

# **Exploiting the biosynthetic potential of myxobacteria for natural product discovery**

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

von

**Alexander Popoff**

Saarbrücken 2020



Tag des Kolloquiums: 15.06.2020

Dekan: Prof. Dr. Guido Kickelbick

Berichterstatter: Prof. Dr. Rolf Müller  
Prof. Dr. Uli Kazmaier

Vorsitz: Prof. Dr. Andriy Luzhetskyy

Akad. Mitarbeiter: Dr. Josef Zapp

Die vorliegende Arbeit wurde von Oktober 2015 bis März 2020 unter Anleitung von Herrn Prof. Dr. Rolf Müller am Helmholtz-Institut für Pharmazeutische Forschung Saarland (HIPS) angefertigt.

Ich bin ein Zwerg auf den Schultern eines Riesen.

frei nach Bernhard von Chartres

## **Danksagungen**

Hiermit möchte ich allen danken, die zur Erstellung dieser Arbeit beigetragen haben. An erster Stelle bedanke ich mich bei meinem Doktorvater Prof. Dr. Rolf Müller, für das in mich gesetzte Vertrauen, die Bereitschaft mich in seiner Arbeitsgruppe aufzunehmen und für die Impulse, die er mir gegeben hat. Des Weiteren möchte ich mich bei Prof. Dr. Ullrich Kazmaier für die Annahme des Co-Referats und Unterstützung als wissenschaftlicher Begleiter bedanken.

Auch meinen Eltern gilt mein Dank, denn ohne deren Unterstützung wäre ich niemals so weit gekommen. Mein besonderer Dank gilt meinen Kollegen Dr. Lena Keller, Dr. Chengzhang Fu und Dr. Bastien Schnell für das Korrekturlesen dieser Arbeit. Eure konstruktive Kritik hat mir sehr geholfen Ordnung in das Chaos dieser Promotion zu bringen.

Ein weiterer Dank gilt meiner ehemaligen Betreuerin Dr. Kirsten Harmrolfs, aber auch den ehemaligen Kollegen Dr. Louise Kjaerulff, Dr. Tony Abou Fayad und Daniel Sauer, die mir vieles gezeigt haben und von denen ich essentielle Techniken erlernen konnte.

Natürlich gilt mein Dank auch meinen anderen Kollegen, insbesondere Maja Remškar für die Kooperation in den Projekten, ohne die es sicherlich nicht diese Ergebnisse gegeben hätte. Auch Joachim Hug möchte ich für die sehr kreativen, oftmals auch kontroversen Gesprächsthemen insbesondere während der Mittagspausen danken. Natürlich möchte ich auch meinen anderen Kollegen Joy Birkelbach, Dr. Domen Pogorevc, Asfandiyar Sikandar, Jan Dastbaz, Dr. Sebastian Adam, Patrick Haack, Sebastian Groß, Sari Rasheed, Nicolas Frank und allen anderen die ich vergessen habe für die Unterstützung, aber auch für das freundschaftliche Miteinander und für die gemeinsamen Unternehmungen danken.

Auch möchten ich meinen Büromitinsassen Hu Zeng, Sebastian Walesch für die gemeinsame Zeit und Markus Neuber zusätzlich noch für die Versorgung mit dem koffeinhaltigen Lebenselixier danken.

## **Abstract**

Myxobacteria are one of the most prolific sources for novel natural products with promising bioactivity against microorganisms or cancer cell lines. This thesis covers the discovery of three new compound classes, the myxoquaterines, the myxoglucamides and the myxopentacin family, the process of the structure elucidation and investigations towards their biosynthetic pathways. The myxoquaterines were discovered using a bioactivity-guided compound isolation strategy. Modifications including the incorporation of a trimethylammonium ion and a prenyl group clearly mark the myxoquaterines as a different compound class than the previously described DKxanthenes. In addition, the isolated myxoquaterines possess intriguing cytotoxic bioactivity against tested cell lines and moderate antimicrobial activity. A genome mining based approach resulted in the discovery of the myxoglucamides. The myxoglucamides are glycolipopeptides with a glucose sugar attached to a hydroxylated lipid chain with a vinyl group and L-aminobutyric acid linked to the C-terminus. The myxopentacines are a serendipitous finding during a genome mining approach. An untypical aminocyclopentyl-carboxylic acid ring and a rare carbonylmethylene moiety underline the unique chemical scaffold of the myxopentacins. Biosynthesis of all three compound classes were investigated by feeding experiments with stable isotope-labeled precursors of biosynthetic building blocks and *in-silico* analysis of the bacterial genomes.

## Zusammenfassung

Myxobakterien sind eine der vielversprechendsten Quellen für neue Naturstoffe im Kampf gegen pathogene Mikroorganismen und Krebserkrankungen. Im Rahmen dieser Doktorarbeit wurden drei neue Naturstoffklassen entdeckt: Die Myxoquaterine, die Myxoglucamide und die Myxopentacine. Zusätzlich beinhaltet diese Arbeit die Strukturaufklärung und Studien zur Aufklärung der entsprechenden Biosynthesewege. Die Myxoquaterine wurden im Rahmen einer bioaktivitätsgeleiteten Isolierung entdeckt. Diese Substanzklasse weist eine große Ähnlichkeit mit den DKxanthenen auf, jedoch zeichnen Modifikationen, wie ein Trimethylammonium Ion und eine Prenylierung diese Moleküle als neue Naturstoffklasse aus. Alle isolierten Myxoquaterin Derivate haben eine starke zytotoxische Bioaktivität. Mittels eines *genome-mining* Ansatzes wurden die Myxoglucamide, eine Gruppe von Glyko-Lipopeptiden, entdeckt. Charakteristisch für diese Substanzen ist ein Glukose-Rest, der an eine hydroxylierte Fettsäurekette gebunden ist. Am C-Terminus dieser Fettsäurekette befindet sich eine Vinyl-Gruppe und eine L-Aminobuttersäure Einheit. Im Rahmen des *genome-mining* Ansatzes wurden als unerwarteter Nebeneffekt auch die Myxopentacine entdeckt. Strukturelle Merkmale dieser Substanzklasse sind die Kette aus Aminocyclopentansäure Blöcken und die Carbonylmethylengruppe. Mittels Fütterungsexperimenten und *in-silico* Analysen wurde die Biosynthese von allen drei entdeckten Substanzklassen untersucht.

## **Authors contribution to this work**

### Chapter 2

The strain MSr 11954 was isolated and cultivated by Dr. Ronald Garcia and the standard-fractionation was performed by Daniel Sauer. The author isolated all myxoquaterine molecules and elucidated their structures, including the relative configuration. Furthermore, the author attempted a Mosher esterification to determine the absolute configuration of one stereocenter. The biosynthetic gene cluster (BGC) was discovered by Maja Remškar and Dr. Yunkun Liu. The author performed an in-depth *in-silico* analysis of the BGC and developed a biosynthesis hypothesis presented in this thesis. In addition, the author designed and performed all feeding experiments presented in this chapter. Joy Birkelbach re-isolated myxoquaterine molecules for further bioactivity testing experiments, which were performed by Irene Kochems, Victoria Schmitt and Stefanie Schmidt.

### Chapter 3

The strain MCy 9003 was isolated by Dr. Ronald Garcia. Genetic manipulations were performed by Maja Remškar. Cultivations and preparations of the extracts were performed by the author and Maja Remškar. The author analyzed the extracts, isolated the myxoglucamides and elucidated the structure. Furthermore, the author performed all experiments for determination of the stereochemistry. In addition, the author performed an in-depth *in-silico* analysis of the BGC and developed a biosynthesis hypothesis presented in this thesis. Feeding experiments were designed by the author and performed by Maja Remškar and the author. Irene Kochems, Victoria Schmitt and Stefanie Schmidt performed bioactivity testing.

### Chapter 4

The author identified the myxopentacin molecules, isolated them and elucidated the structure. Furthermore, the author performed all experiments for determination of the stereochemistry. The BGC was discovered by Maja Remškar and confirmed by creation of an inactivation mutant. The author performed an in-depth *in-silico* analysis of the BGC and developed the biosynthesis hypothesis presented in this thesis. Feeding experiments were designed and performed by the author. Bioactivity testing was performed by Irene Kochems, Victoria Schmitt and Stefanie Schmidt.

## Veröffentlichungen

Teile dieser Arbeit sind derzeit in Vorbereitung zu folgenden Veröffentlichungen:

**Popoff A.\***; Garcia R.\*; Walt C.; Bader C.; Birkelbach J.; Liu Y.; Zaburanyi N.; Nübel Z.; Steglich M.; Overmann J. and Müller R.; Exploring Myxobacterial diversity for novel natural products, to be published

**Popoff A.\***; Remškar M.\*; Garcia R. and Müller R.; Myxoglucamides, a novel myxobacterial glycolipopeptide, discovered by genome mining, to be published

**Popoff A.\***; Remškar M.\*; Hofer W.; Garcia R. and Müller R.; The Myxopentacin family, a group of exceptional peptides, to be published

\*these authors contributed equally

Folgende Publikationen sind während des Zeitraums dieser Arbeit entstanden, jedoch nicht Bestandteil dieser Dissertation:

Yan, F.; Burgard, C.; **Popoff, A.**; Zaburanyi, N.; Zipf, G.; Maier, J.; Bernauer, H. S.; Wenzel, S. C.; Müller, R. (2018) Synthetic biology approaches and combinatorial biosynthesis towards heterologous lipopeptide production, *Chemical. Science* 9, 7510-7519

Pogorevc, D.; Tang, Y.; Hoffmann, M.; Zipf, G.; Bernauer, H. S.; **Popoff, A.**; Steinmetz, H.; Wenzel, S. C. (2019) Biosynthesis and heterologous production of argyrins.

## Conference Contributions (Posters and oral presentations)

**Alexander Popoff**, Maja Remškar, Ronald Garcia, Nestor Zaburanyi and Rolf Müller; Genome mining revealed a new myxobacterial glycolipopeptide containing non-proteinogenic amino acids **Poster presentation** International VAAM Workshop 2019; Jena, Germany

**Alexander Popoff**; Genome mining revealed a new myxobacterial glycolipopeptide **Oral presentation** Summer Symposium of Interdisciplinary Graduate School, 2019; Saarbrücken, Germany.

**Alexander Popoff**; Search for new secondary metabolites in myxobacteria **Oral presentation** Summer Symposium of Interdisciplinary Graduate School, 2016; Saarbrücken, Germany.

**Alexander Popoff**, Maja Remškar, Ronald Garcia, Nestor Zaburanyi and Rolf Müller; Genome mining revealed a new myxobacterial glycolipopeptide containing non-proteinogenic amino acids **Poster presentation** 9<sup>th</sup> HIPS Symposium 2019; Saarbrücken, Germany

**Alexander Popoff**, Ronald Garcia, Maja Remškar, Rolf Müller New DKxanthen-like secondary metabolite from a myxobacterium shows antimicrobial and cytotoxic activity **Poster presentation** International VAAM Workshop 2018; Frankfurt, Germany

**Alexander Popoff**, Ronald Garcia, Maja Remškar, Rolf Müller New DKxanthen-like secondary metabolite from a myxobacterium shows antimicrobial and cytotoxic activity **Poster presentation** 8<sup>th</sup> HIPS Symposium 2018; Saarbrücken, Germany

# Contents

Chapter 1 Introduction .....	13
1.1 Nature as source of pharmaceuticals.....	13
1.2 Bacteria as sources for pharmaceuticals.....	13
1.3 Current role of natural products for the pharma industry .....	14
1.4 Non ribosomal peptides .....	15
1.5 Other peptide type natural products .....	19
1.6 Polyketides .....	20
1.7 Terpenes.....	22
1.8 From a microorganism to the drug .....	24
1.8.1 Accessing new natural products .....	24
1.8.2 Structure elucidation of natural products .....	29
1.9 Outline of this work.....	38
1.10 References.....	39
Chapter 2 Myxoquaterines.....	43
2.1 Activity guided discovery .....	43
2.2 Structure elucidation.....	45
2.2.1 Difference between variant A and B.....	51
2.2.2 Stereochemical analysis.....	51
2.3 Biosynthesis of myxoquaterines .....	53
2.3.1 Implications from biosynthesis towards stereochemistry.....	57
2.4 Bioactivity .....	58
2.5 Conclusions .....	59
2.6 References.....	59
2.8 Supplementary information .....	61
2.8.1 Experimental procedures.....	61
2.8.2 Supplementary data/additional experimental results .....	66
Chapter 3– Myxoglucamide.....	113
3.1 Genome mining for compound discovery.....	113
3.1.1 Identification of the myxoglucamides .....	114
3.2 Structure elucidation of Myxoglucamides .....	116
3.2.1 Stereochemical assignment.....	121
3.3 Biosynthesis of myxoglucamides .....	123
3.4 Bioactivity .....	127
3.5 Conclusions .....	127
3.6 References.....	127
3.7 Supplementary information.....	129

3.7.1 Experimental procedures.....	129
3.7.2 Supplementary data/additional experimental data .....	133
Chapter 4 – Myxopentacin .....	174
4.1 Discovery of the Myxopentacin family.....	174
4.2 Structure elucidation of the Myxopentacins .....	175
4.2.1 Stereochemistry .....	180
4.3 Biosynthesis.....	181
4.3.1 ACPC moiety .....	181
4.3.1 AGOOA moiety.....	187
4.3.3 Identification and analysis of the myxopentacin gene cluster .....	188
4.4 Identification of additional myxopentacin variants .....	192
4.5 Bioactivity .....	194
4.6 Conclusions .....	194
4.7 References.....	195
4.8 Supplementary information .....	196
4.8.1 Experimental procedures.....	196
Discussion .....	247
Evaluation of activity-guided compound isolation vs. genome mining .....	247
Considerations regarding Biosynthesis .....	252
Elucidation of the biosynthesis – many questions – few answers .....	252
PLP catalyzed condensation-considerations regarding the resulting stereochemistry.....	255
Function of natural products .....	257
Final words .....	261
References.....	261

# Chapter 1 Introduction

## 1.1 Nature as source of pharmaceuticals

The concept of using plants or fungi as pharmaceuticals might be as old as mankind itself. The oldest evidence for the use of natural pharmaceuticals already in the Neolithic Period was found together with the Iceman "Ötzi", who was discovered 1991 in the eastern Alps. Together with the mummified body, a leather bag with remains of the birch fungus *Fomitopsis betulina* was found, which is not known to be a typical meal but offers antibacterial effects<sup>[1]</sup>. Throughout early history, the medical use of plants is described in documents like the Jewish Talmud and the Chinese Pen T`Sao written by the Emperor Shen Nung 2500 BC. Even Homer mentioned plants from the genus *Artemisia* and their positive effects on health.<sup>[2]</sup>

Usage of natural pharmaceuticals was and still is not only limited to the human species. Animals too are known to consume plants for other purposes than nutrition. This so-called self-medication is a behavior well known for example for dogs eating grass to provoke vomiting. An interesting observation of self-medication was made by the Anthropologist Barbara Fruth observing Bonobos swallowing leaves of *Manniophytun fulvum* proposing that the primates use this plant to treat parasite infections<sup>[3]</sup>.

Without knowledge of the composition of plants and fungi, their usage was only based on empirical observations until advances in chemistry enabled scientists to isolate pure compounds from plants. A milestone was the isolation of the alkaloid morphine by Friedrich Sertürner between 1803-5 followed by the isolation of strychnine in 1818 by Pierre Joseph Pelletier and Joseph Bienaimé Caventou. In 1820, the two chemists were also able to isolate quinine from *Cinchona calisaya*, a drug that was used for the treatment of malaria and that still is used for malaria protection.<sup>[4][5]</sup>

A real turning point in the field of pharmacology was the discovery of penicillin from *Penicillium notatum* by Alexander Fleming in 1928.<sup>[6]</sup> His discovery not only provided us with the strongest weapon to fight bacterial infections but it also brought microorganisms in the focus as sources of new pharmaceuticals. This led to the golden age of antibiotics where thousands of new antibiotics were discovered from fungi as well as from bacteria.

## 1.2 Bacteria as sources for pharmaceuticals

Even though the credit for discovering the first antibiotic went to Alexander Fleming, Selman Waksman was the first who systematically started screening microbes for the production of antibiotics. Motivated by the discovery of the antibiotic gramicidin from *Bacillus brevis* by his former student René Dubos in 1939,<sup>[7]</sup> Waksman started to screen soil microbes (mostly bacteria from the subclass *Actinomycetales*) for their antibacterial activity. His first success was the discovery of actinomycin as an antibiotic which later turned out to also offer cytotoxic activity<sup>[8]</sup>. A real breakthrough was the discovery of streptomycin by his PhD student Albert Schatz in 1944 as first antibiotic to treat tuberculosis.<sup>[9]</sup>

From then on, a whole range of natural products with different pharmaceutical properties were isolated from bacteria. Within this domain, actinobacteria (mostly of the genus *Streptomyces*) provide the highest number of isolated natural products from bacteria with 20.000 isolated natural products so far. Approximately 50 % of them were also described to have pharmaceutical properties like cytotoxic or anti-infective activity.

In the last decades, other species gained importance as natural product sources. For example, approximately 1300 natural products were isolated from cyanobacteria between 2001 and 2010. A great number of natural products from cyanobacteria show cytotoxic activity and a few of them are lead molecules for new anti-cancer drugs like dolastatin or curazin.<sup>[10]</sup>

Other significant producers of natural products belong to the genus *Bacillus*. The last resort antibiotic polymyxin is one of the most relevant examples since it is used to treat multiresistant Gram negative bacterial infections<sup>[11]</sup>. Another famous example of a bacillus derived antibiotic is gramicidin mentioned in the beginning of this chapter.

In the last two decades, myxobacteria gained more and more importance as sources for natural products. They offer a highly diverse range of chemical scaffolds containing heteroatoms like chlorine<sup>[12]</sup>. More than 600 natural products were to date discovered from myxobacteria.<sup>[13]</sup>

Myxobacteria are gram negative bacteria, which usually live in soil and were first described by Ronald Thaxter in 1892<sup>[14]</sup>. Myxobacteria belong to the order of *Myxococcales* which is part of the class of *Deltaproteobacteria* and can be divided into three suborders: *Nannocystineae*, *Sorangineae* and *Cystobacterineae*. Based on 16S sequence data these suborders can be divided into 16 families like *Myxococcaceae*.<sup>[15]</sup>

In contrast to many other prokaryotes, myxobacteria can also appear in multicellular forms like fruiting bodies, a fact that led to their first classification as fungi<sup>[16]</sup>. Especially under nutrient-poor conditions, they aggregate to form fruiting bodies. In these cell aggregates, the majority of the cells lyse while the remaining cells turn into spores. In these spores, metabolism is reduced to a minimum and the spores can resist challenging environmental conditions like temperatures up to 55 °C.<sup>[17]</sup>

Myxobacteria commonly appear in very competitive habitats such as soil, which is enriched with other microorganism like fungi and actinomycetes. To be able to compete with their neighbors and for their predatory lifestyle, myxobacteria developed several ways to compete like secreted endopeptidases and glucosaminidases with bacteriolytic activities<sup>[18]</sup>.

Before diving deeper in the field of natural product research, a more specific definition needs to be given of what natural products are and what kind of natural products offer pharmaceutical properties: The term natural product refers to all organic molecules produced by a living organism including nucleic acids and intermediates of glycolysis. Usually they were referred to as primary metabolites. But a small portion of the metabolites are different. They are only produced by certain species and beneficial but not essential for the producer organism. This distinct group of metabolites is called secondary metabolites.<sup>[19]</sup> The majority of natural product derived pharmaceuticals are based on secondary metabolites.

### ***1.3 Current role of natural products for the pharma industry***

Despite the success of natural products as lead scaffolds in drug discovery, most pharmaceutical companies have dramatically reduced or even discontinued their effort in natural products drug discovery programs. Instead, their interest has shifted towards combinatorial chemistry approaches using large compound libraries in combination with high throughput screenings<sup>[20]</sup>. Are natural products outdated as a source for new drugs? The answer is clearly no. Within the newly approved pharmaceuticals coming to the market, natural products or natural product derived compounds are highly represented.<sup>[21]</sup> In 2018, 16 % of all FDA approved drugs were natural products or compounds derived from natural products<sup>[22]</sup>.

Are natural products therefore an equal or superior source for pharmaceuticals compared to combinatorial chemistry?

To provide an answer to this question, it is crucial to compare the building or development process: In combinatorial chemistry approaches, large libraries of small synthetic compounds are screened to bind and inhibit a specific target protein. The compounds are generated from smaller building blocks linked by amide, ester, ether or carbon bonds.<sup>[23]</sup>

Not unlike molecules resulting from combinatorial chemistry approaches, natural products also consist of building blocks linked by amide, ester or carbon bonds. A natural product is the result of an evolutionary process. In contrast to synthetic molecules, natural products are the result of an evolutionary driven and thus unobtainable number of selection cycles. In addition, they are selected based on their ability to act on a target organism instead of an isolated target protein. Furthermore, the evolutionary process itself is expected to be different: While a combinatorial drug is build up based on selection cycles, the selection in natural product evolution takes place on complete molecules rather than on their building blocks. [24]

The high complexity of natural products has fascinated chemists for more than two centuries. The total synthesis of a complex natural product is regarded as supreme discipline among synthetic chemists. This chemical complexity is achieved by highly sophisticated enzymatic assembly chains that select and link the different building blocks. In many cases, tailoring steps follow the chain assembly to modify the building blocks and increase the chemical diversity. Based on their biosynthesis, natural products can be divided into several classes. The most abundant classes produced in bacteria are polyketides, non-ribosomal peptides, post-translationally modified ribosomal peptides and terpenes. Interestingly, hybrid molecules being derived from two different assembly chain classes quite often occur [25]

## ***1.4 Non ribosomal peptides***

Non-ribosomal peptides (NRPs) are a highly variable class of natural products. They consist of amino acids as building blocks and can be linear or cyclic. A subgroup of NRPs are lipopeptides, peptides that contain an alkyl or polyene moiety and usually show a higher hydrophobicity. Examples of different kinds of NRPs including structures with deviations from the typical NRPs biosynthesis logic are given in figure 1.

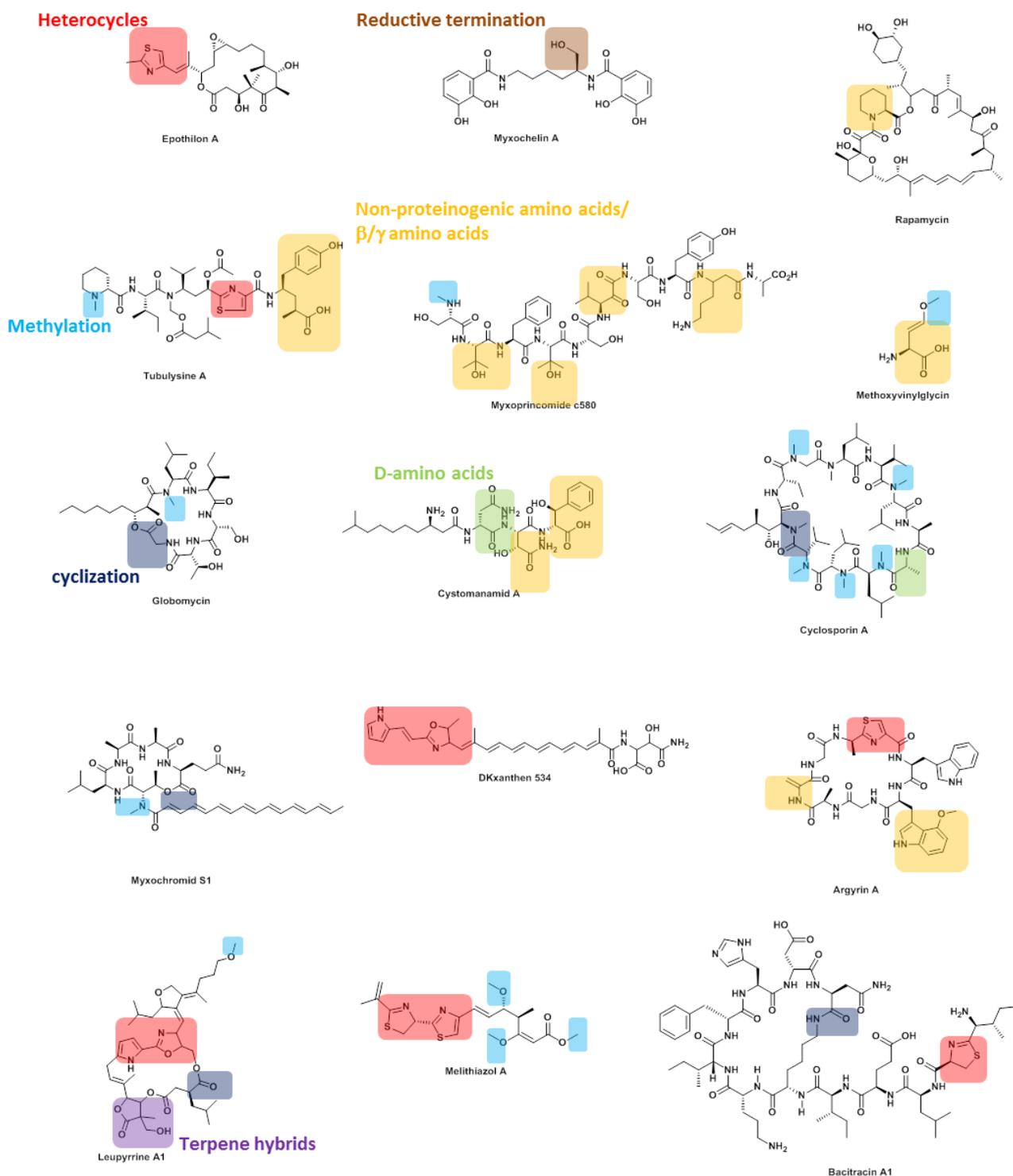


Figure 1: Examples of NRPs or NRP hybrid molecules. Following structural features that represent deviations from classical NRPS biosynthesis logic result in a higher variability of the molecules and are highlighted in different colors: Heterocycles (red); reduced C-terminus (brown); non-proteinogenic amino acids, including  $\beta$ - and  $\gamma$ -amino acids (yellow); D-amino acids (green); N- and O-methylation (blue); different cyclizations (dark blue); terpene hybrids (violet)

## Principles of non-ribosomal peptide biosynthesis

As indicated by the name, NRPs are not synthesized by the ribosome. Instead this group of peptides is synthesized in large modular enzyme complexes called non-ribosomal peptide synthetases (NRPSs).<sup>[26]</sup> Each



*acid of the amino acid under release of AMP. (B) Chain elongation: The PCP bound amino acid 2 on module 2 performs a nucleophilic attack on the carbonyl group of amino acid 1 in module 1. The reaction results in the formation of a peptide bond catalyzed by the C-domain in module 2. The dipeptide is bound to the PCP domain of module 2 and can act as target for further elongation steps from following module bound peptide chains. (C) Termination of the biosynthesis: The serine residue of the TE domain performs a nucleophilic attack on the carbonyl group of the thioester. The thioester bond is cleaved and a new ester is formed with the serine from the TE domain. In a second step the peptide is released as a lactone or carboxylic acid after a second nucleophilic attack from an intramolecular hydroxy-group or from a water molecule*

Termination of the biosynthesis is usually mediated by cleavage of the thioester bond by a thioesterase (TE) domain comparable to polyketides. The Serine performs a nucleophilic attack on the carbonyl group of the thioester forming an intermediate. A second nucleophilic attack of an intramolecular hydroxy-group or a water molecule releases the carbonyl group as cyclic lactone or as carboxylic acid from the serine residue (figure 2C).<sup>[29]</sup>

In some cases, the termination of the biosynthesis is not dependent on the work of a TE: An example for a TE independent release can be found the myxochelin biosynthesis. The PCP bound thioester is reduced by a reductase domain encoded in the last module forming an aldehyde which is further reduced by an aldoreductase into an alcohol. The reduction results in the release from the assembly chain.<sup>[30]</sup>

### **Variations in NRPS biosynthesis**

Characteristic for NRPs is their higher variability compared to ribosomal peptides. This is achieved by cyclization, implementation of D-amino acids, methylation of amide bonds, incorporation of unusual amino acids, heterocyclization and polyketide hybrid elements (figure 1).

Typical cyclic peptides are the depsipeptides, where the C-terminus is cyclized with an intramolecular hydroxy-group forming a lactone. Examples are the myxochromides, where the molecule is cyclized by a condensation between the C-terminal carboxyl group and the alcohol from threonine<sup>[31]</sup> and globomycins, where an ester bridge is formed between a hydroxy group in the polyketide part of this molecule and the carboxyl group of the C-terminal glycine<sup>[32]</sup>. Another type of cyclization occurs between the N- and C-terminus by an amide bond formation, which can be found for example in the cyclic NRP argyirin<sup>[33]</sup>. Also cyclizations between the C-terminus and amine residues within the peptide chain were observed. For example in the antibiotic bacitracin the  $\epsilon$ -NH<sub>2</sub> of the lysine on position 7 forms an amide bond with the C-terminus to give a cycle.<sup>[34]</sup> Many NRPs contain D-amino acids in addition to the proteinogenic amino acids. So called epimerisation domains (E-domains) included in the modules catalyze the conversion of the PCP-domain bound amino acid into its D-amino acid analogue.<sup>[35]</sup> A far less common way of introducing D-Amino acids is the specific recognition by the A-domain like in the cyclosporin biosynthesis<sup>[36]</sup>.

N-methylation like observed in cyclosporine is another common modification of NRPs. The methylation is SAM dependent and catalyzed by a methyl transferase domain inside the module of the corresponding amino acid.<sup>[37]</sup> SAM dependent methylations are not restricted to nitrogen. O-Methylation can as well be found in natural products. A mechanism first described in the biosynthesis of melithiazole from *Melittangum lichenicola*.<sup>[38]</sup>

Typical heterocycles found in NRPs are pyrroles, oxazoles, oxazolines, oxazolidines and their respective sulfur analogues thiazoles variants and methylvariants. Cyclization for all mentioned groups goes in a similar manner. Cysteine, serine or threonine need to be linked by a peptide bond to an amino acid or to an acetate unit in case on a NRPs-PKs hybrid. Heterocyclase domain (HC-domain) are a subtype of C-domains. But instead of a condensation reactions they catalyze the cyclization reactions to form a heterocycle.<sup>[39]</sup> The oxygen or sulfur performs a nucleophilic attack on the carbonyl group forming a tetrahedral intermediate. The oxygen

from the carbonyl group is eliminated to form the oxazoline/ thiazoline ring. These rings can be further reduced or oxidized to form oxazolidines/thiazolidines or aromatic oxazoles/ thiazoles.<sup>[40]</sup>

Beside the 20 proteinogenic amino acids more than 500 amino acids exist in nature. Most of them can also be found in NRPs<sup>[27]</sup>. For example in many NRPs incorporation of amino-butyric acid (Abu) is observed which can be the result of low A-domain specificity like for example in some myxochromides and argyriins that have variants with alanine and Abu.<sup>[41]</sup> In addition also A-domain specific towards non-canonical amino acids like pipercolic acid were described. For example in the biosynthesis of rapamycin where pipercolic acid is incorporated by an NRPs type mechanism<sup>[42]</sup>.

Another option to include non-proteinogenic amino acids inside NRPS scaffold is the modification of a canonical amino acid in the peptide chain or during its integration. For example in the case of methoxyvinylglycine (MVG) glutamate is strongly modified by hydroxylation, decarboxylation and dehydration<sup>[43]</sup>.

Most non-ribosomal peptides found in myxobacteria are non-ribosomal peptide-polyketide (NRPs-PKS) hybrid molecules<sup>[44]</sup>. These NRPS, PKS hybrids can be further distinguished. One main group are the lipopeptides, linear or cyclic peptides with a fatty acid chain that is often attached to the C- or N-terminus of the peptide. Prominent examples for this group are the myxochromides or the cystomanamides<sup>[31,45]</sup>

Another group of NRPS-PKS hybrid molecules contain a larger polyketide part within the scaffold as seen in DKxanthenes or epothilones. The feature of these molecules are that the biosynthetic assembly line contains large polyketide modules enclosing NRPS modules and vice versa.<sup>[12]</sup> The ketosynthases of such hybrid PKS Clusters can therefore recognize the PCP bound peptide or amino acid and catalyze the C-C bond formation<sup>[46]</sup>. In the other direction from a PKS to a NRPS module the C-domains can catalyze condensations between the ACP bound acyl chain and the subsequent amino acid being bound to the NRPS module<sup>[47]</sup>.

A special case of NRPS-PKS hybrid molecules are NRPs containing  $\beta$  and  $\gamma$  amino acids like myxoprincomide or tubulysine. A characteristic feature of these hybrid synthetases are single PKS modules within the NRPs assembly line.<sup>[48,49]</sup>

Terpene NRP hybrid molecules do also exist but are a lot more uncommon than NRPs-PKS hybrid molecule. An example for such a hybrid is the cytotoxic leupyrrine from *Sorangium cellulosum* which is actually a NRPS-PKS-terpene hybrid molecule. During biosynthesis the pyrrol moiety derived from proline is prenylated with geranylphosphate<sup>[50]</sup>

## 1.5 Other peptide type natural products

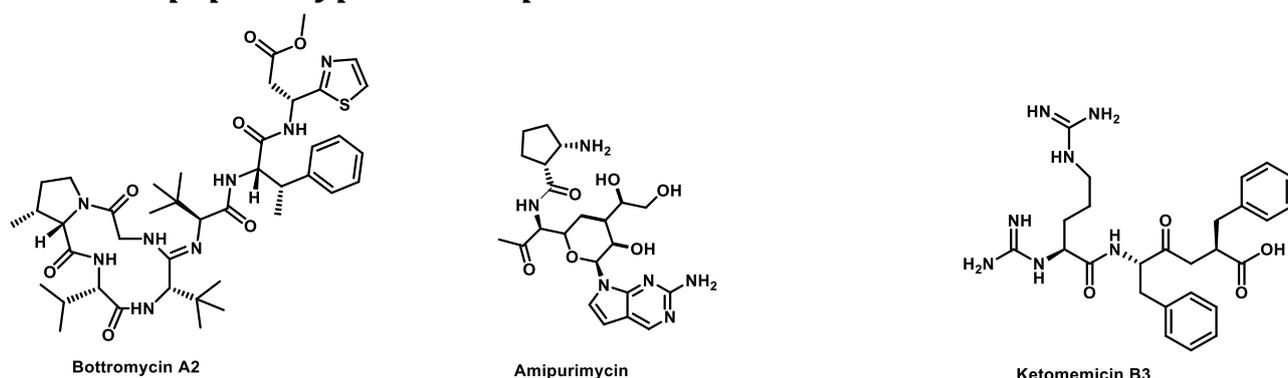


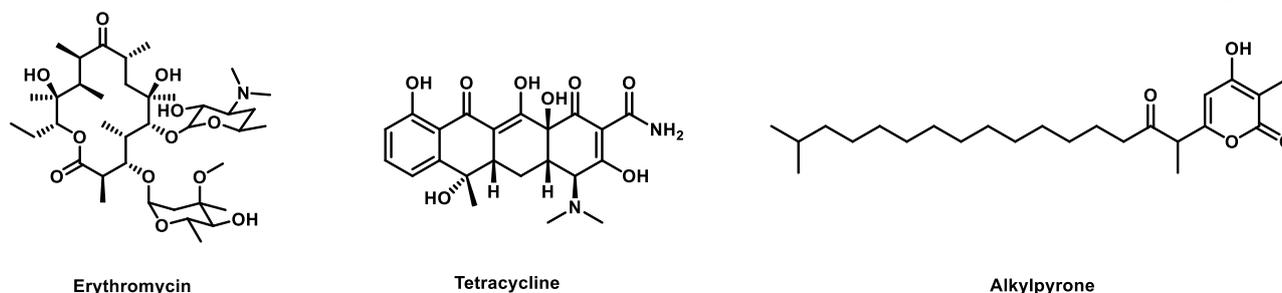
Figure 3: Examples of peptide type natural products: Bottromycin A2 a ribosomal synthesized and post-translational modified peptide (Ripp), amipurimycin a nucleotide antibiotic where a GRASP ligase conducts peptide bond formation and ketomemycin B3 a pseudopeptide where two amino acids (phenylalanine) are fused by an acetate linker.

Ribosomal peptides represent the biggest group in the natural product peptide class beside NRPs. Ribosomal synthesized and post-translational modified peptides (RiPPs) are a large growing number of natural products

that can be found in all domains of life. As their name suggests they are peptides synthesized by ribosomes and modified after their translation that are smaller than 10 kDa and bear a wide range of biological functions like antibacterial effects, cytotoxicity or as biosurfactant. During their biosynthesis the RiPPs are translated from the respective RNA sequence like a proteinogenic peptide into a precursor peptide which consists out of a N-terminal leader peptide and a unmodified core peptide. The leader peptide promotes recognition by enzymes performing post translational modifications like methylation, epimerization or oxidation of the amino acids inside the peptide. The leader peptide is proteolytically cleaved off. In some cases there is a C-terminal follower peptide that is crucial for cyclizations reactions. The follower peptide is as well cleaved off proteolytically like the leader peptide as well.<sup>[51]</sup> Rippps can be easily mistaken with NRPs since their peptide chain structure often shows high grades of similarities as in the case of bottromycin (figure 3)<sup>[52]</sup>. Formation of peptide bond or peptide like bonds is not limited to NRPS or ribosomes. ATP Grasp ligase enzymes can also form peptide bonds like for example in bacilysin biosynthesis<sup>[53]</sup> or for example in the attachment of aminocyclopentyl carboxylic acid (ACPC) to the nucleotide scaffold in the amipurimycin (figure 3) biosynthesis<sup>[54]</sup>. In case of cyclic dipeptides also a ribosomal independent pathway involving amino acyl t-RNA is known to conduct the biosynthesis<sup>[55]</sup>. An exceptional way of linking two amino acids can be found in biosynthesis of the pseudo peptide ketomemycin (figure 3): Here two phenylalanine units were linked via acetate by a claison type condensation reaction forming an exceptional carbonyl-methylene moiety<sup>[56]</sup>.

## 1.6 Polyketides

Polyketides are a large group of natural products being found in plants, fungi and bacteria. The term originates from the word ketide that stands for a methylene group attached to carbonyl group. Polyketides can be linear, cyclic or polycyclic. Examples for different kind of bacteria derived polyketides are shown in figure 4.



Erythromycin

Tetracycline

Alkylpyrone

Figure 4: Examples for polyketides: The antibiotic erythromycin a representative of a type I polyketide, the antibiotic tetracycline a type II polyketide and alkylpyrone an example for type III polyketides.

Biosynthesis of polyketides takes place in a similar manner as fatty acid synthesis: In both cases large multi-enzyme complexes sharing same catalytic activities, conduct the biosynthesis<sup>[57]</sup> and have the same evolutionary origin<sup>[58]</sup>. Fatty acid synthase (FAS) and Polyketide synthase (PKS) use malonate as building block, but in PKS also variants like methylmalonate and ethylmalonate were common building blocks. The PKS as well as the FAS contain an acyl-carrier protein domain (ACP-domain), an acyl transferase domain (AT-domain), a ketoreductase domain (KR-Domain) a dehydrogenase domain (DH-domain), and an enolyreductase domain (ER-domain).

Initiation of the biosynthesis of Polyketides and fatty acids starts with the loading of the starter unit which is mostly acetyl-CoA to the ACP-domain. The starter unit is covalently bound via a phosphopantetheine linker to the ACP-domain (figure 5A).<sup>[59]</sup> In addition to acetyl-CoA various other starter units like isovaleryl-CoA, 2-methylbutyryl-CoA or isobutyryl-CoA can be found in fatty acids or polyketides. For chain elongation in

polyketides the ACP-bound starter unit is transferred to the KS-domain of the following module while a building block (usually malonyl-CoA) is loaded onto the ACP domain via the AT-domain. Outgoing from here, the building block reacts with the KS-domain transiently bound starter block in a claisen condensation to form a new C-C bond. During this reaction the free carboxylic acid from the malonyl-CoA building block is released as CO<sub>2</sub> (figure 5B). The formed ketone can be reduced depending on the domains contained in the module. A KR-domain will reduce the ketone to an alcohol that can be further reduced by a DH-domain to a double bond. The double bond can be further reductive hydrated into a saturated bond by the ER-domain. For elongation of the acyl/polyketide chain this process is repeated several times (figure 5C and D).<sup>[60]</sup>

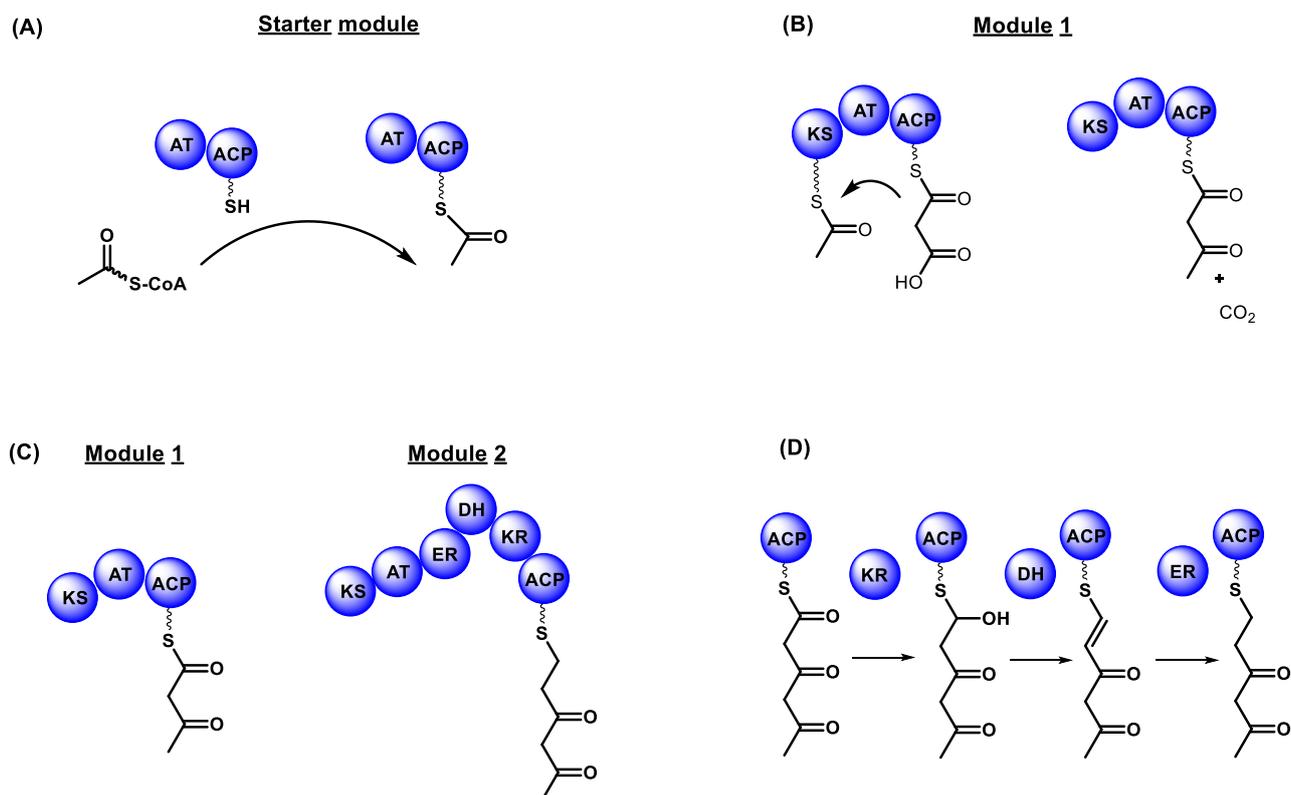


Figure 5: Description of polyketide biosynthetic logic: (A) Initiation of biosynthesis by transfer of the starter unit (here acetyl-CoA) to the acyl carrier protein catalyzed by the acyl transferase domain (AT-domain). (B) First elongation step in the first module: The starter unit is transiently transferred to the ketosynthase domain (KS-domain) of module 1. The KS catalyzes the claisen type condensation of the ACP bound malonate. During the reaction carbon dioxide is cleaved off from the malonate. (C) Elongation step catalyzed by module 2: Module 2 contains additional domains that chemically reduce the elongated polyketide chain as displayed in (D): The carbonyl group is reduced into an alcohol by the ketoreductase domain (KR-domain), the dehydratase domain (DH-domain) reduces the alcohol into a double bond that is further reduced by the enoyl-reductase domain (ER-domain)

Termination of biosynthesis is conducted by a thioesterase domain (TE-domain) by cleaving off the thioester bond and releasing either a linear acyl chain with a carboxylic acid (fatty acid) or a cyclic macrolide. In many biosynthetic pathways the TE-domain catalyzes also the cyclization between the carboxylic acid and a hydroxyl group on the acyl chain to give a macrolide polyketide.<sup>[61]</sup>

Polyketides can be divided into type I, II and III:

### **Type I polyketides**

Bacterial type I PKS are organized in modules with each module bearing the necessary catalytic domains (KS-domain, AT-domain, ACP-domain e.g.). After a chain elongation step the growing polyketide chain is passed to the next module until the chain elongation is terminated by a module containing a TE-domain. Instead of malonyl CoA type I PKS can use methyl-malonyl-CoA and ethyl-Malonyl CoA. In combination with limited/incomplete reduction steps a high structural variability in comparison to fatty acids is obtain. <sup>[62]</sup> In some biosynthetic pathways certain polyketide modules were found to act iteratively <sup>[63]</sup>. Resulting products are usually linear chains with a high degree of saturation or be cyclized via an ester bridge, so called lactones or macrolides.

### **Type II polyketides**

Products of the type II polyketide pathway are usually cyclic polyphenoles with strong antibacterial or cytotoxic activity. For long time type II PKs were expected only to occur in gram positive actinomyces till recent results also demonstrated the occurrence in gram negative myxobacteria<sup>[64]</sup>.

In contrast to the usual type I polyketide, type II PK synthases can be considered as minimal PKS which only consists out of two KS domains (KS $\alpha$  and KS $\beta$ ) and an ACP domain that work iteratively <sup>[65]</sup>. The variation or of known structures is achieved by tailoring reactions like oxygenation, methylation or glycosylation<sup>[66]</sup>. The length of the carbon chain during synthesis is determined by the KS $\beta$  unit<sup>[67]</sup>. Another unique feature of type II PK synthase units is that only malonyl coA is used as building block <sup>[66]</sup>.

### **Type III polyketides**

Until 1999 it was thought that the occurrence of type III polyketides is only limited to plants and fungi until their presence was also proven in bacteria<sup>[68]</sup>. Type III polyketide synthases form a homodimeric complex where each monomer catalyze the whole set of reactions from initiation over elongation and termination of the chain elongation. Usually three to five malonyl-CoA units were used for chain elongation to yield a product containing one or two highly unsaturated rings. In contrast to type II polyketides, type III polyketides can also utilize methylmalonyl- and ethylmalonyl-CoA as substrates as observed for example in germicidin biosynthesis<sup>[69]</sup>. In some type III PKs like the alkylpyrones e.g. long alkyl chains are the starter unit that are simply elongated by two malonyl- or methyl-malonyl-CoA units followed by cyclization<sup>[70]</sup>.

## ***1.7 Terpenes***

Terpenes are a large and diverse compound family. A unique feature of this natural product family is that they can be found in all domains of life. While most classes of natural products are limited to bacteria, fungi and plants, terpenes like cholesterol can be also found in mammals. Especially in mammals the group of sterols plays a significant role as hormones and is an important part of the cell wall.

Plants produce a high variety of terpenes beside phytosterols the plant analogue of animal cholesterol. Most famous examples for plant derived terpenes with pharmacological use are taxol and artemisin. Compared to plants or fungi bacteria do not appear to produce terpenes as often as other secondary metabolite classes like polyketides and non-ribosomal peptides. Nonetheless some bacterial terpenes are quite known like geosmin (figure 6) for example which is responsible for the earthy smell of the soil. [71]

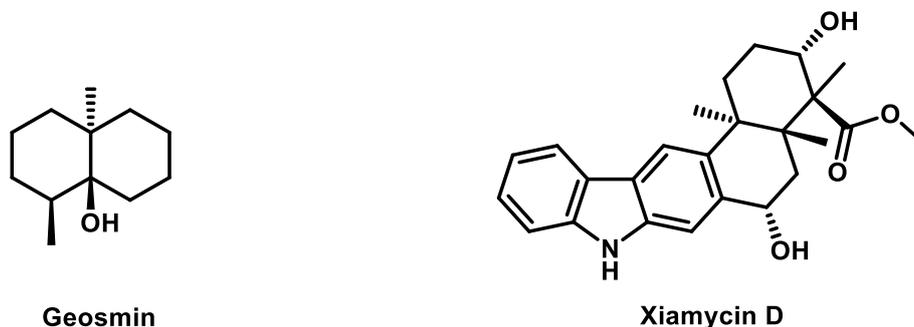


Figure 6: Examples of bacteria derived terpenes

Terpenes are derived from C<sub>5</sub> isoprene units which lead to scaffolds with a Carbon numbers being multiples of 5. Therefore terpenes are classified based on their carbon number. Terpenes are derived from dimethylallylphosphate (DMAPP) and isopentenyl diphosphate (IPP) which enzymatically isomerize into each other [72]. IPP itself is derived either by the mevalonate pathway (figure 7A) or the mevalonate independent pathway (figure 7B) with methyl-erythritol phosphate (MEP) as key intermediate. [73]

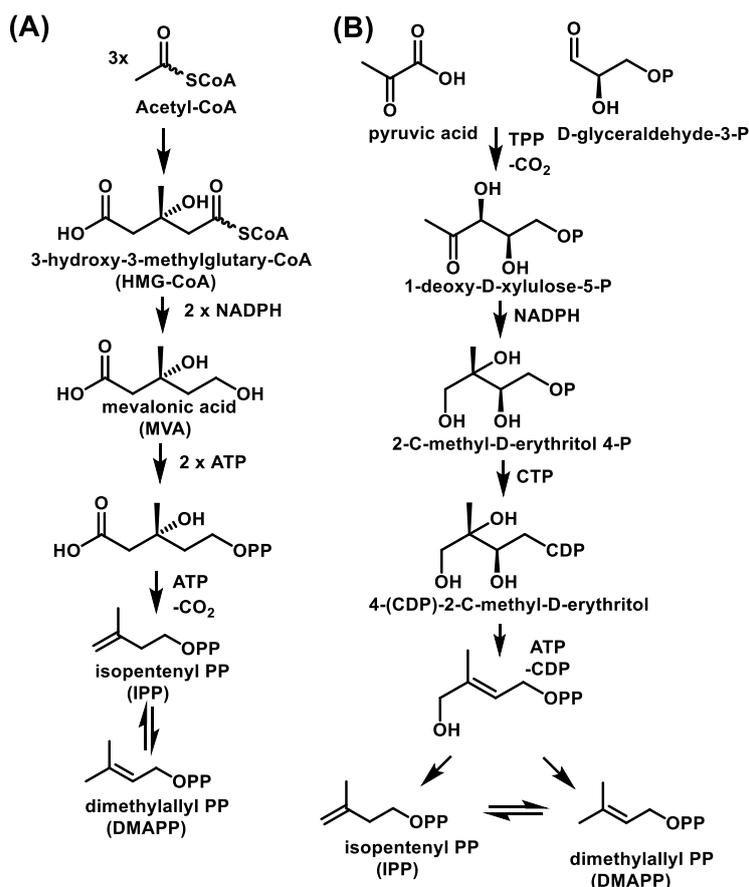


Figure 7: Biosynthesis pathways of Dimethylallyl-Pyrophosphate (DMAPP): (A) Mevalonic acid pathway; (B) Mevalonic acid independent pathway

In the mevalonate pathway mevalonic acid, the key intermediate and precursor of IPP is made out of three acetyl-CoA units by claison condensation. On the other side during the mevalonate independent pathway

MEP is the product of a thiaminpyrophosphate dependent reaction of pyruvate with Glyceraldehyde-3-Phosphate [73,74]. The formation of a monoterpene requires the combination of a DMAPP and an IPP unit. By losing the diphosphate DMAPP turns into a resonance stabilized allylic cation enabling a nucleophilic attack from the IPP resulting in the formation of geranyl-diphosphate (figure 8). Elongation of the monoterpene chain works in a similar manner by the cleavage of the diphosphate and the nucleophile attack of an IPP unit forming. [75]

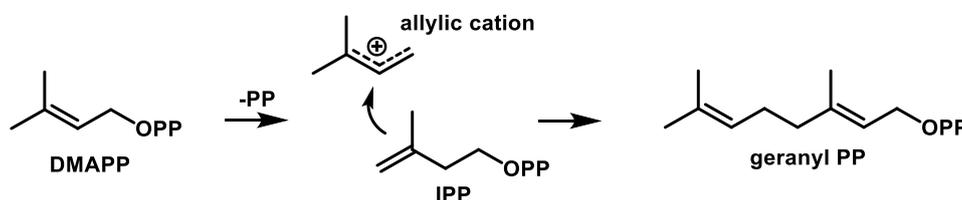


Figure 8: Formation of geranyl-pyrophosphate from dimethylallyl-Pyrophosphate (DMAPP) and Isopentenyl-pyrophosphate (IPP)

The extreme variations of different terpenes derived from the same scaffold are achieved by E-Z isomerizations, or from tailoring reactions involving reductions as well as hydroxylation reactions. The high variations of cyclized terpenes that came from the same isoprene chain, are the result of cationic intermediate states allowing different cyclization reactions to occur. [76] In addition terpene-amino acid hybrid structures were often observed like in terpene-indole alkaloids [77]. But also in bacteria these hybrids are not that uncommon like it is the case with the indole terpenes like xiamycin D (figure 6) isolated from streptomyces spp [78].

## 1.8 From a microorganism to the drug

The development of a new drug takes over 12 years and costs approximately 2.6 billion \$ and is a highly complex process consisting of several steps, in particular the hit identification, the hit to lead optimization, followed by clinical trials [79]. Academic institutes, which are often focused on basic research, or smaller companies usually do not possess the financial capability, lack human resources and knowhow to cover all steps of the process. Therefore, division of work is a common strategy in drug development. Academia's contribution in the process of drug development is primarily concentrated on the hit identification step. As described previously natural products are a great source for new pharmaceuticals. Before a natural product can be used for hit identification, it has to be identified and purified from a complex mixture of primary and secondary metabolites represented by the bacterial broth. As soon as the natural product is isolated, the structure needs to be thoroughly determined in order to allow mechanistic studies of the mode of action and also the development of synthetic routes for a better access to the compound. The following section focuses on the identification of novel natural products from bacteria, the structure elucidation and highlights difficulties and strategies to overcome them.

### 1.8.1 Accessing new natural products

The main difficulty of natural product isolation is to identify one molecule of interest among often several thousand other molecules and to separate it from this complex mixture. A variety of methods and techniques is available that assist in this process. Beside basic techniques like precipitation, crystallization, liquid/liquid

partitioning and column chromatography, high performance liquid chromatography (HPLC) is the most valuable technique for both identification and separation. HPLC was developed based on column chromatography. The use of closer packed columns and higher flow rates provided a higher reproducibility while UV/VIS or refractory detection enabled a real-time monitoring of the separation.<sup>[80]</sup>

The introduction of mass spectrometry or to be precise the coupling of mass spectrometers with HPLC machines outgoing from the late 70s had a massive impact on natural product research. The mass spectrometer serves as detector comparable to conventional UV/VIS detectors adding the dimension of the mass to the chromatogram. In addition, a mass spectrometer provides a higher sensitivity than UV detectors and detection is not limited to molecules with UV activity. But like with UV/VIS detectors, not every molecule can be detected by MS, since this technique is strongly dependent on the individual ionization behavior. Utilizing the additional mass dimension the isolation process is simplified since target compounds can be separated in a more focused manner.<sup>[81]</sup>

For the compound identification process, the coupling of HPLC with high resolution mass spectrometers (LC-*hrMS*) was a major progress. It allows better differentiation of molecules with similar masses, early dereplication and determination of the sum formula. Dereplication is of growing importance since the great number of discovered natural products carries the risk of rediscovery<sup>[82]</sup>. In order to avoid this, it is crucial to “quickly identify known chemotypes”<sup>[83]</sup> in the crude extract. The whole process is summarized in the term dereplication. In order to do proper dereplication, information about all described natural products or a subset of them linked to a certain species of interest are vital. Databases like the Dictionary of Natural Products (DNP)<sup>[84]</sup> or Scifinder<sup>[85]</sup>, which contain most of today's described natural products, can provide this information including the sum formula, structure and spectroscopic data. A classic dereplication strategy would make use of these databases after a natural product is isolated and spectroscopic data like NMR spectra or IR-spectra were acquired in order to unequivocally identify a known compound. Since isolation of natural products is a time-consuming task, it is advantageous to implement the dereplication step already at an earlier stage when working with the crude extract. To achieve this, analytical information of the molecules contained in the crude extract have to be obtained. One option here is the acquisition of NMR spectra of crude extracts in combination with principal component analysis (PCA), which is a commonly used method for compound identification<sup>[86]</sup>. However, due to the low sensitivity of NMR spectroscopy, this methodology is bound to fail with compounds in low abundance. In order to identify molecules with lower titers in the crude extract, mass spectrometry is the method of choice due to its higher sensitivity. With *hrMS* equipment, the sum formula can be determined. Comparison of the high-resolution mass or the sum formula of a compound to a database might be sufficient in some cases, but as a differentiation between constitutional isomers is not possible, more information about the compounds is required in most cases. These additional analytical information are the retention time and the fragmentation pattern. The strong dependency of the retention time in turn requires all entries of the database to be obtained with the same or a highly comparable analytical setup. More like the fingerprint of a molecule, the MS fragmentation pattern can also be used for dereplication. With modern computer algorithms for Clustering of MS/MS spectra, also derivatives of known compounds can be identified.<sup>[87][88]</sup>

To identify natural products with bioactivity in a crude extract, a fractionation is the method of choice. In classical fractionation approaches the crude extract is fractionated using chromatographic techniques like normal phase column chromatography, size exclusion chromatography or reverse phase flash chromatography. The resulting fractions are tested again and the active fractions are further fractionated until a pure compound can be correlated to the initial biological activity (figure 9A).<sup>[89]</sup>

With LC-*hr*MS analytical setups, the classical fractionation is further developed into the activity guided compound isolation. Here, only a small portion of a crude extract is fractionated using LC-*hr*MS. The collected fractions were conventionally tested for biological activity but due LC-*hr*MS, the masses of the natural products being present in the fractions are known (figure 9B). Using this approach the number of potentially biologically active compounds can be narrowed down to small number or even to a single molecule from a complex mixture. The potential targets can be isolated from the initial extract in a more focused manner or production can be optimized before purification. In addition with the provided retention time and *hr*mass data offer the possibility of early dereplication.

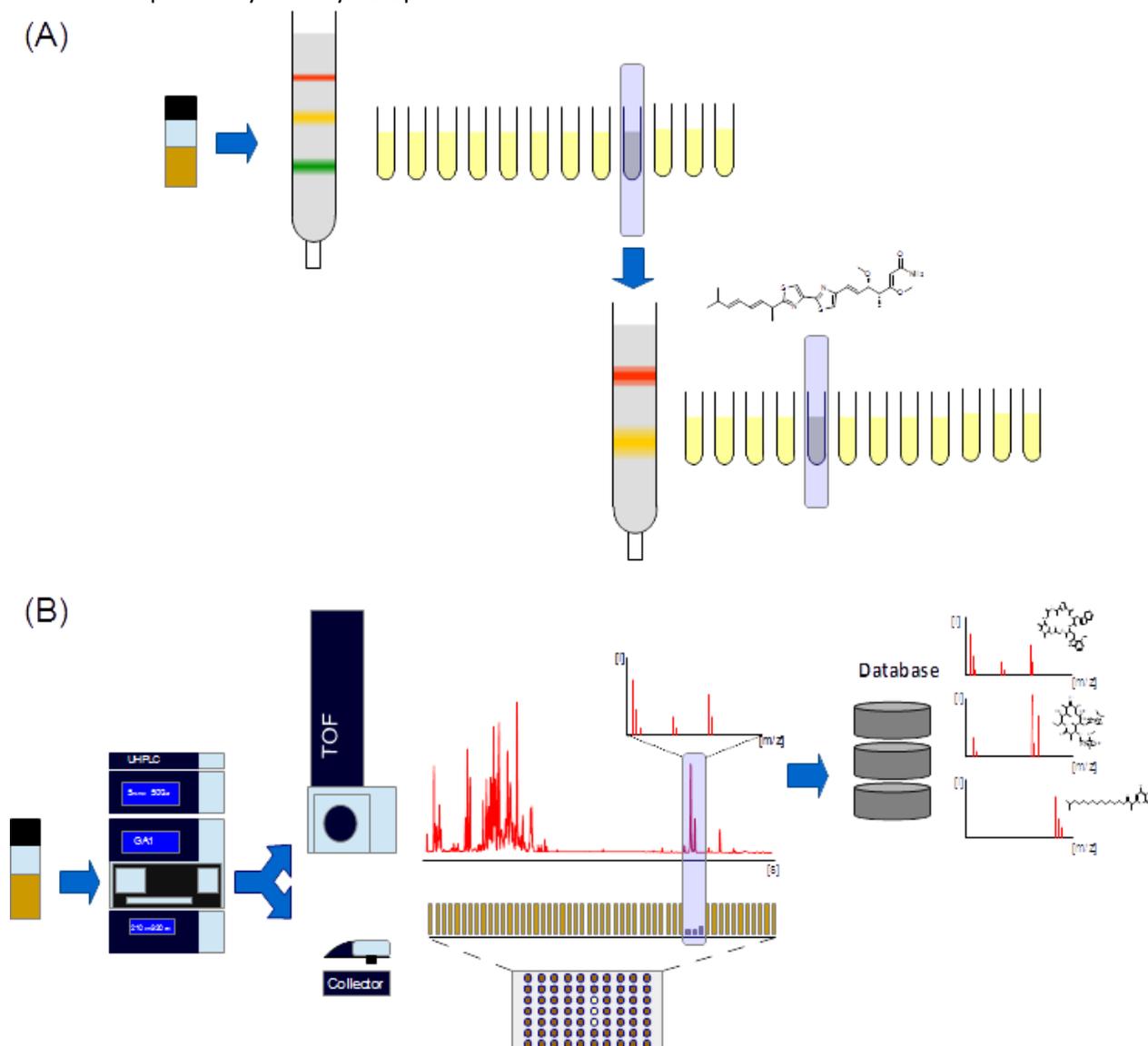


Figure 9: (A) Classical fractionation strategy for identification and isolation of bioactive compounds: A bacterial extract is fractionated by column chromatography. The collected fractions are tested for bioactivity. Active fractions are further fractionated until one fraction contains the active molecule. (B) Activity guided compound isolation: A crude extract is fractionated by HPLC. The outgoing flow is split. One portion of the flow is injected into a high resolution mass spectrometer while the other portion goes into a fraction collector. The collected fractions were tested for bioactivity. Since the fractions are collected to defined time intervals an alignment with the chromatogram is possible. Active fractions can be correlated to a retention time window therefore potential target masses can be identified. Using the high resolution mass and retention time dereplication using databases is possible to avoid re-discovery. Potential new target masses can be determined and in the following be isolated from the bacterial extract.

The classic fractionation and also activity guided compound isolation as natural product discovery strategy has several drawbacks. Natural products provide a wide array of bioactivities, but during the screening process, only a small range of potential biological activities can be tested. Natural products with other biological properties might simply be missed out. Low abundance of natural products can be a second drawback since production in bacteria often underlies a strong variance. Natural products being produced in high titers are usually discovered easily. Unfortunately, these "low hanging fruits" are often already described scaffolds<sup>[90]</sup> whereas low abundant novel metabolites might not be detected in this method. To overcome these drawbacks, new discovery approaches using modern analytic setups and cultivation techniques were developed. These new approaches can be applied in addition to classical approaches or as complete replacement.

The biodiversity approach is focused on the bacterial source. Nowadays, the majority of bacterial natural products are derived from actinobacteria. Expanding the screening focus to a more diverse range of bacterial species will increase the chance of new discoveries. This is shown by the increasing number of natural products isolated from cyanobacteria and myxobacteria in the last two decades.<sup>[13,91]</sup> Also the vast number of so far uncultured bacteria implies a great pool of novel promising scaffolds. Novel sampling strategies from specialized environments like insect colonies or sampling of human microbiome associated bacteria are promising approaches for natural product discovery.<sup>[92]</sup> Of course not every novel isolate is a good source. Some bacteria are simply less gifted producers of natural products, while other bacterial species have promising potential to produce many different natural products with respect to the genome size and natural habitat. For example analysis of a great number of genomes from anaerobic bacteria could show a correlation between biosynthetic potential and genome size. In addition, the biosynthetic potential strongly was reflected by the habitat. While soil associated bacteria had the greatest potential, isolates from extreme environment showed the lowest<sup>[93]</sup>

The metabolic approach focuses on the whole metabolome. The emphasis of this approach is more on the discovery of novel chemical entities. In the last couple of years, metabolic approaches could demonstrate their ability to identify low abundant metabolites. Based on LC-*hr*MS datasets of cultures and media blanks, principle component analysis (PCA) followed by selected MS<sup>2</sup> fragmentation and clustering algorithms were successfully applied to discover several new natural products.<sup>[94,95]</sup>

With reduced costs for genome sequencing, genomic approaches gain more and more importance. The genome mining approach is a more diverse approach and not exclusively focused on natural product discovery. The term can rather be applied to everything related to genome based analysis of an organism. In natural product research, genome mining can be divided into identification of biosynthetic genes and genome mining based compound discovery.

