family protein	*Vitiosangium sp*. GDMCC 1.1324	94 / 89 %	WP_108077922.1
Orf2	225	sulfotransferase family protein	*Chloroflexi bacterium* AL-N1	74 / 57 %	NOK60153.1
Orf3	267	3'(2'),5'-bisphosphate nucleotidase CysQ	*Vitiosangium sp*. GDMCC 1.1324	94 / 87 %	WP_108077923.1
Orf4	418	GTP-binding protein	*Pyxidicoccus fallax*	96 / 93 %	WP_169347789.1
Orf5	188	adenylyl-sulfate kinase	*Pyxidicoccus fallax*	94 / 90 %	WP_169347795.1
Orf6	302	sulfate adenylyltransferase subunit CysD	*Vitiosangium sp*. GDMCC 1.1324	98 / 97 %	WP_108075484.1
CpcF	88	acyl carrier protein	*Dyella sp*. L4-6	67 / 44 %	WP_114847001.1
CpcE	417	HlyD family secretion protein	*Vitiosangium sp*. GDMCC 1.1324	94 / 89 %	WP_108077906.1
CpcD	546	DHA2 family efflux MFS transporter permease subunit	*Vitiosangium sp*. GDMCC 1.1324	93 / 86 %	WP_108077907.1
CpcC	1223	PKS (KS_{17-426}, $AT_{523-818}$, $ACP_{913-977}$, $TE_{1002-1217}$)	*Vitiosangium sp*. GDMCC 1.1324	88 / 81 %	WP_108077908.1
CpcB	5360	NRPS (C_{43-479}, $A_{485-1011}$, $PCP_{1022-1086}$, $C_{1107-1541}$, $A_{1545-2069}$, $PCP_{2081-2145}$, $C_{2166-2596}$, $A_{2602-3128}$, $PCP_{3140-3204}$, $C_{3225-3660}$, $A_{3664-4188}$, $PCP_{4198-4262}$, $C_{4283-4719}$, $A_{4723-5247}$, $PCP_{5258-5322}$)	*Archangium violaceum*	79 / 70 %	WP_052519231.1
CpcA	562	AMP-binding protein	*Vitiosangium sp*. GDMCC 1.1324	96 / 92 %	WP_108077857.1

CpcG	316	transketolase	*Pyxidicoccus fallax*	88 / 81 %	WP_169346720.1
CpcH	347	transketolase	*Vitiosangium sp*. GDMCC 1.1324	95 / 92 %	WP_108077855.1
CpcI	69	lysine biosynthesis protein LysW	*Pyxidicoccus fallax*	95 / 87 %	WP_169346718.1
CpcJ	281	RimK family alpha-L-glutamate ligase	*Vitiosangium sp*. GDMCC 1.1324	96 / 94 %	WP_108077854.1
CpcK	347	N-acetyl-gamma-glutamyl-phosphate reductase	*Vitiosangium sp*. GDMCC 1.1324	95 / 90 %	WP_108077853.1
CpcL	287	phage major capsid protein	*Vitiosangium sp*. GDMCC 1.1324	99 / 97 %	WP_108077852.1
CpcM	310	sulfotransferase	*Pyxidicoccus fallax*	91 / 84 %	WP_169346714.1
CpcN	258	short-chain dehydrogenase/reductase (SDR) family oxidoreductase	*Vitiosangium sp*. GDMCC 1.1324	95 / 90 %	WP_108077850.1
CpcO	284	[LysW]-aminoadipate kinase	*Vitiosangium sp*. GDMCC 1.1324	95 / 90 %	WP_108077849.1
CpcP	356	M20/M25/M40 family metallo-hydrolase	*Vitiosangium sp*. GDMCC 1.1324	95 / 91 %	WP_108077848.1
Orf7	91	acyl-CoA dehydrogenase family protein	*Vitiosangium sp*. GDMCC 1.1324	92 / 91 %	WP_108077859.1
Orf8	148	hypothetical protein MEBOL_002563	*Melittangium boletus* DSM 14713	77 / 67 %	ATB29114.1
Orf9	190	hypothetical protein Q664_04790	*Archangium violaceum* Cb vi76	81 / 73 %	KFA94092.1
Orf10	752	PAS domain S-box protein	*Cystobacter fuscus*	78 / 66 %	WP_002623053.1

*aa: amino acids

Table S4. 7. *Substrate specificity analysis of A domains from the cystopipecotide NRPS from Cystobacter sp. SBCb004[a].*

A domain	Stachelhaus[1] sequence	Neareast Stachelhaus code	NRPSPredictor2[2]	Consensus
A_CpcA	ssyyaGlAVK	pip[b]	hydrophobic-aliphatic[c]	-
A_CpcB_module1	DVQYIAqVaK	pro	pro	pro
A_CpcB_module2	DAFFLGgTFK	ile	hydrophobic-aliphatic	-
A_CpcB_module3	DVQYIAqVaK	pro	pro	pro
A_CpcB_module4	DALfvGGTFK	val	val	val
A_CpcB_module5	DALvLGLtnK	hpg[d]	ile	-

[a] Predicted substrate specificities were retrieved from outcomes of the antiSMASH[3] gene cluster analysis. [b] pip: pipecolic acid. [c] hydrophobic - aliphatic: gly, ala, val, leu, ile, abu (2-amino-butyric acid), iva (isovaline). [d] hpg: 4-hydoxy-phenyl-glycine. -: no consensus.

Table S4. 8. *Substrate specificity analysis of AT domain from the cystopipecotide PKS from C*ystobacter *sp. SBCb004[a].*

AT domain	ATSignature (Top 3 matches, score)	minova (Top 3 predication, score)
AT_CpcC	Malonyl-CoA (58.3 %), Isobutyryl-CoA (54.2 %), Methylmalonyl-CoA (54.2 %)	Malonyl-CoA (115.7), Methoxymalonyl-CoA (64.3), Methylmalonyl-CoA (62.9)

[a] Predicted substrate specificities were retrieved from outcomes of the antiSMASH[3] gene cluster analysis.

Table S4. 9. *Comparison of BGC13 and cystopipecotide BGC (CPC)* in silico*.*

Protein	Congener in CPC	Putative function	Identity/similarity (aa* level)
BGC13_A	CpcA	AMP-binding protein	88 %/ 94 %
BGC13_B	CpcB	NRPS (C_{-432}, $A_{438-966}$, $PCP_{978-1042}$, $C_{1067-1497}$, $A_{1501-2025}$, $PCP_{2037-2100}$, $C_{2121-2551}$, $A_{2557-3085}$, $PCP_{3099-3163}$, $TE_{3183-3434}$)	78 %/ 85 %
BGC13_D	CpcD	DHA2 family efflux MFS transporter permease subunit	82 %/ 91 %
BGC13_E	CpcE	HlyD family secretion protein	82 %/ 90 %

BGC13_F	CpcF	acyl carrier protein	86 %/ 93 %
BGC13_G	CpcG	transketolase	86 %/ 91 %
BGC13_H	CpcH	transketolase	86 %/ 90 %
BGC13_I	CpcI	lysine biosynthesis protein LysW	89 %/ 95 %
BGC13_J	CpcJ	RimK family alpha-L-glutamate ligase	93 %/ 96 %
BGC13_K	CpcK	N-acetyl-gamma-glutamyl-phosphate reductase	90 %/ 95 %
BGC13_L	CpcL	phage major capsid protein	97 %/ 98 %
BGC13_M	CpcM	sulfotransferase	83 %/ 90 %
BGC13_N	CpcN	short-chain dehydrogenase/reductase (SDR) family oxidoreductase	88 %/ 94 %
BGC13_O	CpcO	[LysW]-aminoadipate kinase	86 %/ 93 %
BGC13_P	CpcP	M20/M25/M40 family metallo-hydrolase	90 %/ 95 %
BGC13_ORF7	Orf7	acyl-CoA dehydrogenase family protein	95 %/ 97 %
BGC13_ORF6	ORF6	sulfate adenylyltransferase subunit CysD	96 %/ 98 %
BGC13_ORF5	ORF5	adenylyl-sulfate kinase	89 %/ 95 %
BGC13_ORF4	ORF4	GTP-binding protein	91 %/ 94 %

*aa: amino acids.

Table S4. 10. *Distribution and diversity of putative HHMPA biosynthetic operon containing myxobacteria.*

Strain	Organisation of NRPS-PKS	Position of PCP_L domain	Type
Myxococcus fulvus MCy8288	6 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Myxococcus sp. MCy10608	6 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Corallococcus sp. MCy10984	5 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1

Corallococcus sp. MCy12228	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Cystobacterineae strain MCy10653	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Cystobacterineae strain MCy12733	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Cystobacterineae strain MCy12716	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Cystobacterineae strain MCy10649	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Cystobacterineae strain MCy10585	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Coallococcus coralloides MCy10705	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Corallococcus coralloides MCy12446	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Corallococcus sp. MCy11561	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Corallococcus sp. MCy8332	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Myxococcaceae strain MCy12473	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Corallococcus sp. MCy9049	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Myxococcus sp. MCy11578	3 NRPS modules	Adjacent to A_L domain	B1
Cystobacterineae strain MCy10644	3 NRPS modules	Adjacent to A_L domain	B1
Cystobacterineae strain MCy9003	3 NRPS modules	Adjacent to A_L domain	B1
Pyxidicoccus sp. MCy8408	1 NRPS modules	Adjacent to A_L domain	B1

Myxococcus fulvus MCy9270	incomplete NRPS sequence	Adjacent to A_L domain	unknown
Myxococcus fulvus MCy9280	incomplete NRPS sequence	Adjacent to A_L domain	unknown
Cystobacter sp. SBCb004	5 NRPS modules, 1 PKS module	downstream	A2
Vitiosangium cumulatum MCy10943	5 NRPS modules, 1 PKS module	downstream	A2
Archangium sp. MCy8383	4 NRPS modules	downstream	B2
Cystobacter sp. SBCb004	3 NRPS modules	downstream	B2
Vitiosangium cumulatum MCy10943	3 NRPS modules	downstream	B2
Cystobacterineae strain MCy10597	3 NRPS modules	downstream	B2
Corallococcus coralloides MCy6431	3 NRPS modules	downstream	B2
Cystobacterineae strain MCy5730	incomplete NRPS sequence	downstream	unknown
Pyxidicoccus fallax MCy8396	incomplete NRPS sequence	downstream	unknown
Cystobacterineae strain MCy8401	incomplete NRPS sequence	downstream	unknown
Cystobacter armeniaca MCy5465	incomplete NRPS sequence	downstream	unknown
Cystobacterineae strain MCy8369	incomplete NRPS sequence	unknown	unknown
Archangium minus MCy1093	incomplete NRPS sequence	unknown	unknown

A_L: loading A domain. PCP_L: loading PCP domain.

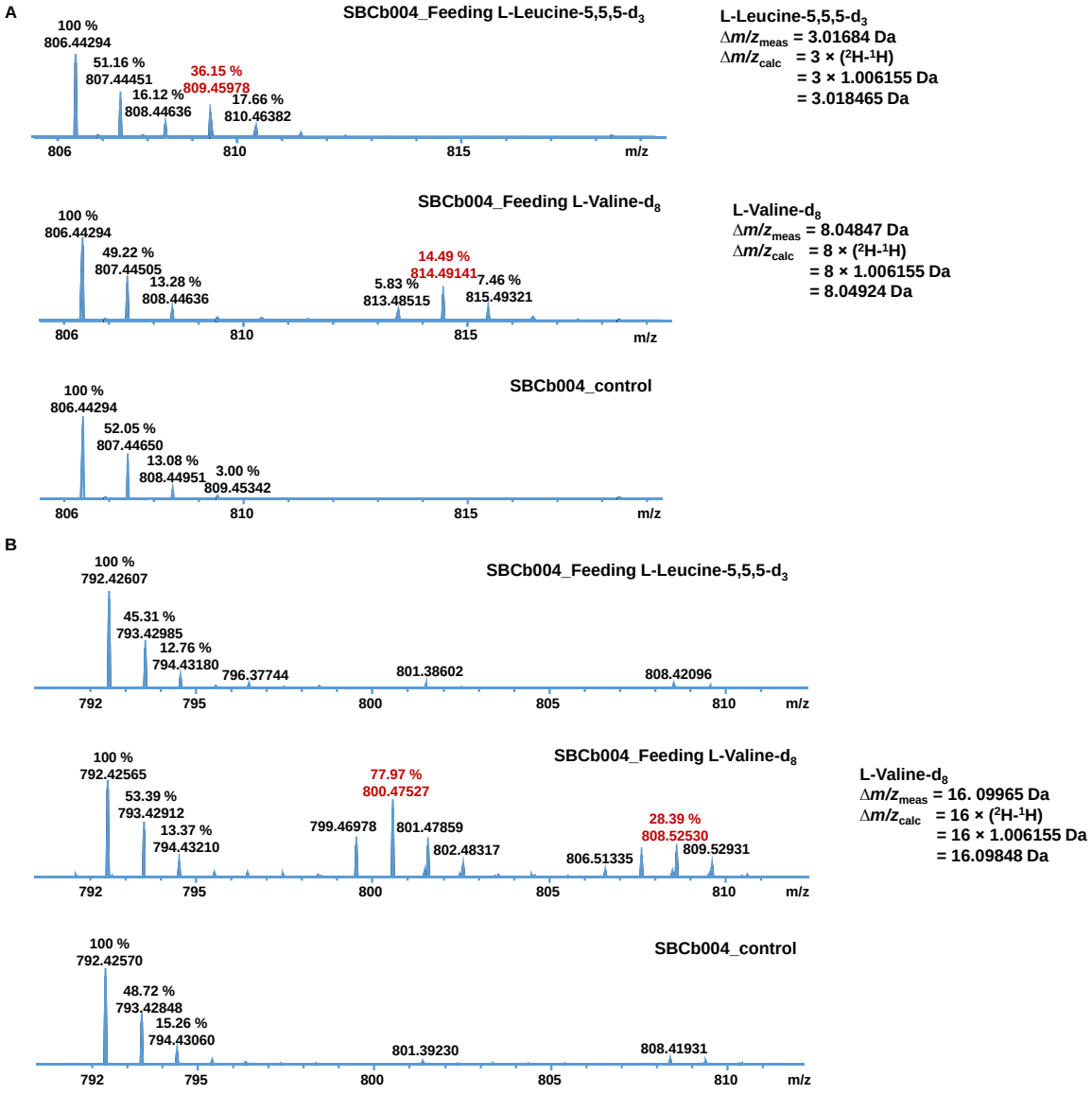


Figure S4. 23. Feeding studies with L-Leucine-5,5,5-d$_3$ and L-Valine-d$_8$. *(A) Mass spectra of cystopipecotide A (806 m/z) with L-Leucine-5,5,5-d$_3$ and L-Valine-d$_8$ fed to the culture. (B) Mass spectra of cystopipecotide B (792 m/z) with L-Leucine-5,5,5-d$_3$ and L-Valine-d$_8$ fed to the culture.*

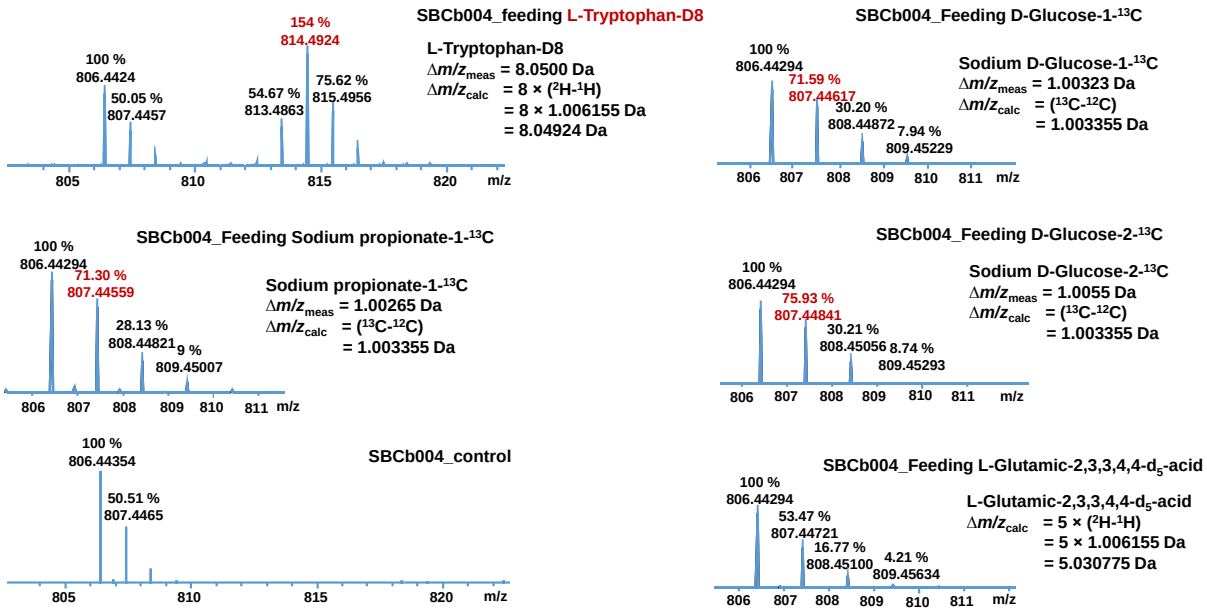


Figure S4. 24. Mass spectra of cystopipecotide A (806 m/z) with L-Tryptophan-d8, D-Glucose-1-13C, D-Glucose-2-13C, Sodium propionate-1-13C and L-Glutamic-2,3,3,4,4-d5-acid fed to the culture.

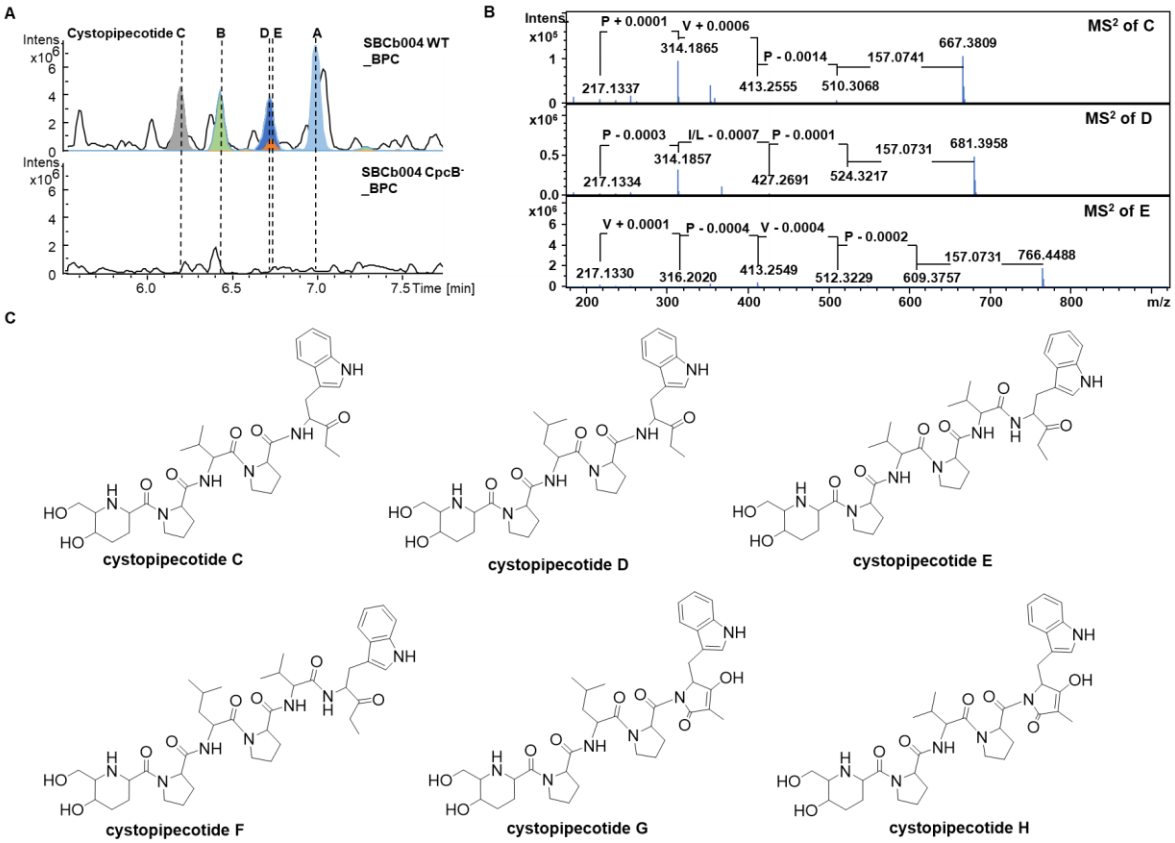


Figure S4. 25. Gene inactivation of cpcB and discovery of cystopipecotide derivatives. *(A) HPLC-MS analysis of cystopipecotide production in* Cystobacter *sp. SBCb004 wild type strain (WT) and the mutant SBCb004 cpcB*$^{-}$*. The extracted ion chromatograms of cystopipecotides A-E are filled with different colours. (B) MS*^{2} *spectra of cystopipecotides C-E. (C) Predicted structures of cystopipecotides C-H.*

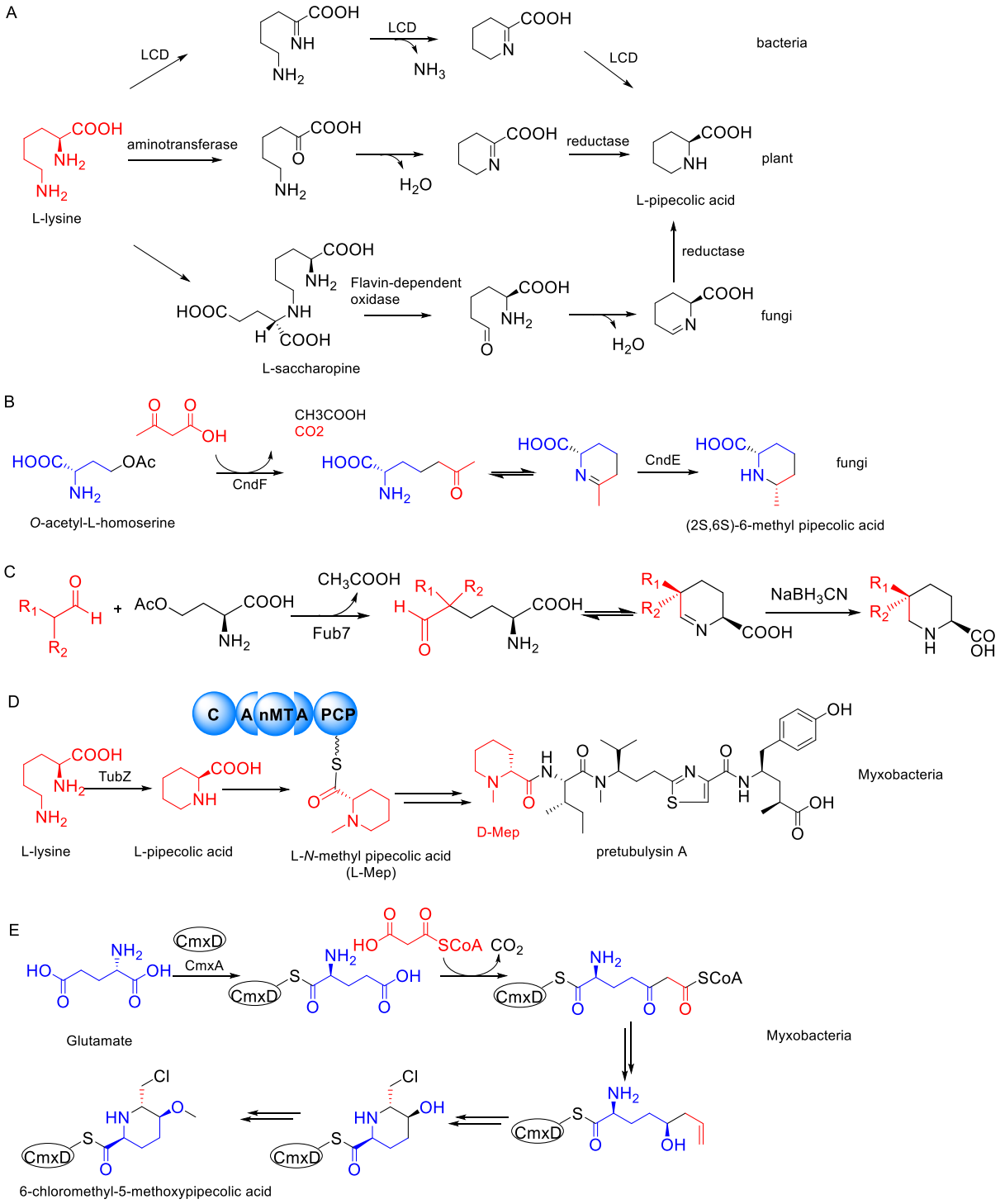


Figure S4. 26. Overview of biosynthetic and chemoenzymatic synthetic routes towards pipecolic acid and its derivatives. *(A) Biosynthesis of L-pipecolic acid from L-lysine. LCD: lysine-cyclodeaminase. (B) Biosynthesis of γ-substituted pipecolate from* O*-acetyl-L-homoserine and acetoacetate. (C) Chemoenzymatic synthesis of 5-substituted L-pipecolic acid. R_1, R_2: alkyl. (D) Biosynthesis of D-*N*-methyl pipecolic acid (D-Mep) in tubulysin. (E) Putative biosynthetic pathway of 6-chloromethyl-5-methoxypipecolic acid (CMPA) from L-glutamic acid.*

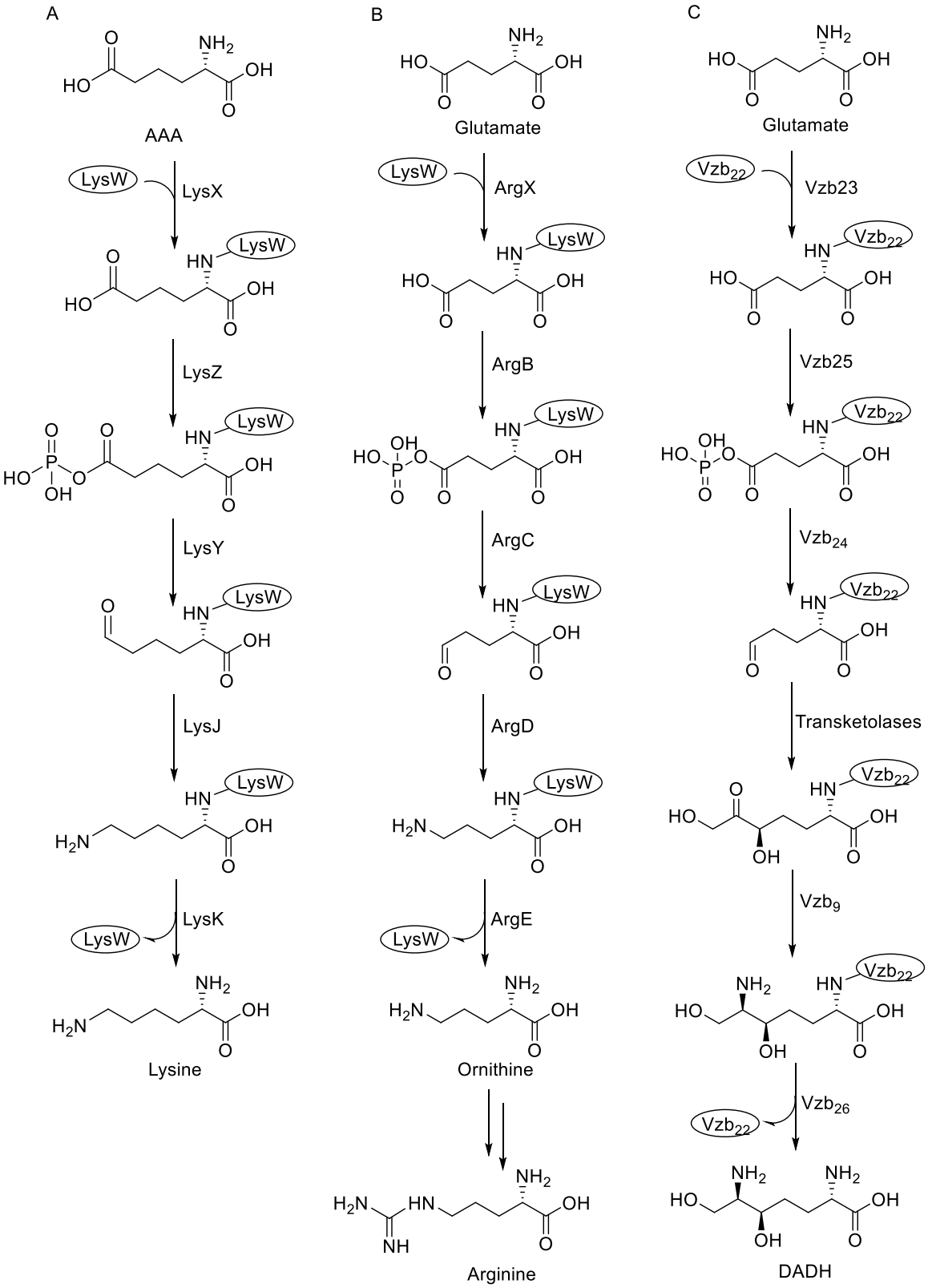


Figure S4. 27. Comparison of LysW mediated biosyntheses. *(A) Lysine biosynthesis via α-aminoadipate (AAA) pathway in* Thermus thermophilus*. (B) Arginine biosynthesis mediated by LysW in* Sulfolobus acidocaldarius*. (C) (2S,6R)-diamino-(5R,7)-dihydroxy-heptanoic acid (DADH) biosynthesis catalysed by AAA pathway enzymes and transketolases in* Streptomyces *sp. SANK 60404.*

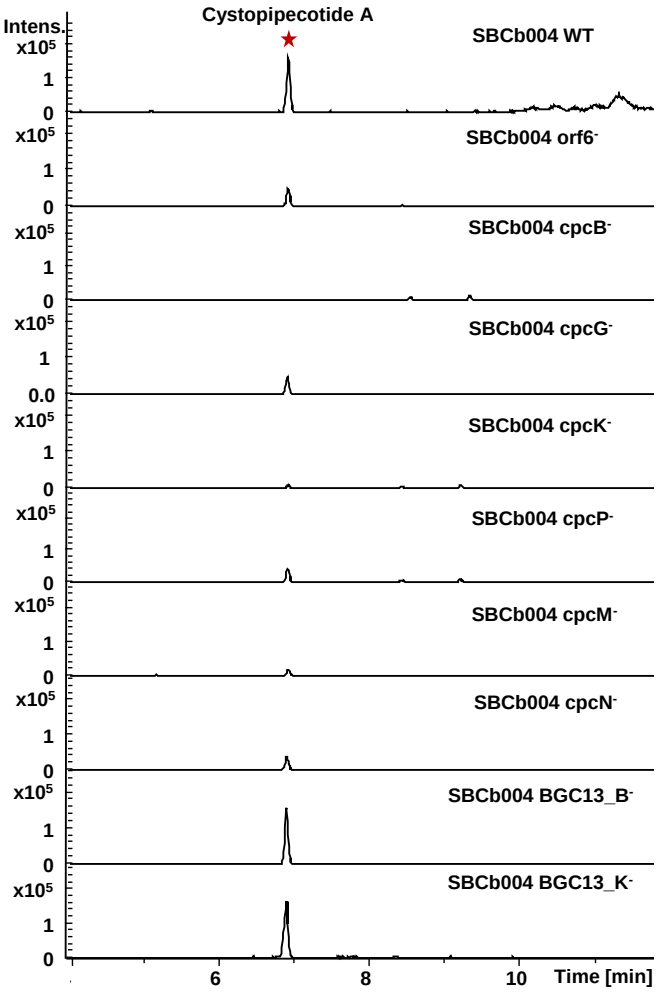


Figure S4. 28. The extracted ion chromatograms of cystopipecotide A *in the SBCb004 wild type strain (WT) and the mutants SBCb004 orf6*$^{-}$*, SBCb004 cpcB*$^{-}$*, SBCb004 cpcG*$^{-}$*, SBCb004 cpcK*$^{-}$*, SBCb004 cpcP*$^{-}$*, SBCb004 cpcM*$^{-}$*, SBCb004 cpcN*$^{-}$*, SBCb004 BGC13_B*$^{-}$ *and SBCb004 BGC13_K*$^{-}$*.*

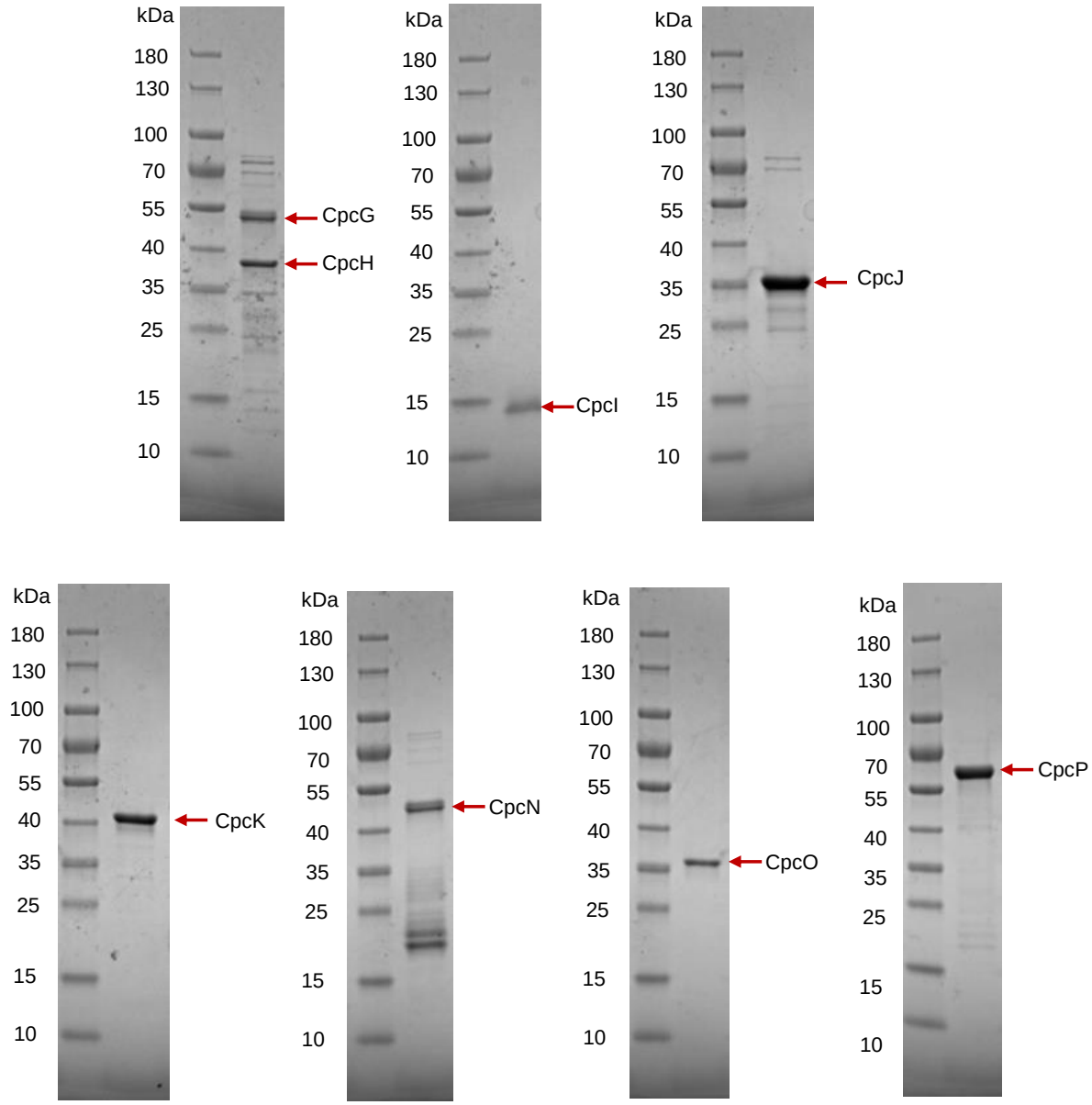


Figure S4. 29. SDS-PAGE analysis *of CpcG/H, CpcI, CpcJ, CpcK, CpcN, CpcO and CpcP.*

S 4.3 Bioactivity Data

Table S4. 11. *Minimum-Inhibitory Concentration (MIC, µg/mL) of Cystopipecotides A and B.*

Test organisms	Cystopipecotide A	Cystopipecotide B
Staphylococcus aureus Newman	> 64	> 64
Mycobacterium smegmatis mc2-155	> 64	> 64
Bacillus subtilis DSM10	> 64	> 64
Citrobacter freundii DSM30039	> 64	> 64
Acinetobacter baumannii DSM30007	> 64	> 64
Candida albicans DSM1665	> 64	> 64
Escherichia coli Δ*acrB* JW0451-2	> 64	> 64
Escherichia coli DSM1116	> 64	> 64
Cryptococcus neoformans DSM11959	> 64	> 64
Pseudomonas aeruginosa PA14 (DSM19882)	> 64	> 64
Pichia anomala DSM6766	> 64	> 64

Table S4. 12. *Half-Inhibitory Concentration (IC_{50}, µg/mL) of Cystopipecotides A and B against Cancer Cell Lines.*

Test cancer cell line	Cystopipecotide A	Cystopipecotide B
HepG2	> 37	2.58
HCT-116	36.11	10.53
KB3.1	18.44	5.36

S 4.4 References

1. Fu, J. *et al.* Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* **30,** 440–446; 10.1038/nbt.2183 (2012).
2. Wright, G. D. & Thompson, P. R. Aminoglycoside phosphotransferases: proteins, structure, and mechanism. *Frontiers in bioscience: a journal and virtual library* **4,** D9-21 (1999).
3. Deuschle, U., Kammerer, W., Gentz, R. & Bujard, H. Promoters of Escherichia coli: a hierarchy of in vivo strength indicates alternate structures. *EMBO J.* **5,** 2987–2994 (1986).
4. Chai, Y. *et al.* Discovery of 23 natural tubulysins from *Angiococcus disciformis* An d48 and *Cystobacter* SBCb004. *Chem. Biol.* **17,** 296–309; 10.1016/j.chembiol.2010.01.016 (2010).
5. Hoffmann, T., Müller, S., Nadmid, S., Garcia, R. & Müller, R. Microsclerodermins from terrestrial myxobacteria: an intriguing biosynthesis likely connected to a sponge symbiont. *Journal of the American Chemical Society* **135,** 16904–16911; 10.1021/ja4054509 (2013).
6. Vollbrecht, L. *et al.* Argyrins, immunosuppressive cyclic peptides from myxobacteria. II. Structure elucidation and stereochemistry. *J. Antibiot.* **55,** 715–721; 10.7164/antibiotics.55.715 (2002).
7. Marfey, P. Determination of D-amino acids. II. Use of a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzene. *Carlsberg Res. Commun.* **49,** 591–596; 10.1007/BF02908688 (1984).
8. B'Hymer, C., Montes-Bayon, M. & Caruso, J. A. Marfey's reagent: past, present, and future uses of 1-fluoro-2,4-dinitrophenyl-5-L-alanine amide. *J. Sep. Sci.* **26,** 7–19 (2003).
9. Panter, F., Krug, D., Baumann, S. & Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **9,** 4898–4908; 10.1039/C8SC01325J (2018).
10. Wang, H. *et al.* RecET direct cloning and Redαβ recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nat. Protoc.* **11,** 1175–1190; 10.1038/nprot.2016.054 (2016).
11. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature*; 10.1038/nature03959 (2005).
12. Rothberg, J. M. & Leamon, J. H. The development and impact of 454 sequencing. *Nat. Biotechnol.* **26,** 1117–1124; 10.1038/nbt1485 (2008).
13. Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6,** 493–505 (1999).

14. Röttig, M. *et al.* NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **39,** W362-W367; 10.1093/nar/gkr323 (2011).

15. Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.,* W81-W87; 10.1093/nar/gkz310 (2019).