The identification of biosynthetic genes is quite effective since biosynthesis of natural products follows certain rules despite their high variety. Especially polyketides, NRPs and RiPPs are produced by multi-enzyme complexes with highly conserved domains<sup>[25,51,60]</sup>. This fact can be used to search bacterial genomes for biosynthetic core enzymes. In the past, cosmid libraries of genomes were screened by southern blot hybridization using homologues as probes. Since full genome sequences are now available at lower cost, genomes can be screened *in silico* using alignment algorithms to search for biosynthetic enzymes.<sup>[96]</sup> Considering the size of bacterial genomes, especially from myxobacteria, a manual search for single conserved domains by alignment turns into a tedious task. Using the fact that genes involved in the biosynthesis of natural products are usually clustered into biosynthetic gene clusters, a BGC can be recognized if conserved domains of biosyn-

thetic core enzymes are located in close proximity or even in the same open reading frame (ORF) <sup>[97]</sup>. Computational services like antiSMASH use this relation to screen a microbial genome for BGCs using known and conserved biosynthetic enzyme domains as reference. As output, the BGCs present in a genome are annotated including proposed functional features of the domains and a rough estimation of the structural output. <sup>[98]</sup> Having annotated clusters from a genome, these BGCs can be correlated to a natural product produced by the microorganism or vice versa. Experimental proof that a certain cluster is really responsible for the production of a natural product usually involves genetic manipulations like knockout, inactivation or heterologous expression of the BGC. With the increase in genomic information, many studies have demonstrated that the biosynthetic potential of a certain bacterial strain is much higher than the observed metabolic output. Reasons for this discrepancy can be diverse. The product metabolite of a BGC simply might evade detection by the used analytical setup or can be hidden among other metabolites. Another reason for a missing metabolite is that the BGC is so far downregulated that it is below the detection limit under the applied cultivation conditions. <sup>[88,95]</sup> In order to access these cryptic BGCs, diverse approaches exist, that can be categorized in “one strain many compounds” (OSMAC) approaches and genome mining based compound discovery approaches.

The OSMAC approach tries to alter the metabolic profile and the bioactivity profile of a natural product producer by using different cultivation conditions. The basic idea here is, that different cultivation conditions will affect the BGC expression. For example one BGC expression is downregulated in medium 1, while in medium 2 the same BGC expression is upregulated due to a yet unknown mechanism. Also co-cultivation or usages of chemical elicitors were shown to increase or induce production of certain natural products. Since OSMAC approaches are mostly undirected, they require often a high-throughput screening. <sup>[99]</sup>

Genome mining based compound discovery approaches try to overcome this discrepancy to identify and yield new natural products. Most of these approaches require a proper analytical setup (mostly LC-MS devices) since they are based on a sophisticated interplay between genomics and metabolomics with the aim to connect a BGC with one or several peaks in the chromatogram. The most widely used method is the inactivation of the BGC for example by disruption of a crucial domain. The corresponding metabolite(s) will disappear and can be identified by comparing the chromatograms from wild type and knock-out mutant. In case of low abundant metabolites or metabolites hidden among other metabolites in LC-MS chromatograms, statistical approaches like PCA are useful to identify the differences. <sup>[49]</sup>

For “silent” BGCs, insertion of a putative promoters like *tn5* or the inducible vanillate promoter are a way to circumvent restriction of regulatory mechanisms that prevent the BGC from being expressed. If successful, new metabolites can be observed that have not been detected in the WT strain. Recent studies demonstrated that this approach successfully led to the discovery of new natural products <sup>[64]</sup>.

Activation by promoter insertion or domain disruption by integration of a foreign sequence requires manipulation of the genome. A well-established method is the targeted gene replacement by homologous recombination. Originally a mechanism to repair and exchange DNA, this mechanism is captured to introduce changes in the DNA of living cells. In order to integrate a DNA fragment into the genome in a targeted manner, a donor DNA sequence is needed usually provided in form of a plasmid. This donor DNA consists of the sequence fragment that should be integrated and two flanking homology regions that are homologous to the target in the host DNA. The integration sequence should have a selection marker to identify successful homologous recombinants since efficiency is usually low and requires a double strand break to happen before in the target region. <sup>[100]</sup>

Genome mining compound discovery is based on the knowledge about the biosynthesis of natural products, which was collected over the last 50 years. Prediction of biosynthetic pathways is only reliable if the function

of the involved proteins has been described. For newly discovered compounds with structural features not covered in previous research, it is crucial to elucidate their biosynthetic pathways in order to improve the prediction and identification of unknown BGCs. In addition, in depth knowledge about the biosynthesis enables synthetic biology approaches to generate new derivatives of a compound with improved pharmacological properties.

## 1.8.2 Structure elucidation of natural products

Once a natural product is identified and isolated, the structure determination is a crucial part for the characterization of a new molecule. Spectroscopic techniques like IR- or UV-VIS spectroscopy, which are commonly used in organic chemistry, will provide information about certain structural features. However, for the *de novo* structure analysis of a new natural product, these pieces of information are often insufficient. High-resolution mass spectrometry can provide the sum formula and by using MS<sup>2</sup> fragmentation, a partial structure elucidation is possible. Nevertheless, elucidation of whole structures of novel natural products can rarely be achieved only by mass spectrometry since most often, unknown fragments are created that barely are connectable.<sup>[101]</sup>

Nowadays, two techniques are commonly used for the *de novo* elucidation of complex natural products. One technique is X-ray diffraction spectroscopy. Based on the diffraction pattern of a crystal, an electron density map can be created giving the spatial structure of a molecule. A requirement for this method is the possibility to grow large enough crystals of a molecule, which is in most cases a major hindrance.

In contrast, nuclear magnet resonance (NMR) can be used with fewer restrictions. Only the low sensitivity requires decent amounts of the sample (which is also the case for X-ray crystallography) and acquired datasets need extensive analysis for *de novo* structure elucidation. Due to simple sample preparation and the universal applicability, NMR is the most powerful technique for structure elucidation in natural product research.<sup>[102]</sup> Most structures of natural products in the last decades were determined by NMR underlining the effectivity of this technique. In addition, the use of NMR spectroscopy is not limited to structure elucidation. In natural product research, the technique is also used to assist in the elucidation of biosynthetic pathways (e.g. for investigating incorporation profiles of isotope labelled precursors) and in the analysis of target binding to determine the mode of action of a natural product (e.g. STD-NMR).

### 1.8.2.1 Fundamentals of NMR spectroscopy

Like every kind of spectroscopy, NMR spectroscopy is based on the interaction between electromagnetic waves and matter. The focus of this spectroscopic method lies on the nuclei of atoms, as indicated by the name, and basically translates the magnetic properties of certain nuclei into chemical information. To understand how this chemical information are yielded, additional explanations about the basic principles and measurement methods are necessary:

Every nuclei has a positive charge due to the presence of protons and neutrons. In addition, some nuclei have an intrinsic angular momentum or spin  $P$ . A spinning positive charge is creating a magnetic moment  $\mu$ , which is proportional to the spin  $P$ .

$$\mu = \gamma * P \quad (1)$$

The gyromagnetic ratio  $\gamma$  is a proportional factor in equation 1. It is individual for every nucleus and determines the sensitivity of a certain isotope in NMR. Also the spin  $P$  is individual for every nucleus and depends on its spin quantum number  $I$ , which determines if a certain nucleus is NMR active. Nuclei with an odd mass and therefore an odd number of either neutrons or protons, like  $^1\text{H}$ ,  $^{13}\text{C}$  or  $^{15}\text{N}$ , have spin quantum numbers of  $I = \frac{1}{2}$ . Nuclei with an even mass composed of odd numbers of protons and neutrons, like  $^2\text{H}$  and  $^{14}\text{N}$ , have a spin quantum number of  $I = 1$ . Nuclei with equal numbers of protons and neutrons like  $^{12}\text{C}$  or  $^{16}\text{O}$  have spin quantum number of  $I = 0$  and therefore zero spin and no magnetic moment  $\mu$  (NMR inactive).

In a magnetic field  $B_0$ , a nucleus with a magnetic moment  $\mu$  is oriented. Depending on its spin quantum number  $I$ , it can take  $(2I+1)$  different orientations. Nuclei with a spin quantum number of  $\frac{1}{2}$  can consequently take 2 different orientations. The different orientations are also termed Zeeman-levels. The energy of the Zeeman-levels is dependent on the gyromagnetic constant  $\gamma$ , the magnetic spin quantum number  $m$  (which is equal to the spin quantum number  $I$ ), and the magnetic field  $B_0$  as shown in formula 2.

$$E = -m\gamma \hbar B_0 \quad (2)$$

As mentioned previously, all spectroscopic methods are based on the interaction of matter with electromagnetic waves. This interaction can only take place if the resonance condition is fulfilled, which means for NMR spectroscopy that the energy of the electromagnetic wave must be equal to the energy difference between two Zeeman-levels. The energy of electromagnetic waves depends on their frequency  $\nu$ , while the energy difference between two Zeeman levels is defined by the gyromagnetic constant  $\gamma$  and the magnetic field  $B_0$ . Therefore, for nuclei with magnetic spin quantum numbers of  $\frac{1}{2}$ , the resonance condition is expressed in formula 3.

$$\nu = \left| \frac{\gamma}{2\pi} \right| * B_0 \quad (3)$$

In order to acquire NMR spectra, several measurement methods are possible. In the beginning of NMR spectroscopy, the field sweep method and the frequency sweep methods were used. Both methods can be categorized as continuous wave (CW) methods where the amount of energy which a sample absorbs are measured continuously, while either the magnetic field ( $B_0$ ) or the excitation frequency is altered in order to fulfill the resonance conditions for each nucleus in the sample. These methods are time consuming since only one nucleus can be measured at a time.

The development of pulsed or FT-NMR (Fourier-Transform) mostly replaced the CW methods in the 60s. In this method, all nuclei of one kind (e.g.  $^1\text{H}$ ,  $^{13}\text{C}$  or  $^{15}\text{N}$ ) are excited at once by one unselective radio-frequency pulse  $B_1$  and their relaxation is detected as a function of time. To understand the principle of this measurement method, introduction of the magnetization vector model is necessary.

A nucleus with a magnetic moment  $\mu$  and a magnetic spin quantum numbers of  $\frac{1}{2}$  that is oriented in a magnetic field  $B_0$  can also be considered as a vector in a coordinate system with the magnetic field  $B_0$  as z-axis. The vector is rotating around the z-axis with the Larmor frequency, which is equal to its resonance condition. The vector shows either in the same direction as the magnetic field  $B_0$  or in the opposite direction; the two directions represent both Zeeman-levels.

A macroscopic sample inside a magnetic field contains a great number vectors rotating around the z-axis and being aligned with or against the magnetic field  $B_0$ . They average into the nuclei magnetization vector  $M_0$  (figure 10).

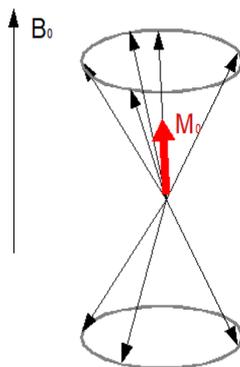


Figure 10: Precession cone showing the alignment of nuclei dipoles inside the magnetic field  $B_0$ . Since the number of vectors aligned with the magnetic field is higher than vectors aligned against the magnetic field, the magnetization  $M_0$  is parallel to the magnetic field  $B_0$ .

As mentioned previously, in the FT-NMR method, all nuclei of one kind (e.g.  $^1\text{H}$ ,  $^{13}\text{C}$  or  $^{15}\text{N}$ ) are excited at once by one unselective radio-frequency pulse  $B_1$  (termed “hard pulse”). This hard pulse  $B_1$  is focusing all vectors into the nuclei magnetization vector  $M_0$  and turns them into the direction of the x,y-plane. The turning angle  $\Theta$  is dependent on the pulse length  $\tau$  and the intensity of the pulse  $B_1$ .

$$\Theta = \gamma B_1 \tau_p \quad (4)$$

The rotation of the vectors induces a current in the detector, which consists of the receiver coils placed in x or y direction (depending whether receiver and excitation coil are the same). The signal detected by the receiver coil is dependent of the turning angle of the vectors. After the nuclei magnetization vector  $M_0$  is turned into the x,y plane by the pulse  $B_1$ , it slowly starts turning back into the z direction (‘relaxation’). On its way back, the nuclei magnetization vector splits into its single vectors. Relocation into z-direction reduces the signal intensity detected in the receiver coils until no signal is detected anymore, which resulted in the name free induction decay (FID) for the detected signal. The FID consists of overlapping sinus functions, each from a different kind of nucleus. A Fourier Transformation, which is a mathematical function that transforms a function of time into its constituent frequencies, will produce the resulting NMR spectrum (figure 11).

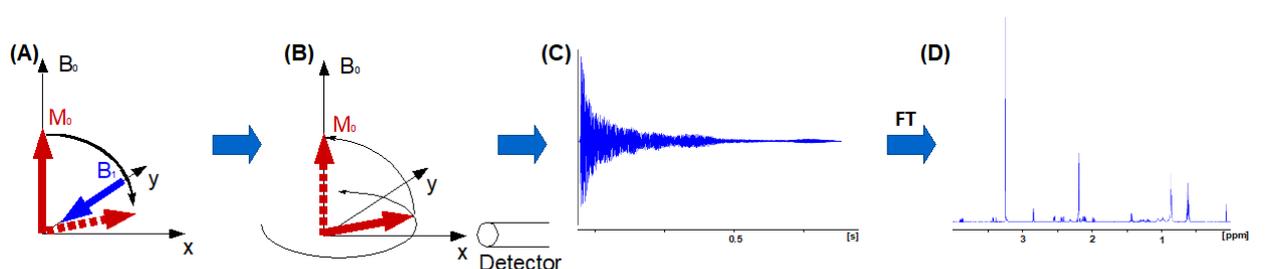


Figure 11: Principle of pulse spectrometry: (A) A pulse  $B_1$  is turning the magnetization vector  $M_0$  into the x,y-plane. (B) After the pulse, the magnetization vector  $M_0$  is precessing and moving back into its initial position. The rotation is inducing an electric current in the detector coil. The amplitude of the signal is decreasing over time due to the return of  $M_0$ . The detected signal over time is called free induced decay (FID) as shown in (C). Performing a Fourier Transformation over the time domain will give the NMR spectrum as displayed in (D).

As mentioned in the beginning of this chapter, NMR spectroscopy basically translates the magnetic properties of certain nuclei in a sample into chemical information. There are basically three pieces of information that can be obtained from the signals in a 1D NMR experiment: The chemical shift value of the nuclei in a molecule, the splitting pattern of the peaks and the relative intensity of the signals.

The chemical shift value is the most important information. It allows precise conclusions concerning the neighborhood of the observed nucleus and its association with functional groups. The chemical shift value is directly related to the resonance condition, which is only defined by the gyromagnetic constant  $\gamma$  and the strength of the magnetic field  $B_0$ . According to this consideration, the resonance condition and therefore the chemical shifts of nuclei of the same kind should be equal. However, in reality, all electrons near a specific nucleus in a molecule reduce the effective magnetic field  $B_0$  that is exerting its influence on the nucleus. As a result, the resonance frequency of that nucleus changes. Therefore, the individual resonance frequency of chemically non-equivalent nuclei is slightly different and provides information about the direct vicinity of the nuclei.

An increase of the electron density in the vicinity of a nucleus results in higher shielding of the nucleus leading to a lower resonance frequency. Vice versa, electronegative substituents with a certain electron 'pulling power' like oxygen in the vicinity result in a deshielding of the nucleus leading to a higher resonance frequency. Since the absolute resonance frequency is an arbitrary value that highly depends on the magnetic field as well as the instrument and sample-specific environment (e.g. the probe, the tube and the temperature), the frequency is expressed as a difference value compared to a reference standard. Trimethylsilan (TMS) is used as a reference standard and has been assigned a chemical shift of zero. The silica in TMS has a lower electronegativity than carbon, therefore giving a stronger shielding (and lower resonance frequency) than most organic substituents. The scale is commonly expressed as parts per million (ppm), a value that is independent of the spectrometer frequency. The chemical shift  $\delta$  is defined as follows:

$$\delta [ppm] = \frac{\Delta \nu}{\nu_{reference}} \quad (5)$$

Chemical shift values can be estimated by using empirically determined increment rules<sup>[103]</sup> or NMR chemical shift prediction software (for example Chemdraw or ACD-Lab NMR Suite). Using large NMR databases in combination with machine learning algorithms, the prediction of chemical shifts is constantly improving.<sup>[104]</sup> In addition, modern computational approaches using quantum chemical calculations, like density field theory (DFT), can be employed to calculate NMR chemical shift values<sup>[105]</sup>.

Signals in NMR spectra often show a splitting pattern instead of being one distinct signal. This splitting results from interaction with neighboring nuclei, which is termed coupling. Coupling effects can be divided into direct and indirect/scalar coupling. While the direct coupling describes an interaction between two nuclei through space, the indirect/scalar coupling is a through-bond interaction between nuclei. The splitting pattern for example in a proton spectrum is mostly the result of germinal (two bonds apart,  $^2J$ ) and vicinal (three bonds apart,  $^3J$ ) proton couplings. Also coupling between for example a proton and a carbon or a proton and more distant protons take place. But these type of couplings are more important in two-dimensional NMR experiments that will be referred to later in this chapter.

The phenomenon of indirect/ scalar coupling can be explained using the vector model: The magnetic spin of a certain nucleus is influencing the magnetic spins of neighboring nuclei in a mutual relationship. Depending on the direction of the magnetization vector of a nucleus  $M_0$  pointing towards or away from the direction of the magnetic field, it increases or decreases the effective magnetic field acting on the neighboring nuclei. Since in a macroscopic sample both cases occur despite the small energy gap between both directions, the resonance frequency of the neighboring nuclei is increased as well as decreased by the same factor. Instead of one larger peak, two smaller peaks with half of the intensity are observed in the NMR spectrum (figure 12A). Depending on the number of magnetic equivalent coupling partners, the signal of a proton is split

for example into a doublet, triplet or quartet. For every magnetic non-equivalent coupling partner, a resonance signal is split again, for example into a triplet of a doublet (figure 12B). For structure elucidation, analysis of the proton splitting pattern allows a conclusion about the number of neighboring atoms.

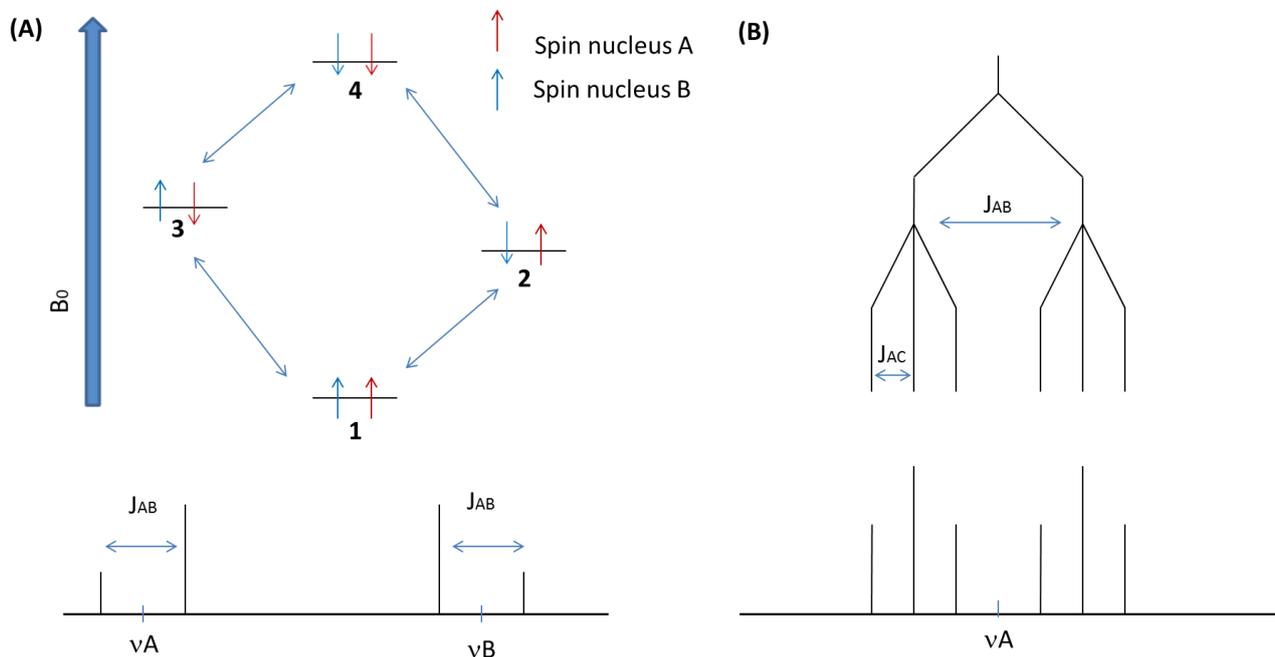


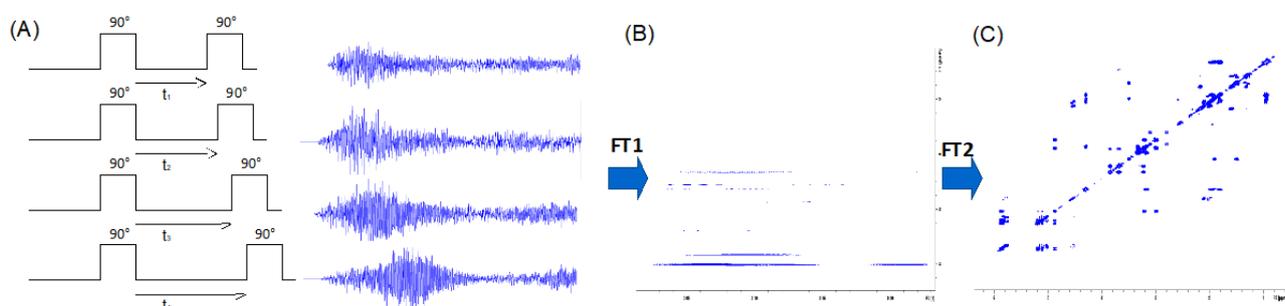
Figure 12: (A) Possible orientations in a two spin system in the magnetic field  $B_0$ . The four allowed transitions between the four states result in four signals for two nuclei in the NMR spectrum. The signals differ  $\frac{1}{2} J$  from the respective resonance frequency  $\nu$  of the nuclei. (B) Example of a splitting pattern from a system with three different magnetic non-equivalent coupling partners. Nucleus A couples with nucleus B leading to a doublet splitting. Coupling with two magnetic equivalent nuclei C results in a further splitting of the doublet into a doublet of a triplet.

From the distance between the peaks in a doublet, triplet or quartet, the coupling constant can be determined. Two coupling partners share one coupling constant, while more coupling partners will give more coupling constants that might differ from each other. The magnitude of a coupling constant depends on the distance between coupling partners, the angle and the respective electron density. The strong dependency of the coupling constants on the rotational angle was investigated by Martin Karplus and is summarized in the Karplus diagram that is a valuable tool to determine the conformation of a molecule. <sup>[106]</sup>

The intensities of NMR signals are proportional to the number of nuclei they derive from. Therefore, the integrated intensity of a signal allows to draw conclusions about the ratio of atoms that give rise to the signal. In organic chemistry, the two most commonly observed atoms are  $^1\text{H}$  and  $^{13}\text{C}$ . Even if the before mentioned statements are true for both nuclei, it has to be mentioned that the most widely used one-dimensional  $^{13}\text{C}$  experiment is a fully proton decoupled experiment. This NMR spectroscopy technique eliminates the splitting of the carbon signals due to coupling with neighboring protons and since the coupling with other  $^{13}\text{C}$  atoms is negligible due to their low natural abundance, all resulting signals appear as single peaks. This technique simplifies the spectrum but as a result, the area under the signals is not proportional anymore to the number of carbons giving rise to the signal. Nevertheless, the intensity of the signals in a  $^{13}\text{C}$  NMR spectrum still provides a certain degree of information as for example quaternary carbons routinely show a lower intensity than proton-attached carbons.

### 1.8.2.2 2D NMR experiments

Natural products are often highly complex molecules. Exact knowledge about neighboring atoms or atoms located in close proximity to each other is crucial for structure elucidation. This knowledge can be obtained from 2D NMR experiments. In addition, problems of overlapping signals, which complicate the structure elucidation, are often reduced or solved by the addition of a second dimension. Most 2D NMR methods are based on indirect or on direct couplings. Due to the coupling effects, a nucleus has not one but several slightly differing resonance frequencies depending on the number and nature of coupling partners. Therefore, the corresponding magnetization vectors have also different Larmor frequencies. The excitation pulse is focusing the vectors, but due to different rotation velocities, they split up. By application of a second pulse, the direction of rotation is inverted and the vectors refocus after a certain time depending on their Larmor frequency. This so called spin echo phenomenon finds its applications in 2D NMR experiments. By modulating the time between the pulses and/or the detection of the FID in small increments, a second dimension is added to the experiment. For each time increment, an FID is detected which is transferred into a 1D spectrum by Fourier Transformation. A second Fourier Transformation is applied over the time increment domain giving a 2D NMR spectrum (figure 13).<sup>[107]</sup>



*Figure 13(A) The pulse program of a 2D NMR experiment, here H,H correlation spectroscopy (COSY) is displayed. The mixing time between both 90° pulses is increasing by a fixed increment leading to different FIDs due to different spin echoes. (B) A first Fourier Transformation gives a 2D spectrum that is frequency resolved in F1 axis but time resolved in F2 axis. (C) The second Fourier Transformation provides a 2D spectrum frequency resolved in both axis.*

Most 2D NMR experiments can be divided into heteronuclear coupling experiments (HSQC and HMBC) and homonuclear coupling experiments (COSY, TOCSY and NOESY/ ROESY). The Heteronuclear Single Quantum Coherence (HSQC) indicates which carbon a proton is bound to, while Heteronuclear Multiple Bond Correlation (HMBC) indicates more distant carbons. The H,H Correlation SpectroscopY (COSY) shows coupling of neighboring (germinal and vicinal) protons (figure 14). In some cases (e.g. double bonds and aromatic systems), coupling to more distant can be observed with this experiment.

The Total Correlation SpectroscopY (TOCSY) experiment is often crucial in structure elucidation of peptides and oligosaccharides. The principle of this experiment resembles the H,H COSY experiment, but by application of a so called "spin-lock", all protons within a spin system are excited. In the 2D spectrum, all proton signals within one spin system show cross peaks (figure 14).<sup>[108]</sup>

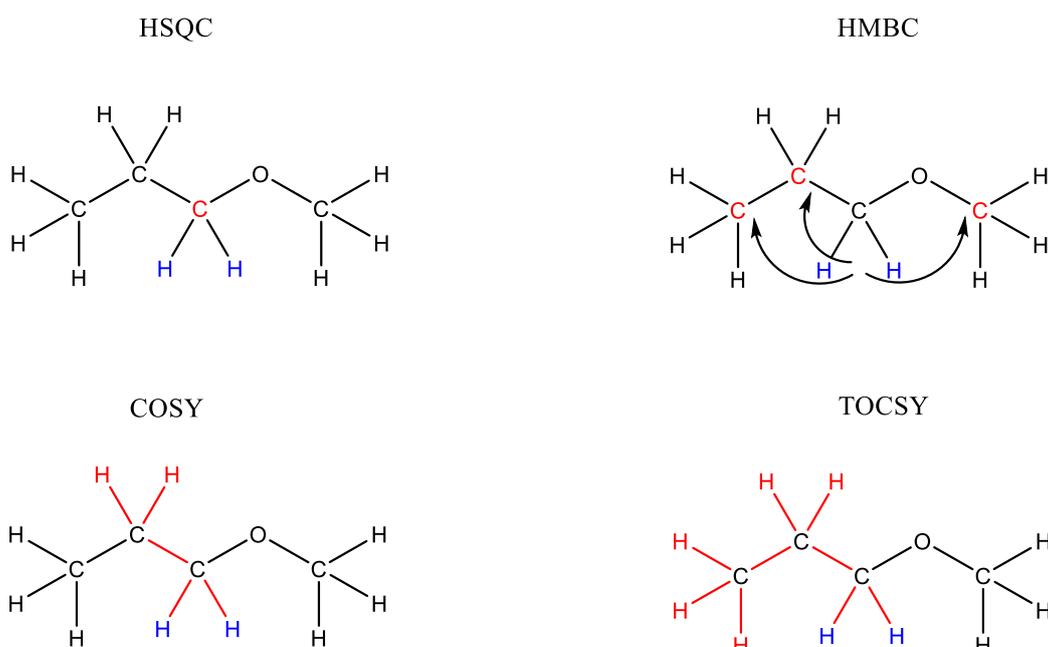


Figure 14: Types of scalar couplings indicated by standard 2D NMR experiments. Excited protons are displayed in blue and coupling atoms are displayed in red. Arrows indicate more distant coupling partners in the HMBC.

The Nuclear Overhauser Effect Spectroscopy (NOESY) is indicating direct spin couplings between protons. The Nuclear Overhauser effect (NOE) itself is based on the saturation of one nucleus, which is then transferring its magnetization through dipolar interactions to other nuclei through space. The intensity of dipolar interactions through space is inversely proportional to the radius (by the factor  $r^{-6}$ ) enabling distance measurements inside a molecule. In a similar manner, the Rotating frame Overhauser Effect Spectroscopy (ROESY) can be applied, even though this method uses a different pulse program. <sup>[109][103]</sup>

### 1.8.2.3 Methods for elucidation of the stereochemistry

Natural products are three dimensional objects. The bioactivity is often mediated by specific interaction with other three dimensional molecules (e.g. proteins). Therefore, determination of the natural product is a crucial step in the process of structure elucidation and also for future development of the natural product into a drug candidate. Stereochemical analysis distinguishes between absolute and relative configuration.

#### The relative configuration.

The relative configuration describes the spatial relations between atoms inside a molecule including double bond configuration without a reference point. While the orientation of double bonds is described as cis/ trans or E/Z, the relative stereodiscriptors syn/anti or threo/erythro are used to describe the relation of two neighboring stereocenters.

Standard NMR Experiments ( $^1\text{H}$ ,  $^{13}\text{C}$ , HSQC and HMBC) mostly provide information about the planar structure of a molecule and therefore, additional methods are necessary. Most NMR based methods used for the elucidation of the relative configuration include the analysis of coupling constants,  $^{13}\text{C}$ -shift analysis and NOE correlations.

The dependency of the magnitude of the coupling constants from the angle between two coupling nuclei, which was first described by Karplus, is widely employed. It can be used in the analysis of double bond configuration, but also neighboring stereocenters can be put into a relation using coupling constants. In cases of unspecific values or occurrence of rotamers (that lead to averaged values), a simple analysis of the vicinal coupling of protons is not sufficient. For acyclic molecules with defined substituents (methyl, oxygen, nitrogen or chlorine), the more detailed *J*-based analysis, also known as Murata's method, can be used. This method also considers  $^2J_{C,H}$  and  $^3J_{C,H}$ -coupling constants and NOE correlations to state the relative configuration<sup>[110]</sup>. The values of interproton spin-coupling constants ( $^3J_{H,H}$ ) can often be extracted from  $^1H$  NMR experiments. The required carbon-proton spin-coupling constants ( $^{2,3}J_{C,H}$ ) can be acquired using HETCOR, HETLOC or HSQMBC 2D NMR experiments<sup>[111]</sup>.

Due to different shielding effects, the relative configuration of stereocenters is influencing the chemical shift values. Using this relation, the relative configuration of stereocenters within specific structural motifs can be determined by comparing the experimental shift values to shift values of known structures. Especially for carbohydrate analysis, this approach is widely used to identify the sugar residue.<sup>[112]</sup> For various acyclic polyol compounds, Kishis universal NMR database method can be employed. Created from empirical analysis of carbon shift deviations in certain structural motifs, this database provides rules to determine the relative positions of neighboring stereocenters based on the variance from the average  $^{13}C$  shift values for the listed structural motifs.<sup>[113]</sup>

Intramolecular distances obtained from NOE measurements provide valuable information for the assignment of the relative configuration. For smaller rigid molecules, NOE-obtained distance restraints can be directly used to determine the relative configuration. For larger flexible molecules, the analysis requires more effort. Conformational variability needs to be considered in addition to the restraints constrains. A variety of computational methods and programs (e.g. fc-rDG/DDD calculations, ACD/Labs' Structure Elucidator expert system and SPARTAN software (wavefun Inc.)) were designed to fulfill this task.<sup>[114]</sup>

The described approaches can often not be applied to connect remote stereocenters. Here, the application of residual dipolar coupling (RDC) experiments can help to overcome this problem. Due to free movement of the molecules in isotropic media (nearly all organic solvents), other dipolar interactions than the scalar spin-spin coupling and NOEs are averaged to zero. In anisotropic media on the other hand, for example specific polymers, a partial alignment of the molecules enables direct magnetic spin-spin interactions. These additional anisotropic interactions result in increased couplings compared to values in isotropic media. Since the magnitude of magnetic spin-spin interactions is dependent on the angle of the spin in the magnetic field and on the distance of the interaction partner by the factor of  $r^{-3}$ , valuable constrains for the stereochemical configuration analysis are provided<sup>[115]</sup>. Determination of the configuration of the complete molecule is usually done by the singular value decomposition (SVD) method. The SVD method is based on calculated RDCs from a modelled structure, which are then compared to the experimentally acquired ones<sup>[116]</sup>.

### **The absolute configuration.**

In contrast to the relative configuration, the absolute configuration describes the spatial orientation of atoms in a chiral molecule from a fixed reference point. Nomenclature rules like the universal Cahn-Ingold-Prelog rules or the Fisher nomenclature for amino acids and sugars are employed to specify the absolute configuration. The determination of the absolute configuration requires an additional chiral component being present in the experiment. Standard NMR experiments cannot fulfill this task. Therefore, a variety of spectroscopic

methods, often in combination with non-spectroscopic methods, are used to assign the absolute configuration of stereocenters in natural products. Commonly used methods are chemical derivatization in combination with NMR (Mosher method) or HPLC-MS (advanced Marfey's method), spectroscopic methods using polarized light, and bioinformatics approaches.

Most chemical methods are based on the chemical derivatization of the chiral natural product. Using both enantiomers of a chiral derivatization reagent, the chiral natural product can be transformed into two diastereomers. While enantiomers are mirror inverted molecules that do show identical physicochemical properties in an achiral environment, diastereomers can be separated from each other.

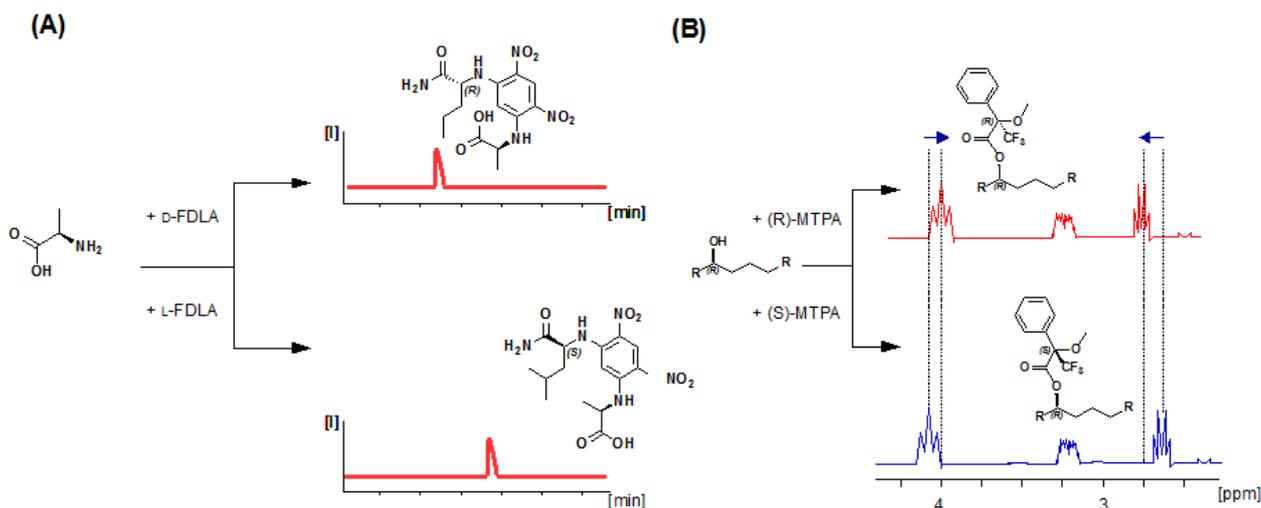


Figure 15: (A) Derivatization of an amino acid using *D*- and *L*-FDLA to create diastereomers. The diastereomers can be separated chromatographically. Using the respective standards, the absolute configuration of the amino acid can be determined by comparison of the retention time. (B) A chiral secondary alcohol is derivatized using (*R*) and (*S*)  $\alpha$ -methoxy- $\alpha$ -trifluoromethylphenylacetic acid MTPA to create two diastereomers. Protons in close proximity to the ester bond experience a shift in their resonance frequency due to different shielding effects from MTPA. By determining if the shift differences in between the diastereomers are positive or negative, an assignment of the absolute configuration of the alcohol can be made

For the assignment of the absolute configuration of a stereocenter containing a secondary alcohol or amine, derivatization into a Mosher ester using  $\alpha$ -methoxy- $\alpha$ -trifluoromethylphenylacetic acid (MTPA) is often the method of choice. The derivatization reagent contains an additional stereocenter. Depending on the enantiomer that is used for ester formation, different shielding effects apply to the chiral molecule. By comparing differences of the chemical shifts of the esters of both enantiomers, the absolute configuration of the respective stereocenter can be determined (figure 15B).<sup>[117]</sup> The Mosher ester relies on the formation of a stable product, which is not always achievable. For peptides, the advanced Marfey's method using 1-fluoro-2-4-dinitrophenyl-5-L-leucine amide (FDLA) as derivatization reagent is the method of choice to determine the stereochemistry of amino acid residues using HPLC-MS (figure 15A)<sup>[118]</sup>.

Polarized light itself is chiral and therefore interacts differently with two enantiomers of a molecule. This phenomenon, called optical rotatory dispersion or optical rotation, is a valuable tool to identify enantiomers. The dependency of the rotation angle from the wavelength is expressed in a circular dichroism (CD) spectrum. Usually a CD spectrum has negative and positive areas. The magnitude of the rotational angle in a CD spectrum of a molecule has the identical shape as the UV-VIS spectrum of the molecule. A rotational angle that is first negative and turns positive with increasing wavelength is called positive Cotton effect, while a first positive rotational angle is referred to as negative cotton effect. This effect is excessively used to investigate

the three dimensional structure of molecules. In molecules containing two chromophores connected by a stereocenter, the CD spectra usually show characteristic cotton effects. Taking advantage of this effect, the exciton chirality method is widely used to determine the configuration of stereocenters. This non-empirical method focuses on the charge transfer transitions between two spin systems (usually the chromophores). Depending on the spatial orientation, the rotational strength that describes the rotational angle at a given frequency varies strongly and results in positive or negative cotton effects.<sup>[119]</sup> Using modern computational methods, the rotational strength and therefore the CD spectrum of a molecule can be calculated using time dependent density field theory (TDDFT)<sup>[120]</sup>.

Recent developments in genome sequencing, the development of bioinformatic tools, and a deeper understanding of biosynthetic mechanisms have made it possible to predict the configuration of specific secondary alcohols in polyketides based on the DNA sequence of the ketoreductase domains.<sup>[121]</sup> In a comparable manner, the stereochemistry of amino acids in NRPSs can be predicted based on the presence of an epimerase domain the respective module in the BGC and amino acid specific C- and A-domains<sup>[122]</sup>.

## **1.9 Outline of this work**

The value of natural products for the development of new drugs and especially novel anti-infectives is undeniable. With the background of the increasing drug resistance, this value comes hand in hand with a constant need for new natural products. Therefore, the main focus of this work was the identification and characterization of new natural products from myxobacteria in combination with the elucidation of the biosynthetic pathways responsible for their production. The rationale of using myxobacteria as source for novel natural products is that they are relatively underexplored compared to actinobacteria. In addition relatively large genomes with a great number of biosynthetic gene cluster (BGC) clearly indicate a great potential. And the rising number of described natural products discovered from myxobacteria clearly proved this type of bacteria to be a promising source.

In the first step, two discovery strategies were chosen to fulfill this task. The first strategy was a classical cultivation of a newly isolated strain followed by an activity-guided compound isolation. This approach yielded the cytotoxic myxoquaterin compound family. The second strategy involved a genome mining approach with the goal to unveil the hidden metabolome of a second strain with a poor production profile under laboratory conditions. This approach yielded the myxoglucamides. In addition, a third compound class, the myxopentacines, was discovered by genetic manipulation of the strain.

In the second part of this study, all isolated metabolites were structurally characterized by multidimensional NMR experiments and *hr*MS measurements. For the determination of the relative and absolute configuration, various NMR experiments as well as chemical degradation and derivatization were used.

In part three, the biosynthesis of the discovered natural products was examined by *in silico* analysis in combination with feeding experiments.

For the myxoglucamides, the biosynthetic gene cluster (BGC) was already known due to the preceding genetic manipulation. The BGC of the myxoquaterines was identified based on the structural similarity to the DKxanthen compound family. Also the myxopentacin BGC was identified based on structure similarities. Here, similar substructures were found in the natural products amipurimycin and ketomemycin. By searching for core biosynthetic enzymes from both biosynthetic pathways, the corresponding BGC was identified. A gene inactivation experiment confirmed the link between BGC and myxopentacin production.

The results are subdivided based on the compound classes. Chapter 2 covers the myxoquaterin compound class, while chapter 3 is about the myxoglucamides. Chapter 4 describes everything related to the myxopentacin family.

## 1.10 References

- [1] M. Pleszczyńska, M. K. Lemieszek, M. Siwulski, A. Wiater, W. Rzeski, J. Szczodrak, *World journal of microbiology & biotechnology* **2017**, *33*, 83.
- [2] B. B. Petrovska, *Pharmacognosy reviews* **2012**, *6*, 1.
- [3] B. Fruth, N. B. Ikombe, G. K. Matshimba, S. Metzger, D. M. Muganza, R. Mundry, A. Fowler, *American journal of primatology* **2014**, *76*, 146.
- [4] A. P. Klockgether-Radke, *Anesthesiol. Intensivmed. Notfallmed. Schmerzther.* **2002**, *37*, 244.
- [5] M. Delepine, *J. Chem. Educ.* **1951**, *28*, 454.
- [6] A. Fleming **1929**, 226.
- [7] R. J. Dubos, *The Journal of experimental medicine* **1939**, *70*, 11.
- [8] a) S. A. Waksman, H. B. Woodruff, *Exp. Biol. Med.* **1940**, *45*, 609; b) H. Eagle, G. E. Foley, *The American Journal of Medicine* **1956**, *21*, 739.
- [9] A. Schatz, E. Bugle, S. A. Waksman, *Experimental Biology and Medicine* **1944**, *55*, 66.
- [10] L. T. Tan, *Phytochemistry* **2007**, *68*, 954.
- [11] T. Velkov, K. D. Roberts, R. L. Nation, P. E. Thompson, J. Li, *Future microbiology* **2013**, *8*, 711.
- [12] K. J. Weissman, R. Müller, *Nat. Prod. Rep.* **2010**, *27*, 1276.
- [13] J. Bérdy, *The Journal of antibiotics* **2012**, *65*, 385.
- [14] R. Thaxter, *Botanical Gazette* **1892**, *17*, 389.
- [15] R. Garcia, K. Gerth, M. Stadler, I. J. Dogma Jr., R. Müller, *Mol. Phylogenet. Evol.* **2010**, *57*, 878.
- [16] M. Dworkin, D. Kaiser (Eds.) *Myxobacteria II*, ASM Press, Washington, D.C., **1993**.
- [17] M. Dworkin in *Prokaryotic Development* (Eds.: Y.V. Brun, L. Shimkets), ASM Press, Washington, DC, **2000**, pp. 221–242.
- [18] S. Sudo, M. Dworkin, *J. Bacteriol.* **1972**, *110*, 236.
- [19] J. W. Bennett, R. Bentley in *Advances in Applied Microbiology* (1989), pp. 1–28.
- [20] J. Y. Ortholand, A. Ganesan, *Curr. Opin. Chem. Biol.* **2004**, *8*, 271.
- [21] D. A. Dias, S. Urban, U. Roessner, *Metabolites* **2012**, *2*, 303.
- [22] G. de la Torre, Beatriz, F. Albericio, *Molecules* **2019**, *24*, 809.
- [23] R. Liu, X. Li, K. S. Lam, *Current opinion in chemical biology* **2017**, *38*, 117.
- [24] M. J. Stone, D. H. Williams, *Mol. Microbiol.* **1992**, *6*, 29.
- [25] S. C. Wenzel, R. Müller, *Nat. Prod. Rep.* **2007**, *24*, 1211.
- [26] S. C. Wenzel, R. Müller, *Curr. Opin. Chem. Biol.* **2005**, *9*, 447.
- [27] C. T. Walsh, R. V. O'Brien, C. Khosla, *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 7098.
- [28] T. Stachelhaus, A. Huser, M. A. Marahiel, *Chem. Biol.* **1996**, *3*, 913.
- [29] a) T. A. Keating, C. T. Walsh, *Curr. Opin. Chem. Biol.* **1999**, *3*, 598; b) K. M. Hoyer, C. Mahlert, M. A. Marahiel, *Chem. Biol.* **2007**, *14*, 13; c) R. M. Kohli, J. Takagi, C. T. Walsh, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 1247.
- [30] N. Gaitatzis, B. Kunze, R. Müller, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 11136.
- [31] S. C. Wenzel, B. Kunze, G. Höfle, B. Silakowski, M. Scharfe, H. Blöcker, R. Müller, *ChemBioChem* **2005**, *6*, 375.
- [32] M. Inukai, R. Enokita, A. Torikata, M. Nakahara, S. Iwado, M. Arai, *J. Antibiot.* **1978**, *31*, 410.
- [33] L. Vollbrecht, H. Steinmetz, G. Höfle, *J. Antibiot.* **2002**, *55*, 715.
- [34] J. Porath, *Nature* **1953**, *172*, 871.
- [35] S. A. Sieber, M. A. Marahiel, *Chem Rev* **2005**, *105*, 715.
- [36] K. Hoffmann, E. Schneider-Scherzer, H. Kleinkauf, R. Zocher, *J. Biol. Chem.* **1994**, *269*, 12710.
- [37] T. Velkov, J. Horne, M. J. Scanlon, B. Capuano, E. Yuriev, A. Lawen, *Chemistry & biology* **2011**, *18*, 464.

- [38] S. Weinig, H.-J. Hecht, T. Mahmud, R. Müller, *Chem. Biol.* **2003**, *10*, 939.
- [39] C. Rausch, I. Hoof, T. Weber, W. Wohlleben, D. H. Huson, *BMC Evol. Biol.* **2007**, *7*, 78.
- [40] R. S. Roy, A. M. Gehring, J. C. Milne, P. J. Belshaw, C. T. Walsh, *Nat. Prod. Rep.* **1999**, *16*, 249.
- [41] a) D. Pogorevc, Y. Tang, M. Hoffmann, G. Zipf, H. S. Bernauer, A. Popoff, H. Steinmetz, S. C. Wenzel, *ACS Synth. Biol.* **2019**, *8*, 1121; b) C. Burgard, N. Zaburanyi, S. Nadmid, J. Maier, H. Jenke-Kodama, E. Luxemburger, H. S. Bernauer, S. C. Wenzel, *ACS Chem. Biol.* **2017**, *12*, 779.
- [42] A. König, T. Schwecke, I. Molnar, G. A. Böhm, P. A. Lowden, J. Staunton, P. F. Leadlay, *Eur. J. Biochem.* **1997**, *247*, 526.
- [43] J. B. Patteson, Z. D. Dunn, B. Li, *Angew. Chem. Int. Ed.* **2018**.
- [44] S. C. Wenzel, R. Müller in *Comprehensive Natural Products Chemistry II, Vol 2: Structural Diversity II - Secondary Metabolite Sources, Evolution and Selected Molecular Structures* (Ed.: B.S. Moore), Elsevier, Oxford, **2010**, pp. 189–222.
- [45] L. Etzbach, A. Plaza, R. Garcia, S. Baumann, R. Müller, *Org. Lett.* **2014**, *16*, 2414.
- [46] H. W. Chen, S. O'Connor, D. E. Cane, C. T. Walsh, *Chem. Biol.* **2001**, *8*, 899.
- [47] Gatto, G. J. Jr., S. M. McLoughlin, N. L. Kelleher, C. T. Walsh, *Biochemistry* **2005**, *44*, 5993.
- [48] H. Steinmetz, N. Glaser, E. Herdtweck, F. Sasse, H. Reichenbach, G. Höfle, *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 4888.
- [49] N. S. Cortina, D. Krug, A. Plaza, O. Revermann, R. Müller, *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 811.
- [50] H. B. Bode, S. C. Wenzel, H. Irschik, G. Höfle, R. Müller, *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 4163.
- [51] P. G. Arnison, M. J. Bibb, G. Bierbaum, A. A. Bowers, T. S. Bugni, G. Bulaj, J. A. Camarero, D. J. Campopiano, G. L. Challis, J. Clardy et al., *Nat. Prod. Rep.* **2013**, *30*, 108.
- [52] Y. Hou, M. D. Tianero, J. C. Kwan, T. P. Wyche, C. R. Michel, G. A. Ellis, E. Vazquez-Rivera, D. R. Braun, W. E. Rose, E. W. Schmidt et al., *Org. Lett.* **2012**, *14*, 5050.
- [53] G. Özcengiz, İ. Ögülür, *New Biotechnology* **2015**, *32*, 612.
- [54] A. J. Romo, T. Shiraishi, H. Ikeuchi, G.-M. Lin, Y. Geng, Y.-H. Lee, P. H. Liem, T. Ma, Y. Ogasawara, K. Shin-Ya et al., *J. Am. Chem. Soc.* **2019**.
- [55] A. K. Mishra, J. Choi, S.-J. Choi, K.-H. Baek, *Molecules (Basel, Switzerland)* **2017**, *22*.
- [56] J. Kawata, T. Naoe, Y. Ogasawara, T. Dairi, *Angew. Chem. Int. Ed.* **2017**.
- [57] S. Smith, S. C. Tsai, *Nat. Prod. Rep.* **2007**, *24*, 1041.
- [58] H. Jenke-Kodama, A. Sandmann, R. Müller, E. Dittmann, *Mol. Biol. Evol.* **2005**, *22*, 2027.
- [59] P. W. Majerus, A. W. Alberts, P. R. Vagelos, *PNAS* **1964**, *51*, 1231.
- [60] C. Hertweck, *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 4688.
- [61] R. S. Gokhale, D. Hunziker, D. E. Cane, C. Khosla, *Chem. Biol.* **1999**, *6*, 117.
- [62] J. Staunton, K. J. Weissman, *Nat. Prod. Rep.* **2001**, *18*, 380.
- [63] P. Meiser, K. J. Weissman, H. B. Bode, D. Krug, J. S. Dickschat, A. Sandmann, R. Müller, *Chem. Biol.* **2008**, *15*, 771.
- [64] F. Panter, D. Krug, S. Baumann, R. Müller, *Chem. Sci.* **2018**, *9*, 4898.
- [65] B. Shen in *Topics in Current Chemistry, Vol. 209* (Eds.: F. J. Leeper, J. C. Vederas), Springer-Verlag Berlin Heidelberg; Springer e-books, Berlin, Heidelberg, **2000**, pp. 1–51.
- [66] C. Hertweck, A. Luzhetskyy, Y. Rebets, A. Bechthold, *Nat. Prod. Rep.* **2007**, *24*, 162.
- [67] R. McDaniel, S. Ebert-Khosla, D. A. Hopwood, C. Khosla, *Science* **1993**, *262*, 1546.
- [68] N. Funa, Y. Ohnishi, I. Fujii, M. Shibuya, Y. Ebizuka, S. Horinouchi, *Nature* **1999**, *400*, 897.
- [69] J. A. Chemler, T. J. Buchholz, T. W. Geders, D. L. Akey, C. M. Rath, G. E. Chlipala, J. L. Smith, D. H. Sherman, *J. Am. Chem. Soc.* **2012**, *134*, 7359.
- [70] a) C. Nakano, H. Ozawa, G. Akanuma, N. Funa, S. Horinouchi, *Journal of bacteriology* **2009**, *191*, 4916; b) J. J. Hug, F. Panter, D. Krug, R. Müller, *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 319.
- [71] N. N. Gerber, H. A. Lechevalier, *Appl. Microbiol.* **1965**, *13*, 935.
- [72] M. E. Diaz, J. G. Mayoral, H. Priestap, M. Nouzova, C. Rivera-Perez, F. G. Noriega, *Insect biochemistry and molecular biology* **2012**, *42*, 751.
- [73] W. Eisenreich, M. Schwarz, A. Cartayrade, D. Arigoni, M. H. Zenk, A. Bacher, *Chem. Biol.* **1998**, *5*, R221-R233.
- [74] T. Kuzuyama, H. Seto, *Proceedings of the Japan Academy. Series B, Physical and biological sciences* **2012**, *88*, 41.

- [75] a) A.A. Kandutsch, H. Paulus, E. Levin, K. BLOCH, *Journal of Biological Chemistry* **1964**, 239, 2507; b) K. Ogura, T. Koyama, S. Seto, *Biochemical and Biophysical Research Communications* **1969**, 35, 875.
- [76] D. V. Banthorpe, B. V. Charlwood, M. J. Francis, *Chem. Rev.* **1972**, 72, 115.
- [77] S. E. O'Connor, J. J. Maresh, *Nat. Prod. Rep.* **2006**, 23, 532.
- [78] Z. Xu, M. Baunach, L. Ding, C. Hertweck, *Angew. Chem. Int. Ed.* **2012**, 51, 10293.
- [79] R. C. Mohs, N. H. Greig, *Alzheimer's & dementia (New York, N. Y.)* **2017**, 3, 651.
- [80] O. Sticher, *Nat. Prod. Rep.* **2008**, 25, 517.
- [81] M. S. Lee, E. H. Kerns, *Mass Spectrom. Rev.* **1999**, 18, 187.
- [82] C. R. Pye, M. J. Bertin, R. S. Lokey, W. H. Gerwick, R. G. Linington, *Proc. Natl. Acad. Sci. U.S.A.* **2017**, 114, 5601.
- [83] J. A. Beutler, A. B. Alvarado, D. E. Schaufelberger, P. Andrews, T. G. McCloud, *J. Nat. Prod.* **1990**, 53, 867.
- [84] "Dictionary of Natural Products", can be found under <http://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml>.
- [85] S. W. Gabrielson, *J. Med. Libr. Assoc.* **2018**, 106.
- [86] S. L. Robinette, R. Brüsweiler, F. C. Schroeder, A. S. Edison, *Accounts of chemical research* **2012**, 45, 288.
- [87] A. T. Aron, E. Gentry, K. L. McPhail, L. F. Nothias, M. Nothias-Esposito, A. Bouslimani, D. Petras, J. M. Gauglitz, N. Sikora, F. Vargas et al., *Reproducible Molecular Networking Of Untargeted Mass Spectrometry Data Using GNPS*, **2019**.
- [88] D. Krug, R. Müller, *Nat. Prod. Rep.* **2014**, 31, 768.
- [89] D. D. Baker, M. Chu, U. Oza, V. Rajgarhia, *Nat. Prod. Rep.* **2007**, 24, 1225.
- [90] A. Flemming, *Nature reviews. Drug discovery* **2013**, 12, 826.
- [91] J. Herrmann, A. A. Fayad, R. Müller, *Nat. Prod. Rep.* **2017**, 34, 135.
- [92] a) C. Beemelmans, H. Guo, M. Rischer, M. Poulsen, *Beilstein journal of organic chemistry* **2016**, 12, 314; b) A. Zipperer, M. C. Konnerth, C. Laux, A. Berscheid, D. Janek, C. Weidenmaier, M. Burian, N. A. Schilling, C. Slavetinsky, M. Marschal et al., *Nature* **2016**, 535, 511.
- [93] A. C. Letzel, S. J. Pidot, C. Hertweck, *Nat. Prod. Rep.* **2013**, 30, 392.
- [94] F. Panter, D. Krug, R. Müller, *ACS Chem. Biol.* **2019**, 14, 88.
- [95] T. Hoffmann, D. Krug, S. Hüttel, R. Müller, *Anal. Chem.* **2014**, 86, 10780.
- [96] N. Ziemert, M. Alanjary, T. Weber, *Nat. Prod. Rep.* **2016**, 33, 988.
- [97] M. H. Medema, R. Kottmann, P. Yilmaz, M. Cummings, J. B. Biggins, K. Blin, I. de Bruijn, Y. H. Chooi, J. Claesen, R. C. Coates et al., *Nat. Chem. Biol.* **2015**, 11, 625.
- [98] K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema, T. Weber, *Nucleic Acids Res.* **2019**, W81-W87.
- [99] K. Ochi, T. Hosaka, *Appl. Microbiol. Biotechnol.* **2013**, 97, 87.
- [100] M. R. Capecchi, *Science* **1989**, 244, 1288.
- [101] a) T. Kind, O. Fiehn, *Bioanalytical reviews* **2010**, 2, 23; b) K. Levsen, H.-M. Schiebel, B. Behnke, R. Dötzer, W. Dreher, M. Elend, H. Thiele, *J. Chromatogr.* **2005**, 1067, 55.
- [102] J. Prichystal, K. A. Schug, K. Lemr, J. Novak, V. Havlicek, *Analytical chemistry* **2016**, 88, 10338.
- [103] T.D.W. Claridge, *High-resolution NMR techniques in organic chemistry*, Elsevier, Amsterdam, London, **2009**.
- [104] a) S. G. Spanton, D. Whittern, *Magnetic resonance in chemistry : MRC* **2009**, 47, 1055; b) E. Jonas, S. Kuhn, *Journal of cheminformatics* **2019**, 11, 50.
- [105] G. L. Stoychev, A. A. Auer, F. Neese, *Journal of chemical theory and computation* **2018**, 14, 4756.
- [106] M. Karplus, *Adv. Ceram. Mater.* **1963**, 85, 2870.
- [107] W. P. Aue, E. Bartholdi, R. R. Ernst.
- [108] L. Braunschweiler, R.R. Ernst, *Journal of Magnetic Resonance (1969)* **1983**, 53, 521.
- [109] A. W. Overhauser, *Phys. Rev.* **1953**, 92, 411.
- [110] N. Matsumori, D. Kaneno, M. Murata, H. Nakamura, K. Tachibana, *J. Org. Chem.* **1999**, 64, 866.
- [111] D. Menche, *Nat. Prod. Rep.* **2008**, 25, 905.
- [112] K. Bock, C. Pedersen in *Advances in Carbohydrate Chemistry and Biochemistry*, pp. 27–66.
- [113] Y. Kobayashi, J. Lee, K. Tezuka, Y. Kishi, *Organic letters* **1999**, 1, 2177.

- [114]a) M. Köck, J. Junker, *Journal of Molecular Modeling* **1997**, *3*, 403; b) Y. D. Smurnyy, M. E. Elyashberg, K. A. Blinov, B. A. Lefebvre, G. E. Martin, A. J. Williams, *TETRAHEDRON* **2005**, *61*, 9980.
- [115]A. Saupe, G. Englert, *Phys. Rev. Lett.* **1963**, *11*, 462.
- [116]Losonczy, J. A., et al.
- [117]T. R. Hoye, C. S. Jeffrey, F. Shao, *Nat. Protoc.* **2007**, *2*, 2451.
- [118]a) P. Marfey, *Carlsberg Res. Commun.* **1984**, *49*, 591; b) K.-i. Harada, K. Fujii, K. Hayashi, M. Suzuki, Y. Ikai, H. Oka, *Tetrahedron Letters* **1996**, *37*, 3001.
- [119]N. Harada, K. Nakanishi, *Acc. Chem. Res.* **1972**, *5*, 257.
- [120]A. E. Nugroho, H. Morita, *Journal of natural medicines* **2014**, *68*, 1.
- [121]D. H. Kwan, F. Schulz, *Molecules* **2011**, *16*, 6092.
- [122]T. Stachelhaus, H. D. Mootz, M. A. Marahiel, *Chem. Biol.* **1999**, *6*, 493.

## Chapter 2 Myxoquaterines

### 2.1 Activity guided discovery

The growing problem of antibiotic resistance is a well-known phenomenon of our present time. In order not to lose the lead, antibiotics granted us against infective diseases, it is necessary to continuously develop novel antibiotics. Since the majority of anti-infective drugs are based on or derived from natural products, the search for new compounds with antimicrobial activity is an important contribution. So far, most of the known antibiotics were discovered from *Streptomyces* species,<sup>[1]</sup> while other bacterial clades, such as Myxobacteria, are still underexplored. These soil dwelling bacteria have turned out to be a valuable source of new chemical scaffolds with antimicrobial activity.<sup>[2]</sup> Therefore we continuously screen soil samples for novel myxobacterial isolates. New strains are added to our strain-collection, cultivated and tested for antimicrobial activity. In our ongoing screening effort, one myxobacterial strain from our collection came to our attention. The strain MSr 11954 showed a *Streptomyces*-like phenotype, but 16s RNA sequence analysis and whole genome sequencing confirmed that it represents a novel genus inside the clade of *Sorangineae* (figure 1).

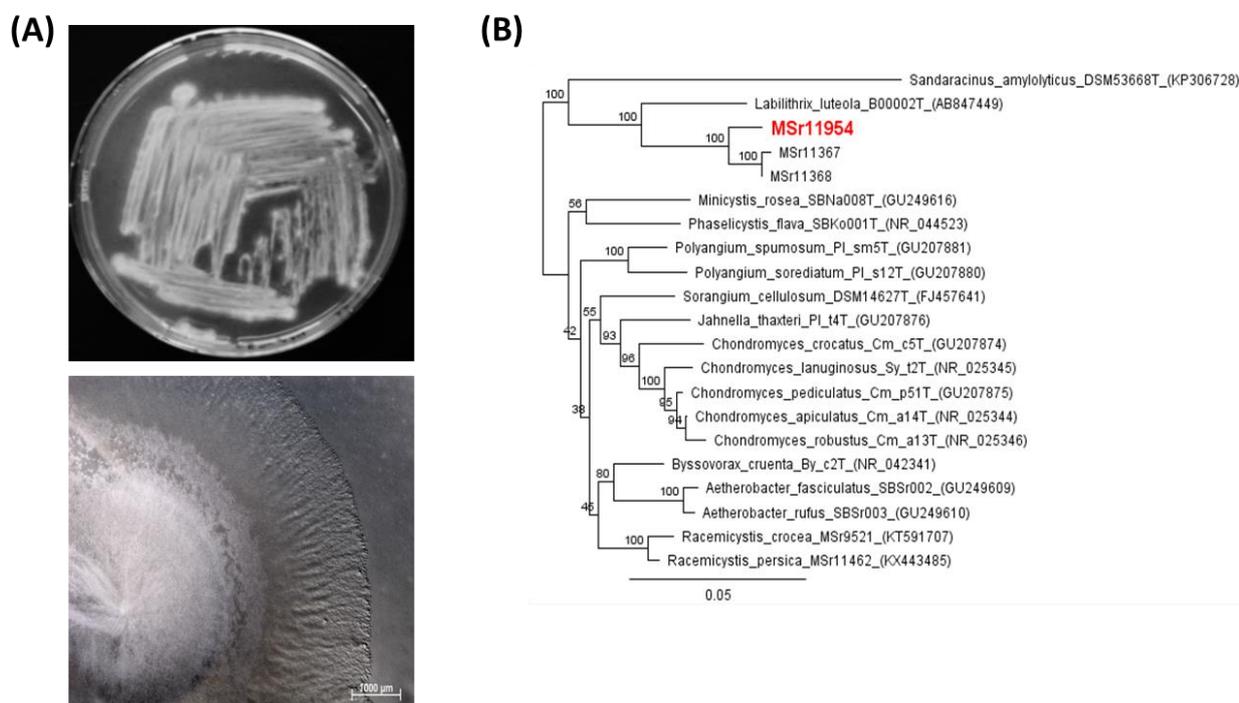


Figure 1: (A) phenotype of MSr 11954 on agar plate (B) phylogenetic analysis of MSr 11954 (pictures and phylogenetic analysis kindly provided by Dr. Ronald Garcia)

Organic extracts of initial liquid cultivations showed fungicidal and anti-gram positive bioactivity against our indicator strains (*M. hiemalis* and *M.luteus*). In order to identify the bioactive compounds produced by MSr 11954, a LC-hrMS<sup>2</sup> coupled fractionation was performed. Since the focus was set on the antifungal bioactivity, *M. hiemalis* was chosen as indicator strain. Knowing the collection time of the fractions with bioactivity, it was possible to correlate the bioactivity with a group of compounds represented by specific masses in the LC-hrMS<sup>2</sup> chromatogram as shown in figure 2.

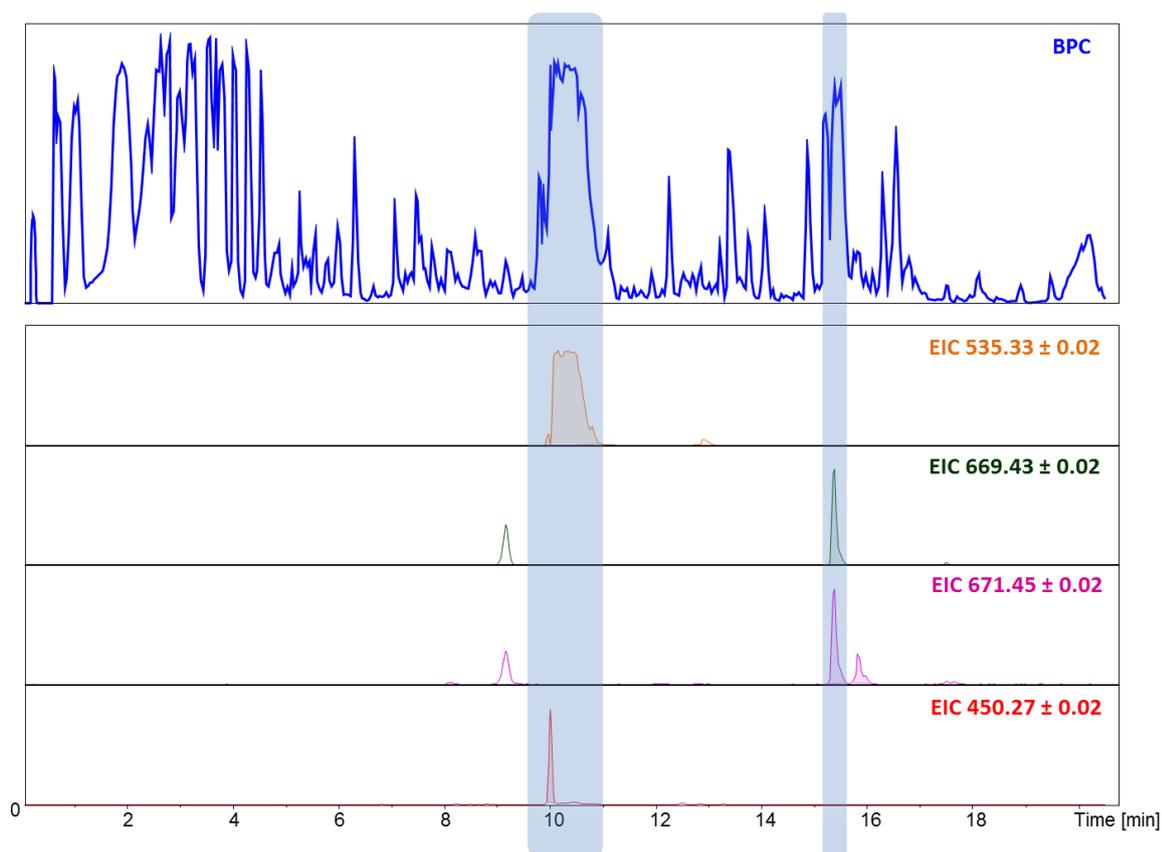


Figure 2: Results of fractionation showing the base peak chromatogram BPC of the crude extract displayed as blue chromatogram. Areas of fractions showing activity against *M. hiemalis* are marked in blue. Masses found in the active areas are displayed as their extracted ion chromatogram (EIC) traces.

The identified signals exhibited some common features: (1) The UV spectra of the signals showed absorption maxima around 340 nm. (2) The double charged ion mass  $[M+2H]^{2+}$  was more intense than the corresponding single charged mass peak  $[M+H]^+$ . (3) The retention time of the signals showed a strong dependency on the pH of the chosen eluent system.

Dereplication using our *inhouse* database “Myxobase”<sup>[3]</sup> and the dictionary of natural products<sup>[4]</sup> showed no significant hits for the identified masses. Therefore, we decided to isolate the molecules for further characterization. To do so, 12 l of MSr 11954 were fermented with XAD-16 adsorber resin. The molecules were isolated using liquid/liquid partitioning, flash chromatography and semi preparative HPLC.

During the isolation process, double peaks for each of the molecules were observed in the HPLC chromatograms (figure 3). Since the double peaks in each case had the same mass, they were expected to be stereoisomers. Except for the molecule with the mass to charge ratio  $m/z = 450.27$ , both peaks could be isolated. In the following the first peak is referred to as variant A while the second peak is variant B. Structure elucidation using 2D *hr*NMR measurements and *hr*MS measurements revealed that the isolated compounds belong to one compound family. Except for one derivative with the mass-to-charge ratio  $m/z = 450.27$ , all derivatives contained a characteristic quaternary ammonium ion. Therefore, this compound family was named myxoquaterines.

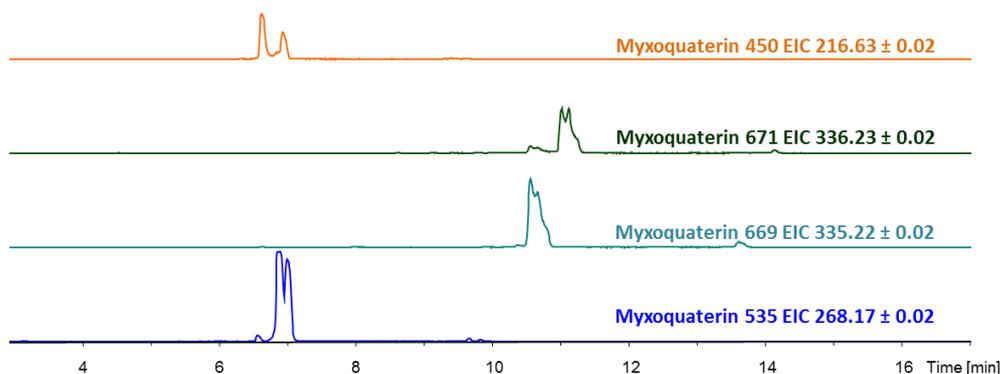


Figure 3: LC-hrMS chromatogram of a crude extract of MSr 11954 showing the extracted ion chromatogram (EIC) traces of myxoquaterin 450 (orange) myxoquaterin 671 (green), myxoquaterin 669 (turkis) and myxoquaterin 535 (blue)

## 2.2 Structure elucidation

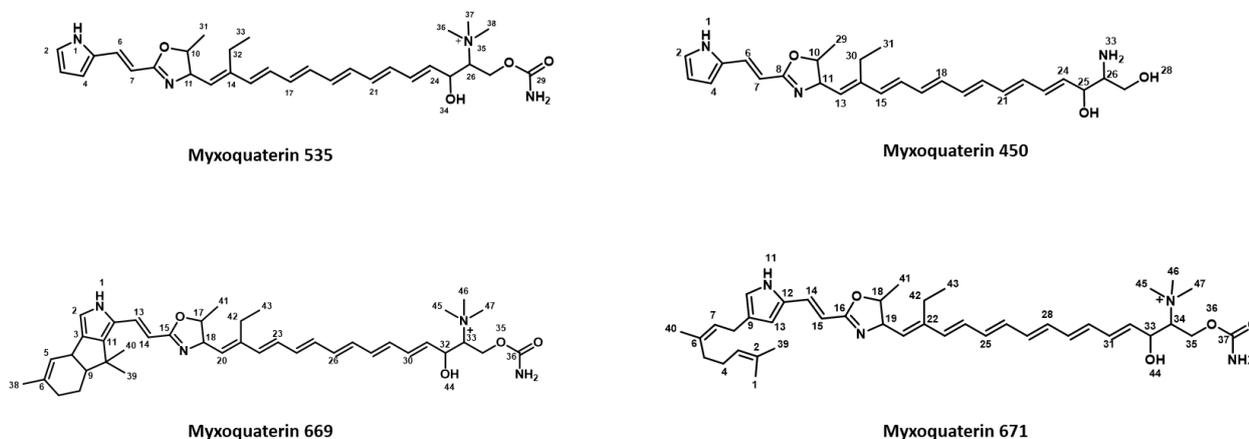


Figure 4: Planar structures of Myxoquaterine variants 535 (1), 450 (2), 669 (3) and 671 (4)

### Myxoquaterine 535

Myxoquaterin 535 (**1**) was isolated as brown amorphous solid with a characteristic fishy smell. HRESIMS of **1** displayed an  $[M]^+$  peak at  $m/z$  535.32886 (calc. for  $C_{31}H_{43}N_4O_4$ , 535.3278,  $\Delta = 1.8$  ppm), consistent with the molecular formula  $C_{31}H_{43}N_4O_4$ .

The  $^1H$  NMR spectrum of **1** in  $CD_3OD$  exhibits five signals characteristic for hetero-aromatic moieties at  $\delta$  6.16 ( $H-7$ ),  $\delta$  6.17 ( $H-3$ ),  $\delta$  6.43 ( $H-4$ ),  $\delta$  6.90 ( $H-2$ ) and  $\delta$  7.26 ( $H-6$ ). In addition, two signals at  $\delta$  5.36 and  $\delta$  5.75 in combination with a group of overlapping signals between  $\delta$  6.23-6.45 indicate the presence of a polyene chain. Furthermore, three signals at  $\delta$  3.62 ( $H-26$ ),  $\delta$  4.56 ( $H-10$ ) and  $\delta$  4.71 ( $H-25$ ) are characteristic for methine protons substituted with oxygen or nitrogen. Additionally, a fourth methine group was identified in the HSQC spectrum ( $\delta_{C-10}$  83.2,  $\delta_{H-10}$  4.39) overlapping with signals from a diastereotopic methylene group ( $\delta_{C-27}$  61.6,  $\delta_{H-27}$  4.47/4.39). In both cases, the proton chemical shift values indicate hetero-aromatic substituents. The HSQC spectrum shows the presence of a second diastereotopic methylene group ( $\delta_{C-32}$  20.8,  $\delta_{H-32}$  2.39/2.45). Apart from this, three signals characteristic for methyl protons were found. One signal at  $\delta$  3.37 ( $H_3-36,37,38$ ) is an intense singlet indicating presence of several magnetic equivalent methyl groups attached to a nitrogen atom. The two other signals, a doublet at  $\delta$  1.43 ( $H_3-31$ ) and a triplet at  $\delta$  1.19 ( $H_3-33$ ), indicate two methyl groups connected to a hydrocarbon scaffold based on the chemical shift value and the

multiplicity.

Examination of COSY crosspeaks indicate that the three signals at  $\delta$  6.17 (H-3),  $\delta$  6.43 (H-4) and  $\delta$  6.90 (H-2) belong to a spin system. In addition, all three protons show HMBC correlations to a quaternary carbon at  $\delta$  129.7 (C-5). N-HMBC correlations from H-2 to a nitrogen at  $\delta$  149.3 clearly reveal these signals to be derived from a pyrrole ring. The methyl group at  $\delta$  1.43 (H<sub>3</sub>-31) is part of a spin system consisting of H-10, H-11 and H-13 as demonstrated by COSY and TOCSY correlations. In combination with HMBC correlations to a quaternary carbon at  $\delta$  166.8 and the characteristic shift values, this structural moiety can be deduced as methyloxazoline ring with an adjacent sp<sup>2</sup> hybridized methine, as shown by COSY correlations between H-11 and H-13. Further examination of TOCSY crosspeaks revealed the presence of a larger spin system consisting of a polyene chain, as indicated by the characteristic proton resonances of H-15 to H-24. A selective 1D TOCSY experiment with the resonance of H-24 ( $\delta$  5.75) as selective excitation frequency could deduce that all protons from H-15 to H<sub>2</sub>-27a/b are part of this spin system. The sequential order of H-24 to H-27a/b was established based on COSY correlations. The characteristic shift values of methine proton H-25 ( $\delta$  4.71) and the diastereotopic methylene protons H<sub>2</sub>-27a/b ( $\delta$  43.9/4.47) are indicative for adjacent oxygen, while the proton and carbon chemical shift values of CH-26 ( $\delta_{\text{C-26}}$  76.5,  $\delta_{\text{H-26}}$  3.62) indicate a neighboring nitrogen. Strong HMBC correlations from H-25 to three chemically equivalent methyl carbons at  $\delta$  54.9 (C-36, C-37, C-38) imply that the methyl residues are bound to a nitrogen. Indeed, the N-HMBC spectrum showed strong correlations from H<sub>3</sub>-36,37,38 to a nitrogen at  $\delta$  53.0 (N-30). The carbon and proton chemical shift values resemble those observed for the trimethyl-ammonium ion found in ergothioneine<sup>[5]</sup>, suggesting the same structural feature is present in myxoquaterine 535. This finding is supported by MS<sup>2</sup> fragmentation, since MS<sup>2</sup> fragmentation of the major ion peak 535.32 [M]<sup>+</sup> shows an intense signal at 476.25 *m/z*, which corresponds to a fragment where the trimethyl-ammonium group is cleaved off. Additionally, the native positive charge of the molecule from the trimethyl-ammonium ion leads to detection of the formic acid adducts 625.37 [M<sup>+</sup> + [2FA]]<sup>-</sup> and 579.36 [M-H+[FA]]<sup>-</sup> in MS ESI negative mode (supplementary figure 2). Apart from the trimethyl-ammonium ion, the molecule is modified with a carbamate group at position 27 as indicated by HMBC correlations from H-27a/b to a carbon at  $\delta$  157.7 (C-29). The presence of this functional group was corroborated by MS<sup>2</sup> fragmentation: From the major peak 535.32 [M]<sup>+</sup>, a fragment ion of 415.23 *m/z* can be observed which corresponds to the molecule with the trimethyl-ammonium and the carbamate being cleaved off (figure 5B).

The connectivity between the pyrrole ring, the methyloxazoline ring and the polyene chain was established by inspection of the remaining signals: H-6 ( $\delta$  7.62) and H-7 ( $\delta$  6.16) are connected by a double bond, as shown by COSY crosspeaks and characteristic shift values. The large coupling constant of 16 Hz suggests an *E* configuration. In addition, starting from H-6, HMBC correlations can be observed to the carbon resonances of C-5 and C-8, indicating the double bond as link between the pyrrole ring and the methyloxazoline. An identical scaffold can be found in the DKxanthen compound family<sup>[6]</sup>; comparison of the chemical shift values suggest the same structural motive.

The sp<sup>2</sup> hybridized methine proton at  $\delta$  5.36 (H-13) connected to the methyloxazoline ring, shows an HMBC correlation to the corresponding carbon of position 15 ( $\delta$  136.6) and vice versa. Since H-15 and H-11 show HMBC correlations to a quaternary carbon at  $\delta$  144.2 (C-14), it is indicated C-14 is located between position 13 and 15. Moreover, HMBC correlations from the ethyl group ( $\delta_{\text{H3-33}}$  1.19 and  $\delta_{\text{H-32a/b}}$  2.39/2.45) to the same quaternary carbon show that the ethyl group is connected to C-14. The complete structure strongly resembles the DKxanthenes, with major differences at the C-terminal part with a quaternary ammonium ion and a carbamate group instead of asparagine.

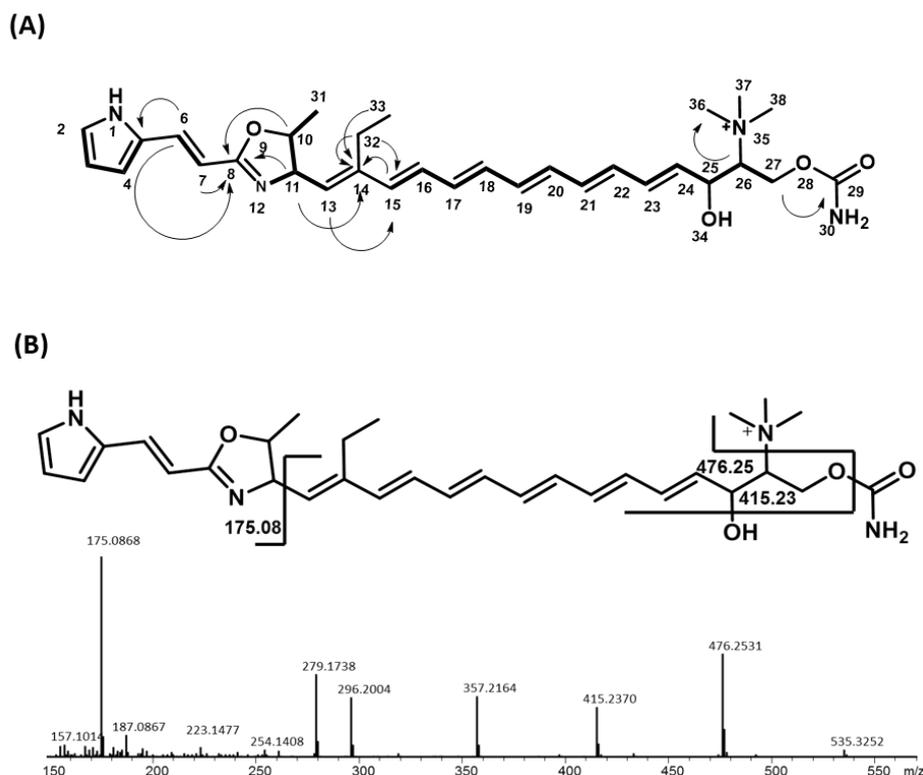


Figure 5: (A) key COSY (bold line) and HMBC (arrow) correlations of myxoquaterin 535; (B) hrMS<sup>2</sup> spectrum and observed fragments of Myxoquaterin 535. MS<sup>2</sup> spectrum is extracted from LC-hrMS measurement of an extract of MSr 11954.

## Myxoquaterin 450

Myxoquaterin 450 (**2**) was isolated as a brown amorphous solid. HRESIMS of **2** displayed an [M+H]<sup>+</sup> peak at  $m/z$  450.27642 (calc. for C<sub>27</sub>H<sub>36</sub>N<sub>3</sub>O<sub>3</sub>, 450.27512,  $\Delta$  = 3 ppm), consistent with the molecular formula C<sub>27</sub>H<sub>36</sub>N<sub>3</sub>O<sub>3</sub>.

Comparison of NMR features for **2** showed it to be highly similar to **1**, with the main difference that the singlet of the trimethyl-ammonium moiety was missing and the chemical shift values of the diastereotopic methylene protons H-27a/b as well as H-26 and H-25 were shifted upfield compared to **1**. In combination with the 2D NMR experiments HSQC and HMBC, the corresponding carbon chemical shift values were obtained. These data, together with the obtained sum formula, show that the trimethyl-ammonium moiety at position 26 is replaced by an amine and that the carbamate group attached to position 27 is missing in **2**.

Table 1: spectroscopic data of myxoquaterin 535 variant A and myxoquaterin 450 acquired in CD<sub>3</sub>OD

535			450		
position	$\delta_c^a$	$\delta_H^b$ (J in Hz)	position	$\delta_c^c$	$\delta_H^d$ (J in Hz)
1	<sup>15</sup> N 149.3 <sup>e</sup>		1	NH	
2	123.2	6.90 d,d (1.3, 2.6)	2	123.2	6.89 d,d (2.3,1.2)
3	110.5	6.17 d,d (2.7,3.4)	3	110.6	6.16 d,d (2.6,3.8)
4	114.4	6.43 m	4	114.5	6.42 d,d (3.9,1.7)
5	129.7		5	129.6	

6	132.7	7.26 d (16.0)	6	132.5	7.25 d (15.8)
7	106.8	6.16 d (16.2)	7	107.2	6.16 d (15.9)
8	166.8		8	166.7	
9			9	O	
10	83.2	4.39 m	10	83.2	4.39 d,q (6.4,6.6)
11	71.0	4.56 d,d (9.7,7.3)	11	70.9	4.56 d,d (9.7,7.2)
12	N		12	N	
13	131.9	5.36 d (9.5)	13	132.0	5.36 d (9.7)
14	144.2		14	144.2	
15	136.6	6.23 d (15.3)	15	136.5	6.23 d (15.3)
16	135.4	6.43 m	16	135.1	6.44 m
17-21	132-136	6.30-6.50	17-22	131.3-134.5	6.29-6.54
22	131.9	6.32 m			
23	135.9	6.45 m	23	134.8	6.44 m
24	131.8	5.75 d,d (7.6,15.0)	24	132.2	5.69 d,d (15.3,7.2)
25	71.3	4.71 d,d (8.1,8.1)	25	70.5	4.22 m
26	76.5	3.62 m	26	58.6	3.04 m
27a	61.6	4.47 m	27a	60.1	3.72 m
27b		4.39 m	27b		3.59 m
28	O		28	OH	
29	157.7		29	19.8	1.44 d (6.2)
30	<sup>15</sup> N 53.0 <sup>e</sup>		30a	20.8	2.40 m
31	19.8	1.43 d (6.4)	30b		2.44 m
32a	20.8	2.39 m	31	14.7	1.12 t (7.6)
32b		2.45 m	33	N	
33	14.9	1.19 t (7.7)			
35	N 53.0				
36-38	54.9	3.37 s			

<sup>a</sup>acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

<sup>b</sup>acquired at 700 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

<sup>c</sup>acquired at 125 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

<sup>d</sup>acquired at 500 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

<sup>e</sup><sup>15</sup>N resonance obtained from <sup>15</sup>N-HMBC acquired at 71 MHz

## Myxoquaterine 669

Myxoquaterine (**3**) was isolated as brownish solid that had a characteristic fish smell. HRESIMS of **3** displayed an [M]<sup>+</sup> peak at  $m/z$  669.44082 (calc. for C<sub>41</sub>H<sub>57</sub>N<sub>4</sub>O<sub>4</sub>, 669.43743,  $\Delta$  = 5 ppm), consistent with the molecular formula C<sub>41</sub>H<sub>57</sub>N<sub>4</sub>O<sub>4</sub>.

The <sup>1</sup>H NMR spectrum of **3** showed significant similarities to the spectra of the previously described molecules **1** and **2**. Major differences were found for signals corresponding to the pyrrole ring. There, only one singlet at  $\delta$  6.47 (H-2) was observed. However, several new signals appeared: One signal with a chemical shift value characteristic for an sp<sup>2</sup> hybridized methine group at  $\delta$  5.66 (H-5) and several multiplets between  $\delta$  3.59-1.28 were identified. Examination of the HSQC spectrum identified these signals as two methine groups ( $\delta_{C-4}$  38.6,  $\delta_{H-4}$  3.59 and  $\delta_{C-9}$  56.8,  $\delta_{H-9}$  2.17) and two diastereotopic methylene groups ( $\delta_{C-8}$  30.6,  $\delta_{H-7}$  1.86/1.96 and  $\delta_{C-8}$  23.3,  $\delta_{H-8}$  1.74/1.28). In addition, three new methyl singlets at  $\delta$  1.67 (H<sub>3</sub>-38),  $\delta$  1.33 (H<sub>3</sub>-39) and  $\delta$  1.27 (H<sub>3</sub>-40) appeared in the <sup>1</sup>H NMR spectrum.

Analysis of TOCSY correlations showed that the new signals from the methylene and methine protons form a new spin system in addition to the spin system identified in **1** and **2**. The respective connectivity was elucidated by analysis of COSY crosspeaks and HMBC correlations to be a six membered terpenoic ring (figure 6). H-5 and H-7 show reciprocal HMBC correlations to their corresponding carbon resonances at  $\delta$  122.8 (C-5) and  $\delta$  30.6 (C-7). An  $sp^2$  hybridized carbon at  $\delta$  134.5 (C-6) is located in between as indicated by HMBC correlations from H-7a/b and H-8a/b. Me-38 is bound to C-6 as shown by HMBC correlations from  $\delta$  1.67 (H<sub>3</sub>-38) to  $\delta$  134.5 (C-6) and neighboring carbons at  $\delta$  122.8 (C-5) and  $\delta$  30.6 (C-7). Close inspection of the HMBC spectrum revealed the presence of an additional quaternary carbon at  $\delta$  43.1 (C-10). Correlations from H-9, H<sub>3</sub>-39 and H<sub>3</sub>-40 to this quaternary carbon C-10 suggest that these positions are adjacent to C-10. The number of identified carbons as well as the elucidated structural features are typical for monoterpene ring systems [7].

The connectivity of the terpenoic moiety to the rest of the molecule was established by HMBC correlations: HMBC correlations from H-2 and H-13 to two additional quaternary carbons in the pyrrole ring compared to **1** and **2** suggest that this is the linker region to the terpenoic moiety. Indeed, HMBC correlations of the terpenoic proton at  $\delta$  2.17 (H-9) to both  $\delta$  134.1 (C-3) and  $\delta$  142.7 (C-11) as well as further HMBC correlations from H<sub>3</sub>-39 and H<sub>3</sub>-40 to C-11 suggest the six membered terpenoic ring is connected to the pyrrole via a five membered ring with C-10 as additional atom in the ring (figure 6). No HMBC correlations were observed from H-4 or H-5 to C-3, but the determined sum formula, and other observed HMBC correlations do not allow any other connectivity.

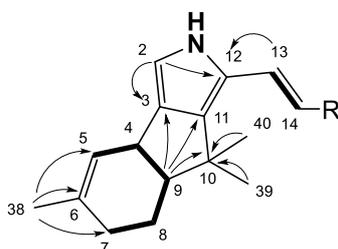


Figure 6: key COSY (bold line) and HMBC (arrows) correlations of Myxoquaterine 669

Table 2: Spectroscopic data of myxoquaterin 669 variant A and myxoquaterine 671 variant A acquired in CD<sub>3</sub>OD

Myxoquaterin 669			Myxoquaterin 671		
position	$\delta C^a$	$\delta H^b$	position	$\delta^{13}C^a$	$\delta^1H^b$
1	N		1	17.6	1.61 s
2	115.1	6.47 s	2	131.9	
3	134.1		3	125.2	5.11 t 6.5
4	38.6	3.59 m	4	27.4	2.12
5	122.8	5.66 m	5	40.5	2.05
6	134.5		6	135.7	
7a	30.6	1.96 m	7	124.9	5.32 t 6.6
7b		1.86 m	8	26.1	3.15 d 7.1
8a	23.3	1.74 m	9	126.4	
8b		1.28 m	10	120.8	6.64 s
9	56.8	2.17 m	11	NH	
10	43.1		12	129.7	
11	142.7		13	114.1	6.25 s
12	121.3		14	132.5	7.18 d 16.1

13	130.9	7.25 d 15.9	15	106.6	6.11 d 16.8
14	105.2	5.99 d 16.0	16	166.7	
15	167		17	O	
16	O		18	71.0	4.54 m
17	83.2	4.38 m	19	83.1	4.38 m
18	70.9	4.54 d,d 9.6, 7.3	20	N	
19	N		21	132.1	5.36 d 9.2
20	132.2	5.36 d 9.6	22	144.1	
21	144.2		23	136.7	6.23 d 14.7
22	136.8	6.23 d 15.4	24-31	131.7-136.9	6.46-6.43
23	135.5	6.43 m	32	131.6	5.75 d,d 14.9,7.8
24-29	132.0-136.7	6.29-6.45	33	71.0	4.71 d,d 8.2,8.8)
30	135.7	6.45 m	34	76.2	3.64 m
31	131.4	5.75 d,d 14.9,7.8	35a	61.3	4.48 m
32	71.0	4.71 d,d 8.26,8.28	35b	61.3	4.39 m
33	76.3	3.63 m	36	O	
34a	61.4	4.47 d,d 14.1, 2.5	37	n.d.	
34b		4.38 m	38	N	
35	O		39	25.7	1.67 s
36	157.3		40	15.8	1.67 s
37	N		41	19.9	1.43 d 5.9
38	23.6	1.67 s	42	20.8	2.4 m
39	23.9	1.33 s			2.37 m
40	30.1	1.27 s	43	14.	1.11 t 7.9
41	19.9	1.44 d 6.1	44	OH	
42a	20.7	2.38 m	45	54.9	3.37 s
42b	20.5	2.47 m	46	54.9	3.37 s
43	14.6	1.12 t 7.6	47	54.9	3.37 s
44	N				
45	55.0	3.37 s			
46	55.0	3.37 s			
47	55.0	3.37 s			

<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

<sup>b</sup> acquired at 700 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

### Myxoquaterine 671

Myxoquaterin 671 (**4**) was isolated as a brown solid with a characteristic fish like odor. HRESIMS of **4** displayed an [M]<sup>+</sup> peak at  $m/z$  671.45425 (calc. for C<sub>41</sub>H<sub>57</sub>N<sub>4</sub>O<sub>4</sub>, 671.45308,  $\Delta$  = 1.7 ppm), consistent with the molecular formula C<sub>41</sub>H<sub>59</sub>N<sub>4</sub>O<sub>4</sub>.

<sup>1</sup>H NMR data showed major similarities to the spectrum of **3**. The identified differences include three methyl singlets at  $\delta$  1.61 (H<sub>3</sub>-1) and 1.67 (H<sub>3</sub>-39 and H<sub>3</sub>-40) as well as three methylene protons at  $\delta$  2.12 (H<sub>2</sub>-4), 2.05 (H<sub>2</sub>-5) and 3.15 (H<sub>2</sub>-8) and two methine protons at  $\delta$  5.11 (H-3) and 5.32 (H-7). In addition, two signals at  $\delta$  6.64 (H-10) and  $\delta$  6.25 (H-12) were found and assigned to the pyrrole ring based on the characteristic shift values. All newly found proton signals from except H-10 and H-12 from thy pyrrole ring form a spin system, as indicated by inspection of COSY and TOCSY crosspeaks. Based on the shift values including the carbon resonances obtained from HSQC and HMBC spectra, the signals were deduced as geranyl-chain. Connectivity to the pyrrole could be established by inspection of the HMBC spectrum: HMBC correlations from H-8 in the geranyl moiety to carbon resonances at  $\delta$  126.4 (C-9) and  $\delta$  120.8 (C-10) clearly showed the connectivity of the geranyl chain to the pyrrole ring. Further HMBC correlations from pyrrole protons H-9 and H-13 to C-9

clearly indicate C-9 is also part of the pyrrole ring.

### 2.2.1 Difference between variant A and B

NMR spectra of all myxoquaterine variants B (second peak for all mass traces in figure 3) strongly resemble its counterparts from variant A. The only NMR chemical shift deviations between the two variants were located in the area of the ethyl group and the methyloxazoline ring. A comparison of ROESY data for both variants of **1** identified variants A and B as *E/Z* isomers. As indicated by specific ROE correlations, in particular between H-32a/b and H-11 and H-15 as well as between H-15 and H-13, the double bond between the methyloxazoline ring and the ethyl group is *E*-configured in variant A (figure 7A). In contrast, this double bond is *Z*-configured in variant B as indicated by the ROE correlations between H-32a/b and H-13 as well as between H-15 and H-11 (figure 7B). The ROE correlation between H-32a/b and H-16 can be observed in both variants. Since all A variants of the myxoquaterine derivatives show deviations of less than 0.3 ppm for the carbon resonances in this area, it was concluded that all A-variants are *E*-configured. The same accounts for the B-variants that were concluded to be *Z*-configured. A conversion of variant A into variant B was not observed.

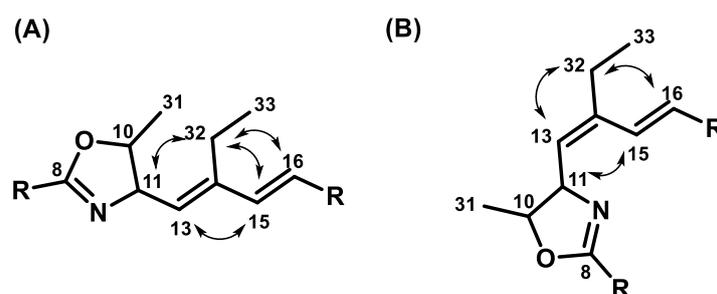


Figure 7: Part of myxoquaterine structures that differ in variant A and B. (A) Variant A, arrows indicate key ROESY correlations. (B) Variant B, arrows indicate key ROE correlations.

### 2.2.2 Stereochemical analysis

The configuration of the double bonds in the polyene could not be determined completely since overlapping signals impeded an analysis of all coupling constants. All values that were determined ( $^3J(\text{H-15};\text{H-16}) = 15.3$  Hz and  $^3J(\text{H-23};\text{H-24}) = 15.0$  Hz in myxoquaterine 535) indicate an *E*-configuration.

**1**, **2** and **4** contain four stereocenters, while **3** contains two additional ones within the cyclized terpene moiety. In all myxoquaterine derivatives, the stereocenters appear as pairs where two stereocenters are connected via a single bond. Therefore, the determination of the relative configuration of the neighboring stereocenters was achieved using ROESY data in combination with homo- and heteronuclear coupling constants.

For **1**, the relative configuration between the quaternary ammonium at C-26 and the hydroxyl group at C-25 was determined by *J*-based configuration analysis, including hetero nuclear  $^{2/3}J_{\text{CH}}$  couplings<sup>[8]</sup>. This method was described for various substituents at neighboring stereocenters, including hydroxyl groups, amines and alkylated amines. Its application was described for a scaffold that was similar to the one found in **2** with the free amine.<sup>[9]</sup> But since **2** was isolated in a low yield, the method was applied on **1**. The hetero nuclear coupling constants were obtained using a dipsi2Hetloc NMR experiment.

Hetloc (dipsi2)		
$^3J(\text{H-25,H-26})$	8.1 Hz	large
$^3J(\text{H-25,C-27})$	2.4 Hz	small
$^3J(\text{C-24,H-26})$	1.2 Hz	small
$^2J(\text{C-25,H-26})$	5.0 Hz	large
$^2J(\text{C-26,H-25})$	4.8 Hz	large

acquired at 700 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

Using the reference data<sup>[8,9]</sup>, the orientation of the hydroxyl group to the quaternary ammonium ion can be assigned as gauche (figure 8). In addition, the observed ROESY correlations between H-24 and H-27a/b support this assignment. Therefore, the relative configuration between these two chiral centers was determined as erythro.

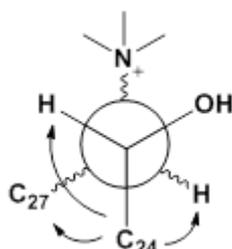


Figure 8: Newman projection of the quaternary ammonium and the hydroxyl group showing assignment obtained by comparison with the reference data. ROE correlations (arrows) support the assignment.

For the assignment of the relative configuration between C-10 and C-11, another approach based on the comparison of carbon chemical shifts was used. The methyloxazoline ring and the proximate pyrrole ring system can also be found in the myxobacterial compound family of the DKxanthenes<sup>[6]</sup>. Comparison of carbon resonances showed differences of maximal 0.5 ppm implying the same relative configuration of the compounds. Analysis of ROE correlations showed a strong correlation between methyl protons H<sub>3</sub>-31 and H-11 whereas no correlations was observed between H-10 and H-11 indicating that H-10 and H-11 are arranged in anti-configuration (figure 9). Similar observations were described for the DKxanthenes.

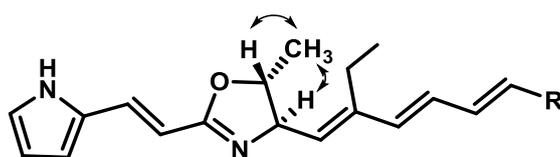


Figure 9: ROE correlations (arrows) indicate anti-configuration of both hydrogens in the methyloxazoline moiety.

In **3**, the stereocenters at C-4 and C-9 connect the outer terpene ring with the middle ring. Strong ROE correlations can be observed between H-4 and H-9 indicating that both protons point in the same direction. In addition, the methyl protons H<sub>3</sub>-40 show strong correlations with both protons (figure 10). This finding suggests that the outer ring is oriented in a half-chair conformation. Due to missing ROE correlations to other non-terpenic parts of the molecule, a further analysis was not possible.

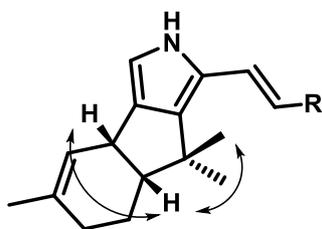


Figure 10: ROESY correlations (arrows) indicate both hydrogens are oriented in the same direction in the terpenic rings of myxoquaterin 669

To determine the absolute configuration of the hydroxyl group in **1**, we attempted a Mosher esterification. Unfortunately, the derivatization with the chiral reagent was unsuccessful. Even though LC-MS reaction control indicated a product formation with the expected mass within minutes, the product degraded within a few hours. In order to analyze the reaction product, 2D NMR datasets of the reaction mixtures of the putative R- and S-MTPA esters were acquired immediately. The NMR data showed alteration of carbon resonances in the pyrrole ring while the carbon resonance at the hydroxyl group C-25 was unchanged. These findings suggest that the MTPA-chlorides reacted at the pyrrole ring instead of the hydroxyl group. Therefore, this method could not be used to assign the stereochemistry of the hydroxyl group.

### 2.3 Biosynthesis of myxoquaterines

Illumina sequencing of isolated genomic DNA of MSr 11954 to obtain the genome sequence yielded 68 contigs, which were assembled using abyss-pe v.2.10<sup>[10]</sup>. Total size of the genome was 12.2 Mbp. Analysis of the genome sequence using the antiSMASH tool<sup>[11]</sup> identified a BGC with 35 % identity with the DKxanthen BGC. Cluster borders, automated annotation of open reading frames and functional domains was provided by antiSMASH. Additionally, all open reading frames expected to be in the Cluster were analysed using the automated antiSMASH annotation and manual domain analysis tools provided by Pfam and HHpred<sup>[12][13]</sup>. The analysis showed that the cluster mostly contains type I PKS genes. Two ORFs, *mxqB* and *mxqG*, also contain NRPS type catalytic domains. In addition, some ORFs contain domains expected to be involved in tailoring reactions like methylation or carbamoylation.

Four ORFs do not contain biosynthesis genes but transporter genes and a regulatory gene were identified. A majority of the genes show significant homology to genes from the DKxanthen biosynthetic pathway (table 6)<sup>[6,14]</sup>. Therefore, the biosynthesis of the myxoquaterines is expected to happen in a similar manner: The biosynthesis is initiated by adenylation of proline by MxqB, an adenylation domain-containing protein, and oxidation by the dehydrogenase MxqA, to generate pyrrole-carboxylic acid as starter unit. In the second step, the starter unit is transferred onto the ACP domain of MxqE. Like described for the DKxanthen assembly chain, no elongation step is expected to happen here. Instead, MxqF conducts the first elongation step. Like in DKxanthen biosynthesis, a discrete DH domain is missing suggesting another DH domain to act in-trans. The methyloxazoline moiety is formed by MxqG, a NRPS-PKS hybrid. Here, ethyl-malonyl-CoA is used as polyketide extender unit. The ethyl-malonyl-CoA is provided by MxpD, a crotonyl-CoA reductase carboxylase (CCR), as indicated by the antiSMASH annotation. In further steps of the biosynthesis, MxqM and MxqN elongate the polyketide chain in a similar manner as in the DKxanthen biosynthesis. The last elongation step is conducted by MxqO. Biosynthetic domains for polyketide elongation found by antiSMASH (KS,AT and ACP) suggest formation of a  $\beta$ -keto thioester (figure 12). A thioester domain is missing suggesting an alternative release mechanism. Myxoquaterin 450 (**2**) is an intermediate of the biosynthesis after chain release. For-

mation of **2** out of the expected thioester would require following reactions according to typical PKS biosynthetic logic: (1) Reduction of the  $\beta$ -ketone into an alcohol, (2) an oxidation of the  $\alpha$ -position into an  $\alpha$ -ketone followed by (3) a transamination reaction and (4) reduction/release of the thioester into an alcohol. In addition to the PKS domains, a domain was annotated by the antiSMASH tool as aminotransferase in MxqO. Search for other domains without assigned functions that could catalyze the required reactions yielded MxqQ, which contains an oxidoreductase domain according to antiSMASH and MxqH that was annotated as reductase domain by antiSMASH. Other necessary domains with a catalytic function to form the released intermediate **2** could not be identified. Search for homologues of reductive domains known to conduct a TE independent reductive release using corresponding domains from myxochelin, myxalamid A, saframycin and gramicidin biosynthesis did not yield any hit<sup>[15]</sup>.

Therefore, a release mechanism similar to the fungal mycotoxin fumonisin and the antibiotic prodigiosin was considered<sup>[16]</sup>. In both cases, an intermediate is released from an ACP domain by a PLP mediated transfer of an amino acid. Feeding experiments with deuterated L-serine showed a shifted isotope pattern indicating incorporation. In addition, feeding experiments with 1-<sup>13</sup>C labelled L-serine did not show any differences between supplemented culture and control culture suggesting that formation of the C-C bond between the polyketide chain and L-serine involves decarboxylation of L-serine as described in fumonisin and prodigiosin and the domain in MxqO with the PLP-binding site acts as PLP-dependent acyltransferase.

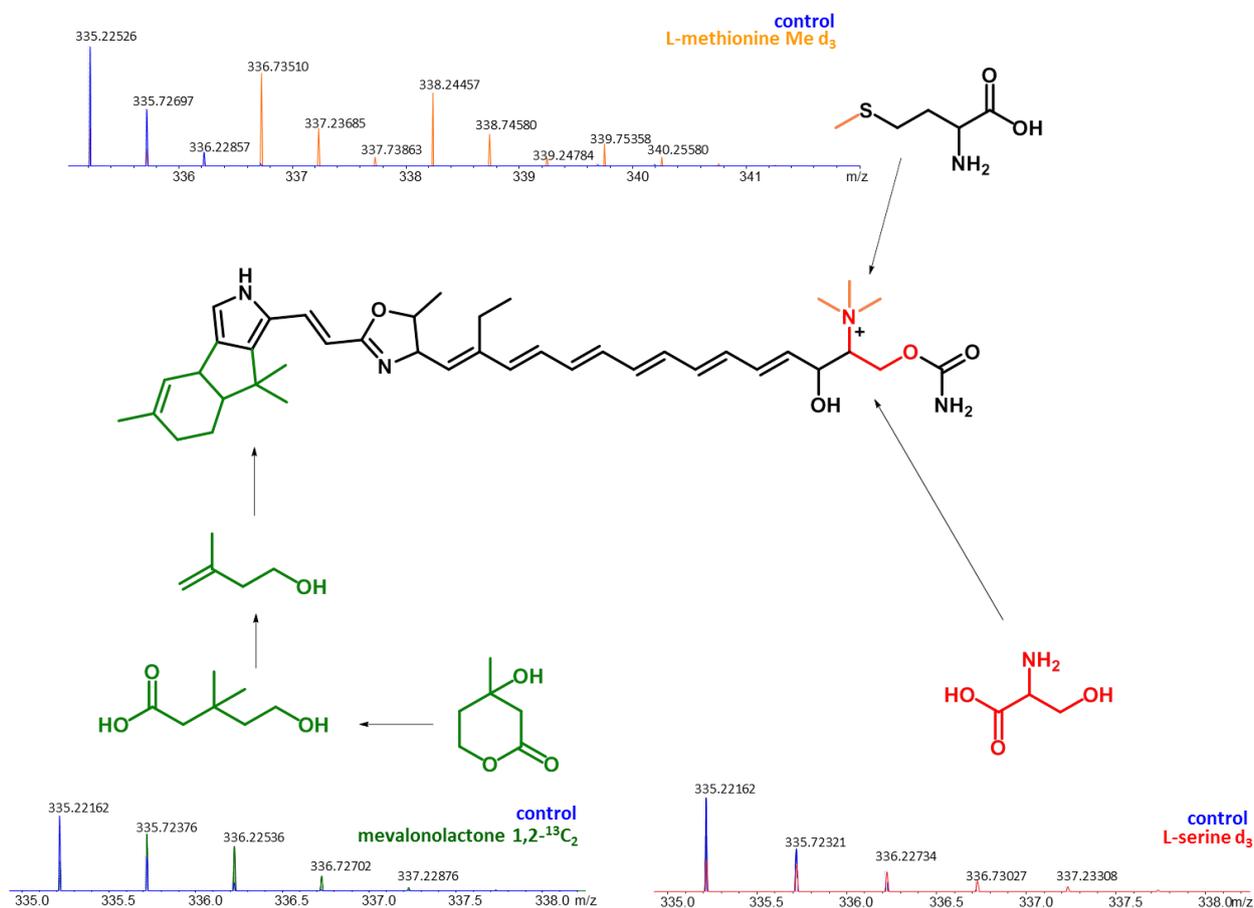


Figure 11: Isotope pattern of Myxoquaterin 669 showing double charged mass  $[M+2H]^{2+}$  extracted from LC-hrMS from the control culture (blue signals) and a culture supplemented with L-methionine Me  $d_3$  (orange signals), mevalonolactone 1,2-<sup>13</sup>C<sub>2</sub> (green signals) and L-serine  $d_3$  (red signals).

The intermediate **2** is further processed by two tailoring reactions: A carbamoyltransferase MxqR transfers a carbamate residue to the primary alcohol. The amine is methylated threefold by a SAM dependent methyltransferase MxqP. Feeding of methionine-Me  $d_3$  clearly showed the incorporation of the deuterated methyl group into the molecule.

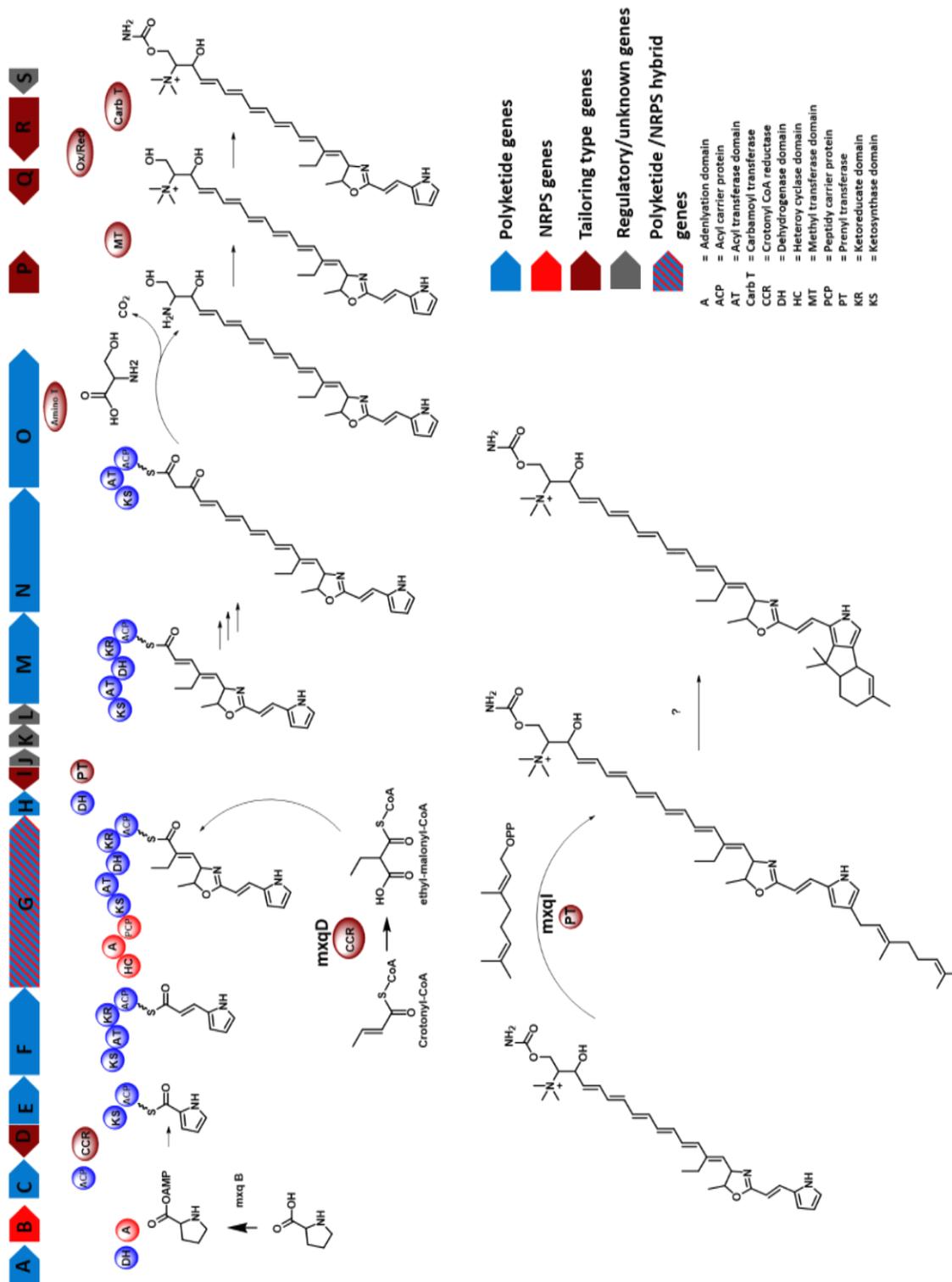


Figure 12: Scheme of the myxoquaterine biosynthesis showing the myxoquaterine biosynthetic gene cluster and catalytic domains conducting key reactions of the biosynthesis.

Myxoquaterine 669 (**3**) and 671 (**4**) contain an additional terpene moiety connected to the pyrrole ring. The terpene moiety is expected to be derived from a prenylation reaction with geranyl-pyrophosphate based on the characteristic geranyl-structure. Subsequently, a cyclization reaction turns **4** into **3**. Even though BLAST

search did not yield clear results for Mxql, the structure-based sequence comparison tool HHpred revealed similarities of Mxql with domains functioning as aromatic prenyltransferases.

*Table 4: List of genes in the myxoquaterin biosynthetic gene Cluster in MSr 11954*

Gene	Size <sup>a</sup>	Protein homologue <sup>b</sup>	Homologue in DKxanthen BGC	Proposed function	Putative catalytic domains
<i>mxqS</i>	382	KYG07499.1 (53/86)		Fis family transcriptional regulator	
<i>mxqR</i>	743	WP_075096512.1(53/69)		carbamoyltransferase/ carbamoyltransferase	aspartate
<i>mxqQ</i>	290	APR84655.1 (60/72)		Putative oxidoreductase/ short-chain dehydrogenase	Ox/Red
<i>mxqP</i>	293	PYP37896.1 (33/50)		methyltransferase domain-containing protein	MT
<i>mxqO</i>	1425	WP_073601403.1(43/56)		type I polyketide synthase/ PLP-dependent acyltransferase domain in the end	KS;AT;ACP;PLP binding domain;
<i>mxqN</i>	1897	WP_011554295.1(43/58)	dkxH/ dkxl	type I polyketide synthase	KS; AT; DH; KR; ACP
<i>mxqM</i>	1923	WP_015351541.1(41/53)	dkxH/ dkxl	type I polyketide synthase	KS; AT; DH; KR; ACP
<i>mxqL</i>	409	WP_105480911.1 (42/56)		HlyD family efflux transporter periplasmic adaptor subunit	
<i>mxqK</i>	254	WP_105480910.1(55/69)		ABC transporter ATP-binding protein/ lipoprotein-releasing system	
<i>mxqJ</i>	374	WP_106829348.1(44/64)		FtsX-like permease family protein/ transporter protein	
<i>mxqI</i>	459	ABF89221.1 (37/52)		FAD-binding protein/ geranylgeranyl reductase/ P450	prenyltransferase
<i>mxqH</i>	430	WP_121288086.1(45/60)		FAD-binding oxidoreductase	DH, Redox enzyme
<i>mxqG</i>	3049	WP_011554302.1(63/74)	dkxG	hybrid non-ribosomal peptide synthetase/ type I polyketide synthase	ACP; KR; DH; AT; KS; PCP; AT; HC
<i>mxqF</i>	1408	WP_095988363.1(56/66)	dkxF	type I polyketide synthase	ACP KR KS AT
<i>mxqE</i>	815	WP_013376470.1(60/70)	dkxE	polyketide synthase	KS ACP
<i>mxqD</i>	435	RLB43495.1 (61/73)		crotonyl-CoA carboxylase/ reductase/ PKS ER	CCR reductase
<i>mxqC</i>	84	WP_002619219.1(63/78)		acyl carrier protein	ACP
<i>mxqB</i>	508	WP_106092715.1(54/67)	dkxA	amino acid adenylation domain-containing protein/ prolyl-AMP ligase PltF	A
<i>mxqA</i>	381	WP_015351545.1(55/69)	dkxB	acyl-CoA dehydrogenase/ enoyl reductase	DH

<sup>a</sup> Sizes are in amino acids; <sup>b</sup> Accession numbers and percentage of identity/similarity are given in parentheses.

Inspection of metabolic profiles of closely related strains (figure 1) revealed that compounds of the myxoquaterine compound family can be found in the strain MSr 11367. A closer look into the production profile of MSr 11367 showed that the prenylated myxoquaterine derivatives **3** and **4** were not present in the crude extract (figure 13B). A BGC with high similarities to the myxoquaterine cluster in MSr 11954 was identified by antiSMASH analysis of the genome. A detailed comparison of both clusters showed a different organization of the ORFs (figure 13A). In accordance with the missing myxoquaterine derivatives, a homologue of the putative prenyltransferase gene *mxqI* was not present in MSr 11367. In combination, these results strongly supports the hypothesis that Mxql is responsible for transferring the geranyl unit onto the pyrrole moiety.

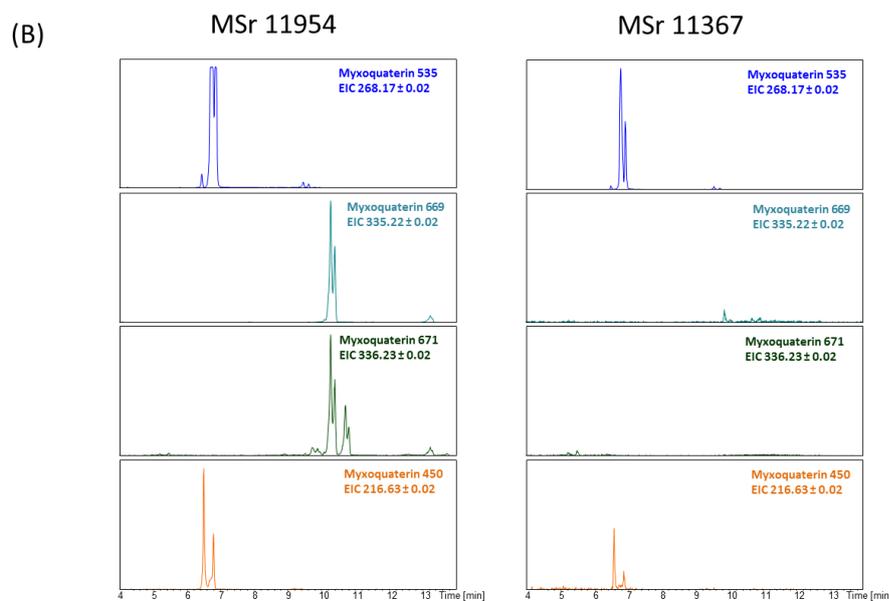
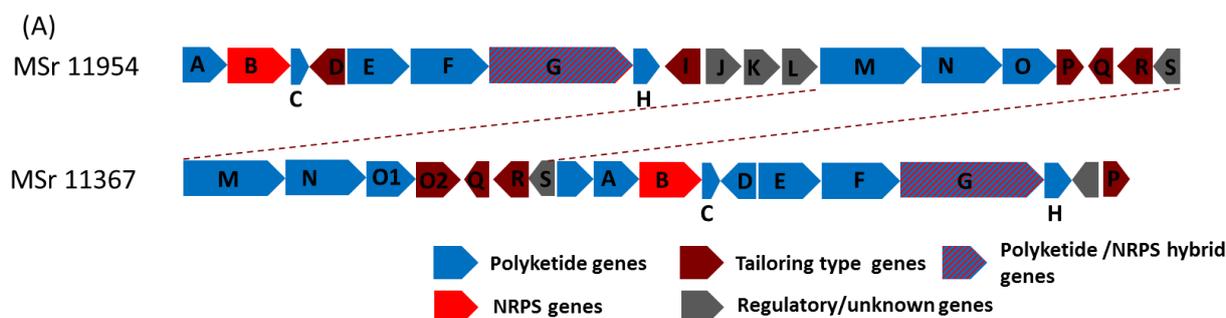


Figure 13: (A) Organization of the myxoquaterine BGC in MSr 11954 and the alternative myxoquaterine producer MSr 11367. (B) Myxoquaterine production profile in MSr 11954 and MSr 11367 showing the extracted ion chromatogram (EIC) of the double charged ions of Myxoquaterin 535, 669, 671 and 450.

In order to analyze if the prenylation reaction is a tailoring reaction, cultures of MSr 11954 were supplemented with structurally similar DKxanthen 534 and DKxanthen 561. A prenylation of these molecules would clearly indicate the prenylation to be a tailoring reaction. However, a prenylation was not observed since expected products could not be identified in the crude extracts by LC-*hr*MS measurements. A control experiment showed incorporation of the supplemented DKxanthenes into the bacterial cells.

A spontaneous cyclization of **4** into **3** was not observed. Therefore it was concluded that cyclization is conducted enzymatically as described in the biosynthesis of other terpenes<sup>[17]</sup>. In order to identify a gene that could conduct the cyclization, the BGC was screened for homologues of XiaF, a terpene cyclase in the biosynthesis xiamycin, a bacterial indolo terpene<sup>[18]</sup>. Since xiamycin is also a terpene-heteroaromate hybrid, a similar mechanism can also be expected for formation of **4**. However, a homologue of XiaF was not found. Also homologues of geosmin synthase and methylisborneol synthase genes, both well described terpene cyclases were not found in the myxoquaterine BGC<sup>[19]</sup>.

### 2.3.1 Implications from biosynthesis towards stereochemistry

Considering the absolute configuration of methyloxazoline, an *in silico* prediction can be made based on the underlying biosynthetic logic: Taking into account that the methyloxazoline ring is formed out of an acetate

unit and threonine catalyzed by a heterocyclase domain<sup>[20]</sup>. Due to a lack of an E-domain in this module the threonine can be expected to be in its L-form. In order to undergo cyclization the hydroxyl group must direct towards the carbonyl-group resulting in a 120° turn of the hydroxyl group. In the cyclized product the stereo center of the L-threonine is expected to not be changed. Therefore both stereocenters in the methyl-oxazoline group must be *R,R*-configured (figure 14).

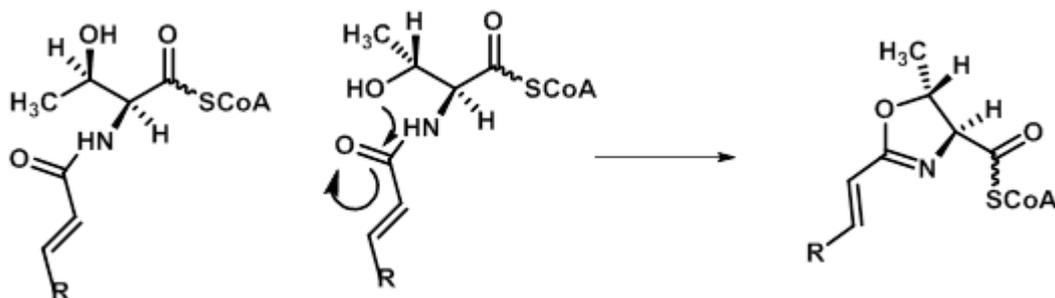


Figure 14: Heterocyclization reaction of the methyl-oxazoline moiety with stereochemistry of educt and product

An experimental determination is missing so far. A possible option to counter this would be usage of the natural chromophore systems within the molecule. Depending on the absolute configuration, different Cotton effects can be expected. Unfortunately, no comparable molecule with described absolute configuration was found by literature research that could be used for comparison of the CD spectra. An *in silico* prediction based on the BGC of the absolute configuration of the hydroxyl group or the neighboring quaternary ammonium ion/amine cannot be made since it is not exactly known which KR domain is reducing the ketone into an alcohol and whether the stereo center of the serine is preserved.

## 2.4 Bioactivity

All isolated myxoquaterine derivatives were tested against a broad array of fungal and bacterial organisms and showed weak activity against gram positive test strains (*S. aureus*, *B. subtilis* and *M. luteus*) except myxoquaterin 450 (2). Also against fungal test strains (*P. anomala*, *C. albicans* and *M. hiemalis*) weak bioactivity was found. No gram negative bioactivity was found except for myxoquaterine variant A against *E. coli*  $\Delta$  TolC. Interestingly, all tested myxoquaterine derivatives showed strong cytotoxic activity against KB3.1 and HCT-116 cell lines. A remarkable test result is the nearly in every case increased bioactivity of variant A compared to variant B.

Table 5: Biological activity of isolated myxoquaterin variants

Test organism	Myxoquaterin 535		Myxoquaterin 669		Myxoquaterin 671		Myxoquaterin 450
	Var. A	Var. B	Var. A	Var. B	Var. A	Var. B	
	[ $\mu\text{g/ml}$ ]						
<i>P. anomala</i>	>64	64	16	64	64	>64	64
<i>C. albicans</i>	>64	64	64	64	>64	>64	64
<i>M. hiemalis</i>	64	64	64	64	64	64	64

<i>S. aureus</i>	64	64	64	64	64	64	>64
<i>E. coli</i> $\Delta$ TolC	32	>64	> 64	> 64	>64	>64	>64
<i>E. coli</i> DSM 1116	>64	> 64	>64	>64	>64	>64	>64
<i>B. subtilis</i>	16	64	64	64	>64	>64	>64
<i>M. luteus</i>	16	>64	64	64	64	>64	>64
IC <sub>50</sub> HCT-116	0.14	2.96	0.00093	0.144	0.65	2.28	0.01
IC <sub>50</sub> KB 3.1	0,377	0,038	0,078	0,164	0,67	2,36	0.009

Regarding the functional groups of the myxoquaterine molecules, it seemed likely that the cytotoxic activity is mediated by the quaternary ammonium ion and carbamoyl group<sup>[21]</sup>. However, bioactivity of myxoquaterine 450 (**2**), which is missing these functional groups, gives rise to the assumption that the cytotoxicity is elicited from other structural features.

## 2.5 Conclusions

Exploration of a novel genus within the clade of *Sorangineae* resulted in the discovery of the myxoquaterine compound family. Organic extracts of the strain MSr 11954 showed antifungal bioactivity, and were analyzed by a standard fractionation and a group of molecules could be correlated to the observed anti-fungal activity of the obtained fractions. The molecules were isolated from a large scale fermentation and their structures were elucidated using 2D NMR experiments. Bigger parts of the core structure resemble the DKxanthen compound family. A characteristic structure difference of the molecules to the DKxanthenes was a trimethylammonium ion. In addition, two of the isolated myxoquaterine molecules were found prenylated. An experimental assignment of the absolute configuration was not successful, but the relative configuration of neighboring stereocenters was determined.

The biosynthesis was examined by feeding experiments and *in silico* analysis of the BGC. These studies revealed that bigger parts of the biosynthesis happen in a similar manner to the DKxanthen biosynthesis. But in contrast to the DKxanthen biosynthesis, release occurs in a PLP dependent fashion and tailoring reactions turn the molecule in a different molecule than the DKxanthenes. The myxoquaterines showed a promising cytotoxic activity against tested cell lines and moderate antimicrobial activity against the used test strains.

## 2.6 References

- [1] J. Bérdy, *J. Antibiot. (Tokyo)* **2012**, *65*, 441.
- [2] J. Herrmann, A. A. Fayad, R. Müller, *Nat. Prod. Rep.* **2017**, *34*, 135.
- [3] T. Hoffmann, D. Krug, N. Bozkurt, S. Duddela, R. Jansen, R. Garcia, K. Gerth, H. Steinmetz, R. Müller, *Nat. Commun.* **2018**, *9*, 803.
- [4] "Dictionary of Natural Products", can be found under <http://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml>.
- [5] H. Song, W. Hu, N. Naowarajna, A. S. Her, S. Wang, R. Desai, L. Qin, X. Chen, P. Liu, *Sci. Rep.* **2015**, *5*, 1.
- [6] P. Meiser, H. B. Bode, R. Müller, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19128.
- [7] A. G. Pacheco, P. Machado de Oliveira, D. Piló-Veloso, A. Flávio de Carvalho Alcântara, *Molecules (Basel, Switzerland)* **2009**, *14*, 1245.
- [8] N. Matsumori, D. Kaneno, M. Murata, H. Nakamura, K. Tachibana, *J. Org. Chem.* **1999**, *64*, 866.
- [9] G. Bifulco, P. Dambruoso, L. Gomez-Paloma, R. Riccio, *Chem. Rev.* **2007**, *107*, 3744.
- [10] S. D. Jackman, B. P. Vandervalk, H. Mohamadi, J. Chu, S. Yeo, S. A. Hammond, G. Jahesh, H. Khan, L. Coombe, R. L. Warren et al., *Genome research* **2017**, *27*, 768.

- [11] K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema, T. Weber, *Nucleic Acids Res.* **2019**, W81-W87.
- [12] S. El-Gebali, J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J. Richardson, G. A. Salazar, A. Smart et al., *Nucleic acids research* **2019**, *47*, D427-D432.
- [13] L. Zimmermann, A. Stephens, S.-Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A. N. Lupas, V. Alva, *Journal of molecular biology* **2018**, *430*, 2237.
- [14] P. Meiser, K. J. Weissman, H. B. Bode, D. Krug, J. S. Dickschat, A. Sandmann, R. Müller, *Chem. Biol.* **2008**, *15*, 771.
- [15] L. Du, L. Lou, *Nat. Prod. Rep.* **2010**, *27*, 255.
- [16] Y.-L. Du, K. S. Ryan, *Nat. Prod. Rep.* **2019**, *36*, 430.
- [17] J. S. Dickschat, *Nat. Prod. Rep.* **2015**, *33*, 87.
- [18] S. Kugel, M. Baunach, P. Baer, M. Ishida-Ito, S. Sundaram, Z. Xu, M. Groll, C. Hertweck, *Nat Commun*, **8**, 1.
- [19] a) J. Jiang, X. He, D. E. Cane, *J. Am. Chem. Soc.* **2006**, *128*, 8128; b) M. Komatsu, M. Tsuda, S. Omura, H. Oikawa, H. Ikeda, *PNAS* **2008**, *105*, 7422.
- [20] A. W. Truman, *Beilstein J. Org. Chem.* **2016**, *12*, 1250.
- [21] a) H. Nagamune, T. Maeda, K. Ohkura, K. Yamamoto, M. Nakajima, H. Kourai, *Toxicology in Vitro* **2000**, *14*, 139; b) A. K. Ghosh, M. Brindisi, *Journal of medicinal chemistry* **2015**, *58*, 2895.

## 2.8 Supplementary information

### 2.8.1 Experimental procedures

#### Strains

strains	suborder	family	genus	source
MSr 11954	<i>Sorangineae</i>	Novel/unclassified	Novel/unclassified	isolated by Dr. Ronald Garcia
MSr 11367	<i>Sorangineae</i>	Novel/unclassified	Novel/unclassified	isolated by Dr. Ronald Garcia

#### Media

##### CyHv4 media

Soytone	0.2% (m/v)
Casitone	0.3 % (m/v)
Glucose	0.2 % (m/v)
Soluble starch	0.5 % (m/v)
Maltose monohydrate	0.2 % (m/v)
Yeast extract	0.15 % (m/v)
CaCl <sub>2</sub> · 2H <sub>2</sub> O	0.1 % (m/v)
MgSO <sub>4</sub> · 7H <sub>2</sub> O	0.1 % (m/v)
HEPES	25 mM (m/v)

The ingredients were dissolved in MiliQ H<sub>2</sub>O. The pH was adjusted to 7.2 with 10 N KOH.

##### CTT media

Casiton	1 % (m/v)
TRIS Buffer pH 8 100 mM	10 % (v/v)
KH <sub>2</sub> PO <sub>4</sub> Buffer pH 7.6 100 mM	1 % (v/v)
MgSO <sub>4</sub> · 7H <sub>2</sub> O	0.2 % (v/v)

The pH was adjusted to 7.6 with 10 N KOH.