5. Zooshikella

The Discovery of Zoomarinepane: A Result of Varying Cultivation Conditions to Provoke Changes in the Metabolome of *Zooshikella marina* sp. Uxx12806

Joy Birkelbach, Ronald Garcia, Asfandyar Sikandar, Daniel Krug, and Rolf Müller

Affiliation

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

Contributions and Acknowledgements

The Author's Effort

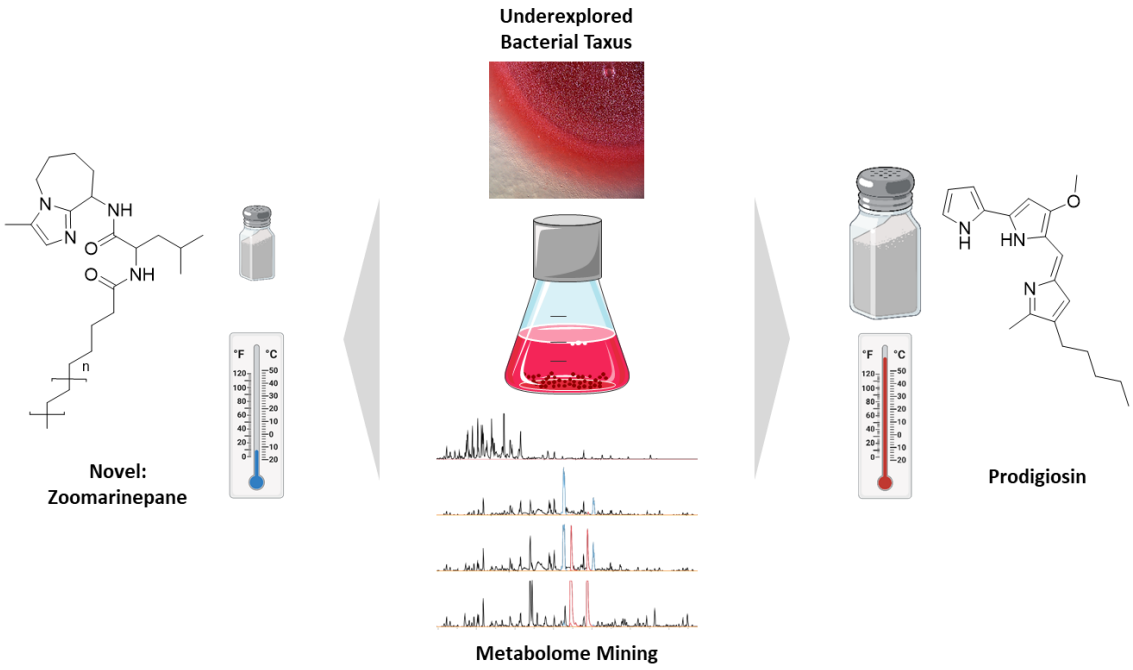
This author significantly contributed to the concept of this study, the design and performance of experiments, and the interpretation of the results. Cultivation, statistical analysis, compound purification, *de novo* structure elucidation, and the proposal of the biosynthetic model was conducted by the author. Furthermore, the author conceived and wrote this manuscript.

Contributions by Others

Ronald Garcia contributed by isolating the strain *Zooshikella marina* sp. Uxx12806. Asfandyar Sikandar contributed to the domain structure design and biosynthesis discussion. Daniel Krug and Rolf Müller contributed to the conception and supervision of this study, editing and proofreading of this manuscript. We thank Judith Hoffmann for the verification and proofreading of the structure elucidation and Sebastian Walesch for scientific discussions and proofreading of the manuscript.

5.3. Abstract

Natural products have proven to supply a great diversity of chemical scaffolds, which often exhibit biological activities and provide novel sources of drug candidates. As chemical diversity correlates with taxonomic diversity, it is fundamentally important to investigate diverse taxa. In order to find novel chemical scaffolds and possibly bioactive natural products we investigated the metabolome of the novel isolate *Zooshikella marina* sp. Uxx12806, which belongs to an underexplored bacterial phylum. We were able to show how varying cultivation conditions effect the production of its secondary metabolites. While the known compound prodigiosin was upregulated under high salinity and high temperature conditions, the production of the herein discovered novel compound family zoomarinepane was upregulated under low salinity and low temperature conditions. Following the UHPLC-HR-MS and -MS^2 based discovery of zoomarinepane, we elucidated its chemical structure using NMR, FT-ICR-MS and –MS^2, and Marfey's analysis discovering a unprecedented imidazo[1,2-a]azepin amine moiety. The intriguing structure has led us to propose a putative biosynthetic pathway and to speculate on the formation of the ring moiety, using the retro-biosynthetic approach to identify the NRPS gene cluster responsible for the biosynthesis of zoomarinepane.



Graphical abstract.

5.4. Introduction

Microbial natural products are known to be prolific sources of a wide spectrum of active pharmaceutical ingredients with antifungal, antibacterial, anticancer, or anti-parasitic activities[1–3]. In the past, natural products have also proven their worth in many other sectors, like nutrition, agriculture, aquaculture, or pigments[4–7]. Nevertheless, the rediscovery rate of known natural products has been a problem for industry and academia in recent decades, which resulted from the fact that only very few taxa were screened for bioactive natural products, using only a limited number of microbiological, molecular biological, and analytical methods[2,5]. This rediscovery problem confounds the fact that the distribution of already discovered compounds does not correlate with the biosynthetic potential of their producers. Indeed, it has been shown that the biosynthetic potential of bacteria is far from being adequately explored, and it is recommended to investigate underexplored taxonomically distant strains and employ advanced biotechnological and analytical methods in order to discover novel chemical scaffolds[8–11].

The marine habitat is one of nature's niches that harbours a large bacterial diversity, which has been underrepresented as natural product producer in recent years, but offers great potential[4,12,13]. In this study, we discovered a stain from the genus *Zooshikella*, which is a marine-derived gammaproteobacterium. So far, only three species belonging to this genus have been validly published, *Z. ganghwensis*[14], *Z. marina*[15] and *Z. harenae*[16], and one species that has not been approved as a novel species, *Z. rubidus*[17]. Apart from their ability to produce the red pigment prodigiosin, which not only causes the pink-red colour of the strains but also displays a broad range of bioactivities with an antitumor, antibacterial and algicidal profile, the metabolome of this genus has not been explored yet[17,18]. The strains were isolated from different marine habitats, ranging from sand samples retrieved from an oyster in the North Sea, Germany[16], to the sediment of the shore of the Yellow Sea, South Korea[14], and to beach sand from the Shivrajpur–Kachigad beach, India[15]. Previous experiments showed that the cells differ in colour and growth characteristics when cultivating at different conditions and different media, showing optimal growth at 30-35 °C, at pH 7-8 and 2-4% (w/v) NaCl[14,15]. Cultivation experiments with other prodigiosin producers like *Serratia*, *Streptomyces* or *Vibrio* showed that the production of the pigment depends on different factors, like iron and salt concentration and the cultivation temperature, showing different effects whether the strain was derived from marine or terrestrial environment[18–23]. Additionally, it has been shown that varying cultivation conditions can have opposite effects on the transcription and production of prodigiosin compared to other natural products in *Streptomyces*[22]. Considering that in routine cultivation usually only less than 10% of biosynthetic gene clusters (BGC) are translated into secondary metabolites[2,5,24], our interest into *Zooshikella* sp. was piqued as its metabolome

has not yet been explored in detail, let alone the impact of cultivation experiments on the metabolome.

Therefore, this study investigated the effect of altered cultivation conditions on the metabolome of *Zooshikella marina* sp. Uxx12806. The effect of different temperatures and salt concentrations on the metabolome was demonstrated, showing similar effects on the production of prodigiosin compared to marine-derived *Streptomyces* and halophilic *Vibrio*[21,22]. Furthermore, we were able to detect, isolate and characterise a novel compound family, which showed a differential production profile compared to prodigiosin. This compound family, named zoomarinepane, shows an unprecedented chemical scaffold with a imidazo[1,2-a]azepin amine moiety. Based on an *in silico* analysis of the putative *zmp* biosynthetic gene cluster, we devised a possible biosynthetic pathway for the biosynthesis of zoomarinepane.

5.5. Results and Discussion

In the course of our ongoing search for novel bacterial species to find novel natural products and possibly novel bioactive compounds, we discovered the Gram-negative gammaproteobacterium *Zooshikella marina sp.* Uxx12806. It was isolated from a marine sample because its ability to glide on agar stood out (**Figure 5. 1**).

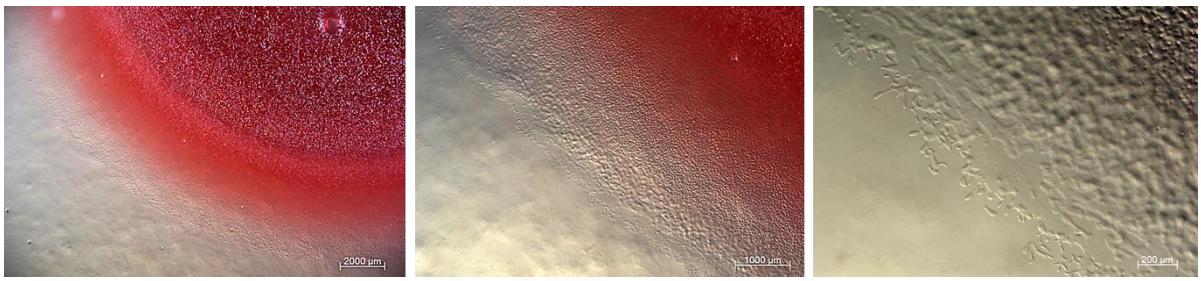


Figure 5. 1. **Zooshikella marina** ***sp. Uxx12806 on TSY6-SWS agar****.*

Genome and MS based dereplication. According to its 16S ribosomal RNA sequence Uxx12806 belongs to the family of *Zooshikellaceae*, its position in the phylogenetic tree can be viewed in **Figure S5. 1**. Genomic DNA was submitted to Illumina sequencing, which resulted in 56 contigs and an estimated genome size of six mbp. Analysis of the sequenced genome with antiSMASH 7.0.0[25] revealed 13 putative BGC (and eight putative BGC regions at contig edges, referred in parentheses), of which four (seven) BGCs encode non-ribosomal peptide synthetases (NRPS), two

encode polyketide synthases (PKS) and NRPS hybrids, and seven (one) encode others like saccharides, fatty acids, and a betalactone (**Table S5. 1**). Out of these 13 BGCs, four could be identified that are likely to encode the biosynthetic machineries of known compounds althiomycin[26], ectoine[27], griseobactin[28], and prodigiosin[29] with a similarity score of > 40%, and a high consensus of KownClusterBlast and MIBiG comparison. Several BGC could not be annotated with known compounds, which sparked our interest to investigate the metabolome of this strain (**Figure 5. 2**).

After a first cultivation of *Z. marina* sp. Uxx12806 in TSY6-SWS medium the crude extract was submitted to UHPLC-MS and -MS^2 analysis and the received data was dereplicated against our in-house database Myxobase[30], the Npatlas database[31], the dictionary of natural products[32], and LipidMaps[33]. Under the herein tested conditions, we were able to detect the production of the known compound prodigiosin[34], which was verified by MS^2-fragmentation analysis (**Figure S5. 2**), and a large number of metabolites that could not be mapped to known compounds.

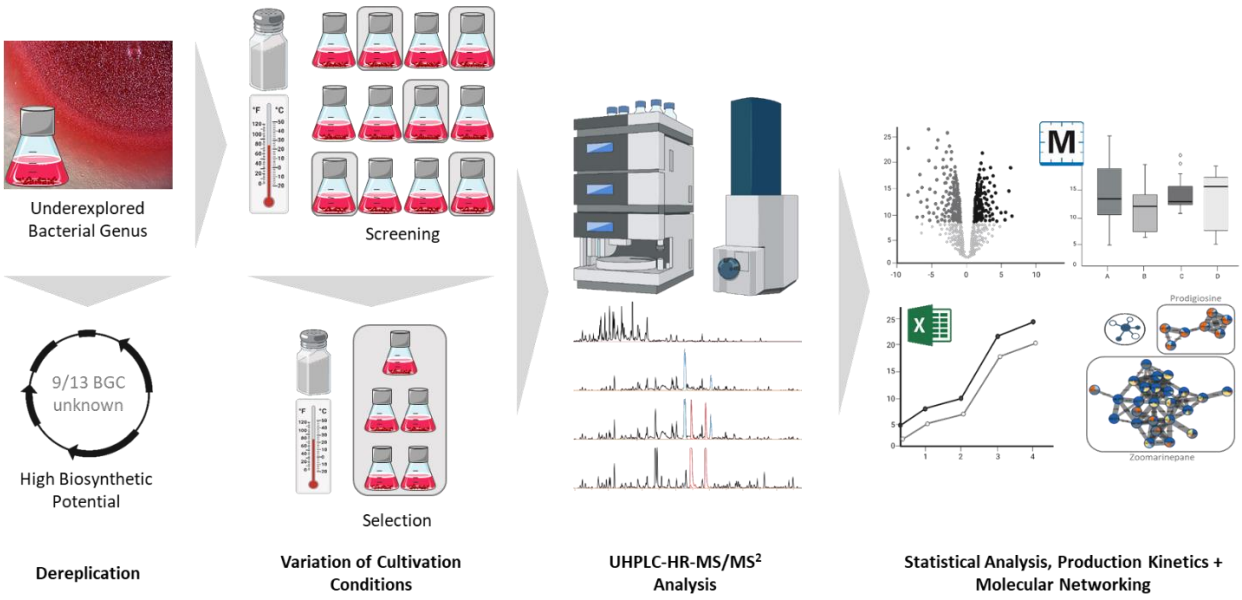


***Figure 5. 2. The metabolome mining workflow towards the discovery of zoomarinepane**, using different cultivation conditions to produce changes in the metabolome.*

Variation of cultivation conditions to produce changes in the metabolome. In order to test the influence of altered cultivation conditions on prodigiosin production and possibly enhance the production of novel compounds, the salt supplementation and temperature was changed when cultivating *Z. marina* sp. Uxx12806 (**Figure 5. 2**). For a rough screening of salinity tolerance, different versions of TSY6-SWS medium were prepared with sea salt concentrations ranging from 1 to 50 g/L. Furthermore, cultivation in standard TSY6-SWS medium was carried out at temperatures between 18 and 37 °C. Inspection of the cultures showed a great range in culture density as well as colour (**Figure 5. 3**), especially when cultivating with different salt concentrations. However, cultivation with 1 g/L sea salt did not show any growth after 10 days.

The evaluation and dereplication of UHPLC-HR-MS and -MS^2 analysis of each crude extract showed great differences in the secondary metabolite profile (**Figure S5. 3**), especially highlighted by the production levels of prodigiosin which showed similar effects to reported halophilic prodigiosin producers[21,22]. In this study, prodigiosin was produced best with higher salt concentrations and at medium temperatures. Furthermore, the known natural product althiomycin could be identified based on UHPLC-MS features using our in-house library Myxobase and published data[30,35]. Althiomycin was only produced at 18 °C in small amounts during these experiments (**Figure S5. 4**). A further UHPLC-MS feature with *m/z* 405.3223 at a retention time of 8.55 min, later named zoomarinepane A, sparked our interest, as its production levels highly depended on salt concentration in the opposite manner as for the prodigiosins (**Figure S5. 3**).

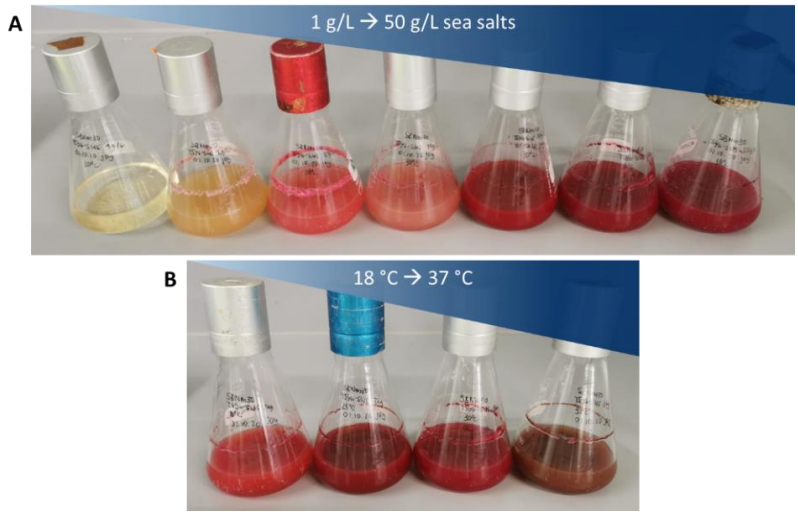


Figure 5. 3.* Zooshikella marina *sp. Uxx12806 cultivated (A) with different sea salt concentrations and (B) at different temperatures. *A: sea salt concentrations from left to right 1, 2, 5, 10, 20, 36, 50 g/L; cultivated at 30 °C. B: Temperatures from left to right 18, 25, 30, 37 °C; cultivated with 36 g/L sea salt.*

Selected cultivation conditions for statistical- and production kinetic analysis. To further investigate this trend, a low salt (2 g/L) and a high salt (50 g/L) condition, a low (18 °C) temperature and a high (37 °C) temperature condition, as well as the standard condition (36 g/L, 30 °C) were selected for a repeated cultivation experiment for statistical and production kinetic analysis. The end point UHPLC-MS analysis showed prodigiosin to be produced best at the standard condition and higher salt concentrations, whereas the production level at the low salt concentration condition was lowest, almost at the detection minimum (**Figure 5. 4**, left). Production kinetic experiments showed a similar trend. Prodigiosin production increased during the first three days, primarily for the standard condition. A lag time and slower increase in production could be observed for low temperature experiments, and almost no production of prodigiosin could be observed for the low salt condition (**Figure 5. 4**, right). Althiomycin, on the other hand, was only produced in detectable amounts during the low salt condition at end-point analysis (**Figure 5. 4**, left). However, production kinetic experiments showed low production with the standard condition at the starting time point only. At later time points, it was not detectable with this condition. The highest production level for althiomycin was after 5 days during the low salt condition, which showed a decrease of detectable amounts afterwards (**Figure 5. 4**, right). Nevertheless, as althiomycin was produced at the low temperature condition in previous experiments (**Figure S5. 3**) the production seems to depend on other parameters as well, which should be investigated in further experiments. Zoomarinepane A showed an inversed production trend compared to the prodigiosin production. It was best produced in the low salt condition and the least in the high salt condition (**Figure 5. 4**, left). Production kinetic experiments showed overall similar trends of zoomarinepane A production at all cultivation conditions. However, the steepest increase of production after three days could be observed at the high temperature experiment, which was surpassed by the low salt condition after seven days (**Figure 5. 4**, right).

A t-test ANOVA analysis was used to calculate the significance level of the production levels of all detected features. The biggest log2 fold change with the best significance level when analysing the production of prodigiosin congeners versus the production of zoomarinepane congeners could be shown in the low salt condition compared to the standard condition (**Figure S5. 7**-**Figure S5. 12**).

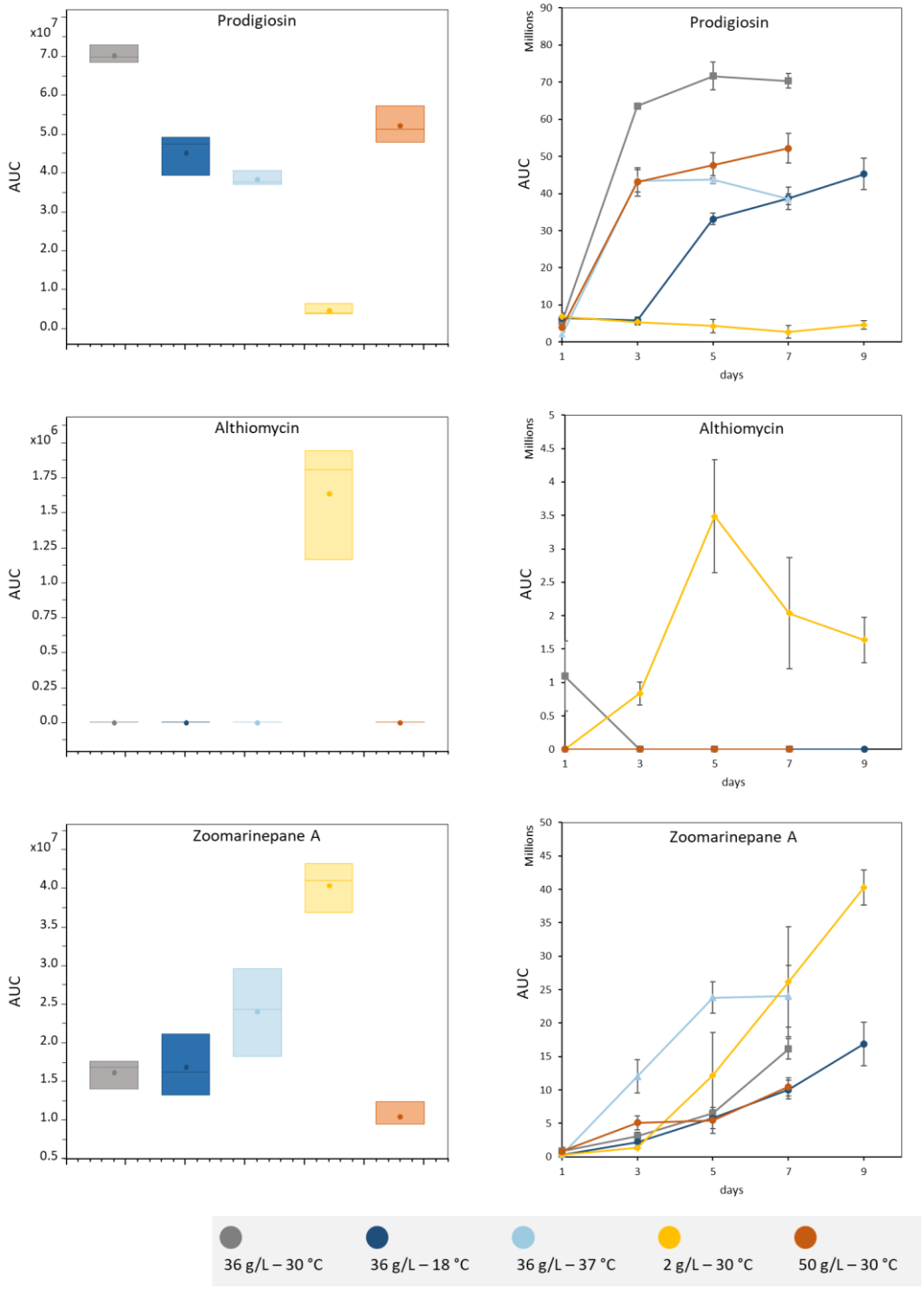


Figure 5. 4. Production level of* Z. marina *sp. Uxx12806 compounds when cultivating at different temperatures and with different sea salt concentrations. *Left side: End-point analysis of the production levels of prodigiosin, althiomycin and zoomarinepane A displayed in box plots. Right side: Time-point analysis of the production kinetics of prodigiosin, althiomycin and zoomarinepane displayed in line charts. All production levels were analysed by the area under the curve (AUC) of the extracted ion chromatograms calculated using MetaboScape. Due to slow growth, the 2 g/L and 18 °C conditions were cultivated for nine instead of seven days.*

The end-point UHPLC-MS^2 analysis was used for feature-based molecular networking[36] that allowed us to search for further derivatives of the detected compounds (**Figure S5. 6**). Here, too, the production level of each compound during each condition was analysed and compared. The molecular cluster of prodigiosin congeners showed best production at the high salt and standard condition experiments, whereas zoomarinepane congeners overall showed best production at the low temperature and low salt conditions. The end-point production level for althiomycin was too low for MS^2 experiments.

Selection of zoomarinepane as a target compound. Due to the interesting production characteristics of the zoomarinepane compound family, we investigated the HR-MS^2 spectra of congeners that were produced in highest amounts, named zoomarinepane A-D, in order to get further insights into their structural features. The MS^2-fragmentation pattern showed fragments that matched leucine (*m/z* 166.134-279.218, **Figure S5. 13**) in case of zoomarinepane A-C, and valine (*m/z* 166.134-265.202, **Figure S5. 13**) in case of zoomarinepane D. The C-terminal fragment (*m/z* 166.134) could be detected in all congeners, however, did not match any proteinogenic amino acid or known compound fragment. The N-terminal fragment varied in size. These unknown features of the hitherto unknown natural product prompted us to elucidate its structure and find insights into its possibly interesting biosynthetic origin.

Structure elucidation. ESI-HR-MS of zoomarinepane A (**1**) displayed a $[M+H]^+$ signal at *m/z* 405.3221 (calcd for $C_{23}H_{41}N_4O_2$, 405.3224, Δ=0.74 ppm) consistent with the molecular formula $C_{23}H_{40}N_4O_2$ including six double-bond equivalents (DBE). Evaluation of the ^{1}H NMR spectrum of **1** showed two signals characteristic for α-protons of a peptide ($\delta_{H\text{-}2'}$ 4.48 and $\delta_{H\text{-}2''}$ 5.05), as well as a signal for a deshielded aromatic proton ($\delta_{H\text{-}1'''}$ 6.60) and signals for a set of deshielded diastereotopic methylene protons ($\delta_{H\text{-}6''}$ 3.89 and 4.12). Furthermore, several overlapping signals could be observed (δ_H 0.9-2.3), which could be resolved and assigned to the molecule by analysing the 2D NMR spectra HSQC, COSY and HMBC as follows (**Table S5. 5**).

From a spin system obtained from COSY and HMBC correlations, an octanoic acid moiety could be derived starting from a methyl group CH_3-8 ($\delta_{H\text{-}8}$ 0.89, $\delta_{C\text{-}8}$ 14.6) through a sequence of six methylene groups CH_2-7 – CH_2-2 (δ_H 1.27-2.27, δ_C 23.8-37.1), and ending with the carbonyl C-1 ($\delta_{C\text{-}1}$ 176.6; **Figure 5. 5**). Furthermore, based on characteristical shifts as well as COSY and HMBC correlations, the α-proton H-2' ($\delta_{H\text{-}2'}$ 4.48, $\delta_{C\text{-}2'}$ 53.4) proved to be part of a leucine moiety (**Figure 5. 5**). The other α-proton, H-2'' ($\delta_{H\text{-}2''}$ 5.05, $\delta_{C\text{-}2''}$ 50.6), however, could not be assigned to any

proteinogenic amino acid. Instead, COSY and HMBC correlations indicated it was connected to a chain of four diastereotopic methylene groups ($\delta_{H\text{-}3''}$ 1.70 and 1.90, $\delta_{H\text{-}4''}$ 1.89 and 2.02, $\delta_{H\text{-}5''}$ 1.57 and 1.88, $\delta_{H\text{-}6''}$ 3.89 and 4.12), whereas the terminal position of this spin system, position 6'', showed deshielded shifts ($\delta_{H\text{-}6''}$ 3.89 and 4.12, $\delta_{C\text{-}6''}$ 45.5) indicating a connection to a nitrogen atom. Additionally, H-2'' showed HMBC correlations to a fully substituted aromatic carbon ($\delta_{C\text{-}1''}$ 149.6), which in turn was part of a 5-methyl-imidazole moiety, the structure of which was confirmed based on characteristical carbon shifts and HMBC correlations from the methyl protons H-3''' ($\delta_{H\text{-}3'''}$ 2.19) to C-1''' ($\delta_{C\text{-}1'''}$ 123.6) and from H-1''' ($\delta_{H\text{-}1'''}$ 6.60) to C-1''. Furthermore, an HMBC correlation from the deshielded methylene protons H-6'' to C-2''' ($\delta_{C\text{-}2'''}$ 130.5) was observed revealing an imidazo[1,2-a]azepin amine moiety (**Figure 5. 5**). Finally, the sequence of the molecule was established through HMBC correlations from α-protons to carbonyl carbons, showing the sequence octanoic acid - leucine - imidazo[1,2-a]azepin amine (**Figure 5. 5**), being in accordance with the calculated DBE.

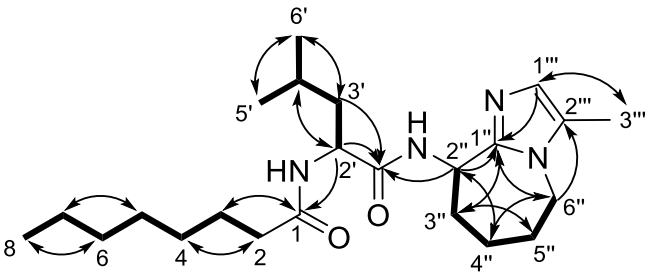


Figure 5. 5. Key COSY (bold lines, back) and HMBC (arrows) correlations of zoomarinepane A (1).

The absolute configuration of the leucine moiety was elucidated using Marfey's method. Therefore, the D- and L-FDLA (1-fluoro-2,4-dinitrophenyl-5-L/D-leucinamide) derivatives of **1** after hydrolysis were analysed using UHPLC-ESI-HR-MS in comparison to an authentic standard (**Figure S5. 14**)[37,38]. Leucine proved to be L-configured (**Figure 5. 6**).

The other stereogenic center, the α-position of the imidazo[1,2-a]azepin amine moiety, could neither be solved by Marfey's analysis nor by NMR experiments as we did not possess an authentic standard or observe NOE correlations to this position. Indeed, after close evaluation of NMR and LC-MS data, which showed signal doubling or peak splitting in all experiments (**Figure S5. 15**A and B), it was assumed that zoomarinepanes exist as an equilibrium of diastereomers caused by the exchange and inversion of the acidic α-proton H-2'' (**Figure S5. 15**C). To prove this hypothesis, a solution of **1** in methanol-d_4 (+ 0.1 % TFA-*d*) was submitted to FT-ICR-MS and $-MS^2$ measurements, which showed the exchange of all acidic protons, including the α-proton H-2'' (**Figure S5. 16**).

ESI-HR-MS analysis of zoomarinepane B (**2**) and C (**3**) showed a $[M+H]^+$ signal at *m/z* 433.3536 (calcd for $C_{25}H_{45}N_4O_2$, 433.3537, *Δ*=0.21 ppm) and at *m/z* 461.3856 (calcd for $C_{27}H_{49}N_4O_2$, 461.3850, *Δ*=1.4 ppm), which correspond to molecular formulae containing C_2H_4 or C_4H_8 more than **1**, respectively. Consistent with this finding, 1D and 2D NMR data of **2** and **3** showed high resemblance to those of **1**, however exhibited two or four more CH_2 units, respectively. Based on COSY and HMBC correlations, the fatty acid chains were found to be elongated by one C_2H_4 unit in case of **2**, or by two C_2H_4 units in case of **3**, and thus feature a decanoic and dodecanoic acid moiety, respectively (**Table S5. 6** and **Table S5. 7**).

The molecular formula of zoomarinepane D (**4**) $C_{22}H_{38}N_4O_2$ was supported by ESI-HR-MS data with a $[M+H]^+$ signal at *m/z* 391.3071 (calcd for $C_{22}H_{39}N_4O_2$, 391.3068, *Δ*=0.8 ppm). Compared to **1**, the only differences in the HSQC spectrum that could be observed were signals characteristical for valine instead of leucine. Indeed, the analysis of COSY and HMBC spectra proved this finding (**Table S5. 8**). Therefore, **4** exhibits the sequence octactanoic acid - valine - imidazo[1,2-a]azepin amine.

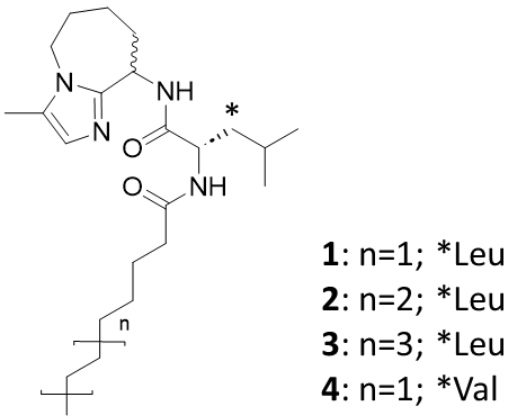


Figure 5. 6. Chemical structures of zoomarinepane A-D (1-4). *The wavy bond represents the inversion of the stereogenic centre in solution.*

Proposal for the biosynthetic pathway of zoomarinepane. Based on the chemical structures of zoomarinepane a non-ribosomal peptide synthetase (NRPS) biosynthetic origin seemed most likely. Therefore, a feeding experiment with isotopically labelled amino acids was conducted, which showed the incorporation of leucine, lysine and threonine (**Figure S5. 17**-**Figure S5. 18**). Furthermore, one of the NRPS encoding BGCs fit the expected number of modules: a fatty acid loading module and three amino acid elongation modules. The respective BGC, named *zpm*, was dereplicated using ClusterBlast, KnownClusterBlast and the MIBiG data base[25,39], and revealed a low agreement with BGC of published compounds. The putative *zmp* BGC spans a region about 30 kb , whereas the core BGC consists of three genes. A detailed analysis of the genes in the *zmp* BGC in *Z. marina* sp. Uxx12806 was performed using the BLAST algorithm, showing the highest similarity to NRP synthetases of *Zooshikella ganghwensis* (**Table S5. 10**). Using the retro-biosynthetic approach, the following proposal for the biosynthesis of zoomarinepane, including the intriguing cyclisation of lysine and threonine to form the imidazo[1,2-a]azepin amine moiety, was elaborated (**Figure 5. 7**A-C). The assembly line starts with the recognition and activation of a fatty acid as a starter molecule by a fatty acid adensosine monophosphate ligase (AL), the fatty acid is then transferred to the first acyl carrier protein (ACP). The activated substrate undergoes three NRPS based elongation cycles, each consisting of condensation (C), adenylation (A) and transfer to a peptidyl carrier protein (PCP). The assembly line is terminated by a thioesterase (TR) domain with predicted oxidoreductive activity (**Figure 5. 7**B, **Table S5. 10** and **Table S5. 11**). This typical NRPS assembly line forms the linear peptide fatty acid-leu-lys-thr, before it undergoes cyclisation and oxidation. However, *in silico* analysis of the adenylation domain specificities is not consistent with the incorporated amino acids (**Table S5. 11**). Based on the Stachelhaus code, the adenylation domains in module one and two should prefer to load hydrophobic-aliphatic amino acids, showing the best hit for alanine in module one and glycine in module two. However, zoomarinepane possess a leucine moiety that is introduced through module one, and a lysine moiety that is introduced through module two. Both incorporated amino acids were verified by feeding experiments. The third adenylation domain should prefer to load a hydrophilic amino acid with an aromatic side chain, while threonin is incorporated into zoomarinepane according to feeding experiments. A relaxed substrate specificity of the A domain may be responsible for such observation, which can be directly influenced by the adjacent C domain, as shown by extensive biochemical experiments on microcystin biosynthesis[40].

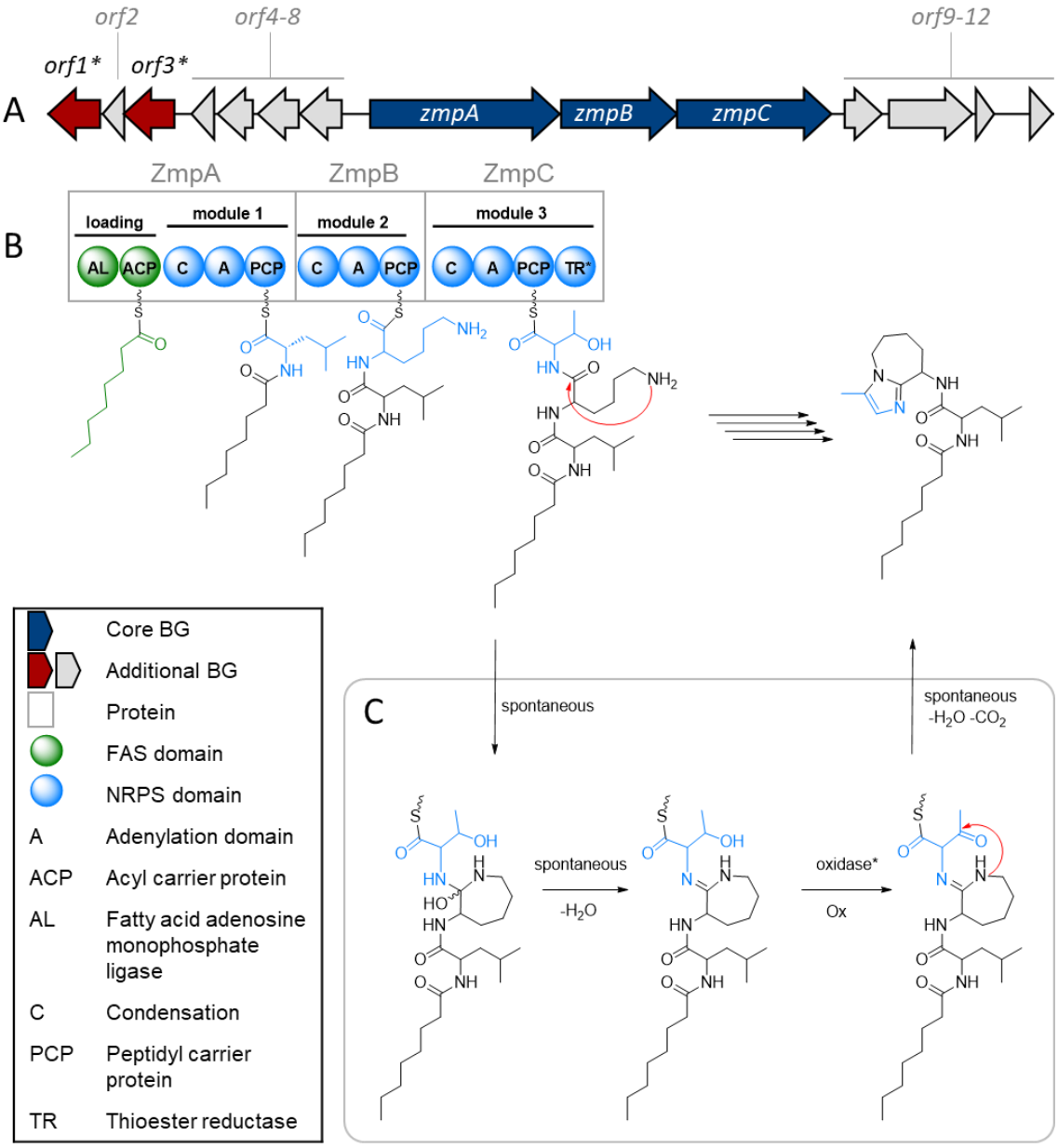


Figure 5. 7. The putative biosynthetic gene cluster (A) and the proposed biosynthetic pathway (B) leading to zoomarinepane, including a proposal for the oxidation and cyclisation of the imidazo[1,2-a] azepin amine moiety (C). **the oxidation is proposed be conducted either by upstream-located* orf1 *(FAD dependent monooxygenase),* orf3 *(CYP-450) or by the TR domain, which is predicted to belong to the SDR family with oxidoreductive activity. The function of additional genes in grey (*orf2, orf4-8, orf9-12*) remains elusive.*

The cyclisation of the lysine moiety is proposed to happen spontaneously, similar to cyclised lysine in mycobactin[41] or the cyclised ornithine in potensibactin[42]. The next step, the oxidation of the threonine hydroxy group, however, is unlikely to occur spontaneously, but rather by a tailoring enzyme such as an upstream-located cytochrome p450 (**Figure 5. 7**A and C), or the termination domain, which is predicted to exhibit oxidoreductive activity (**Figure 5. 7**B and C). To gain further insights into the potential role of the termination domain (TR, module 3, **Figure 5 .7**B), we generated a 3D model using AlphaFold[43,44]. This model was subsequently used to find structural homologues via DALI server[45]. The TR model share high structural homology to the NADPH-dependent terminal reductase (R) domain from the myxalamid pathway, MxaA R (pdb: 4u7w; **Figure 5 .8**A)[46]. The overlay of the MxaA R-NDPH complex structure with the TR AlphaFold model revealed the NDAPH- and putative substrate binding sites. As expected the NADPH-binding region contains the well-conserved nucleotide-binding motif TGxxGxxG located in close proximity to the proposed catalytic triad (**Figure 5 .8**B). MxaA R is responsible for the release of a linear substrate from pPant-tethered carrier protein (PCP), which is reflective by the presence of a long and narrow substrate binding tunnel. In comparison, the ZmpC TR predicted substrate binding region is compact and wide (**Figure 5 .8 C**). This suggests that TR acts on a fully cyclised PCP-bound intermediate, with p450(s) likely responsible for the oxidation of the threonine hydroxyl group. So far, only reductive release of natural products from the NRPS assembly line followed by spontaneous cyclisation have been reported[47,48], which makes the case of zoomarinepane particularly interesting to investigate further through biochemical experiments. Finally, the decarboxylation is proposed to happen spontaneously, similar to siphonazole[49], or mediated by the termination domain as well.