### General analytic procedures

#### Standard LC-MS analysis

##### amazon

The amaZon is an Ion trap MS/MS mass spectrometer manufactured by Bruker Daltonics coupled with a Dionex HPLC from Thermo Scientific. This machine was used for standard analytics using following template gradient programs:

##### 6 min gradient:

The following reversed phase HPLC methods were used for the mass spectrometry coupled analysis: Column: ACQUITY BEH 50 x 2.1 mm, 1.7 μm, 130 Å, a flow rate of 0.6 ml/ min and a column temperature of 45 °C. Following gradient was applied with H<sub>2</sub>O + 0.1 % FA as eluent A and ACN + 0.1 % FA as eluent B. 0-1 min: 5 % Eluent B; 1-7 min: linear increase of eluent B from 5 to 95 %; 7-8.5 min: 95 % eluent B; 8.5-9 min: linear decrease of eluent B from 95 to 5 % B; 9-11.5 min: re-equilibration with 5 % eluent B. The detection was usually performed in the positive ESI MS mode, using a scan range from 200-2000 m/z.

#### 9 min gradient:

The following reversed phase HPLC methods were used for the mass spectrometry coupled analysis: Column: ACQUITY BEH 50 x 2.1 mm, 1.7  $\mu\text{m}$ , 130  $\text{\AA}$ , a flow rate of 0.6 ml/ min and a column temperature of 45 °C. Following gradient was applied with H<sub>2</sub>O + 0.1 % FA as eluent A and ACN + 0.1 % FA as eluent B. 0-0.5 min: 5 % Eluent B; 0.5-9.5 min: linear increase of eluent B to 95 %; 9.5-10.5 min: 95 % eluent B; 10.5-11 min: linear decrease of eluent B; 11-12.5 min: re-equilibration with 5 % eluent B. The detection was usually performed in the positive ESI MS mode, using a scan range from 200-2000 m/z.

#### 18 min gradient:

The following reversed phase HPLC methods were used for the mass spectrometry coupled analysis: Column: ACQUITY BEH 50 x 2.1 mm, 1.7  $\mu\text{m}$ , 130  $\text{\AA}$ , a flow rate of 0.6 ml/ min and a column temperature of 45 °C. Following gradient was applied with H<sub>2</sub>O + 0.1 % FA as eluent A and ACN + 0.1 % FA as eluent B. 0-0.5 min: 5 % Eluent B; 0.5-18.5 min: linear increase of eluent B to 95 %; 18.5-20.5 min: 95 % eluent B; 20.5-20.8 min: linear decrease of eluent B; 20.8-22.5 min: re-equilibration with 5 % eluent B. The detection was usually performed in the positive ESI MS mode, using a scan range from 200-2000 m/z.

#### maXis 4G

The maXis 4G is a high resolution TOF mass spectrometer manufactured by Bruker Daltonics coupled with a Dionex HPLC from Thermo Scientific. This machine was used for high resolution LC-MS measurements. Following template method was used:

Column: ACQUITY BEH 100 x 2.1 mm 1.7  $\mu\text{m}$  130  $\text{\AA}$ , a flow rate of 0.6 ml/ min and a column temperature of 45 °C. Following gradient was applied with H<sub>2</sub>O + 0.1 % FA as eluent A and ACN + 0.1 % FA as eluent B. 0-0.5 min: 5 % Eluent B; 0.5-18.5 min: linear increase of eluent B to 95 %; 18.5-20.5 min: 95 % eluent B; 20.5-20.8 min: linear decrease of eluent B; 20.8-22.5 min: re-equilibration with 5 % eluent B. The detection was usually performed in the positive MS mode or MS<sup>2</sup> mode, using a scan range from 150-2500 m/z.

#### Spectroscopic methods

NMR were acquired with a Bruker Ascend 700 NMR spectrometer equipped with a 5mm TCI cryoprobe and a Bruker UltraShield 500 NMR spectrometer equipped with a 5mm TCI cryoprobe. Optical rotation was measured using a Jasco polarimeter. CD spectra were acquired using a Jasco J-1500 CD spectrometer.

#### Cultivation of MSr 11954

Dr. Ronald Garcia performed small-scale cultivation (100 ml) for 12 days in at 30°C.

Dr. Ronald Garcia using the same conditions performed large-scale cultivation (6 x 2 L) in CyHv4 medium for 14 days.

#### Cultivation of MSr 11367

Dr. Ronald Garcia performed small-scale cultivation (100 ml) at 30 °C for 12 days .

#### Cultivation of DK1622

For cultivation of DK1622, a pre-culture was inoculated from an Agar plate into 100 ml CTT media. The pre-culture was used for inoculation of 5 x 1.5 l cultures. The cultures were cultivated at 30 °C in 1.5 l CTT media

in 5 l flasks and inoculated with 1.0 % from the pre-culture. 2 days after inoculation, 2 % XAD-16 suspension was added to the culture. 8 days after inoculation, the cultures were harvested.

### **Standard fractionation**

Standard fractionations of crude extracts were performed by Daniel Sauer. Following analytical setup was used for the standard fraction: A maXis 2G, which is a high resolution TOF mass spectrometer manufactured by Bruker Daltonics coupled with a Dionex HPLC from Thermo Scientific and a Advion Triversa Nanomate nano-ESI system for fraction collection. Following HPLC method was used: Column: ACQUITY BEH 100 x 2.1 mm 1.7  $\mu\text{m}$  130  $\text{\AA}$ , a flow rate of 0.6 ml/ min and a column temperature of 45 °C. Following gradient was applied with H<sub>2</sub>O + 10 mM AmFo as eluent A and ACN + 10 mM AmFo as eluent B. 0-0.5 min: 5 % Eluent B; 0.5-18.5 min: linear increase of eluent B to 95 %; 18.5-20.5 min: 95 % eluent B; 20.5-20.8 min: linear decrease of eluent B; 20.8-22.5 min: re-equilibration with 5 % eluent B. The detection was performed in MS<sup>2</sup> mode, using a scan range from 145-2000 m/z. For fractionation, 5  $\mu\text{l}$  of a crude extract were injected into the HPLC. Fractions were collected from 0.62-20.63 min in a 96 well plate. Each fraction was collected for 12.6 s. The collected fractions were tested for their bioactivity.

### **Genome sequencing**

Genomic DNA was isolated by a standard chloroform/phenol extraction protocol and was performed by Dr. Ronald Garcia and Irene Cochems. The genomic DNA was submitted to pacBio Inc. and Illumina for sequencing. The raw sequencing data from Illumina sequencing were assembled from by Nestor Zaburanyi.

### **Biological testing**

The testing of all crude extracts, Sephadex fractions, HPLC fractions and pure compounds were performed by Viktoria Schmitt, Irene Cochems and Stefanie Schmidt according to existing protocols.<sup>[1]</sup>

### **Preparation of crude extract from MSr 11954 small scale cultivations**

XAD-16 and cells were frozen at -20 °C and lyophilized. After lyophilization, metabolites were extracted with 50 ml MeOH/acetone 1:1 under constant shaking for 2.5 h. The supernatant was filtrated into a round bottom flask and the solvent was evaporated under reduced pressure. The residue was dissolved in MeOH, transferred into a 4 ml Vial and stored at 20 °C.

Myxoquaterines were purified from a 24 l batch of MSr 11954 cultivated in CyHv4 media containing XAD-16 adsorber resin. The cells and the XAD resin were separated from the media by centrifugation and lyophilized. The dried cell mass and resin was extracted three times with MeOH/acetone (1:1). The solvents were poured through a funnel with glass wool and the solvent was evaporated. The dried extract was re-dissolved in n-hexane and MeOH + 10% H<sub>2</sub>O and transferred into a separatory funnel. The layers were separated and the liquid/liquid extraction was repeated with n-Hexane. The MeOH was dried and re-dissolved in H<sub>2</sub>O and DCM. The layers were separated in a separatory funnel and the aqueous layer was re-extracted with DCM. The precipitate between the layers was collected and dissolved in acetone and MeOH. Myxoquaterines were contained in the DCM layer and in the dissolved precipitate. The DCM layer and the dissolved precipitate were dried *in vacuo*.

The DCM layer was fractionated using C-18 reverse phase silica. Following system was used: Biotage with a 25 g Snapfit column, Versa Flash Spherical C18 silica 45-75  $\mu\text{m}$  70 A, Solvent System: A= H<sub>2</sub>O + 10mM AmFo ; B= MeOH + 10mM AmFo; A flow rate of 20 ml/min was used and 45 ml were collected per tube. The column was equilibrated with with 3 column volumes (CV) 70:30 Solvent A/B. The sample was loaded on the column using vacuum dried iSolute beads. Following gradient was used: 2 CV 70:30 Solvent

A/B; 35 CV linear increases from 30-100 % Solvent B; 5 CV 100 % Solvent B. The column was afterwards flushed with 200 ml Acetone.

An aliquot of every fraction (including the waste from the flushes) was taken and measured by LC-MS to screen for the Myxoquaterines. Fractions containing the target masses were pooled and dried *in vacuo*.

For further purification of Myxoquaterine 450, 535a and 535b HPLC with following setup was used:

Dionex Ultimate 3000; Column: XSelect CSH 130 (C18) 250 x 10 mm, 5  $\mu$ m; flowrate 5 ml/min and column temperature 40 °C with H<sub>2</sub>O + 0.1 % FA as eluent A and ACN + 0.1% FA as eluent B. The following gradient was applied: 0-2 min 5 % eluent B; 2-4 min linear increase of eluent B to 20 %; 4-20 min linear increase of eluent B to 24 %; 20-23 min linear increase of eluent B to 95 %; 23-24 min 95 % eluent B; 24-25 min linear decrease of eluent B to 5 %; 25-28 min re-equilibration with 5 % eluent B. Myxoquaterine 450 was collected from 15.2-15.7 min. Myxoquaterine 535 variant A was collected from 17.7-18.6 min; Myxoquaterine 535 variant B was collected from 18.8-19.4 min. The collected fractions were dried and dissolved in 4 ml MeOH.

Myxoquaterine 450 still contained impurities as controlled by LC-MS and NMR measurements. In order to obtain clean myxoquaterine 450, the collected fraction was further purified using following HPLC setup:

Dionex Ultimate 3000; Column: Luna C-18 (2) 250 x 4.6 mm, 5  $\mu$ m; flowrate 1.5 ml/min and column temperature 40 °C with H<sub>2</sub>O + 0.1 % FA as eluent A and ACN + 0.1% FA as eluent B. The following gradient was applied: 0-2 min 5 % eluent B; 2-22 min linear increase of eluent B to 95 %; 22-24 min 95 % eluent B; 24-25 min linear decrease of eluent B to 5 %; 25-30 min re-equilibration with 5 % eluent. Myxoquaterine 450 was collected from 10.4-10.75 min. The collected fraction was dried and dissolved in 4 ml MeOH.

Fractions containing myxoquaterine 669 and 671 were combined, dried and dissolved in MeOH. For further purification HPLC with following setup was used:

Dionex Ultimate 3000; Column: XSelect CSH 130 (C18) 250 x 10 mm, 5  $\mu$ m; flowrate 5 ml/min and column temperature 40 °C with H<sub>2</sub>O + 0.1 % FA as eluent A and ACN + 0.1% FA as eluent B.; 0-2 min linear increase of eluent B from 5 % to 33 %; 2-22min linear increase of eluent B to 45 %; 22-23 min linear increase of eluent B to 95 %; 23-26 min 95 % eluent B; 26-27 min linear decrease of eluent B to 5 %; 27-31 min re-equilibration with 5 % eluent B. Myxoquartarine 669a was collected from 16.7-17.3 min, Myxoquartarine 669 b was collected from 17.3-18.0 min, Myxoquartarine 671a was collected from 20.2-20.75 min and Myxoquartarine 671b was collected from 20.9-21.4 min. The collected fractions were dried and dissolved in 4 ml MeOH.

### **Isolation of DKxanthen 534 and 560**

DKxanthen 534 and DKxanthen 560 were purified from a 9l batch of DK1622 cultivated in CTT media containing XAD-16 adsorber resin. The cells and the XAD resin were separated from the media by centrifugation and lyophilized. The dried cell mass and resin was extracted three times with MeOH/acetone (1:1). The solvents were poured through a funnel with glass wool and the solvent was evaporated. The dried extract was re-dissolved in n-hexane and MeOH + 10% H<sub>2</sub>O and transferred into a separatory funnel. The layers were separated and the liquid/liquid extraction was repeated n-hexane. The MeOH was dried and re-dissolved in H<sub>2</sub>O and ethylacetate and acidified with 1 ml 1N HCl. The layers were separated in a separatory funnel and the aqueous layer was re-extracted with ethyl acetate. The ethyl acetate layer was dried *in vacuo*.

The DKxanthenes were mostly present in the ethyl acetate layer. For further purification the ethyl acetate layer was fractionated using flash chromatography. Following system was used:

Biotage with a 25 g Snapfit column, Versa Flash Spherical C18 silica 45-75  $\mu$ m 70 A, Solvent System: A= H<sub>2</sub>O + 0.1 % FA; B= MeOH + 0.1 % FA; A flow rate of 20 ml/min was used and 45 ml were collected per tube. The column was equilibrated with 3 column volumes (CV) 95:5 Solvent A/B. The sample was loaded on the column using vacuum dried iSolute beads. Following gradient was used: 2 CV 95:5 Solvent A/B; 30 CV linear increases from 5-100 % Solvent B; 5 CV 100 % Solvent B.

An aliquot of every fraction (including the waste from the flushes) was taken and measured by LC-MS to screen for DKxanthenes. Fractions containing DKxanthenes were pooled and dried *in vacuo*.

For further purification HPLC with following setup was used:

Dionex Ultimate 3000; Column: Luna C18 (2) 250 x 10 mm, 5 µm; flowrate 5 ml/min and column temperature 40 °C with H<sub>2</sub>O + 0.1 % FA as eluent A and ACN + 0.1% FA as eluent B. The following gradient was applied: 0-2 min 5 % eluent B; 2-4 min linear increase of eluent B to 25 %; 4-24 min linear increase of eluent B to 40 %; 24-25 min linear increase of eluent B to 95 %; 25-28 min 95 % eluent B; 28-29 min linear decrease of eluent B to 5 %; 29-32 min re-equilibration with 5 % eluent B. DKxanthen 534 was collected from 17.8-18.4 min and DKxanthen 560 was collected from 21.8-22.25 min. The collected fractions were dried and dissolved in 4 ml MeOH.

### **Mosher Ester formation of myxoquaterine 535 variant A**

For Mosher esterification different reaction conditions were carried out. Every reaction was performed in dried solvents. Myxoquaterine 535 variant A was lyophilized to remove remaining humidity.

Reaction	solvent	reagents
1	DCM	DMAP (4EQ), DIC (4EQ), (S)-MTPA (4EQ), 0.37 µM Myxoquartarin 535 variant A
2	DCM	DMAP (52EQ), (R)-MTPA-Cl (16 EQ), 0.5 µM Myxoquartarin 535 variant A
3	Pyridine d <sub>5</sub>	(S)-MTPA-Cl (50 EQ), 0.46 µM Myxoquartarin 535 variant A (R)-MTPA-Cl (50 EQ), 0.46 µM Myxoquartarin 535 variant A

Reaction progress was controlled using over a period of up to 7 days by LC-MS.

### **Feeding experiments**

The feeding study was performed with MSr 11954 in CyHv4 media in 10 ml cultures in 50 ml flasks. The cultures were inoculated with 10 % from a preculture of MSr 11954 in CyHv4 media.

Isotope labelled substrates	concentration	Solvent for application
L-methionine-Me d <sub>3</sub>	0.25 M	MeOH/dH <sub>2</sub> O 1:1
L-serine d <sub>3</sub>	0.25 M	
L-serine 1- <sup>13</sup> C	0.25 M	
mevalonolactone 1,2- <sup>13</sup> C <sub>2</sub>	0.25 M	
L-cysteine <sup>13</sup> C <sub>3</sub> <sup>15</sup> N	0.1 M	

To test substrate incorporation 10 µl of the labelled substrate-solution was added to the cultures two times over two days. 24 h after last addition of substrates 500 µl of XAD-16 suspension were added. 9 days after inoculation the cultures were harvested by centrifugation. The supernatant was discarded and the cells and the XAD-16 were frozen and lyophilized before extraction as previously described. The dried extract was dissolved in 1 ml MeOH.

### **DKxanthen feeding**

The Feeding study was performed with MSr 11954 in CyHv4 media in 10 ml cultures in 50 ml flasks. The cultures were inoculated with 10 % from a preculture of MSr 11954 in CyHv4 media. 100 µl of a solution of DKxanthen 534 and DKxanthen 560 (c= 1mg/ml) were added 5 times to the cultures over a time period of 3 days. After 4 days XAD-16 was added to the cultures except one culture. The culture without XAD was cultivated for 13 days, while cultures with XAD were harvested by centrifugation after 16 days. The supernatant

was discarded and the cells and the XAD were frozen and lyophilized before extraction as previously described. The dried extract was dissolved in 1 ml MeOH.

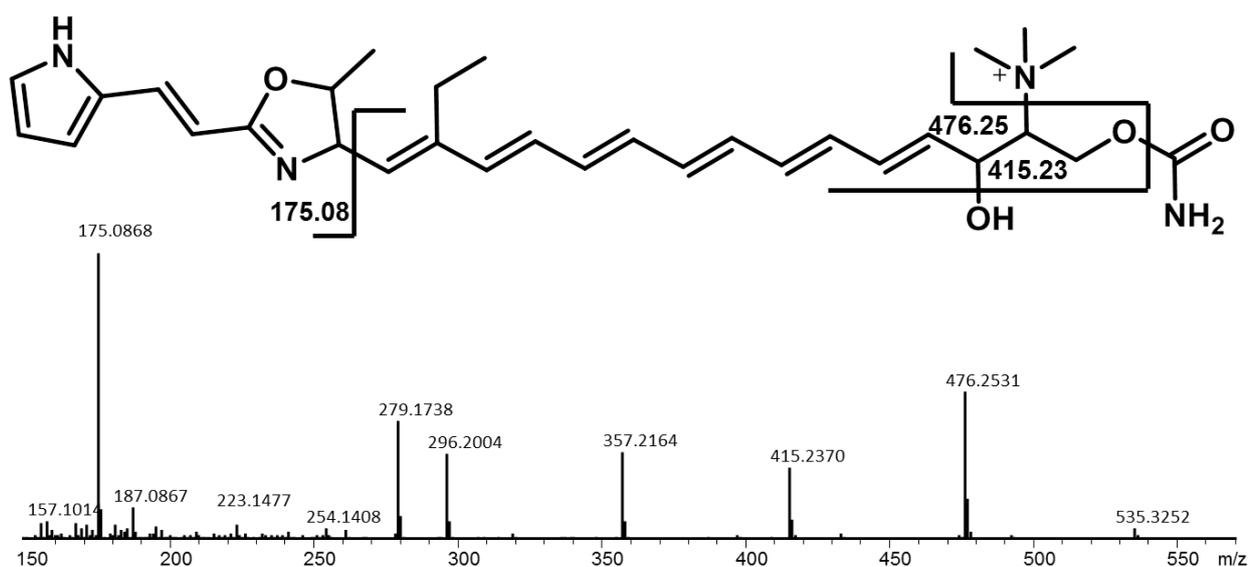
The culture without XAD was centrifuged to separate cells and media. The cells were washed with CyHv4 media and centrifuged again. The procedure was repeated again. The cells were freeze dried and extracted like described previously. The dried extract was dissolved in 1 ml MeOH.

The media was extracted by liquid/liquid partitioning using ethyl-acetate. The ethyl-acetate was dried *in vacuo* and dissolved in 1 ml MeOH. The supernatant was discarded and the cells and the XAD were frozen and lyophilized before extraction as previously described. The dried extract was dissolved in 1 ml MeOH.

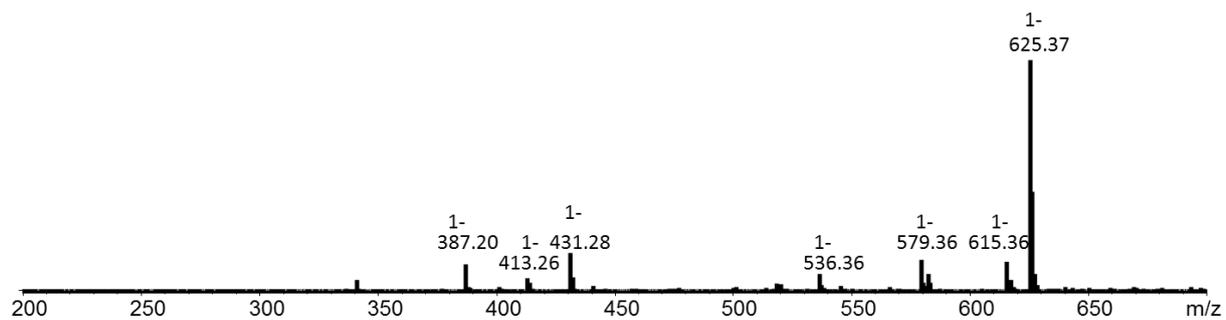
## 2.8.2 Supplementary data/additional experimental results

### Analytical data myxoquaterine 535

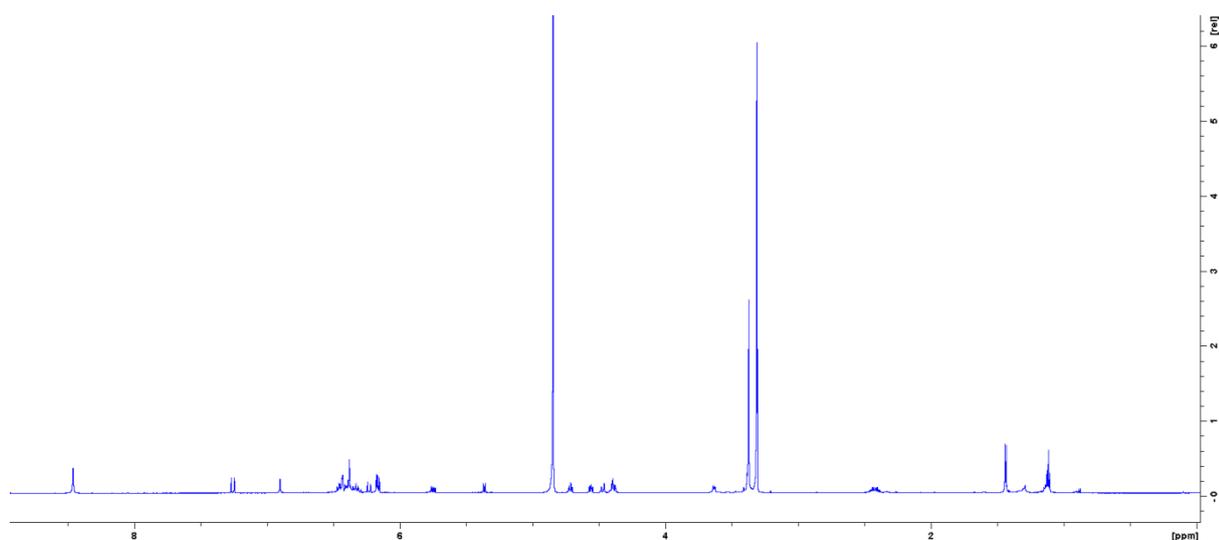
Myxoquaterine 535 was isolated as brown/yellowish amorphous solid that turned into a powder after lyophilization. *hrMass*: 535.32886 [M]<sup>+</sup>; 268.16919 [M+H]<sup>2+</sup>; Specific optical rotation variant A:  $[\alpha]_D^{20}$  45.3° g<sup>-1</sup> ml dm<sup>-1</sup>; specific optical rotation variant B:  $[\alpha]_D^{20}$  102.82° g<sup>-1</sup> ml dm<sup>-1</sup>; UV<sub>max</sub> Variant A: 190, 331, 346, 358, 385 nm; UV<sub>max</sub> Variant B: 190, 332, 346, 359, 383 nm (UV data obtained by a Jasco J-1500 CD spectrometer in MeOH)



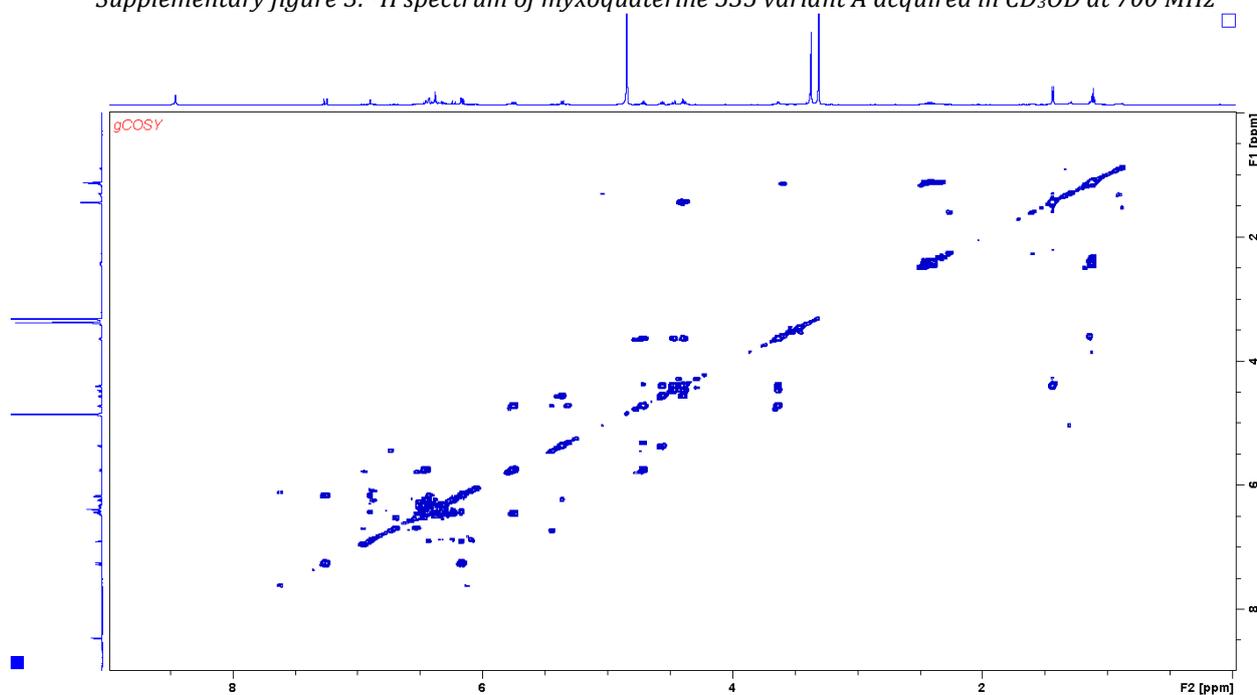
Supplementary figure 1: *hrMS*<sup>2</sup> spectrum and observed fragments of myxoquaterine 535. *MS*<sup>2</sup> spectrum is extracted from LC-*hrMS* measurement of an extract of MSr 11954.



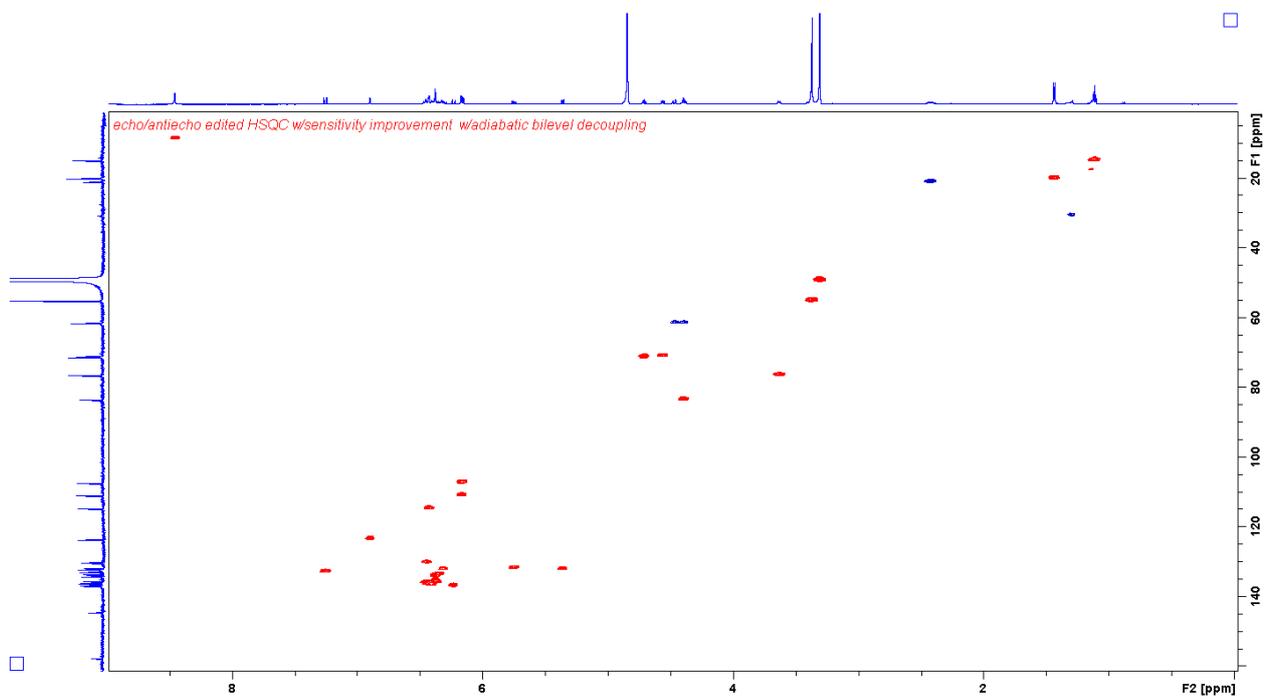
Supplementary figure 2: MS spectrum of myxoquaterine 535 obtained in MS negative mode. MS spectrum is obtained from LC-MS measurement of isolated myxoquaterin 535.



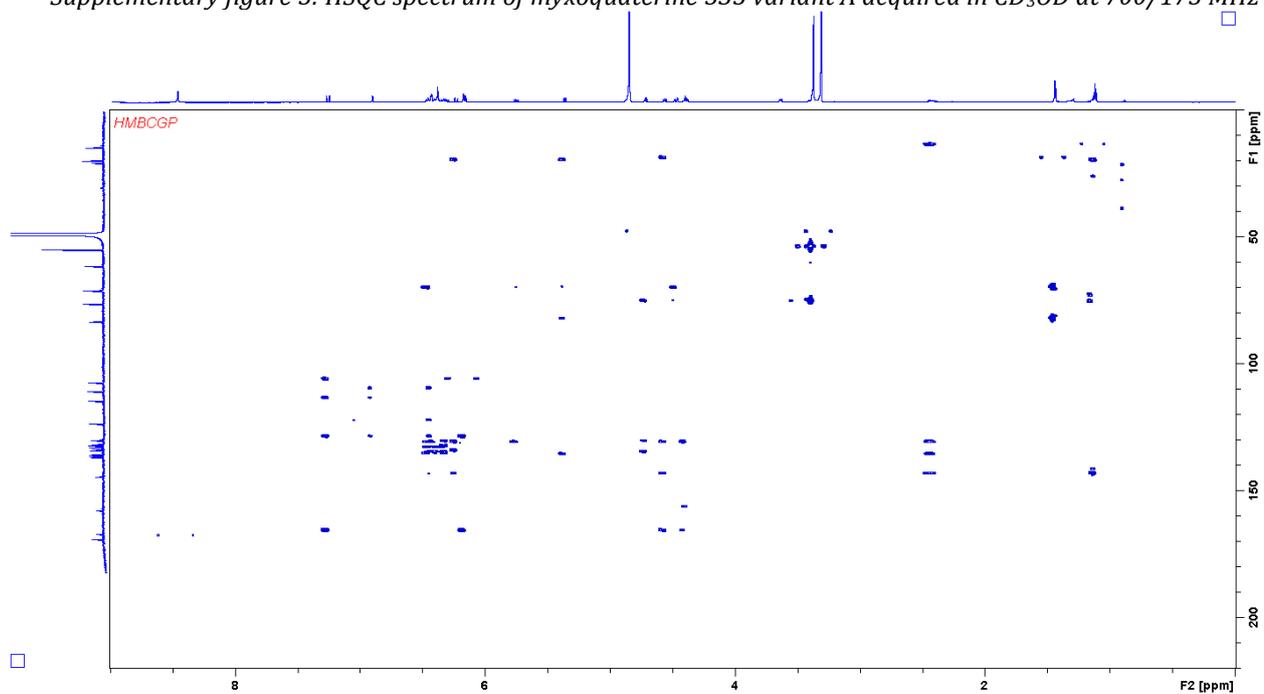
Supplementary figure 3:  $^1\text{H}$  spectrum of myxoquaterine 535 variant A acquired in  $\text{CD}_3\text{OD}$  at 700 MHz



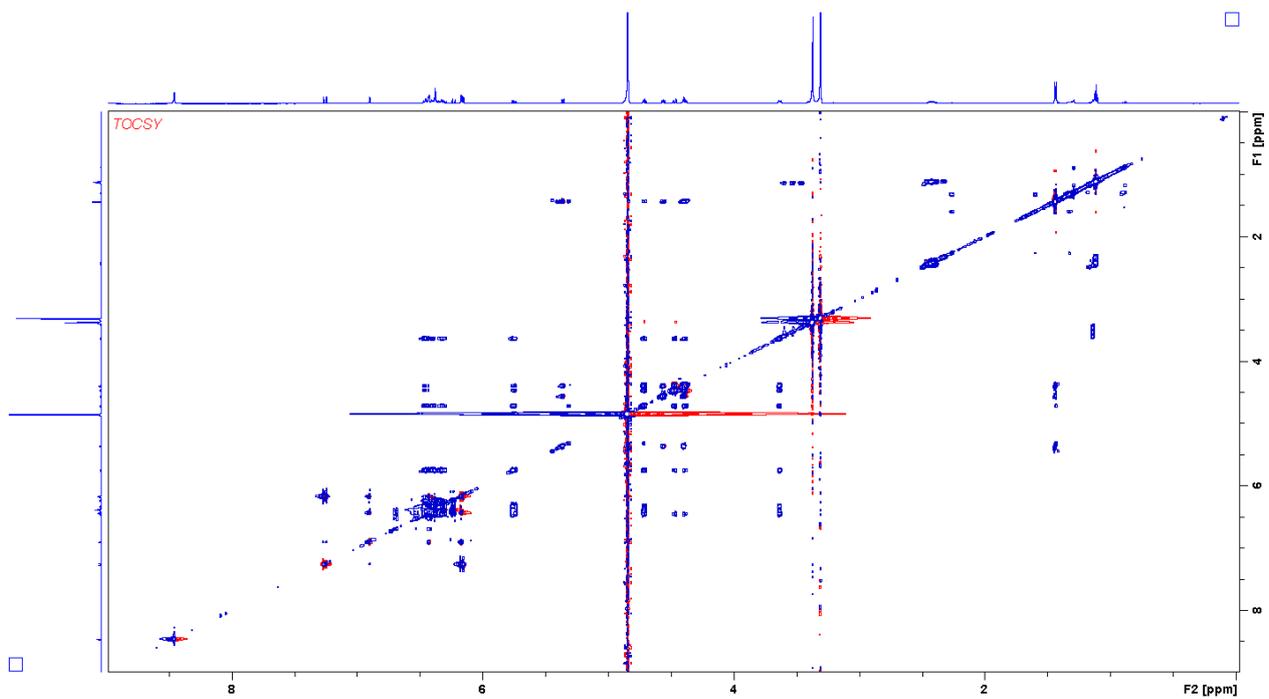
Supplementary figure 4: COSY spectrum of myxoquaterine 535 variant A acquired in  $\text{CD}_3\text{OD}$  at 700 MHz



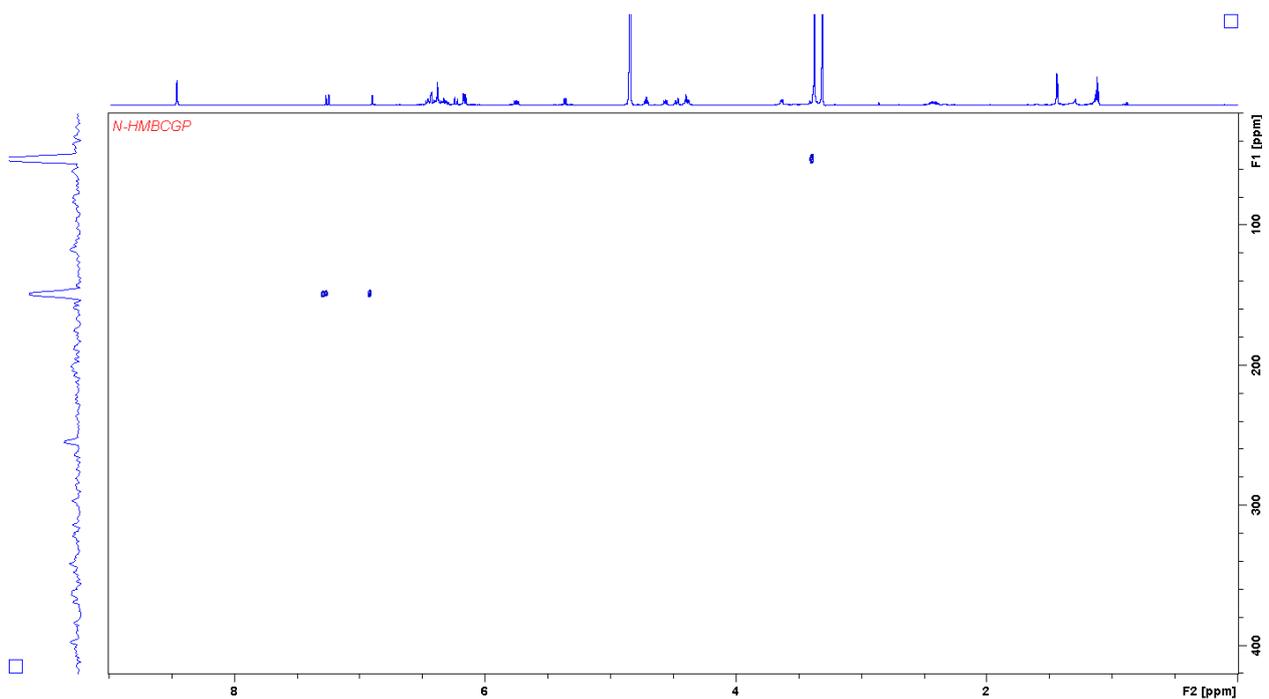
Supplementary figure 5: HSQC spectrum of myxoquaterine 535 variant A acquired in CD<sub>3</sub>OD at 700/175 MHz



Supplementary figure 6: HMBC spectrum of myxoquaterine 535 variant A acquired in CD<sub>3</sub>OD at 700 MHz

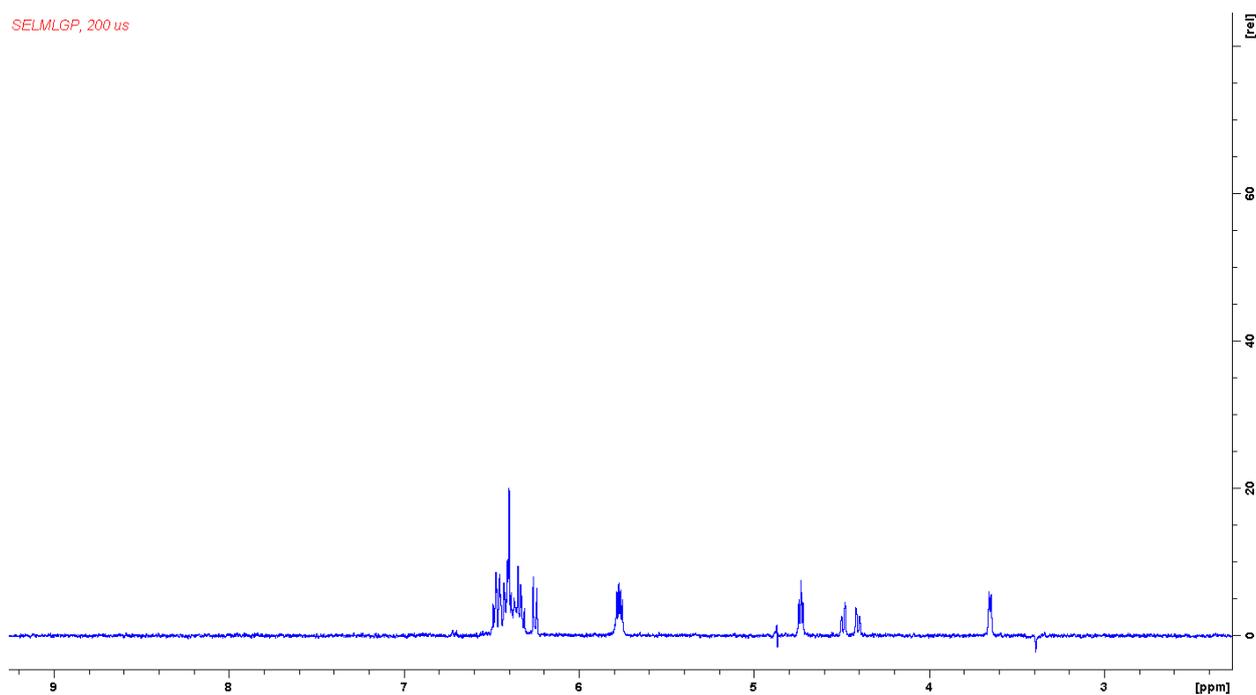


Supplementary figure 7: TOCSY spectrum of myxoquaterine 535 variant A acquired in  $CD_3OD$  at 700/175 MHz

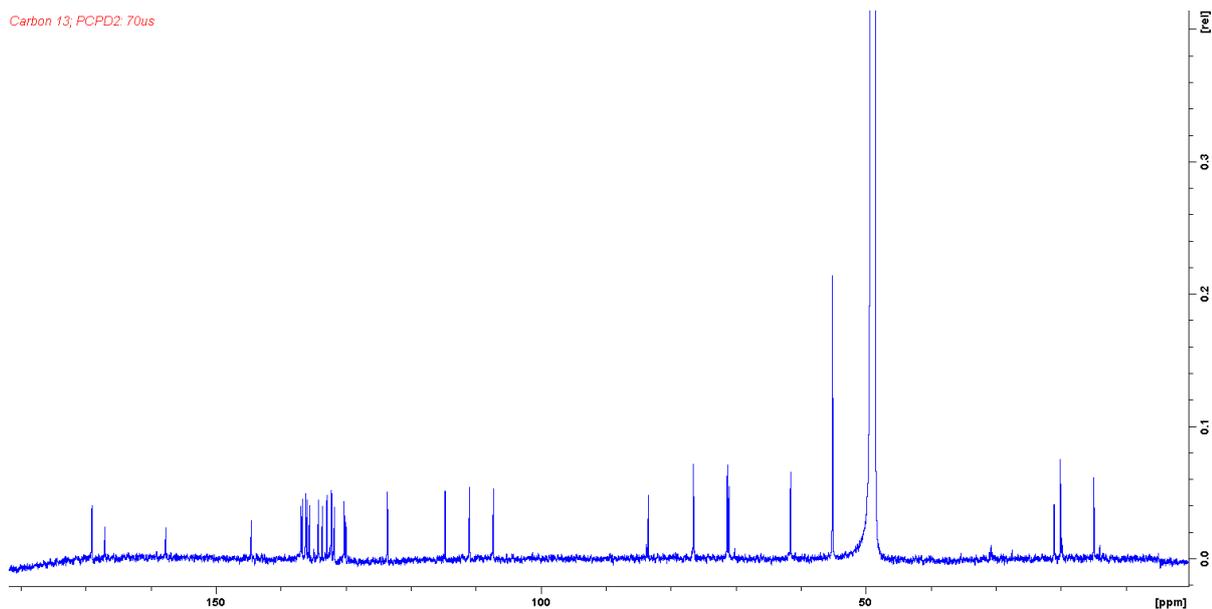


Supplementary figure 8:  $^{15}N$ -HMBC spectrum of myxoquaterine 535 variant A acquired in  $CD_3OD$  at 700/71 MHz

SELMLGP, 200 us



Supplementary figure 9: selective 1D TOCSY with excitation at 5.75 ppm and 200 ms mixing time of myxoquaterine 535 variant A acquired in CD<sub>3</sub>OD at 700 MHz



Supplementary figure 10: <sup>13</sup>C spectrum of myxoquaterine 535 variant A acquired in CD<sub>3</sub>OD at 175 MHz



Supplementary figure 11: Structure of myxoquaterine 535

Supplementary table 1: Spectroscopic data of myxoquaterine 535 variant A in CD<sub>3</sub>OD.

position	$\delta_C^a$	$\delta_H^b$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
1	<sup>15</sup> N 149.3 <sup>e</sup>			
2	123.2	6.90 d,d (1.3, 2.6)	3,4	3,4,5

3	110.5	6.17 d,d (2.7,3.4)	2,4	2,4,5
4	114.4	6.43 m	3,2	2,3,5
5	129.7			
6	132.7	7.26 d (16.0)	7	4,5,7,8
7	106.8	6.16 d (16.2)	6	6,8
8	166.8			
9				
10	83.2	4.39 m	11,31	8,11,13
11	71.0	4.56 d,d (9.7,7.3)	10,13	7,8,10,13,14,31
12	N			
13	131.9	5.36 d (9.5)	11	10,11,15,32,33
14	144.2			
15	136.6	6.23 d (15.3)	16	13,14,16
16	135.4	6.43 m	15	
17-21	132-136	6.30-6.50		
22	131.9	6.32 m		
23	135.9	6.45 m		
24	131.8	5.75 d,d (7.6,15.0)	23,25	22,25,26
25	71.3	4.71 d,d (8.1,8.1)	24,26	23,24,26,36-38
26	76.5	3.62 m	25, 27a/b	25,36-38
27a	61.6	4.39 m	26, 27b	25,29
27b		4.47 m	26, 27b	25,26,29
28	O			
29	157.7			
30	<sup>15</sup> N 53.0 <sup>e</sup>			
31	19.8	1.43 d (6.4)	10	10,11
32a	20.8	2.39 m	32b, 33	13,14,15,33
32b		2.45 m	32a, 33	13,14,15,33
33	14.9	1.19 t (7.7)	32 a/b	14,32
34	OH			
35	N			
36-38	54.9	3.37 s		26

<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

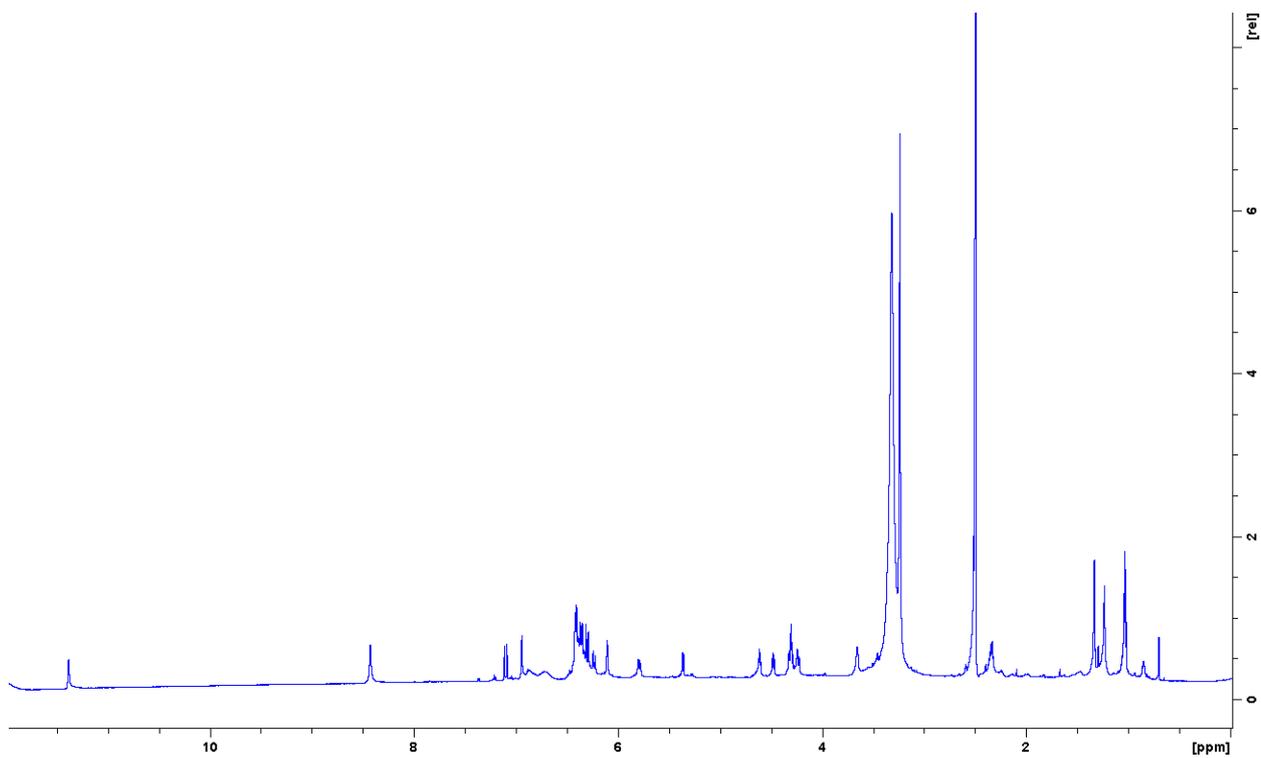
<sup>b</sup> acquired at 700 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

<sup>c</sup> proton showing COSY correlation to indicated proton.

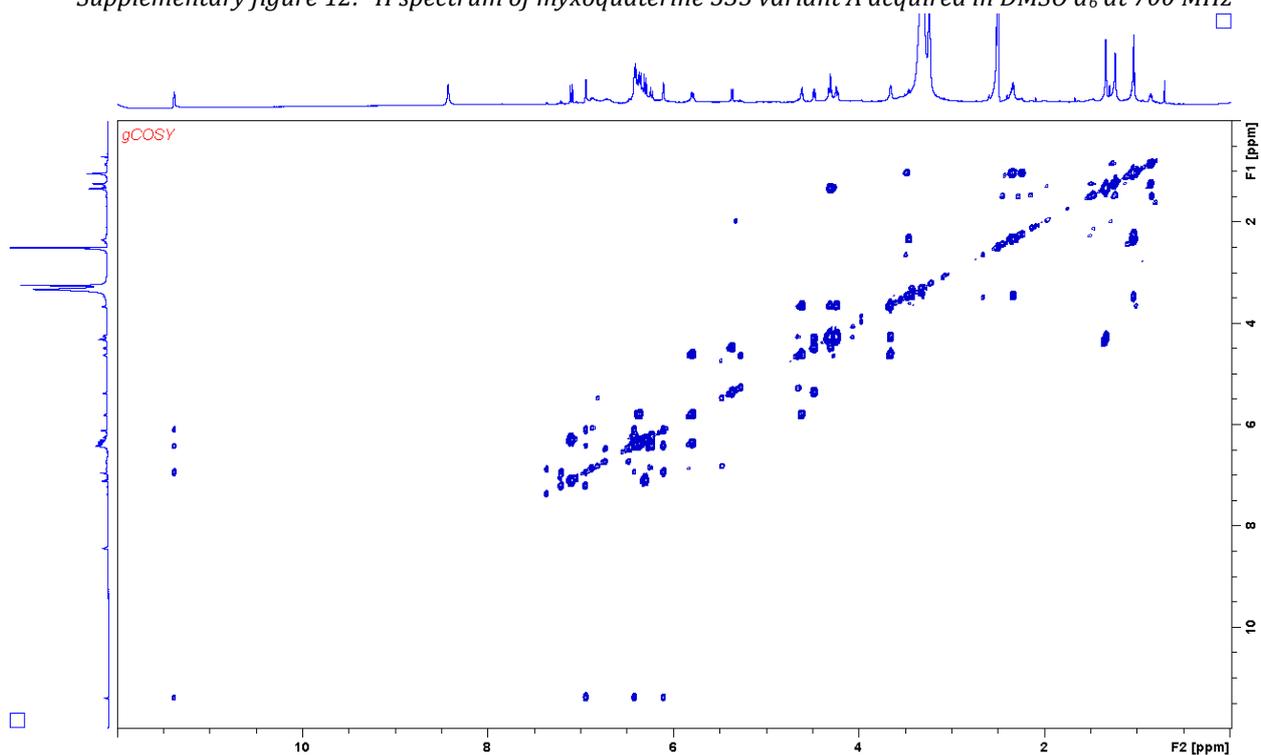
<sup>d</sup> proton showing HMBC correlations to indicated carbons.

<sup>e</sup> <sup>15</sup>N resonance obtained from <sup>15</sup>N-HMBC acquired at 71 MHz

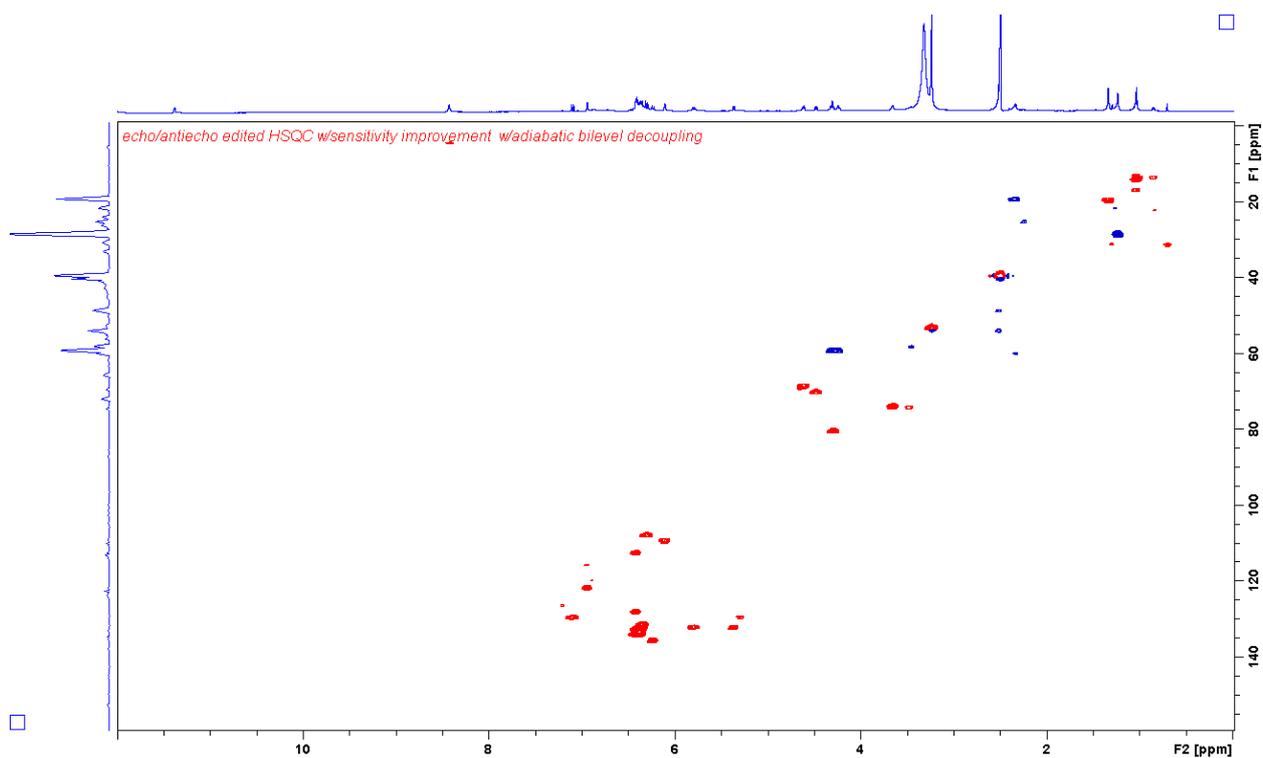
NMR spectra were acquired with a Bruker Ascend 700 NMR spectrometer equipped with a 5mm TCI cryoprobe



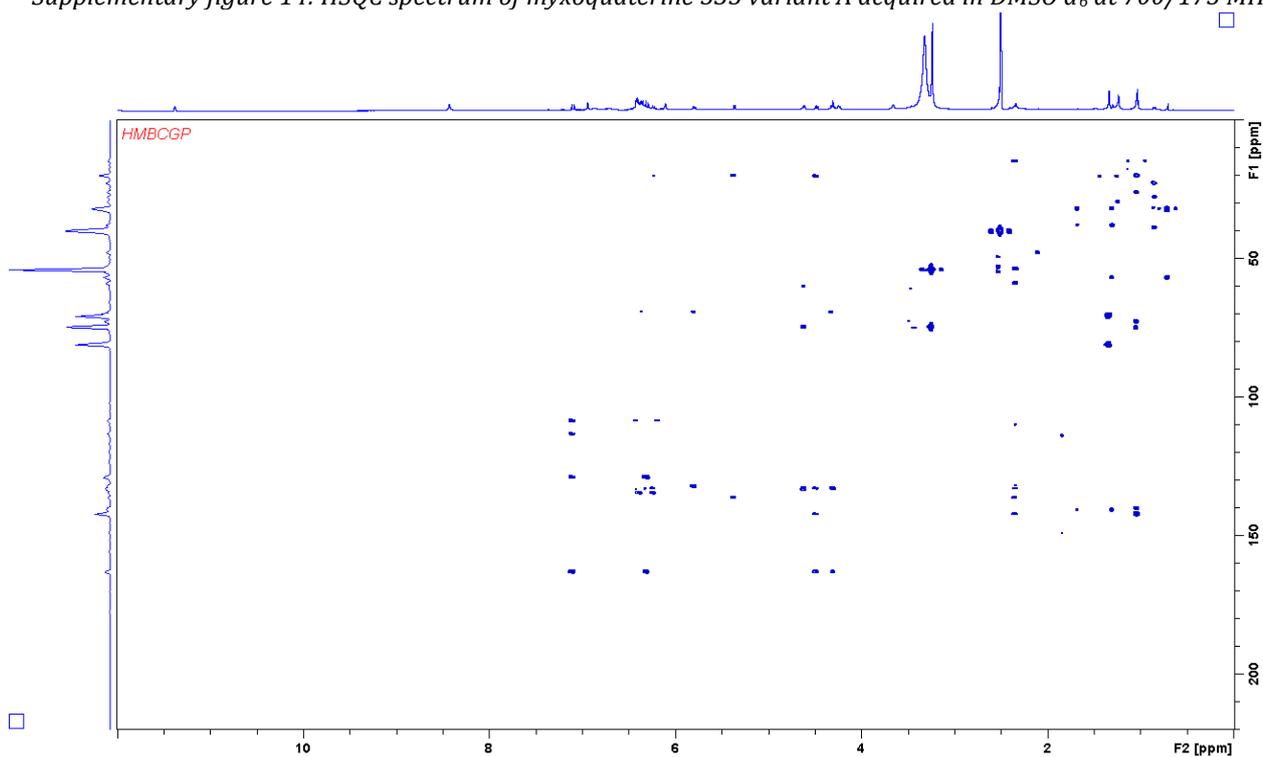
Supplementary figure 12:  $^1\text{H}$  spectrum of myxoquaterine 535 variant A acquired in  $\text{DMSO } d_6$  at 700 MHz



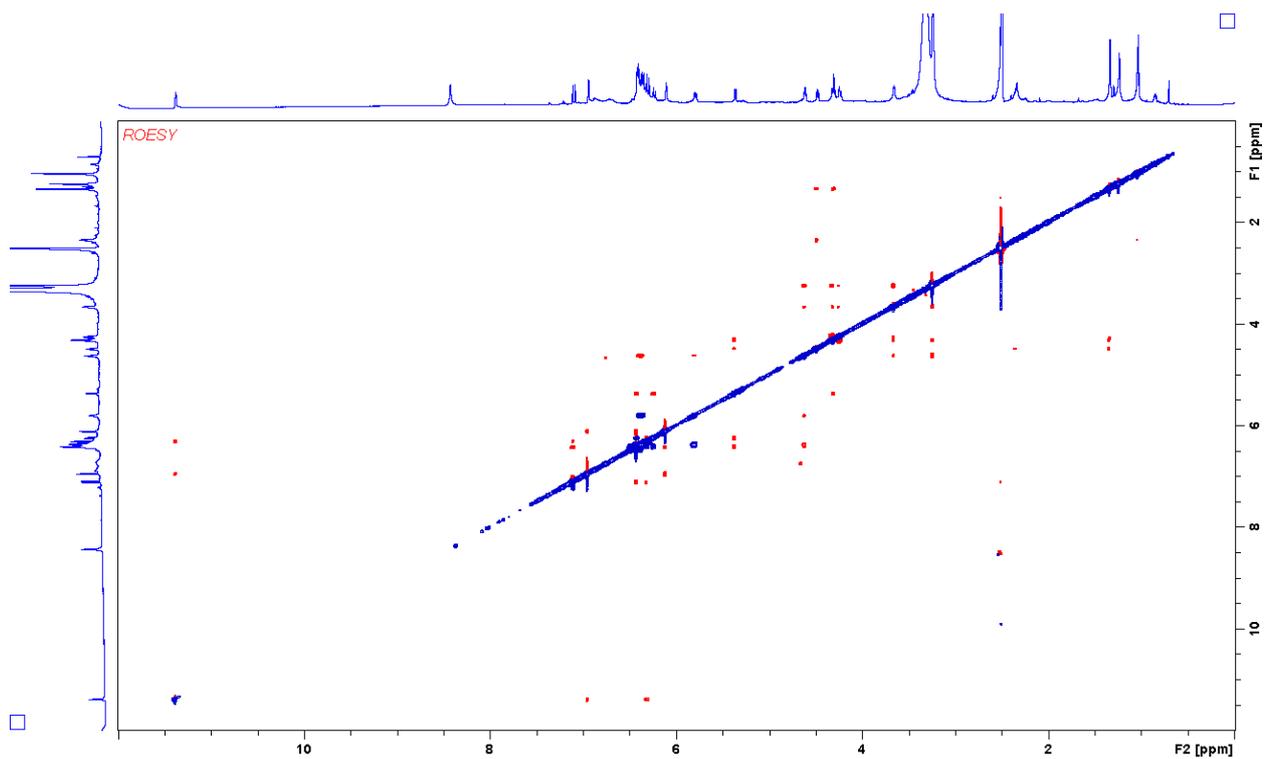
Supplementary figure 13: COSY spectrum of myxoquaterine 535 variant A acquired in  $\text{DMSO } d_6$  at 700 MHz



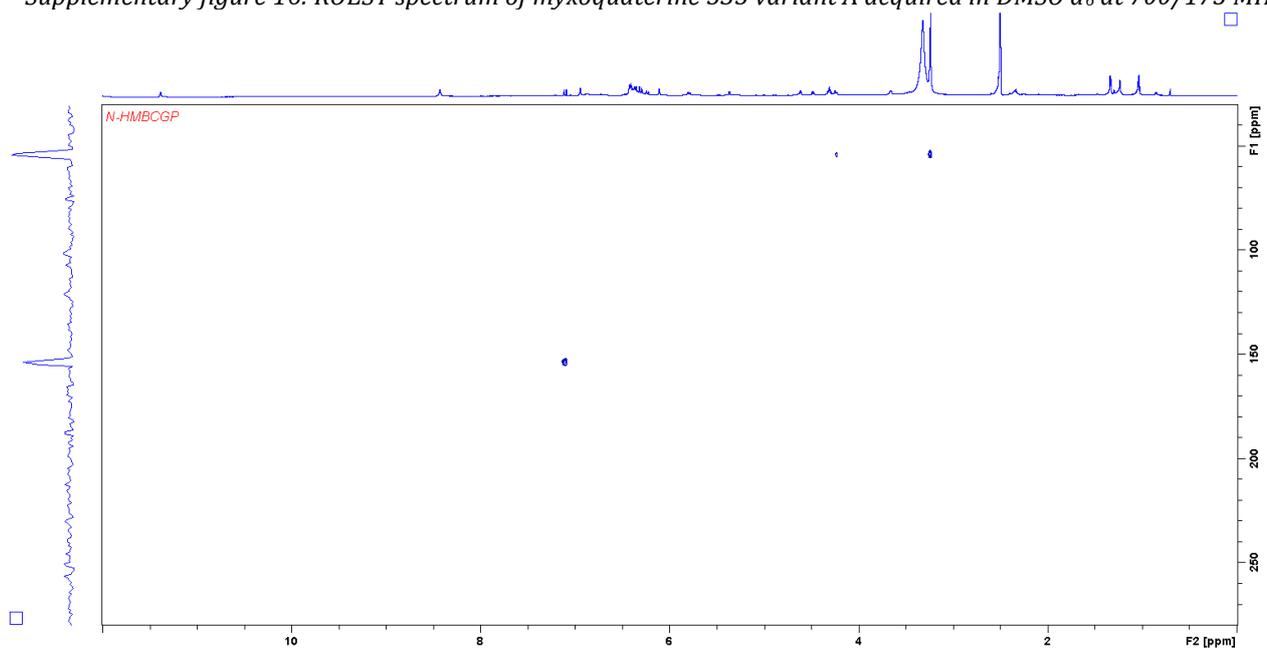
Supplementary figure 14: HSQC spectrum of myxoquaterine 535 variant A acquired in DMSO  $d_6$  at 700/175 MHz



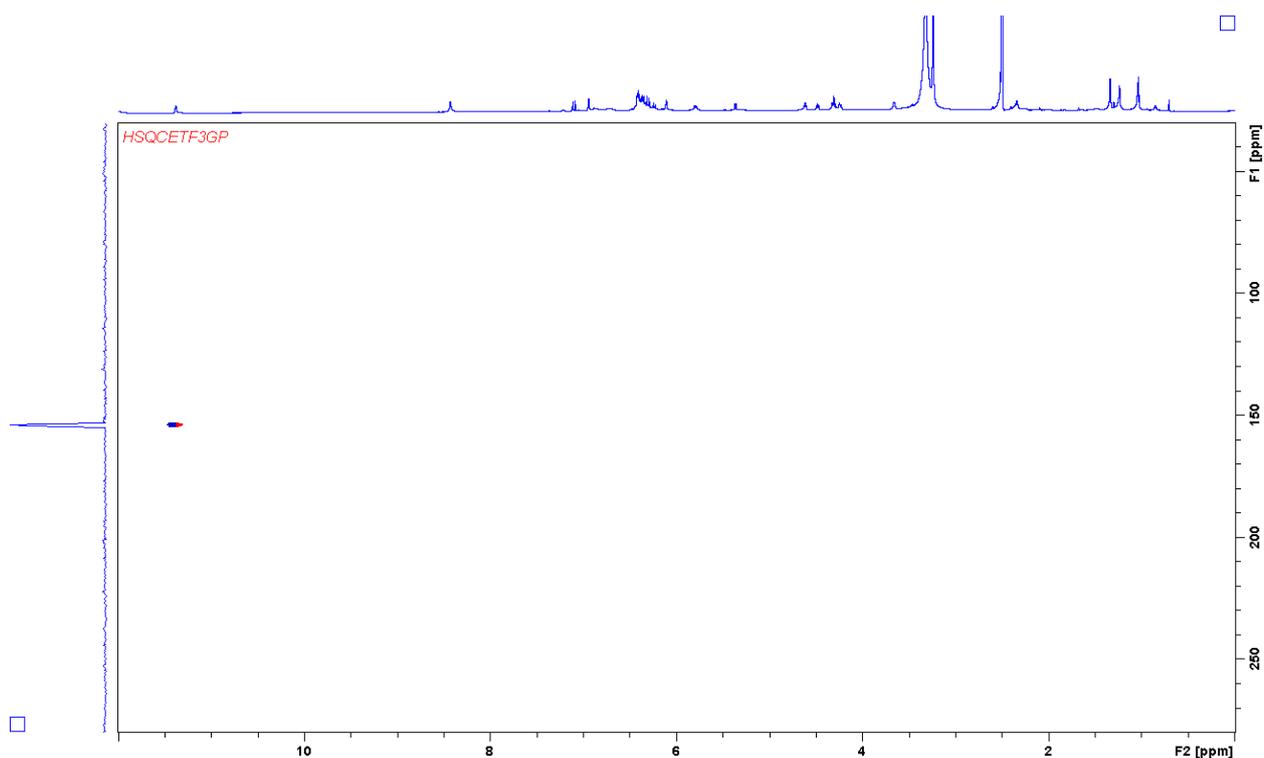
Supplementary figure 15: HMBC spectrum of myxoquaterine 535 variant A acquired in DMSO  $d_6$  at 700/175 MHz



Supplementary figure 16: ROESY spectrum of myxoquaterine 535 variant A acquired in DMSO  $d_6$  at 700/175 MHz



Supplementary figure 17:  $^{15}\text{N}$  HMBC spectrum of myxoquaterine 535 variant A acquired in DMSO  $d_6$  at 700/175 MHz



Supplementary figure 18:  $^{15}\text{N}$  HSQC spectrum of myxoquaterine 535 variant A acquired in  $\text{DMSO } d_6$  at 700/71 MHz

Supplementary table 2: Spectroscopic data of myxoquaterine 535 variant A in  $\text{DMSO } d_6$ .

position	$\delta_c^a$	$\delta_H^b$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
1	$^{15}\text{N } 154.3^e$	11.38		
2	121.8	6.95 s	3	3,5
3	109.2	6.11 m	2,4	
4	112.6	6.42 m	3	2
5	128.3			
6	129.5	7.10 d (16.3)	7	4,5,7,8
7	107.9	6.31 d (16.3)	6	5,6,8
8	162.5			
9				
10	80.5	4.29 m	11,31	8,14
11	70.2	4.48 d,d (8.9,7.3)	11,13	8,31,13,14
12				
13	132.5	5.36 d (9.2)		
14	141.6			
15	135.7	6.23 d (15.3)	16	
16	133.9	6.42 m		
17-22				
23	131.4	6.36 m	24	
24	132.2	5.79 d,d (7.3,14.2)	23,25	
25	68.6	4.62 d,d (7.3,7.3)	24,26	24,26,27
26	73.9	3.67 m	25	
27a	59.3	4.31 m	26	26
27b	59.3	4.23 m	26	29
29	154.9			
30	N			
31	19.6	1.33 d (6.0)	10	10,11
32a	19.4	2.34 m	33	15,33
32b	19.4	2,34 m	33	15,33

33	14.1	1.03 t (7.5)	32	14,16
36	53.4	3.24 s		26
37	53.4	3.24 s		26
38	53.4	3.24 s		26

<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal DMSO d<sub>6</sub> at δ 39.52 ppm.