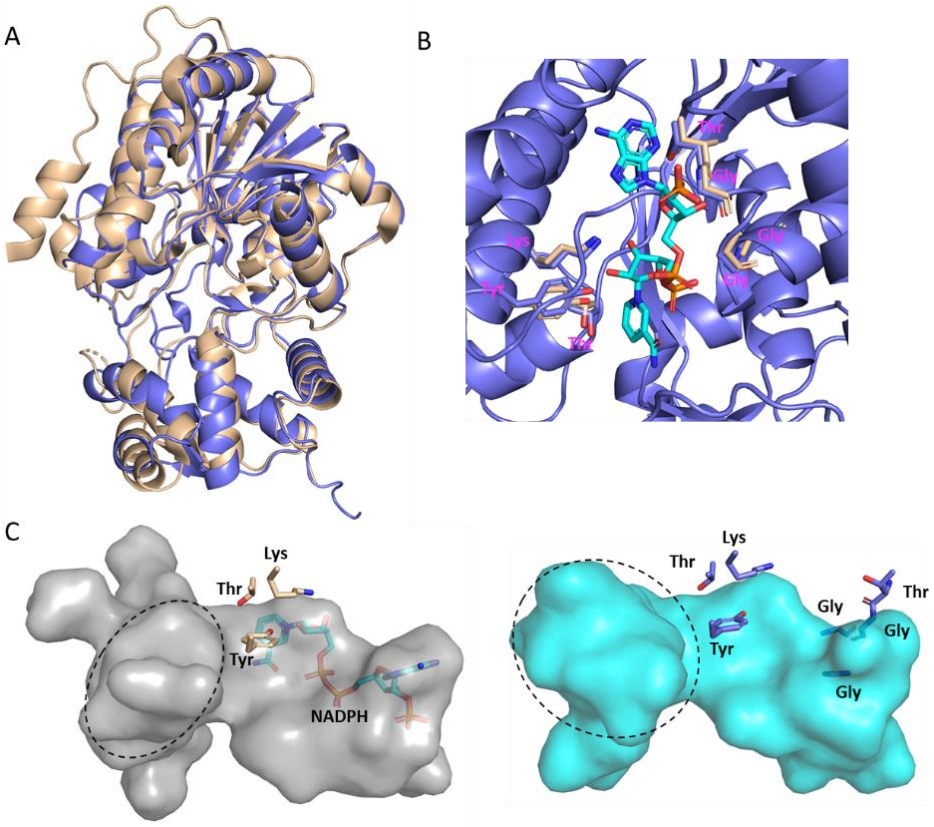


Figure 5. 8. Structural analysis of the ZmpC TR domain*. (A) Superposition of the TR model (slate) predicted by AlphaFold[43,44] with structural homologue identified from DALI (MxaA R; pdb: 4u7w; wheat)[45,46]. Cα RMSD was approx. 0.84 Å over the entire length of the protein. (B) The close-up view of the active-site and co-factor binding residues illustrate high level of homology shared by TR and MxsA R. For clarity only the active site residues and nucleotide-binding motif (TGxxGxxG; Thr1154, Gly1155, Gly1158 and Gly1161) of MxaA R are shown. (C) Surface representation of the substrate and co-factor binding cavities of TR and Mxa R calculated using CavitOmiX. Dotted circles highlight the difference in the size of the substrate binding pockets. The active-site residues, nucleotide-bind motif and NDAPH (cyan) are shown as sticks.*

In order to unambiguously identify the zoomarinepane BCG, further experiments are required, such as an inactivation of the BGC by knock out or the heterologous expression of zoomarinepane. As *Z. marina* sp. Uxx12806 was not genetically amenable, alternative producers could be sought out. Considering that two published *Zooshikella* strains harbour BGCs, which highly resemble the *zmp* BGC (**Figure S5. 19**)[50], it would be interesting to screen their metabolome for zoomarinepanes and investigate their genetic accessibility. Further proof of the biosynthetic pathway could be provided by biochemical experiments. Especially a feeding experiment with a SNAC-ester of the linear peptide to the termination domain or to the cytochrome P450 could give further insights into the cyclisation and oxidation steps required for the formation of the imidazo[1,2-a]azepin amine moiety.

Bioactivity. Zoomarinepane A and B were tested against a panel of bacterial and fungal pathogens as well as cancer cell lines. However, neither exhibited a significant antimicrobial nor cytotoxic activity (**Table S5. 12**). Therefore, the biological activity of zoomarinepanes is yet unknown and requires further investigation.

5.6. Experimental Section

UHPLC-MS and MS² measurements. UPLC-MS measurement were performed on a Dionex (Germering, Germany) Ultimate 3000 RSLC system equipped with Waters (Eschborn, Germany) BEH C_{18} column (100 × 2.1 mm, 1.7 µm) equipped with a Waters VanGuard BEH C_{18} 1.7 µm guard column. The flow rate was set to 0.6 mL/min, and the column temperature was 45 °C. The 18 minute gradient was as follows: 0-0.5 min, 95% water with 0.1% formic acid (A) and 5% acetonitrile with 0.1% formic acid (B); 0.5-18.5 min, 5-95% B; 18.5-20.5 min, 95% B; 20.5-21 min, 95-5% B; 21-22.5 min, 5% B. For the 9 minute gradient following method was used: 0-0.5 min, 95% water with 0.1% formic acid (A) and 5% acetonitrile with 0.1% formic acid (B); 0.5-9.5 min, 5-95% B; 9.5-10.5 min, 95% B; 10.5-10.8 min, 95-5% B; 10.8-12.5 min, 5% B. UV-vis 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 Bruker Daltonics (Bremen, Germany) maXis 4G HR-qToF mass spectrometer equipped with an Apollo II ESI source. Mass spectra were acquired in centroid mode ranging from 150-2500 *m/z* at a 2 Hz full scan rate. Mass spectrometry source parameters were set to 500 V as end plate offset; 4000 V as capillary voltage; nebuliser gas pressure 1 bar; dry gas flow of 5 L/min and a dry temperature of 200 °C. Ion transfer and quadrupole settings were set to funnel RF 350 Vpp.; multipole RF 400 Vpp as transfer settings and ion energy of 5 eV as well as a low mass cut of 300 *m/z*. Collision cell was set to 5.0 eV and pre-pulse storage time was set to 5 µs. Spectra acquisition rate was set to 2 Hz.

For MS^2 experiments, CID (collision-induced dissociation) energy was ramped from 35 eV for 500 *m/z* to 45 eV for 1,000 *m/z*. MS full scan acquisition rate was set to 2 Hz and MS² spectra acquisition rates were ramped from 1 to 4 Hz for precursor ion intensities of 10 kcts to 1,000 kcts.

Calibration was done automatically before every LC-MS run by injection of a sodium formiate solution and calibration on the respective clusters formed in the ESI source. All MS analyses were acquired in the presence of the lock masses ($C_{12}H_{19}F_{12}N_3O_6P_3$, $C_{18}H_{19}O_6N_3P_3F_2$ and $C_{24}H_{19}F_{36}N_3O_6P_3$) which generate the [M+H]+ ions of 622.0289; 922.0098 and 1221.9906.

NMR measurements. NMR data was recorded on an AVANCE III 700 MHz NMR (^{1}H at 700 MHz, ^{13}C at 175 MHz) equipped with a 5 mm inverse TCI cryoprobe (Bruker, Billerica, MA, USA). Shift values (δ) were calculated in ppm, and coupling constants (J) were calculated in Hz. Spectra were recorded in methanol-d_4, or chloroform-d and adjusted to the solvent signals (methanol-d_4 δ_H 3.31, δ_C 49.2, and chloroform-d δ_H 7.27, δ_C 77.0). Measurements were conducted in 5 mm Shigemi tubes (Shigemi Inc., Allison Park, PA 15101, USA). For the two-dimensional experiments HMBC, HSQC, ROESY, and gCOSY standard pulse programs were used. HMBC experiments were optimised for $^{2,3}J_{C\text{-}H}$ = 6 Hz, and HSQC were optimised for $^{1}J_{C\text{-}H}$ = 145 Hz.

Bacterial strains and culture conditions. *Zooshikella marina* sp. Uxx12806 was cultivated at 18, 25, 30, or 37 °C and 180 rpm in 100 mL TSY6-SWS-36 medium (300 mL shaking flask, see **Table S5. 2**) or at 30 °C and 180 rpm in 100 mL (300 mL shaking flask) TSY6-SWS-1, -2, -5, -10, -20, and -50 medium (**Table S5. 2**) for seven days in duplicates. For statistical analysis and feature based molecular networking, and for the analysis of production levels *Z. marina* sp. Uxx12806 was cultivated at 18, 30, or 37 °C and 180 rpm in 100 mL TSY6-SWS-36 medium (300 mL shaking flask, see **Table S5. 2**) or at 30 °C and 180 rpm in 100 mL (300 mL shaking flask) TSY6-SWS-2, and -50 medium (see **Table S5. 2**) for seven to ten days in triplicates. For the compound isolation of zoomarinepanes *Z. marina* sp. Uxx12806 was cultivated for seven days at 30 °C and 180 rpm in six times 1.5 L TSY6-SWS-36 medium (total: 9 L) in 5 L Erlenmeyer cultivation flasks. All cultures were inoculated with 1% (v/v) seed culture (grown for 24 hours), and to all cultures 2% Amberlite XAD-16 resin was added.

Preparation and sequencing of genomic DNA. *Zooshikella marina* sp. Uxx12806 was grown in 100 mL TSY6-SWS medium at 30 °C for seven days. The culture was spun down at 8288 x g (Eppendorf Centrifuge 5804R) for 10 minutes at room temperature (RT) and the supernatant was discarded. The cells were washed in 15 mL and resuspended in 5 mL of SET buffer. 50 µL RNAse stock solution was added and the reaction mixture incubated at 37 °C for 30-60 minutes. After adding 300 µL Proteinase K stock solution, the mixture was inverted several times. 600 µL 10% SDS was added and incubated for two hours at 55 °C until it turned viscous. After that, 6 mL phenol:chloroform:isoamylalcohol (25:24:1) was added and swung for one hour with 5 rpm. Then, the mixture was centrifuged for five minutes at 8000 rpm and RT, followed by transferring the upper aqueous phase into a new tube by using an end-cut 1 mL Eppendorf pipette tip in order to avoid DNA shearing. The extraction step was repeated with the upper phase. The clean upper

phase was extracted with 5 mL chloroform:isoamylalcohol (24:1), swung for one hour and centrifuged at 8000 rpm and RT for 10 minutes. After transferring 4 mL of the upper layer into a new tube using an end-cut 1 mL Eppendorf pipette tip, 440 µL of 3 M sodium acetate pH 5.2 and 11 mL of pure ice cold ethanol was added. The tube was inverted several times, until cotton like DNA appeared. The DNA was collected by rolling it on the end of a Pasteur pipette. After washing it several times with 70% ethanol the DNA was air dried and suspended in 1 mL sterile MQ water. In order to solve the DNA its suspension was incubated at 55 °C. The quantity and quality of DNA was tested using a NanoDrop 2000c spectrometer (Thermo Fisher Scientific) at 260 nm inspecting the 260/280 and 260/230 ratio. The solution was stored at -20 °C. Illumina sequencing of gDNA was conducted by the NGS service of HZI.

Metabolomics analysis. Cultures were grown in triplicates. Cells and medium blank with XAD-16 resin were harvested by centrifugation and extracted by maceration in 50 mL acetone:methanol (1:1, v/v) for three hours. After drying the extracts, they were dissolved in 1 mL methanol. Crude extracts and medium blanks were submitted to UHPLC-MS measurement in a 1:200 (v/v) dilution. Data was processed and annotated using DataAnalysis 5.3 and MetaboScape 2022b (both softwares from Bruker Daltonics, Billerica, MA, USA) and our in-house database Mxbase[30], parameters are shown in **Table S5. 3**. Blanks were removed, and a t-test/ANOVA was run with MetaboScape.

Feature based molecular networking. Cells or medium blank with XAD-16 resin were harvested by centrifugation and extracted by maceration in 50 mL acetone:methanol (1:1, v/v) for three hours. After drying the extracts, they were dissolved in 1 mL methanol. Crude extracts and medium blanks were submitted to UHPLC-MS^2 measurement in a 1:200 (v/v) dilution. Data was processed and annotated using MetaboScape 2021b (Bruker) and our in-house database Mxbase[30], parameters are shown in **Table S5. 3**. Compounds with MS^2 spectra were exported with the built-in "Export for GNPS/Sirius" function and uploaded to the GNPS server[51] at University of California San Diego via FileZilla FTP upload to ftp:// ccms-ftp01.ucsd.edu to be analysed using GNPS feature based molecular networking[36,51], parameters shown in **Table S5. 4**. The production level of the compounds produced under each condition was compared with the other conditions, and presented in pie charts using Cytoscape 3.7.1. Singletons and molecular clusters that contained more than 60% medium blank were removed.

Compound isolation. *Zooshikella marina* sp. Uxx12806 was grown in 9 L TSY6-SWS medium supplemented with 2% Amberlite XAD-16 resin at 30 °C for 10 days. Cells and XAD-16 resin were pelleted with 15970 x g and 4 °C for 15 minutes. The supernatant was discarded and the pellet lyophilised. The pellet was extracted four times with 30 mL/g methanol:acetone (1:1, v/v), each time stirring with 250 rpm for 30-60 min. Afterwards the solvent was removed via a rotary evaporator with appropriate pressure and a water bath temperature of 40 °C. The crude extract was partitioned between hexane and methanol + 5 % deionised water (1:1, v/v; each layer 30 mL/g dry pellet), and the methanol extract was subsequently partitioned between dichloromethane (DCM) and deionised water (1:1, v/v; each layer 30 mL/g dry pellet), then between ethyl acetate and deionised water (1:1, v/v; each layer 30 mL/g dry pellet). The DCM extract was fractionated with normal phase flash chromatography using the IsoleraTM Spektra One (Biotage®) system. As stationary phase high-purity grade silica gel (pore size 60 Å, 70-230 mesh, 63-200 µm in a 25 g SNAP cartridge (Biotage®)) was used with a gradient of ethyl acetate (A) and methanol (B). The gradient parameters were as follows: 5 column volumes (CV) with 0% B, 35 CV 0-80% B, 40 CV 80-100% B and 45 CV 100% B.

The fraction containing zoomarinepane A-D was separated by semi-preparative HPLC (Waters xSelect® Peptide CSH C_{18}, 5 μm, 250 × 10 mm; 5 mL/min; 40 °C; MS detector HCTplus ion trap (Bruker) using a gradient as follows: 0-2 min, 95% H_2O with 0.1% formic acid (A) and 5% acetonitrile with 0.1% formic acid (B); 2-3 min, 5% B to 30% B; 3-22 min, 30% to 55% B; 22-23 min, 55% to 95% B; 23-25 min, 95% B; 25-26 min, 95% to 5% B; 26-26.5 min, 5% B. 3.5 mg **1** (t_R=11.5 min), 2.6 mg **2** (t_R=16.7 min), 0.5 mg **3** (t_R=22.1 min) and 0.6 mg **4** (t_R=9.9 min) were obtained.

Structure elucidation. *Zoomarinepane A (**1**)*. Yellow solid; *M* = 404.60; UV (41% ACN+0.1% FA) λ_{max} 200, 225 nm; for ^{1}H and ^{13}C NMR data see **Table S5. 5** and **Figure S5. 20**-**Figure S5. 24**; ESI-HR-MS *m/z* 405.3221 (calcd 405.3224).

*Zoomarinepane B (**2**)*. Yellow solid; *M* = 432.65; UV (49% ACN+0.1% FA) λ_{max} 195, 225 nm; for ^{1}H and ^{13}C NMR data see **Table S5. 6** and **Figure S5. 25**-**Figure S5. 29**; ESI-HR-MS *m/z* 433.3536 (calcd 433.3537).

*Zoomarinepane C (**3**)*. Yellow solid; *M* = 460.71; UV (55% ACN+0.1% FA) λ_{max} 195, 225 nm; for ^{1}H and ^{13}C NMR data see **Table S5. 7** and **Figure S5. 30**-**Figure S5. 33**; ESI-HR-MS *m/z* 461.3856 (calcd 461.3850).

*Zoomarinepane D (**4**)*. Yellow solid; *M* = 390.57; UV (39% ACN+0.1% FA) λ_{max} 200, 220 nm; for ^{1}H and ^{13}C NMR data see **Table S5. 8** and **Figure S5. 34**-**Figure S5. 38**; ESI-HR-MS *m/z* 391.3071 (calcd 391.3068).

FT-ICR-MS measurement of proton exchange. After incubating a solution of **1** (10 µM in methanol-d_4 + 0.1% deuterated trifluoric acid (TFA-*d*)) for one hour, the sample was measured with direct infusion electrospray-fourier transform ion cyclotron resonance (DI-FT-ICR)-MS and –MS/MS on the Bruker SolariX XR 7T. A control experiment was conducted after dissolving the compound in methanol instead. The mass spectrometer was externally calibrated to a mass accuracy below 1 ppm before measurements. Therefore, the instrument was switched from standby to operate and an LC/MS calibration standard for ESI-TOF (Agilent) was injected. Calibration files for the tune mix measurements were saved and used to evaluate the performance of the instrument during the period of measurements. The samples were injected with a preinstalled syringe pump at a flow rate of 3.0 μL/min. 32 scans were performed accumulating for 500 ms. The total m/z ranged from 150 to 2000. The data size was set to 8 M with a 2 s transient. Collision RF Amplitude was optimised, parameters are described in **Table S5. 9**.

***In silico* analysis of the genome of *Zooshikella marina* sp. Uxx12806.** The prediction of all the BGCs was done in antiSMASH 7.0.0.[25] employing the detection strictness “relaxed” and enabling all extra features. (**Table S5. 1**).

Proposal for the biosynthetic pathway of zoomarinepane. The catalytic domains of the core BGC and their substrate specificity were analysed using CCD[52], Pfam 35.0[53], NaPDoS2[54], and the Stachelhaus code[55]. Furthermore, all proteins that were not predicted to be related to NRPS domains were analysed using the BLAST algorithm[56] (**Table S5. 10** and **Table S5. 11**).

Feeding experiment. Feeding studies of *Zooshikella marina* sp. Uxx12806 were performed in 10 mL TSY6-SWS medium in a 50 mL Erlenmeyer shaking flask with the supplementation of 1 mL of a 25 mM amino acid stock solution of L-threonine-$^{13}C_4$,^{15}N, L-lysine dihydrochloride-$^{13}C_6$,$^{15}N_2$, L-leucine-d_3, L-histidine-$^{15}N_3$, and L-alanine-$^{13}C_3$,^{15}N, respectively. The amino acid stock solutions were prepared in D_2O except for L-leucine-d_3 that was dissolved in D_2O:methanol-d_4 (1:1 v/v) due

to solubility issues. Therefore, equal aliquots (250 μL) of the stock solution were added to the culture after 24, 36, 48, and 60 h. A control experiment without isotopically labelled precursor was done in parallel. After 72 h Amberlite XAD-16 resin (10% v/v) was added to the culture. The culture was harvested by centrifugation after 90 h. Cells including XAD were extracted with 40 mL acetone:methanol (1:1 v/v). The solvents were evaporated, and the extract was resolved in methanol and analysed by UHPLC-MS in a 1:100 (v/v) dilution injecting 2 µL.

Antimicrobial activity tests. All microorganisms were obtained from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung für Mikroorganismen und Zellkulturen, DSMZ). the Coli Genetic Stock Center (CGSC) or were part of our internal collection and were handled according to standard sterile microbiological procedures and techniques. Minimum inhibitory concentrations (MIC) of **1**-**4** were determined by microbroth dilution assays. Methanolic stocks (1 mg/mL) were used to prepare serial dilutions (0.125 to 64 µg/mL) in sterile 96 well plates and the test strain was added to a final suspension of 10^4 to 10^5 colony forming units (CFU)/mL. Afterwards the plates were incubated for 24 h at 30 °C or 37 °C, depending on the microorganism. The lowest concentration at which no growth was observed by visual observation was considered as the MIC..

Cytotoxicity assay. Cell lines were obtained from the DSMZ and were cultured under conditions as recommended by the depositor. Cells were seeded at 6×10^3 cells per well of the 96-well plates in 180 μL complete medium and treated with compounds in serial dilution after 2 h of equilibration. Both compounds and the internal solvent control were tested in duplicate. After 5 d of incubation, 20 μL of 5 mg/mL MTT (thiazolyl blue tetrazolium bromide) in PBS (phosphate-buffered saline) was added per well, and samples were further incubated for 2 h at 37 °C. The medium was then discarded, and cells were washed with 100 μL PBS before 100 μL 2-propanol/10 N HCl (250:1) was added in order to dissolve formazan granules. The absorbance at 570 nm was measured using a microplate reader, and cell viability was expressed as percentage relative to the respective methanol control. The half-maximal inhibitory concentration (IC_{50}) values were determined by sigmoidal curve fitting by plotting % relative growth versus applied concentrations using 0% (full inhibition) and 100% (no inhibition) as borders.

5.7. Conclusion

In this study, different cultivation conditions altering the temperature profile and salt content of the cultivation medium were applied to the strain *Zooshikella marina* sp. Uxx12806, which is an underexplored bacterial species with high biosynthetic potential. The conditions were found to lead to the up and down regulation of various compounds. The herein discovered and characterised zoomarinepane compound family proved to be of particular interest, as its production profile is inverse compared to the known compound prodigiosin. This effect shows how simple variations in cultivation conditions can lead to different metabolite profiles and immensely influence the production profile of certain compounds. The chemical structure of zoomarinepane were elucidated using NMR, Marfey's, and FT-ICR-MS/MS analysis, showing an imidazo[1,2-a]azepin amine moiety unprecedented in natural products. Therefore, a proposal for the biosynthesis of this intriguing scaffold was devised. In order to gain further insights into the formation of the ring, further molecular biological or biochemical experiments should be conducted. In addition, it would be interesting to investigate the effects of the altered cultivation conditions not only on the metabolome, but also on the transcriptome and proteome[57,58]. This could underpin the biosynthesis proposal and uncover regulatory effects for the production of secondary metabolites[22,59].

5.8. Reference

1. Aldholmi, M., Marchand, P., Ourliac-Garnier, I., Le Pape, P. & Ganesan, A. A Decade of Antifungal Leads from Natural Products: 2010–2019. *Pharmaceuticals* **12,** 182; 10.3390/ph12040182 (2019).

2. Walesch, S. *et al.* Fighting antibiotic resistance-strategies and (pre)clinical developments to find new antibacterials. *EMBO Rep.,* e56033; 10.15252/embr.202256033 (2022).

3. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *Journal of natural products*; 10.1021/acs.jnatprod.9b01285 (2020).

4. Demain, A. L. Importance of microbial natural products and the need to revitalize their discovery. *J. Ind. Microbiol. Biotechnol.* **41,** 185–201; 10.1007/s10295-013-1325-z (2014).

5. Katz, L. & Baltz, R. H. Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **43,** 155–176; 10.1007/s10295-015-1723-5 (2016).

6. Rather, I. A. *et al.* Diversity of Marine Bacteria and Their Bacteriocins: Applications in Aquaculture. *Reviews in Fisheries Science & Aquaculture* **25,** 257–269; 10.1080/23308249.2017.1282417 (2017).

7. Domínguez-Borbor, C. *et al.* The marine symbiont Pseudovibrio denitrificans, is effective to control pathogenic Vibrio spp. in shrimp aquaculture. *Aquaculture* **508,** 127–136; 10.1016/j.aquaculture.2019.04.077 (2019).

8. Gavriilidou, A. *et al.* Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nat. Microbiol.* **7,** 726–735; 10.1038/s41564-022-01110-2 (2022).

9. Hoffmann, T. *et al.* Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. *Nat. Commun.* **9,** 803; 10.1038/s41467-018-03184-1 (2018).

10. Pye, C. R., Bertin, M. J., Lokey, R. S., Gerwick, W. H. & Linington, R. G. Retrospective analysis of natural products provides insights for future discovery trends. *Proc. Natl. Acad. Sci. USA* **114,** 5601–5606; 10.1073/pnas.1614680114 (2017).

11. Bader, C. D., Neuber, M., Panter, F., Krug, D. & Müller, R. Supercritical Fluid Extraction Enhances Discovery of Secondary Metabolites from Myxobacteria. *Anal. Chem.* **92,** 15403–15411; 10.1021/acs.analchem.0c02995 (2020).

12. Montaser, R. & Luesch, H. Marine natural products. A new wave of drugs? *Future Med. Chem.* **3,** 1475–1489; 10.4155/fmc.11.118 (2011).

13. Subramani, R. & Aalbersberg, W. Marine actinomycetes. An ongoing source of novel bioactive metabolites. *Microbiological research* **167,** 571–580; 10.1016/j.micres.2012.06.005 (2012).

14. Yi, H., Chang, Y.-H., Oh, H. W., Bae, K. S. & Chun, J. Zooshikella ganghwensis gen. nov., sp. nov., isolated from tidal flat sediments. *Int.J.Syst.Evol.Microbiol.* **53,** 1013–1018; 10.1099/ijs.0.02521-0 (2003).

15. Ramaprasad, E. V. V., Bharti, D., Sasikala, C. & Ramana, C. V. Zooshikella marina sp. nov. a cycloprodigiosin- and prodigiosin-producing marine bacterium isolated from beach sand. *Int. J. Syst. Evol. Microbiol.* **65,** 4669–4673; 10.1099/ijsem.0.000630 (2015).

16. Pira, H. *et al.* Zooshikella harenae sp. nov., Isolated from Pacific Oyster Crassostrea gigas, and Establishment of Zooshikella ganghwensis subsp. marina subsp. nov. and Zooshikella ganghwensis subsp. ganghwensis subsp. nov. *Diversity* **13,** 641; 10.3390/d13120641 (2021).

17. Lee, J. S. *et al.* Exceptional production of both prodigiosin and cycloprodigiosin as major metabolic constituents by a novel marine bacterium, Zooshikella rubidus S1-1. *Appl. Environ. Microbiol.* **77,** 4967–4973; 10.1128/AEM.01986-10 (2011).

18. Han, R., Xiang, R., Li, J., Wang, F. & Wang, C. High-level production of microbial prodigiosin: A review. *Journal of basic microbiology* **61,** 506–523; 10.1002/jobm.202100101 (2021).

19. Silverman, M. P. & Munoz, E. F. Effect of iron and salt on prodigiosin synthesis in Serratia marcescens. *Journal of bacteriology* **114,** 999–1006; 10.1128/jb.114.3.999-1006.1973 (1973).

20. Siva, R., Subha, K., Bhakta, D., Ghosh, A. R. & Babu, S. Characterization and enhanced production of prodigiosin from the spoiled coconut. *ABAB* **166,** 187–196; 10.1007/s12010-011-9415-8 (2012).

21. Gallardo, K., Candia, J. E., Remonsellez, F., Escudero, L. V. & Demergasso, C. S. The Ecological Coherence of Temperature and Salinity Tolerance Interaction and Pigmentation in a Non-marine Vibrio Isolated from Salar de Atacama. *Front. Microbiol.* **7,** 1943; 10.3389/fmicb.2016.01943 (2016).

22. Sevcikova, B. & Kormanec, J. Differential production of two antibiotics of Streptomyces coelicolor A3(2), actinorhodin and undecylprodigiosin, upon salt stress conditions. *Arch. Microbiol.* **181,** 384–389; 10.1007/s00203-004-0669-1 (2004).

23. Suryawanshi, R. K., Patil, C. D., Borase, H. P., Salunke, B. K. & Patil, S. V. Studies on production and biological potential of prodigiosin by Serratia marcescens. *ABAB* **173,** 1209–1221; 10.1007/s12010-014-0921-3 (2014).

24. Nett, M., Ikeda, H. & Moore, B. S. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat. Prod. Rep.* **26,** 1362–1384; 10.1039/b817069j (2009).

25. Blin, K. *et al.* antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.*; 10.1093/nar/gkad344 (2023).

26. Gerc, A. J., Song, L., Challis, G. L., Stanley-Wall, N. R. & Coulthurst, S. J. The insect pathogen Serratia marcescens Db10 uses a hybrid non-ribosomal peptide synthetase-polyketide synthase to produce the antibiotic althiomycin. *PLoS ONE* **7,** e44673; 10.1371/journal.pone.0044673 (2012).

27. Reshetnikov, A. S. *et al.* Diversity and phylogeny of the ectoine biosynthesis genes in aerobic, moderately halophilic methylotrophic bacteria. *Extremophiles : life under extreme conditions* **15,** 653–663; 10.1007/s00792-011-0396-x (2011).

28. Patzer, S. I. & Braun, V. Gene cluster involved in the biosynthesis of griseobactin, a catechol-peptide siderophore of Streptomyces sp. ATCC 700974. *J. Bacteriol.* **192,** 426–435; 10.1128/JB.01250-09 (2010).

29. Wang, X. *et al.* Identification and Verification of the Prodigiosin Biosynthetic Gene Cluster (BGC) in Pseudoalteromonas rubra S4059. *Microbiol Spectr* **9,** e0117121; 10.1128/Spectrum.01171-21 (2021).

30. 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.* **31,** 768–783; 10.1039/c3np70127a (2014).

31. Linington, R. G. Npatlas - The Natural Products Atlas. Available at https://www.npatlas.org (2018).

32. Dictionary of Natural Products. Available at http://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml (2019).

33. Sud, M. *et al.* LMSD: LIPID MAPS structure database. *Nucleic Acids Res.* **35,** D527-32; 10.1093/nar/gkl838 (2007).

34. Kim, D. *et al.* Biosynthesis of antibiotic prodiginines in the marine bacterium Hahella chejuensis KCTC 2396. *J. Appl. Microbiol.* **102,** 937–944; 10.1111/j.1365-2672.2006.03172.x (2007).

35. Cortina, N. S., Revermann, O., Krug, D. & Müller, R. Identification and characterization of the althiomycin biosynthetic gene cluster in Myxococcus xanthus DK897. *Chembiochem : a European journal of chemical biology* **12,** 1411–1416; 10.1002/cbic.201100154 (2011).

36. Nothias, L.-F. *et al.* Feature-based molecular networking in the GNPS analysis environment. *Nat Methods* **17,** 905–908; 10.1038/s41592-020-0933-6 (2020).

37. Marfey, P. Determination of D-amino acids. II. Use of a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzene. *Carlsberg Res. Commun.* **49,** 591–596; 10.1007/BF02908688 (1984).

38. B'Hymer, C., Montes-Bayon, M. & Caruso, J. A. Marfey's reagent: past, present, and future uses of 1-fluoro-2,4-dinitrophenyl-5-L-alanine amide. *J. Sep. Sci.* **26,** 7–19 (2003).

39. Terlouw, B. R. *et al.* MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res.* **51,** D603-D610; 10.1093/nar/gkac1049 (2023).

40. Meyer, S. *et al.* Biochemical Dissection of the Natural Diversification of Microcystin Provides Lessons for Synthetic Biology of NRPS. *Cell Chem. Biol.* **23,** 462–471; 10.1016/j.chembiol.2016.03.011 (2016).

41. Quadri, L. E., Sello, J., Keating, T. A., Weinreb, P. H. & Walsh, C. T. Identification of a Mycobacterium tuberculosis gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. *Chem. Biol.* **5,** 631–645 (1998).

42. Dejong, C. A. *et al.* Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching. *Nat. Chem. Biol.* **12,** 1007–1014; 10.1038/nchembio.2188 (2016).

43. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **50,** D439-D444; 10.1093/nar/gkab1061 (2022).

44. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589; 10.1038/s41586-021-03819-2 (2021).

45. Holm, L. Dali server: structural unification of protein families. *Nucleic Acids Res.* **50,** W210-5; 10.1093/nar/gkac387 (2022).

46. Barajas, J. F. *et al.* Comprehensive Structural and Biochemical Analysis of the Terminal Myxalamid Reductase Domain for the Engineered Production of Primary Alcohols. *Chemistry & biology* **22,** 1018–1029; 10.1016/j.chembiol.2015.06.022 (2015).

47. Kopp, F., Mahlert, C., Grunewald, J. & Marahiel, M. A. Peptide macrocyclization: The reductase of the nostocyclopeptide synthetase triggers the self-assembly of a macrocyclic imine. *J. Am. Chem. Soc.* **128,** 16478–16479; 10.1021/ja0667458 (2006).

48. Wilson, D. J., Shi, C., Teitelbaum, A. M., Gulick, A. M. & Aldrich, C. C. Characterization of AusA: a dimodular nonribosomal peptide synthetase responsible for the production of aureusimine pyrazinones. *Biochemistry* **52,** 926–937; 10.1021/bi301330q (2013).

49. Mir Mohseni, M. *et al.* Discovery of a mosaic-like biosynthetic assembly line with a decarboxylative off-loading mechanism through a combination of genome mining and imaging. *Angew. Chem. Int. Ed.* **55,** 13611–13614; 10.1002/anie.201606655 (2016).

50. Gilchrist, C. L. M. & Chooi, Y.-H. Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics (Oxford, England)*; 10.1093/bioinformatics/btab007 (2021).

51. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34,** 828–837; 10.1038/nbt.3597 (2016).

52. Marchler-Bauer, A. *et al.* CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res.* **37,** D205-10; 10.1093/nar/gkn845 (2009).

53. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49,** D412-D419; 10.1093/nar/gkaa913 (2021).

54. Klau, L. J. *et al.* The Natural Product Domain Seeker version 2 (NaPDoS2) webtool relates ketosynthase phylogeny to biosynthetic function. *J. Biol. Chem.* **298,** 102480; 10.1016/j.jbc.2022.102480 (2022).

55. Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6,** 493–505; 10.1016/S1074-5521(99)80082-9 (1999).

56. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410; 10.1016/S0022-2836(05)80360-2 (1990).

57. Boldt, J. *et al.* Bursts in biosynthetic gene cluster transcription are accompanied by surges of natural compound production in the myxobacterium Sorangium sp. *Microbial Biotechnology*; 10.1111/1751-7915.14246 (2023).

58. Remmers, I. M. *et al.* Orchestration of transcriptome, proteome and metabolome in the diatom Phaeodactylum tricornutum during nitrogen limitation. *Algal Research* **35,** 33–49; 10.1016/j.algal.2018.08.012 (2018).

59. Neubacher, N. *et al.* Symbiosis, virulence and natural-product biosynthesis in entomopathogenic bacteria are regulated by a small RNA. *Nat. Microbiol.* **5,** 1481–1489; 10.1038/s41564-020-00797-5 (2020).

Supporting Information

The Discovery of Zoomarinepane: A Result of Variing Cultivation Conditions to Provoce Changes in the Metabolome of *Zooshikella marina* sp. Uxx12806

Joy Birkelbach, Ronald Garcia, Daniel Krug, and Rolf Müller

Affiliation

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

S 5.1 Strain Description

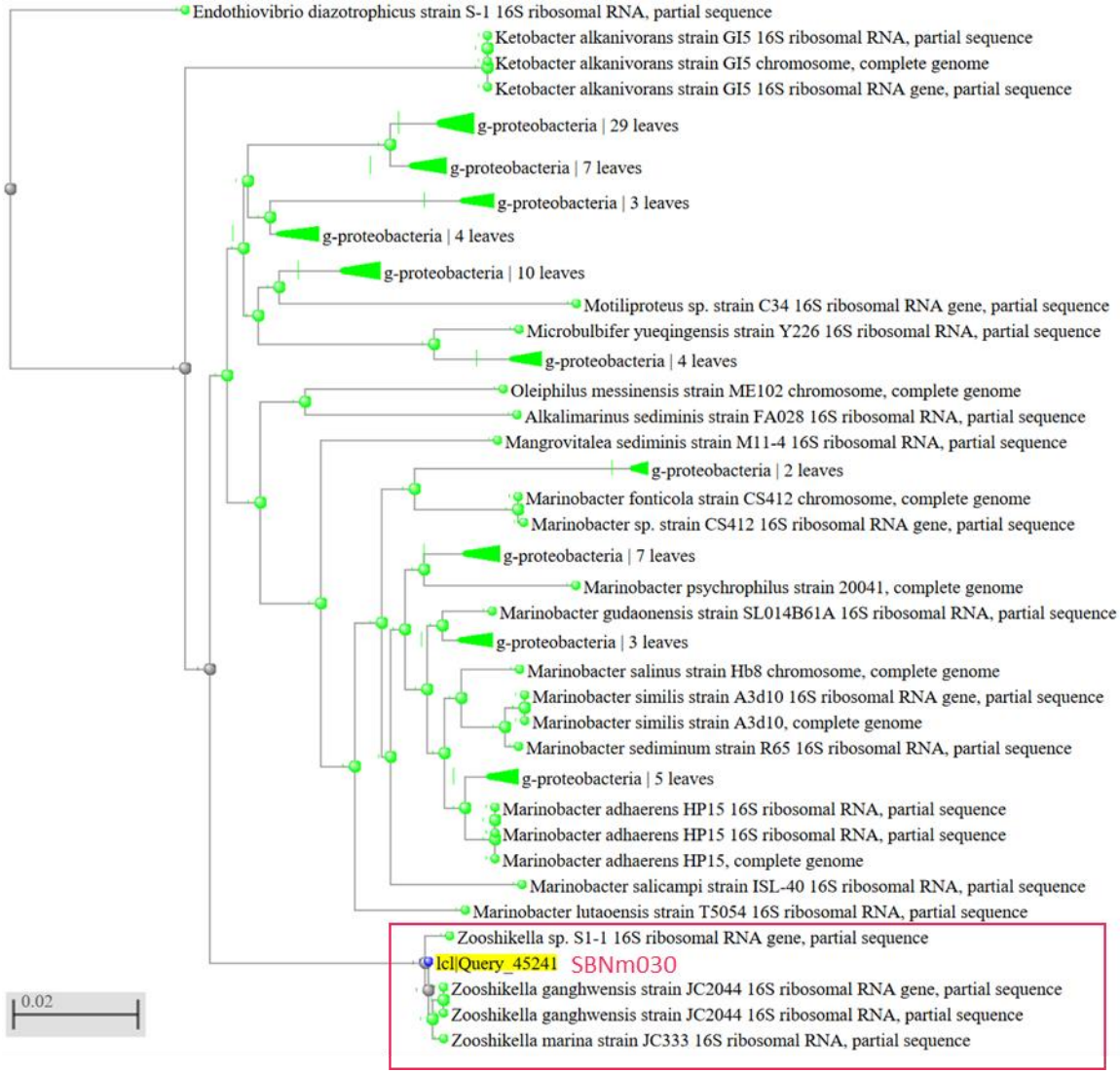


Figure S5. 1. *Phylogenetic cladogram with the position of* Zooshikella marina *sp. Uxx12806.*

S 5.2 Genome and MS Based Dereplication

Table S5. 1. *Predicted Biosynthetic Gene Clusters (BGC) in* Zooshikella marina *sp. Uxx12806 by antiSMASH version 7.0.0[1]. Out of 56 contigs, 20 regions were identified as secondary metabolite regions.*

BGC region	Type	Known compound	Similarity [%]
1.1	NRPS	Saframycin Mx 1	66[a]
2.1	T1PKS, NRPS	**Althiomycin[2]**	62
2.2[b]	NRPS	-	-
3.1	NRPS	-	-
3.2	NRP-metallopore, NRPS	Griseobactin[3]	46
4.1	Prodigiosin	**Prodigiosin[4]**	42
5.1	NRPS-like	-	-
6.1	Saccharide, crocagin	Emulsan	13[a]
7.1[b]	Betalactone	-	-
11.1	NAGGN	-	-
12.1	Saccharide, LAP	-	-
14.1	Arylpolyene, halogenated	APE Vf	35[a]
15.1[b]	NRPS	-	-
20.1	NRPS, T1PKS	-	-
22.1[b]	NRPS-like, indole	Minimycin	80
23.1	Ectoine	Ectoine[5]	83
27.1[b]	NRPS	Type R and F tailocins	6[a]
31.1[b]	NRPS	-	-
48.1[b]	NRPS	-	-
54.1[b]	NRPS	-	-

PKS: polyketide synthase. NRPS: non-ribosomal peptide synthetase. PKS-NRPS: hybrid polyketide synthase-non-ribosomal peptide synthetase. Known compounds (**bold**) which could be identified in Uxx12806 extracts. [a]KnownClusterBlast, MIBiG and BLAST of core BGC against NCBI data base low consensus (<40%) or incomplete match. [b]Region on contig edge.

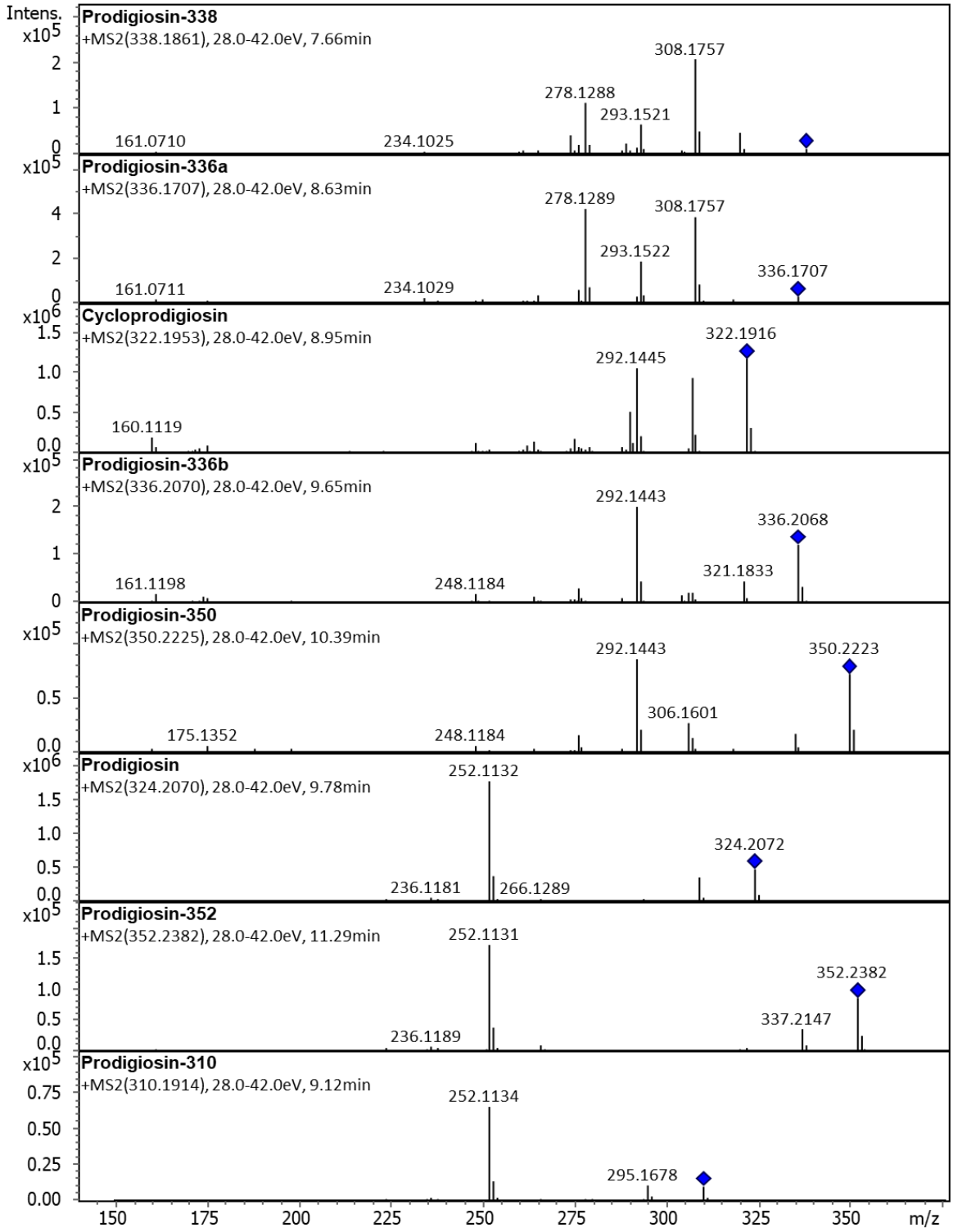


Figure S5. 2. HR-MS² spectra of prodigiosin and analoga. *Product ions that are characteristical for prodigiosins can be observed upon collisionally induced dissociation[6]. The selected precursor ions which* m/z *were selected for fragmentation are marked by a blue rombus.*

S 5.3 Media Composition

All media were prepared with deionised water and sterilised by autoclaving at 121 °C and 2 bar for 20 minutes. The pH was adjusted to pH 7.2 with NaOH before autoclaving.