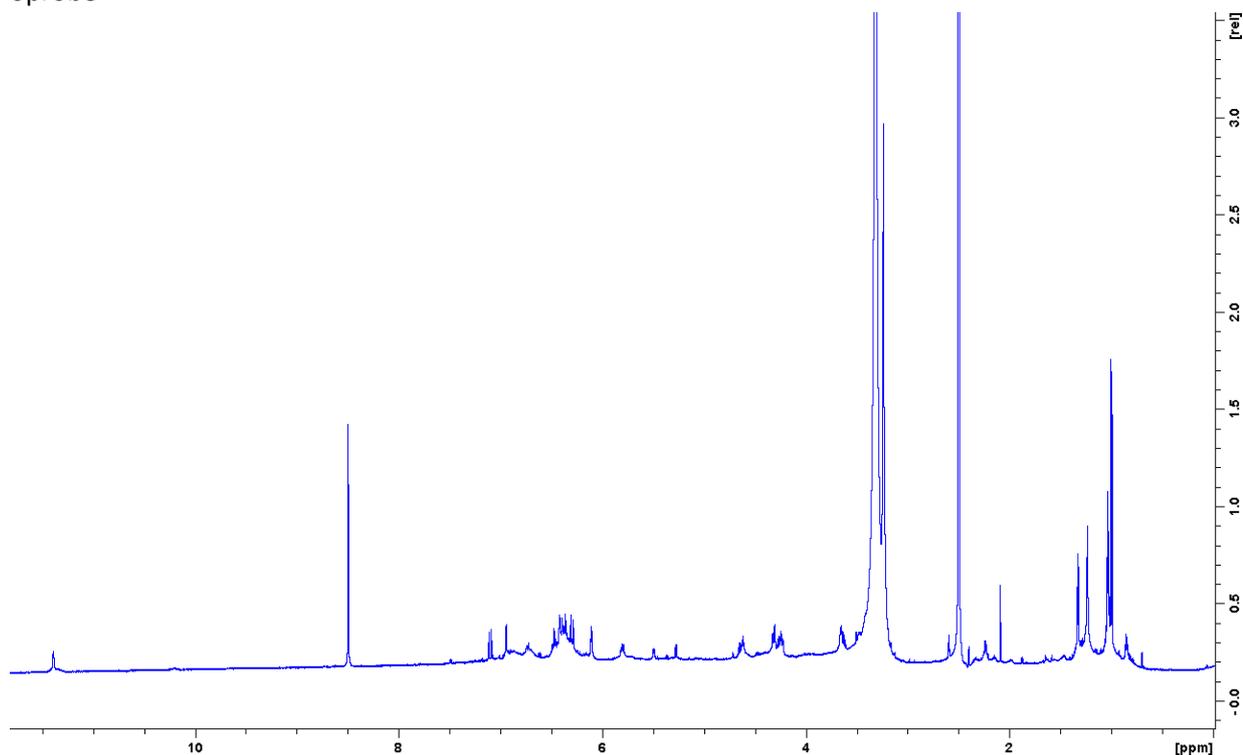
<sup>b</sup> acquired at 700 MHz, referenced to solvent signal DMSO d<sub>6</sub> at δ 2.50 ppm.

<sup>c</sup> proton showing COSY correlation to indicated proton.

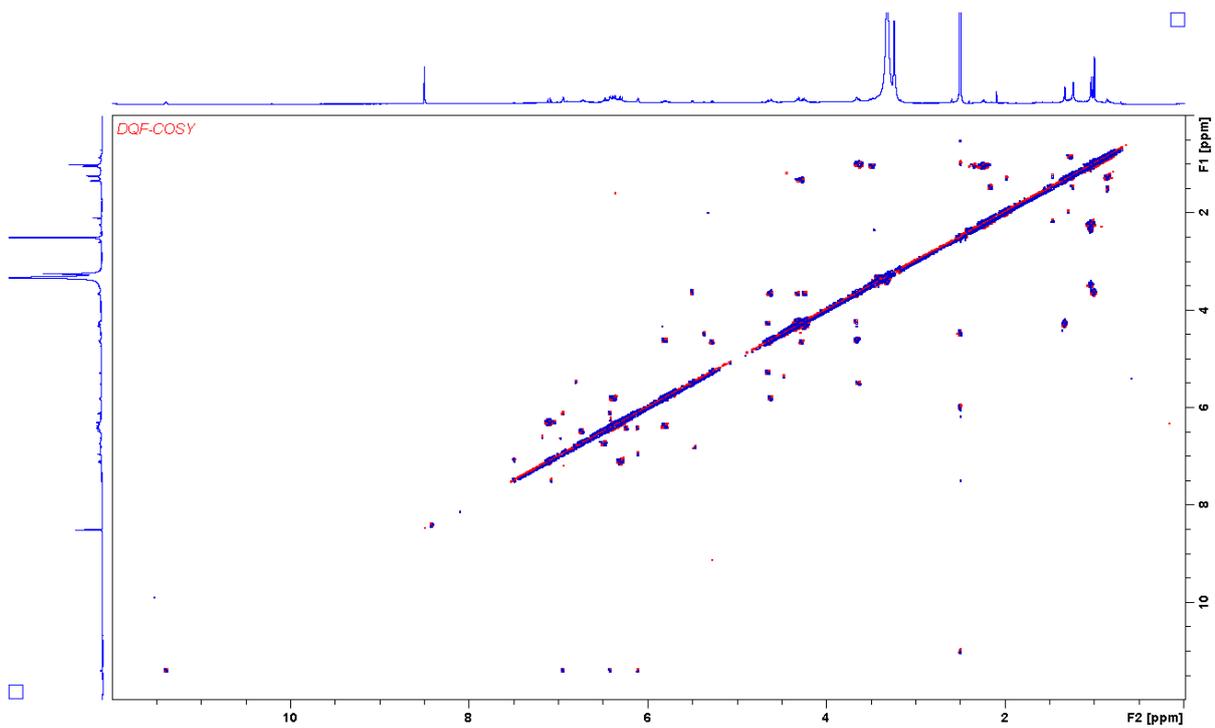
<sup>d</sup> proton showing HMBC correlations to indicated carbons.

<sup>e</sup> <sup>15</sup>N resonance obtained from <sup>15</sup>N-HMBC acquired at 71 MHz

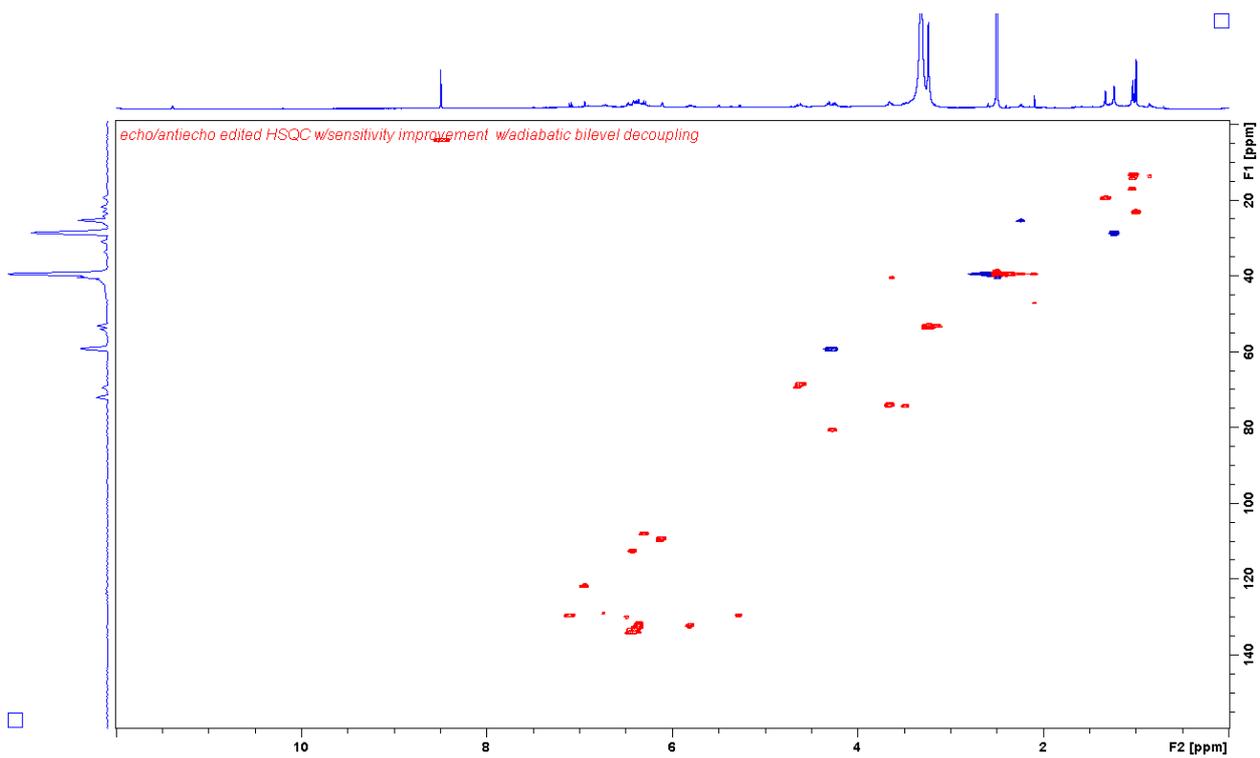
NMR spectra were acquired with a Bruker Ascend 700 NMR spectrometer equipped with a 5mm TCI cryoprobe



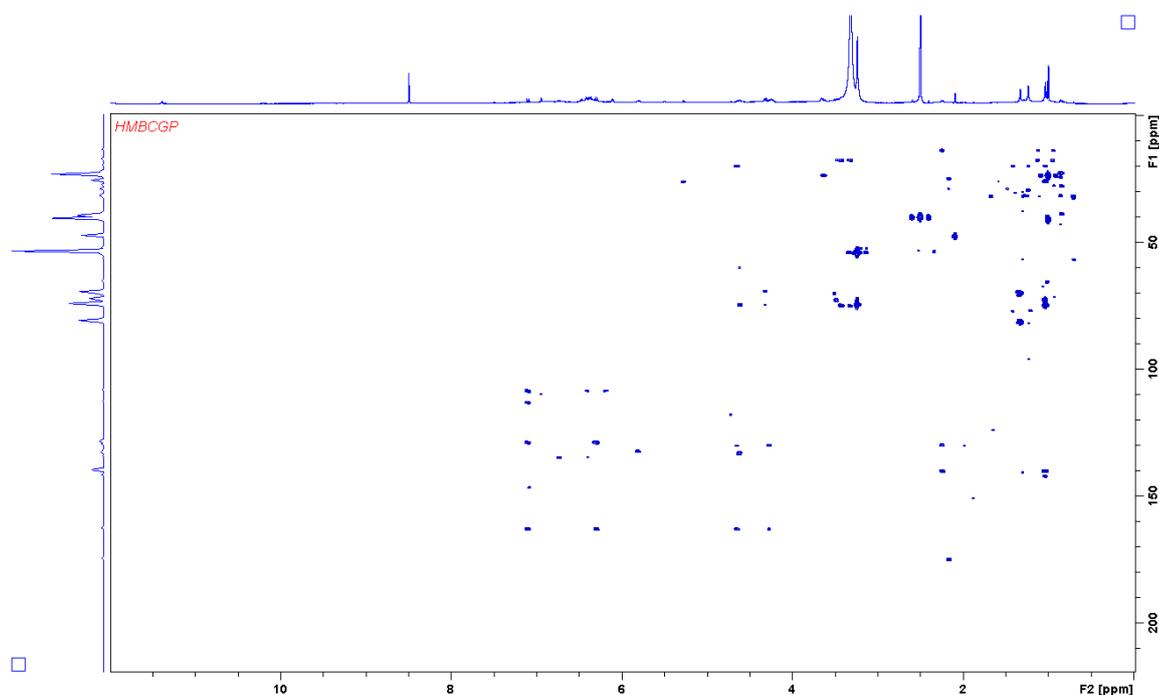
Supplementary figure 19: <sup>1</sup>H spectrum of myxoquaterine 535 variant B acquired in DMSO d<sub>6</sub> at 700 MHz



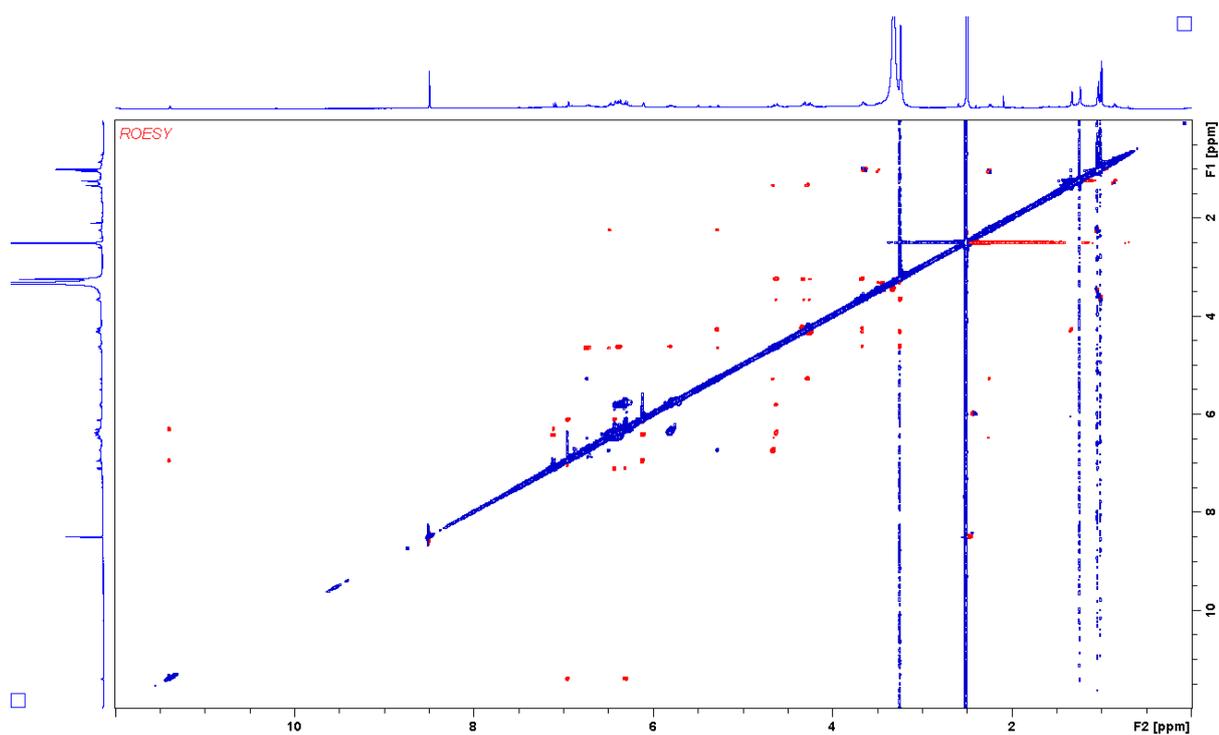
Supplementary figure 20: COSY spectrum of myxoquaterine 535 variant B acquired in DMSO d<sub>6</sub> at 700 MHz



Supplementary figure 21: HSQC spectrum of myxoquaterine 535 variant B acquired in DMSO d<sub>6</sub> at 700/175 MHz



Supplementary figure 22: HMBC spectrum of myxoquaterine 535 variant B acquired in DMSO  $d_6$  at 700/175 MHz



Supplementary figure 23: ROESY spectrum of myxoquaterine 535 variant B acquired in DMSO  $d_6$  at 700/175 MHz

Supplementary table 3: spectroscopic data of myxoquaterine 535 variant B in DMSO  $d_6$

position	$\delta_C^a$	$\delta_H^b$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
1	N	11.39 s	2,3,4	
2	122.6	6.95 m	3	3,5
3	109.0	6.12 m	2,4	
4	112.7	6.42 m	3	

5	128.2			
6	129.5	7.10 d (16.2)	7	4,5,7,8
7	108.0	6.30 d (16.3)	6	5,8
8	162.3			
9	o			
10	80.6	4.27 m	11	6,8
11	69.4	4.65 m	10,14	6,8,31
12	N			
13	129.6	5.28 d (9.0)	11	15,32
14	139.9			
15	128.9	6.74 d (15.0)		
16-22	129.5-134.2	6.28-6.52		
23	131.4	6.36 m	24	
24	132.2	5.79 d,d (7.3, 14.2)	23,25	23,25
25	68.6	4.62 d,d (7.2,7.3)	24,26	24,26,27a/b
26	73.9	3.67 m	25, 27a/b	
27a	59.3	4.23 m	26,27b	26,29
27b	59.3	4.31 m	26,27b	29
28	O			
29	155.2			
30	N			
31	19.5	1.33 d (6.1)	10	10,11
32	25.5	2.24 q (7.2)	33	33
33	13.3	1.03 t (7.4)	32	14,32
34	OH			
35	N			
36-38	53.3	3.24 s		26

<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal DMSO d<sub>6</sub> at  $\delta$  39.52 ppm.

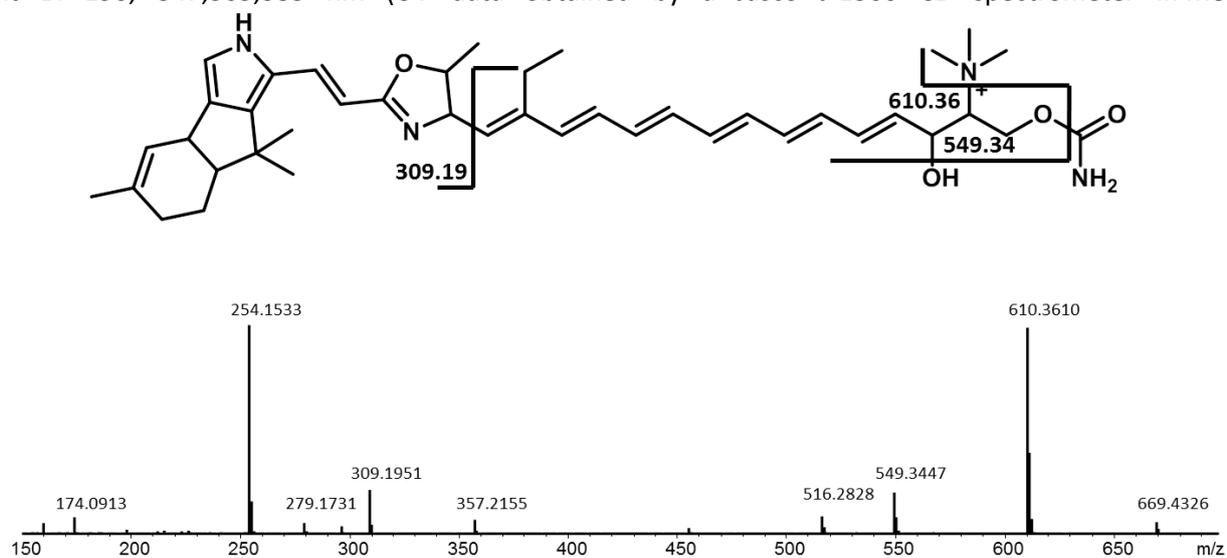
<sup>b</sup> acquired at 700 MHz, referenced to solvent signal DMSO d<sub>6</sub> at  $\delta$  2.50 ppm.

<sup>c</sup> proton showing COSY correlation to indicated proton.

<sup>d</sup> proton showing HMBC correlations to indicated carbons.

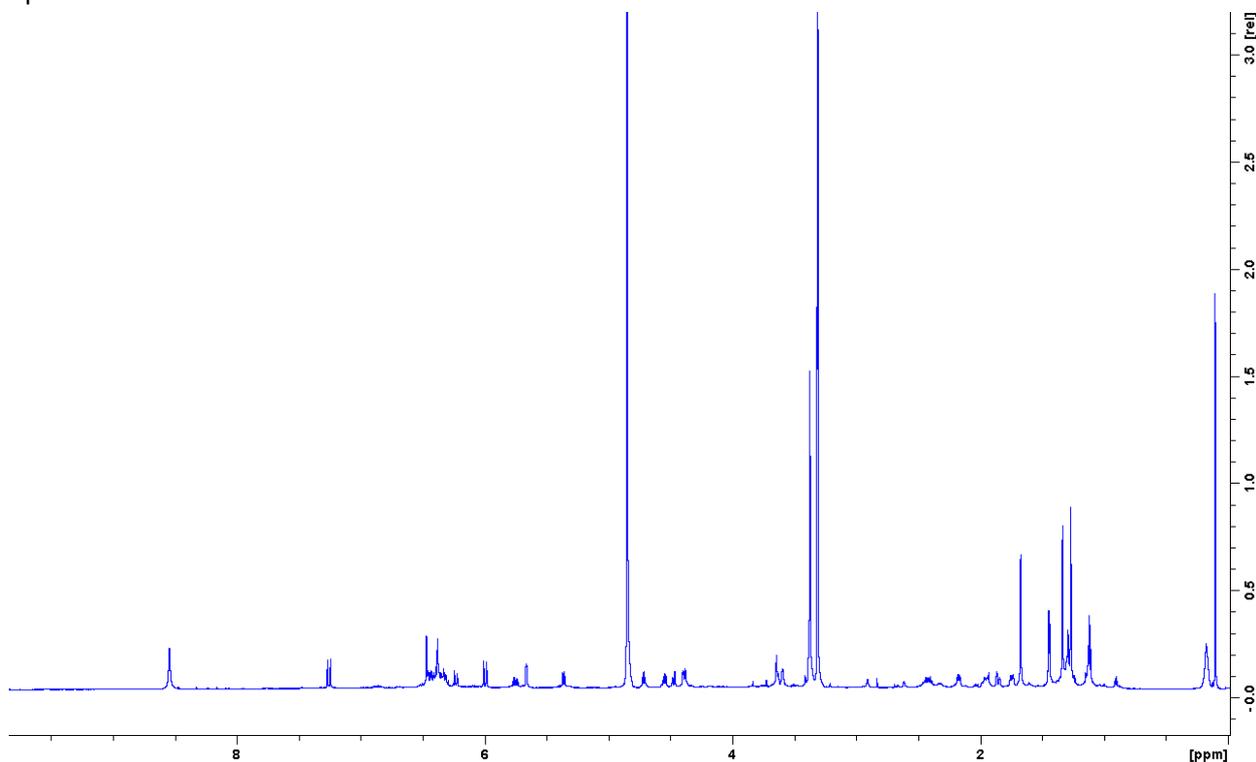
### Analytical data myxoquaterin 669

Myxoquaterin 669 was isolated brown yellowish amorphous solid that turned into a powder after lyophilization. *hrMass*: 669.44082 [M]<sup>+</sup>; 335.22330 [M+H]<sup>2+</sup>; specific optical rotation variant A:  $[\alpha]_D^{20}$  38.1 °g<sup>-1</sup>ml dm<sup>-1</sup>; specific optical rotation variant B:  $[\alpha]_D^{20}$  86.3 °g<sup>-1</sup>ml dm<sup>-1</sup>; UV<sub>max</sub> Variant A: 192, 347, 364, 383 nm; UV<sub>max</sub> Variant B: 190, 347, 365, 385 nm (UV data obtained by a Jasco J-1500 CD spectrometer in MeOH)

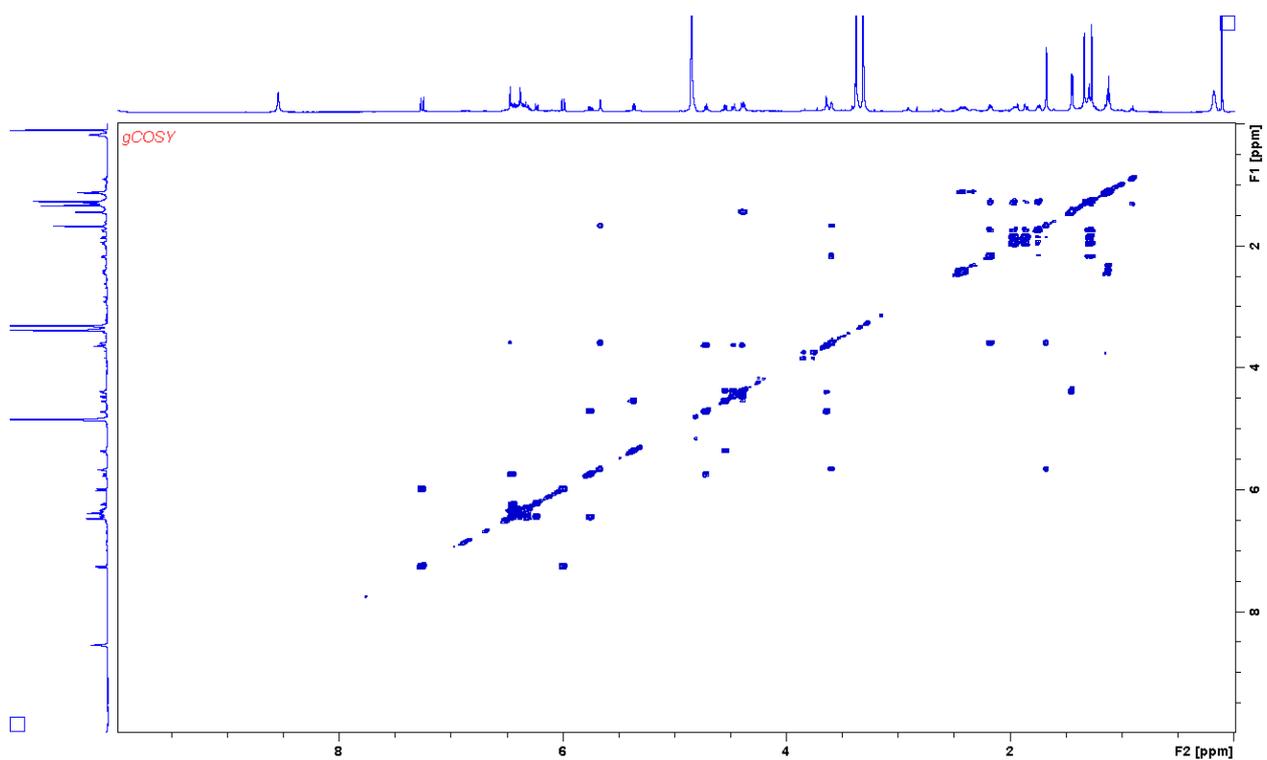


Supplementary figure 24: *HrMS*<sup>2</sup> spectrum and observed fragments of myxoquaterine 669. *MS*<sup>2</sup> spectrum is extracted from *LC-hrMS* measurement of an extract of *MSr* 11954.

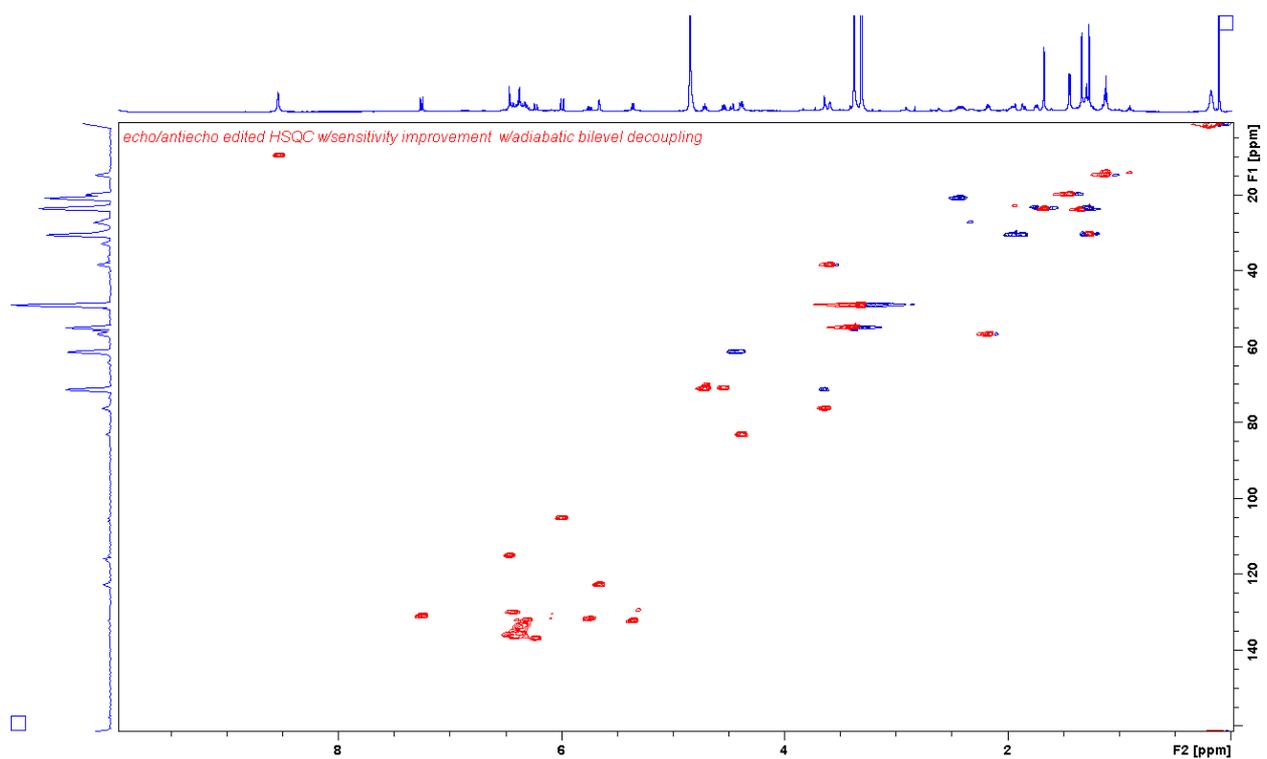
NMR spectra were acquired with a Bruker Ascend 700 NMR spectrometer equipped with a 5mm TCI cryoprobe



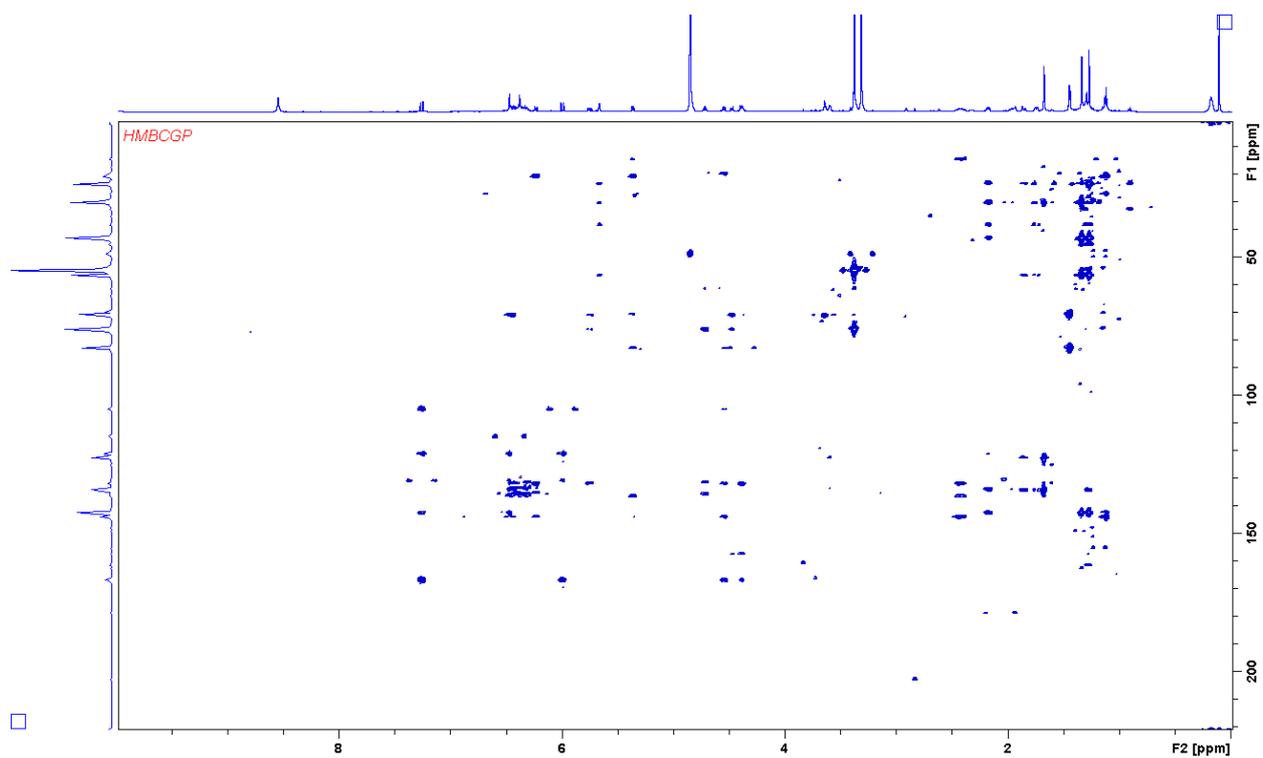
Supplementary figure 25: <sup>1</sup>H spectrum of myxoquaterine 669 variant A acquired in CD<sub>3</sub>OD at 700 MHz



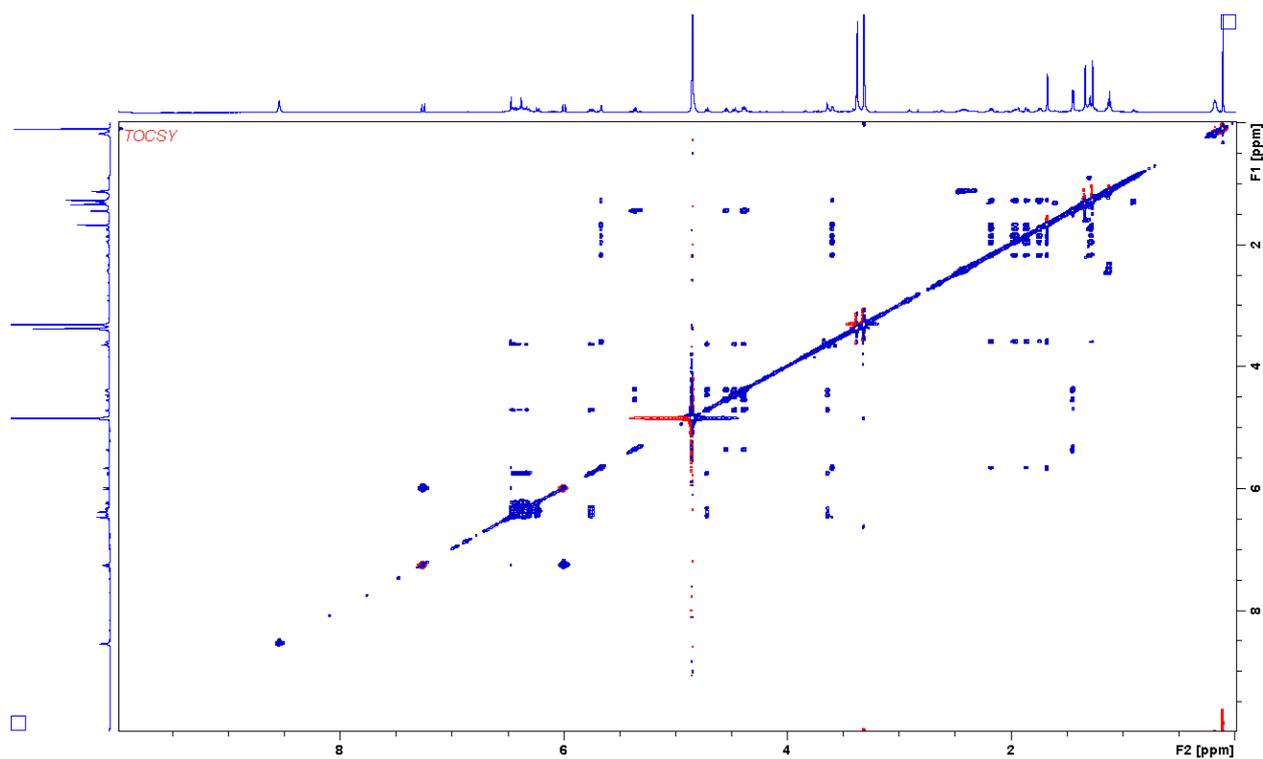
Supplementary figure 26: COSY spectrum of myxoquaterine 669 variant A acquired in CD<sub>3</sub>OD at 700 MHz



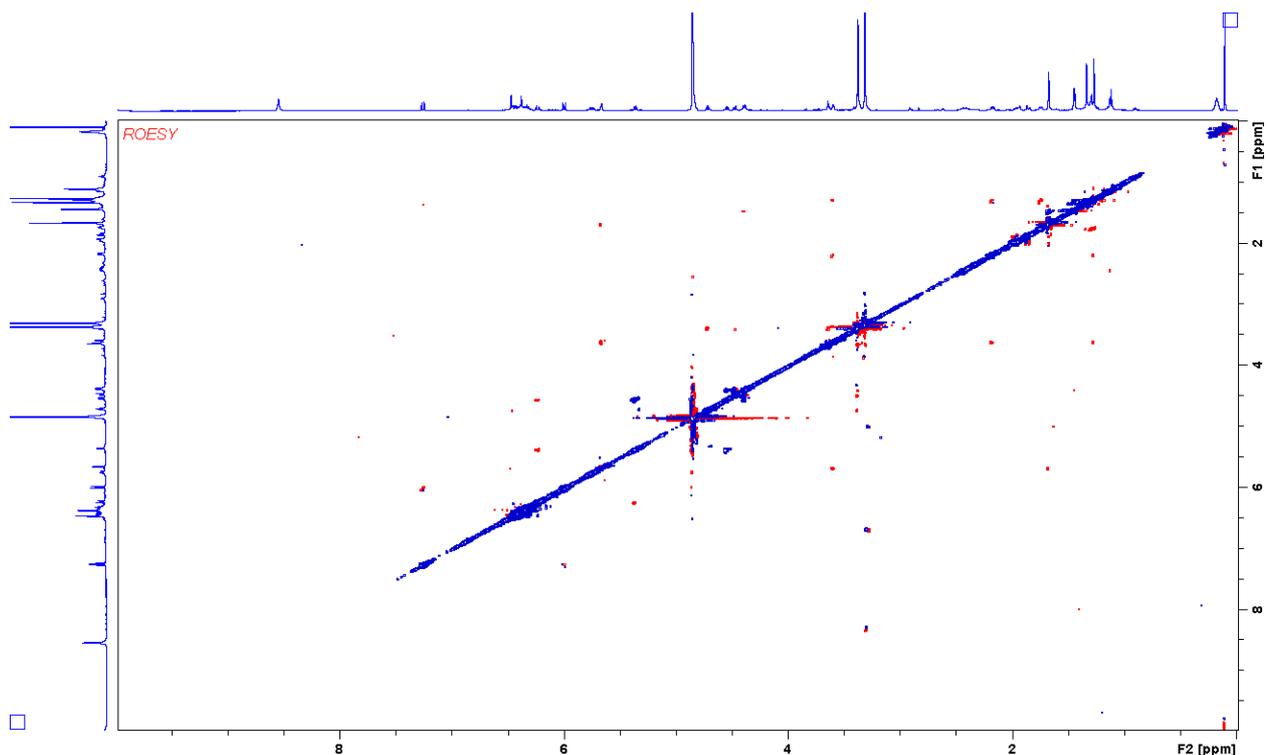
Supplementary figure 27: HSQC spectrum of myxoquaterine 669 variant A acquired in CD<sub>3</sub>OD at 700/175 MHz



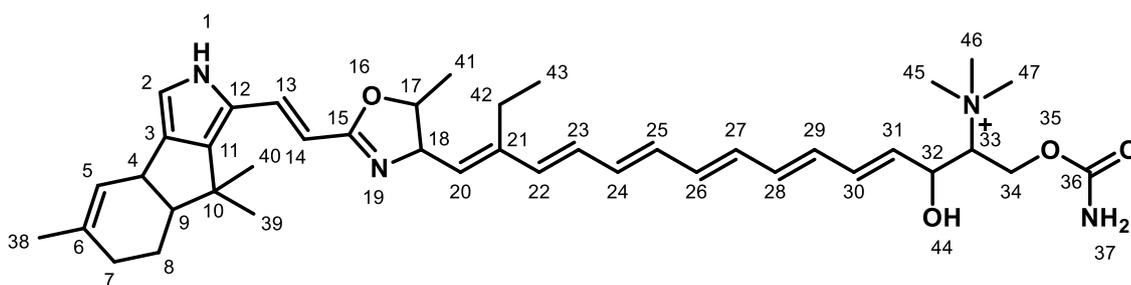
Supplementary figure 28: HMBC spectrum of myxoquaterine 669 variant A acquired in  $CD_3OD$  at 700/175 MHz



Supplementary figure 29: TOCSY spectrum of myxoquaterine 669 variant A acquired in  $CD_3OD$  at 700 MHz



Supplementary figure 30: ROESY spectrum of myxoquaterine 669 variant A acquired in CD<sub>3</sub>OD at 700 MHz



Supplementary figure 31: Structure of myxoquaterine 669

Supplementary table 4: Spectroscopic data of myxoquaterine 669 variant A in CD<sub>3</sub>OD.

position	$\delta_C^a$	$\delta_H^b$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
1	N			
2	115.1	6.47 s		3,12,11
3	134.1			
4	38.59	3.59 m	5,9	5
5	122.8	5.66 m	4,38	4,7,9,38
6	134.5			
7a	30.6	1.96 m	7b, 8a/b	
7b		1.86 m	7a, 8a/b	5,8,9
8a	23.3	1.74 m	7a/b,8b,9	4,6,9
8b		1.28 m	7a/b,8a,9	
9	56.8	2.17 m	8a/b,4	3,4,7,8,10,11
10	43.1			
11	142.7			
12	121.3			
13	130.9	7.25 d 15.9	14	11,12,14,15
14	105.2	5.99 d 16.0	13	12,13,15
15	167			
16	O			
17	83.2	4.38 m	18,41	20,15

18	70.9	4.54 d,d 9.6, 7.3	17,20	15,21,41
19	N			
20	132.2	5.36 d 9.6	18	17,18,42,43
21	144.2			
22	136.8	6.23 d 15.4	23	20,21,23,42
23	135.5	6.43 m		
24-29	132.0-136.7	6.29-6.45		
30	135.7	6.45 m		
31	131.4	5.75 d,d 14.9,7.8		
32	71.0	4.71 d,d 8.26,8.28	31,33	33, 34
33	76.3	3.63 m	32, 34	32, 34
34a	61.4	4.47 d,d 14.1, 2.5	33,34b	33,36
34b		4.38 m	33,34b	36
35	O			
36	157.3			
37	N			
38	23.6	1.67 s		5,6,7
39	23.9	1.33 s		9,10,11,40
40	30.1	1.27 s		9,10,11,39
41	19.9	1.44 d 6.1		
42a	20.7	2.38 m	42b, 43	20,21,22
42b	20.5	2.47 m	42a, 43	21,42
43	14.6	1.12 t 7.6	42 a/b	21,42
44	OH			
45	55.0	3.37 s		33
46	55.0	3.37 s		33
47	55.0	3.37 s		33

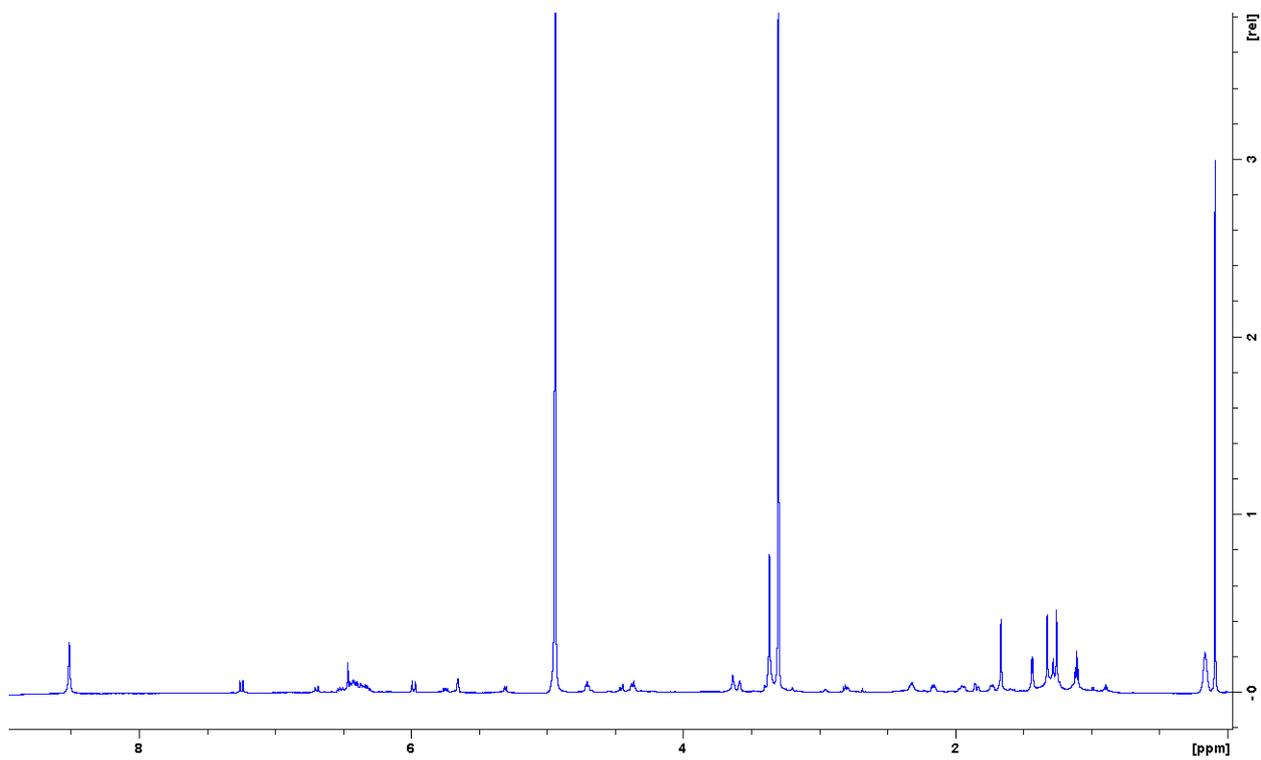
<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

<sup>b</sup> acquired at 700 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

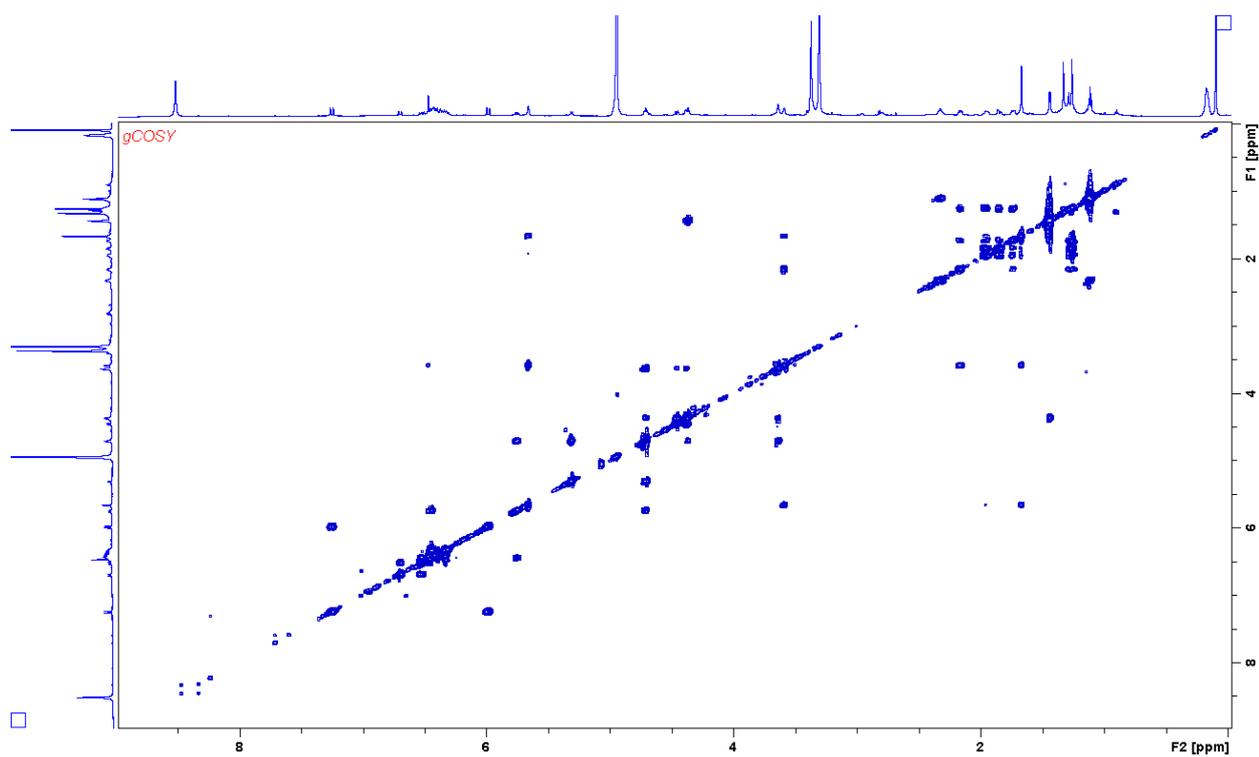
<sup>c</sup> proton showing COSY correlation to indicated proton.

<sup>d</sup> proton showing HMBC correlations to indicated carbons.

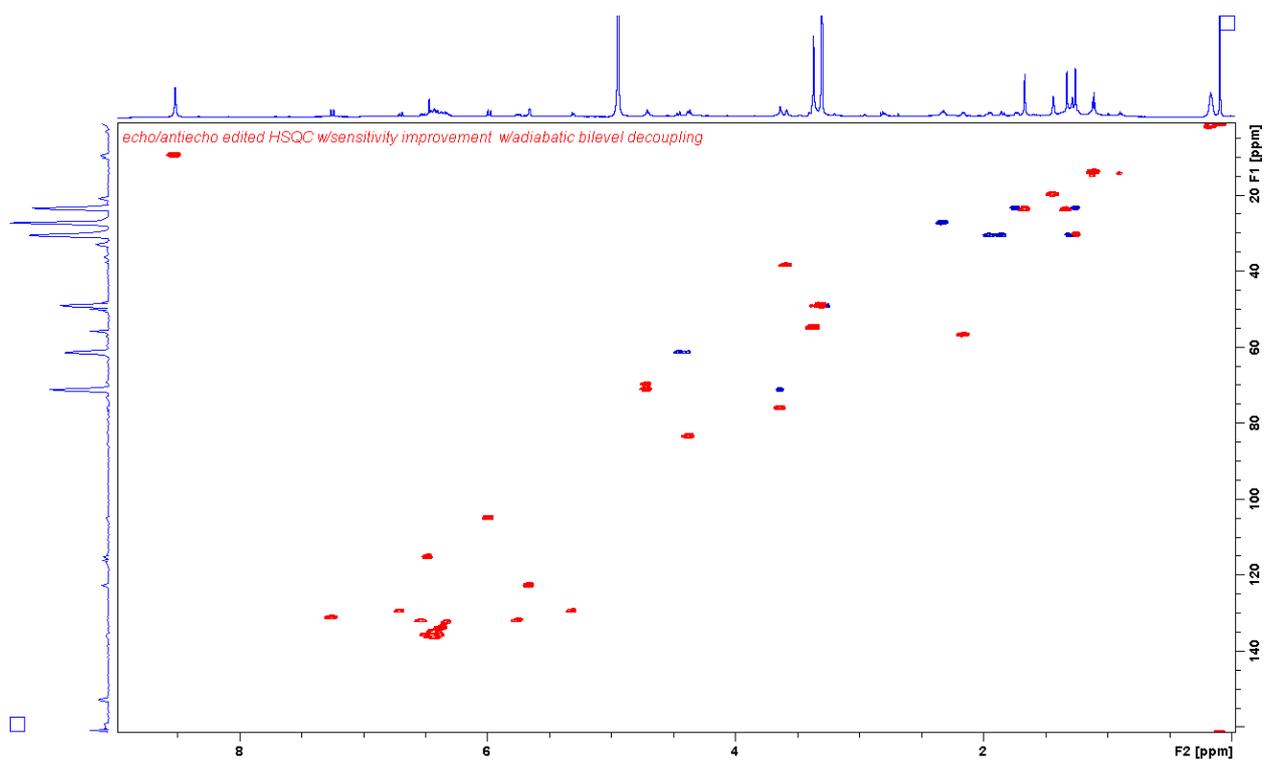
NMR spectra were acquired with a Bruker Ascend 700 NMR spectrometer equipped with a 5mm TCI cryoprobe



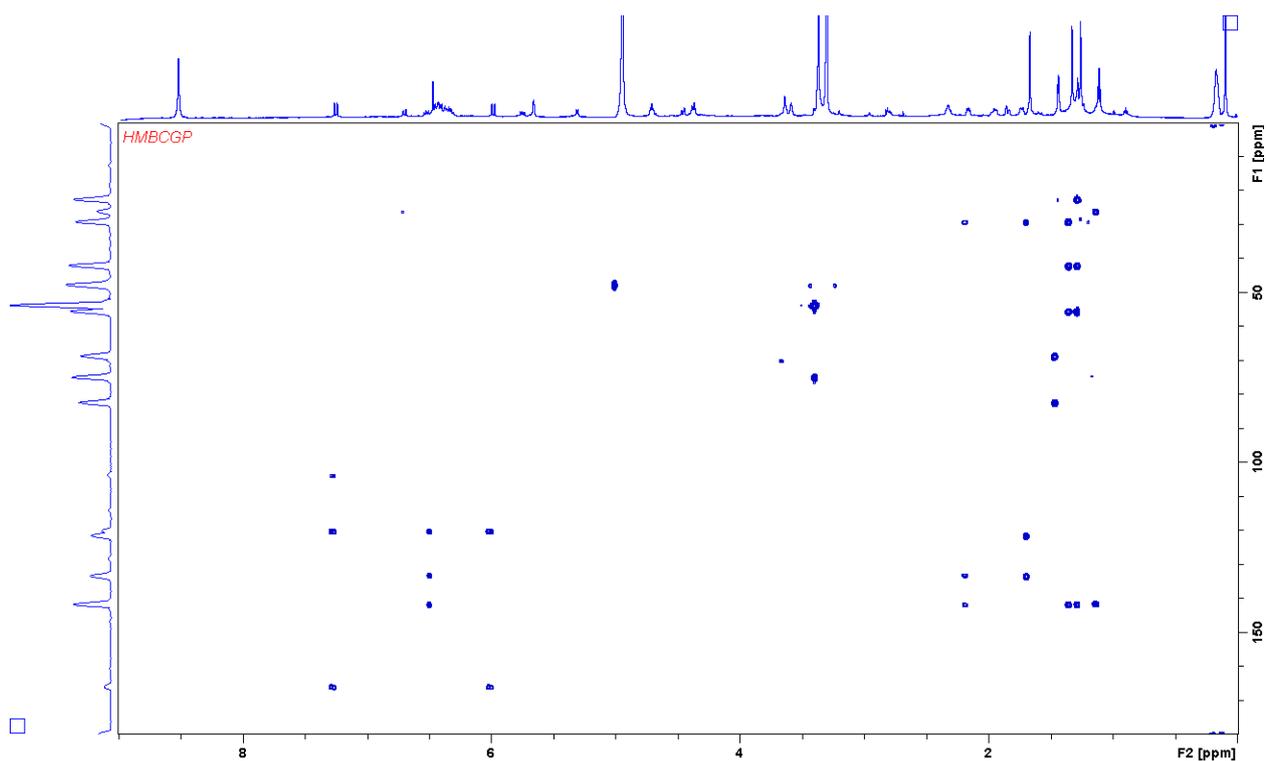
Supplementary figure 32:  $^1\text{H}$  spectrum of myxoquaterine 669 variant B acquired in  $\text{CD}_3\text{OD}$  at 700 MHz



Supplementary figure 33: COSY spectrum of myxoquaterine 669 variant B acquired in  $\text{CD}_3\text{OD}$  at 700 MHz



Supplementary figure 34: HSQC spectrum of myxoquaterine 669 variant B acquired in CD<sub>3</sub>OD at 700/175 MHz



Supplementary figure 35: HMBC spectrum of myxoquaterine 669 variant B acquired in CD<sub>3</sub>OD at 700/175 MHz

Supplementary table 5: Spectroscopic data of Myxoquaterine 669 variant B in CD<sub>3</sub>OD.

position	$\delta^{13}\text{C}^{\text{a}}$	$\delta^1\text{H}^{\text{b}}$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
1	N			
2	115.1	6.47 s		3,12,11
3	134.1			
4	38.6	3.59 m	5	
5	122.8	5.66 m	4,38	4,7,9,38

6	134.4			
7a	30.5	1.95 m	7b, 8a/b	
7b		1.85 m	7a, 8a/b	5,6
8a	23.3	1.74 m	7a/b,8b,9	4,6,9
8b		1.26 m	7a/b,8a,9	
9	56.6	2.16 m	8a/b	3,4,7,8,10,11
10	43.1			
11	142.8			
12	121.1			
13	131.0	7.25 d	14	11,12,14,15
14	105.2	5.99 d	13	12,13,15
15	166.9			
16	O			
17	83.4	4.37 m	18,41	20,15
18	69.7	4.72 d,d	17,20	15,17,20,21
19	N			
20	129.2	5.31 d	18	17,22,41
21	142.4			
22	129.2	6.69 d	23	20,21,42
23	131.5	6.53 d,d	22	
24-31	131.8-136.9	6.27-6.48		
32	71.1	4.70 m	31,33	33
33	76.0	3.64 m	32, 34	45,46,47
34a	61.4	4.46 m	33,34b	32, 36, 45,46,47
34b		4.38 m	33,34b	36
35	O			
36	157.6			
37	N			
38	23.6	1.67 s		5,6,7
39	23.8	1.33 s		9,10,11,40
40	30.4	1.26 s		9,10,11,39
41	19.7	1.44 d	17	17,18
42a	27.3	2.33 m	43	20,21,22,43
42b		2.31 m	43	
43	13.7	1.12 t	42 a/b	21,42
44	OH			
45	54.8	3.37 s		33
46	54.8	3.37 s		33
47	54.8	3.37 s		33

<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

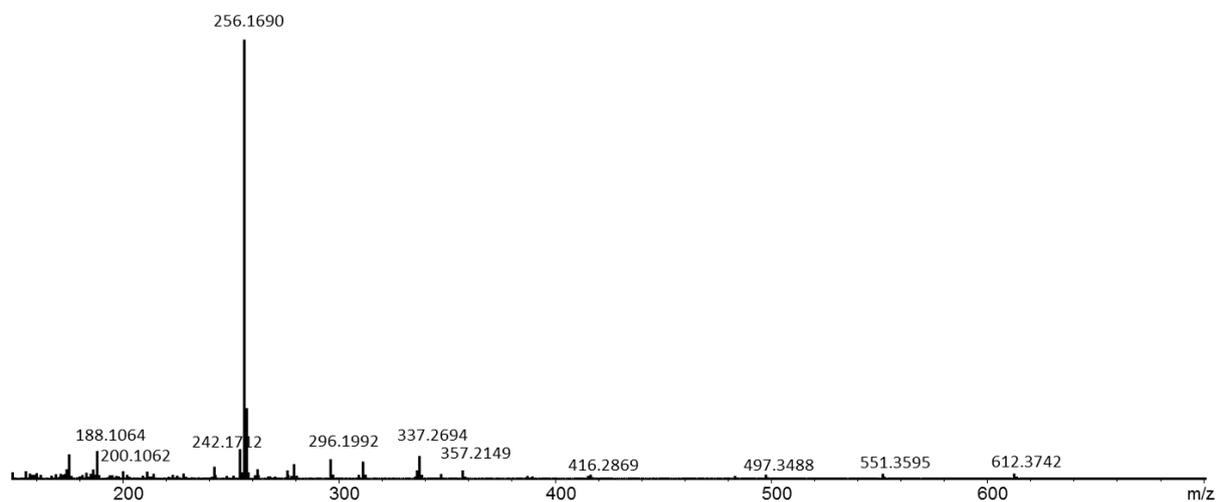
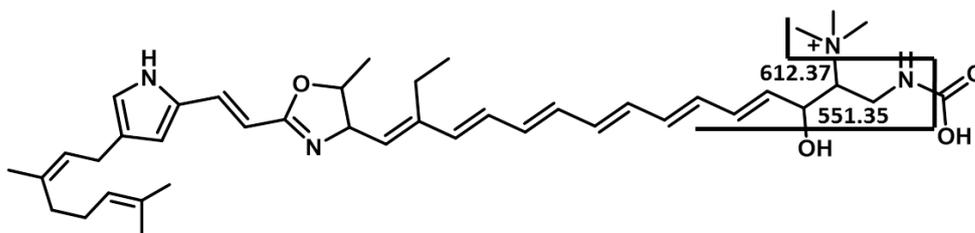
<sup>b</sup> acquired at 700 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

<sup>c</sup> proton showing COSY correlation to indicated proton.

<sup>d</sup> proton showing HMBC correlations to indicated carbons.

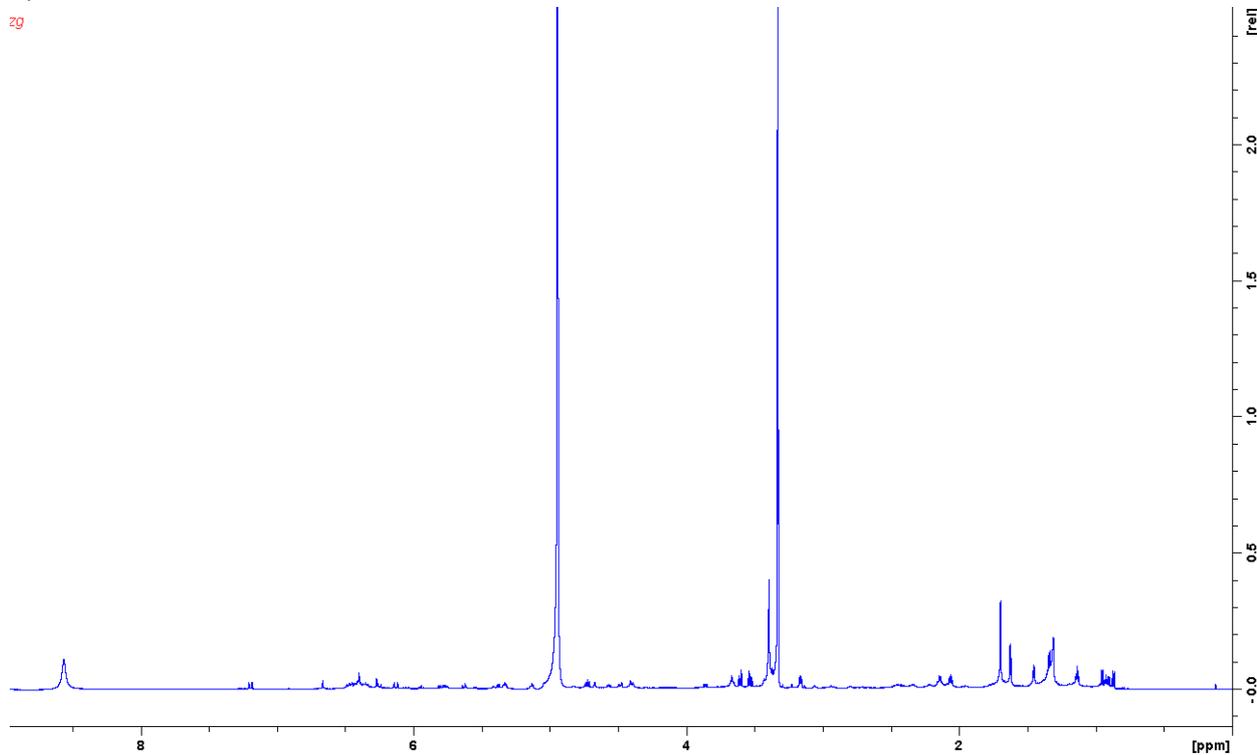
### Analytical data myxoquaterin 671

Myxoquaterin 671 was isolated brown yellowish amorphous solid that turned into a powder after lyophilization. *hr*Mass: 671.45425[M]<sup>+</sup>; 336.23055 [M+H]<sup>2+</sup>; specific optical rotation variant A:  $[\alpha]_D^{20}$  14.1 °g<sup>-1</sup>ml dm<sup>-1</sup>; specific optical rotation variant B:  $[\alpha]_D^{20}$  45.9 °g<sup>-1</sup>ml dm<sup>-1</sup> UV<sub>max</sub> Variant A: 193, 346, 364,384 nm; UV<sub>max</sub> Variant B: 192, 347,365,385 nm (UV data obtained by a Jasco J-1500 CD spectrometer in MeOH)

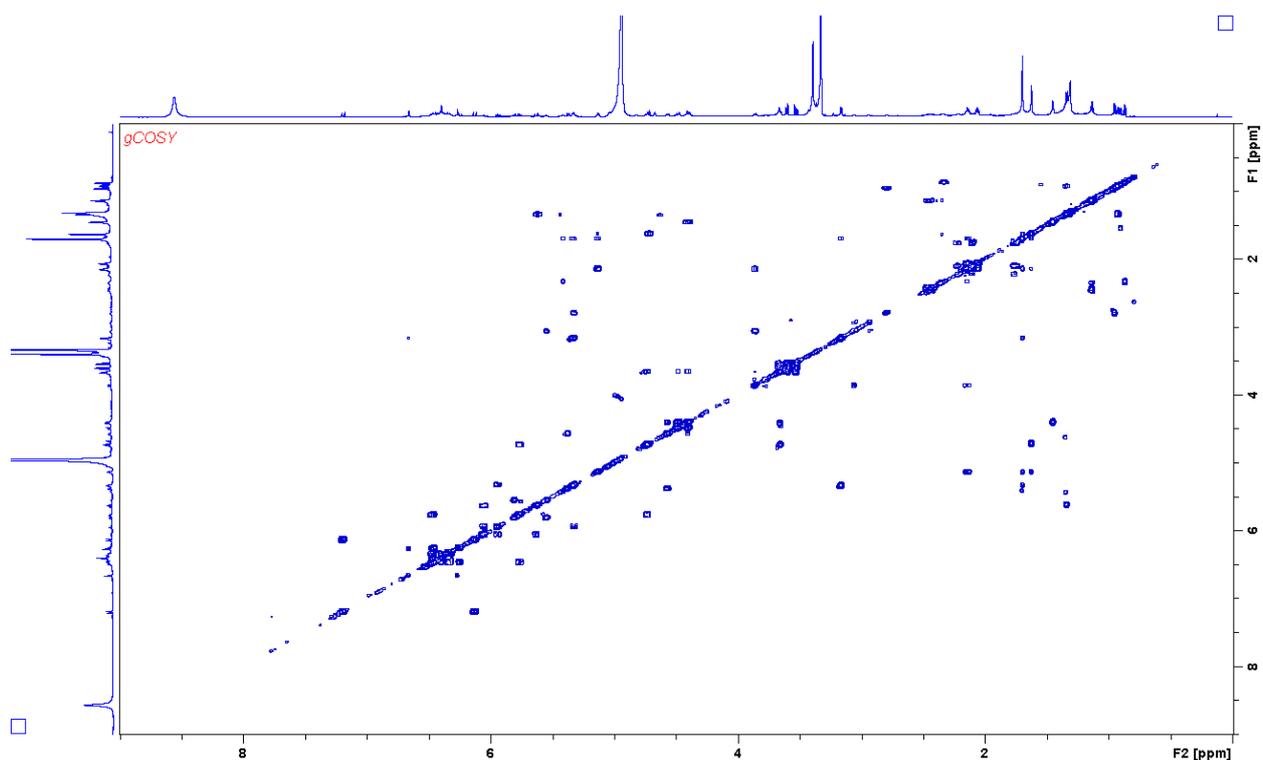


Supplementary figure 36: *hrMS*<sup>2</sup> spectrum and observed fragments of myxoquaterine 671. *MS*<sup>2</sup> spectrum is extracted from LC-*hrMS* measurement of an extract of MSr 11954.

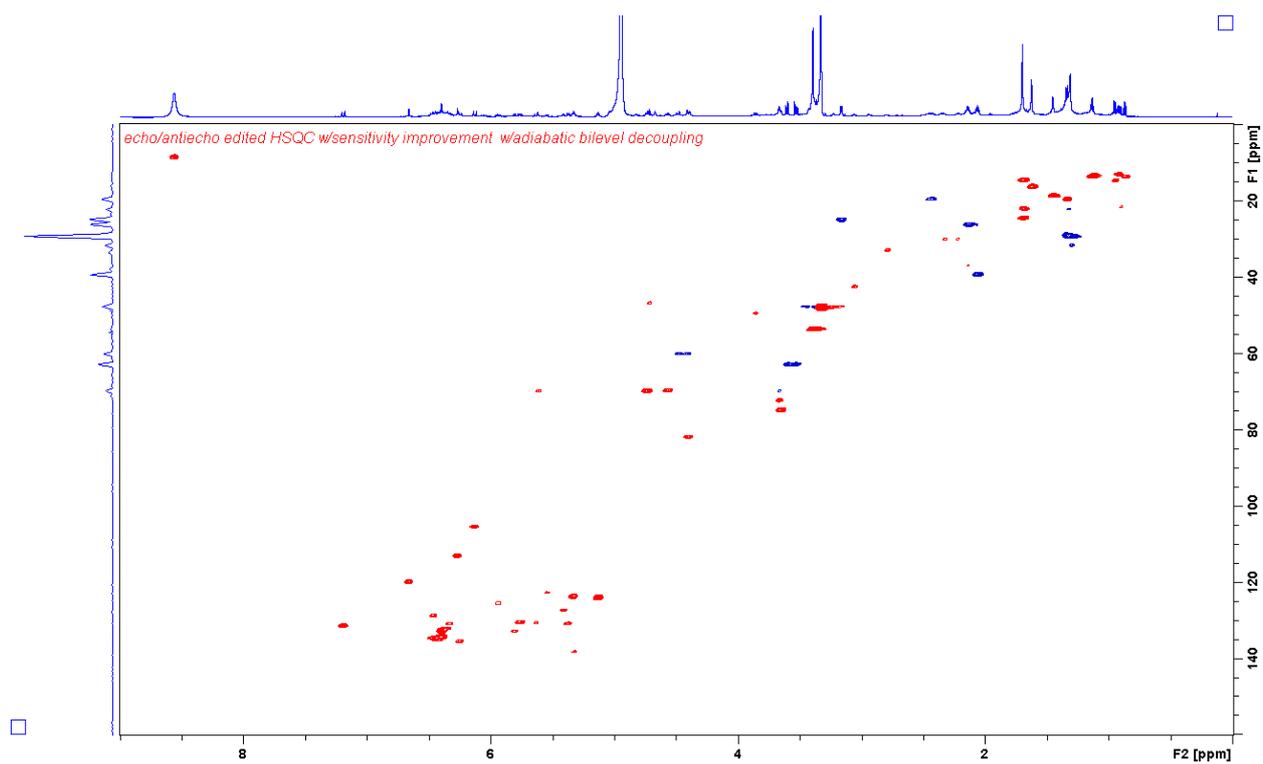
NMR spectra were acquired with a Bruker Ascend 700 NMR spectrometer equipped with a 5mm TCI cryoprobe



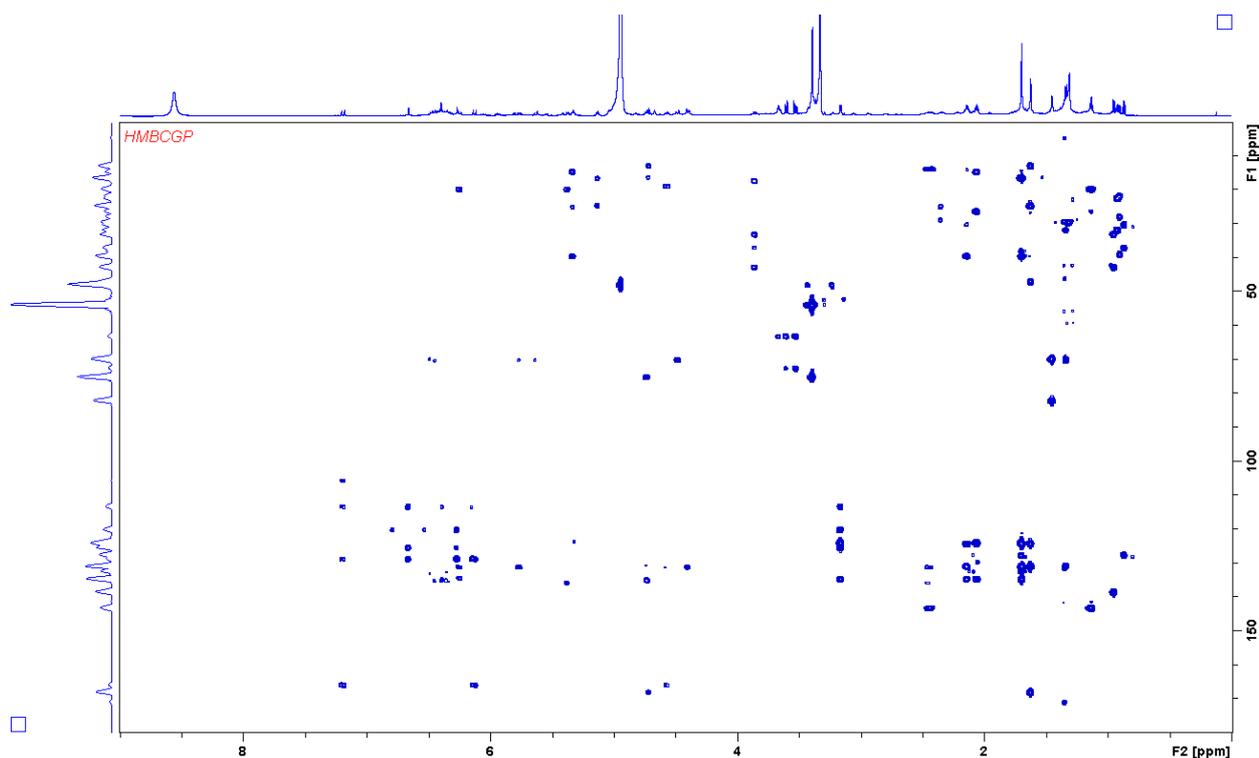
Supplementary figure 37: <sup>1</sup>H spectrum of myxoquaterine 671 variant A acquired in CD<sub>3</sub>OD at 700 MHz



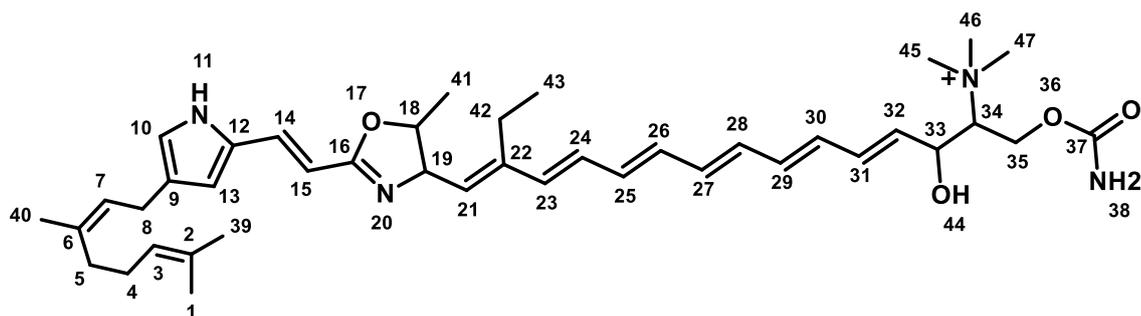
Supplementary figure 38: COSY spectrum of myxoquaterine 671 variant A acquired in  $CD_3OD$  at 700 MHz



Supplementary figure 39: HSQC spectrum of myxoquaterine 671 variant A acquired in  $CD_3OD$  at 700/175 MHz



Supplementary figure 40: HMBC spectrum of myxoquaterine 671 variant A acquired in CD<sub>3</sub>OD at 700/175 MHz



Supplementary figure 41: Structure of Myxoquaterine 671

Supplementary table 6: Spectroscopic data of myxoquaterine 671 variant A in CD<sub>3</sub>OD.

position	$\delta_C^a$	$\delta_H^b$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
1	17.6	1.61 s		2,3
2	131.9			
3	125.2	5.11 t 6.5	4	
4	27.4	2.12	3,5	
5	40.5	2.05	4	2,3,4
6	135.7			
7	124.9	5.32 t 6.6	8	
8	26.1	3.15 d 7.1	7	6,7,9,10
9	126.4			
10	120.8	6.64 s		9,12,13
11	N			
12	129.7			
13	114.1	6.25 s		9,10,12
14	132.5	7.18 d 16.1	15	16
15	106.6	6.11 d 16.8	14	12,16
16	166.7			
17	O			
18	71.0	4.54 m	19,41	

19	83.1	4.38 m	18,21	
20	N			
21	132.1	5.36 d 9.2	19	
22	144.1			
23	136.7	6.23 d 14.7		
24-31	131.7-136.9	6.46-6.43		
32	131.6	5.75 d,d 14.9,7.8	33,31	
33	71.0	4.71 d,d 8.2,8.8)	32,34	
34	76.2	3.64 m	33,35a/b	
35a	61.3	4.48 m	35a, 34	33
35b	61.3	4.39 m	35b, 34	
36	O			
37	n.d.			
38	N			
39	25.7	1.67 s		1,2,5,8
40	15.8	1.67 s		
41	19.9	1.43 d 5.9	18	18,19
42	20.8	2.4 m		
		2.37 m		
43	14.7	1.11 t 7.9	42a/b	22,42
44	OH			
45	54.9	3.37 s		34
46	54.9	3.37 s		34
47	54.9	3.37 s		34

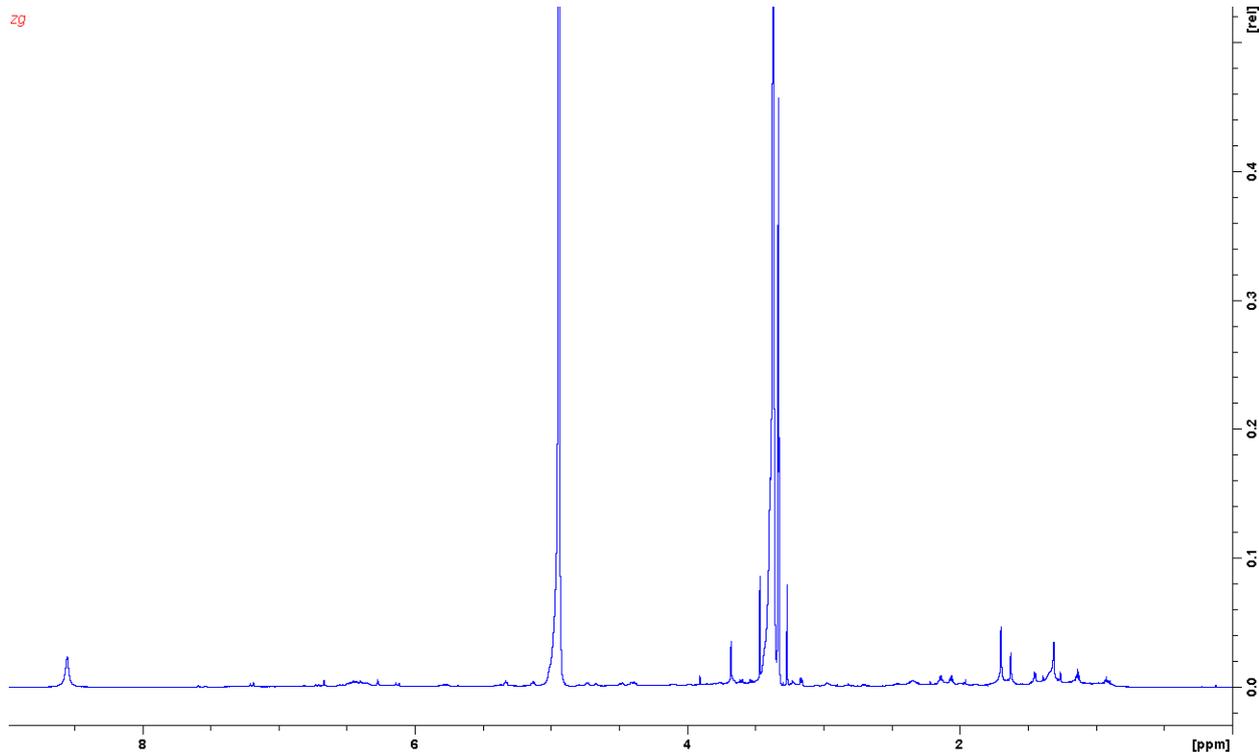
<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

<sup>b</sup> acquired at 700 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

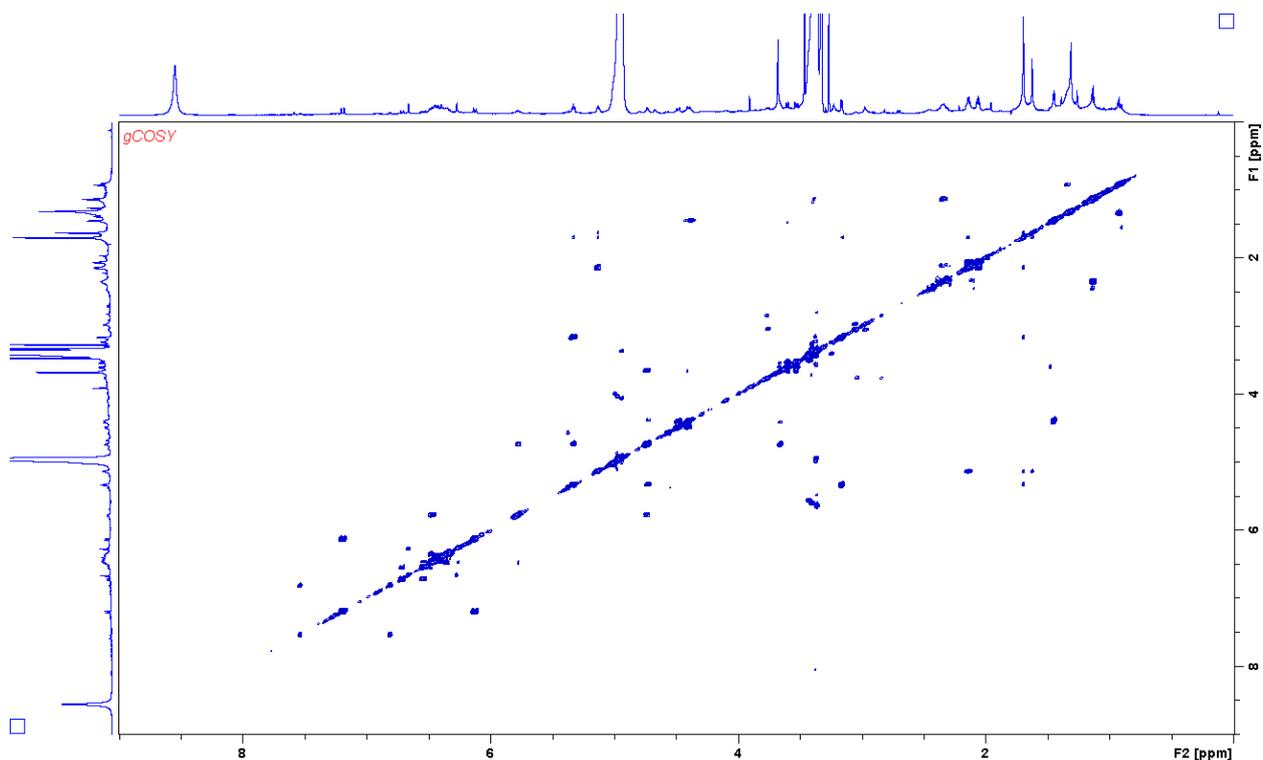
<sup>c</sup> proton showing COSY correlation to indicated proton.

<sup>d</sup> proton showing HMBC correlations to indicated carbons.

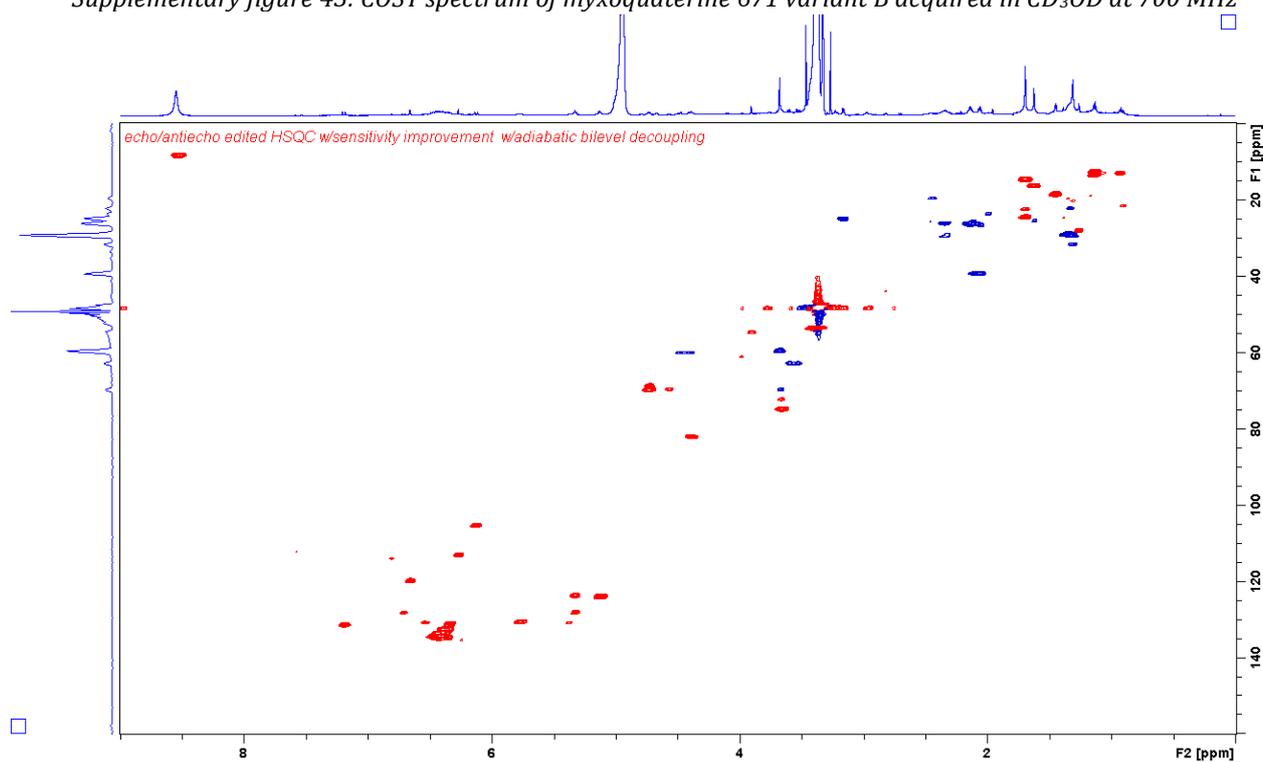
NMR spectra were acquired with a Bruker Ascend 700 NMR spectrometer equipped with a 5mm TCI cryoprobe



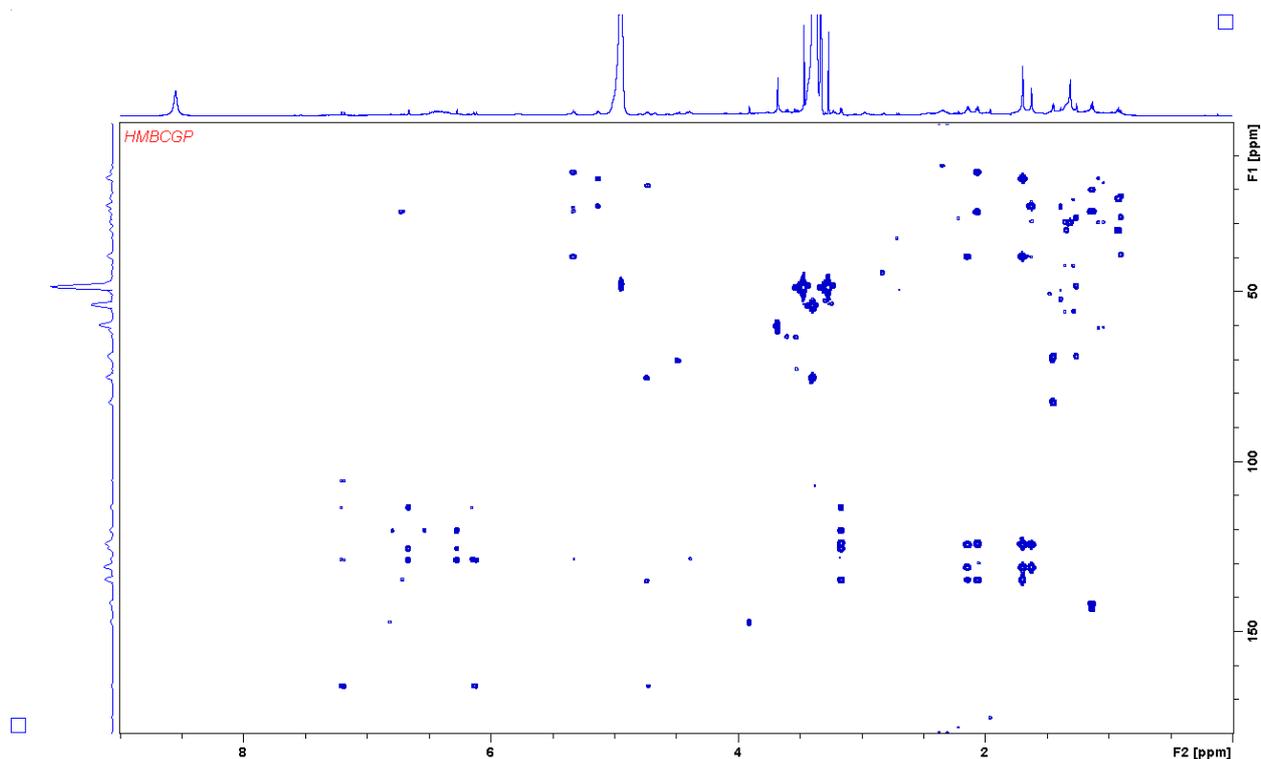
Supplementary figure 42: <sup>1</sup>H spectrum of myxoquaterine 671 variant B acquired in CD<sub>3</sub>OD at 700 MHz



Supplementary figure 43: COSY spectrum of myxoquaterine 671 variant B acquired in CD<sub>3</sub>OD at 700 MHz



Supplementary figure 44: HSQC spectrum of myxoquaterine 671 variant B acquired in CD<sub>3</sub>OD at 700/175 MHz



Supplementary figure 45: HMBC spectrum of myxoquaterine 71 variant B acquired in CD<sub>3</sub>OD at 700/175 MHz

Supplementary table 7: Spectroscopic data of myxoquaterine 671 variant B in CD<sub>3</sub>OD.

position	$\delta_c^a$	$\delta_H^b$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
1	17.6	1.61 s		2,3,39
2	131.9			
3	125.2	5.11 t 6.9	4	
4	27.4	2.12	3,5	2,3,5,6
5	40.6	2.05	4	3,4,6,7,40
6	135.6			
7	124.9	5.31 m	8	
8	26.1	3.14 d 7.3	7	6,7,9,10,13
9	126.5			
10	121.2	6.65 s		9,12,13
11	N			
12	129.7			
13	114.3	6.25 s		9,10,12
14	132.6	7.18 d 15.7	15	12,13,15,16
15	106.3	6.11 d 16.1	14	12,16
16	166.9			
17	O			
18	69.8	4.71 m	19,41	16
19	83.4	4.37 m	18,21	21
20	N			
21	129.3	5.31 d 9.2	19	
22	142.3			
23	129.4	6.70 d 15.1	24	21
24	132.0	6.53 d,d 14.9, 9.4	23	
25-31	131.5-136.9	6.40-6.45		
32	131.7	5.75 d,d 15.2,7.6	33	
33	71.0	4.71 m	32,34	
34	76.0	3.64 m	33,35a/b	33

35a	61.4	4.47 d,d 2.0, 14.1	35a, 34	33,34
35b		4.38 m	35b, 34	
36	O			
37	n.d.			
38	N			
39	25.7	1.67 *		
40	15.8	1.67 *		
41	19.6	1.43 d 6.0	18	18,19
42	27.3	2.33 m	43	22, 41
43	13.8	1.11 t 7.6	42	22,42
44	OH			
45	54.8	3.37 s		34
46	54.8	3.37 s		34
47	54.8	3.37 s		34

<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

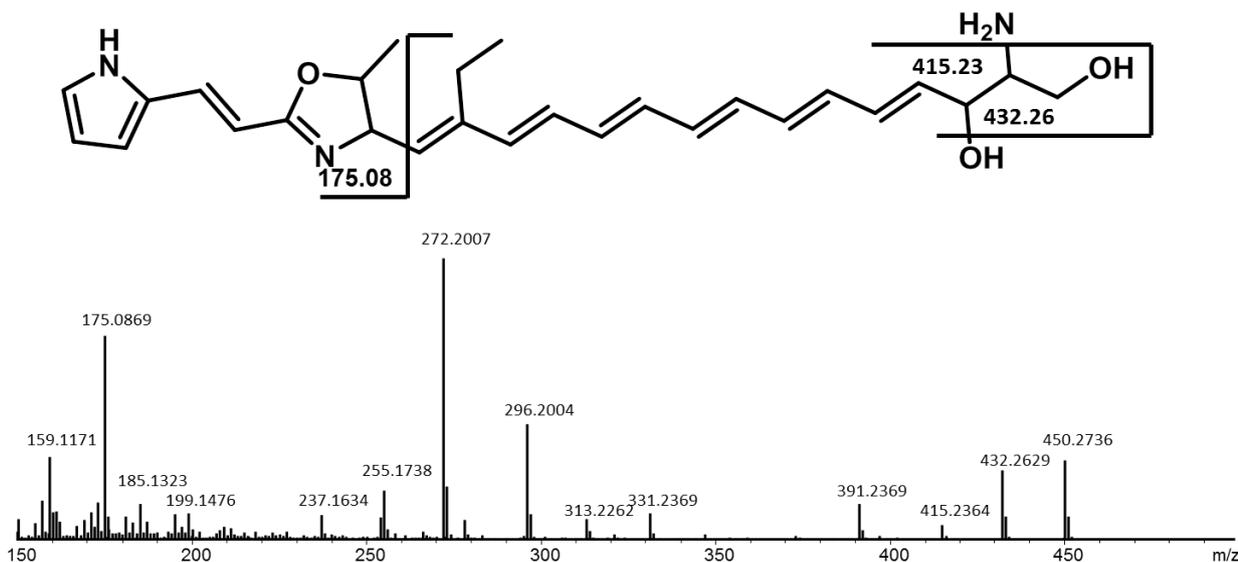
<sup>b</sup> acquired at 700 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

<sup>c</sup> proton showing COSY correlation to indicated proton.

<sup>d</sup> proton showing HMBC correlations to indicated carbons.

### Analytical data myxoquaterine 450

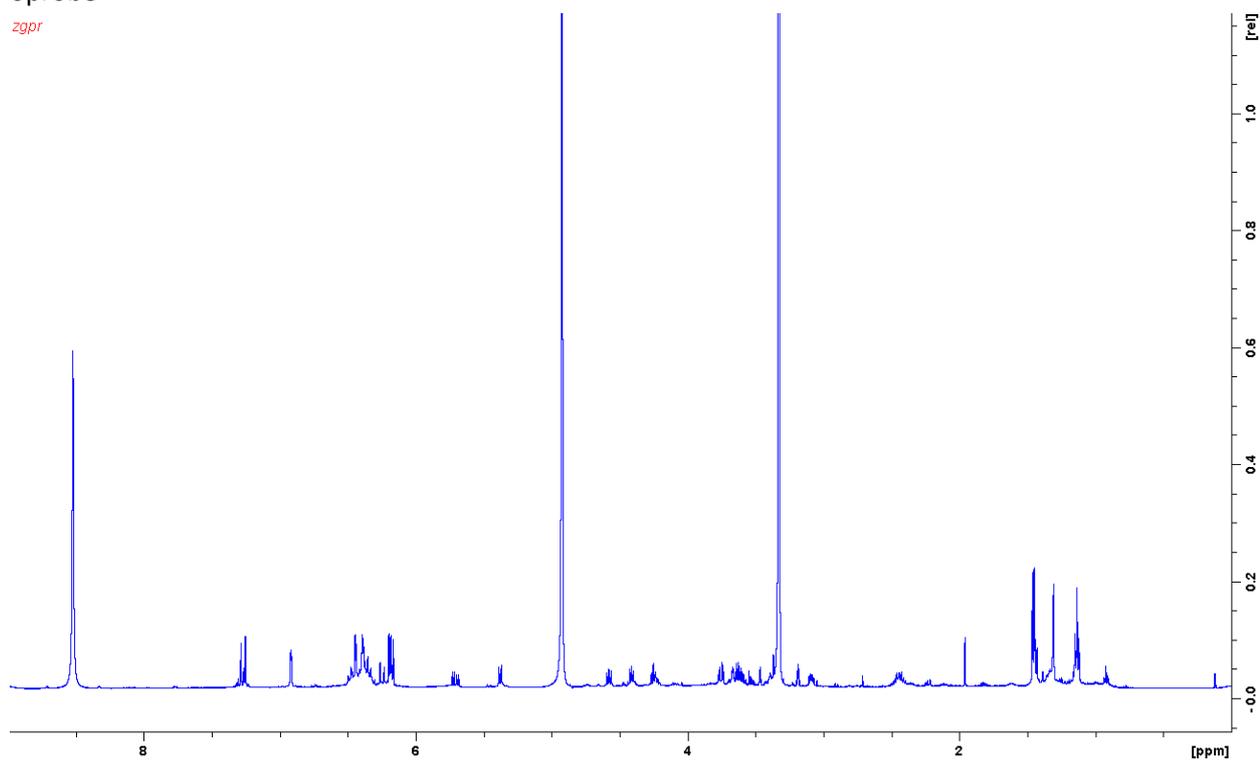
Myxoquaterine 450 was isolated brown yellowish amorphous solid that turned into a powder after lyophilization. *hrMass*: 450.27642 [M]<sup>+</sup>; 216.63939 [M-H<sub>2</sub>O+H]<sup>2+</sup>; specific optical rotation:  $[\alpha]_D^{20}$  53.3 °g<sup>-1</sup>ml dm<sup>-1</sup>; UV<sub>max</sub> 190, 326, 345, 358, 379 nm (UV data obtained by a Jasco J-1500 CD spectrometer in MeOH)



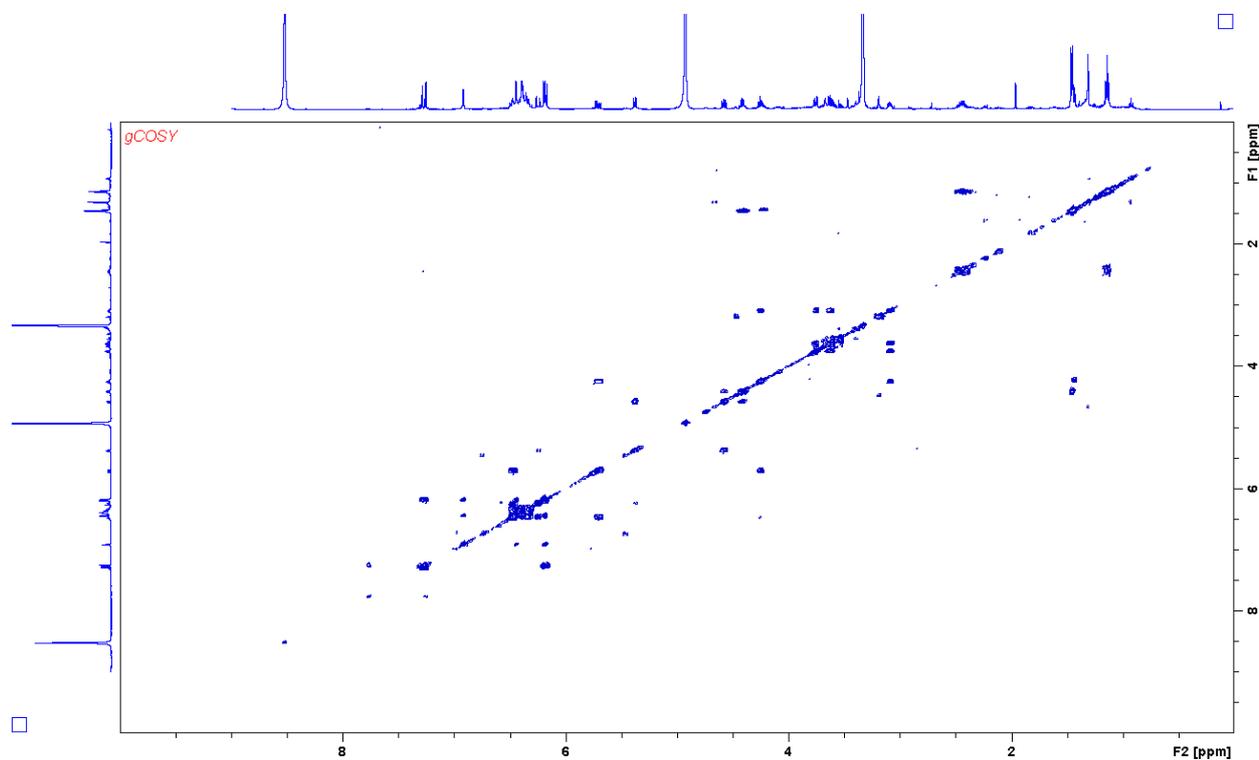
Supplementary figure 46: *hrMS*<sup>2</sup> spectrum and observed fragments of myxoquaterine 450. *MS*<sup>2</sup> spectrum is extracted from LC-*hrMS* measurement of an extract of MSr 11954.

NMR spectra were acquired with a Bruker UltraShield 500 NMR spectrometer equipped with a 5mm TCI cryoprobe

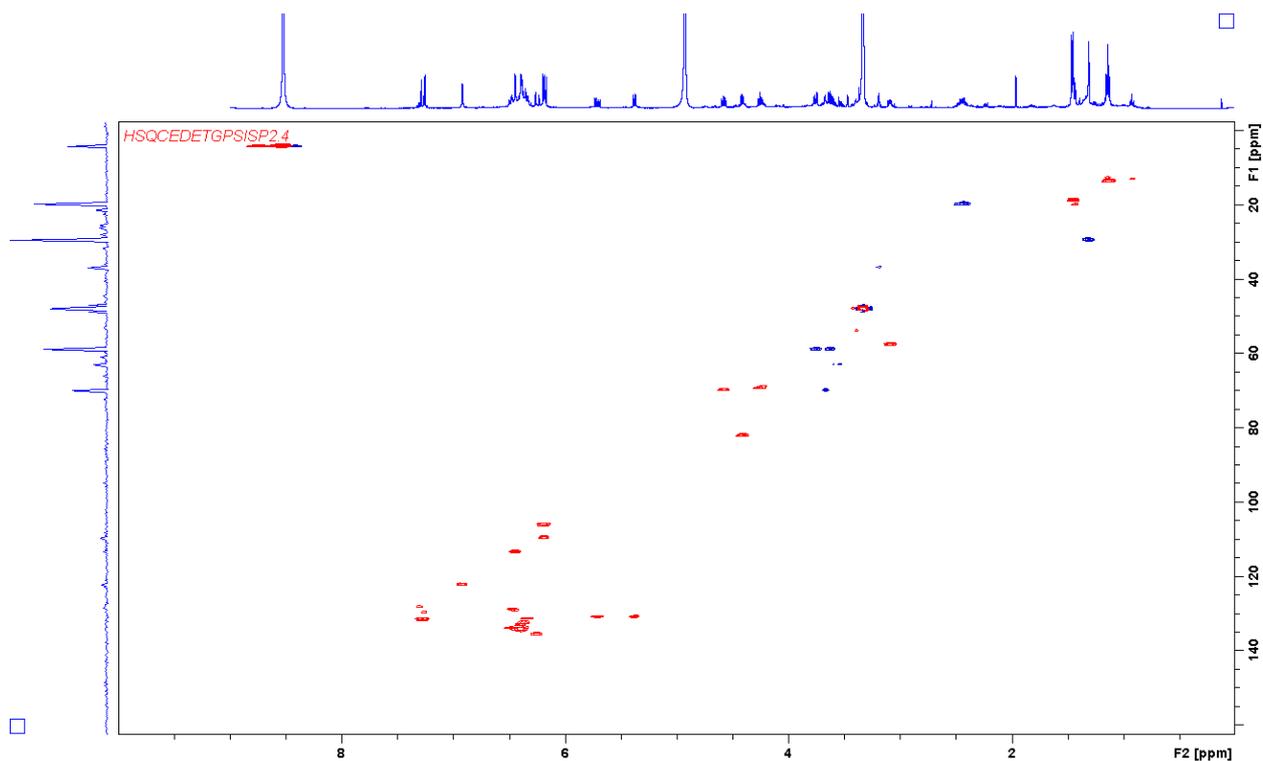
zgpr



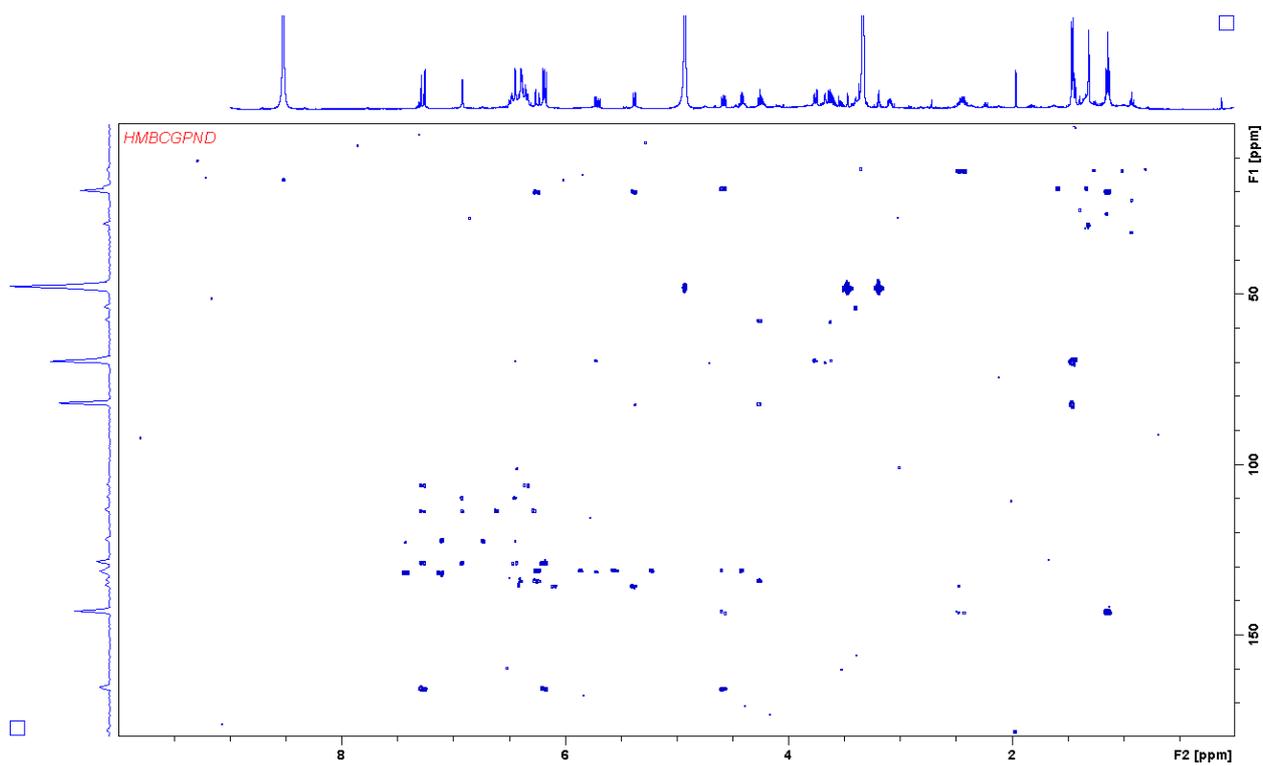
Supplementary figure 47:  $^1\text{H}$  spectrum of myxoquaterin 450 acquired in  $\text{CD}_3\text{OD}$  at 500 MHz



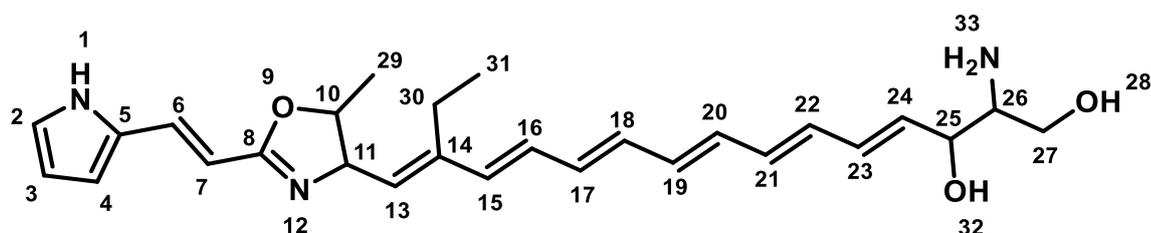
Supplementary figure 48: COSY spectrum of myxoquaterine 450 acquired in  $\text{CD}_3\text{OD}$  at 500 MHz



Supplementary figure 49: HSQC spectrum of myxoquaterine 450 acquired in CD<sub>3</sub>OD at 500/125 MHz



Supplementary figure 50: HMBC spectrum of myxoquaterine 450 acquired in CD<sub>3</sub>OD at 500/125 MHz



Supplementary figure 51: structure of myxoquaterine 450

Supplementary table 8: Spectroscopic data of myxoquaterine 450 in CD<sub>3</sub>OD.

position	$\delta_c^a$	$\delta_H^b$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
1	N			
2	123.2	6.89 d,d (2.3,1.2)	3,4	3,4,5
3	110.6	6.16 d,d (2.6,3.8)	2,4	5
4	114.5	6.42 d,d (3.9,1.7)	3	2,3,5
5	129.6			
6	132.5	7.25 d (15.8)	6	4,5,7,8
7	107.2	6.16 d (15.9)	7	2,4,6,8
8	166.7			
9	O			
10	83.2	4.39 d,q (6.4,6.6)	11,29	6,8,11
11	70.9	4.56 d,d (9.7,7.2)	10,13	10,13,14,29
12	N			
13	132.0	5.36 d (9.7)	11	10,11,15,30
14	144.2			
15	136.5	6.23 d (15.3)	16	13,14,16
16	135.1	6.44 m		
17-22	131.3-134.5	6.29-6.54		
23	134.8	6.44 m		
24	132.2	5.69 d,d (15.3,7.2)	23,25	25,26
25	70.5	4.22 m	24,26	26,27
26	58.6	3.04 m	25	25
27a	60.1	3.59 m	27b, 26	
27b		3.72 m	27a, 26	
28	OH			
29	19.8	1.44 d (6.2)		8,10
30a	20.8	2.40 m	31	13,14,15,31
30b		2.44 m		
31	14.7	1.12 t (7.6)	30a/b	14,30

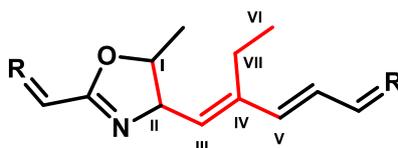
<sup>a</sup> acquired at 125 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

<sup>b</sup> acquired at 500 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

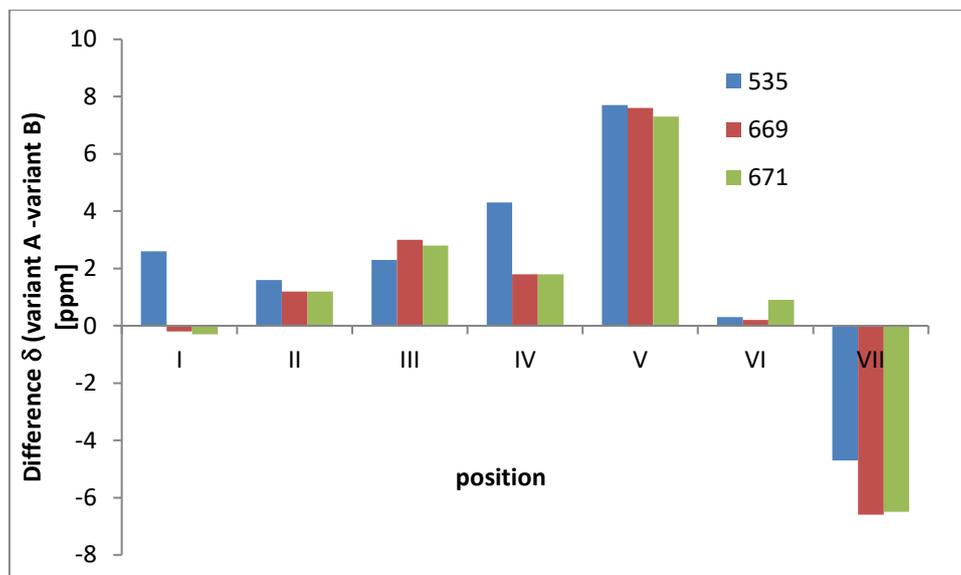
<sup>c</sup> proton showing COSY correlation to indicated proton.

<sup>d</sup> proton showing HMBC correlations to indicated carbons.

### Differences between myxoquaterin variant A and B

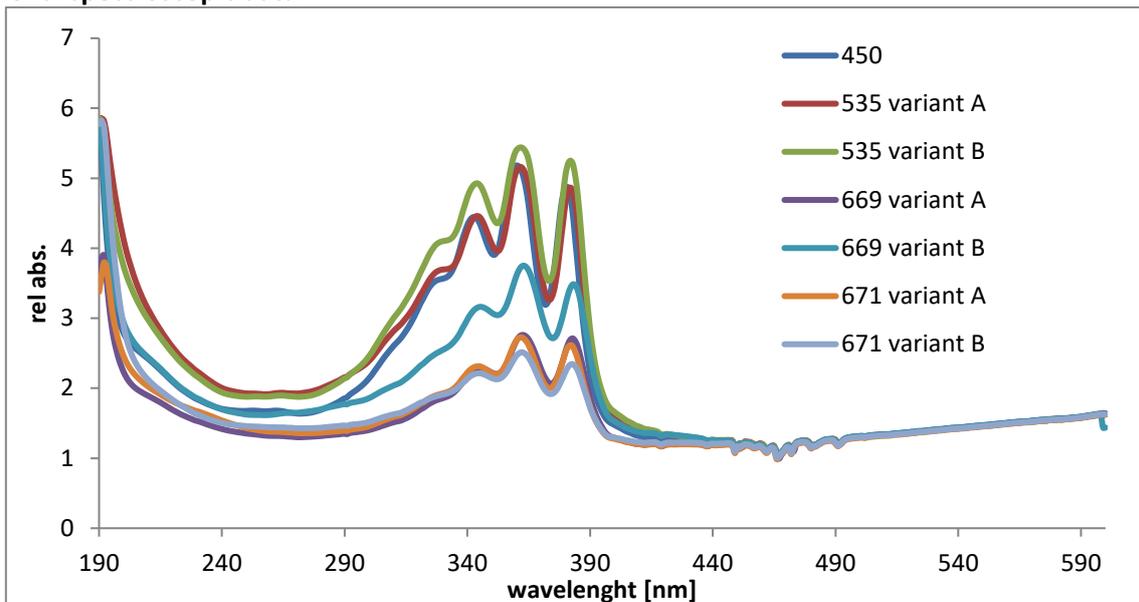


Supplementary figure 52: Structural moiety in the myxoquaterine molecules that show variations in the chemical shifts.

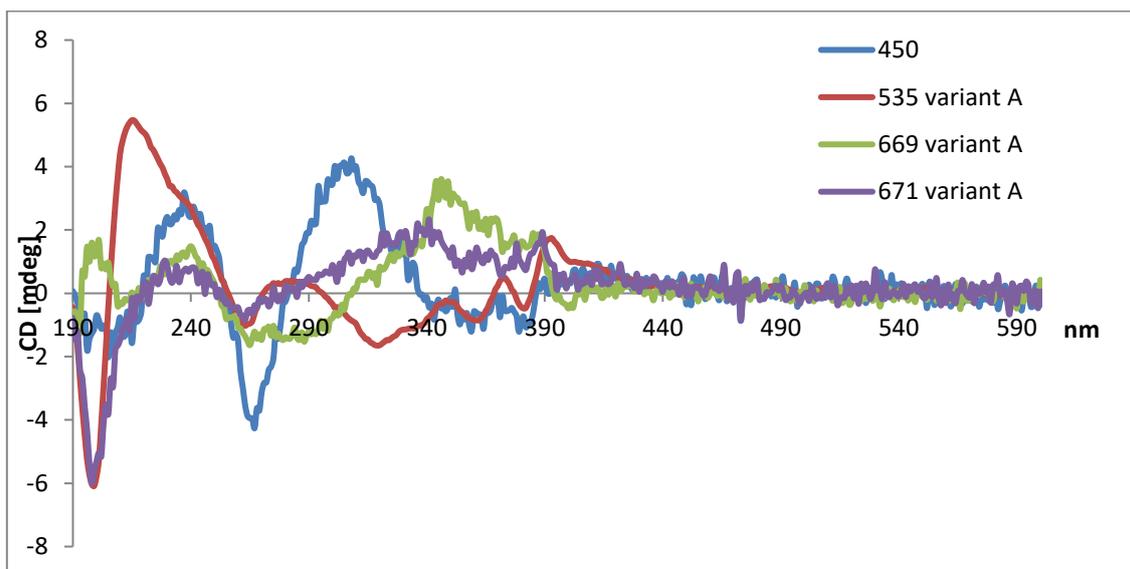


Supplementary figure 53: Shift differences between myxoquaterine variants A and B.

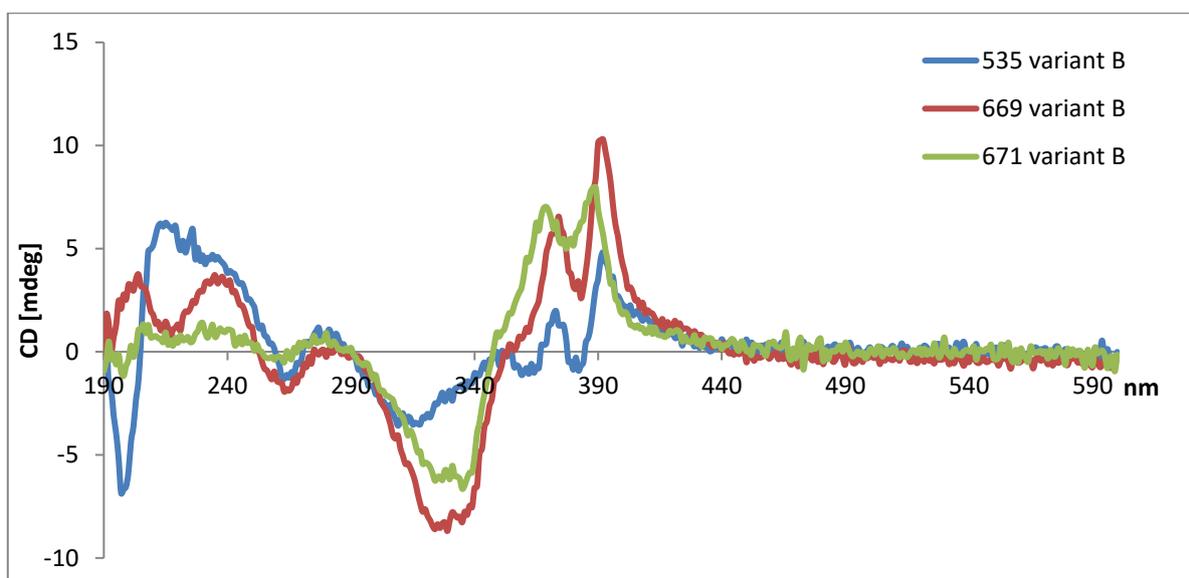
**Additional spectroscopic data:**



Supplementary figure 54: UV spectra of isolated myxoquaterine variants acquired with a Jasco J-1500 CD-spectrometer in MeOH



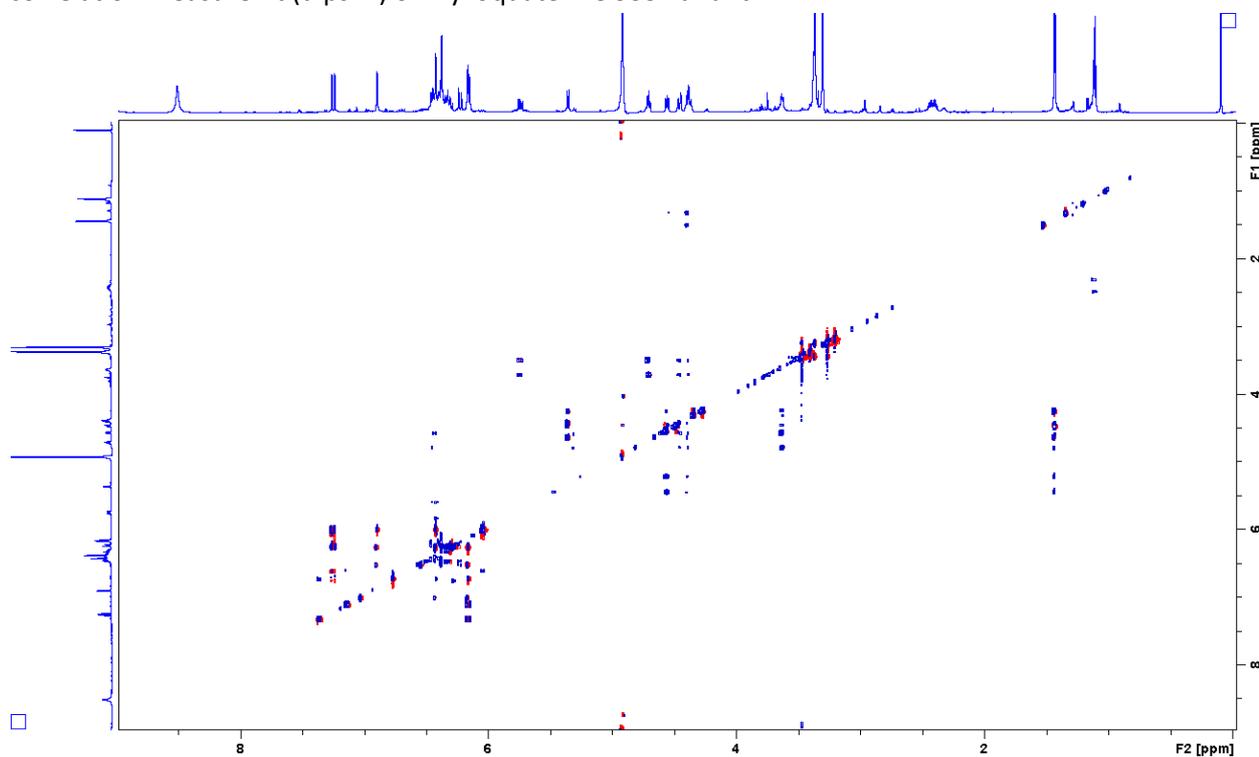
Supplementary figure 55: CD-spectra of myxoquaterine A-variants acquired with a Jasco J-1500 CD-spectrometer in MeOH



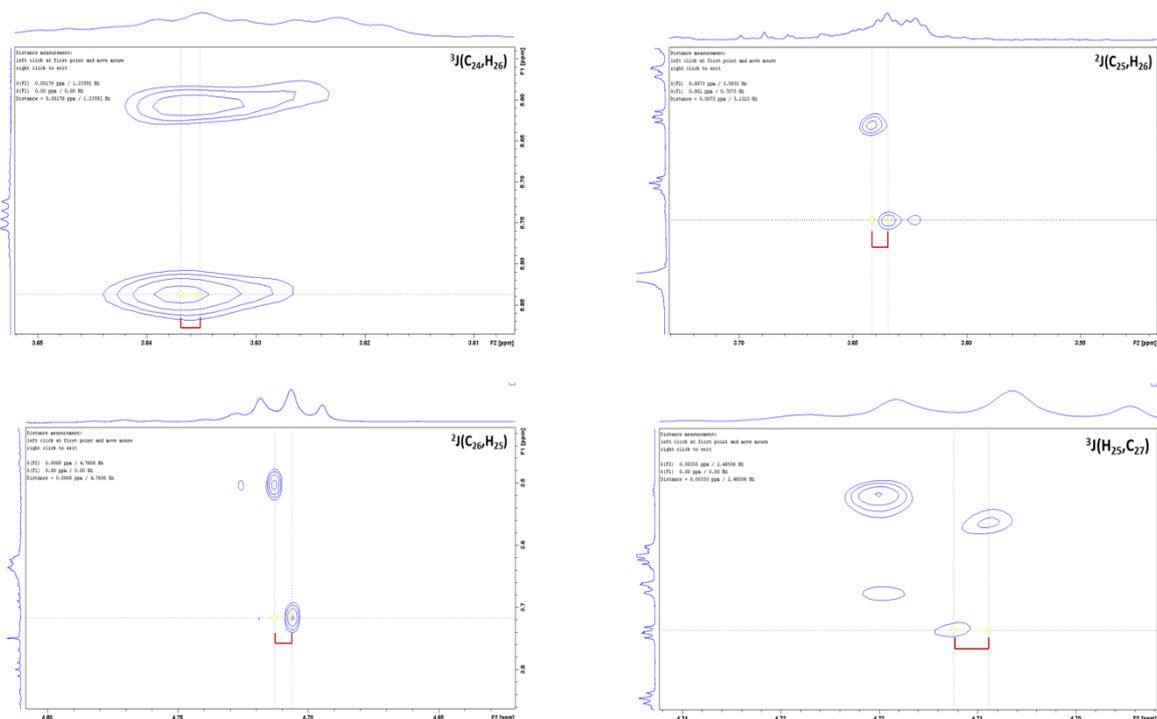
Supplementary figure 56: CD-spectra of myxoquaterine B-variants acquired with a Jasco J-1500 CD-spectrometer in MeOH

#### Determination of configuration of the Ammonium and hydroxy group

The relative configuration of the ammonium group and the hydroxy group was determined by analysis of coupling constants in myxoquaterine 535 variant A. The H,H coupling constants were determined from the  $^1\text{H}$  spectrum of myxoquaterine 535 variant A. Hetero coupling constants were determined by a hetero correlation measurement (dipsi 2) of myxoquaterine 535 variant A.



Supplementary figure 57: dipsi2 (Hetloc) spectrum of myxoquaterine 535 obtained in  $\text{CD}_3\text{OD}$  at 700 MHz



Supplementary figure 58: Determination of respective coupling constants from dipsi2 (Hetloc) spectrum

### Mosher esterification

Determination of the absolute configuration of the hydroxy was tried by Mosher esterification:



[M]<sup>+</sup> 751.36  
[M+1H]<sup>2+</sup> 376.18

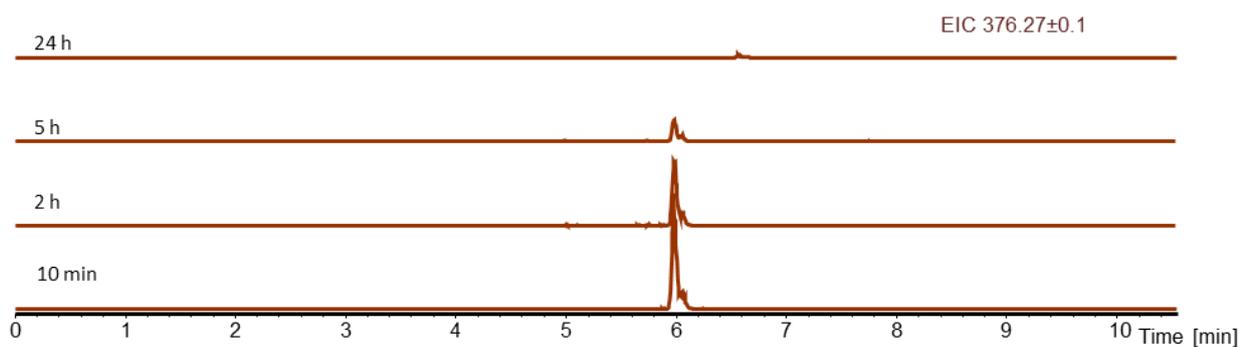
Supplementary figure 59: Expected product of Mosher esterification of Myxoquaterine variant A.