Table S5. 2. *Cultivation media composition for the cultivation of* Zooshikella marina *sp. Uxx12806.*

Medium	Composition	Supplier	Quantity [g/L]
	Tryptone	BD	6.0
	Potato starch	Roth	4.0
	Yeast extract	BD	1.0
	HEPES	Roth	2.38
TSY6-SWS	Sea salts	ATI	36.0
TSY6-SWS-1 [a]	Sea salts	ATI	1.0
TSY6-SWS-2 [a]	Sea salts	ATI	2.0
TSY6-SWS-5 [a]	Sea salts	ATI	5.0
TSY6-SWS-10 [a]	Sea salts	ATI	10.0
TSY6-SWS-20 [a]	Sea salts	ATI	20.0
TSY6-SWS-50 [a]	Sea salts	ATI	50.0

[a] Composition of the medium corresponds to TSY6-SWS except for the salinity.

S 5.4 Cultivation Experiments

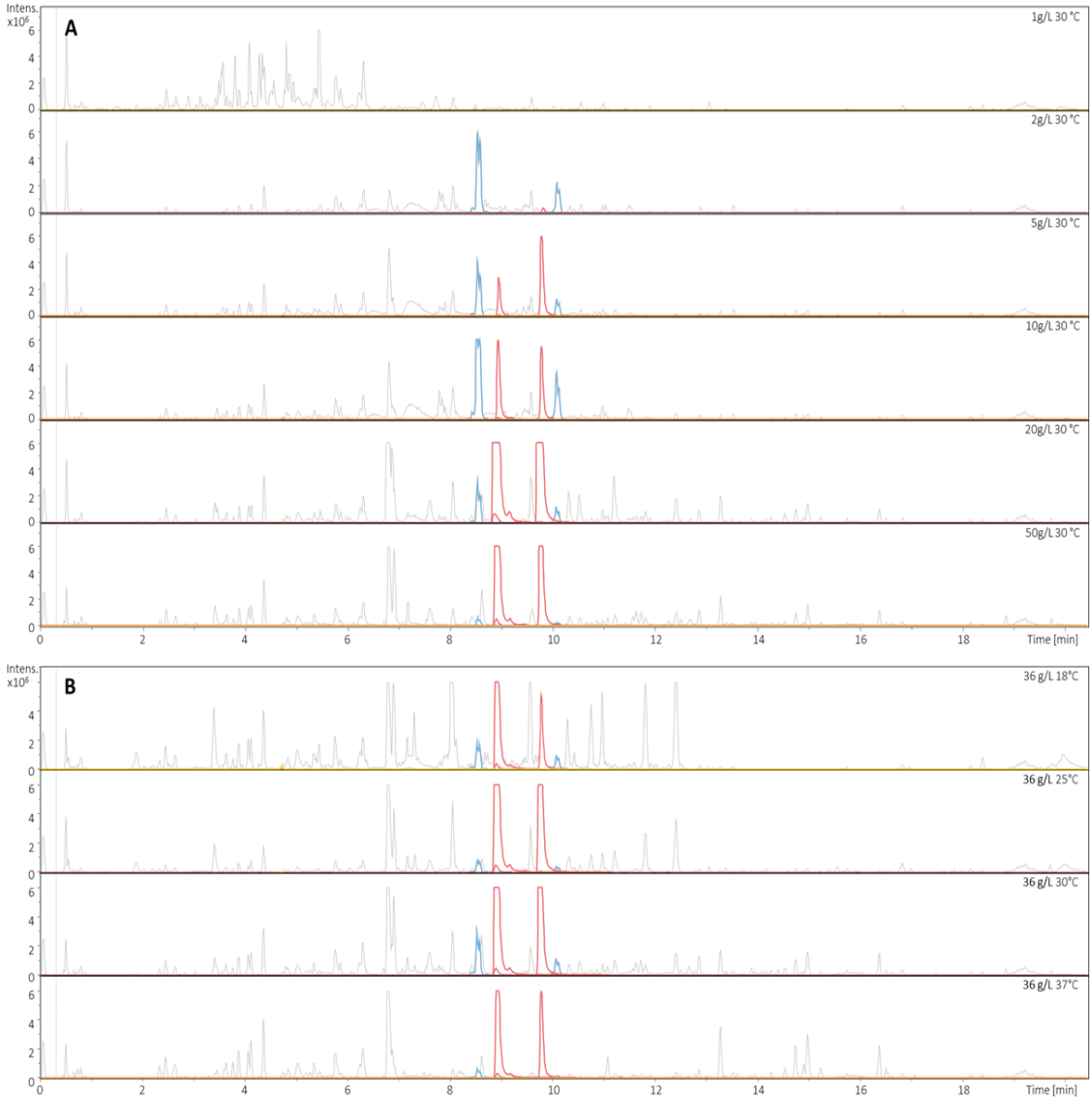


Figure S5. 3. LC-MS chromatograms of* Zooshikella marina *sp. Uxx12806 crude extracts after cultivation with different sea salt concentrations (A) and at different temperatures (B). *The EICs of prodigiosin (red,* m/z *322.1928 ± 0.02, R_t 9.78 min) and cycloprodigiosin (red,* m/z *324.2080 ± 0.02, R_t 8.93 min), zoomarinepane A (blue,* m/z *405.3223 ± 0.02, R_t 8.55 min) and B (blue,* m/z *433.3535 ± 0.02, R_t 10.09 min), and althiomycin (yellow,* m/z *422.0588 ± 0.02w, R_t 4.73 min; only present at 18 °C) are highlighted.*

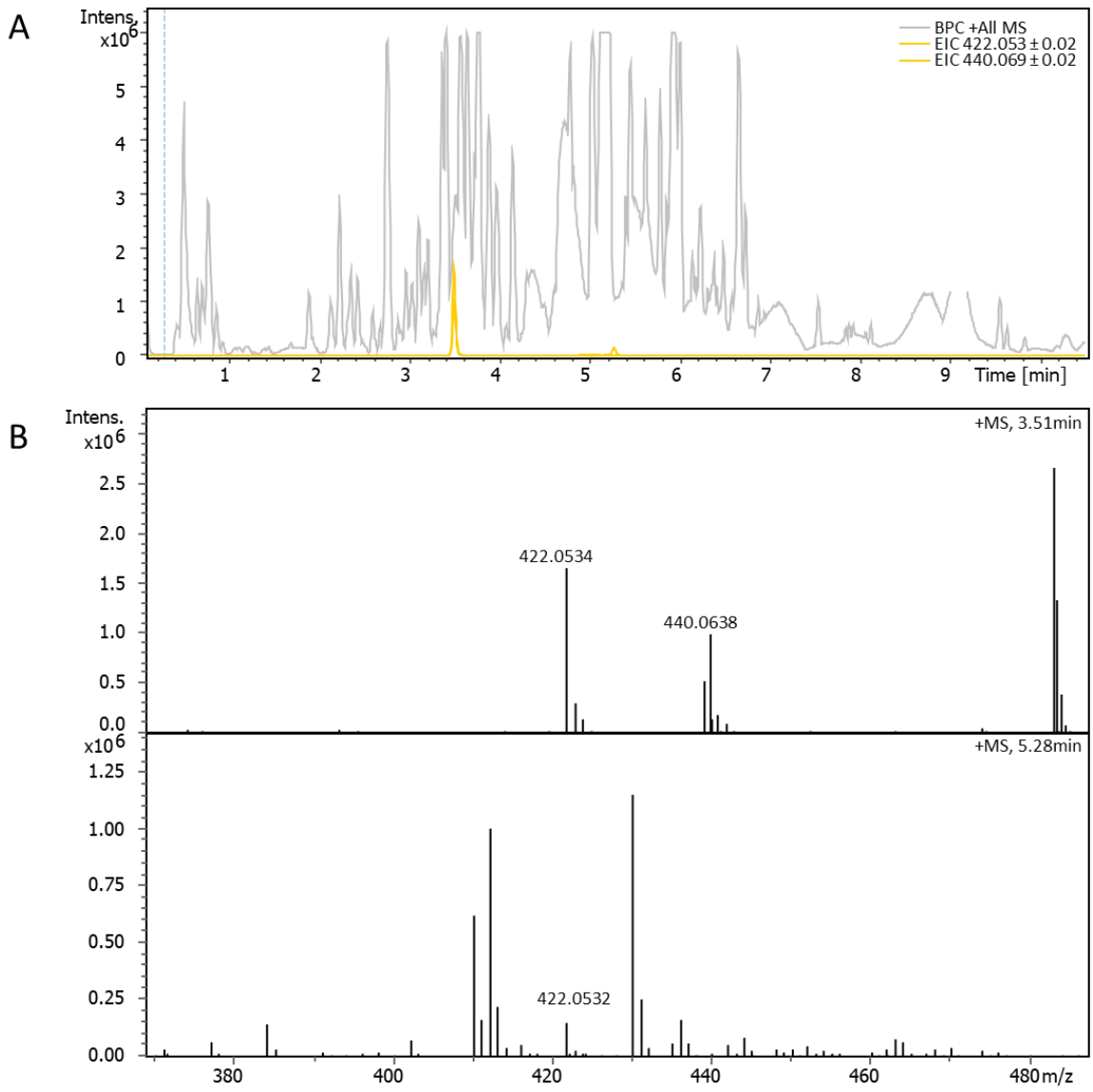


***Figure S5. 4. UHPLC-MS analysis of the production of althiomycin and analoga**. **A:** Base peak chromatogram of* Zooshikella marina *sp. Uxx12806 crude extract after cultivation in TSY6-SWS-2 medium for five days. The extracted ion chromatograms (EIC) of althiomycin (yellow, m/z 440.069 ± 0.02, R_t 3.51 min) and the dehydrated analogon (yellow, m/z 422.053 ± 0.02, R_t 5.28 min) were highlighted. **B:** MS spectra of allthiomycin and the dehydrated althiomycin analogon. Shown data match those of published althiomycin[7]. UHPLC-MS method: 9-minute gradient.*

S 5.5 Statistical Analysis

Table S5. 3. *MetaboScape processing and annotation parameters.*

Parameter	Value
Processing	
Minimum number of featurers for extraction	2
Minimum number of featurers for result	2
Max *m/z*	2000.0
Min *m/z*	150.0
Max retention time [s]	1230.0
Min retention time [s]	30.0
Intensity threshold	25000.0
Extraction minimum cluster length	3
Min seed cluster length	5
Recalibration	Formiate cluster
Annotation, narrow (wide)	
m/z range [ppm]	2 (3)
Retention time range [min]	0.1 (0.2)
mSigma	10 (20)
MS/MS score	900 (800)
CCS	2.0 (5.0)

Table S5. 4. *GNPS feature based molecular networking processing parameters.*

Parameter	Value
Precursor Ion Mass Tolerance [Da]	0.02
Fragment Ion Mass Tolerance [Da]	0.02
Min Pairs Cos	0.7
Min Matched Fragment Ions	6
Max shift between precursors [Da]	500
Network TopK	10
Max Connected Component Size	100
Library Min Matched Peaks	3
Score Threshold	0.7
Maximum Analog Search Mass Difference [Da]	100.0

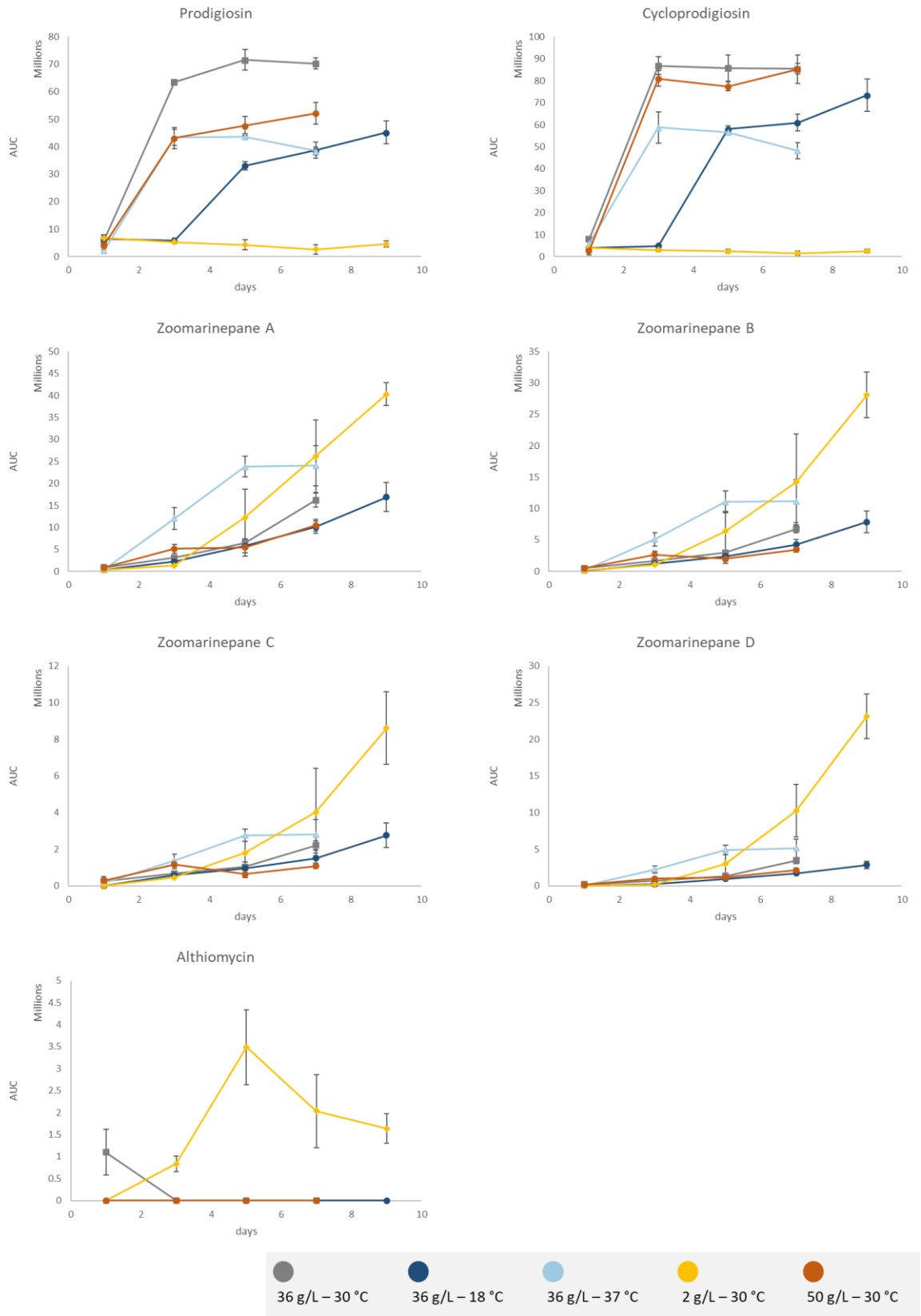


Figure S5. 5. Production level of selected natural products during cultivation of* Zooshikella marina *sp. Uxx12806 at different temperatures and with different sea salt concentrations at specific times. *Due to slow growth, the 2 g/L and 18 °C conditions were cultivated for nine instead of seven days*

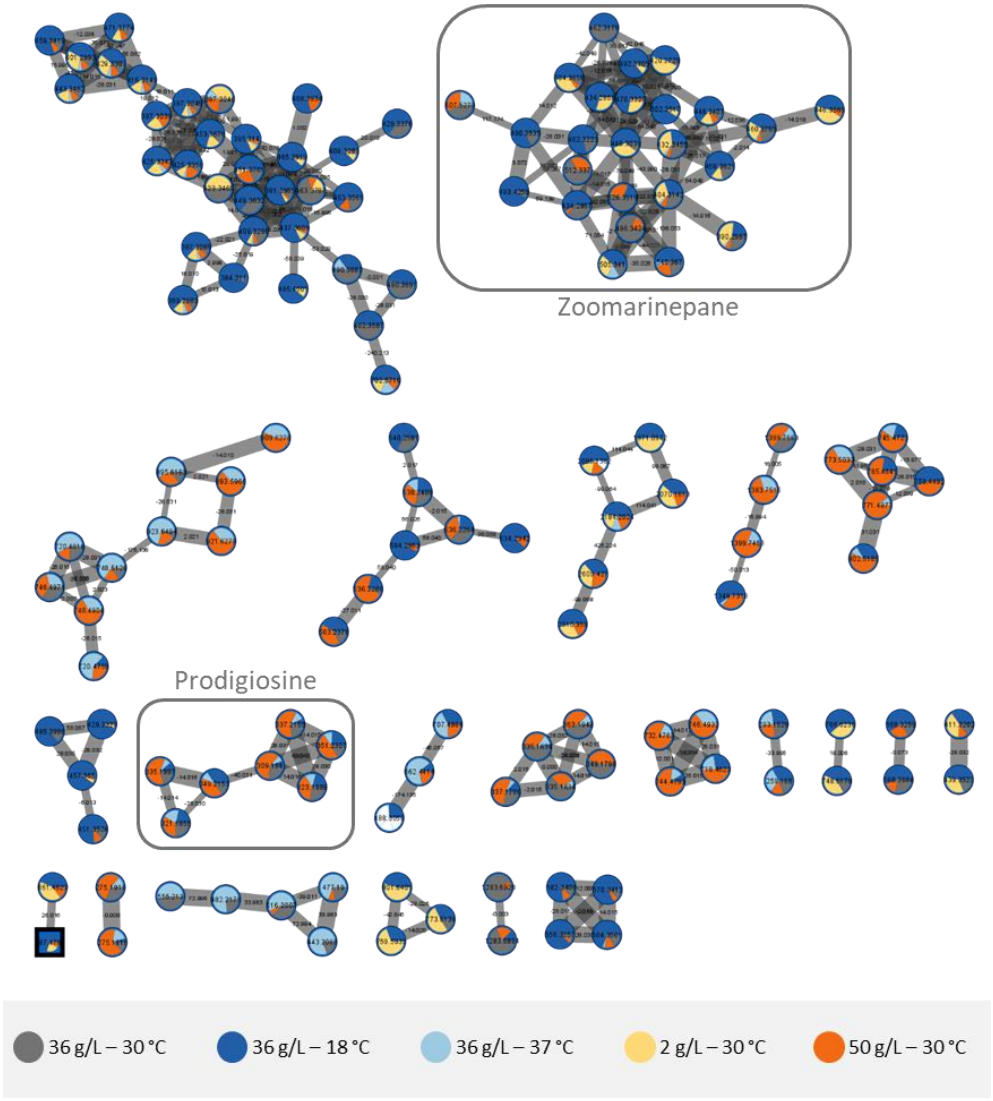


Figure S5. 6. Feature based molecular networks of crude extracts after cultivation of* Zooshikella marina *sp. Uxx12806 at different temperatures and with different sea salt concentrations , generated with GNPS. *The molecular clusters of prodigiosine and zoomarinepane are highlighted in a grey box, respectively. The production level of compounds [area under the curve, AUC] during each condition compared to other conditions is represented by pie charts. The thickness of the edges depend on the cosine score, which represents how closely related two congeners are. Singletons and molecular clusters that contained more than 60% medium blank were removed.*

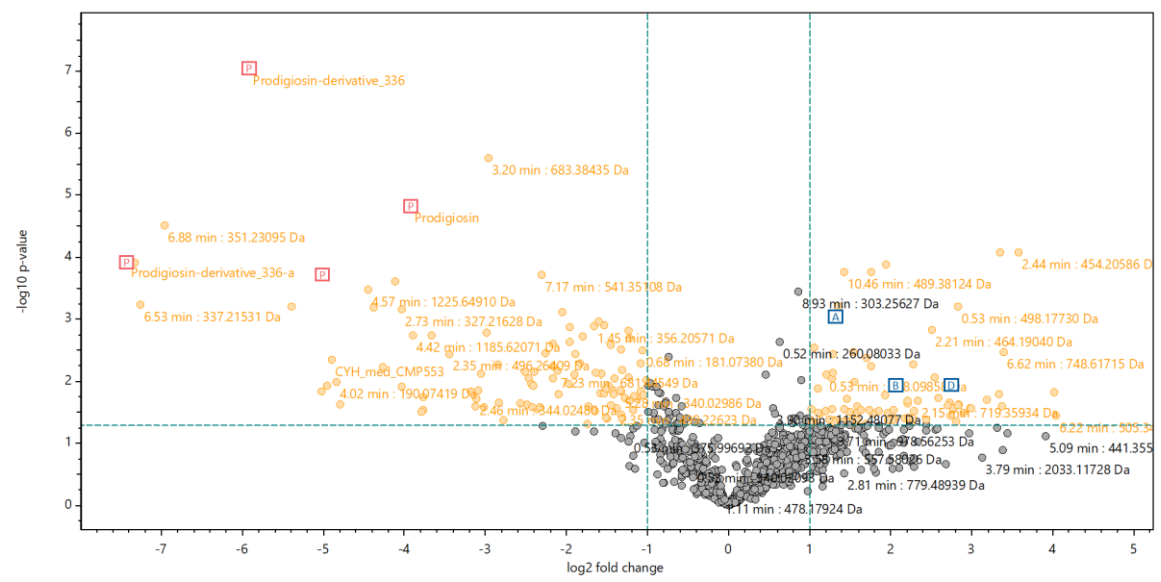


Figure S5. 7. Vulcano plot of significance (-log10 p-value) versus fold change (log2 fold change) of cultivation conditions 2 g/L – 30 °C versus 36 g/L – 30 °C. *Data points appearing towards the top of the plot indicate high significance, data points towards the left side of the plot indicate downregulation, and towards the right side upregulation of features. Highlighted in red: prodigiosin derivatives (P). Highlighted in blue: zoomarinepanes (A, B, and D).*

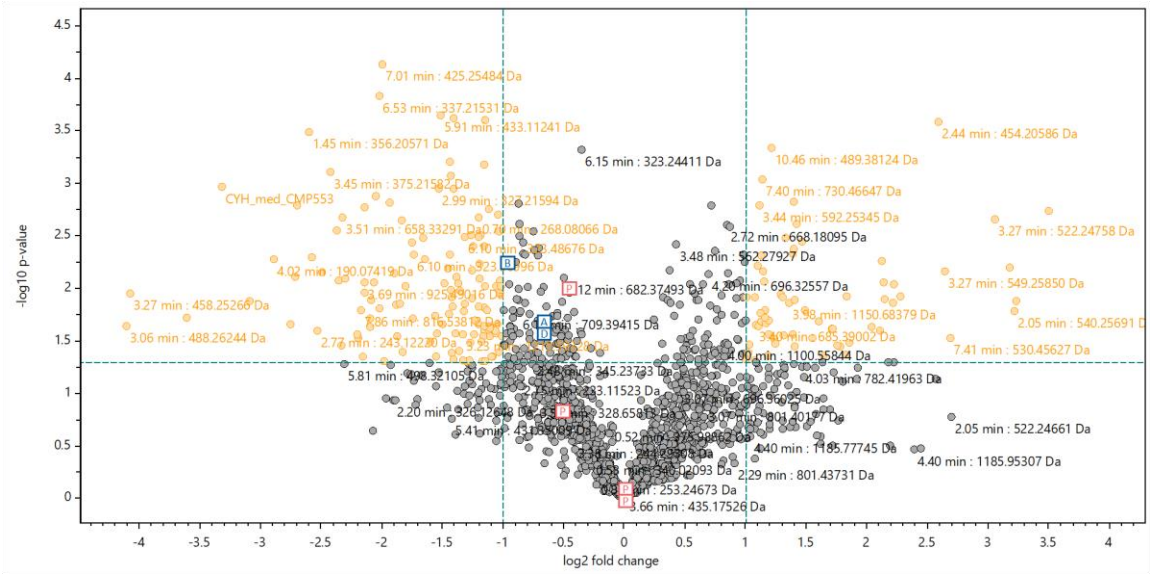


Figure S5. 8. Vulcano plot of significance (-log10 p-value) versus fold change (log2 fold change) of cultivation conditions 50 g/L – 30 °C vs. 36 g/L – 30 °C. *Data points appearing towards the top of the plot indicate high significance, data points towards the left side of the plot indicate downregulation, and towards the right side upregulation of features. Highlighted in red: prodigiosin derivatives (P). Highlighted in blue: zoomarinepanes (A, B, and D).*

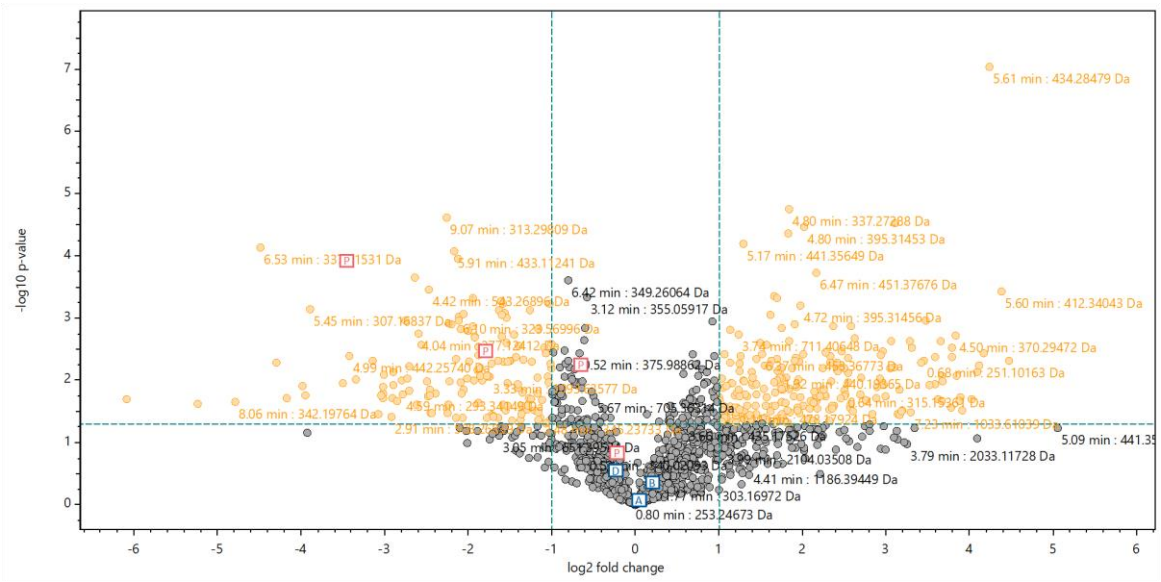


Figure S5. 9. Vulcano plot of significance (-log10 p-value) versus fold change (log2 fold change) of cultivation conditions 36 g/L – 18 °C vs. 36 g/L – 30 °C. *Data points appearing towards the top of the plot indicate high significance, data points towards the left side of the plot indicate downregulation, and towards the right side upregulation of features. Highlighted in red: prodigiosin derivatives (P). Highlighted in blue: zoomarinepanes (A, B, and D).*

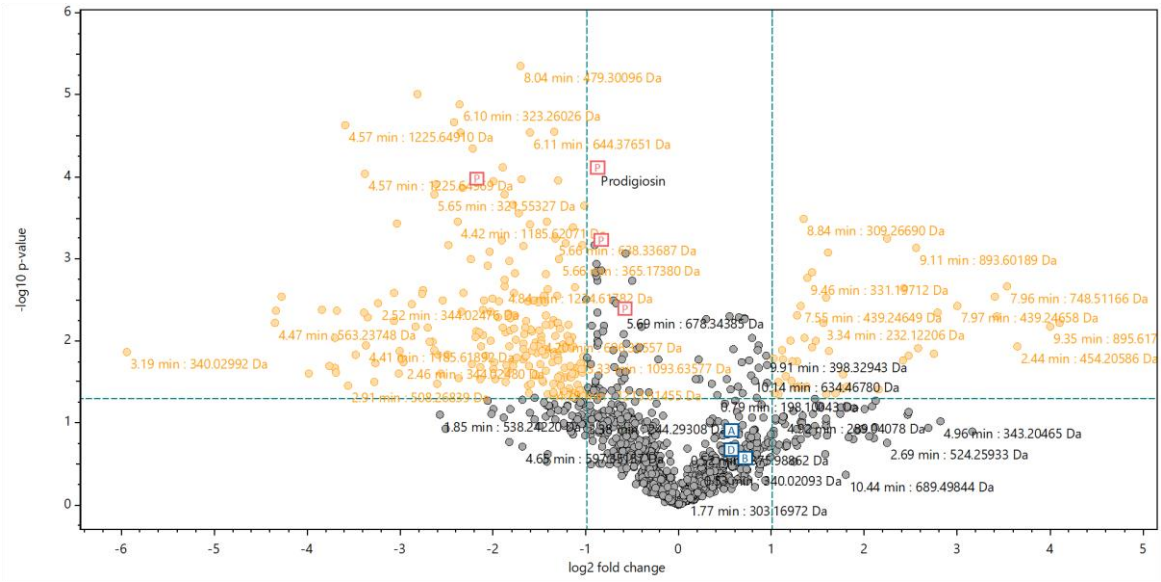


Figure S5. 10. Vulcano plot of significance (-log10 p-value) versus fold change (log2 fold change) of cultivation conditions 36 g/L – 37 °C vs. 36 g/L – 30 °C. *Data points appearing towards the top of the plot indicate high significance, data points towards the left side of the plot indicate downregulation, and towards the right side upregulation of features. Highlighted in red: prodigiosin derivatives (P). Highlighted in blue: zoomarinepanes (A, B, and D).*

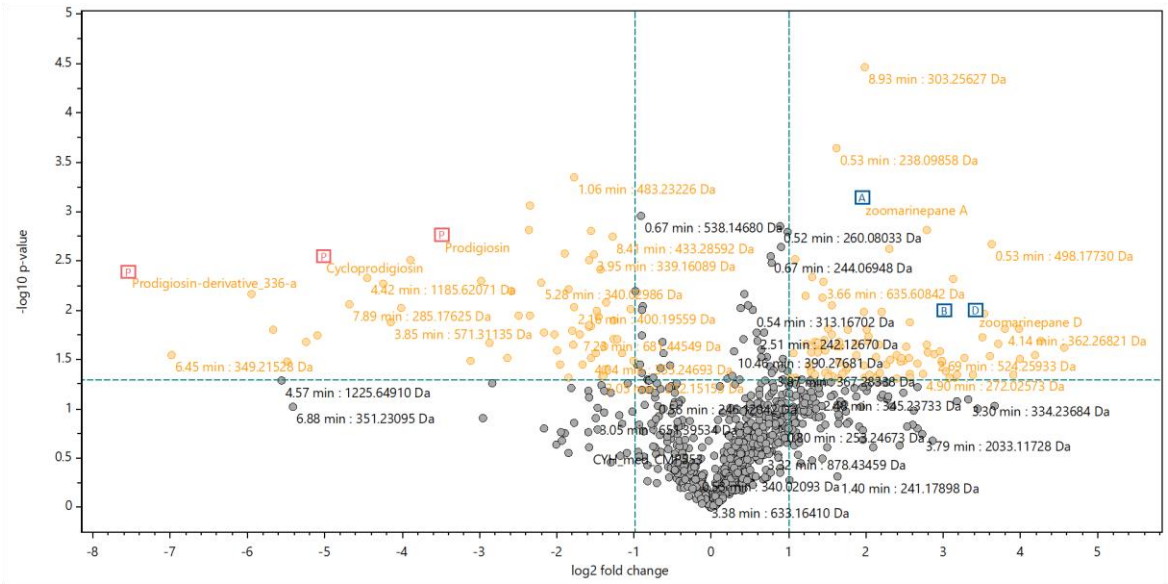


Figure S5. 11. Vulcano plot of significance (-log10 p-value) versus fold change (log2 fold change) of cultivation conditions 2 g/L – 30 °C vs. 50 g/L – 30 °C. *Data points appearing towards the top of the plot indicate high significance, data points towards the left side of the plot indicate downregulation, and towards the right side upregulation of features. Highlighted in red: prodigiosin derivatives (P). Highlighted in blue: zoomarinepanes (A, B, and D).*

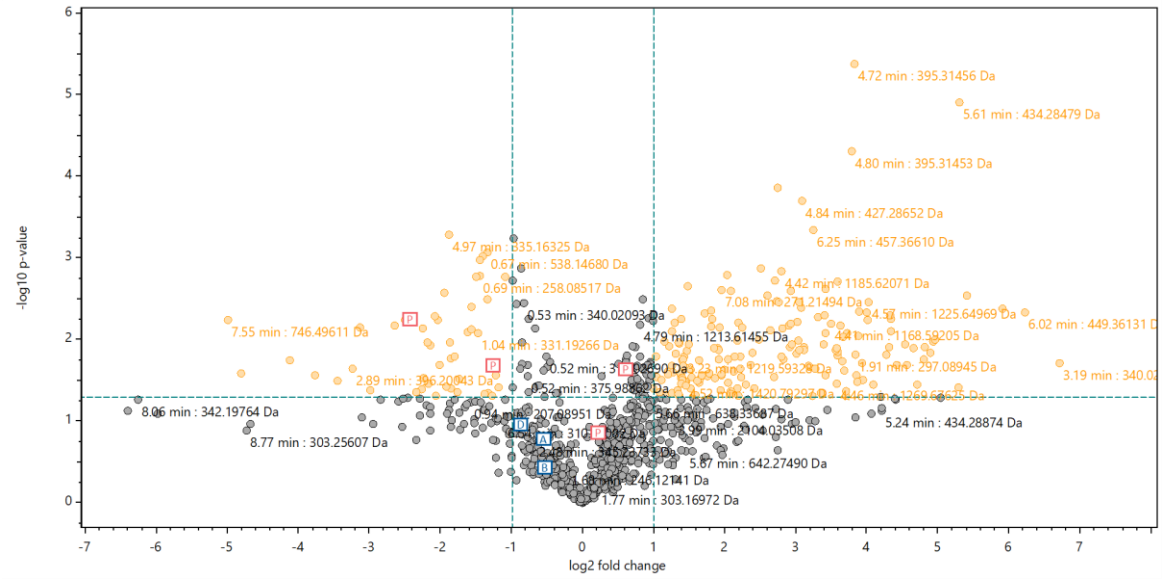


Figure S5. 12. Vulcano plot of significance (-log10 p-value) versus fold change (log2 fold change) of cultivation conditions 36 g/L – 18 °C vs. 36 g/L – 37 °C. *Data points appearing towards the top of the plot indicate high significance, data points towards the left side of the plot indicate downregulation, and towards the right side upregulation of features. Highlighted in red: prodigiosin derivatives (P). Highlighted in blue: zoomarinepanes (A, B, and D).*

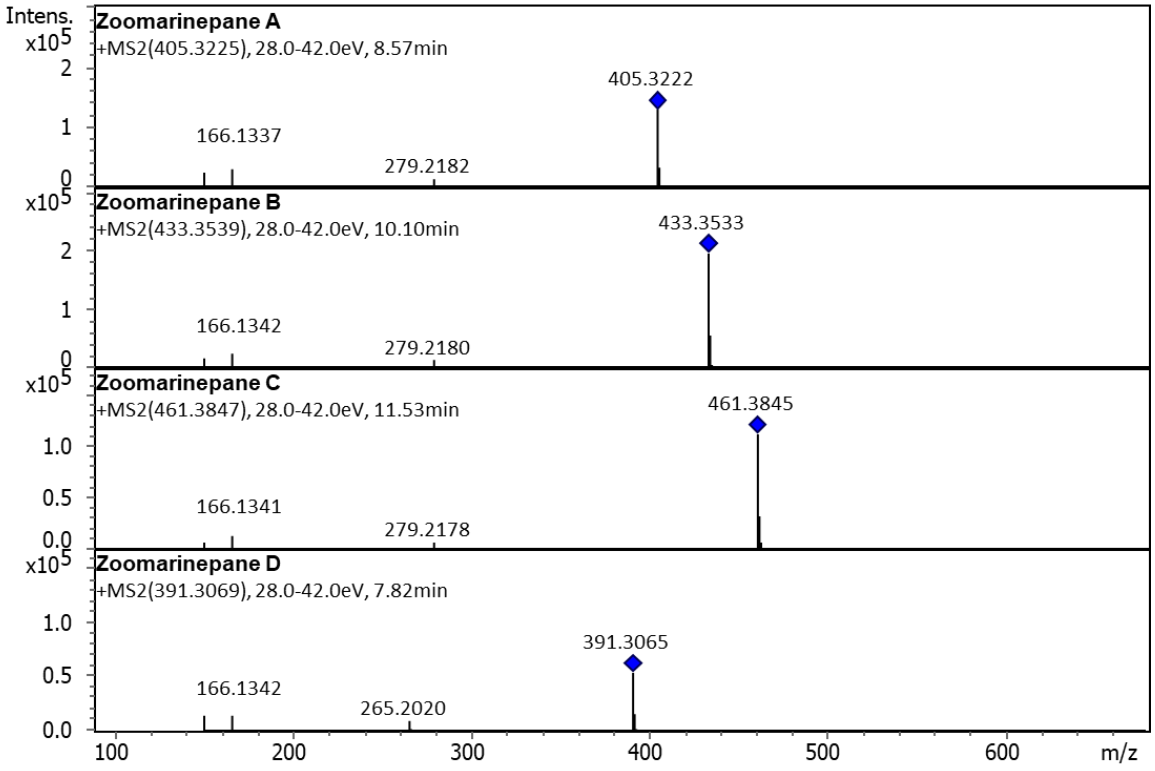
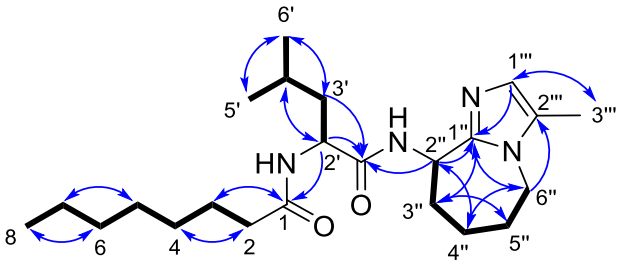


Figure S5. 13. HR-MS² spectra of zoomarinepane A-D. *Characteristical product ions can be observed upon collisionally induced dissociation. The selected precursor ions which* m/z *were selected for fragmentation are marked by a blue rhombus.*

S 5.6 Structure Elucidation

Table S5. 5. *NMR Spectroscopic Data for compound zoomarinepane A[a] (**1**).*

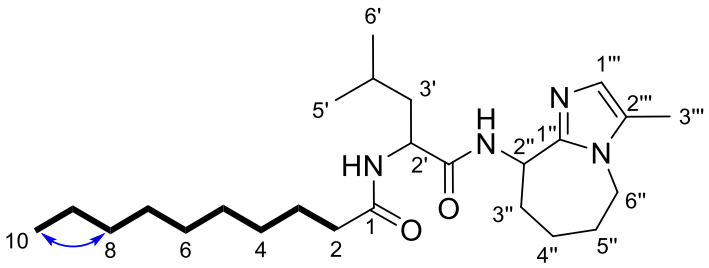


Position	δ_C[b], type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]
1	176.6, C	-	-	-
2 [f]	37.1, CH_2	2.27, t (7.4)	3	1, 3, 4/5
3 [f]	27.2, CH_2	1.62, m	2, 4/5	1, 2, 4/5
4/5 [f,g]	30.3, CH_2	1.32, m	3, 6	3, 6
6 [h]	33.1, CH_2	1.27, m	4/5, 7 [h]	5, 7, 8
7 [h]	23.8, CH_2	1.31, m	6 [h], 8	4/5, 8
8	14.6, CH_3	0.89, t (6.9)	7	6, 7
1'	174.2, C	-	-	-
2' [f]	53.4, CH	4.48, dd (9.7, 5.4)	3'	1, 1', 3', 4'
3' [f]	42.2, CH_2	1.62, m	2', 4'	1', 2', 4'
4' [f,i]	26.2, CH	1.68, m	3', 5', 6'	2', 3', 5', 6'
5'/6' [f,j]	23.6, CH_3	0.97, dd (11.5, 6.4)	4'	3', 4', 5'/6'
	22.0, CH_3	0.92, dd (11.5, 6.4)	4'	3', 4', 5'/6'
1''	149.6, C	-	-	-
2'' [f]	50.6, CH	5.05, br dd (9.7, 0.6)	3''ab	1', 1'', 3'', 4''
3''a [f,i]	33.8, CH_2	1.70, m	2'', 3''b, 4''b	1'', 2'', 5'' [i]
3''b [f,j]		1.90, m	2'', 3''a	1'', 2'', 5''
4''a [f,j]	28.1, CH_2	1.89, m	4''b, 5''a	5'', 6'' [j]
4''b [f]		2.02, m	3''a, 4''a, 5''a	5'', 6''
5''a [f,k]	29.2, CH_2	1.57, m	4''ab, 5''b [k], 6''ab	3'', 6'' [k]
5''b [f,k]		1.88, m	5''a [k], 6''ab	3'', 6''
6''a [f]	45.5, CH_2	3.89, m	5''ab, 6''b	1'', 4'', 5'', 2'''
6''b [f]		4.12, m	5''ab, 6''a	1'', 4'', 5'', 2'''
1''' [f]	123.6, CH	6.60, s	3'''	1'', 2''', 3'''
2'''	130.5, C	-	-	-

3''' [f]	9.6, CH_3	2.19, s	1'''	1''', 2'''

[a]Recorded in methanol-d_4. [b]Acquired at 175 MHz, adjusted to the solvent signal of methanol-d_4 (δ_c 49.15 ppm). [c]Acquired at 700 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 3.31 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Protons showing HMBC correlations to indicated carbons. [f]Signal doubling can be observed due to the equilibrium of two diastereomers caused by the acidic proton at position 2''. [g, h, i, j, k]Overlapping signals.

Table S5. 6. *NMR Spectroscopic Data for compound zoomarinepane B[a] (**2**).*

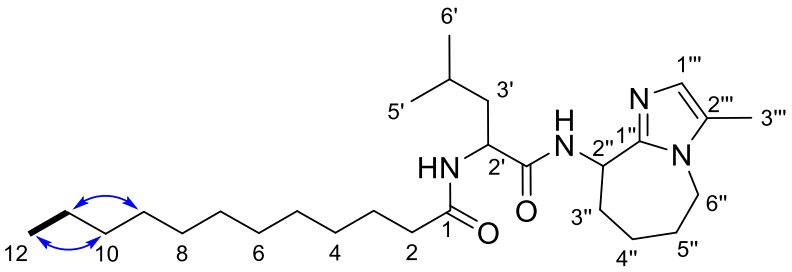


Position	δ_C[b], type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]
1	173.2, C	-	-	-
2	36.9, CH_2	2.23, m	3	1, 3, 4
3	25.8, CH_2	1.64, m	2, 4	1, 2, 5
4	29.5, CH_2	1.31, m	3	2, 3, 6, 7
5	29.4, CH_2	1.28, m	-	6, 7
6, 7 [f]	29.6, CH_2	1.25, m	-	-
8	32.0, CH_2	1.25, m	-	6, 7
9	22.8, CH_2	1.28, m	10	8, 10
10	14.3, CH_3	0.88, t (7.1)	9	8, 9
1'	171.7, C	-	-	-
2'	51.9, CH	4.61, m	3'ab, 7'	1', 3'
3'a	42.4, CH_2	1.55, m	2', 3'b, 4', 5'/6'	1', 2', 4'
3'b	42.4, CH_2	1.70, m	2', 3'a, 4'	1', 2', 4'
4'	25.1, CH	1.66, m	3'ab, 5'/6'	2', 3'
5'/6' [f]	23.2, CH_3	0.96, m	4', 5'/6'	3', 4', 5'/6'
	22.2, CH_3	0.97, m	3'a, 4', 5'/6'	3', 4', 5'/6'
7'	n.a., NH	6.17, br m (8.1)	2'	-
1''	148.6, C	-	-	-
2''	49.7, CH	4.88, m	3''ab, 8''	1', 1'', 3'', 4''
3''a	33.4, CH_2	1.39, m	2'', 3''b	- [g]

3''b		-	2'', 3''a, 4''ab	- [g]
4''a	28.0, CH_2	2.07, m	3''b, 4''b, 5''ab	- [g]
4''b		1.87, m	3''b, 4''a, 5''ab	- [g]
5''a	28.2, CH_2	1.40, m	4''ab, 5''b, 6''ab	4''
5''b		2.00, m	4''ab, 5''a, 6''ab	4''
6''a	44.3, CH_2	4.05, m	5''ab, 6''b	1'', 5'', 2'''
6''b		3.67, m	5''ab, 6''a	1'', 5'', 2'''
7''	n.a., N	-	-	-
8''	n.a., NH	7.73, br d (5.5)	2''	1', 1''
1'''	123.3, CH	6.61, s	3'''	1'', 2'''
2'''	128.5, C	-	-	-
3'''	9.8, CH_3	2.17, m	1'''	1''', 2'''

[a]Recorded in chloroform-d. [b]Acquired at 175 MHz, adjusted to the solvent signal of chloroform-d (δ_c 77.18 ppm). [c]Acquired at 700 MHz, adjusted to the solvent signal of chloroform-d (δ_H 7.27 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Protons showing HMBC correlations to indicated carbons. [f]Overlapping signals. [g]Too weak to be detected. Structural differences to **1** are highlighted in grey.

Table S5. 7. *NMR Spectroscopic Data for compound zoomarinepane C[a] (**3**).*

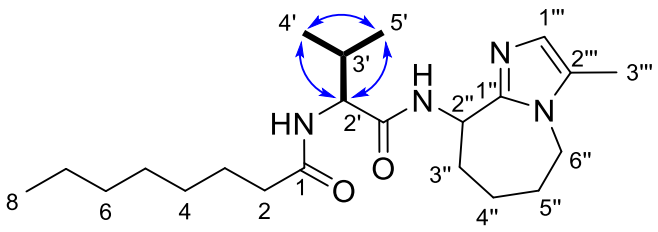


Position	δ_C[b], type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]
1	176.4, C	-	-	-
2	37.1, CH_2	2.26, br dd (13.7, 7.4)	3	1, 3, 4/5/6/7/8 [f]
3	27.0, CH_2	1.62, m	2, 4/5/6/7/8 [f]	1, 2, 4/5/6/7/8 [f]
4/5/6/7/8 [f]	30.5, CH_2	1.32, m	3	3
9 [g]	30.7, CH_2	1.29, m	10 [g]	4/5/6/7/8 [f], 11
10 [g]	32.9, CH_2	1.30, m	9 [g]	11 [g]
11	23.9, CH_2	1.33, m	12	10, 12
12	14.5, CH_3	0.90, m	11	10, 11 [g]
1'	174.0, C	-	-	-

2'	53.4, CH	4.47, m	3'	1, 1', 3'
3'	41.6, CH_2	1.62, m	2', 4'	1', 2', 4', 5'/6'
4'	26.1, CH	1.68, m	3', 5'/6'	3', 5'/6'
5'/6' [f]	23.6, CH_3	0.96, m	4'	3', 4', 5'/6'
	22.1, CH_3	0.94, m	-	5'/6'
1''	149.1, C	-	-	-
2''	50.6, CH	5.03, m	3''ab	1''
3''a	33.7, CH_2	1.92, m	2'', 3''b, 4''b	- [h]
3''b		1.70, m	2'', 3''a	- [h]
4''a	27.9, CH_2	1.88, m	4''b	- [h]
4''b		2.02, m	3''a, 4''a, 5''b	- [h]
5''a	29.3, CH_2	1.61, m	5''b, 6''b	- [h]
5''b		1.87, m	4''b, 5''a, 6''a	- [h]
6''a	45.5, CH_2	4.13, m	5''b, 6''b	1'', 4''
6''b		3.88, m	5''a, 6''a	4''
1'''	123.9, CH	6.59, d (1.0)	3'''	1'', 2'''
2'''	130.2, C	-	-	-
3'''	9.5, CH_3	2.19, m	1'''	1''', 2'''

[a]Recorded in methanol-d_4, the table shows all relevant signals. [b]Acquired at 175 MHz, adjusted to the solvent signal of methanol-d_4 (δ_c 49.15 ppm), due to impurities shifts are deduced from 2D NMR data. [c]Acquired at 700 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 3.31 ppm), due to impurities shifts are deduced from 2D NMR data. [d]Proton showing COSY correlation to indicated protons. [e]Protons showing HMBC correlations to indicated carbons. [f, g]Overlapping signals. [h]Too weak to be detected. Differences to **1** are highlighted in grey.

Table S5. 8. *NMR Spectroscopic Data for compound zoomarinepane D[a] (**4**).*



Position	δ_C[b], type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]
1	176.7, C	-	-	-
2	37.0, CH_2	2.27, m	3	1, 3, 4
3	27.3, CH_2	1.61, m	2, 4, 5	1, 2, 5

4	30.3, CH_2	1.32, m	3, 6	-
5	30.4, CH_2	1.32, m	3, 6	6
6	33.1, CH_2	1.28, m	4, 5	-
7	23.8, CH_2	1.30, m	8	-
8	14.6, CH_3	0.90, m	7	7
1'	173.3, C	-	-	-
2'	60.8, CH	4.17, d (7.9)	3'	1, 1', 4', 5'
3'	31.5, CH	2.09, m	2', 4', 5'	2', 4', 5'
4'[f]	19.9, CH_3	0.94, m	3'	2'[f], 3', 5'
5'[f]	18.9, CH_3	0.94, m	3'	2', 3'[f], 4'[f]
1''	149.4, C	-	-	-
2''	50.7, CH	5.06, br dd (9.4, 1.8)	4''a	1', 1'', 3'', 4''
3''a	33.1, CH_2	1.75, m	3''b	2'', 5''[g]
3''b		1.89, m	3''a, 4''b	2'', 5''
4''a	27.3, CH_2	2.02, m	2'', 4''b	-[h]
4''b		1.85, m	3''b, 4''a	-[h]
5''a	29.4, CH_2	1.66, m	5''b, 6''a	-[h]
5''b		1.79, m	5''a, 6''b	-[h]
6''a	45.5, CH_2	3.92, m	5''a, 6''b	1'', 2'', 4'', 5''
6''b		4.11, m	5''b, 6''a	1'', 2'', 4'', 5''
1'''	124.2, CH	6.56, m	3'''	1'', 2'''
2'''	130.4, C	-	-	-
3'''	9.7, CH_3	2.19, m	1'''	1''', 2'''

[a]Recorded in methanol-d_4, the table shows all relevant signals. [b]Acquired at 175 MHz, adjusted to the solvent signal of methanol-d_4 (δ_C 49.15 ppm), due to impurities shifts are deduced from 2D NMR data. [c]Acquired at 700 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 3.31 ppm), due to impurities shifts are deduced from 2D NMR data. [d]Proton showing COSY correlation to indicated protons. [e]Protons showing HMBC correlations to indicated carbons. [f, g]Overlapping signals. [h]Too weak to be detected. Differences to **1** are highlighted in grey.

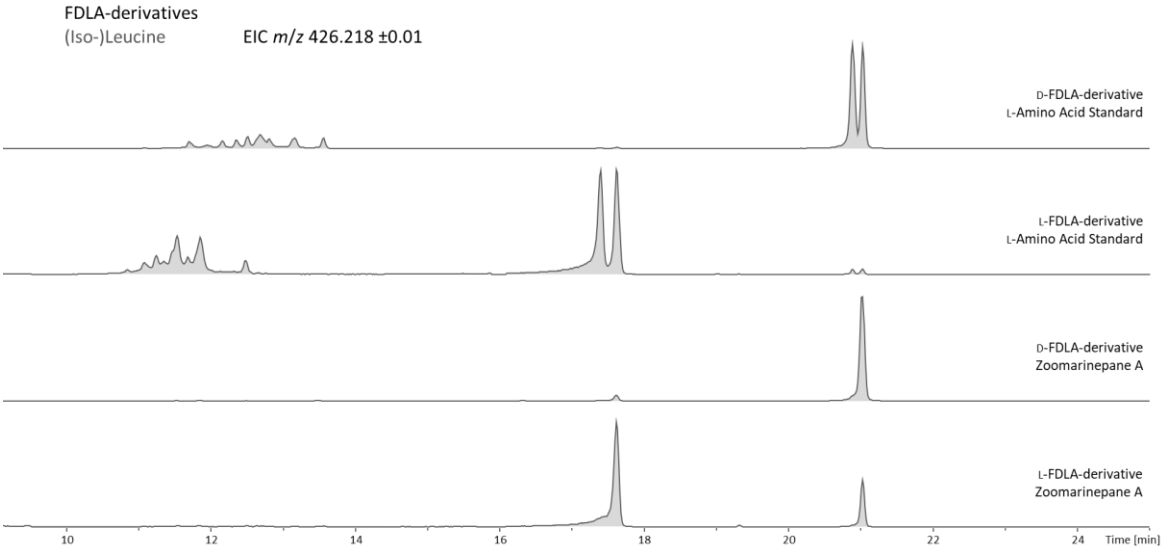


Figure S5. 14. Overlay of the LC-MS chromatograms after Marfey's derivatisation for the characterisation of amino acids using FDLA. *The stereogenic center of leucine is L-configured.*

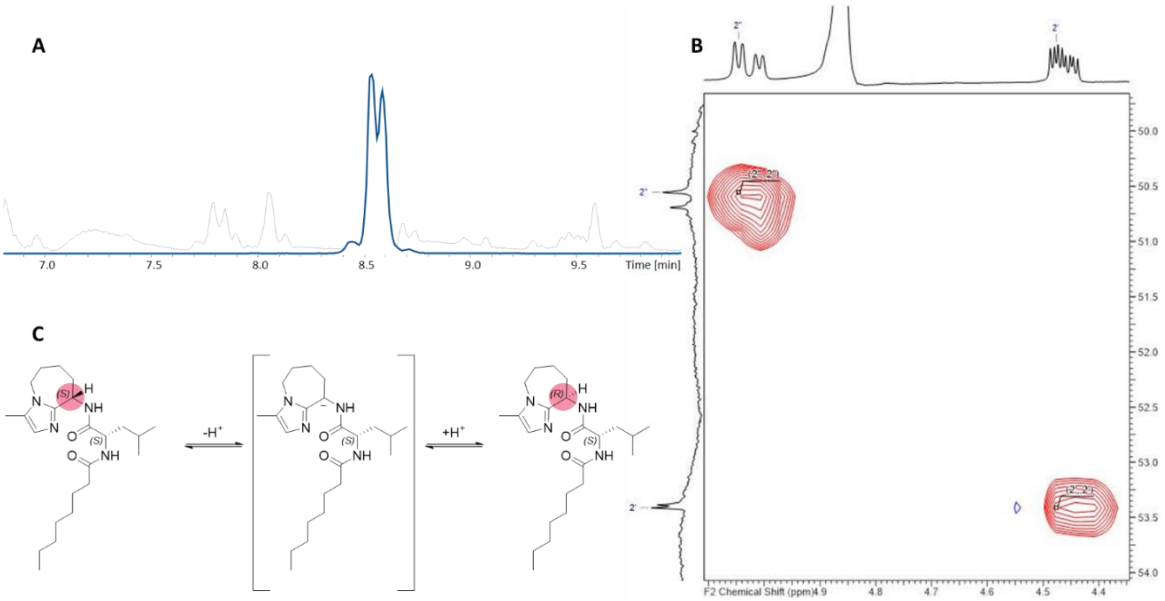


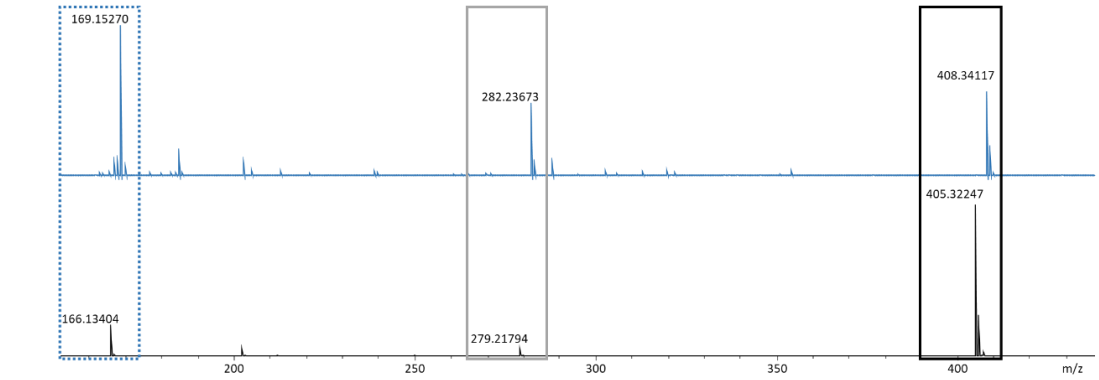
Figure S5. 15. A*: Base peak chromatogram of* Zooshikella marina *sp. Uxx12806 crude extract (grey) and the extracted ion chromatogram of zoomarinepane A (EIC* m/z *405.322 ±0.01, blue), showing a split peak.* ***B****: Detail of the* α*-proton region of the HSCQ spectrum of zoomarinepane A, showing signal doubling in both the* ^{1}H *and the* ^{13}C *dimension.* ***C****: The proposed equilibrium of diastereomers of zoomarinepane A.*

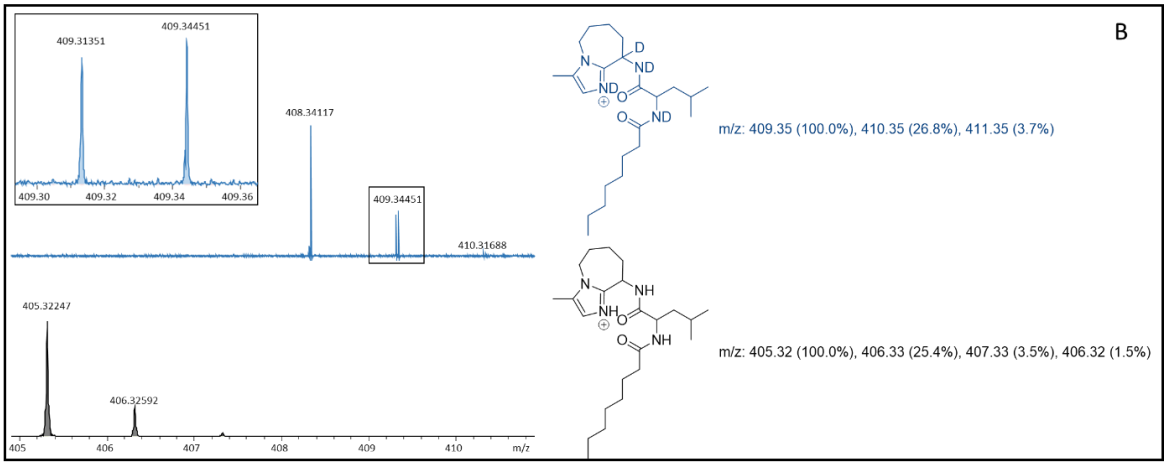
S 5.8 DI-FT-ICR

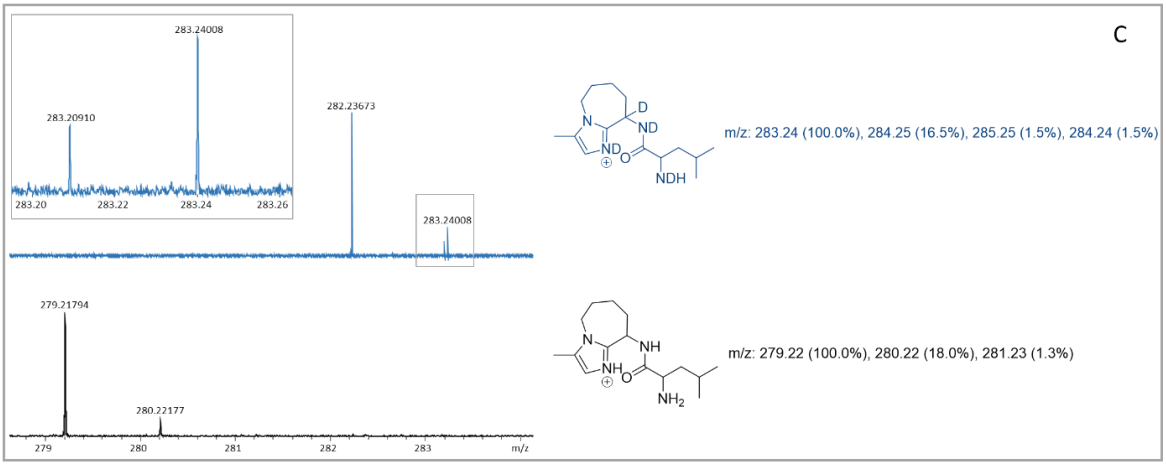
Table S5. 9. *DI-FT-ICR parameters.*

Source	
Capillary voltage	-4500 V
Dry gas	5.0 L/min, 200 °C
Source optics	
Capillary exit	220 V
Deflector plate	200 V
Funnel 1	150 V
Skimmer 1	15 V
Funnel RF Amplitude	100 Vpp
Octopole	
Frequency	5 MHz
RF Amplitude	350 Vpp
Quadrupole	
Q1 Mass	405.00 (409.00) *m/z*

Collision cell		
Collision voltage		-26 V
DC Extract Bias		0.9 V
RF Frequency		2 MHz
Collision RF Amplitude		1500 Vpp
Transfer optics		
Time of Flight		1.000 ms
Frequency		6 MHz
RF Amplitude		350 Vpp
Transfer optics		
Flow		38 %
Quadrupole MS/MS		
Collision Energy		25 V
Q1 Mass		405.00 (409.00) *m/z*
Isolation Window		2.00 *m/z*
Collision RF Amplitude		1500 Vpp
RF Frequency		2 MHz


A
169.15270
282.23673
408.34117
405.32247
166.13404
279.21794
200
250
300
350
400
m/z


B
409.31351
409.34451
408.34117
409.34451
410.31688
409.30
409.32
409.34
409.36
405.32247
406.32592
405
406
407
408
409
410
m/z
m/z: 409.35 (100.0%), 410.35 (26.8%), 411.35 (3.7%)
m/z: 405.32 (100.0%), 406.33 (25.4%), 407.33 (3.5%), 406.32 (1.5%)


C
283.24008
283.20910
282.23673
283.24008
283.20
283.22
283.24
283.26
279.21794
280.22177
279
280
281
282
283
m/z
m/z: 283.24 (100.0%), 284.25 (16.5%), 285.25 (1.5%), 284.24 (1.5%)
m/z: 279.22 (100.0%), 280.22 (18.0%), 281.23 (1.3%)

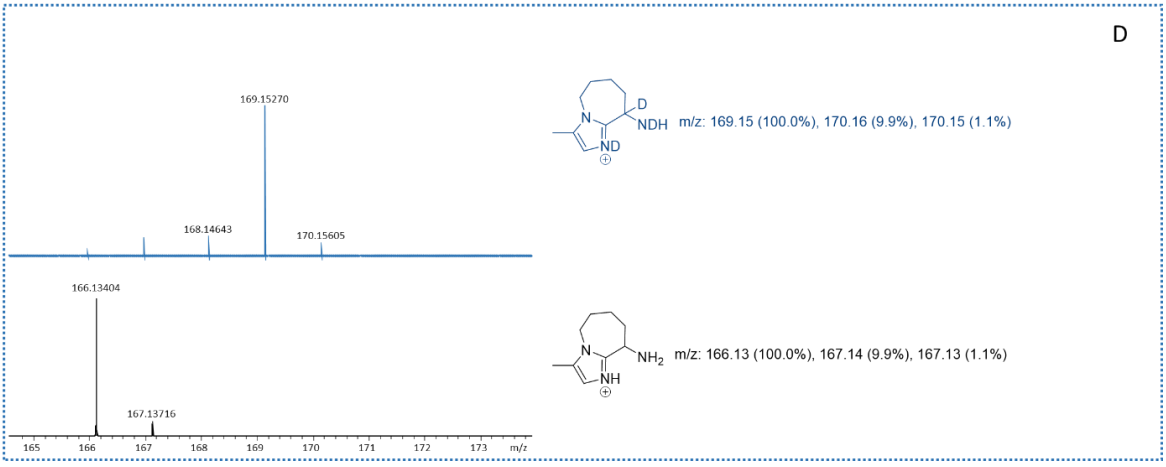

D
169.15270
168.14643
170.15605
166.13404
167.13716
165
166
167
168
169
170
171
172
173
m/z
m/z: 169.15 (100.0%), 170.16 (9.9%), 170.15 (1.1%)
m/z: 166.13 (100.0%), 167.14 (9.9%), 167.13 (1.1%)

Figure S5. 16. FT-ICR-MS^2 spectra of zoomarinepane A showing the exchange of acidic protons with deuterium. *(Blue, top) spectra of **1** in methanol-d$_4$ + 0.1% TFA-d. (Black, bottom) spectra of **1** in methanol. **A**: Overview of the HR-MS^2 spectrum of **1** with key fragments highlighted. **B**: Detailed view of the parent mass of **1** (m/z 405) and the 4x deuterated derivative (m/z 409) with the respective molecular structures. **B**: Detailed view of the key fragment leucine - imidazo[1,2-a]azepin amine (m/z 279) and the 4x deuterated derivative (m/z 283) with the respective molecular structures. **B**: Detailed view of the key fragment imidazo[1,2-a]azepin amine (m/z 166) and the 3x deuterated derivative (m/z 169) with the respective molecular structures.*

S 5.9 Biosynthesis Hypothesis

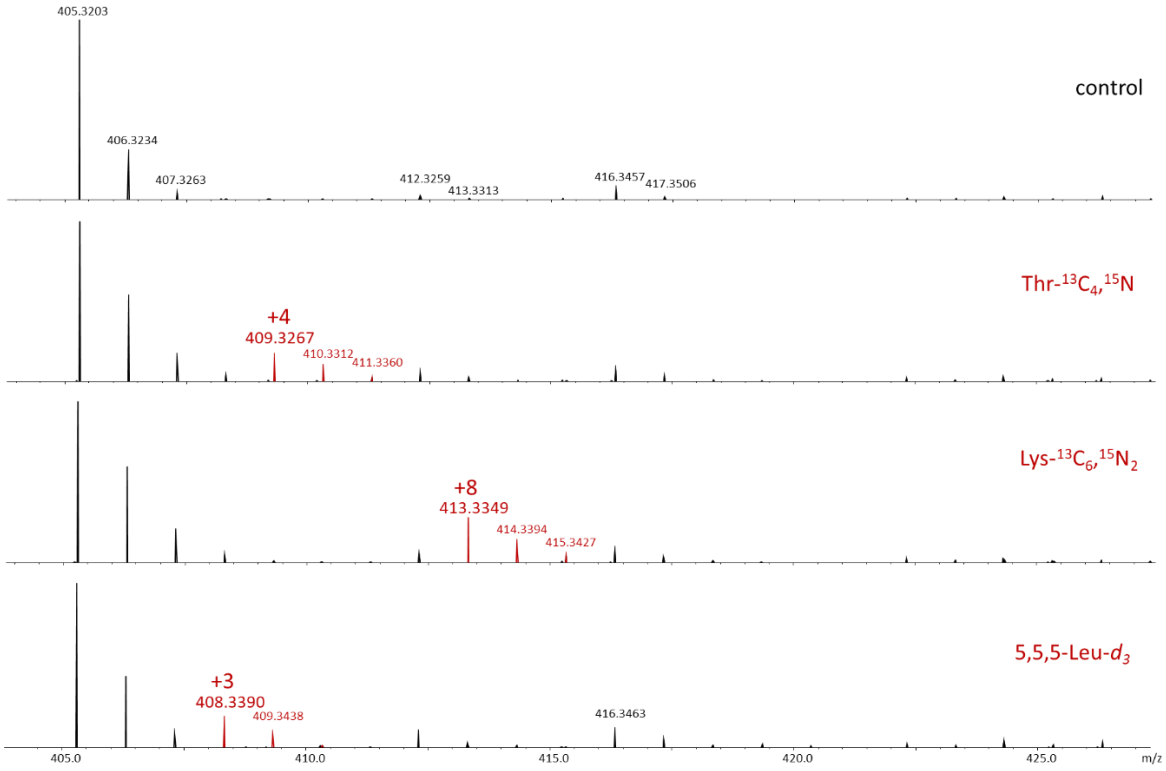


Figure S5. 17. *The exact mass of zoomarinepane A (**1**) as [M+H]+ ion aquired in ESI positive mode after feeding of L-threonine-$^{13}C_4$,^{15}N, L-lysine dihydrochloride-$^{13}C_6$,$^{15}N_2$, and L-leucine-d$_3$; the exact monoisotopic mass of the unlabeled derivative is depicted in black, exact masses of the labelled zoomarinepane derivatives are depicted in red.*

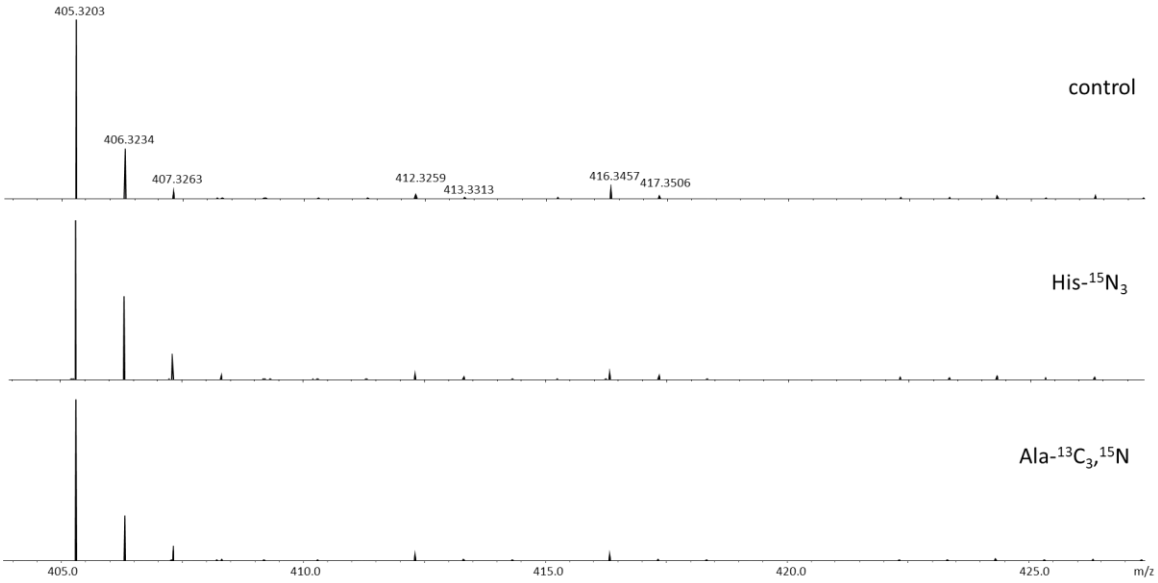


Figure S5. 18. *The exact mass of zoomarinepane A (**1**) as [M+H]+ ion acquired in ESI positive mode after feeding of L-histidine-$^{15}N_3$, and L-alanine-$^{13}C_3$,^{15}N; the exact monoisotopic masses are depicted in black.*

Table S5. 10. *Predicted Proteins Encoded in the putative zoomarinepane BGC of* Zooshikella marina *sp. Uxx12806. CDS were translated into gene sequences and annotated using the BLASTp algorithm[8] against the Non-redundant protein sequences (nr) database at NCBI.*

Protein	Size aa*	Proposed function	Sequence similarity [to source]	Similarity	Accession number of the similar protein
1 [a]	266	biosynthetic-additional	SDR family NAD(P)-dependent oxidoreductase [Zooshikella ganghwensis]	99.25%	WP_212717281.1
2 [a]	191	other	GTP cyclohydrolase I FolE [Zooshikella ganghwensis]	100.00%	WP_038291287.1
3 [a]	240	biosynthetic-additional	dihydromonapterin reductase [Zooshikella ganghwensis]	100.00%	WP_027708467.1
4 [a]	139	other	D-ribose pyranase [Zooshikella ganghwensis]	99.28%	WP_212717279.1

5 [a]	503	transport	ribose ABC transporter ATP-binding protein RbsA [Zooshikella ganghwensis]	99.60%	WP_212717277.1
6 [a]	329	transport	ribose ABC transporter permease [Zooshikella ganghwensis]	100.00%	WP_027708464.1
7 [a]	303	transport	ribose ABC transporter substrate-binding protein RbsB [Zooshikella ganghwensis]	100.00%	WP_212717275.1
8 [a]	305	biosynthetic-additional	ribokinase [Zooshikella ganghwensis]	99.34%	WP_212717273.1
9 [a]	334	regulatory	substrate-binding domain-containing protein [Zooshikella ganghwensis]	100.00%	WP_212717270.1
Orf1	503	biosynthetic-additional	FAD-dependent monooxygenase [Zooshikella ganghwensis]	98.81%	WP_212717269.1
Orf2	205	biosynthetic-additional	class I SAM-dependent methyltransferase [Zooshikella ganghwensis]	99.51%	WP_212717267.1
Orf3	498	biosynthetic-additional	cytochrome P450 [Zooshikella ganghwensis]	99.80%	WP_212717265.1
Orf4	41	other	no significant similarity found		
Orf5	219	biosynthetic-additional	class I SAM-dependent methyltransferase [Zooshikella ganghwensis]	99.09%	WP_212717263.1
Orf6	359	other	hypothetical protein [Zooshikella ganghwensis]	100.00%	WP_212717261.1
Orf7	397	other	hypothetical protein [Zooshikella ganghwensis]	98.99%	WP_212717259.1
Orf8	405	other	hypothetical protein [Zooshikella ganghwensis]	99.75%	WP_212717257.1

ZmpA	1845	biosynthetic	non-ribosomal peptide synthetase [Zooshikella ganghwensis]	99.46%	WP_212717255.1
ZmpB	1128	biosynthetic	non-ribosomal peptide synthetase [Zooshikella ganghwensis]	99.82%	WP_212717253.1
ZmpC	1502	biosynthetic	non-ribosomal peptide synthetase [Zooshikella ganghwensis]	99.47%	WP_212717251.1
Orf9	67	biosynthetic-additional	MbtH family NRPS accessory protein [Zooshikella ganghwensis]	100.00%	WP_212717249.1
Orf10	364	biosynthetic-additional	methyltransferase [Zooshikella ganghwensis]	99.72%	WP_212717556.1
Orf11	813	other	M20/M25/M40 family metallo-hydrolase [Zooshikella ganghwensis]	98.40%	WP_212717239.1
Orf12	177	other	cytochrome b/b6 domain-containing protein [Zooshikella ganghwensis]	99.44%	WP_212717237.1
10 [a]	203	transport	LysE family translocator [Zooshikella ganghwensis]	99.51%	WP_212717235.1
11 [a]	218	regulatory	GntR family transcriptional regulator [Zooshikella ganghwensis]	100.00%	WP_027708459.1
12 [a]	534	other	BCCT family transporter [Zooshikella ganghwensis]	99.81%	WP_051311458.1
13 [a]	374	biosynthetic-additional	mandelate racemase/muconate lactonizing enzyme family protein [Zooshikella ganghwensis]	99.47%	WP_212717233.1
14 [a]	391	other	UxaA family hydrolase [Zooshikella ganghwensis]	99.23%	WP_094786423.1

15 [a]	112	other	UxaA family hydrolase [Zooshikella ganghwensis]	100.00%	WP_027708456.1
16 [a]	245	other	hypothetical protein [Zooshikella ganghwensis]	97.55%	WP_212717226.1
17 [a]	178	other	acyloxyacyl hydrolase [Zooshikella ganghwensis]	100.00%	WP_212717225.1
18 [a]	248	transport	ATP-binding cassette domain-containing protein [Zooshikella ganghwensis]	99.60%	MBU2705636.1
19 [a]	493	transport	glycine betaine ABC transporter substrate-binding protein [Zooshikella ganghwensis]	99.19%	WP_212717221.1
20 [a]	158	other	DUF2569 domain-containing protein [Zooshikella sp. WH53]	87.34%	WP_215821123.1
21 [a]	483	other	phosphatidylserine decarboxylase [Zooshikella ganghwensis]	98.76%	WP_027709052.1
22 [a]	152	other	GNAT family N-acetyltransferase [Zooshikella ganghwensis]	99.34%	WP_212717218.1

*aa: amino acids. The core BGC is highlighted in a grey. Cluster borders (marked with a black line) were determined based on the genetic architecture of the putative *zmp* BGC. [a]genes outside cluster border.

Table S5. 11. *Putative Conserved Domains of the Core Zoomarinepane BGC of* Zooshikella marina *sp. Uxx12806.*

Protein	domain	Proposed function [1]	Substrate (consensus)
ZmpA	AL	Fatty-acid-AMP ligase	Fatty acids
	ACP	Phosphopantetheine attachment site	
	C	Condendsation domain	L-amino acids [2]
	A	Adenylation domain	Alanine (94%) [3]
	PCP	Phosphopantetheine attachment site	
ZmpB	C	Condendsation domain	L-amino acids [2]
	A	Adenylation domain	Glycine (91%) [3]
	PCP	Phosphopantetheine attachment site	
ZmpC	C	Condendsation domain	L-amino acids [2]
	A	Adenylation domain	ohdmTyr (76%) [3]
	PCP	Phosphopantetheine attachment site	
	TR	Thioester reductase, SDR family oxidoreductase	

[1]Analysis with the CCD Conserved Domain Database[9]. [2]Analysis with the Natural Product Domain Seeker[10], catalyse the formation of a peptide bond beteen two L-amino acids. [3]Analysis with Stachelhaus code[11].