Different reaction conditions were tested: The reaction was evaluated based on appearance of expected product mass by LC-MS.

Supplementary table 9: Used conditions and outcome of Mosher esterification

reaction	results
1	No product formation
2	Minor product formation
3	Product main peak in reaction mixture

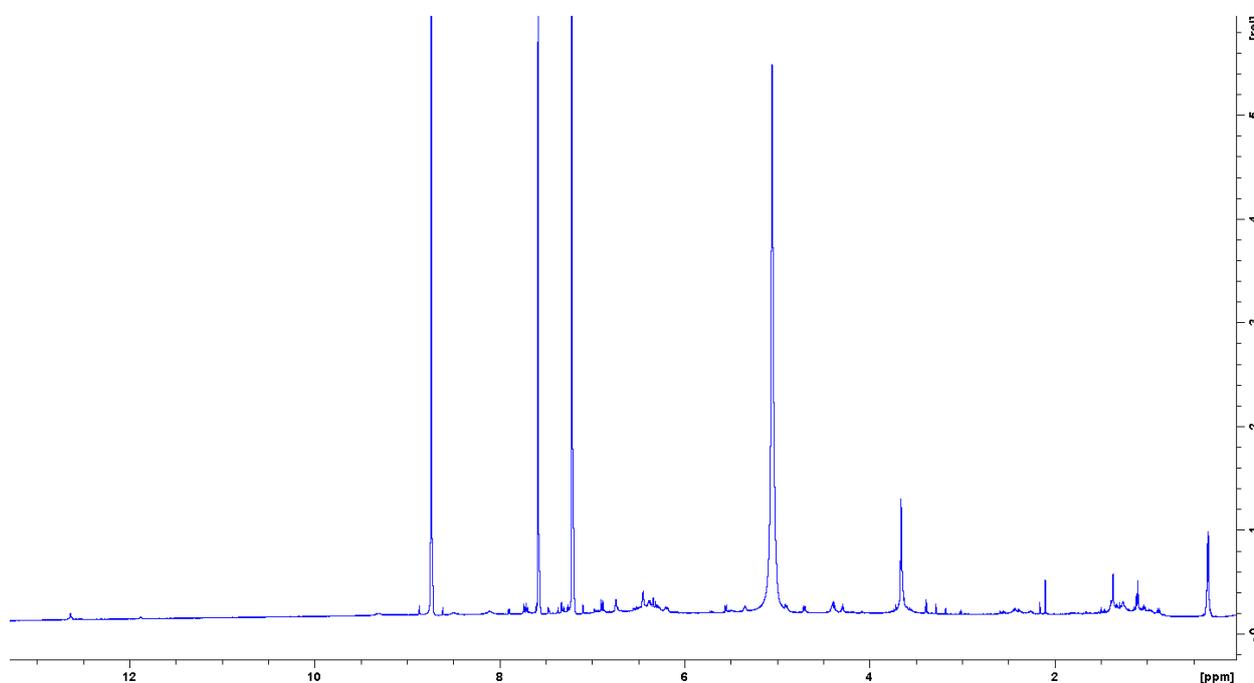
Using Pyridine as solvent and base and  $\alpha$ -methoxy- $\alpha$ -trifluoromethylphenylacetic acid (MTPA)-Chloride as substrate resulted in a good turnover within 10 minutes. The reaction product was not stable and degraded within a few hours.



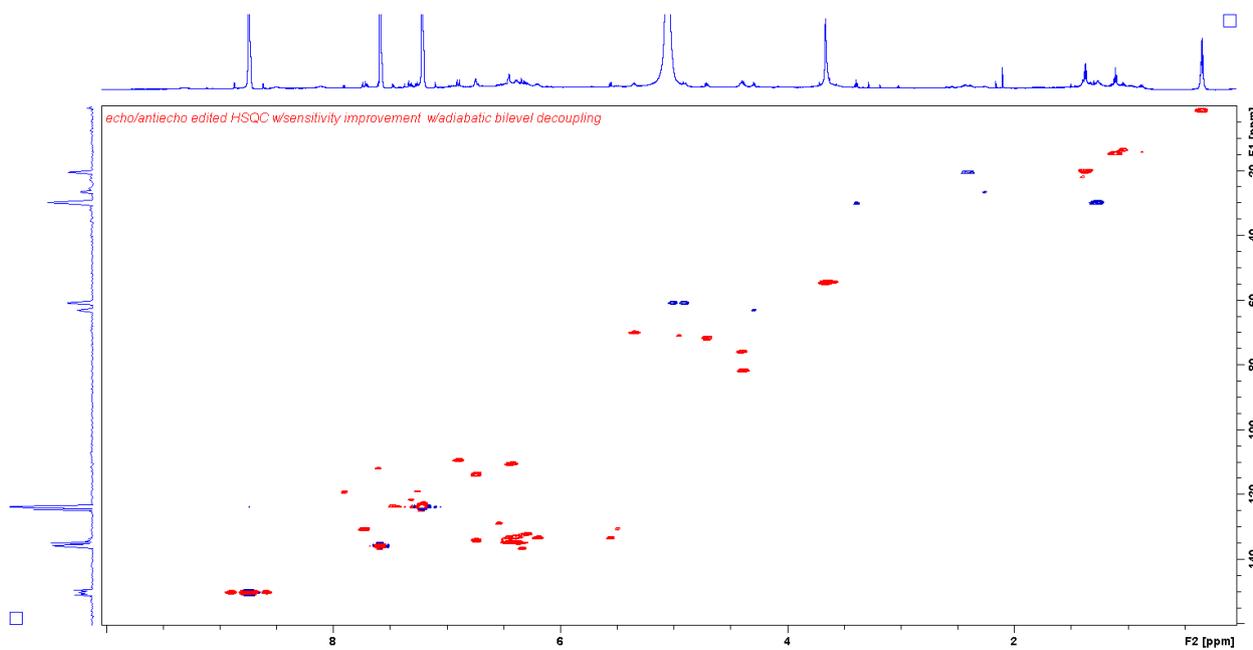
Supplementary figure 60: LC-MS chromatogram showing degradation of the observed product displayed as EIC 376.27 over time.

Therefore, it was tried to acquire immediately NMR datasets of the reaction mixtures in pyridine  $d_5$ . For both MTPA esters a HSQC dataset could be measured. Using this dataset, most chemical shifts could be assigned and compared to myxoquartarine 535 variant A.

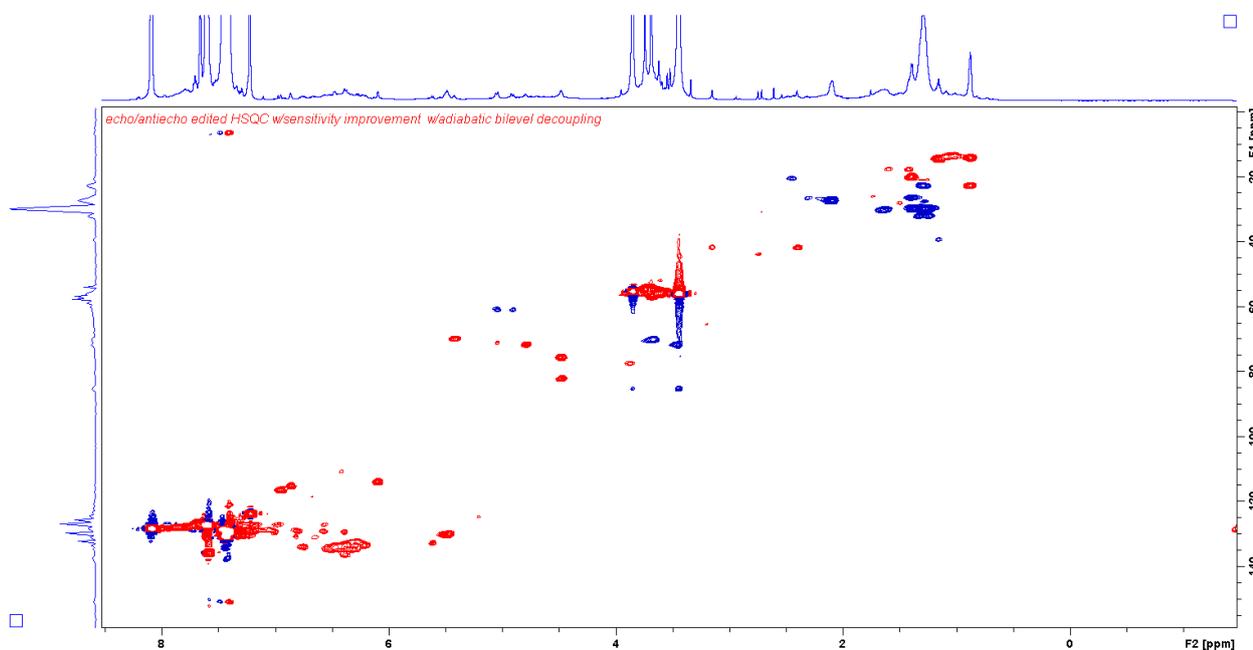
NMR spectra were acquired with a Bruker Ascend 700 NMR spectrometer equipped with a 5mm TCI cryoprobe



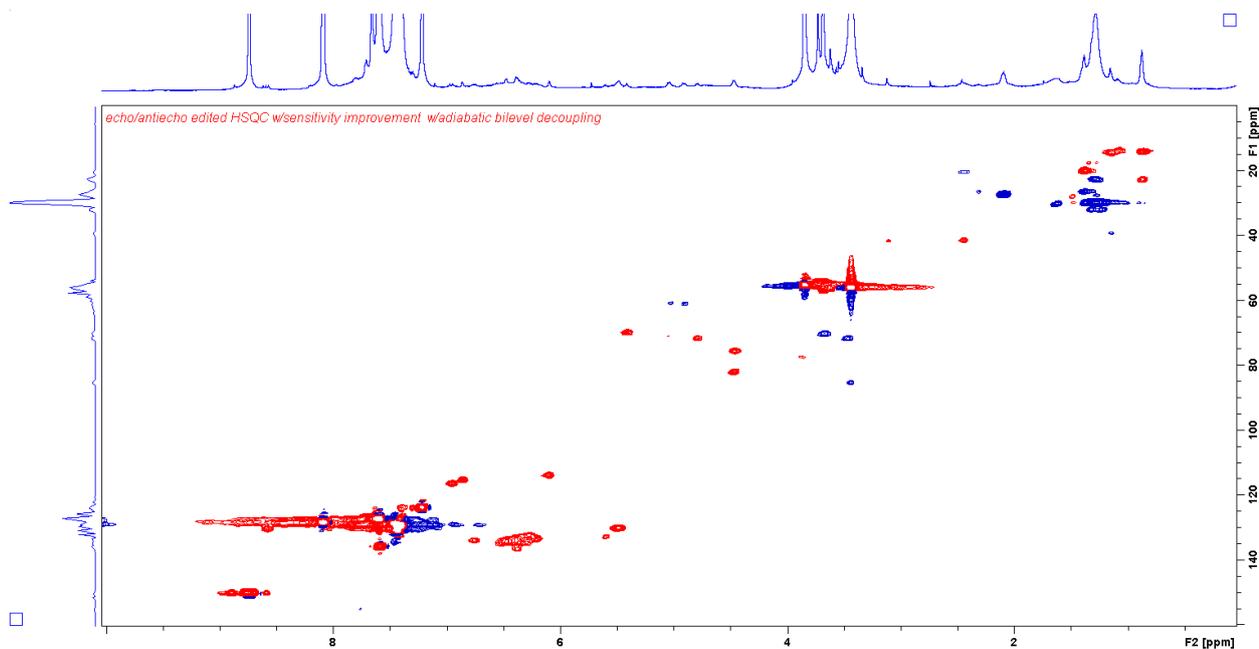
Supplementary figure 61:  $^1\text{H}$  spectrum of myxoquaterine 535 variant A acquired in pyridine  $d_5$  at 700 MHz



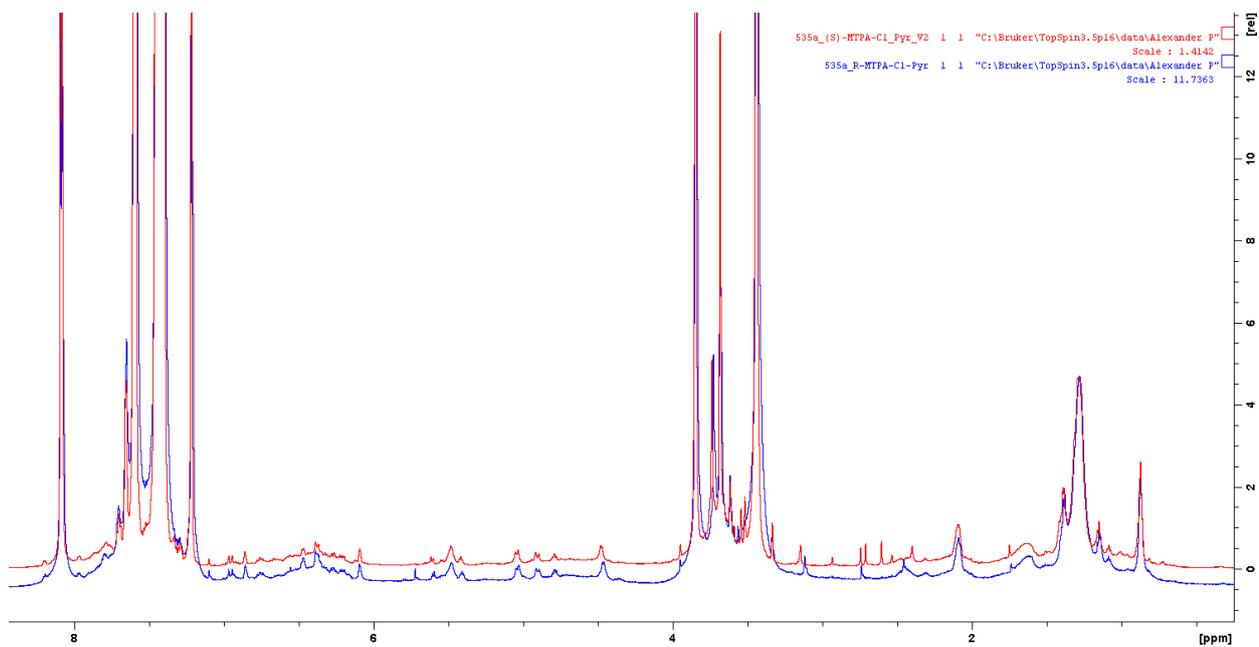
Supplementary figure 62: HSQC spectrum of myxoquaterine 535 variant A acquired in pyridine  $d_5$  at 700/175 MHz



Supplementary figure 63: HSQC of reaction mixture of myxoquaterine variant A derivatized with S-MTPA-Cl acquired in pyridine  $d_5$  at 700/175 MHz



Supplementary figure 64: HSQC spectrum of reaction mixture of myxoquaterine variant A derivatized with R-MTPA-Cl acquired in pyridine  $d_5$  at 700/175 MHz



Supplementary figure 65: Overlapped  $^1H$  spectra of reaction mixtures from derivatization with S-MTPA (displayed in red) and R-MTPA (displayed blue) both spectra acquired in pyridine  $d_5$  at 700 MHz



Supplementary table 10: spectroscopic data of myxoquaterin 535 variant A in pyr *d*<sub>6</sub> and myxoquaterin variant A *R/S*-MTPA reaction mixtures in pyr *d*<sub>6</sub>.

Myxoquaterine 535 variant A			+ (R)-MTPA-Cl		+ (S)-MTPA-Cl	
position	$\delta^{13}\text{C}^a$	$\delta^1\text{H}^b$	$\delta^{13}\text{C}^a$	$\delta^1\text{H}^b$	$\delta^{13}\text{C}^a$	$\delta^1\text{H}^b$
1						
2	n.d.	7.21	n.d.	n.d.	n.d.	n.d.
3	110.7	6.44	114.0	6.09	114.0	6.09
4	113.8	6.74	115.3	6.86	115.3	6.86
5						
6	130.9	7.72	n.d.	n.d.	n.d.	n.d.
7	109.3	6.89	116.5	6.96	116.5	6.95
8	164.0		n.d.		n.d.	
9						
10	81.9	4.38	82.2	4.47	82.3	4.48
11	71.7	4.71	71.8	4.79	71.9	4.79
12						
13	133.5	5.55	132.9	5.60	132.9	5.61
14	n.d.					
33	14.6	1.11	14.5	1.14	14.6	1.16
15-22	131.2-137.8	6.16-6.57	131.2-137.8	6.16-6.57	131.2-137.8	6.16-6.57
23	n.d.	6.73	n.d.	n.d.	n.d.	6.75
24	133.4	6.19	133.5	6.21	132.8	6.21
25	70.1	5.34	70.0	5.41	70.0	5.42
26	75.9	4.40	75.8	4.46	75.8	4.47
27a	60.9	5.01	61.1	5.02	61.05	5.04
27a		4.91		4.90		4.91
29	n.d.		n.d.		n.d.	
31	20.1	1.37	20.1	1.38	20.1	1.39
32a	20.5	2.44	20.5	2.48	20.6	2.48
32b		2.38		2.41		2.43
33	14.6	1.11	14.5	1.14	14.6	1.16
36-38	54.6	3.66	56.8	3.67	56.9	3.68

<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal pyridine *d*<sub>5</sub> at  $\delta$  123.87 ppm.

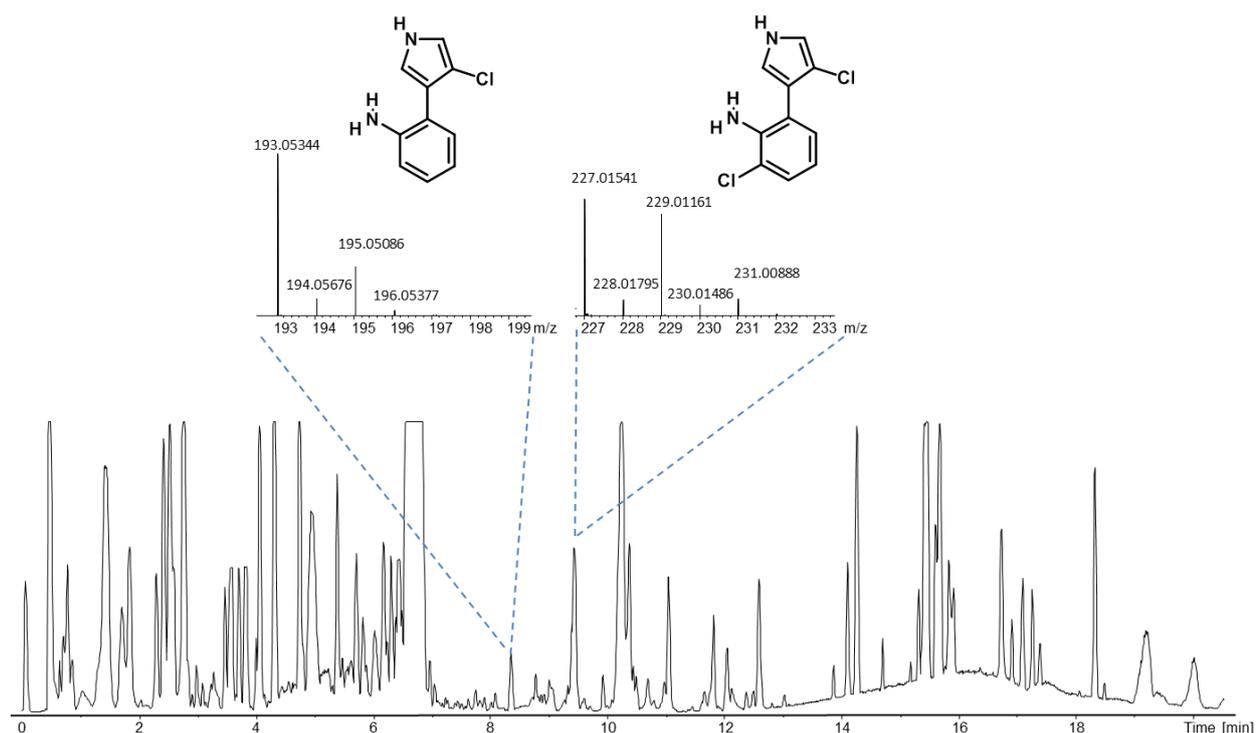
<sup>b</sup> acquired at 700 MHz, referenced to solvent signal pyridine *d*<sub>5</sub> at  $\delta$  7.22 ppm.

Comparison of the chemical shifts showed that major changes can be found in the pyrrole ring suggesting that the MTPA-chloride attaches to the pyrrole and not to the alcohol at position 28.

## Metabolome of Strain MSr 11954

The metabolome of MSr 11954 was screened for known metabolites :

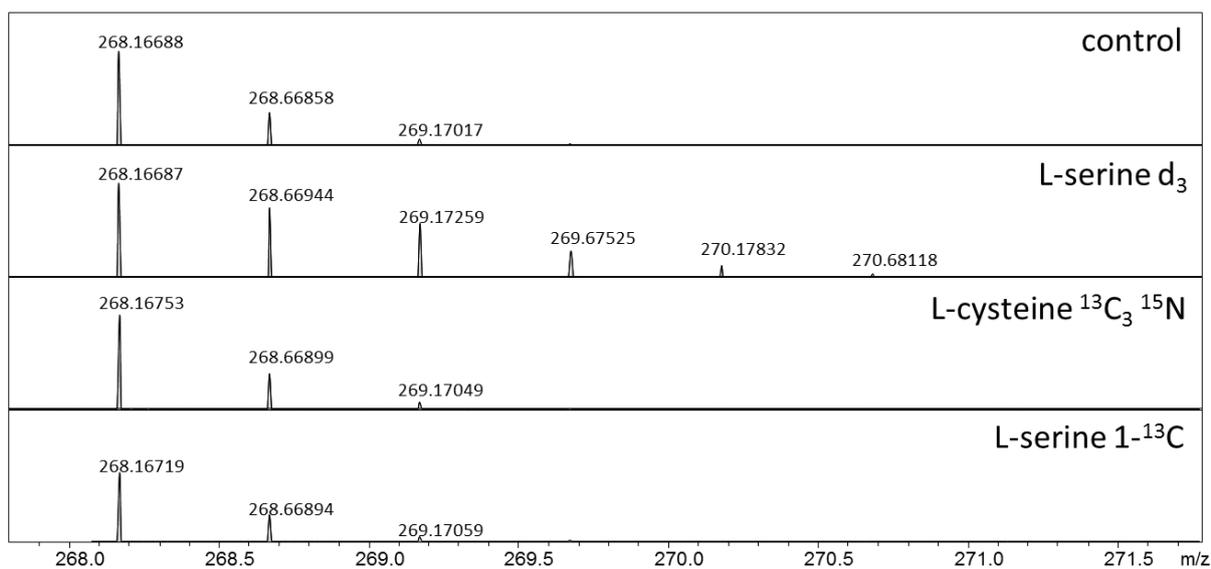
### Amino-pyrrolnitrin variants



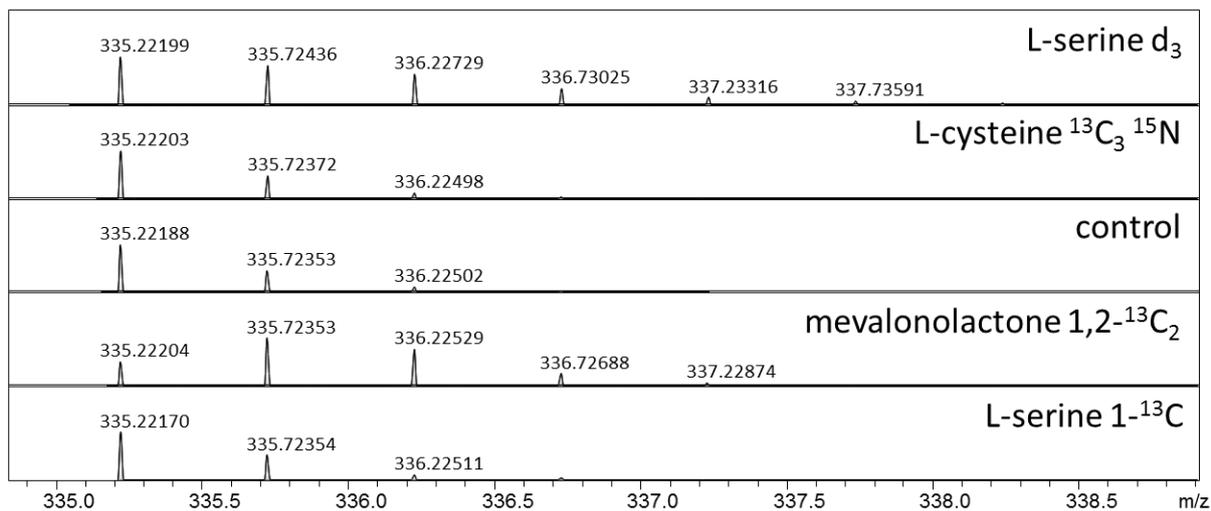
*Supplementary figure 66: LC-hrMS chromatogram of a crude extract of MSr 11954. High resolution Mass spectra of the two known metabolites are displayed on top of their corresponding peaks.*

## Biosynthesis

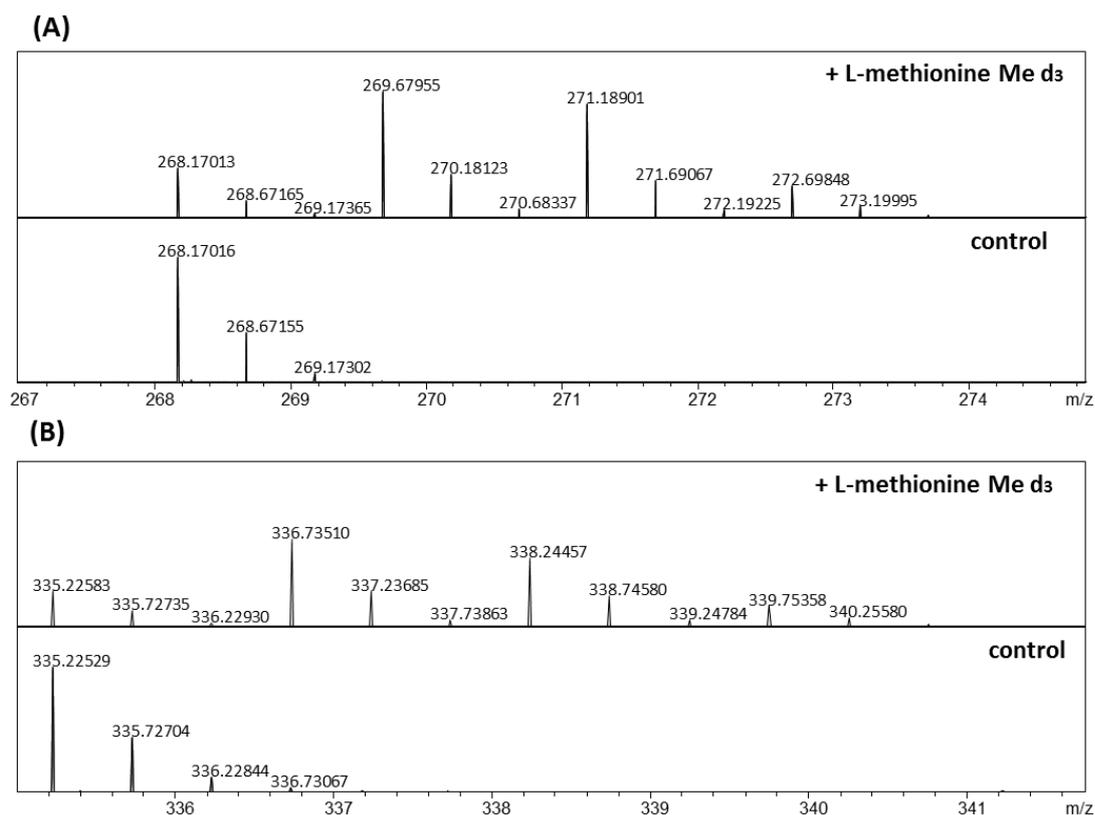
In order to investigate the biosynthesis, various isotope labelled biosynthetic precursors were supplemented to cultures of MSr 11954. Crude extracts of these cultures were measured by LC-hrMS and the isotope patterns of the identified myxoquaterines were investigated for changes compared to untreated control cultures.



Supplementary figure 67: Isotope pattern of myxoquaterine 535 variant A showing double charged mass  $[M+H]^{2+}$  extracted from LC-hrMS from the control culture and cultures supplemented L-serine d<sub>3</sub>, L-cysteine <sup>13</sup>C<sub>3</sub><sup>15</sup>N and L-serine 1-<sup>13</sup>C.



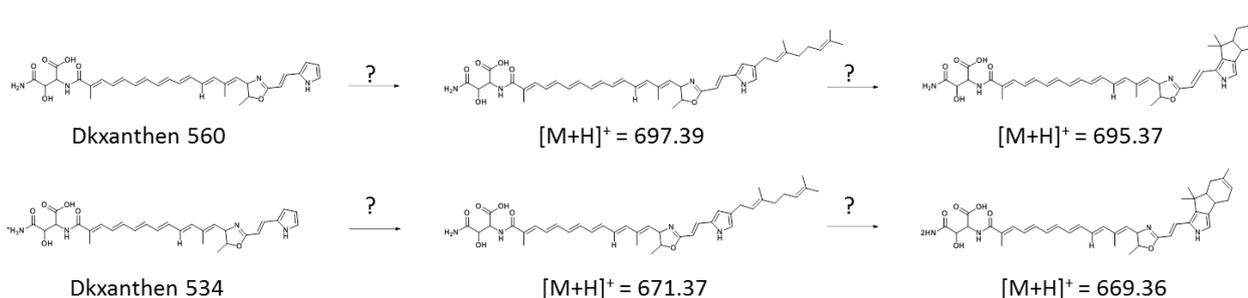
Supplementary figure 68: Isotope pattern of myxoquaterine 669 variant A showing double charged mass  $[M+H]^{2+}$  extracted from LC-hrMS from the control culture and cultures supplemented L-serine d<sub>3</sub>, L-cysteine <sup>13</sup>C<sub>3</sub><sup>15</sup>N, mevalonolactone 1,2-<sup>13</sup>C<sub>2</sub> and L-serine 1-<sup>13</sup>C.



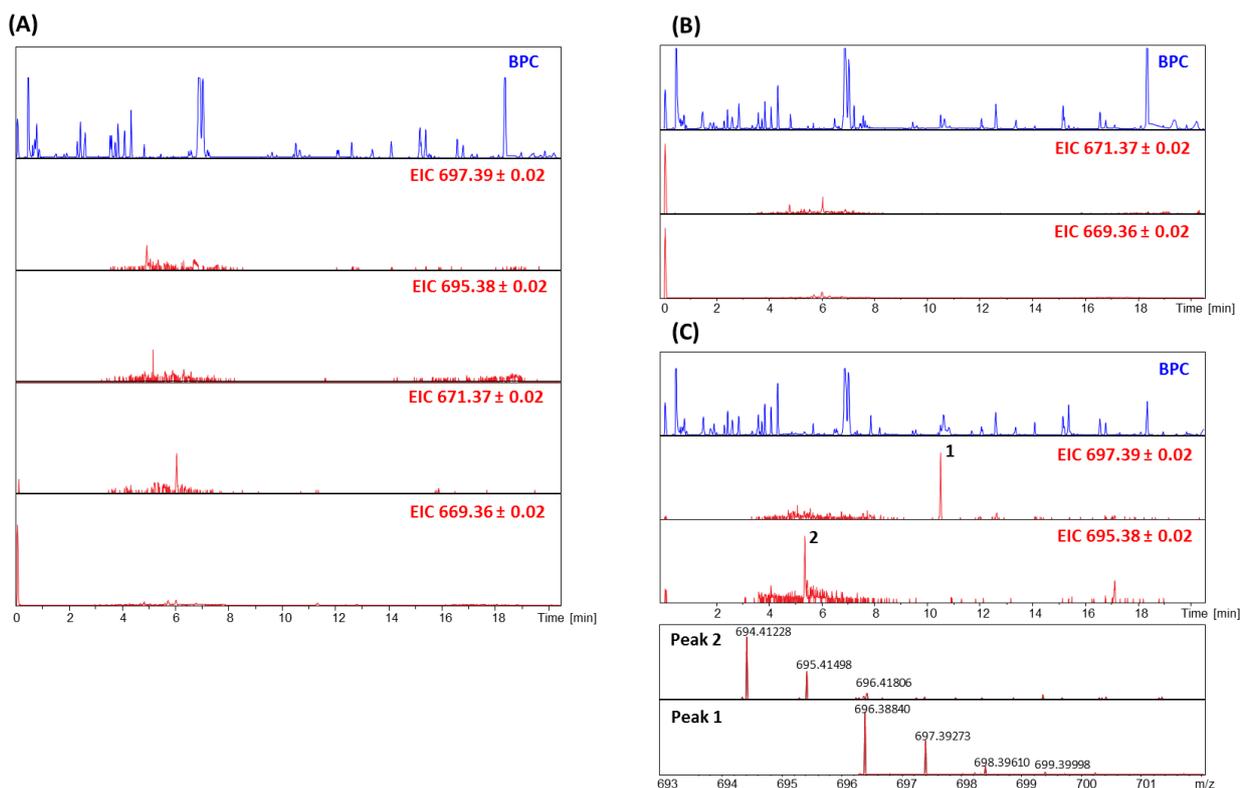
Supplementary figure 69: (A) Isotope pattern of myxoquaterine 535 variant A showing double charged mass  $[M+H]^{2+}$  extracted from LC-hrMS from the control culture and cultures supplemented L-methionine  $d_3$  (B) Isotope pattern of myxoquaterine 669 variant A showing double charged mass  $[M+H]^{2+}$  extracted from LC-hrMS from the control culture and cultures supplemented L-methionine  $d_3$

### Prenylation experiments

To investigate the possible options and achieve a general overview of the prenylation capabilities of MSr 11954 a feeding experiments were conducted: The Myxoquaterines have a high similarity with compounds from DKxanthen family<sup>[2]</sup>. Therefore, prenylation of DKxanthenes by the same aromatic prenyl-transferase was considered to be possible since the pyrrole and methyloxazoline moiety is identical. A prenylation of DKxanthenes can be considered as prove for a post assembly prenylation. DKxanthen 560 and DKxanthen 534 were added to cultures of MSr 11954.



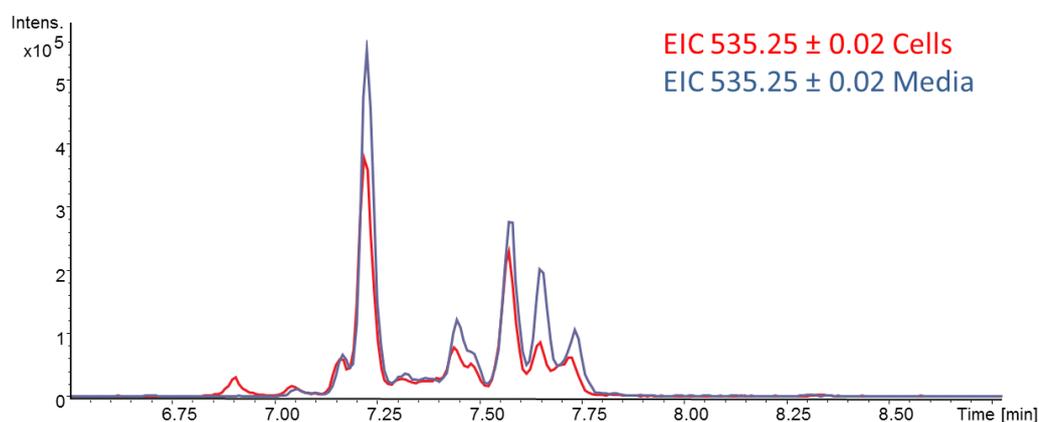
Supplementary figure 70: Expected products in case of prenylation DKxanthen 560 and DKxanthen 534



Supplementary figure 71: (A) LC-hrMS measurement of a control culture showing the base peak chromatogram (BPC) and the extracted ion chromatogram (EIC) of the expected products as negative control. (B) LC-hrMS measurement of a culture supplemented with DKxanthen 534 showing the base peak chromatogram (BPC) and the extracted ion chromatogram (EIC) of the expected products. (C) LC-hrMS measurement of a culture supplemented with DKxanthen 560 showing the base peak chromatogram (BPC) and the extracted ion chromatogram (EIC) of the expected products. hrMass spectrum of peak 2 and peak 1 are displayed below the chromatogram.

he masses of the expected products could not be detected in crude extracts of the feeding cultures. Peaks in the supplementary figure 70C do not originate from the expected product since for both peaks the EIC trace originates from the second isotope peaks of the positive charged ions  $m/z = 694.41m/z$  and  $m/z = 696.38$ . Both ions exclusively appear in the cultures supplemented with DKxanthen 560. Whether they are prenylation products or not cannot be stated so far.

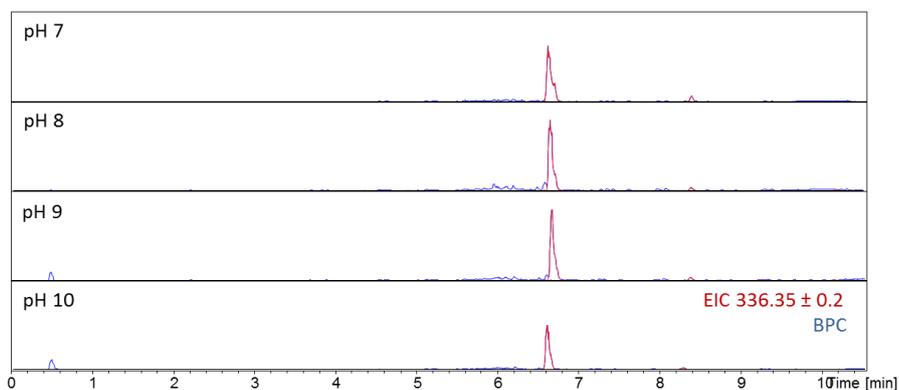
In order to test if not sufficient amounts of DKxanthen are incorporated into cells, one culture was separated from the cultivation media after 13 days by centrifugation. The cells were washed three times with fresh media followed by centrifugation. The media and the cells were extracted separately and measured using LC-hrMS.



Supplementary figure 72: Overlapped LC-hrMS chromatogram of cells and media/supernatant: DKxanthen 534 (detected with  $EIC 535.25 \pm 0.02$ ) indicating presence in cells and media

DKxanthen 534 (the one that was supplemented) was found in the media as well in the cell extract indicating that the supplemented DKxanthenes are incorporated in the cells or the cell wall.

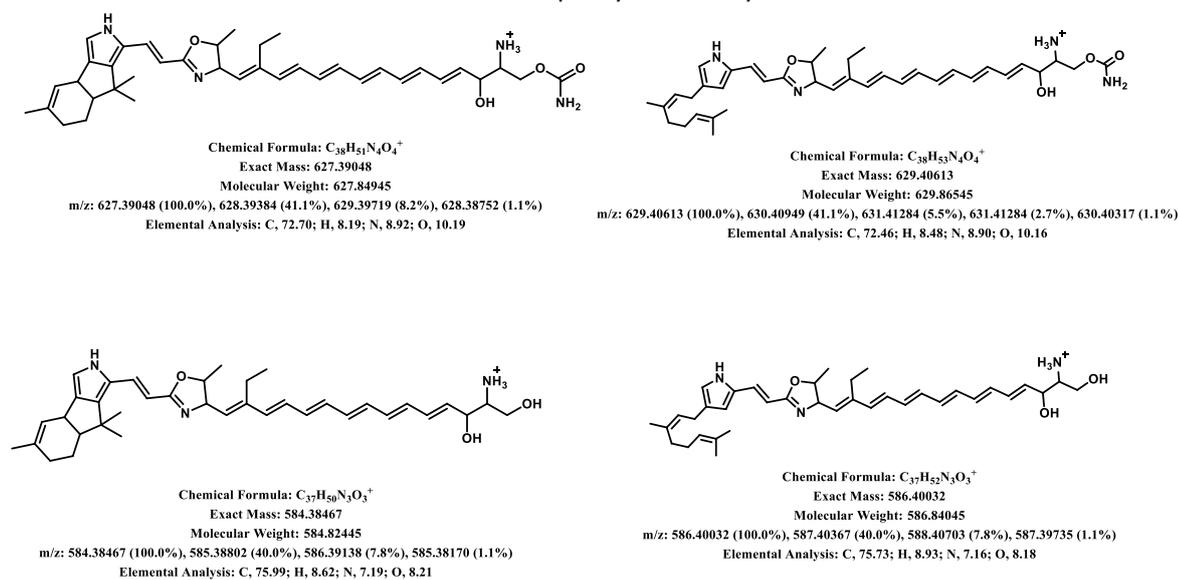
The cyclization of the terpene part was also investigated. The question here was whether the cyclization happens spontaneous or is enzymatically catalyzed. Therefore, a potential turnover of Myxoquartarine 671 into Myxoquartarine 669 was investigated under different pH levels.



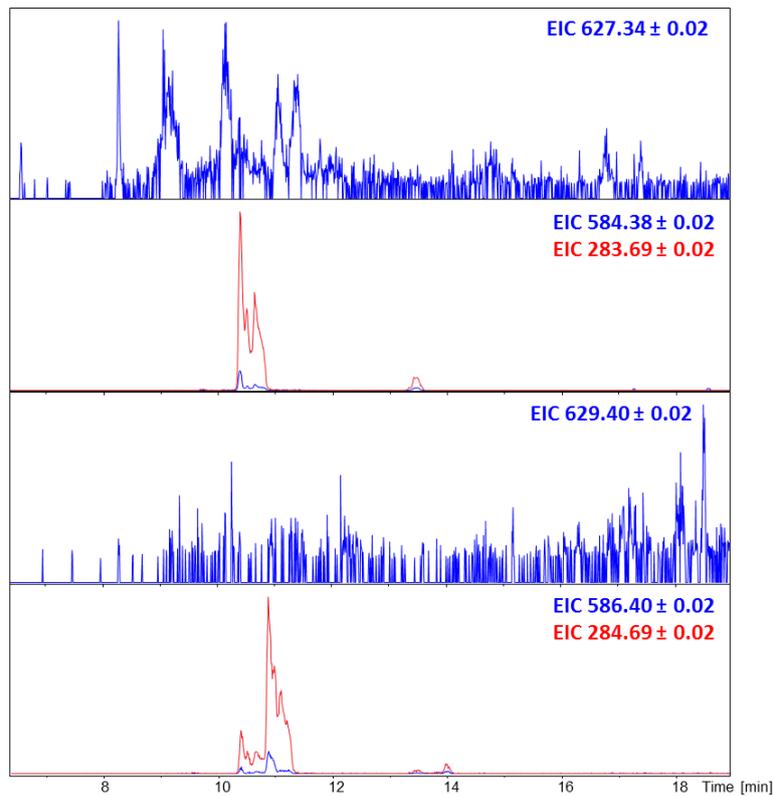
Supplementary figure 73: LC-MS chromatogram of reaction mixtures with different pH levels. In case of cyclization myxoquartarine 671a observed with the EIC 336.35 would disappear.

No cyclization into Myxoquartarine 669 was observed after 24 h under different pH levels in aqueous solution. Therefore, a cyclization might occur under acidic conditions or is completely enzyme dependent.

Crude extract of MSr 11954 was screened for prenylated biosynthetic intermediates.



Supplementary figure 74: Structures and masses of putative prenylated myxoquaterine intermediates

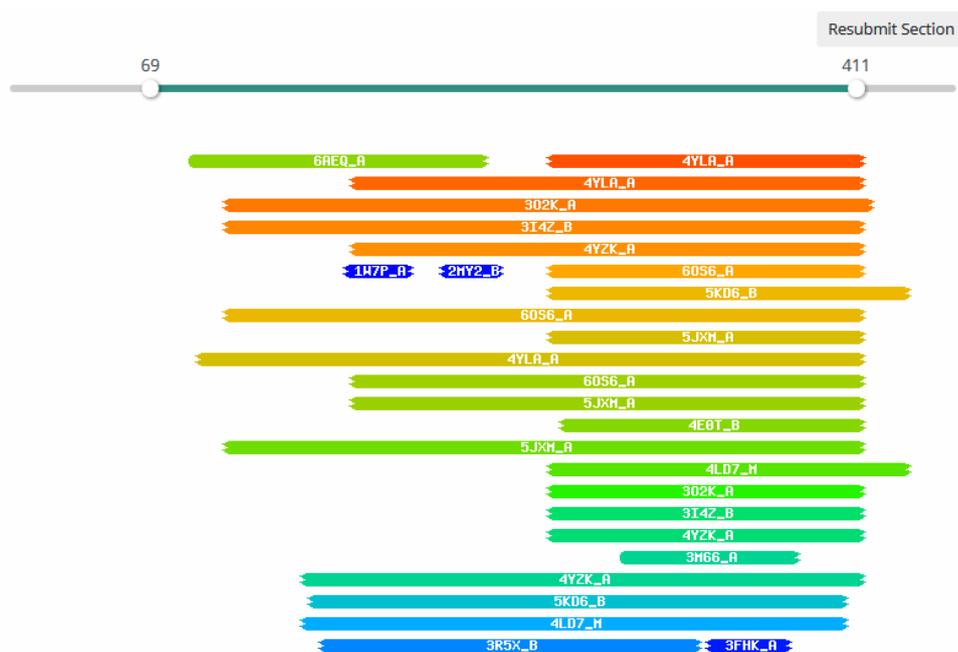


Supplementary figure 75: LC-hrMS of a crude extract of MSr 11954 showing extracted ion chromatograms (EICs) of putative prenylated intermediates.

Masses that would fit prenylated versions of myxoquaterine 450 were identified.

### Identification of Mxql as prenyltransferase

Mxql was analyzed using the HHpred toolkit<sup>[2]</sup>. The six best hits yielded enzymes exclusively classified as prenyltransferases.



Nr	Hit	Name	Probability	E-value	SS	Cols	Target Length
<input type="checkbox"/> 1	4YLA_A	Aromatic prenyltransferase; TRANSFERASE, indole prenyltransferase, PT-fold, indolactam; HET: SO4, DST, ILV; 1.4A {Marinactinospora thermotolerans}; Related PDB entries: 4YL7_A	93.24	0.26	8.6	148	382
<input type="checkbox"/> 2	4YLA_A	Aromatic prenyltransferase; TRANSFERASE, indole prenyltransferase, PT-fold, indolactam; HET: SO4, DST, ILV; 1.4A {Marinactinospora thermotolerans}; Related PDB entries: 4YL7_A	91.42	12	16.4	234	382
<input type="checkbox"/> 3	3O2K_A	Brevianamide F prenyltransferase; PT Barrel, Brevianamide F Prenyltransferase; HET: QRP, MES, DST; 2.4A {Aspergillus fumigatus}; Related PDB entries: 3O24_A	89.91	19	16.4	303	474
<input type="checkbox"/> 4	3I4Z_B	Tryptophan dimethylallyltransferase (E.C.2.5.1.34); prenyl transferase, PT barrel, Alkaloid; HET: GOL, BU2; 1.76A {Aspergillus fumigatus}; Related PDB entries: 3I4X_B 3I4X_A 3I4Z_A	88.67	20	15.3	299	465
<input type="checkbox"/> 5	4YZK_A	Tryptophan dimethylallyltransferase; TRANSFERASE, indole prenyltransferase, PT-fold, indolactam; 1.95A {Streptomyces blastmyceticus}; Related PDB entries: 4YZJ_A 4YZL_A	87.94	14	13.3	235	399
<input type="checkbox"/> 6	6O56_A	CymD prenyltransferase; prenyltransferase, tryptophan, indole, biosynthesis, TRANSFERASE; HET: BEZ, TRP, 6C7; 1.33A {Salinispora arenicola (strain CNS-205)}; Related PDB entries: 6O53_B 6O53_A 6O55_C 6O55_B 6O55_A	85.91	7.4	9.9	145	375

Supplementary figure 76: HHpred search result hits for MxqI.

## References

- [1] a) J. Herrmann, S. Hüttel, R. Müller, *ChemBioChem* **2013**, *14*, 1573; b) S. Hüttel, G. Testolin, J. Herrmann, T. Planke, F. Gille, M. Moreno, M. Stadler, M. Brönstrup, A. Kirschning, R. Müller, *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 12760.
- [2] P. Meiser, H. B. Bode, R. Müller, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19128.
- [3] L. Zimmermann, A. Stephens, S.-Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A. N. Lupas, V. Alva, *Journal of molecular biology* **2018**, *430*, 2237.

## Chapter 3 Myxoglucamide

### 3.1 Genome mining for compound discovery

Genome sequencing in combination with bioinformatics software tools such as antiSMASH<sup>[1]</sup> revealed that the capability of most bacteria to produce secondary metabolites is much higher than the number of identified secondary metabolites<sup>[2]</sup>. The concept of genome mining tries to overcome this discrepancy and to connect the genome with the metabolome. Various studies have proven this concept to be a valuable tool for natural product discovery<sup>[3]</sup>. For this study, the new myxobacterial strain MCy9003 was chosen. The strain belongs to a new genus within the suborder of *Cystobacterinae* and was isolated from a soil sample from the Philippines. Considering its taxonomic distance from other isolates, we expected a differing metabolic profile compared to other *Cystobacterinae*<sup>[4]</sup>. The genome had a size of 11.3 MB and an antiSMASH analysis revealed a total number of 40 Biosynthetic Gene Clusters (BGCs). This enormous biosynthetic potential was in contrast to the observed metabolome. A detailed analysis of the metabolome revealed production of the known and often occurring secondary metabolites myxochelins<sup>[5]</sup> and geosmin<sup>[6]</sup> (supplementary figure 51). However, no promising novel candidates for isolation were identified. Therefore, we choose a genome mining approach for this strain. In this study, we decided to link a BGC to its corresponding metabolites by insertion of an artificial promoter in order to overexpress the BGC. If successful, this approach would increase the yield of the corresponding metabolites, which could then be identified by comparison of the LC-MS chromatograms of wild type and mutant extracts. In addition to the identification, the increased production would also enable the isolation of the corresponding metabolites in sufficient amounts for structure determination and bioactivity testing.

Since manipulations in the genome can have significant effects on the metabolome, which might not be directly connected to the BGC, four selection criteria were established to decide whether the activation of a BGC was regarded as successful or not and if the work with a mutant culture would be continued:

1. The peak area of a target mass must be increased by at least 50 % compared to the WT.
2. Target masses must not be present in the media blank at the same retention time.
3. Target masses must not give any hits for known compounds in the *in-house* database Myxobase. As search criteria, the high resolution mass and the retention time were chosen.
4. Effects of one BGC “activation” must be limited to one target mass or a group of related target masses that have similar MS<sup>2</sup> fragmentation pattern.

Ten BGCs were activated by insertion of an artificial promoter in front of the BGC. The side directed promoter integration was achieved via homologous recombination as described by Panter et al.<sup>[7]</sup>. A vector containing the corresponding homology region, a kanamycin resistance gene as selection marker and either the heterologous tn5 promoter or the inducible vanillate promoter, was used. A region within the first operon within a BGC was used as homology region. The clusters were chosen based on their novelty and architecture. The focus was set on NRPS and PKS type BGCs. The vector was inserted into the cells via electroporation. To select the recombination mutants, cells were cultivated on ager supplemented with kanamycin.

The successful insertion of the promoter was verified by colony PCR using overlapping sets of primers. Mutant and WT strains were cultivated in triplicates and the crude extracts of the cultures were analyzed by LC-*hr*MS. The resulting base-peak chromatograms (BPC) were compared manually. Three out of ten mutant strains fulfilled all four criteria established for the successful activation of the BGC (table 1). Promoter inser-

tion into cluster 18, a BGC identified as NRPS-t1PKS by antiSMASH, led to the discovery of a novel glycolipopeptide family later called myxoglucamides. Activation of Cluster 33 and 34B yielded two different peptide classes. (see supplementary figures 52-57)

*Table 1: List of activated Biosynthetic gene clusters (BGCs), type of Clusters and result of the “activation”*

<b>Cluster</b>	<b>type</b>	<b>Criteria fulfilled</b>	<b>compounds</b>
1	NRPS-t1PKS	No	
8	t1pks	No	
12	NRPS	No	
18	NRPS-t1PKS	Yes	myxoglucamides
24	NRPS-t1PKS	No	
25	NRPS-t1PKS	No	
31B	NRPS-t1PKS	No	
33	NRPS-t1PKS	Yes	peptides 501, 487
34B	NRPS-t1PKS	Yes	peptides 472, 619, 691, 717, 743
35B	NRPS-t1PKS	No	

### 3.1.1 Identification of the myxoglucamides

Mutants of MCy 9003 were created by insertion of two tn5 promoter with different orientation inside BGC no. 18 in front of a NRPS-PKS hybrid gene. PCR analysis confirmed the insertion of the promoter in front of the BGC and one mutant and WT were cultivated in triplicates and extracted using acetone/MeOH 1:1. The crude extracts of were analyzed using LC-*hr*MS and the chromatograms were found to exhibit major differences in the area between 8 and 10.5 min. A closer investigation identified five new peaks corresponding to mass-to-charge ratios between 631.4 and 661.4, which were present in the mutant extracts but not in the wildtype.

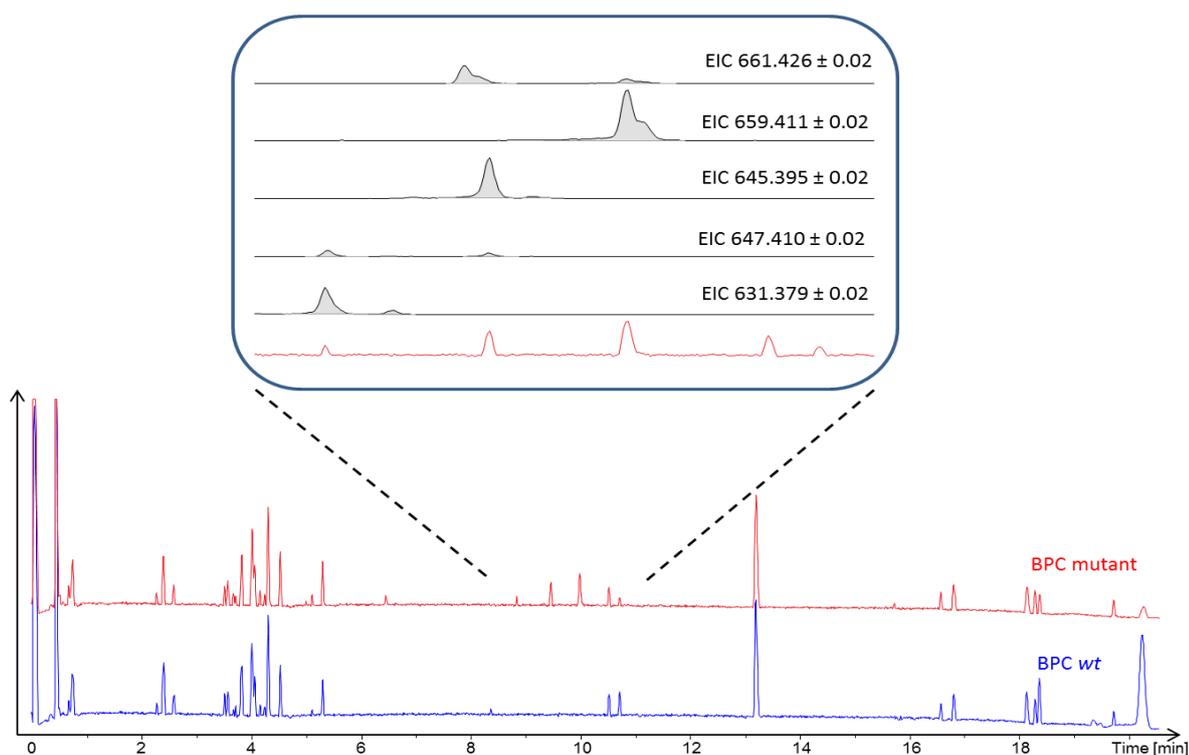


Figure 1: Comparison of extracts from mutant *tn5 NRPS-t1pks 18* and *MCy 9003 WT* showing the base peak chromatograms of WT (blue) and mutant (red). In the detailed view extracted ion chromatograms (EICs) of the five identified masses not present in WT.

Analysis of the MS<sup>2</sup> fragmentation pattern of the newly emerging masses showed a high degree of similarity indicating that they represent a group of related molecules. For further investigation, 9l of the mutant strain were cultivated for 12 days with XAD-16 adsorber resin. The resin and the cells were separated from the supernatant and extracted using MeOH and acetone. The molecules were isolated from the organic extract using a combination of liquid/liquid partitioning, flash chromatography and semipreparative HPLC. Structure elucidation using NMR spectroscopy identified the isolated compounds as a family of glycolipopeptides with differences in the length of their lipid chain and the oxidation level of their peptide moiety. Therefore, the newly isolated compounds were named myxoglucamides.

## 3.2 Structure elucidation of Myxoglucamides

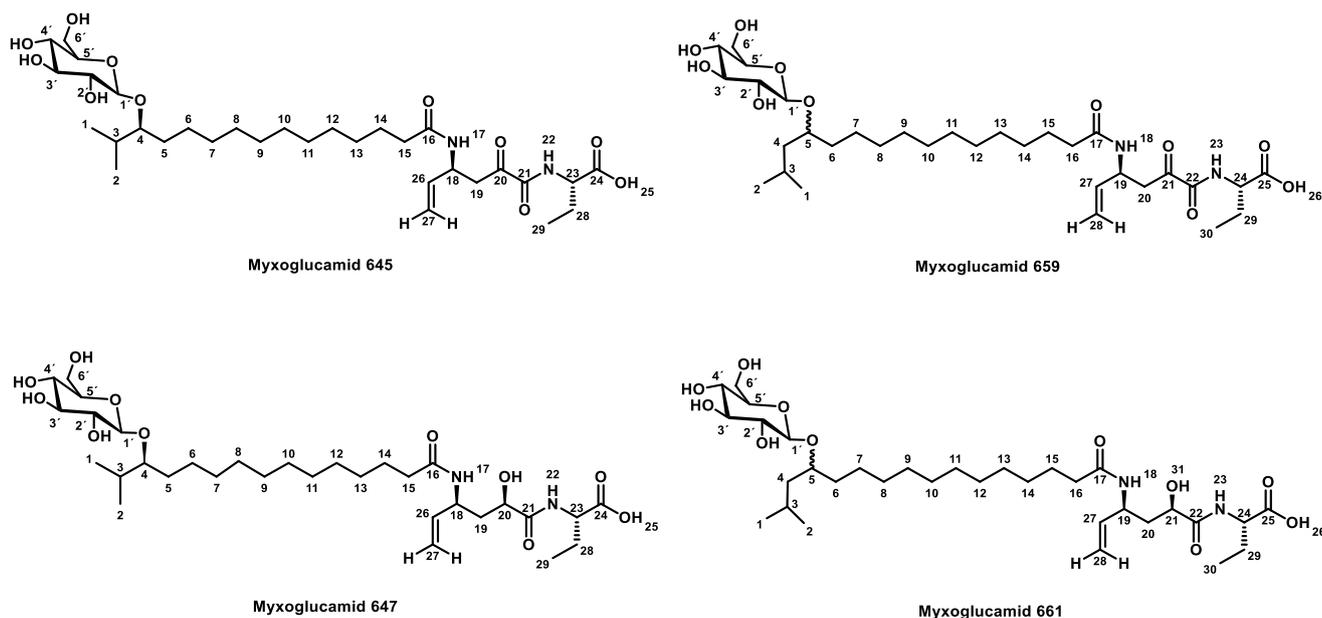


Figure 2: Structures of the isolated myxoglucamide 645 (1), myxoglucamide 659 (2), myxoglucamide 647 (3) and myxoglucamide 661 (4)

### Myxoglucamide 645

Myxoglucamide 645 (1) was isolated as a colorless amorphous solid. The high resolution mass spectrum showed an intense signal for the single charged ion  $[M+H]^+$  at  $m/z = 645.396$ . Using data analysis software, the sum formula of  $C_{32}H_{57}N_2O_{11}$  was determined for protonated mass (calc. 645.39569 measured: 645.39551  $\Delta = 0.28$  ppm).

The  $^1H$  spectrum of 1 exhibited one signal at  $\delta$  5.86 (H-26) characteristic for an olefinic proton and two vinylic protons at  $\delta$  5.19 and 5.11 (H-27a/b). In addition, two proton resonances at  $\delta$  4.87 (H-18) and  $\delta$  4.29 (H-23) were characteristic for  $\alpha$ -protons. Moreover, a group of methylene protons was identified between  $\delta$  1.58–1.27. Beside this, three proton signals at  $\delta$  0.93 (H<sub>3</sub>-1),  $\delta$  0.94 (H<sub>3</sub>-2) and  $\delta$  0.95 (H<sub>3</sub>-29) were found that indicated the presence of three methyl groups based on their characteristic chemical shift values. Furthermore, investigation of the HSQC spectrum revealed presence of two methine groups ( $\delta_{C-4}$  85.2,  $\delta_{H-4}$  3.49), ( $\delta_{C-3}$  31.6,  $\delta_{H-3}$  1.90), a methylene group ( $\delta_{C-15}$  37.1,  $\delta_{H-15}$  2.16) and two diastereotopic methylene groups ( $\delta_{C-19}$  42.7,  $\delta_{H-19}$  2.99/3.21) and ( $\delta_{C-28}$  25.9,  $\delta_{H-28}$  1.80/1.96). Close investigation of the HSQC spectrum showed the presence of an anomeric signal ( $\delta_{C-1'}$  103.4,  $\delta_{H-1'}$  4.30), characteristic a sugar and five signals ( $\delta_{C-2'}$  75.2,  $\delta_{H-2'}$  3.17), ( $\delta_{C-3'}$  78.3,  $\delta_{H-3'}$  3.32), ( $\delta_{C-4'}$  71.7,  $\delta_{H-4'}$  3.31), ( $\delta_{C-5'}$  77.7,  $\delta_{H-5'}$  3.22) and ( $\delta_{C-6'}$  63.0,  $\delta_{H-6'}$  3.68/3.85) that could also be assigned to a pyranose sugar

Based on the characteristic carbon chemical shifts, the pyranose was identified as glucose<sup>[8]</sup>. Examination of the anomeric proton signal at  $\delta$  4.30 (C-1') showed a large coupling constant of  $^3J_{H,H} = 7.7$  Hz between H-1' and H-2' indicating that the glucose is present in its  $\beta$ -anomeric form.

The methine proton at  $\delta$  3.49 (H-4) shows COSY correlations to a multiplet at  $\delta$  1.90 (H-3) and a methylene group at  $\delta$  1.49 (H<sub>2</sub>-5). The corresponding carbon resonance of C-4 at  $\delta$  85.2 shows a strong downfield shift indicative for an adjacent hydroxyl group. H-3 shows COSY correlations to methyl protons at  $\delta$  0.93 (H<sub>3</sub>-1) and  $\delta$  0.94 (H<sub>3</sub>-2) as well as HMBC correlations to  $\delta$  18.4 (C-1) and  $\delta$  18.1 (C-2). Starting from H<sub>2</sub>-5, further COSY correlations can be observed to another methylene group at  $\delta$  1.36 (H<sub>2</sub>-6), which is part of a larger spin

system of the lipid chain (figure 3). The lipid chain consists of nine additional methylene groups as indicated by MS<sup>2</sup> fragmentation data. Unfortunately, seven of the methylene groups show highly overlapping proton and carbon signals in the range between  $\delta$  1.27-1.32 and  $\delta$  30.3-31.0 characteristic for a saturated lipid chain. A methylene group at  $\delta$  1.58 (H<sub>2</sub>-14) shows a COSY correlation to the area of the overlapping signals as well as to additional methylene protons at  $\delta$  2.16 (H<sub>2</sub>-15). Both methylene proton signals H<sub>2</sub>-14 and H<sub>2</sub>-15 also show HMBC correlations to the overlapping carbon region as well as to a carbonyl carbon at  $\delta$  175.7 (C-16) specifying the lipid chain as 13-hydroxy-14-methyl-pentadecanoic acid (HMPA) residue.

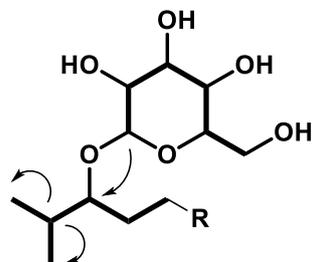


Figure 3: Key COSY (bold line) and HMBC (arrow) correlations of Glucose linked to hydroxy-methylpentadecanoic acid (HMPA)

Examination of COSY crosspeaks showed that the vinylic protons  $\delta$  5.19/5.11 (H-27a/b) correlate to the olefinic proton  $\delta$  5.86 (H-26), indicating both positions form a vinyl group. Further COSY correlations from H-26 to  $\delta$  4.86 (H-18) indicate the vinylic group is attached this  $\alpha$ -position. The  $\alpha$ -proton H-18 shows more COSY correlations to two diastereotopic protons  $\delta$  2.99/3.21 (H-19a/b), demonstrating both pos-18 and pos-19 are neighboring. HMBC correlations from  $\delta$  4.87 (H-18) to a carbon resonance at  $\delta$  196.3 (C-20) in combination with HMBC correlations from  $\delta$  2.99/3.12 (H-19a/b) to C-20 and second carbon resonance at  $\delta$  162.0 (C-21) indicate the presence of an  $\alpha$ -keto-amide group at pos-19. With the previously listed connectivities, this moiety can be deduced as 4-amino-2-oxohex-5-enoic acid block (figure 4). The signal at  $\delta$  4.29 (H-23) was identified as the  $\alpha$ -position of  $\alpha$ -aminobutyric-acid based on characteristic COSY correlations to two diastereotopic protons signals  $\delta$  1.80/1.96 (H-28a/b) that show further correlations to the methyl group (H<sub>3</sub>-29). H-23 and H-29a/b show HMBC correlations to a carbon resonance at  $\delta$  175.3 (C-25), the carboxyl group of the  $\alpha$ -aminobutyric-acid.