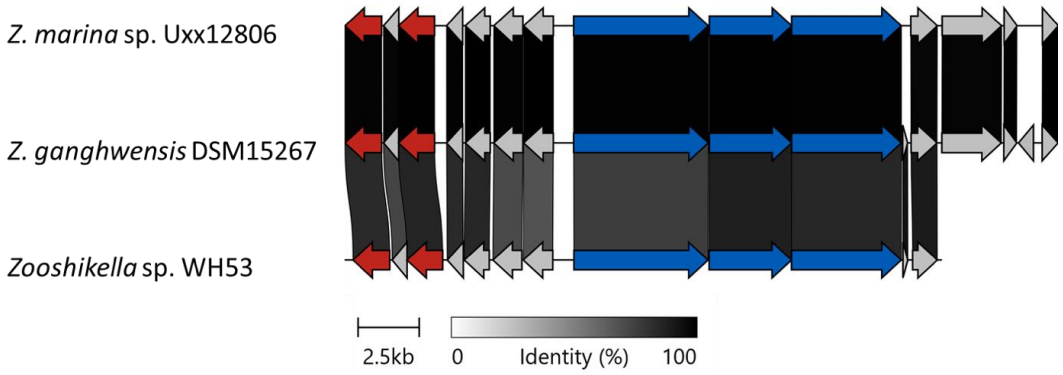


Figure S5. 19. Comparison of the putative* zmp *BGC in* Z. marina *sp. Uxx12806 with the BGCs found in* Z. ganghwensis *DSM15267 and* Zooshikella *sp. WH53. *The similarity of the BGCs was identified using the BLASTp algorithm[8] against the Non-redundant protein sequences (nr) database at NCBI. The published genome data were downloaded from NCBI. The* zmp *BGC was manually annotated and compared using the Clinker algorithm[12]. Core BG are marked in blue (right), additional BG with possibly oxidative activity are marked in red (left), additional BG without assigned functions for the zoomarinepane biosynthesis are marked in grey. Genbank assembly accession number* Zooshikella *sp. WH53: GCA_018837975.1;* Z. ganghwensis *DSM15267: GCA_000428585.1.*

S 5.10 Bioactivity Data

Table S5. 12. *Bioactivity data of zoomarinepane.*

Test organism	Zoomarinepane A	Zoomarinepane B
Staphylococcus aureus Newman	> 64 µg/mL	> 64 µg/mL
Mycobacterium smegmatis mc^2 155	> 64 µg/mL	64 µg/mL
Bacillus subtilis DSM-10	> 64 µg/mL	> 64 µg/mL
Citrobacter freundii DSM-30039	> 64 µg/mL	> 64 µg/mL
Acinetobacter baumannii DSM-30008	> 64 µg/mL	> 64 µg/mL
Candida albicans DSM-1665	> 64 µg/mL	> 64 µg/mL
Escherichia coli JW0451-2 (ΔacrB)	> 64 µg/mL	64 µg/mL
Escherichia coli BW25113 (WT)	> 64 µg/mL	> 64 µg/mL
Cryptococcus neoformans DSM-11959	> 64 µg/mL	> 64 µg/mL
Pseudomonas aeruginosa PA14	> 64 µg/mL	> 64 µg/mL
Pichia anomala	> 64 µg/mL	> 64 µg/mL
Mucor hiemalis	> 64 µg/mL	> 64 µg/mL
CHO-K1 (IC_{50})	29.33 µg/mL	> 37 µg/mL

S 5.11 NMR Spectra

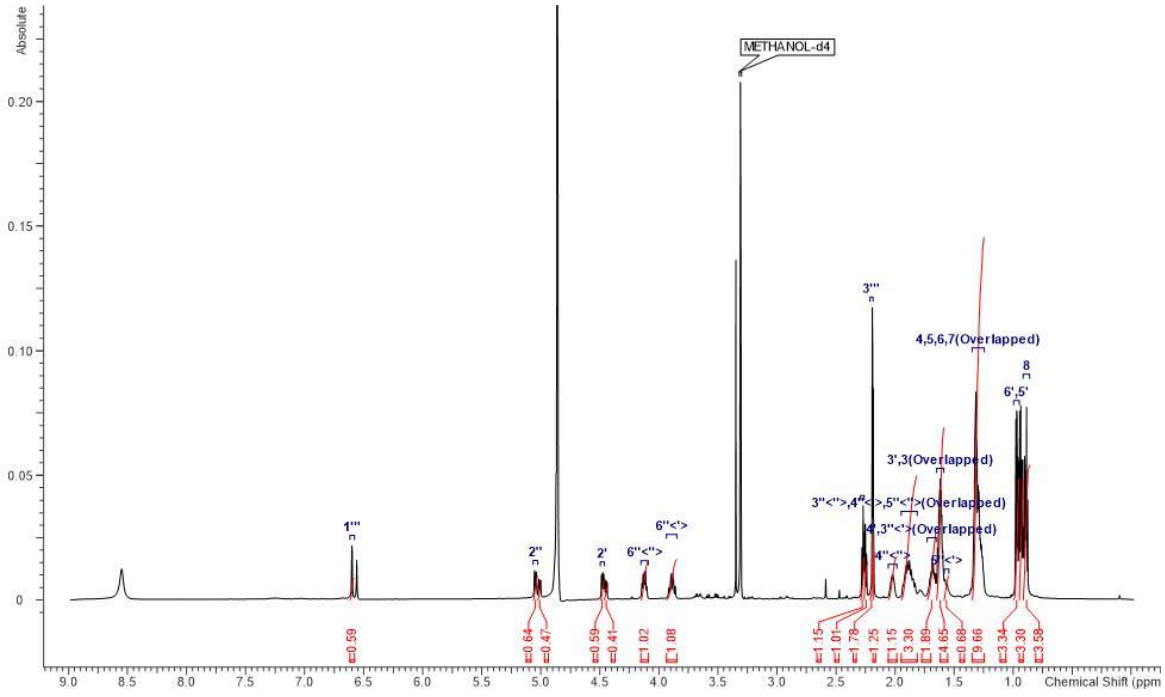


Figure S5. 20. *^{1}H NMR of zoomarinepane A (**1**) recorded in methanol-*d_4 *at 700 MHz.*

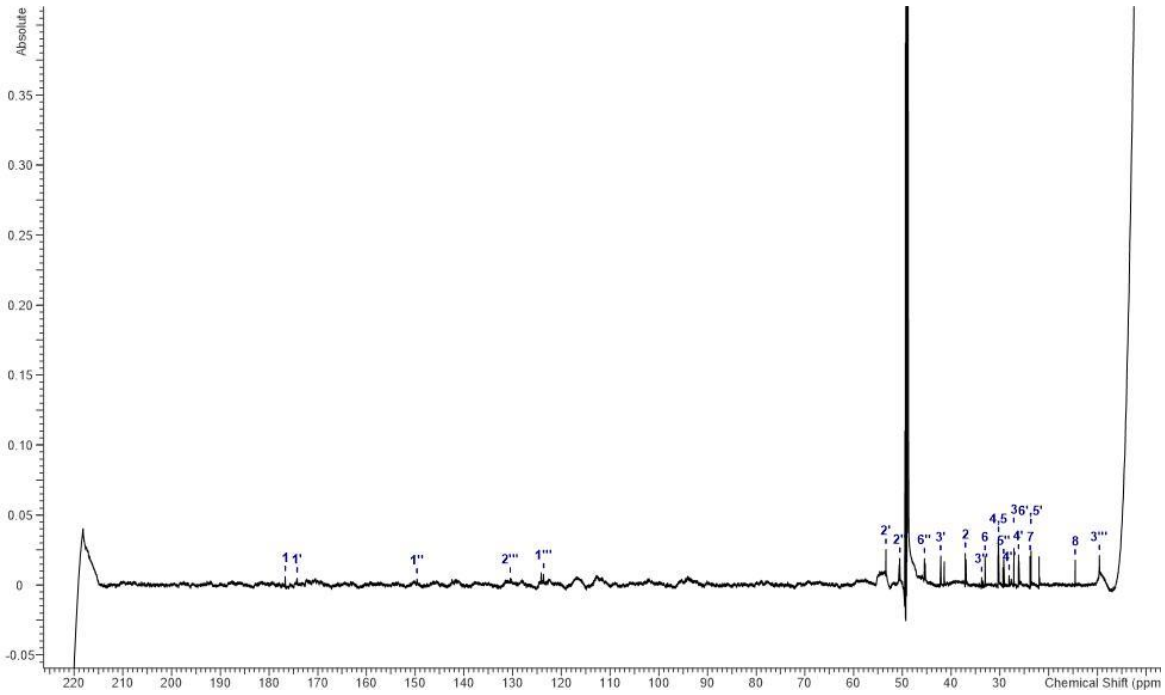


Figure S5. 21. *^{13}C NMR of zoomarinepane A (**1**) recorded in methanol-*d_4 *at 175 MHz.*

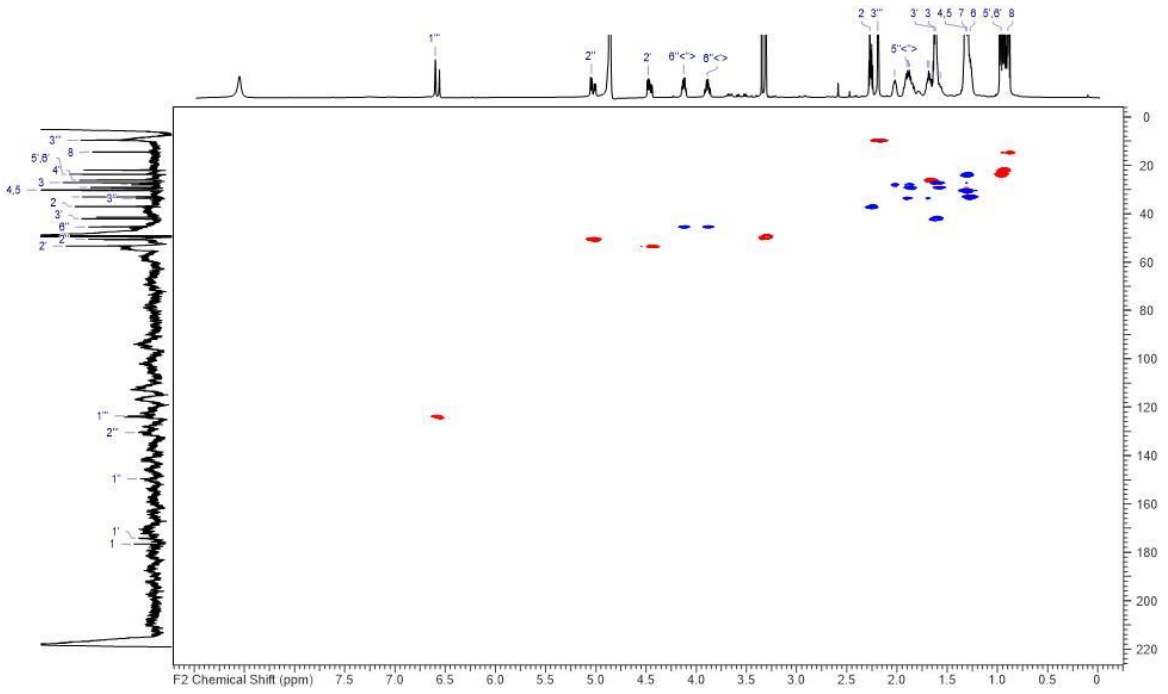


Figure S5. 22. *HSQC NMR of zoomarinepane A (**1**) recorded in methanol-d₄.*

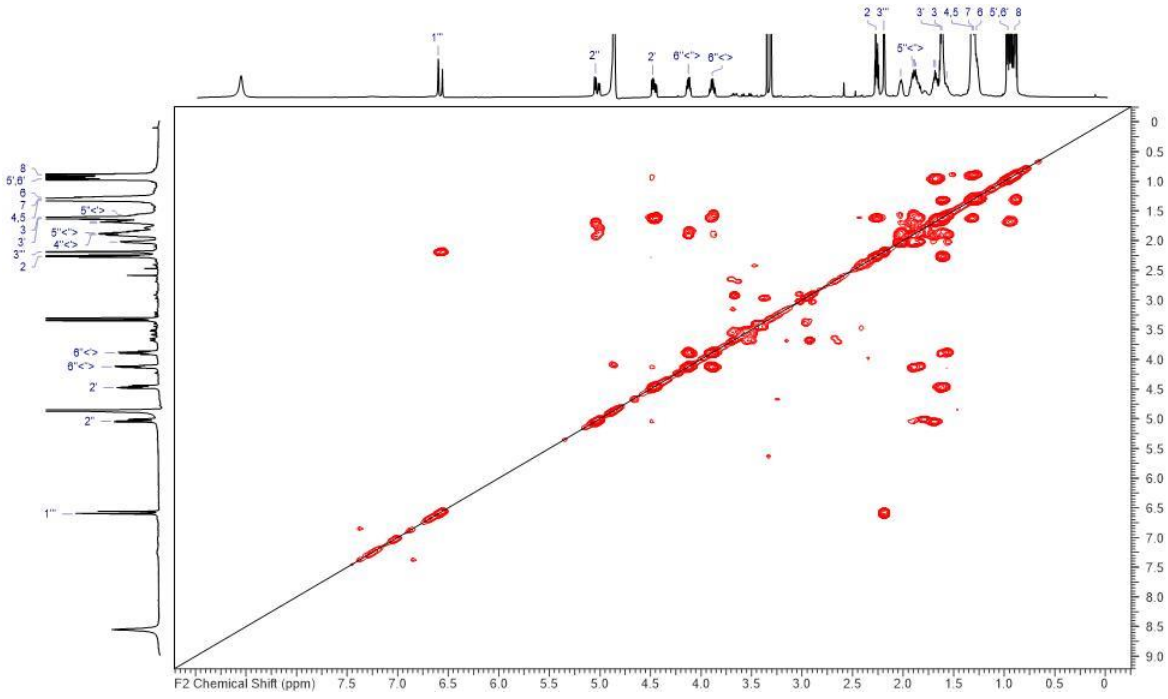


Figure S5. 23. *COSY NMR of zoomarinepane A (**1**) recorded in methanol-d₄.*

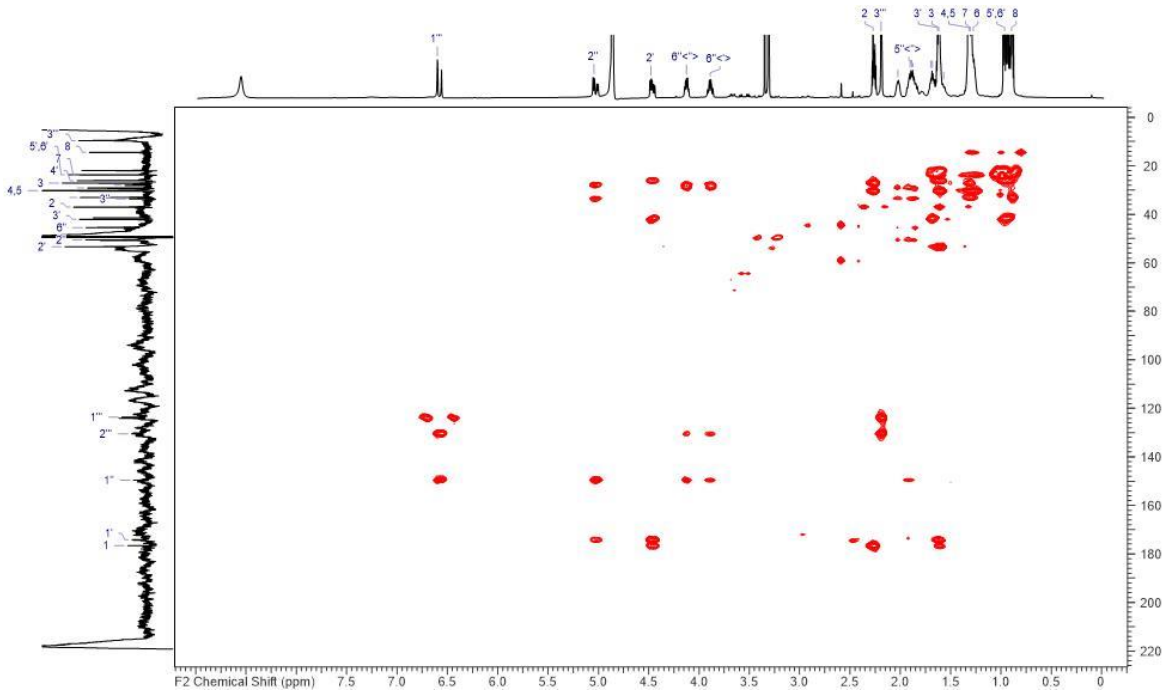


Figure S5. 24. *HMBC NMR of zoomarinepane A (**1**) recorded in methanol-d$_4$.*

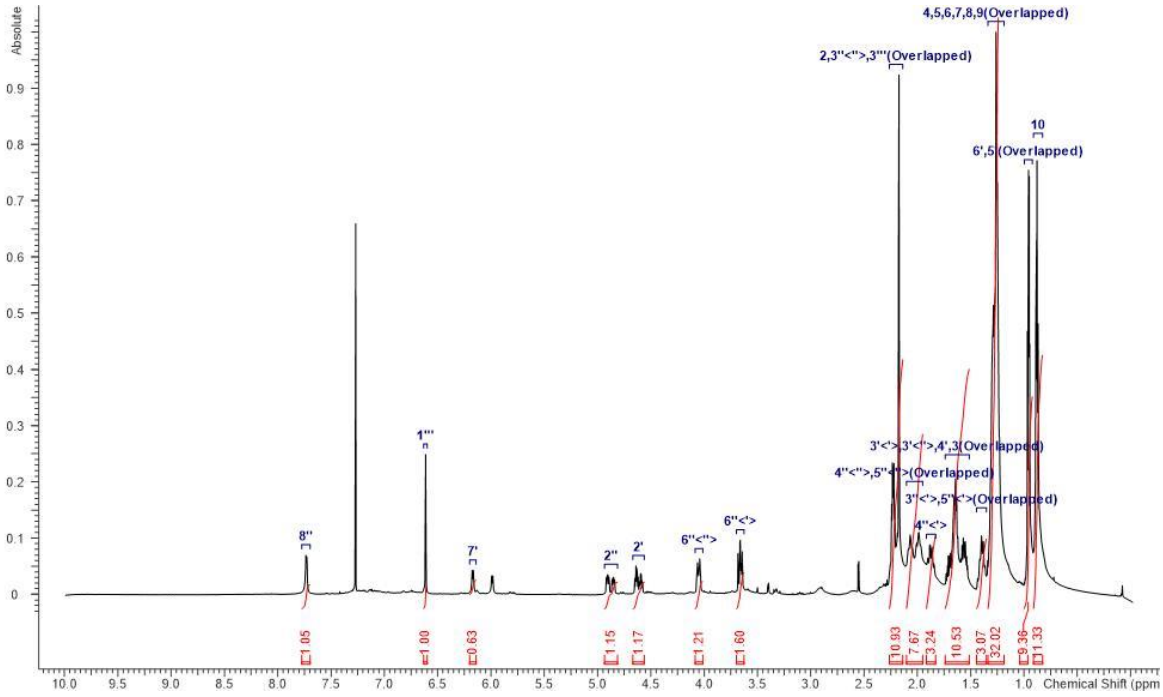


Figure S5. 25. *^{1}H NMR of zoomarinepane B (**2**) recorded in chloroform-d at 700 MHz.*

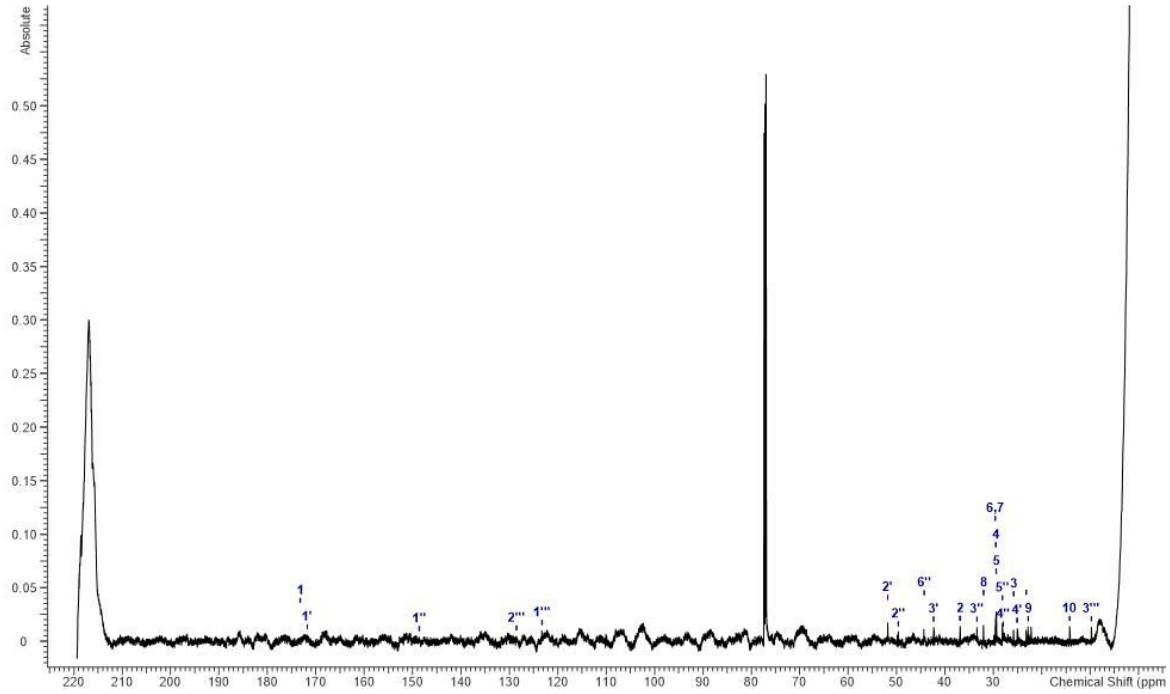


Figure S5. 26. *^{13}C NMR of zoomarinepane B (**2**) recorded in chloroform-d at 175 MHz.*

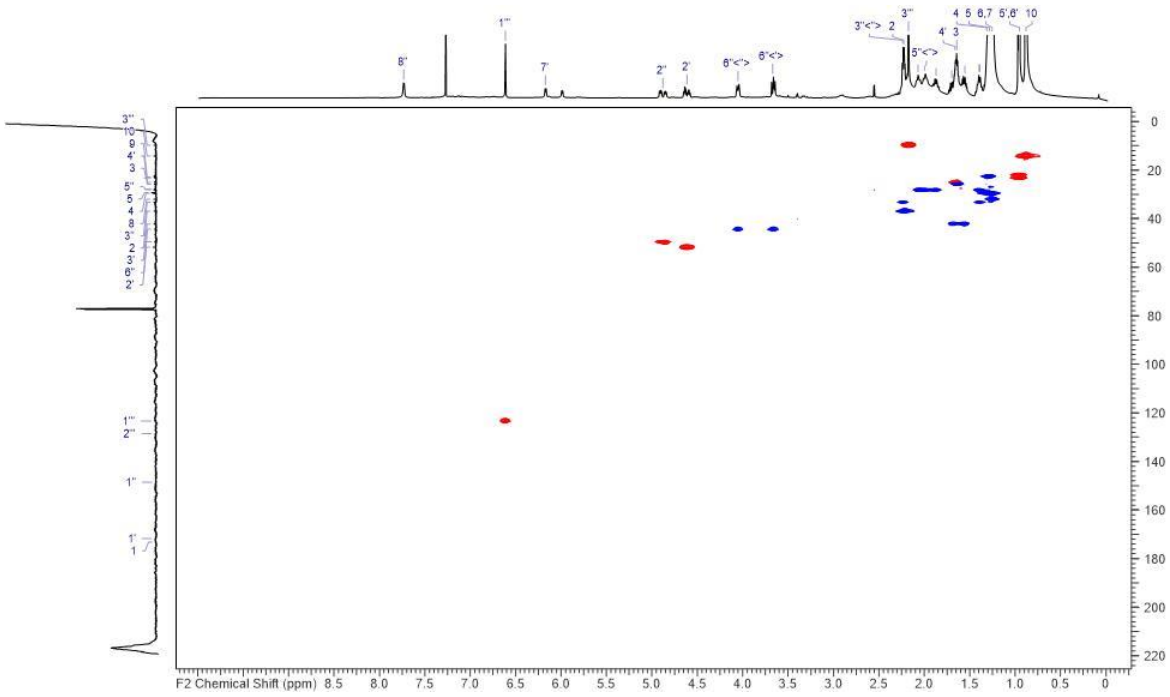


Figure S5. 27. *HSQC NMR of zoomarinepane B (**2**) recorded in chloroform-d.*

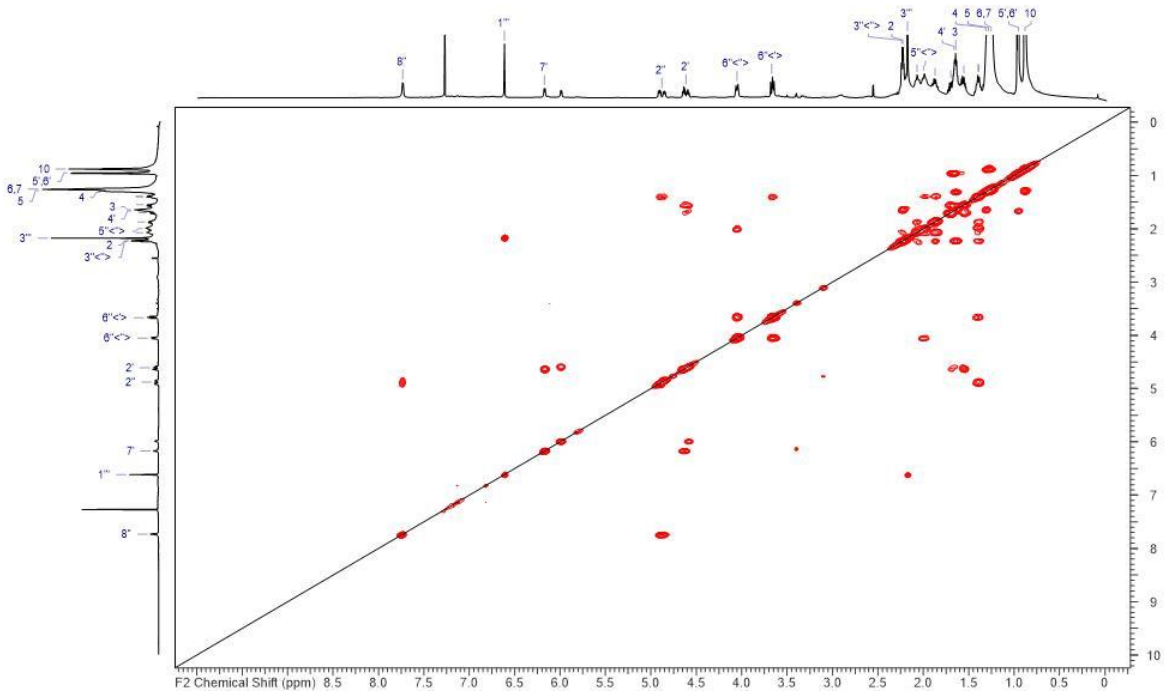


Figure S5. 28. *COSY NMR of zoomarinepane B (**2**) recorded in chloroform-d.*

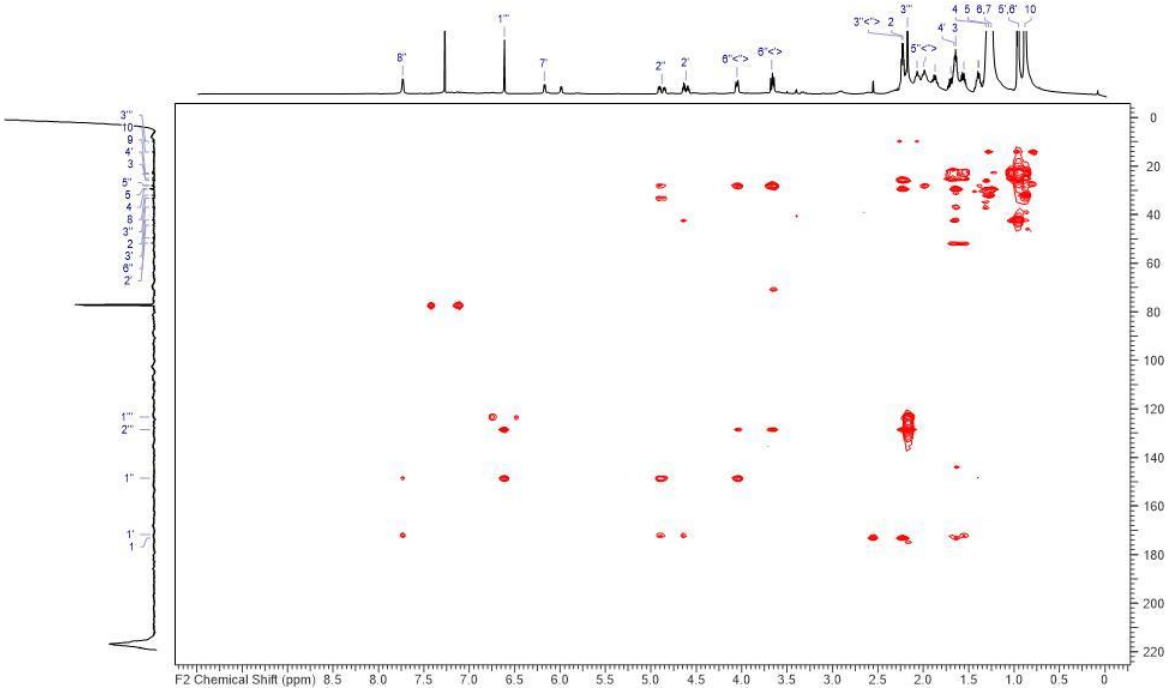


Figure S5. 29. *HMBC NMR of zoomarinepane B (**2**) recorded in chloroform-d.*

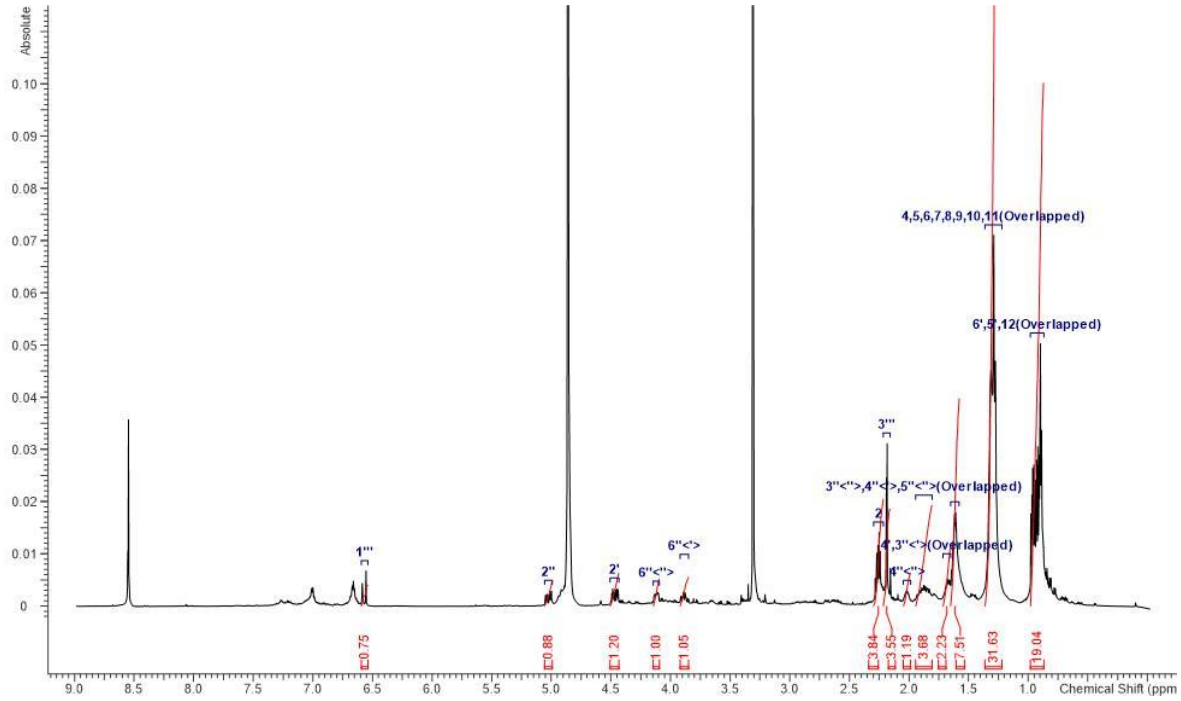


Figure S5. 30. *^{1}H NMR of zoomarinepane C (**3**) recorded in methanol-*d_4 *at 700 MHz.*

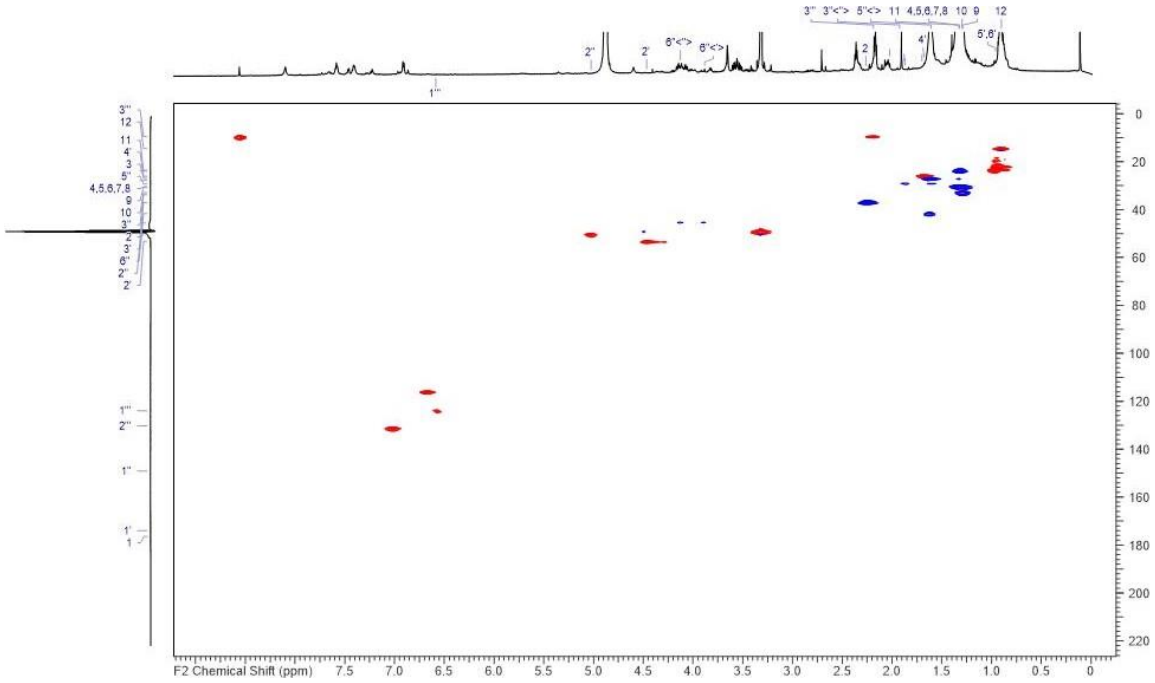


Figure S5. 31. *HSQC NMR of zoomarinepane C (**3**) recorded in methanol-*d_4*.*

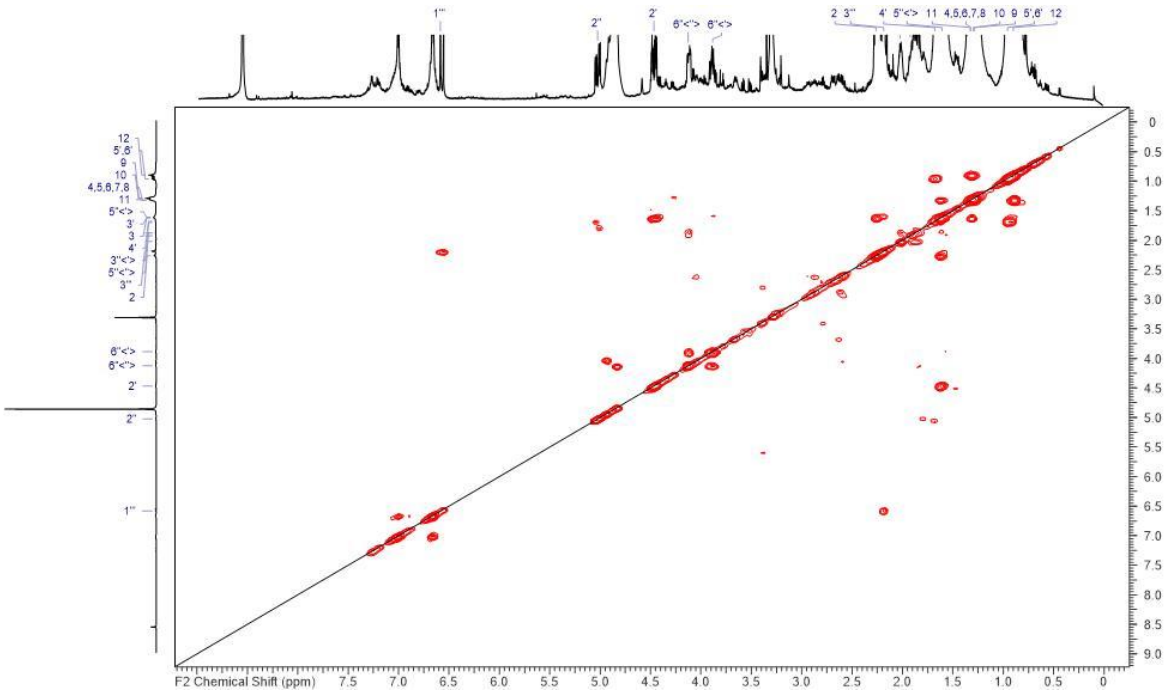


Figure S5. 32. *COSY NMR of zoomarinepane C (**3**) recorded in methanol-d_4.*

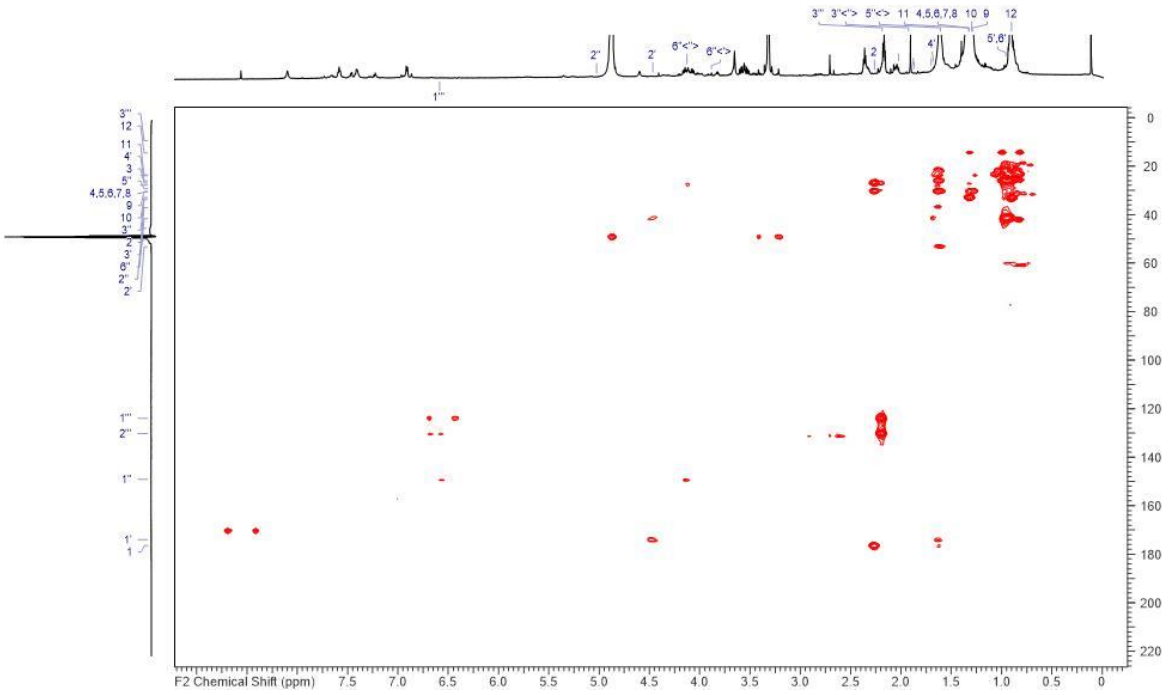


Figure S5. 33. *HMBC NMR of zoomarinepane C (**3**) recorded in methanol-d_4.*

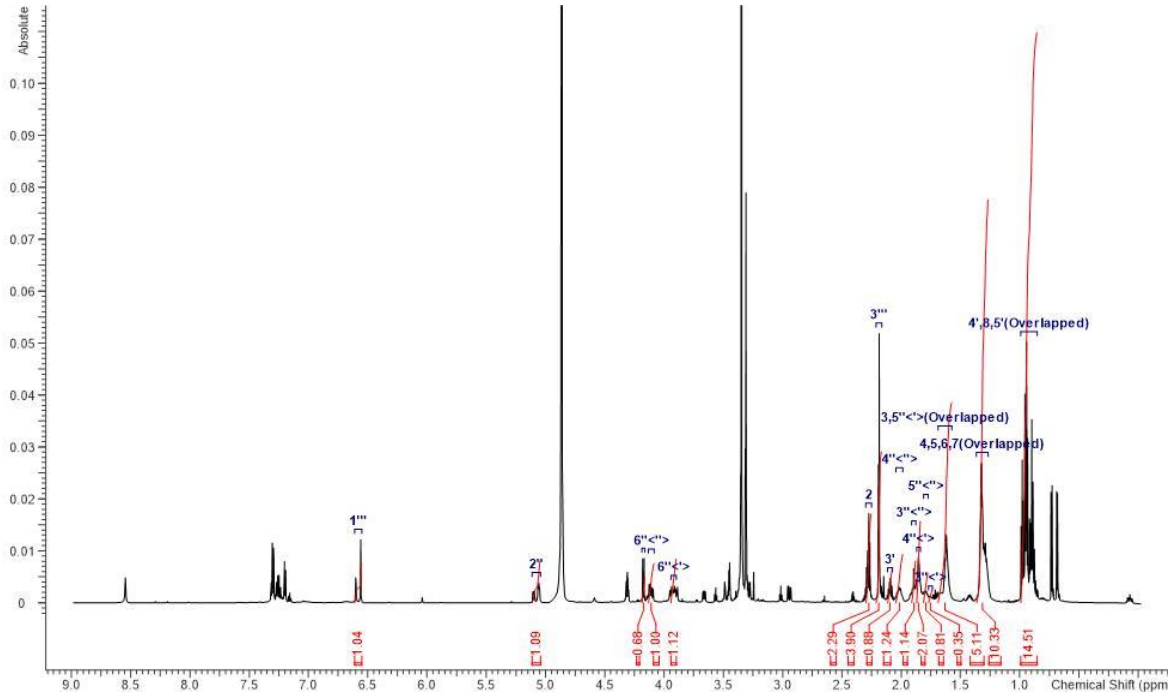


Figure S5. 34. *^{1}H NMR of zoomarinepane D (**3**) recorded in methanol-*d_4 *at 700 MHz.*

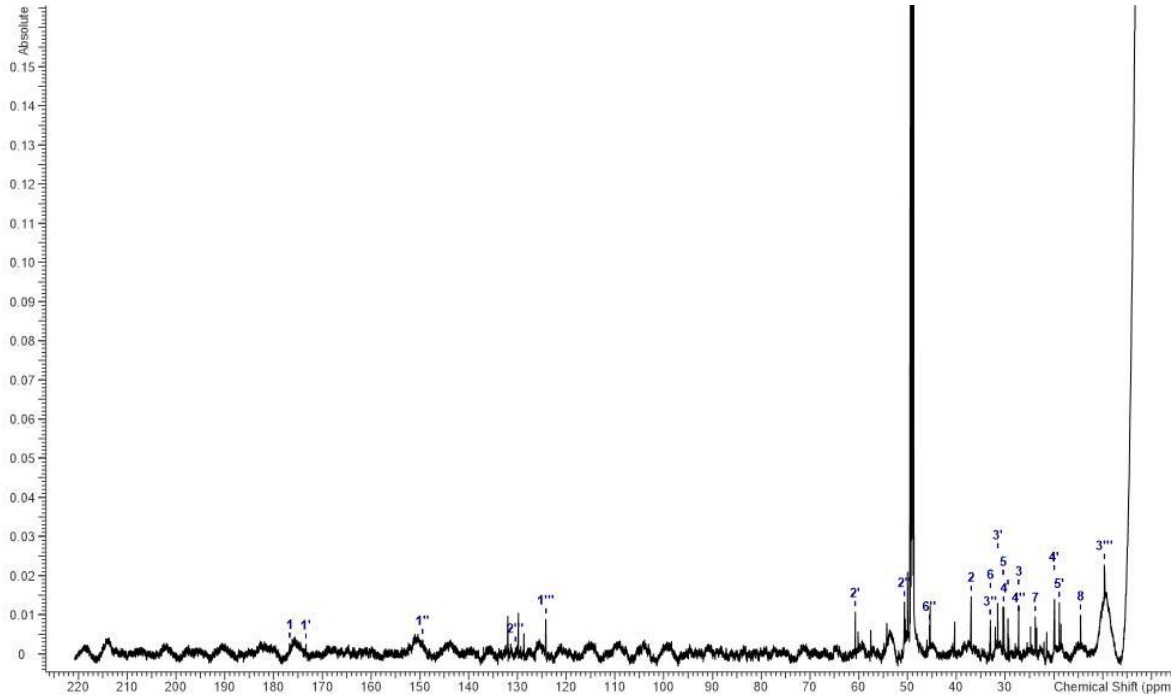


Figure S5. 35. *^{13}C NMR of zoomarinepane D (**3**) recorded in methanol-*d_4 *at 175 MHz.*

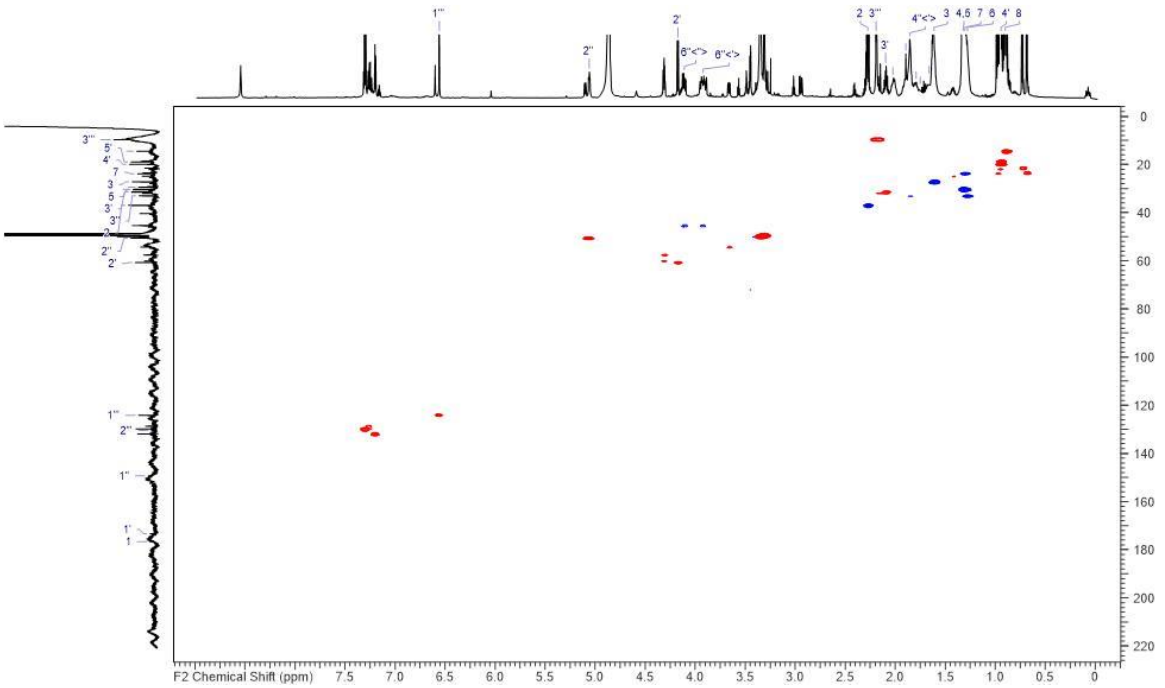


Figure S5. 36. *HSQC NMR of zoomarinepane D (**3**) recorded in methanol-d_4.*

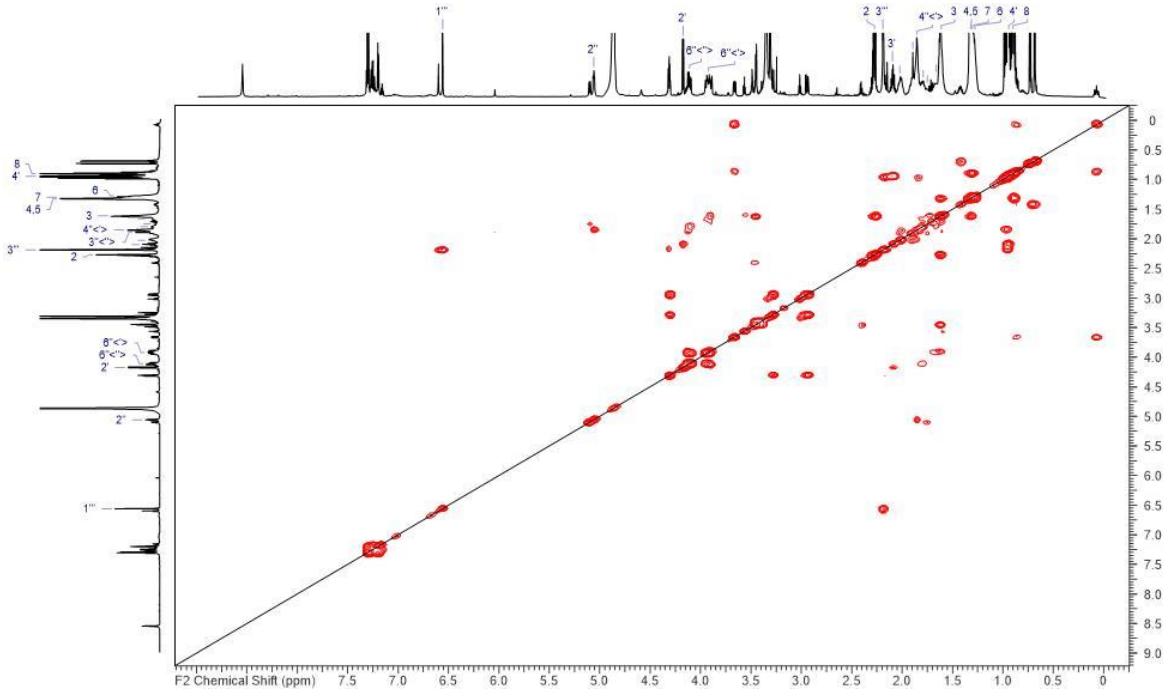


Figure S5. 37. *COSY NMR of zoomarinepane D (**3**) recorded in methanol-d_4.*

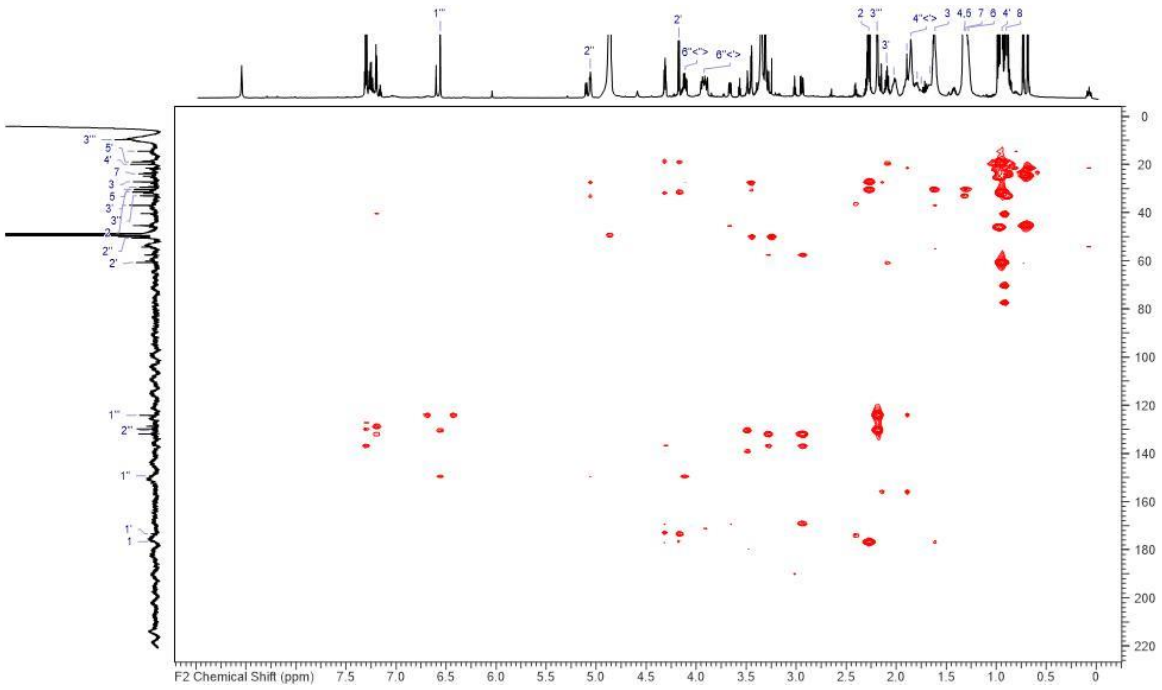


Figure S5. 38. *HMBC NMR of zoomarinepane D (**3**) recorded in methanol-d₄.*

S 5.12 References

1. Blin, K. *et al.* antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.*; 10.1093/nar/gkad344 (2023).
2. Gerc, A. J., Song, L., Challis, G. L., Stanley-Wall, N. R. & Coulthurst, S. J. The insect pathogen Serratia marcescens Db10 uses a hybrid non-ribosomal peptide synthetase-polyketide synthase to produce the antibiotic althiomycin. *PLoS ONE* **7,** e44673; 10.1371/journal.pone.0044673 (2012).
3. Patzer, S. I. & Braun, V. Gene cluster involved in the biosynthesis of griseobactin, a catechol-peptide siderophore of Streptomyces sp. ATCC 700974. *J. Bacteriol.* **192,** 426–435; 10.1128/JB.01250-09 (2010).
4. Wang, X. *et al.* Identification and Verification of the Prodigiosin Biosynthetic Gene Cluster (BGC) in Pseudoalteromonas rubra S4059. *Microbiol Spectr* **9,** e0117121; 10.1128/Spectrum.01171-21 (2021).
5. Reshetnikov, A. S. *et al.* Diversity and phylogeny of the ectoine biosynthesis genes in aerobic, moderately halophilic methylotrophic bacteria. *Extremophiles : life under extreme conditions* **15,** 653–663; 10.1007/s00792-011-0396-x (2011).
6. Lee, J. S. *et al.* Exceptional production of both prodigiosin and cycloprodigiosin as major metabolic constituents by a novel marine bacterium, Zooshikella rubidus S1-1. *Appl. Environ. Microbiol.* **77,** 4967–4973; 10.1128/AEM.01986-10 (2011).
7. Cortina, N. S., Revermann, O., Krug, D. & Müller, R. Identification and characterization of the althiomycin biosynthetic gene cluster in Myxococcus xanthus DK897. *Chembiochem : a European journal of chemical biology* **12,** 1411–1416; 10.1002/cbic.201100154 (2011).
8. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410; 10.1016/S0022-2836(05)80360-2 (1990).
9. Marchler-Bauer, A. *et al.* CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res.* **37,** D205-10; 10.1093/nar/gkn845 (2009).
10. Klau, L. J. *et al.* The Natural Product Domain Seeker version 2 (NaPDoS2) webtool relates ketosynthase phylogeny to biosynthetic function. *J. Biol. Chem.* **298,** 102480; 10.1016/j.jbc.2022.102480 (2022).

11. Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6,** 493–505; 10.1016/S1074-5521(99)80082-9 (1999).

12. Gilchrist, C. L. M. & Chooi, Y.-H. Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics (Oxford, England)*; 10.1093/bioinformatics/btab007 (2021).

Discussion

Modern Natural Products Discovery – Only the Choice between "Rediscovery" of Known Scaffolds or New Scaffolds with No Bioactivity?

On first glance, sceptically viewed, the natural products discovered in this thesis are either derivatives of biologically active known scaffolds (disorazole Zs, Chapter 2 and ajudazols, Chapter 3), or novel scaffolds (cystopipecotides, Chapter 4 and zoomarinepanes, Chapter 5), which did not show any bioactivity in our in-house screenings against the panel of 13 or 14 test organisms, respectively. This fact might raise the question: Are these the only two options for outcomes? Do we, as natural product researchers that are interested in finding novel drug leads, have to choose between known scaffolds or chemical novelty but without bioactivity in standard assays?

This rather bleak picture, however, disregards three arguments that demonstrate a more optimistic perspective: 1. The discovery of new derivatives of biologically active natural products yields many opportunities; 2. The discovery of natural products without initial bioactivity is not the end of the story since often enough promising bioactivity is found later; and 3. There are further approaches applicable to optimise the discovery of novel biologically active natural products. The following sections highlight recent examples of such stories. Furthermore, the downfalls and opportunities for natural products discovery shall be discussed, aiming to convince the sceptical reader why we should still invest in natural products research.

6.1. Why the Discovery of New Derivatives Is a Success Rather Than a Fail

Derivatives of known natural products sometimes reveal their identity relatively late in the discovery process, especially when the producer strain's genome is not sequenced and the applied metabolome analysis neither reveals the known compound itself nor its structural relation to the target compound. This case can be seen in Chapter 2, where a new subfamily belonging to the disorazole scaffold was discovered during a bioactivity-guided workflow. Early de-replication did not show the relation, because the published disorazole families are not produced by this strain,

the biosynthetic gene cluster (BGC) was not identified as the strain was not sequenced, and metabolome analysis did not reveal the identity of the compound family as the structural skeleton is sufficiently different. Only with pure compound and NMR analysis at hand parallels to published disorazoles[1] could be drawn. This may sound like a near-rediscovery story; however, it enabled us to study the diverse chemical scaffold of disorazoles and its biosynthesis in greater details. In addition, disorazole Z is significantly different from the A series of these compounds and could have been given a different name without compromising scientific standards.

Indeed, the discovery of new congeners of a known scaffold of biologically active compounds is often sought out from the beginning because of a several reasons:

1. The investigation of structurally diverse versions of bioactive natural products is often helpful for structure activity relationship (SAR) and mode of action (MoA) studies, sometimes even yielding different activity profiles or physicochemical properties of such congeners[2,3].

2. For the drug development process, the generation of new derivatives is necessary in order to develop a hit into a lead structure, and optimise a lead into a preclinical candidate[2]. However, in many cases, total synthesis is too laborious and ineffective to generate a small library of natural product congeners, which makes biotechnological methods advantageous[2].

3. The discovery of a new producer strain may not only lead to the discovery of new congeners but may open the door to biosynthesis studies, biosynthetic engineering, the generation of non-natural congeners, and the optimisation of production levels[4].

Thus, the study of disorazole Z (Chapter 2) did not only broaden our knowledge about the chemical scaffold and gave further insights into SAR and MoA, the discovery of the biosynthetic pathway and the development of a heterologous expression system enabled the generation of non-natural derivatives and paves the way for biotechnological optimisations (**Figure 6.1**)[5]. Similarly, the potential of biosynthetic engineering efforts was shown in a recent study that generated several new congeners through the heterologous expression of the disorazole A BGC in *Burkholderia thailandensis*[6].

In Chapter 3, the presence of new ajudazole congeners was proposed prior to compound purification and structure elucidation, because the genome of *Cystobacter* sp. SBCb004 was sequenced and BGC analysis showed homologies with the published ajudazol BGC[7]. However, differences in the biosynthesis were proposed, which sparked our interest to discover putatively new congeners. The differences in the biosynthetic pathway that lead to the broadened chemical scaffold could be elucidated, and led to changes in the activity profile: while ajudazole A exhibits

minor activity against fungi and Gram-positive bacteria[8], no anti-microbial but anti-cancer activity could be detected for ajudazole C-H, and a significant loss of cytotoxicity could be shown for glycosylated ajudazole I and J (**Figure 6.1**).

Further myxobacterial compound families that have been broadened or revalued through the discovery of new congeners are cystobactamids, sorangicins, thuggacins, and argyrins. The discovery of new cystobactamid congeners broadened the bioactivity spectrum towards Gram-negative pathogens e.g. *Pseudomonas aeruginosa*[3], neosorangicin brought a congener with better *in vitro* antibacterial activity against Gram-positive pathogens to the table[9], and thuggacin congeners from different producers broadened the chemical scaffold and gave insights into their biosynthetic pathway[10,11]. Furthermore, the impressive story of argyrins shows how the discovery of an alternative producer strain led to the elucidation of the biosynthesis, including the determination of cluster borders and development of a heterologous expression system[12]. This enabled the generation of new derivatives with improved immunosuppressive activity, selective production of certain congeners to facilitate quick purification, and a major increase of the production reaching from less than 10 mg/L in the native producer *Archangium gephyra* Ar 8082[13] up to 400 mg/L in an engineered producer *Myxococcus xanthus* DK1622[14].

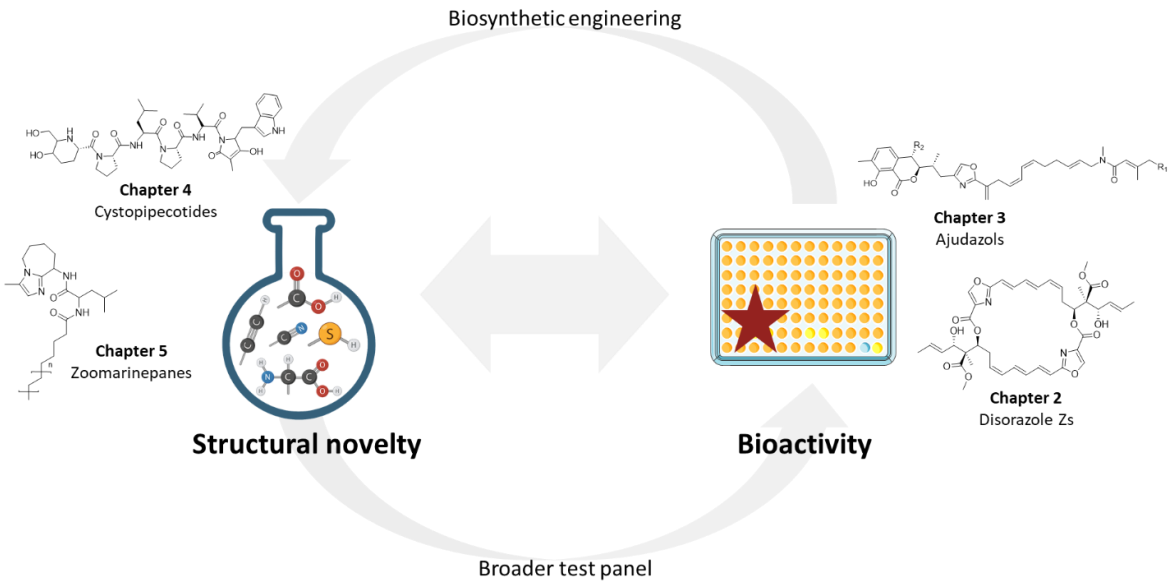


Figure 6.1. The discovery of structurally novel and biologically active natural products covered in this thesis and the opportunities to combine these two discovery aspects. *The star symbolises the observation of a biological activity.*

6.2. Why the Initial Lack of Bioactivity Is Not the End of the Story

Even though the bioactivity-guided approach focusses on an initial bioactivity against a selected test panel, such activity might not be reproducible for the purified compound. For example, synergistic effects of several compounds can lead to a loss of activity after purification, or unquantified high production titres of the "active" component lead to an overestimation of its activity. A prominent example of such case are siderophores, that enhance the activity of other compounds by depriving the pathogen of vital ions[15].

On the other hand, extracts that do not show initial bioactivity might still contain biologically active compounds[16]. The activity can be concealed, because of activity quenching compounds[17], low concentration[18], or because it is not detected through the applied test panel[19]. Nevertheless, natural products are expected to possess great potential to exhibit bioactivities. Often they are produced to fulfil a biological purpose, e.g. to defend the host against competitors, to ensure inter- and intra-species communication or to maintain a symbiosis[20]. Moreover, from a chemical point of view, they also possess great variety of chemical scaffolds, have different physicochemical properties from most chemical libraries, and are evolutionarily optimised to fit molecular targets[21].

Following this logic, genome-guided and metabolome-guided workflows are applied to obtain novel natural products despite a lack of initial bioactivity – or as in case of Chapter 4 and 5 despite the presence of known biologically active compounds, but focusing on the biosynthetic potential to produce many more natural products. In both projects novel compound families were discovered, the cystopipecotides and the zoomarinepanes, with unusual moieties that enrich the chemical scaffolds of natural products. Furthermore, the proposal of their biosynthetic pathways could help to better understand the biosynthesis of other natural products. Those compounds have only been screened against our in-house panel of bacteria, fungi and cancer cell lines. Including them in broader, not only anti-infective screening campaigns might reveal a biological activity (**Figure 6.1**).

Recent examples of compounds without initial bioactivity but successful screening against other targets are sandacrabin B and cyclised thiamyxins. Both were discovered by metabolome-mining techniques revealing interesting structural moieties but only weak activity against microbial pathogens. However, in rather complex and cell-based anti-viral screenings both showed potential application windows when compared to their cytotoxic side activity[22,23].

6.3. How to Optimise the Discovery of Novel Biologically Active Compounds from Myxobacteria

The contemporary workflow to yield new natural products includes the selection of a strain, cultivating it, extracting its metabolites, detecting, analysing and prioritising metabolites for compound purification, and elucidating their structures (**Figure 6.2**). Hereto mentioned bioactivity-, genome- or metabolome-guided workflows are commonly applied to discover new natural products either by a bioactivity-first or a compound-first approach. However, there are many options to optimise this workflow and to intertwine approaches in every step. A few examples of those options will be discussed in the following chapters.

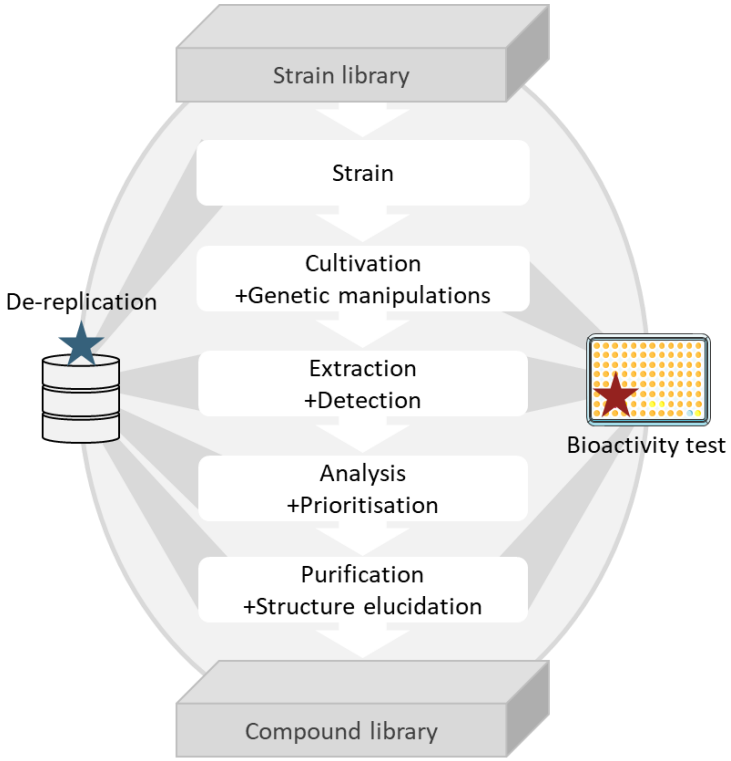


Figure 6.2. Contemporary workflow to discover novel biologically active natural products from myxobacteria. *The star symbolises the observation of a biological activity.*

6.3.1. The Impact of the Strain Selection on Natural Products Discovery

Myxobacteria have proven to be a rich source of biologically active natural products in the past. However, the majority of myxobacterial BGCs have not been connected to their corresponding natural products[24]. Moreover, isolation-independent evaluation of different environmental samples showed that the number of myxobacterial species in these samples is far greater than the

number of isolated strains. This proves the point that the majority of myxobacterial strains have not been studied yet[25–27] and demonstrates their vast untapped potential. In the past, the Myxobacterial Strain Collection at the Helmholtz-Centre for Infection Research has been filled with a large variety of Myxobacteria, especially focussing on *taxa* that produced the most interesting natural products at that point[28]. Nevertheless, in recent years, taxonomic diversity was correlated positively with the diversity of chemical scaffolds produced by bacteria[29,30], thus encouraging us to invest in taxonomically diverse and under-explored species[31]. Due to this fact, our in-house strain collection is constantly filled with novel isolates, which have to be evaluated and prioritised for further investigations. As the selection of a strain is the first step in natural products discovery (**Figure 6.2**), it has a big impact on the outcome of the workflow. Hence, the selection of a strain can be based on their phylogeny-, metabolome-, bioactivity- or genome analysis. In this work, all four parameters were evaluated and de-replicated against published and in-house data in order to select a candidate strain. In Chapter 2, the strain *Sorangium cellulosum* So ce1875, which belongs to a genus proven to host great potential for natural products production[28] was selected because of its bioactivity profile. *Cystobacter* sp. SBCb004 (Chapter 3 and 4) had previously shown great potential to produce potent natural products[12,32,33] and was therefore selected because of its biosynthetic potential. Finally, presented in Chapter 5, the non-myxobacterial strain *Zooshikella marina* sp. Uxx12806, which was added to our strain library because of its myxobacteria-like ability to glide on agar, was chosen as it belongs to an underexplored genus and displays great biosynthetic potential to produce new secondary metabolites.

In our recent review we discussed the value of differential strain isolation as follows:

> "In addition to isolating new bacteria from soil or marine samples, the investigation of animal and human microbiomes shows promise (Hegemann *et al*, 2022). […] Techniques such as microfluidics (Mahler *et al*, 2021) or the "isolation chip" (iChip; Nichols *et al*, 2010) enable high-throughput isolation of microorganisms, only limited by their ability to grow in the chosen environment. Due to the primary in situ cultivation in the environment the original sample was taken from, the iChip greatly improves the number and diversity of isolated strains, compared with conventional Petri dish-based strain-isolation procedures (Nichols *et al*, 2010). Screening of iChip isolates for antibacterial activity resulted in the discovery of the antibiotic teixobactin (Ling *et al*, 2015). Nevertheless, better knowledge of the microbiology of known, but understudied producers of natural products can lead to a constant increase of accessible biodiversity, as demonstrated for Myxobacteria (Garcia *et al*, 2009; Garcia *et al*, 2010)."[31]

Using new isolation techniques and investigating less explored biological samples is likely to lead to novel myxobacterial strain isolations, which may not yet have been cultivated and have great

potential to produce new biologically active natural products. Prominent examples of novel or understudied myxobacterial *taxa* that produce potent compounds are *Pendulasporaceae*[34,35] and *Sandaracinus*[22,36].

6.3.2. The Impact of Cultivation and Genetic Manipulation on Natural Products Discovery

For most microorganisms, about 90% of BGCs could not be linked to a respective natural product under routine cultivation settings[24]. However, studies showed that under standard laboratory conditions most BGCs are transcribed and translated into proteins[37,38], which raises the question why we cannot identify their respective compounds. Either the production yield itself could be too low[39], or due to instabilities[40], insufficient extraction[41] and unsuitable detection methods[42] the produced compound cannot be identified. Even though there are many approaches to tackle the analytical-chemical problems, this paragraph will focus on methods that activate cryptic BGCs, thereby increasing the amount of detectable natural product (**Figure 6.3**).

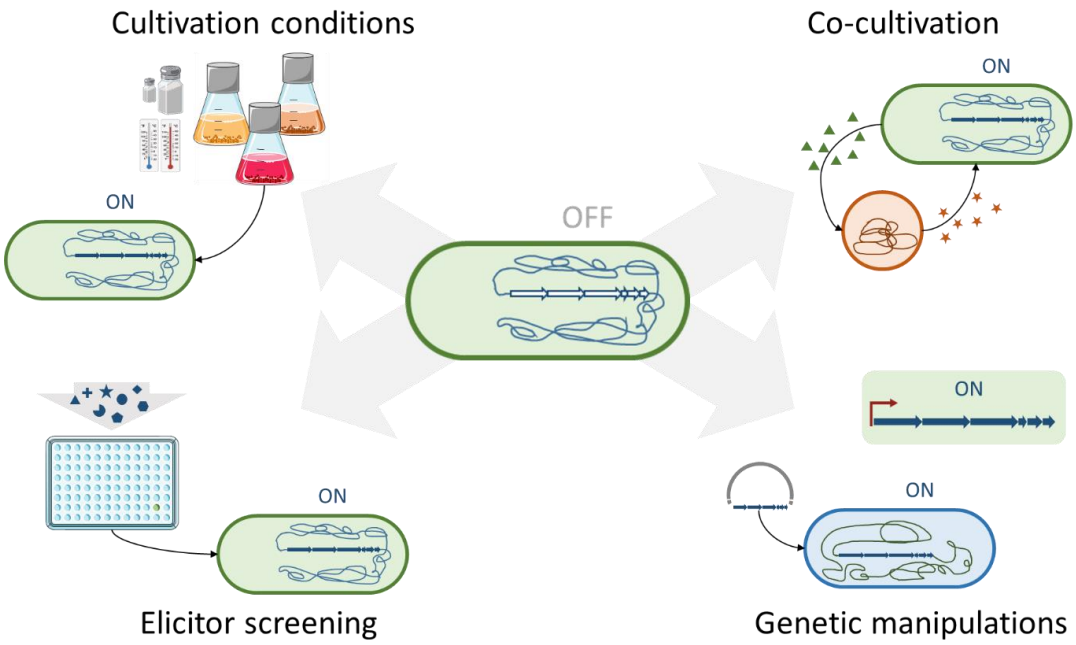


Figure 6.3. Strategies to unlock cryptic BGCs. Top left*: alteration of media and cultivation conditions.* Top right*: The crosstalk in co-culture.* Bottom right*: Activation of a BGC through promoter insertion (top) and heterologous expression (bottom) as examples for genetic manipulations.* Bottom left*: chemical elicitation using a library of small molecules. Adapted from Covington* et al.*[39]*

Media composition and cultivation conditions. The fact that changes at the cultivation level have a major impact on natural products production has been known throughout the history of natural

products science. Nevertheless, starting with the introduction of the one strain many compounds (OSMAC) approach by the Zeeck group, cultivation alterations gained popularity[43,44] and were applied systematically. On the one hand, conditions can be altered by the composition of media or external factors to mimic the natural environment of the strain, for instance choosing similar salinity, trace elements or cultivation temperature. On the other hand, stress-conditions can be applied to trigger natural products production, e.g. over-access or limitation of oxygen supply or media components, or aggravated pH values[45,46]. There are countless possibilities to vary cultivation parameters and one can get very creative to either mimic or aggravate the natural environmental conditions (**Figure 6.3**), whereas both strategies produce many examples that successfully led to the discovery of natural products. A recent study described three unprecedented natural products after cultivating a myxobacterial strain in both aggravated and nature-mimicking conditions, while the production was induced during the mimicking conditions[47]. In a different study, challenging the marine fungus *Spicaria elegans* through high salt supplementation led to the discovery of known and new secondary metabolites[48]. In Chapter 5, a wide cultivation temperature and media salinity range was applied to the cultivation of a *Zooshikella marina* strain. These experiments did not only show the up- and down-regulation of prodigiosin production, but also demonstrated that althiomycin is only produced in detectable amounts under very specific conditions. Furthermore, these experiments led to the discovery of an unprecedented compound family, zoomarinepanes.

Co-cultivation. Considering that microbial environments are mostly complex communities of thousands of organisms, cultivating two or more microbes together is a logical consequence to mimic natural environments. Furthermore, as natural products are known for their role in inter- and intra-species communication, defence mechanisms and symbiotic relationships[20], it is a rational argument that co-cultivation is able to trigger natural products production (**Figure 6.3**)[49]. This approach has already proven successful to enhance the production of antibiotics that could not be detected in pure culture[50,51]. A very successful approach to trigger the production of natural products in *Streptomyces* is the combined-cultivation with mycolic acid producing bacteria[52,53]. This fact inspired recent studies to also co-cultivate Myxobacteria with mycolic acid producers, which triggered the production of several compounds including antibacterial natural products[35]. Even the co-cultivation of two strains from the same genus is able to enhance natural products production, as exemplified by the pairwise co-cultivation of *Sorangium cellulosum* spp., of epothilone producing strains with strains without the biosynthetic machinery to produce epothilones, which enhanced the production of this anti-cancer compound in most cases about 73%[54]. Nevertheless, co-cultivation of two or more microorganisms are highly complex biological

experiments. Therefore, reproducibility and upscaling are difficult to achieve, which may prove to be an obstacle to producing sufficient quantities of a candidate compound[39,55].

Genetic manipulations. Even though genetic manipulations can be used to simply link a BGC to a respective natural product (family), for instance by cluster deactivation as shown for ajudazols (Chapter 3) and cystopipecotides (Chapter 4), molecular biological methods can also be a powerful tool to activate cryptic BGCs. In a first step, BGCs of interest must be prioritised, which can be based on e.g. "the underlying biosynthetic pathway (Hug *et al*, 2019) or, especially in the case of antibiotic discovery, on the presence of self-resistance genes within or nearby the cluster (Johnston *et al*, 2016; Alanjary *et al*, 2017; Panter *et al*, 2018)."[31] The BGC of interest can then be activated in the native producer through the insertion of a constitutively active or inducible promoter or by repressor deletion, or heterologously expressed in another host (**Figure 6.3**)[56,57]. Thus, heterologous expression is not only a helpful tool to investigate biosynthetic pathways or employ biosynthetic engineering of a BGC, as shown for disorazoles (Chapter 2), it is a valuable tool to express cryptic BGCs. However, as discussed in our recent review:

> "One major drawback in heterologous expression is the need to find suitable and genetically tractable hosts that ideally are closely related to the native producers. Application of transposition-based approaches (Fu *et al*, 2008) or the chassis-independent recombinase-assisted genome engineering (CRAGE) method (Wang *et al*, 2019) enable integration of BGCs into different heterologous hosts. The benefits of these approaches include improved production titres, depending on the integration position (Bilyk *et al*, 2017; Pogorevc *et al*, 2019), and an increased diversity of derivatives from an individual BGC, depending on different heterologous hosts (Wang *et al*, 2019). Using heterologous expression even allows in some cases to screen for new natural products without the need to cultivate the original producing bacteria."[31]

This could be a powerful tool to investigate chemical scaffolds that are encoded in myxobacterial strains that are difficult to isolate or difficult to grow.

Chemical elicitation. It has been known for a long time that the addition of small molecules to a culture of microbes has the possibility to change natural products production. For example, it could be shown that the addition of certain antibiotics produces mutations in ribosomes and thereby changes the metabolome of a strain[58]. This effect, called ribosome engineering, inspired studies to elicit natural products production by the addition of small molecules[59]. In contrast to ribosome engineering, however, most chemical elicitors induce the production of secondary metabolites not through genetic modifications, but through interaction with regulators or by acting as repressors or activators themselves[39,49]. Nevertheless, it is difficult or even not possible

to predict the effect of small molecules on different strains; instead it needs to be determined experimentally. Therefore, high-throughput elicitor screening (HiTES) was established within the past decade[60], mainly using strains like *Burkolderia*[61], *Streptomyces* spp.[60], other Actinobacteria[62], or fungi[63]. For this approach, a library of small molecule elicitors is added to microbial cultures and the effect is read-out with selected methods (**Figure 6.3**). Thereby it could be shown that e.g. the antibiotic trimethoprim in sub-inhibitory concentrations globally upregulated secondary metabolite production in *Burkholderia thailandensis*[64]. However, the HiTES approach has not yet been reported for many other *taxa*[47,65], hence offers great opportunities for natural products discovery in myxobacterial strains (cp. Sebastian Walesch, dissertation in preparation).

6.3.3. A Holistic Approach to Discover Novel Biologically Active Natural Products

As described in Chapter 1.4, four different approaches to discover novel compounds have been developed throughout the history of natural products research: the bioactivity-, the genome- and the metabolome-guided workflow, as well as cultivation-based approaches. All workflows have certain benefits and drawbacks. The bioactivity-guided approach is probably the most intuitive and straightforward approach to discover biologically active compounds, however, it is challenged by re-discovery of known natural products or loss of activity during purification. The metabolome- and genome-guided workflows approach the discovery from a different angle using de-replication and prioritisation techniques to discover novel compounds. However, as the biological activity is not a focus from the beginning, the bioactivity of the discovered natural product is often difficult or impossible to predict. Above-mentioned strain selection and cultivation-based techniques (6.3.1-6.3.2) can increase the chance to discover novel chemistry and to unlock cryptic BGCs. However, complex large data sets are created that must be evaluated, de-replicated and prioritised to identify novel biologically active compounds (**Figure 6.2**). Thus, it seems most beneficial to combine techniques from all workflows to achieve the best output – and to use up-to-date analytical and statistical methods to wrestle those large datasets (**Figure 6.4**).

Interplay between genomics, metabolomics and strain isolation. The composition of a strain library is directly connected to their potential to produce natural products. Therefore, it is advisable to fill said strain library with taxonomically different strains to discover chemically diverse scaffolds[29]. In order to optimise the first step of the natural products workflow, recent studies show that metagenomic and metabolomic studies of complex microbial samples in combination with bioinformatic analysis help to prioritise samples and direct microbe isolation. For example, in our lab metagenomic analysis is currently carried out with the aim of prioritising

soil samples with understudied myxobacterial *taxa* for strain isolation (Daniel Krug, personal communication). Furthermore, the metabolomics-bioinformatics pipeline IDBac was published, which enables the prioritisation of strain isolates based on high-throughput matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) analysis, which gives first insights into the phylogeny of single colonies by protein fingerprint and into natural products production by small molecule fingerprint[66]. Nevertheless, it must be noted that the sensitivity of MALDI depends on compound classes, thus, the effectiveness of this method awaits further investigation[67].

Interplay between bioactivity and genomics. As discussed before, the biosynthesis of natural products is often sophisticated and encoded in large BGCs, which results in energy-expensive processes for the producer strain[24]. Therefore, it is assumed that those natural products have a biological purpose and ensure a survival benefit for its producer[68]. The genome-guided workflow is a directed approach that focusses on the biosynthetic potential of a strain. There are several ways to proceed for bioactivity-guided genome mining: First, prioritising a BGC of interest, then choosing a method to connect it to a natural product. It is assumed that orphan BGCs, which have not been associated with known natural products, have great potential to encode for the production of novel biologically active compounds[69]. Even well studied organisms harbour many orphan BGCs, as exemplified by the myxobacterial model strain *M. xanthus* DK1622 that encodes for at least 24 BGCs of which only 9 were connected to their respective natural product[70–72]. Similarly, the strain *Cystobacter* sp. SBCb004, which was studied in Chapter 3 and 4, is considered a rather well known strain. However, only seven BGCs out of 49 in its genome were connected to their respective natural product prior to this study. These examples illustrate the immense biosynthetic potential that already investigated Myxobacteria hold, not to mention the strains that have not yet been explored or discovered.

Several targeted bioactivity-guided genome mining approaches focus on chemical moieties and their respective biosynthetic gene clusters that are associated with biologically active features (chemistry-forward)[73]. Good examples for this approach are the search for new enediynes in *Streptomyces*, which are chemical moieties involved in anti-cancer activity[74], or the follow-up of antibacterial activity and the presence of a putative phenazine BGC in of *Kitaspora* sp. that led to the discovery of new endophenazines[75]. The drawback of these chemistry-forward approaches, however, is that the discovery is limited to known structural features, which rarely possess new mode of actions[76]. Different targeted approaches focus on genes that are predicted to be involved in self-resistance mechanisms or efflux pumps (BGC forward). The reasoning behind this approach

is that the producer needs to protect itself against the toxicity of the natural product by self-resistance mechanisms, and excrete the toxic compound in order to effect its opponent[73]. The genes encoding for these mechanisms are expected to be part of orphan BGCs, which makes them a promising target for genome mining[39]. Prominent examples of these approaches are the fatty acid synthesis inhibitor antibiotics thiotetronic acid, which were discovered through the identification of a duplication of a housekeeping gene, a putative fatty acid synthase resistance gene in *Salinispora* strains[77,78]. Furthermore, pyxidicyclins were discovered due to the presence of a gene encoding a pentapeptide repeat protein in a type II polyketide synthase BGC, which protects the hosts DNA gyrase against gyrase inhibitors[79]. This chemistry-forward approach often results in the discovery of unknown chemical features. However, the mode of resistance and possibly mode of action of the discovered natural products might be predetermined.

Untargeted "chemical genetic" approaches in combination with a bioactivity-guided workflow aim to elicit the expression of BGCs responsible for the production of biologically active natural products. In order to achieve this goal several strategies can be applied, like changing the cultivation modalities, co-cultivation or HiTES (Chapter 6.3.2)[39]. This approach results in complex data sets that require in-depth analysis and de-replication methods from the metabolome analysis workflow.

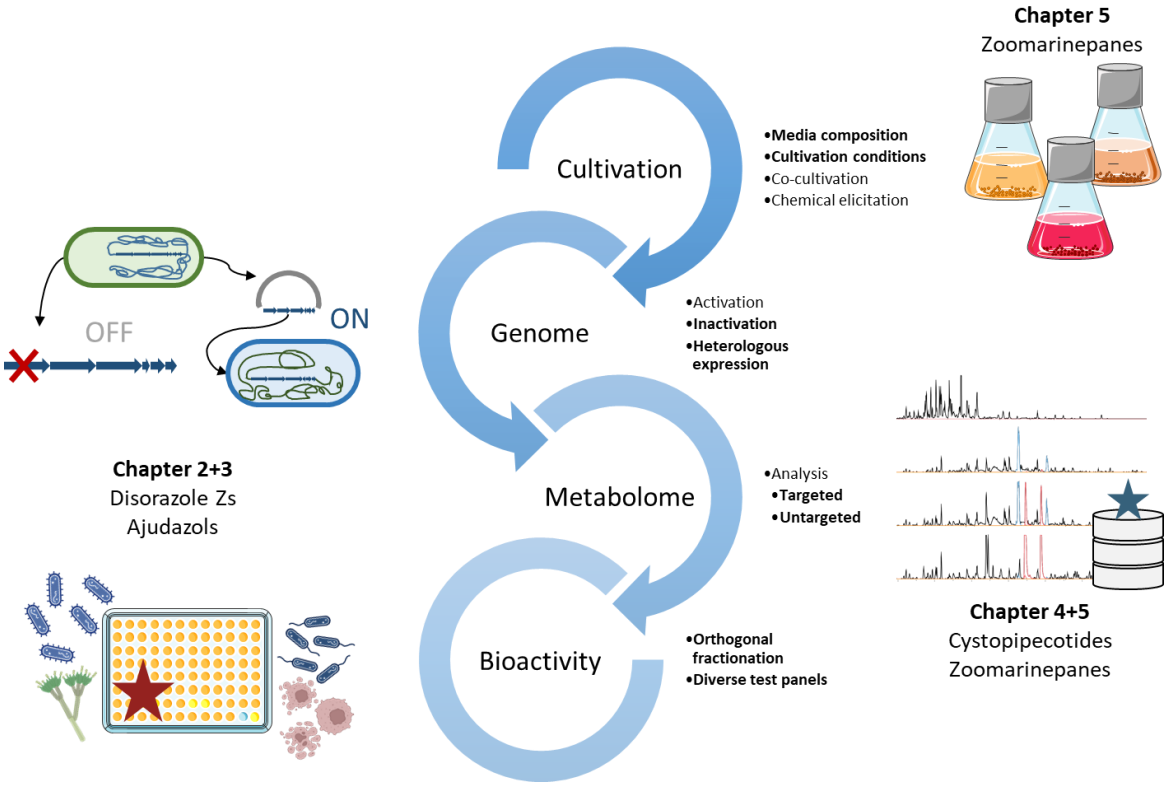


Figure 6.4. Elements of the holistic approach to discover novel biologically active natural products. *Cultivation experiments greatly affect the expression of BGCs, which can be observed through metabolome analysis and bioactivity screening. Different strategies can be applied in the discovery process (bullet points).* Bold*: Strategies that were applied in this work. The star symbolises the observation of a biological activity.*

Interplay between bioactivity and metabolomics. The metabolomics approach is independent of genetics. Thus, no genetic amenability of the strain or BGC analysis is necessary (however might be helpful). As a structure or chemical feature oriented approach, targeted workflows may focus on certain chemical moieties or on binders of molecular (known) targets. An example to apply this strategy is the fishing of natural products from crude extracts with molecular targets (proteins) and submitting the assay to native protein MS measurements[76,80]. However, as discussed before, targeted approaches often lead to a restricted output. In this case the mode of action or chemical moieties are predetermined.

On the other hand, untargeted approaches like OSMAC, HiTES and co-cultivation in combination with a bioactivity-guided workflow are powerful to discover novel biologically active compounds[39,45], whereas the read-out method has a big influence on the bioactivity discovery workflow[39]. In case of screening for e.g. novel antibiotics "the traditional *in vitro* antibiotic susceptibility tests (ASTs) […] do not take into account the *in vivo* environment during an infection. Changes in the availability of nutrients or oxygen, host metabolism and the presence of other microorganisms at the site of infection can greatly affect the *in vivo* bioactivity (Lakemeyer *et al*, 2018)."[31] Therefore, mimicking infection models and a broadening of the screening set-up[81,82] will be beneficial to detect broad ranges of bioactive compounds and enable early prediction accuracy for the translation to animal models[83].

In order to identify a target compound responsible for the bioactivity, it must be detectable by a selected analytical set-up. Most commonly, hyphenation of liquid chromatography with UV/Vis detectors and electrospray high-resolution mass spectrometers (LC-UV/Vis-ESI-HR-MS) in positive ionisation mode are used. However, recent studies show that the selection of analytical systems has a great impact on the detection of natural products[42]. Thus, for instance some terpenes and steroids ionise better using atmospheric pressure chemical ionisation (APCI) or matrix assisted laser desorption ionisation (MALDI)[67,84], and ion mobility MS enables the detection of isomeric and isobaric compounds by semi-orthogonally separating them[85]. Furthermore, nuclear magnetic resonance spectroscopy (NMR)-based techniques are able to give insights into structural moieties and non-ionisable compounds. Hyphenated techniques, like LC-solid phase extraction-NMR-MS

(LC-SPE-NMR-MS), can even be applied on complex fractions or crude extracts to identify a putative target compound[86]. Interesting techniques that can be employed in a high-throughput manner in order to inspect the natural product profile on solid surfaces are MALDI and desorption electrospray ionisation (DESI)[67]. Especially DESI seems a particularly promising ionisation mode, as exemplified by the analysis of predation zones of *M. xanthus* on *Escherichia coli* showing the production of several compounds that are known to be involved in predation[87]. It would be interesting to test this set-up in a high-throughput predation-based or bioactivity-guided metabolomics workflow.

As crude extracts, such as those originating from Myxobacteria, are complex mixtures of several thousand compounds with different production titres, it is essential to use separation techniques to avoid a biased analysis towards high-intensity compounds. Reverse phase C-18 high pressure liquid chromatography (HPLC) is the most commonly used chromatographic technique to separate compounds of myxobacterial crude extracts. However, rather hydrophilic or lipophilic compounds benefit from either different column materials or different chromatographic systems altogether, for instance supercritical fluid chromatography (SFC)[22,88]. In addition to analytical separation, fractionation of crude extracts to test their biological activity is crucial to pinpoint the biological activity to a certain compound (family). Due to the increasingly high rate of known compounds produced by myxobacterial strains, it is most beneficial to use several orthogonal fractionation techniques in parallel in order to apply combinatory analysis to determine bioactive candidates[18,88–90]. Systems that are orthogonal to reverse phase HPLC are for example: normal phase silica, size exclusion (SEC), ion affinity (IAC), centrifugal partition chromatography (CPC), or SFC[91,92].

The resulting huge data sets are impossible to analyse manually. Therefore, several data analysis tools with implemented statistical models, called chemometrics, have been established in recent years that are able to apply matrix filtering and de-replication[93]. Furthermore, biochemometric analysis is able to connect this data with the output from bioactivity assays[90]. A recently published open access tool is NP analyst[94], which is based on bioactivity-based molecular networking after bioassay-guided fractionation of extracts[95]. Such tool will be immensely helpful to quickly identify biologically active components and to de-replicate them against in-house and published databases.

Linking the metabolome and BGC data together. After the identification of a biologically active component during the bioactivity-guided metabolomics workflow, biosynthesis studies and biosynthetic engineering are often the next step to e.g. optimise production or generate

congeners. A common, however manual *in silico* approach to link the chemical structure of the target compound to its putative BGC is called retro-biosynthetic analysis. In Chapter 4 and 5, this approach was used to propose a biosynthetic pathway that leads to the production of cystopipecotides and zoomarinepanes, respectively. On the other hand, after identifying a BGC of interest during the bioactivity-guided genomics workflow, the next step is to link a natural product to the respective BGC. Automated workflows can be used in a more high-throughput manner. Thereby, they are very helpful to link metabolomic data to genomic data and vice versa. Statistical analyses like principal component analysis (PCA) are commonly used to analyse complex datasets[96], e.g. to link natural products to their respective BGC after heterologous expression, activation or inactivation experiments, as shown in Chapter 3. Thus, all approaches are well intertwined in contemporary natural product discovery workflows[97] (**Figure 6.4**), as can be seen in all Chapters of this thesis.

For compounds with characteristic MS^2 fragmentation patterns tools can be used to link metabolome data to genome data without conducting any molecular biological experiments. Examples for such tools are NRPminer[98] to identify non-ribosomal peptides due to their predictable fragmentation pattern, and NPOmix[99] for compounds regardless of their chemical class by using machine learning. Another approach to link metabolome and genome data without conducting genetic manipulations are transcriptome analysis in combination with proteome and metabolome analysis. However, this approach has not yielded successful stories about the global picture of natural products production yet, but only showed correlations of selected sets of compounds[100]. Therefore, it will be interesting to see whether a picture of global natural products production can be drawn with this technique in the future.

6.3.4. Dereplication and Early Structure Predictions

One of the most important steps during the workflow to discover novel biologically active natural products is the de-replication step, meaning the analysis whether compounds are known or related to known scaffolds. This analysis is most valuable as early as possible in the workflow, because purification and structure elucidation experiments are costly and time intensive (**Figure 6.2**)[101]. Further, de-replication is required to be fast, straightforward, broadly and easily applicable, which is realisable through automated workflows.

The earliest point for the application of de-replication in the discovery workflow is the phylogeny level of the respective strain. In order to find chemical novelty, novel strains should be explored. Moreover, as taxonomic diversity correlates with chemical diversity the focus should lie on novel

taxa. For Myxobacteria, we usually compare the phylogeny of novel isolates with strains from our in-house library, other published strains and publicly available 16S rRNA data of uncultivated Myxobacteria. However, open-access tools are available[102,103] and especially useful for other phyla, which are under-represented in our strain collection. Nevertheless, some compounds are produced by strains across different phyla, for instance prodigiosin is produced by many *taxa* such as actinobacterial[104] and *Zooshikella*[105] strains. This emphasises the fact that de-replication should go further than phylogeny analysis.

After the strains genome is sequenced, BGC de-replication against open-access databases, e.g. NCBI[106], and our in-house database should be applied to assess the biosynthetic potential for the production of novel natural products, and to identify homologies to known BGCs. A most helpful tool is antiSMASH[107], which is an open-source online tool that pulls information from different databases to annotate bacterial and fungal BGCs. The BLAST algorithm[108] can be used to compare nucleotide and protein sequences of a query against databases. However, it works rather manually and is not suitable for high-throughput annotations. On the other hand, MIBiG[109] is a community-curated repository of BGCs that is incorporated in the latest antiSMASH versions. It is very useful to screen genomes for BGC similarities to known clusters in a higher throughput manner. Moreover, the antiSMASH pipeline can give first hints on structural moieties produced by orphan BGCs, which is helpful for early de-replication processes. Finally, gene cluster families can be identified and visualised by clustering BGC-similarities using BiGSCAPE[102]. Nonetheless, as shown in chapters 3-5, the prediction of BGC similarities is oftentimes not very accurate. Small similarity values sometimes still account for identical structures, whereas bigger similarity values might not correlate with a known natural product, e.g. for *Zooshikella marina* sp. Uxx12806 the production of prodigiosin was verified by HPLC-MS, but the BGC was annotated with a score of only 42%, whereas a cluster annotated to be responsible for saframycin biosynthesis (66% similarity) could not be verified. Furthermore, data accuracy and similarity indices may change over time with the quality and amount of data input, which was also noticed with the *Zooshikella* data (data not shown).

Therefore, it is crucial not only to de-replicate genome but also metabolome data to avoid misguided work. As metabolome analysis is probably the fastest and easiest method to conduct in a high-throughput manner, it is also the most broadly used technique in natural products research. As discussed above, different analytical set-ups result in the detection of different sets of natural products. Hence, it is essential to choose methods wisely and combine analyses if necessary. The resulting datasets can be analysed using various software. In this work, metabolome analysis was

mostly conducted using HPLC-ESI-HR-MS and –MS^2data and the software MetaboScape (Bruker) was used for media blank subtractions and compound annotations with our in-house library Myxobase[110]. Most useful open-source and community-curated libraries harbouring published data on bacterial natural products (among others), that were also used in this work, are the NPAtlas[111], LipidMaps[112] and the dictionary of natural products[113]. Furthermore, the annotation of natural product families is a key step in modern de-replication processes. Thus, molecular networking by GNPS[114] was applied, where the use of an appropriate library of MS^2 spectra is most helpful. Additionally, manual or software-based analysis of MS^2 fragmentation patterns can give early insights into structural components of target compounds[115]. Further details about structural moieties of target compounds can be gathered in early-step application of NMR techniques, for instance hyphenated LC-SPE-NMR[86]. Partial structures can be manually analysed or annotated with the help of databases like smartNMR[116,117] or NP-MRD[118]. Furthermore, two recent studies shows how NMR-based metabolomics of complex natural products mixtures in combination with the use of databases can lead to the discovery of novel natural products[119,120]. It will be interesting to see how these semi-automated de-replication workflows can make natural products research more efficient by further feeding data into community-curated libraries, and contribute to the discovery of novel biologically active compounds.

6.4. Concluding Remarks

The natural products space is an important pillar for the drug discovery process harbouring great potential of identifying unique structural moieties exhibiting promising affinities to molecular targets. Even though some myxobacterial strains have been explored in detail already, Myxobacteria are still a treasure trove among natural products producers showing a great amount of untouched *taxa* and biosynthetic potential. This work shows how the application of diverse approaches – the bioactivity-, genome-, metabolome- and cultivation-guided workflows – can lead to the discovery of diverse chemical scaffolds. It also highlighted not only the drawbacks, but more importantly the advantages in the discovery of novel congeners and compounds without initial biological activity. Finally, as an outlook, it was proposed that the skilful use of a holistic natural products discovery workflow – diverse cultivation and genomic methods, orthogonal fractionation and bioactivity-metabolome-guided screening with a diverse bioactivity-indicator panel – seems the most promising approach to exploit the myxobacterial potential in-depth, and thus to discover many more novel biologically active natural products in the future.

6.5. References

1. Jansen, R., Irschik, H., Reichenbach, H., Wray, V. & Höfle, G. Disorazoles, highly cytotoxic metabolites from the Sorangicin-producing bacterium *Sorangium cellulosum,* strain So ce12. *Liebigs Ann. Chem.* **1994,** 759–773 (1994).
2. Miethke, M. *et al.* Towards the sustainable discovery and development of new antibiotics. *Nat. Rev. Chem.* **5,** 726–749; 10.1038/s41570-021-00313-1 (2021).
3. Hüttel, S. *et al.* Discovery and Total Synthesis of Natural Cystobactamid Derivatives with Superior Activity against Gram-Negative Pathogens. *Angew. Chem. Int. Ed. Engl.* **56,** 12760–12764; 10.1002/anie.201705913 (2017).
4. Teijaro, C. N., Adhikari, A. & Shen, B. Challenges and opportunities for natural product discovery, production, and engineering in native producers versus heterologous hosts. *Journal of Industrial Microbiology & Biotechnology* **46,** 433–444; 10.1007/s10295-018-2094-5 (2019).
5. Gao, Y. Heterologous Expression for the Discovery and Biosynthesis Studies of Myxobacterial Natural Products. Dissertation. Universität des Saarlandes, 2023.
6. Wang, Z.-J. *et al.* Engineered Biosynthesis of Complex Disorazol Polyketides in a Streamlined Burkholderia thailandensis. *ACS Synth. Biol.*; 10.1021/acssynbio.2c00610 (2023).
7. Buntin, K. *et al.* Production of the antifungal isochromanone ajudazols A and B in *Chondromyces crocatus* Cm c5: biosynthetic machinery and cytochrome P450 modifications. *Angew. Chem. Int. Ed. Engl.* **47,** 4595–4599; 10.1002/anie.200705569 (2008).
8. Kunze, B., Jansen, R., Hofle, G. & Reichenbach, H. Ajudazols, new inhibitors of the mitochondrial electron transport from Chondromyces crocatus. Production, antimicrobial activity and mechanism of action. *J. Antibiot.* **57,** 151–155 (2004).
9. Zaburannyi, N., Herrmann, J., Jansen, R., Mohr, K. & Karwehl, S. *Novel Sorangicin Antibiotc. EP3498714*. Available at https://data.epo.org/publication-server/document?iDocId=5956084&iFormat=0 (2017).
10. Buntin, K. *et al.* Biosynthesis of thuggacins in myxobacteria: comparative cluster analysis reveals basis for natural product structural diversity. *Chem. Biol.* **17,** 342–356; 10.1016/j.chembiol.2010.02.013 (2010).

11. Steinmetz, H. *et al.* Thuggacins, macrolide antibiotics active against *Mycobacterium tuberculosis*: Isolation from myxobacteria, structure elucidation, conformation analysis and biosynthesis. *Chem. Eur. J.* **13,** 5822–5832; 10.1002/chem.200700269 (2007).

12. Pogorevc, D. *et al.* Biosynthesis and Heterologous Production of Argyrins. *ACS Synth. Biol.* **8,** 1121–1133; 10.1021/acssynbio.9b00023 (2019).

13. Sasse, F. *et al.* Argyrins, immunosuppressive cyclic peptides from myxobacteria. I. Production, isolation, physico-chemical and biological properties. *J. Antibiot.* **55,** 543–551; 10.7164/antibiotics.55.543 (2002).

14. Pogorevc, D. & Müller, R. Biotechnological production optimization of argyrins – a potent immunomodulatory natural product class. *Microb. Biotechnol.*; 10.1111/1751-7915.13959 (2021).

15. Gokarn, K. & Pal, R. B. Activity of siderophores against drug-resistant Gram-positive and Gram-negative bacteria. *Infect. Drug. Resist.* **11,** 61–75; 10.2147/IDR.S148602 (2018).

16. Martínez-Fructuoso, L. *et al.* Screen for New Antimicrobial Natural Products from the NCI Program for Natural Product Discovery Prefractionated Extract Library. *ACS Infect. Dis.* **9,** 1245–1256; 10.1021/acsinfecdis.3c00067 (2023).

17. Caesar, L. K. & Cech, N. B. Synergy and antagonism in natural product extracts: when 1 + 1 does not equal 2. *Nat. Prod. Rep.* **36,** 869–888; 10.1039/c9np00011a (2019).

18. Caesar, L. K., Kellogg, J. J., Kvalheim, O. M. & Cech, N. B. Opportunities and limitations for untargeted mass spectrometry metabolomics to identify biologically active constituents in complex natural product mixtures. *J. Nat. Prod.* **82,** 469–484; 10.1021/acs.jnatprod.9b00176 (2019).

19. Cushnie, T. P. T. *et al.* Bioprospecting for Antibacterial Drugs: a Multidisciplinary Perspective on Natural Product Source Material, Bioassay Selection and Avoidable Pitfalls. *Pharm Res* **37,** 125; 10.1007/s11095-020-02849-1 (2020).

20. Atanasov, A. G., Zotchev, S. B., Dirsch, V. M. & Supuran, C. T. Natural products in drug discovery: advances and opportunities. *Nat. Rev. Drug Discov.*; 10.1038/s41573-020-00114-z (2021).

21. Grabowski, K. & Schneider, G. Properties and Architecture of Drugs and Natural Products Revisited. *Current Chemical Biology* **1,** 115–127; 10.2174/187231307779814066 (2007).

22. Bader, C. D. *et al.* Sandacrabins - Structurally Unique Antiviral RNA Polymerase Inhibitors from a Rare Myxobacterium. *Chemistry – A European Journal* **28,** e202104484; 10.1002/chem.202104484 (2022).

23. Haack, P. A. *et al.* Thiamyxins: Structure and Biosynthesis of Myxobacterial RNA-Virus Inhibitors. *Angew. Chem. Int. Ed.* **61,** e202212946; 10.1002/anie.202212946 (2022).

24. Katz, L. & Baltz, R. H. Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **43,** 155–176; 10.1007/s10295-015-1723-5 (2016).

25. Wang, J., Wang, J., Wu, S., Zhang, Z. & Li, Y. Global Geographic Diversity and Distribution of the Myxobacteria. *Microbiology Spectrum* **9,** e0001221; 10.1128/Spectrum.00012-21 (2021).

26. Petters, S. *et al.* The soil microbial food web revisited: Predatory myxobacteria as keystone taxa? *ISME J.*; 10.1038/s41396-021-00958-2 (2021).

27. Mohr, K. I. Diversity of Myxobacteria-We Only See the Tip of the Iceberg. *Microorganisms* **6**; 10.3390/microorganisms6030084 (2018).

28. Landwehr, W., Wolf, C. & Wink, J. Actinobacteria and Myxobacteria-Two of the Most Important Bacterial Resources for Novel Antibiotics. *Curr. Top. Microbiol. Immunol.*; 10.1007/82_2016_503 (2016).

29. Hoffmann, T. *et al.* Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. *Nat. Commun.* **9,** 803; 10.1038/s41467-018-03184-1 (2018).

30. Gavriilidou, A. *et al.* Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nat. Microbiol.* **7,** 726–735; 10.1038/s41564-022-01110-2 (2022).

31. Walesch, S. *et al.* Fighting antibiotic resistance-strategies and (pre)clinical developments to find new antibacterials. *EMBO Rep.,* e56033; 10.15252/embr.202256033 (2022).

32. Chai, Y. *et al.* Discovery of 23 natural tubulysins from *Angiococcus disciformis* An d48 and *Cystobacter* SBCb004. *Chem. Biol.* **17,** 296–309; 10.1016/j.chembiol.2010.01.016 (2010).

33. 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* **1445,** 185–195 (1999).

34. Garcia, R. *et al.* Pendulasporaceae – a unique myxobacterial family with distinct sporulation behaviour and high potential for natural product discovery **Manuscript in preparation** (2023).

35. Neuber, M. Alteration of the metabolite spectrum of myxobacteria through alternative cultivation and extraction techniques. Doctoral Thesis. Saarland University, 2022.

36. Panter, F., Bader, C. D. & Müller, R. The Sandarazols are Cryptic and Structurally Unique Plasmid-Encoded Toxins from a Rare Myxobacterium*. *Angewandte Chemie (International ed. in English)* **60,** 8081–8088; 10.1002/anie.202014671 (2021).

37. Amos, G. C. A. *et al.* Comparative transcriptomics as a guide to natural product discovery and biosynthetic gene cluster functionality. *Proc. Natl. Acad. Sci. USA* **4,** 201714381; 10.1073/pnas.1714381115 (2017).

38. Schley, C., Altmeyer, M. O., Swart, R., Müller, R. & Huber, C. G. Proteome analysis of *Myxococcus xanthus* by off-line two-dimensional chromatographic separation using monolithic poly-(styrene-divinylbenzene) columns combined with ion-trap tandem mass spectrometry. *J. Proteome Res.* **5,** 2760–2768; 10.1021/pr0602489 (2006).

39. Covington, B. C., Xu, F. & Seyedsayamdost, M. R. A Natural Product Chemist's Guide to Unlocking Silent Biosynthetic Gene Clusters. *Annu. Rev. Biochem.* **90,** 763–788; 10.1146/annurev-biochem-081420-102432 (2021).

40. Frank, N. A. Advancing myxobacterial natural product discovery by combining genome and metabolome mining with organic synthesis. Doctoral Thesis. Saarland University, 2022.

41. Bader, C. D., Neuber, M., Panter, F., Krug, D. & Müller, R. Supercritical Fluid Extraction Enhances Discovery of Secondary Metabolites from Myxobacteria. *Anal. Chem.* **92,** 15403–15411; 10.1021/acs.analchem.0c02995 (2020).

42. 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.*; 10.1021/acs.jnatprod.0c00942 (2021).

43. Bode, H. B., Bethe, B., Höfs, R. & Zeeck, A. Big effects from Small Changes: Possible Ways to Explore Nature's Chemical Diversity. *ChemBioChem* **3,** 619–627 (2002).

44. Schiewe, H. J. & Zeeck, A. Cineromycins, γ-Butyrolactones and Ansamycins by Analysis of the Secondary Metabolite Pattern Created by a Single Strain of *Streptomyces*. *J Antibiot* **52,** 635–642; 10.7164/antibiotics.52.635 (1999).

45. Scherlach, K. & Hertweck, C. Triggering cryptic natural product biosynthesis in microorganisms. *Org. Biomol. Chem.* **7,** 1753–1760; 10.1039/b821578b (2009).

46. 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.* **10,** 294; 10.3389/fmicb.2019.00294 (2019).

47. Walt, C., Bader, C. D., Krug, D. & Müller, R. Aggravated Cultivation of Myxobacteria Stimulates Secondary Metabolism. *(Manuscript in preparation)*.

48. 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* **9,** 535–542; 10.3390/md9040535 (2011).

49. Okada, B. K. & Seyedsayamdost, M. R. Antibiotic dialogues: induction of silent biosynthetic gene clusters by exogenous small molecules. *FEMS Microbiol. Rev.* **41,** 19–33; 10.1093/femsre/fuw035 (2017).

50. Adnani, N. *et al.* Coculture of Marine Invertebrate-Associated Bacteria and Interdisciplinary Technologies Enable Biosynthesis and Discovery of a New Antibiotic, Keyicin. *ACS Chem. Biol.* **12,** 3093–3102; 10.1021/acschembio.7b00688 (2017).

51. Pishchany, G. *et al.* Amycomicin is a potent and specific antibiotic discovered with a targeted interaction screen. *Proceedings of the National Academy of Sciences of the United States of America* **115,** 10124–10129; 10.1073/pnas.1807613115 (2018).

52. Onaka, H., Mori, Y., Igarashi, Y. & Furumai, T. Mycolic acid-containing bacteria induce natural-product biosynthesis in *Streptomyces* species. *Appl. Environ. Microbiol.* **77,** 400–406; 10.1128/AEM.01337-10 (2011).

53. Onaka, H. *et al.* Mycolic acid-containing bacteria activate heterologous secondary metabolite expression in *Streptomyces lividans*. *J. Antibiot.* **68,** 594–597; 10.1038/ja.2015.31 (2015).

54. Li, P. F. *et al.* Co-cultivation of *Sorangium cellulosum* strains affects cellular growth and biosynthesis of secondary metabolite epothilones. *FEMS Microbiol. Ecol.* **85,** 358–368; 10.1111/1574-6941.12125 (2013).

55. Bertrand, S. *et al.* Metabolite induction via microorganism co-culture: a potential way to enhance chemical diversity for drug discovery. *Biotechnol. Adv.* **32,** 1180–1204; 10.1016/j.biotechadv.2014.03.001 (2014).

56. Rutledge, P. J. & Challis, G. L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **13,** 509–523; 10.1038/nrmicro3496 (2015).

57. Zhang, X., Hindra & Elliot, M. A. Unlocking the trove of metabolic treasures: activating silent biosynthetic gene clusters in bacteria and fungi. *Curr. Opin. Microbiol.* **51,** 9–15; 10.1016/j.mib.2019.03.003 (2019).

58. Shima, J., Hesketh, A., Okamoto, S., Kawamoto, S. & Ochi, K. Induction of actinorhodin production by rpsL (encoding ribosomal protein S12) mutations that confer streptomycin resistance in Streptomyces lividans and Streptomyces coelicolor A3(2). *Journal of bacteriology* **178,** 7276–7284; 10.1128/jb.178.24.7276-7284.1996 (1996).

59. Ochi, K. *et al.* Ribosome Engineering and Secondary Metabolite Production (Elsevier2004), Vol. 56, pp. 155–184.

60. Craney, A., Ozimok, C., Pimentel-Elardo, S. M., Capretta, A. & Nodwell, J. R. Chemical perturbation of secondary metabolism demonstrates important links to primary metabolism. *Chem. Biol.* **19,** 1020–1027; 10.1016/j.chembiol.2012.06.013 (2012).

61. Seyedsayamdost, M. R. High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *Proc. Natl. Acad. Sci. USA* **111,** 7266–7271; 10.1073/pnas.1400019111 (2014).

62. Moon, K., Xu, F., Zhang, C. & Seyedsayamdost, M. R. Bioactivity-HiTES Unveils Cryptic Antibiotics Encoded in Actinomycete Bacteria. *ACS Chem. Biol.* **14,** 767–774; 10.1021/acschembio.9b00049 (2019).

63. Lee, S. R. & Seyedsayamdost, M. R. Induction of Diverse Cryptic Fungal Metabolites by Steroids and Channel Blockers. *Angewandte Chemie (International ed. in English)* **61,** e202204519; 10.1002/anie.202204519 (2022).

64. Okada, B. K., Wu, Y., Mao, D., Bushin, L. B. & Seyedsayamdost. Mapping the Trimethoprim-Induced Secondary Metabolome of Burkholderia thailandensis. *ACS Chem. Biol.* **11,** 2124–2130; 10.1021/acschembio.6b00447 (2016).

65. Mccurdy, H. D. & MacRae, T. H. Xanthacin. A bacteriocin of Myxococcus xanthus fb. *Can. J. Microbiol.* **20,** 131–135; 10.1139/m74-021 (1974).

66. Costa, M. S., Clark, C. M., Ómarsdóttir, S., Sanchez, L. M. & Murphy, B. T. Minimizing Taxonomic and Natural Product Redundancy in Microbial Libraries Using MALDI-TOF MS and

the Bioinformatics Pipeline IDBac. *J. Nat. Prod.* **82,** 2167–2173; 10.1021/acs.jnatprod.9b00168 (2019).

67. Spraker, J. E., Luu, G. T. & Sanchez, L. M. Imaging mass spectrometry for natural products discovery: a review of ionization methods. *Nat. Prod. Rep.*; 10.1039/c9np00038k (2019).

68. Bode, H. B. & Müller, R. Secondary metabolism in myxobacteria. In *Myxobacteria: Multicellularity and differentiation,* edited by D. Whitworth (ASM Press, Chicago, 2007), pp. 259–282.

69. Chiang, Y.-M., Chang, S.-L., Oakley, B. R. & Wang, C. C. C. Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms. *Current opinion in chemical biology* **15,** 137–143; 10.1016/j.cbpa.2010.10.011 (2011).

70. Goldman, B. S. *et al.* Evolution of sensory complexity recorded in a myxobacterial genome. *Proc. Nat. Acad. Sci. USA* **103,** 15200–15205; 10.1073/pnas.0607335103 (2006).

71. Bader, C. D., Panter, F. & Müller, R. In depth natural product discovery - Myxobacterial strains that provided multiple secondary metabolites. *Biotechnol. Adv.* **39,** 107480; 10.1016/j.biotechadv.2019.107480 (2020).

72. Yang, Y.-J. *et al.* Genome Editing in Model Strain *Myxococcus xanthus* DK1622 by a Site-Specific Cre/loxP Recombination System. *Biomolecules* **8,** 137; 10.3390/biom8040137 (2018).

73. Bauman, K. D., Butler, K. S., Moore, B. S. & Chekan, J. R. Genome mining methods to discover bioactive natural products. *Nat. Prod. Rep.* **38,** 2100–2129; 10.1039/D1NP00032B (2021).

74. Shen, B. *et al.* Enediynes: Exploration of microbial genomics to discover new anticancer drug leads. *Bioorganic & Medicinal Chemistry Letters* **25,** 9–15; 10.1016/j.bmcl.2014.11.019 (2015).

75. Heine, D., Martin, K. & Hertweck, C. Genomics-guided discovery of endophenazines from Kitasatospora sp. HKI 714. *J. Nat. Prod.* **77,** 1083–1087; 10.1021/np400915p (2014).

76. Panter, F., Bader, C. D. & Müller, R. Synergizing the potential of bacterial genomics and metabolomics to find novel antibiotics. *Chem. Sci.,* 5994–6010; 10.1039/D0SC06919A (2021).

77. Tang, X. *et al.* Identification of Thiotetronic Acid Antibiotic Biosynthetic Pathways by Target-directed Genome Mining. *ACS Chem. Biol.* **10,** 2841–2849; 10.1021/acschembio.5b00658 (2015).

78. Yamanaka, K. *et al.* Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. USA* **111,** 1957–1962; 10.1073/pnas.1319584111 (2014).

79. Panter, F., Krug, D., Baumann, S. & Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **9,** 4898–4908; 10.1039/C8SC01325J (2018).

80. Pedro, L. & Quinn, R. J. Native Mass Spectrometry in Fragment-Based Drug Discovery. *Molecules* **21,** 984; 10.3390/molecules21080984 (2016).

81. Farha, M. A., French, S., Stokes, J. M. & Brown, E. D. Bicarbonate Alters Bacterial Susceptibility to Antibiotics by Targeting the Proton Motive Force. *ACS Infect. Dis.* **4,** 382–390; 10.1021/acsinfecdis.7b00194 (2018).

82. Hennessen, F. *et al.* Amidochelocardin Overcomes Resistance Mechanisms Exerted on Tetracyclines and Natural Chelocardin. *Antibiotics* **9**; 10.3390/antibiotics9090619 (2020).

83. Ersoy, S. C. *et al.* Correcting a Fundamental Flaw in the Paradigm for Antimicrobial Susceptibility Testing. *EBioMedicine* **20,** 173–181; 10.1016/j.ebiom.2017.05.026 (2017).

84. Gao, W. *et al.* Quantitative analysis of steroid hormones in human hair using a column-switching LC-APCI-MS/MS assay. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **928,** 1–8; 10.1016/j.jchromb.2013.03.008 (2013).

85. Masike, K., Stander, M. A. & Villiers, A. de. Recent applications of ion mobility spectrometry in natural product research. *1873-264X* **195,** 113846; 10.1016/j.jpba.2020.113846 (2021).

86. Schlotterbeck, G. & Ceccarelli, S. M. LC–SPE–NMR–MS: a total analysis system for bioanalysis. *Bioanalysis* **1,** 549–559; 10.4155/bio.09.50 (2009).

87. Ellis, B. M., Fischer, C. N., Martin, L. B., Bachmann, B. O. & McLean, J. A. Spatiochemically Profiling Microbial Interactions with Membrane Scaffolded Desorption Electrospray Ionization-Ion Mobility- Imaging Mass Spectrometry and Unsupervised Segmentation. *Anal. Chem.*; 10.1021/acs.analchem.9b02992 (2019).

88. Gibitz Eisath, N., Sturm, S. & Stuppner, H. Supercritical Fluid Chromatography in Natural Product Analysis - An Update. *Planta Med.*; 10.1055/s-0037-1599461 (2017).

89. Pauli, G. F., Chen, S.-N., Friesen, J. B., McAlpine, J. B. & Jaki, B. U. Analysis and Purification of Bioactive Natural Products: The AnaPurNa Study. *Journal of natural products* **75,** 1243–1255; 10.1021/np300066q (2012).

90. Inui, T., Wang, Y., Pro, S. M., Franzblau, S. G. & Pauli, G. F. Unbiased evaluation of bioactive secondary metabolites in complex matrices. *Fitoterapia* **83,** 1218–1225; 10.1016/j.fitote.2012.06.012 (2012).

91. Berthod, A. & Faure, K. Separations with a Liquid Stationary Phase: Countercurrent Chromatography or Centrifugal Partition Chromatography. In *Analytical separation science,* edited by J. L. Anderson, A. Berthod, V. Pino & A. Stalcup (Wiley-VCH, Weinheim, 2015), pp. 1177–1206.

92. Wolfender, J.-L., Marti, G., Thomas, A. & Bertrand, S. Current approaches and challenges for the metabolite profiling of complex natural extracts. *Journal of chromatography. A* **1382,** 136–164; 10.1016/j.chroma.2014.10.091 (2015).

93. Yi, L. *et al.* Chemometric methods in data processing of mass spectrometry-based metabolomics: A review. *Anal. Chim. Acta* **914,** 17–34; 10.1016/j.aca.2016.02.001 (2016).

94. Lee, S. *et al.* NP Analyst: An Open Online Platform for Compound Activity Mapping. *ACS Cent. Sci.* **8,** 223–234; 10.1021/acscentsci.1c01108 (2022).

95. Nothias, L.-F. *et al.* Bioactivity-Based Molecular Networking for the Discovery of Drug Leads in Natural Product Bioassay-Guided Fractionation. *J. Nat. Prod.* **81,** 758–767; 10.1021/acs.jnatprod.7b00737 (2018).

96. Kuhnert, N. *et al.* Scope and limitations of principal component analysis of high resolution LC-TOF-MS data: the analysis of the chlorogenic acid fraction in green coffee beans as a case study. *Anal. Methods* **3,** 144–155; 10.1039/c0ay00512f (2011).

97. Avalon, N. E., Murray, A. E. & Baker, B. J. Integrated Metabolomic-Genomic Workflows Accelerate Microbial Natural Product Discovery. *Analytical chemistry* **94,** 11959–11966; 10.1021/acs.analchem.2c02245 (2022).

98. Behsaz, B. *et al.* Integrating genomics and metabolomics for scalable non-ribosomal peptide discovery. *Nat Commun* **12,** 3225; 10.1038/s41467-021-23502-4 (2021).

99. Leão, T. F. *et al.* NPOmix: A machine learning classifier to connect mass spectrometry fragmentation data to biosynthetic gene clusters. *PNAS Nexus* **1,** pgac257; 10.1093/pnasnexus/pgac257 (2022).

100. Boldt, J. *et al. Bursts in biosynthetic gene cluster transcription are accompanied by surges of natural compound production in the myxobacterium Sorangium sp* (2022).

101. Bouslimani, A., Sanchez, L. M., Garg, N. & Dorrestein, P. C. Mass spectrometry of natural products: current, emerging and future technologies. *Nat. Prod. Rep.* **31,** 718–729; 10.1039/c4np00044g (2014).

102. Navarro-Muñoz, J. C. *et al.* A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **16,** 60–68; 10.1038/s41589-019-0400-9 (2020).

103. Ziemert, N. *et al.* The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS ONE* **7,** e34064; 10.1371/journal.pone.0034064 (2012).

104. Tsao, S.-W., Rudd, B. A. M., He, X.-G., Chang, C.-J. & Floss, H. G. Identification of a red pigment from *Streptomyces coelicolor* A3(2) as a mixture of prodigiosin derivatives. *J. Antibiot.* **38,** 128–131; 10.7164/antibiotics.38.128 (1985).

105. Ramaprasad, E. V. V., Bharti, D., Sasikala, C. & Ramana, C. V. Zooshikella marina sp. nov. a cycloprodigiosin- and prodigiosin-producing marine bacterium isolated from beach sand. *Int. J. Syst. Evol. Microbiol.* **65,** 4669–4673; 10.1099/ijsem.0.000630 (2015).

106. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **46,** D8-D13; 10.1093/nar/gkx1095 (2018).

107. Blin, K. *et al.* antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.*; 10.1093/nar/gkad344 (2023).

108. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410; 10.1016/S0022-2836(05)80360-2 (1990).

109. Terlouw, B. R. *et al.* MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res.* **51,** D603-D610; 10.1093/nar/gkac1049 (2023).

110. 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.* **31,** 768–783; 10.1039/c3np70127a (2014).

111. Linington, R. G. Npatlas - The Natural Products Atlas. Available at https://www.npatlas.org (2018).

112. Sud, M. *et al.* LMSD: LIPID MAPS structure database. *Nucleic Acids Res.* **35,** D527-32; 10.1093/nar/gkl838 (2007).

113. CRC Press. Dictionary of Natural Products 30.2. Available at https://dnp.chemnetbase.com (2021).

114. Nothias, L.-F. *et al.* Feature-based molecular networking in the GNPS analysis environment. *Nat Methods* **17,** 905–908; 10.1038/s41592-020-0933-6 (2020).

115. Ernst, M. *et al.* MolNetEnhancer: Enhanced Molecular Networks by Integrating Metabolome Mining and Annotation Tools. *Metabolites* **9**; 10.3390/metabo9070144 (2019).

116. Zhang, C. *et al.* Small molecule accurate recognition technology (SMART) to enhance natural products research. *Scientific Reports* **7,** 14243; 10.1038/s41598-017-13923-x (2017).

117. Reher, R. *et al.* A Convolutional Neural Network-Based Approach for the Rapid Characterization of Molecularly Diverse Natural Products. *J. Am. Chem. Soc.*; 10.1021/jacs.9b13786 (2020).

118. Wishart, D. S. *et al.* NP-MRD: the Natural Products Magnetic Resonance Database. *Nucleic Acids Res* **50,** D665-D677; 10.1093/nar/gkab1052 (2022).

119. Wang, D.-G. *et al.* Constructing a Myxobacterial Natural Product Database to Facilitate NMR-Based Metabolomics Bioprospecting of Myxobacteria. *Analytical chemistry* **95,** 5256–5266; 10.1021/acs.analchem.2c05145 (2023).

120. Egan, J. M., van Santen, J. A., Liu, D. Y. & Linington, R. G. Development of an NMR-Based Platform for the Direct Structural Annotation of Complex Natural Products Mixtures. *Journal of natural products*; 10.1021/acs.jnatprod.0c01076 (2021).