Connectivities between the described structural moieties were established by examination of HMBC correlations. The glucose is linked to position 4 of the HMPA chain via an ether bridge as indicated by strong HMBC from  $\delta$  4.30 (H-1') to  $\delta$  85.2 (C-4) (figure 3). Furthermore, the HMPA chain is connected via its carboxyl terminus to the 4-amino-2-oxohex-5-enoic acid block. HMBC correlations from  $\delta$  2.17 (H-18) to at  $\delta$  175.7 (C-16) clearly indicate this arrangement. In addition, a strong HMBC signal from  $\delta$  4.29 (H-23) to the carbon resonance  $\delta$  162.0 (C-21) can be observed, implying that the  $\alpha$ -aminobutyric-acid is linked via a peptide bond to the 4-amino-2-oxohex-5-enoic acid block. To summarize, the structure can be classified as glycolipopeptide with a glucose, attached to hydroxyl-methylpentadecanoic acid chain, which is further via its carboxyl terminus to an untypical vinylic  $\alpha$ -keto acid and a C-terminal amino butyric acid.

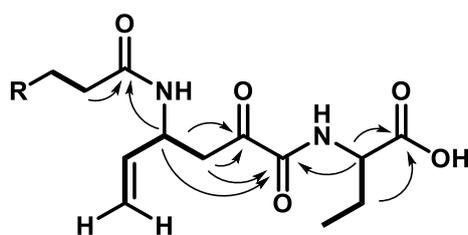


Figure 4: Key COSY (bold line) and HMBC (arrow) correlations of Hydroxy-Methylpentadecanoic acid (HMPA) and peptidic part of Myxoglucamid 645

### Myxoglucamide 659

Myxoglucamide 659 (**2**) was isolated as a colorless amorphous solid. The high resolution mass spectra showed an intense signal for the single charged ion  $[M+H]^+$  at 659.411  $m/z$ . Using data analysis software, the sum formula of  $C_{33}H_{59}N_2O_{11}$  was determined for protonated mass (calc. 659.41081 measured: 659.41134  $\Delta = 0.80$  ppm).

The acquired NMR datasets highly resembled the NMR data of **1** but some signals from the lipid chain were altered and two new diastereotopic methylene protons at  $\delta$  1.28/1.52 (H<sub>2</sub>-4a/b) were identified. Investigation of COSY crosspeaks showed that this methylene group is neighboring H-5 at  $\delta$  3.56, which show further COSY correlations into a bigger spin system comparable to the lipid chain in **1**. The final constitution of the lipid chain was determined by investigation of COSY correlation from pos-1-8. In combination with MS<sup>2</sup> fragmentation data, these findings indicate the lipid chain to be extended by an additional methylene group, specifying this lipid as 14-hydroxy-15-methyl-pentadecanoic acid (HMHA). The pyranose moiety is linked to position 5 as indicating by HMBC correlations from the anomeric proton H-1' to the corresponding carbon resonance at  $\delta$  79.3 (C-5) (figure 5).

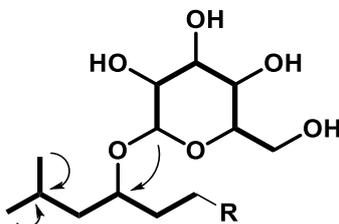


Figure 5: Key COSY (bold line) and HMBC (arrow) correlations of Glucose linked to hydroxy-methylhexadecenoic acid (HMHA)

### Myxoglucamide 647

Myxoglucamide 647 (**3**) was isolated as a colorless amorphous solid. The high resolution mass spectra showed an intense signal for the single charged ion  $[M+H]^+$  at 647.411  $m/z$ . Using data analysis software, the sum formula of  $C_{32}H_{59}N_2O_{11}$  was determined for protonated mass (calc. 647.41134 measured: 647.41073  $\Delta = 0.93$  ppm)

The NMR datasets of **3** has high similarities with the spectra of **1** and **2**. Proton and carbon resonances of pyranose sugar and HMPA chain are identical with its signals from **1** suggesting an identical structure. Differences can be found in peptide part of the molecule. The  $\alpha$ -proton is downfield shifted to  $\delta$  4.58 and the diastereotopic methylene protons H-19a/b and H-19b are also downfield shifted to  $\delta$  1.97/1.77. In addition a new proton signal at  $\delta$  4.03 (H-20) appeared while the resonance observed for the ketone in **1** disappeared, suggesting the ketone being reduced into an alcohol in **3**. Indeed, as indicated by COSY correlations H-20 is located next to methylene protons H-19a/b. Both protons  $\delta$  4.01 (H-20) and  $\delta$  1.97/1.77 (H-19a/b) show HMBC correlations to a carbonyl group at  $\delta$  176.5 (C-21), indicating this moiety to be a 4-amino-2-hydroxy-hex-5-enoic acid (figure 6). The connectivity is identical to **1**, since observed HMBC correlations are similar.

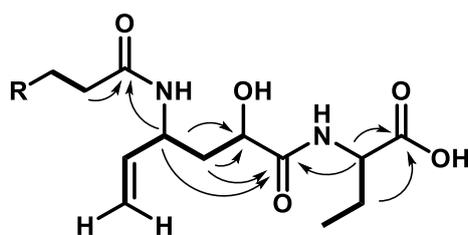


Figure 6: Key COSY (bold line) and HMBC (arrow) correlations of hydroxy-methylpentadecanoic acid (HMPA) and peptidic part of myxoglucamid 647

### Myxoglucamide 661

Myxoglucamide 661 (**4**) was isolated as a colorless amorphous solid. The high resolution mass spectra showed an intense signal for the single charged ion  $[M+H]^+$  at 661.426  $m/z$ . Using data analysis software, the sum formula of  $C_{33}H_{61}N_2O_{11}$  was determined for protonated mass (calc. 661.42699 measured: 661.42636  $\Delta = 0.95$  ppm).

The acquired NMR datasets resembled those of **2** and **3** indicating identical structural features. Chemical shifts and COSY as well as HMBC correlations are identical with the ones observed for the HMHA chain and the pyranose moiety in **2**, implying an identical structure. For the peptide part, observed chemical shifts, COSY and HMBC correlations were identical with the 4-amino-2-hydroxyhex-5-enoic acid moiety and the  $\alpha$ -amino-butyric acid observed in **3**, also implying an identical structure.

Table 2: Spectroscopic data of myxoglucamid 645 and 647 acquired in  $CD_3OD$

position	Myxoglucamid 645			Myxoglucamid 647	
	$\delta_C^a$	$\delta_H^b$ (J in Hz)	$\delta_C^c$	$\delta_H^d$ (J in Hz)	
1	18.4	0.93 d 7.0	17.9	0.93 <sup>e</sup>	
2	18.1	0.94 d 6.6	18.3	0.93 <sup>e</sup>	
3	31.6	1.90 m	31.5	1.90 m	
4	85.2	3.49 m	84.4	3.49 m	
5	31.8	1.49 m	31.7	1.49 m	
6	26.6	1.36 m	26.6	1.35 m	
7-13 <sup>e</sup>	30.3-31.0	1.27-1.32	29.6-31.2	1.26-1.36 m	
14	27.1	1.58 m	26.8	1.60 m	
15	37.1	2.16 t	36.9	2.17 t (7.6)	
16	175.7		175.2		
17	NH		NH		
18	48.6	4.87 m	50.1	4.58 m	
19a	42.7	2.99 d,d 16.8, 8.4	39.8	1.97 m	
19b		3.21 d,d 16.8, 5.3		1.77 m	
20	196.3		69.9	4.03 d,d (3.4,9.5)	
21	162.0		176.5		
22	NH		NH		
23	55.5	4.29 m	54.8	4.31 m	
24	175.2		175.4		
25	OH		OH		
26	138.6	5.86 d,d,d 5.5, 10.4, 5.5	138.2	5.79 d,d,d (17.3,10.5,18)	
27a	115.5	5.19 dt,1.4,17.2	116.4	5.23 d,t (17.0,1.4)	
27b	115.5	5.11 dt 1.4, 10.4		5.15 d,t (10.3,1.2)	
28a	25.9	1.80 m	25.9	1.92 m	
28b		1.96 m		1.78 m	
29	10.5	0.95 t 7.4	9.8	0.93 <sup>e</sup>	
1'	103.4	4.30 d 7.8	103.6	4.30 d (7.6)	
2'	75.5	3.17 d,d	75.2	3.17 d,d (9.0,8.2)	
3'	78.3	3.32 m	78.0	3.33 m	

4'	71.7	3.31 <sup>e</sup>	71.5	3.30 m
5'	77.7	3.22 m	77.4	3.22 m
6'a	63.0	3.85 d,d 11.6, 2.4	62.6	3.85 d,d (11.7,7.6)
6'b	63.0	3.68 d,d 11.5, 5.5	62.6	3.68 d,d (11.6,5.6)

<sup>a</sup> acquired at 125 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

<sup>b</sup> acquired at 500 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

<sup>c</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

<sup>d</sup> acquired at 700 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

<sup>e</sup> overlapping signals

Table 3: Spectroscopic data of myxoglucamid 659 and 661 acquired in CD<sub>3</sub>OD

position	Myxoglucamid 659		Myxoglucamid 661	
	$\delta_C^a$	$\delta_H^b$ (J in Hz)	$\delta_C^c$	$\delta_H^d$ (J in Hz)
1	23.5	0.91 d 6.6	23.3	0.92 d (6.3)
2	23.3	0.93 d 6.6	23.5	0.91 d (6.6)
3	25.5	1.81 m	25.5	1.80 m
4a	45.1	1.28 m	45.1	1.52 m
4b	45.1	1.52 m		1.28 m
5	79.3	3.75 d,d,d,d 6.1,6.1,6.0,6.0	79.3	3.74 m
6a	36.7	1.51 m	36.7	1.59 m
6b	36.7	1.60 m		1.53 m
7	26.2	1.38 m	26.2	1.38 m
8-14	30.4-31.2	1.28-1.32	30.3-31.0	1.33-1.29 m
15	27.1	1.58 m	27.1	1.59 m
16	37.1	2.16 t 7.4	37.2	2.17 t (7.6)
17	175.7		175.4	
18	NH		NH	
19	48.7	4.87 c	50.4	4.58 m
20a	42.7	2.99 d,d 16.8,8.5	40.2	1.97 m
20b	42.7	3.21 d,d 5.4,16.8	40.2	1.78 m
21	196.3		70.3	4.04 d,d (9.4,3.6)
22	162.0		176.9	
23	NH		NH	
24	55.5	4.29 m	54.7	4.33 d,d (7.6,5.0)
25	175.3		175.2	
26	OH		OH	
27	138.6	5.87 d,d,d 17.1, 10.4, 5.5	138.5	5.79 d,d,d (17.1, 10.3,7.0)
28a	115.5	5.11 d,t 10.4,1.4	116.7	5.23 d,t (1.3,17.1)
28b	115.5	5.18 d,t 17.0,1.4	116.7	5.16 d,t (1.3,10.4)
29a	25.9	1.80 m	26.1	1.93 m
29b	25.9	1.96 m	26.1	1.79 m
30	10.5	0.95 t 7.7	10.2	0.95 t (7.6)
1'	103.9	4.29 d 7.8	103.9	4.29 d (7.9)
2'	75.4	3.15d,d 8.7,7.8	75.3	3.15 d,d (7.9,8.9)
3'	78.3	3.33 m	78.3	3.33 m
4'	71.8	3.31d	71.8	3.29 m
5'	77.8	3.23 m	77.8	3.23 d,d,d (9.3,5.4,2.3)
6'a	62.9	3.84 d,d 2.4,11.9	62.9	3.85 d,d (2.6,11.7)
6'b	62.9	3.84 d,d 5.5,11.6	62.9	3.68 d,d (11.8,5.5)

<sup>a</sup> acquired at 125 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

<sup>b</sup> acquired at 500 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

<sup>c</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

<sup>d</sup> acquired at 700 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

### 3.2.1 Stereochemical assignment

The absolute configuration of the stereocenters in myxoglucamide 645 (**1**) was determined as follows. For the sugar moiety and the HMPA chain, a previously published procedure used for the stereodetermination of the iodoglucomides was applied<sup>[9]</sup>: By acid hydrolysis, the molecule was cleaved along its amide and glycosidic bonds liberating the HMPA chain, the sugar moiety, and amino butyric acid (figure 7).

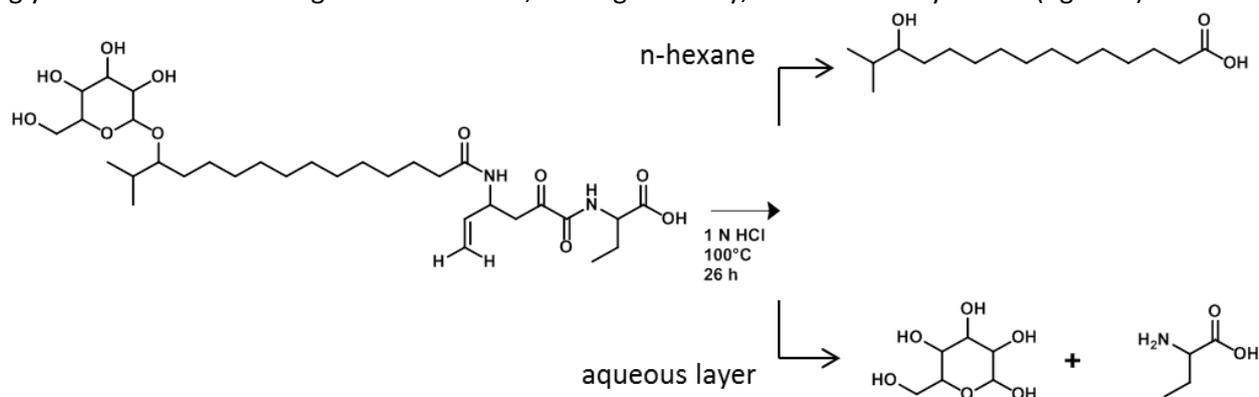


Figure 7: Scheme of the acid hydrolysis and workup of the HMPA, pyranose and amino butyric acid

HMPA was extracted from the hydrolysis mixture using n-hexane and subsequently derivatized with (*R*)- and (*S*)- $\alpha$ -methoxy- $\alpha$ -trifluoromethylphenylacetic acid (MTPA) chlorides resulting in the formation of the respective esters<sup>[10]</sup>. Shift differences between the (*R*)-MTPA ester and (*S*)-MTPA ester were determined by acquiring 2D NMR datasets of both reaction mixtures in pyridine-*d*<sub>5</sub>. Based on the shift differences, the stereocenter HMPA chain can be assigned as (*S*)-configured (figure 8A).

To determine the configuration of the pyranose sugar moiety, the hydrolysis mixture was extracted with ethyl acetate to remove the partially hydrolyzed compound. The remaining aqueous phase contained the pyranose sugar and, to a smaller extent, amino butyric acid as confirmed by NMR in D<sub>2</sub>O. As expected due to similar carbon resonances of the sugar, the pyranose was assigned as glucose since the <sup>13</sup>C chemical shift values were identical with values derived from a D-glucose standard (figure 8B). The positive sign of the optical rotation of  $[\alpha]_D^{20} + 17^\circ \text{ g}^{-1} \text{ ml dm}^{-1}$  indicated that the glucose is D-configured. In mixtures, the optical rotation is averaged based on the molar fraction of the component and their corresponding specific optical rotation<sup>[11]</sup>. Since glucose is present in excess in the measured mixture, it was higher weighted than the amino butyric acid which is present as well.

The configuration of the amino butyric acid was determined using advanced Marfey's Method<sup>[12]</sup>: An aliquot of the hydrolysis mixture was derivatized with D- and L-1-fluoro-2,4-dinitrophenyl-5-Leucine amide (FDLA). A comparison of the retention times with the respective standards showed that the amino butyric acid is L-configured.

In order to elucidate the configuration of the stereocenter at the vinyl group, the 4-amino-2-oxohex-5-enoic acid moiety was degraded using ozonolysis in combination with an oxidative workup and following acid hydrolysis in order to release aspartic acid. Using advanced Marfey's method, the configuration of the released aspartic acid residue was determined to be L. Since the stereocenter is not modified during ozonolysis, the spatial configuration is the same in the initial compound. Therefore, the stereocenter in the 4-amino-2-oxohex-5-enoic acid moiety is (*S*)-configured like the  $\alpha$ -position in L-aspartic acid (figure 8C).

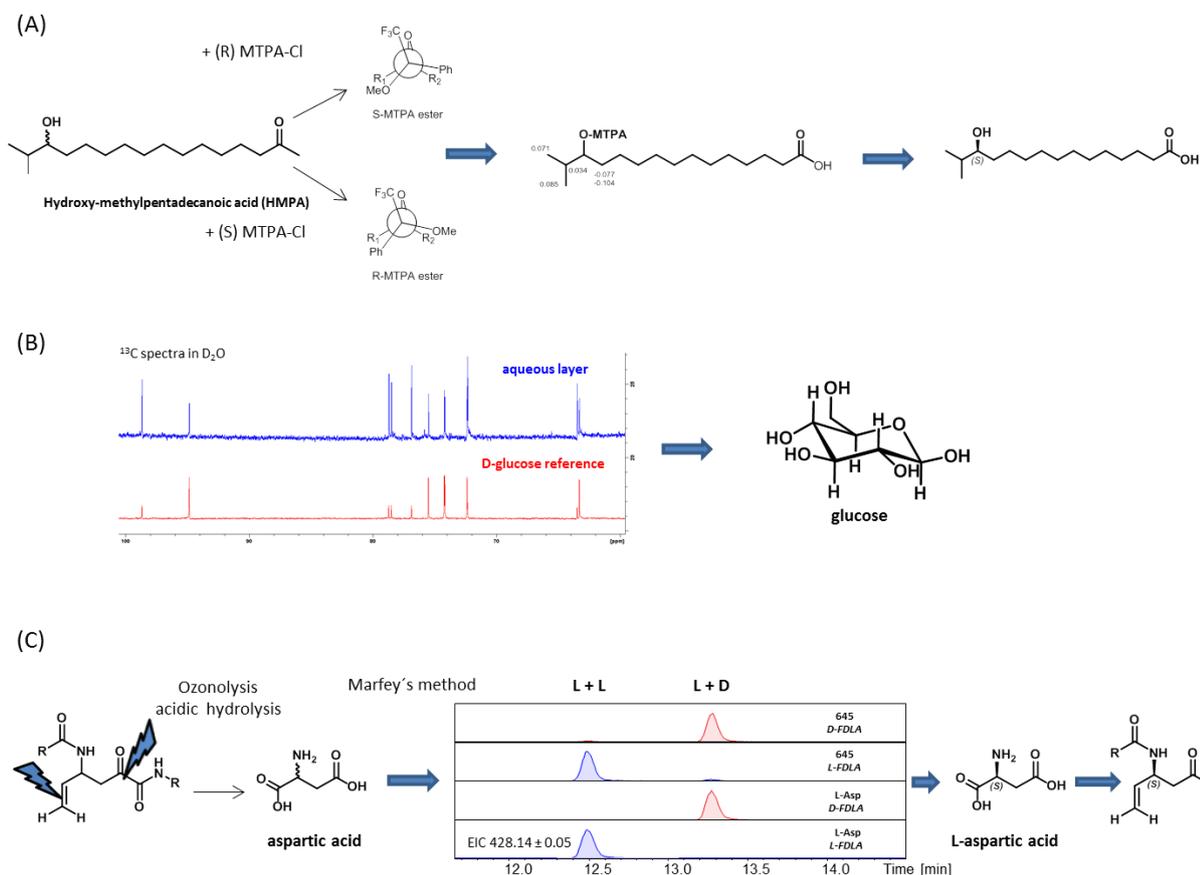


Figure 8: (A): Determination of the absolute configuration of the hydroxy-methylpentadecanoic acid (HMPA): Esters of both MTPA variants were prepared and proton shift differences  $\Delta \delta_{(S-R)}$  were determined. Positive values could be assigned as  $R_1$  according to the Newman projection of the esters while negative values were assigned as  $R_2$ <sup>[10]</sup>. The absolute configuration was therefore assigned as (S). (B): Determination of the pyranose type: <sup>13</sup>C-spectrum of the aqueous layer acquired at 125 MHz in D<sub>2</sub>O displayed in blue and reference <sup>13</sup>C-spectrum of D-glucose acquired at 175 MHz in D<sub>2</sub>O displayed in red. Based on identical carbon resonances, the pyranose could be assigned as glucose. (C): Determination of the stereocenter at the vinyl group: By ozonolysis and workup in H<sub>2</sub>O<sub>2</sub>, the double bond and the  $\alpha$ -keto-group are oxidized and cleaved releasing aspartic acid during acid hydrolysis. Configuration of aspartic acid was elucidated by advanced Marfey's method using derivatization with D- and L- 1-fluoro-2,4-dinitrophenyl-5-Leucine amide (FDLA) and comparison of retention times with an L-aspartic acid standard. Since the chiral carbon is not modified in the procedure, the configuration is identical in the intact molecule.

In all myxoglucamides, the C-terminal amino-butyric acid residue was determined to be L-amino-butyric acid by advanced Marfey's method.

Table 4: Results of Marfey's analysis of myxoglucamide variants compared to the respective amino acid standards

	m/z	RT D-FDLA [min]	RT L-FDLA [min]
L -Abu	398.17	18.16	15.09
D -Abu	398.17	15.10	18.17
Myxoglucamid 645	398.17	18.16	15.09
Myxoglucamid 647	398.17	18.16	15.06
Myxoglucamid 661	398.17	18.16	15.08
Myxoglucamid 659	398.17	18.16	15.09

In myxoglucamid 661 (**4**) and 647 (**3**), the  $\alpha$ -ketone within the 4-amino-2-oxohex-5-enoic acid moiety is reduced to an alcohol resulting in the formation of one additional stereocenter in the molecule. By detailed analysis of ROE correlations, ROE distance measurements and comparison of the obtained distances with distances from a MM2 optimized structure, the relative configuration of both stereocenters was predicted to be 19(*S*), 21(*R*) (figure 9). A *J*-based configuration analysis to confirm the predicted configuration did not succeed due to overlapping signals and insufficient sample amounts.

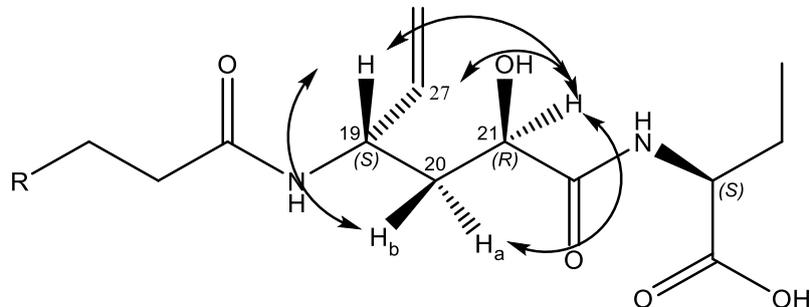


Figure 9: predicted relative configuration of the 4-amino-2-oxohex-5-enoic acid block in myxoglucamide 661. Arrows indicate strong ROE correlations

Considering that biosynthesis of all myxoglucamide derivatives underlie the same logic and are likely to be derived from the same building blocks it is assumed that position 19 in myxoglucamide 661 (**4**) is likely to be (*S*)-configured like in myxoglucamide 645 (**1**) and therefore position 21 is expected to be (*R*)-configured, based on previous assumptions. The same considerations can be adapted to myxoglucamide 647 (**3**) due to the same biosynthetic origin and low deviations in the carbon resonances.

**1** and **3** show deviations of maximal  $\delta$  0.5 for carbon resonances in proximity to position 4 indicating position 4 and all stereo centers in the glucose have the same relative configuration. Assuming the glucose in myxoglucamid 647 is D-glucose like in variant 645 therefore implies that position 4 is also (*S*)-configured to give the same relative configuration. The same accounts for the 4-amino-2-oxohex-5-enoic acid block in **1** and myxoglucamide 659 (**2**). Both molecules show nearly identical carbon resonances in this area. In addition, CD-spectra of both molecules have identical cotton effects indicating an identical configuration. Unfortunately, Mosher esterification of the HMHA chain in **2** and myxoglucamide 661 (**4**) did not succeed. Since enzymes catalyze glycosylation with a high substrate specificity, the glucose moiety in **2** and **4** is also expected to be D-glucose.

### 3.3 Biosynthesis of myxoglucamides

The core biosynthetic gene cluster (BGC) consists of two NRPS modules, one PKS module and a module containing NRPS and PKS type catalytic domains. In addition, genes up- and downstream of the modules possess conserved domains for oxidative functions that are expected to be essential for the biosynthesis.

Using bioinformatic tools antiSMASH, Pfam and HHpred<sup>[1,13,14]</sup>, the tentative functions of the catalytic domains were proposed, the borders of BGC were suggested and a model for the biosynthesis was created. To simplify understanding of the biosynthesis, core biosynthetic genes were assigned as modules, with each module for one elongation step.

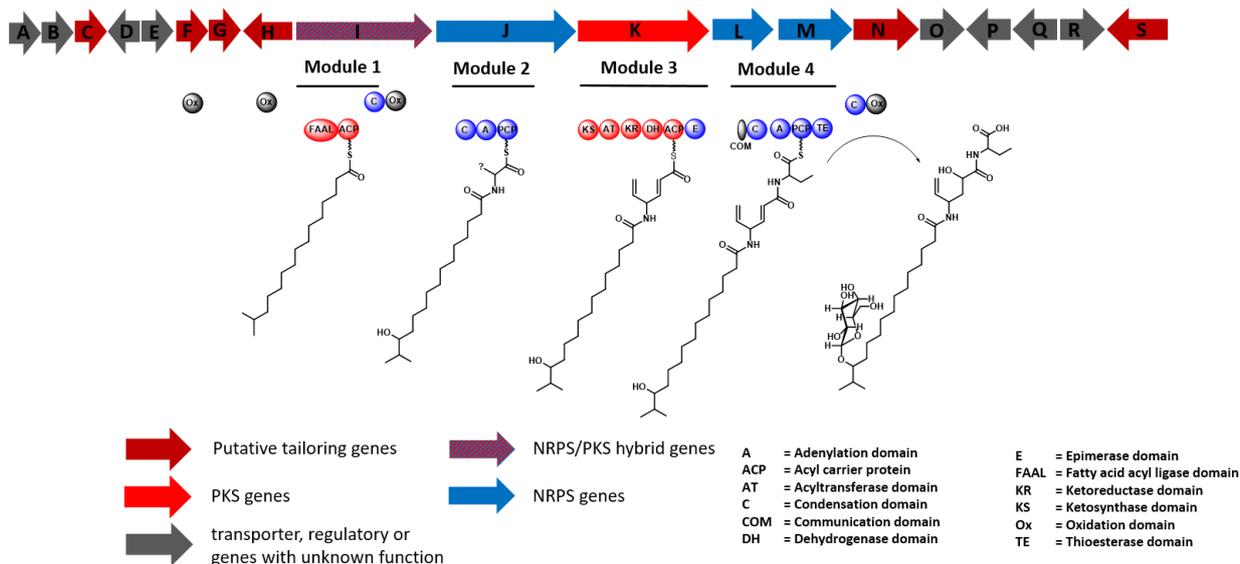


Figure 10: Myxoglucamide gene cluster and proposed model for the myxoglucamide biosynthesis

The biosynthesis of the myxoglucamides is initiated by the loading of a methylpentadecenoic (iso 15:0) or methylhexadecenoic acid (iso 16:0) onto an ACP domain catalyzed by an FAAL domain on module 1 encoded by *mxgl*. In the BGC, no FAS related genes were found. Therefore, it is expected that the iso-fatty acid chains were provided from primary metabolism. Since isovaleryl-CoA and isobutyryl-CoA are known starter units for iso-fatty acids, stable isotope labelled leucine- $d_3$  and valine- $d_8$  were supplemented to the culture. In case of leucine- $d_3$ , incorporation was observed in all myxoglucamid derivatives. But for valine- $d_8$  incorporation was only observed in myxoglucamide variants an iso-15:0 fatty acid (myxoglucamide 645 and 647) (figure 11). Similar observations were made for free carboxylic acids in myxobacteria, supporting the origin of the lipid origin from primary metabolism [15].

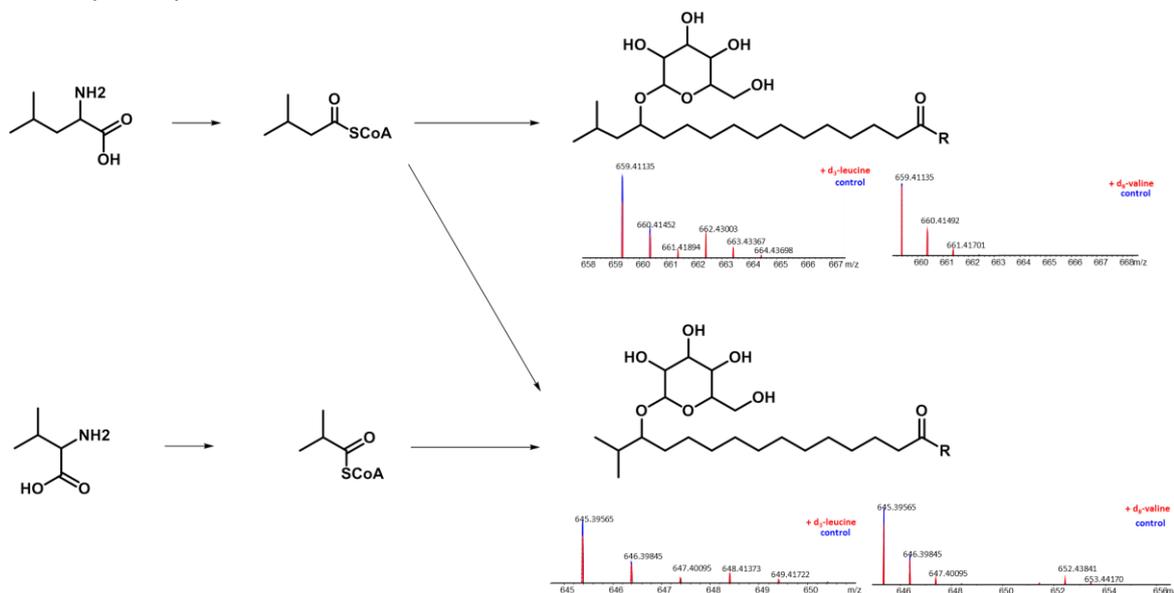


Figure 11: Incorporation of leucine and valine into the HMPA and HMHA chain. Isotope pattern of myxoglucamide 659 and 645 showing single charged mass  $[M+H]^+$  extracted from LC-hrMS from the control culture (blue signals) and a culture supplemented with  $d_3$ -leucine and  $d_8$ -valine (red signals).

Upstream of *mxgl*, three ORFs are located. Analysis using the bioinformatics tool Pfam and HHpred these genes were annotated as a monooxygenase/C-domain in forward direction, a cytochrome P450 and an aldo/ketoreductase. Inside *mxgl* a catalytic C-domain and oxidase domain with homology to tauD enzyme

class are located beside the FAAL domain. The presence of these enzymes bearing oxidative functions in close proximity suggests an involvement in the oxidation step of the iso-fatty acid chains to yield the hydroxyl group. Homologues of CYP124 type cytochromes<sup>[16]</sup> which are described to conduct hydroxylation of  $\omega$ -position of iso-branched fatty chains were not found inside the BGC.

The origin of the 4-amino-2-hydroxyhex-5-enoic acid and 4-amino-2-oxo-hex-5-enoic acid is of NRPS-PKS hybrid nature. In a first step, an amino acid is loaded onto the assembly chain by module 2 and probably modified into the non-proteinogenic amino acid vinylglycine. In a second step, elongation with one acetate unit takes place conducted by PKS catalytic domains in module 3.

Beside the PKS domains an E-domain can be found in module 3, which is atypical for a PKS module. But considering the retro biosynthesis of the 4-amino-2-oxo-hex-5-enoic acid block (figure 12) it would require an amino acid that is (*R*)-configured at the  $\alpha$ -carbon under the assumption that condensation of the peptide bond of fatty acid and NPRS extension follow text book logic and formation of the vinylic moiety does not alter the stereo center. Since most of natural amino acids are (*S*)-configured it would require an epimerization reaction which could be provided by the E-domain. This finding suggest this E-domain to act in trans on module 2.

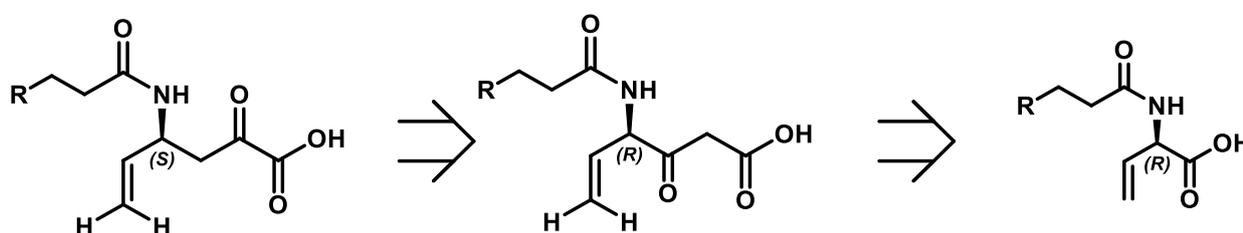


Figure 12: Retro biosynthetic formation of the 4-amino-2-oxo-hex-5-enoic acid block: The configuration changes from (*R*) to (*S*) in the final product due to the reduction of the carboxylic acid to a methylene group that has a lower priority in the CIP nomenclature than the vinylic moiety

Formation of the vinyl group is of certain interest since similar motives are rarely found in natural products. Examples of natural products containing such motives are rhizobitoxine from *Bradyrhizobium elkanii*<sup>[17]</sup>, the fungal metabolite D-vinylglycine<sup>[18]</sup>, aminoethoxyvinylglycine (AVG) from *Streptomyces* sp. NRRL 5331 and methoxyvinylglycine (MVG) from *Pseudomonas aeruginosa*<sup>[19]</sup>. Interestingly the biosynthetic machineries use two different sources for formation of  $\alpha$ -vinyl group. In rhizobitoxine and aminoethoxyvinylglycine (AVG) the  $\alpha$ -vinylic amino acid is derived from homoserine<sup>[17,20]</sup>. Homologues of characteristic enzymes from both biosynthetic pathways like the homoserine dehydrogenase (AVG biosynthesis) and the dihydrorhizobitoxine desaturase RtxC (rhizobitoxine biosynthesis) were not found inside the BGC. Also feeding of isotope labelled aspartic acid as a precursor of homoserine did not show any incorporation.

In MVG biosynthesis glutamic acid is utilized for the formation of the  $\alpha$ -vinylic amino acid. Feeding with stable isotope labelled glutamic acid and glutamine did not show incorporation. Therefore it can be concluded that formation of the  $\alpha$ -vinylic moiety takes place by a so far unknown mechanism with a yet to be identified precursor.

Feeding experiments using <sup>13</sup>C labelled acetate showed acetate incorporation. In order to differ between acetate incorporation into the fatty acid chain and acetate incorporation into the rest of the molecule, fragment 231.13 *m/z* containing the C-terminal peptide part of myxoglucamid 661 was observed using MS<sup>2</sup> fragmentation. The isotope pattern indicates acetate incorporation. Hydroxyl-malonyl-CoA as building block instead of malonyl-CoA was considered unlikely since <sup>13</sup>C<sub>3</sub>glycine, a potential precursor of hydroxyl-malonyl-CoA, did not show incorporation and characteristic biosynthetic genes of hydroxy-malonyl CoA were not found in proximity or inside the BGC<sup>[21]</sup>. A ketoreductase (KR) and dehydrogenase (DH) domain inside MxgK

are reducing the formed ketide into an alkene. Downstream of *mxgK* an oxidoreductase/luciferase like oxidase similar to the one identified in the myxoprincomide BGC<sup>[22]</sup> is located. It is expected to oxidize the  $\alpha$ -position of the alkene into the  $\alpha$ -ketone the 4-amino-2-oxohex-5-enoic acid block.

Biosynthesis of the peptide chain finishes with the condensation of L-amino butyric acid. The last module is split into two genes together containing a C-domain, an A-domain and a TE domain. Termination of the assembly chain and release is conducted by the TE domain.

At both ends of the BGC two glycosyltransferase (GT) genes, *mxgC* and *mxgS*, can be found. Sequence comparison of the first GT using the Pfam toolkit<sup>[13]</sup> suggested this glycosyltransferase to belong into GT family 28 with UDP-glucuronosyl or UDP-N-acetylglucosamine as proposed substrates. The second GT can be classified into family 1<sup>[23]</sup>. Which GT is actually adding the glucose remains elusive.

Table 5: Annotated genes found in the myxoglucamid biosynthesis

Gene	Size <sup>a</sup>	Protein homolog <sup>b</sup>	Proposed function	Catalytic domains
<i>mxgA</i> (-)	393	WP_002615006.1 (100/78)	efflux RND transporter periplasmic adaptor subunit	
<i>mxgB</i> (-)	1048	WP_013374292.1 (99/88)	acriflavine resistance protein B	
<i>mxgC</i> (-)	420	WP_094306277.1 (99/57)	glycosyltransferase	GT
<i>mxgD</i> (+)	492	WP_120584626.1 (99/66)	unknown	
<i>mxgE</i> (-)	745	WP_043394394.1 (99/71)	S9 family peptidase	
<i>mxgF</i> (-)	368	WP_108068418.1 (99/84)	aldo/keto reductase	
<i>mxgG</i> (-)	464	WP_120613650.1 (96/88)	cytochrome P450	
<i>mxgH</i> (+)	455	WP_120613649.1 (100/73)	Monoxygenase / C-domain	
<i>mxgI</i> (-)	1509	WP_089130892.1 (99/40)	Polyketide synthetase	FAAL; ACP; <sub>L</sub> CL; Ox
<i>mxgJ</i> (-)	1082	WP_084668121.1 (99/58)	Non-ribosomal peptide synthetase	<sub>L</sub> CL; A; PCP
<i>mxgK</i> (-)	2242	WP_120613648.1 (100/81)	Polyketide synthetase	KS; AT; KR; DH; ACP; E
<i>mxgL</i> (-)	490	WP_120643780.1 (92/82)	non-ribosomal peptide synthetase	COM; <sub>D</sub> CL
<i>mxgM</i> (-)	827	WP_120643780.1 (99/83)	non-ribosomal peptide synthetase	A; PCP; TE
<i>mxgN</i> (-)	877	WP_120615411.1 (100/76)	Luciferae like oxidoreductase	<sub>D</sub> CL; Ox
<i>mxgO</i> (-)	235	WP_052518234.1 (99/76)	lipase	
<i>mxgP</i> (+)	209	WP_120206138.1 (95/75)	CoA pyrophosphatase	
<i>mxgQ</i> (+)	297	WP_046716342.1 (100/87)	undecaprenyl-diphosphate phosphatase	
<i>mxgR</i> (-)	156	WP_120205724.1 (100/90)	unknown	
<i>mxgS</i> (+)	415	WP_121718591.1 (99/85)	glycosyltransferase	GT

<sup>a</sup> Sizes are in amino acids; <sup>b</sup> Accession numbers and percentage of identity/similarity are given in parentheses.

### 3.4 Bioactivity

The isolated myxoglucamide variants were tested against an array of gram positive (*B. subtilis*, *S. aureus*), and fungal (*M. hiemalis*, *P. anomala*) test strains in concentrations up to 64 µg/ml but no antimicrobial activity was found. In addition myxoglucamides 661 and 647 were tested against gram negative *E. coli* WT and *E. coli* acrB in concentrations up to 64 µg/ml but no antimicrobial activity was found. Against the tested cell lines (KB 3.1 and HCT 116) moderate cytotoxic activity was found.

Table 6: Cytotoxicity test results

	Myxoglucamide 645 [µg/ml]	Myxoglucamide 647 [µg/ml]	Myxoglucamide 659 [µg/ml]	Myxoglucamide 661 [µg/ml]
KB 3.1 IC <sub>50</sub>	> 111	n.d.	111	n.d.
HCT 116 IC <sub>50</sub>	100	4.15	100	85.13

### 3.5 Conclusions

Here, the discovery of a new glycolipopeptide by application of a genome mining approach is reported. Two promoters were inserted in front of a putative BGC. Comparison of mutants and wild type by LC-MS revealed production of a group of new molecules. These molecules were isolated and the structures were determined using 2D NMR experiments. Remarkable features of the structures are a glycosylated iso-fatty acid chain linked via an amide bond to a vinylic moiety. For one isolated myxoglucamide molecule, the absolute configuration was determined experimentally for every stereocenter. In combination with these data and further experimental results, a prediction for the stereocenters in the other isolated myxoglucamide molecules was mostly possible.

The biosynthesis was examined by *in silico* analysis and feeding experiments and a biosynthetic schema was created. So far, two open questions remain towards the biosynthesis: The ω-hydroxylation step of the iso-fatty acid chain remains elusive and origin and formation of the vinylic moiety cannot be explained so far. The myxoglucamides were tested for bioactivity but no antimicrobial activity was found. Two myxoglucamides showed cytotoxic activity in higher concentrations.

### 3.6 References

- [1] K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema, T. Weber, *Nucleic Acids Res.* **2019**, W81-W87.
- [2] D. Krug, G. Zurek, O. Revermann, M. Vos, G. J. Velicer, R. Müller, *Appl. Environ. Microbiol.* **2008**, *74*, 3058.
- [3] O. N. Sekurova, O. Schneider, S. B. Zotchev, *Microbial Biotechnology* **2019**, *12*, 828.
- [4] T. Hoffmann, D. Krug, N. Bozkurt, S. Duddela, R. Jansen, R. Garcia, K. Gerth, H. Steinmetz, R. Müller, *Nat. Commun.* **2018**, *9*, 803.
- [5] B. Kunze, N. Bedorf, W. Kohl, G. Höfle, H. Reichenbach, *J. Antibiot.* **1989**, *42*, 14.
- [6] N. N. Gerber, H. A. Lechevalier, *Appl. Microbiol.* **1965**, *13*, 935.
- [7] F. Panter, D. Krug, S. Baumann, R. Müller, *Chem. Sci.* **2018**, *9*, 4898.
- [8] G. R. Pettit, Y. Tang, J. C. Knight, *J. Nat. Prod.* **2005**, *68*, 974.

- [9] F. S. Tareq, J. H. Kim, M. A. Lee, H.-S. Lee, Y.-J. Lee, J. S. Lee, H. J. Shin, *Organic letters* **2012**, *14*, 1464.
- [10] T. R. Hoye, C. S. Jeffrey, F. Shao, *Nat. Protoc.* **2007**, *2*, 2451.
- [11] C. R. Cantor, P. R. Schimmel, *2: Techniques for the study of biological structure and function*, Freeman, New York, **1980**.
- [12] R. Bhushan, H. Bruckner, *Amino Acids* **2004**, *27*, 231.
- [13] S. El-Gebali, J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J. Richardson, G. A. Salazar, A. Smart et al., *Nucleic acids research* **2019**, *47*, D427-D432.
- [14] L. Zimmermann, A. Stephens, S.-Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A. N. Lupas, V. Alva, *Journal of molecular biology* **2018**, *430*, 2237.
- [15] J. S. Dickschat, H. B. Bode, R. M. Kroppenstedt, R. Müller, S. Schulz, *Org. Biomol. Chem.* **2005**, *3*, 2824.
- [16] A. Greule, J. E. Stok, J. J. de Voss, M. J. Cryle, *Nat. Prod. Rep.* **2018**.
- [17] T. Yasuta, S. Okazaki, H. Mitsui, K. Yuhashi, H. Ezura, K. Minamisawa, *Applied and environmental microbiology* **2001**, *67*, 4999.
- [18] D. B. Berkowitz, B. D. Charette, K. R. Karukurichi, J. M. McFadden, *Tetrahedron, asymmetry* **2006**, *17*, 869.
- [19] J. B. Patteson, Z. D. Dunn, B. Li, *Angew. Chem. Int. Ed.* **2018**.
- [20] M. Fernández, Y. Cuadrado, J. F. Aparicio, J. F. Martín, *Microbiology* **2004**, *150*, 1467.
- [21] D. Chen, Q. Zhang, P. Cen, Z. Xu, W. Liu, *Appl. Environ. Microbiol.* **2012**, *78*, 5093.
- [22] N. S. Cortina, D. Krug, A. Plaza, O. Revermann, R. Müller, *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 811.
- [23] L. L. Lairson, B. Henrissat, G. J. Davies, S. G. Withers, *Annual review of biochemistry* **2008**, *77*, 521.

### 3.7 Supplementary information

#### 3.7.1 Experimental procedures

##### Strains

strains	suborder	family	genus	source
MCy 9003	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	Isolated by Dr. Ronald Garcia from a soil sample from Philippines
MCy 9003 mutant tn5 NRPS-t1pks 1	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant tn5 t1pks 8	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant tn5 NRPS 12	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant tn5 NRPS-t1pks 18	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant tn5 NRPS-t1pks 24	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant tn5 NRPS-t1pks 25	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant tn5 NRPS-t1pks 31	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant tn5 NRPS-t1pks 33	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant tn5 NRPS-t1pks 34	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant van NRPS-t1pks 35	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar

##### Media

###### CyS-agar

Casitone	0.25 %
Yeast extract	0.1 %
Soluble starch	0.1 %
HEPES	10 mM
CaCl <sub>2</sub> ·2H <sub>2</sub> O	0.1 %
Agar	1.5 %

The ingredients were dissolved in MiliQ H<sub>2</sub>O. The pH was adjusted to 7.0 with KOH.

###### CyS media

Casitone	0.25 %
Yeast extract	0.1 %
Soluble starch	0.1 %
HEPES	10 mM
CaCl <sub>2</sub> ·2H <sub>2</sub> O	0.1 %

The ingredients were dissolved in MiliQ H<sub>2</sub>O. The pH was adjusted to 7.2 with KOH.

## General analytic procedures

### Standard LC-MS analysis

#### amazon

The amaZon is an Ion trap MS/MS mass spectrometer manufactured by Bruker Daltonics coupled with a Dionex HPLC from Thermo Scientific. This machine was used for standard analytics using following template gradient programs:

#### 6 min gradient:

The following reversed phase HPLC methods were used for the mass spectrometry coupled analysis: Column: ACQUITY BEH 50 x 2.1 mm, 1.7  $\mu\text{m}$ , 130  $\text{\AA}$ , a flow rate of 0.6 ml/ min and a column temperature of 45 °C. Following gradient was applied with H<sub>2</sub>O + 0.1 % FA as eluent A and ACN + 0.1 % FA as eluent B. 0-1 min: 5 % Eluent B; 1-7 min: linear increase of eluent B from 5 to 95 %; 7-8.5 min: 95 % eluent B; 8.5-9 min: linear decrease of eluent B from 95 to 5 % B; 9-11.5 min: re-equilibration with 5 % eluent B. The detection was usually performed in the positive ESI MS mode, using a scan range from 200-2000 m/z.

#### 9 min gradient:

The following reversed phase HPLC methods were used for the mass spectrometry coupled analysis: Column: ACQUITY BEH 50 x 2.1 mm, 1.7  $\mu\text{m}$ , 130  $\text{\AA}$ , a flow rate of 0.6 ml/ min and a column temperature of 45 °C. Following gradient was applied with H<sub>2</sub>O + 0.1 % FA as eluent A and ACN + 0.1 % FA as eluent B. 0-0.5 min: 5 % Eluent B; 0.5-9.5 min: linear increase of eluent B to 95 %; 9.5-10.5 min: 95 % eluent B; 10.5-11 min: linear decrease of eluent B; 11-12.5 min: re-equilibration with 5 % eluent B. The detection was usually performed in the positive ESI MS mode, using a scan range from 200-2000 m/z.

#### 18 min gradient:

The following reversed phase HPLC methods were used for the mass spectrometry coupled analysis: Column: ACQUITY BEH 50 x 2.1 mm, 1.7  $\mu\text{m}$ , 130  $\text{\AA}$ , a flow rate of 0.6 ml/ min and a column temperature of 45 °C. Following gradient was applied with H<sub>2</sub>O + 0.1 % FA as eluent A and ACN + 0.1 % FA as eluent B. 0-0.5 min: 5 % Eluent B; 0.5-18.5 min: linear increase of eluent B to 95 %; 18.5-20.5 min: 95 % eluent B; 20.5-20.8 min: linear decrease of eluent B; 20.8-22.5 min: re-equilibration with 5 % eluent B. The detection was usually performed in the positive ESI MS mode, using a scan range from 200-2000 m/z.

#### maXis 4G

The maXis 4G is a high resolution TOF mass spectrometer manufactured by Bruker Daltonics coupled with a Dionex HPLC from Thermo Scientific. This machine was used for high resolution LC-MS measurements. Following template method was used:

Coloumn: ACQUITY BEH 100 x 2.1 mm 1.7  $\mu\text{m}$  130  $\text{\AA}$ , a flow rate of 0.6 ml/ min and a column temperature of 45 °C. Following gradient was applied with H<sub>2</sub>O + 0.1 % FA as eluent A and ACN + 0.1 % FA as eluent B. 0-0.5 min: 5 % Eluent B; 0.5-18.5 min: linear increase of eluent B to 95 %; 18.5-20.5 min: 95 % eluent B; 20.5-20.8 min: linear decrease of eluent B; 20.8-22.5 min: re-equilibration with 5 % eluent B. The detection was usually performed in the positive MS mode or MS<sup>2</sup> mode, using a scan range from 150-2500 m/z.

#### Spectroscopic methods

NMR were acquired with a Bruker Ascend 700 NMR spectrometer equipped with a 5mm TCI cryoprobe and a Bruker UltraShield 500 NMR spectrometer equipped with a 5mm TCI cryoprobe. Optical rotation was measured using a Jasco polarimeter. CD spectra were acquired using a Jasco J-1500 CD spectrometer.

#### **Cultivation of MCy 9003 in liquid media**

MCy 9003 was cultivated in a flask at 30 °C with 180 rpm on a shaker. The cultures were either inoculated from a liquid pre-culture or from a piece of agar from an already grown agar culture by cutting out the agar piece under sterile conditions and dropping it into the flask with media. For extract preparation the adsorber resin XAD-16 was added to the culture one day after inoculation. The cultures were incubated between 9 and 12 days.

#### **Cultivation of MCy 9003 on Agar**

The Agar plate was inoculated with a piece from a grown Agar plate or a drop of liquid media and cultivated at 30 °C. For Agar plates VY/2-Agar, Cy-Agar or CyS-Agar was used.

#### **Genome sequencing**

Genomic DNA was isolated by Dr. Ronald Garcia and Irene Cochems and submitted to pacBio Inc. for sequencing. The raw sequencing were assembled by Nestor Zaburanyi

#### **Biological testing**

The testing of all crude extracts, Sephadex fractions, HPLC fractions and pure compounds were performed by Viktoria Schmitt, Irene Cochems and Stefanie Schmidt according to existing protocols.<sup>[1]</sup>

#### **Cultivation of mutants**

MCy 9003 mutants were cultivated in a baffled 300 ml flask at 30 °C with 180 rpm on a shaker in 100 ml CyS media with 75 µg/ml Kanamycin. The cultures were either inoculated from a liquid pre-culture or from a piece of agar from an already grown agar culture by cutting out the agar piece under sterile conditions and dropping it into the flask with media. For extract preparation the adsorber resin XAD-16 was added to the culture one day after inoculation. The cultures were incubated between 9 and 12 days.

#### **Preparation of crude extract from MCy 9003**

XAD and cells were frozen at -20 °C and lyophilized. After lyophilization metabolites were extracted with 50 ml MeOH/acetone 1:1 under constant shaking for 2.5 h. The supernatant was filtrated into a round flask and the solvent was evaporated under reduced pressure and the residue was dissolved in MeOH and transferred into a 4 ml Vial and stored at 20 °C.

#### **Isolation of Myxoglucamides**

The molecules were isolated from a 9 L culture of MCy 9003 mutant tn5 NRPS-t1pks 18 in CyS media. The cells and the adsorber resin XAD-16 were extracted three times with 800 ml acetone/ MeOH 1:1. The solvents were poured through a funnel stuffed with glass wool and the solvents were evaporated. The dried residues were

dissolved in 500 ml H<sub>2</sub>O and extracted twice with 400 ml ethyl-acetate in a separatory funnel. Myxoglucamides were mostly found in the aqueous layer. The aqueous layer was dried *in vacuo* and further fractionated using reverse phase flash chromatography. Following setup was used:

Biotage with a 25 g Snapfit column, Versa Flash Spherical C18 silica 45-75  $\mu\text{m}$  70 Å, Solvent System: A=H<sub>2</sub>O + 0.1 % FA ; B= MeOH + 0.1 FA; A flow rate of 20 ml/min was used and 50 ml were collected per tube. The column was equilibrated with with 3 column volumes (CV) 95:5 Solvent A/B. The sample was loaded on the column using iSolute beads. Following gradient was used: 4 CV 95:5 Solvent A/B; 40 CV linear increases from 5-100 % Solvent B; 5 CV 100 % Solvent B.

An aliquot of every fraction (including the waste from the flushes) was taken and measured by LC-MS to screen for the myxoglucamides. Fractions containing myxoglucamides were dried and further purified by HPLC. Following setup was used:

Dionex Ultimate 3000 coupled with Bruker High capacity trap mass spectrometer (HCT); Column: Luna C18(2) 250 x 10 mm, 5  $\mu\text{m}$ ; flowrate 5 ml/min and column temperature 40 °C with H<sub>2</sub>O + 0.1 % FA as eluent A and ACN + 0.1% FA as eluent B. The following gradient was applied: 0-2 min 5 % eluent B; 2-4 min linear increase of eluent B to 43 %; 4-24 min linear increase of eluent B to 46 %; 24-25 min linear increase of eluent B to 95 % eluent B; 25-28 min 95 % eluent B 28-29 min linear decrease of eluent B to 5 %; 29-33 min re-equilibration with 5 % eluent B. Myxoglucamid 647 was collected from 10.8-11.1 min; Myxoglucamid 661 was collected from 13.0-13.4 min; Myxoglucamid 645 was collected from 14.4-14.85 min and Myxoglucamid 659 was collected from 18.0-18.7 min. The collected fractions were dried and dissolved in MeOH

#### **Acid Hydrolysis of Myxoglucamide 645**

3.7 mg myxoglucamide 645 were dissolved in 1N HCl and were incubated at 100 °C for 26 h. The Hydrolysis mixture was left to cool down to room temperature and extracted first with n-hexane followed by ethyl-acetate. The remaining aqueous layer was mixed with MeOH and dried under nitrogen flow. All extracts were controlled by LC-MS and NMR to observe hydrolysis products.

#### **Mosher Ester formation 13-Hydroxy-14-Methylpentadecenoic acid (HMA)**

13-Hydroxy-14 methylpentadecenoic acid was isolated from the hydrolysis mixture by extraction with n-hexane.

The (*S*)-MTPA ester was prepared as followed:

100  $\mu\text{g}$  of HMA (0.37  $\mu\text{M}$ ) were dissolved in 120  $\mu\text{l}$  pyridine *d*<sub>5</sub> and 1.5  $\mu\text{l}$  (*R*)-MTPA-Cl (9.6  $\mu\text{M}$ ) were added to the reaction mixture. The reaction progress was observed via LC-MS. The expected product mass 487.30 [M-H]<sup>-</sup> was observed in ESI negative mode. After 3.5 h 450  $\mu\text{l}$  pyridine *d*<sub>5</sub> were added to the reaction mixture and the reaction mixture was transferred into a NMR tube.

The (*R*)-MTPA ester was prepared as followed:

120  $\mu\text{g}$  of HMA (0.44  $\mu\text{M}$ ) were dissolved in 120  $\mu\text{l}$  pyridine *d*<sub>5</sub> and 1.5  $\mu\text{l}$  (*S*)-MTPA-Cl (9.6  $\mu\text{M}$ ) were added to the reaction mixture. The reaction progress was observed via LC-MS. The expected product mass 487.30 [M-H]<sup>-</sup> was observed in ESI negative mode. After 5 h 450  $\mu\text{l}$  pyridine *d*<sub>5</sub> were added to the reaction mixture and the reaction mixture was transferred into a NMR tube.

#### **Marfey's Analysis**

Between 50 and 200  $\mu\text{g}$  compound were placed in a 2 ml glass vial. 100  $\mu\text{l}$  6N HCl were added and the sample was heated to 110 °C for 45 min under nitrogen atmosphere. In a next step the sample was dried and dissolved in 110  $\mu\text{l}$  H<sub>2</sub>O. 50  $\mu\text{l}$  of the sample were transferred into two 1.5 ml plastic tubes. 20  $\mu\text{l}$  1N NaHCO<sub>3</sub> with pH 9 were added to each tube followed by addition of 1 % Marfey's reagent in acetone (D-FDLA and L-FDLA). The reaction mix was incubated for 2 h under constant shaking with 700 rpm. The reaction was stopped by addition of 10  $\mu\text{l}$  2 N HCl. After the reaction was stopped 300  $\mu\text{l}$  ACN were added and the sample

was centrifuged and an aliquot of the supernatant was measured by LC-*hr*MS analysis using maXis4G analytical setup. Standards were handled in the similar manner skipping the hydrolysis with HCl. Following gradient was used: Column: ACQUITY BEH 100 x 2.1 mm 1.7  $\mu$ m 130 Å flowrate 0.6 ml/min and column temperature 45 °C with H<sub>2</sub>O + 0.1 % FA as eluent A and ACN + 0.1% FA as eluent B. % Eluent B; 0-1 min: linear increase of eluent B to 10 %; 1-15 min: linear increase of eluent B to 35 %; 15-22 min: linear increase of eluent B to 55 %; 15-22 min: linear increase of eluent B to 55 %; The detection was performed in positive MS mode.

### **Ozonolysis of Myxoglucamid 645**

200  $\mu$ g Myxoglucamid 645 dissolved in 10 ml MeOH were placed in a two necked flask and cooled down to -70°C using an acetone bath with dry ice. Ozone was introduced into the solution for 5 min and 1 ml 35 % H<sub>2</sub>O<sub>2</sub> were added to the mixture afterwards. The MeOH was removed *in vacuo* and aqueous remains were dried under N<sub>2</sub>. The reaction mixture was further analyzed using Marfey's method as described previously. As standard amino acid L-aspartate was used.

### **Feeding experiments**

The Feeding study was performed with MCy 9003 mutant tn5 NRPS-t1pks 18 in CyS media in 10 ml cultures in 50 ml flasks. Following isotope labelled substrates were used:

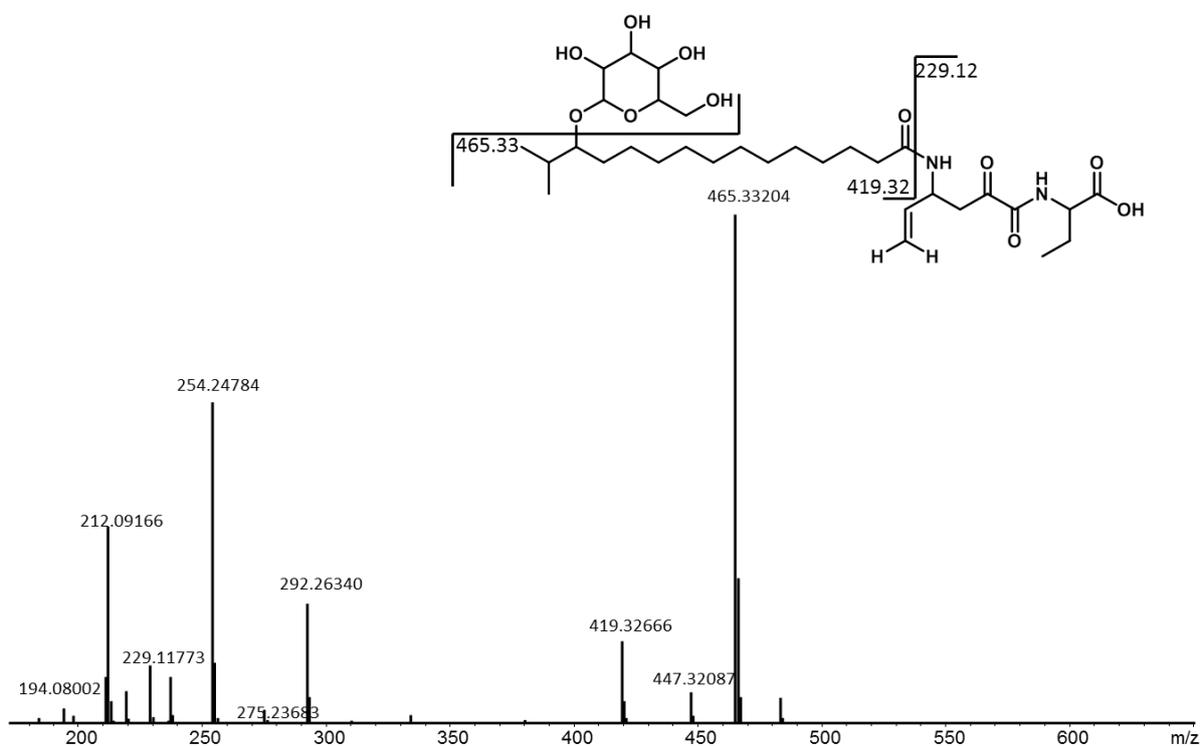
Isotope labelled substrates	concentration	Solvent for application
L -valine $d_8$	0.25 M	MeOH/dH <sub>2</sub> O 1:1
L -leucine $d_3$	0.25 M	
1- <sup>13</sup> C-acetat	0.5 M	
2- <sup>13</sup> C-acetat	0.5 M	
<sup>13</sup> C <sub>3</sub> -Glycine	0.5 M	
D / L aminobutyric acid $d_6$	0.25 M	
L methionine <sup>15</sup> N <sub>1</sub> <sup>13</sup> C <sub>6</sub>	0.25 M	
L -glutamine <sup>15</sup> N <sub>2</sub> <sup>13</sup> C <sub>5</sub>	0.1 M	
L -glutamic acid $d_5$	0.25 M	
L -aspartic acid <sup>13</sup> C <sub>4</sub> <sup>15</sup> N <sub>1</sub>	0.25 M	

To test substrate incorporation 10  $\mu$ l of the labelled substrate-solution was added to the cultures two times over two days. 24 h after last addition of substrates 500  $\mu$ l of XAD-16 suspension were added. 9 days after inoculation the cultures were harvested by centrifugation. The supernatant was discarded and the cells and the XAD were frozen and lyophilized before extraction as previously described. The dried extract was dissolved in 1 ml MeOH.

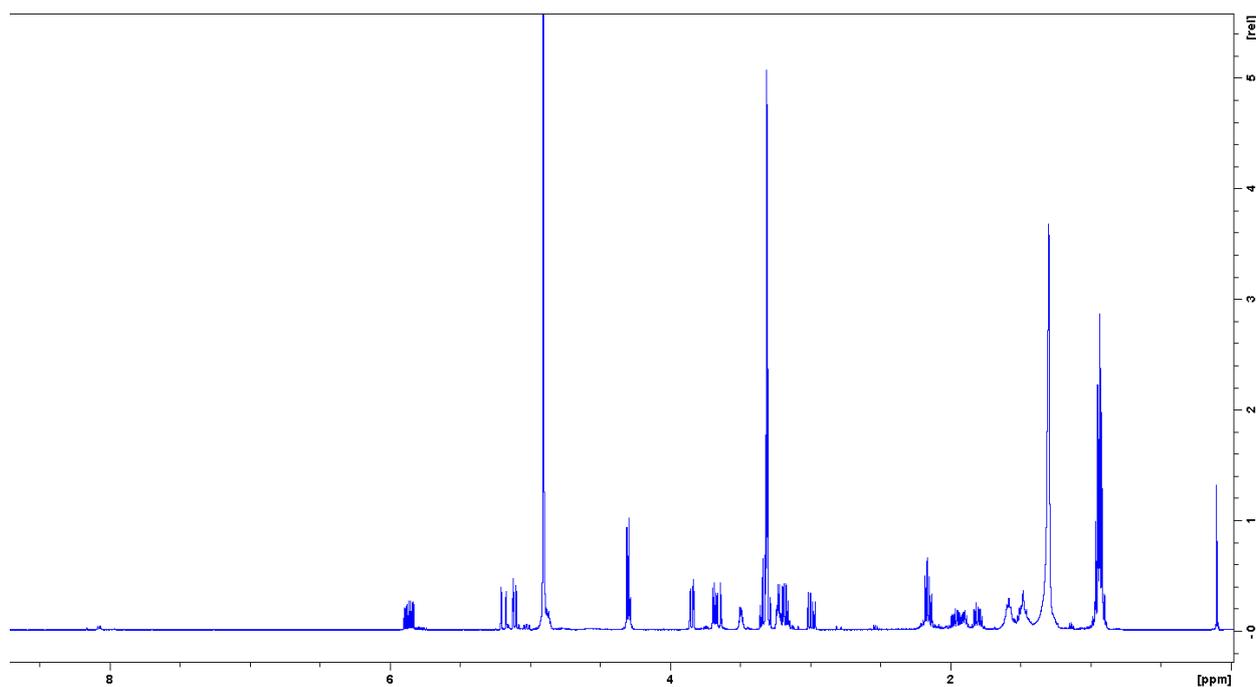
## **3.7.2 Supplementary data/additional experimental data**

### **Analytical data myxoglucamide 645**

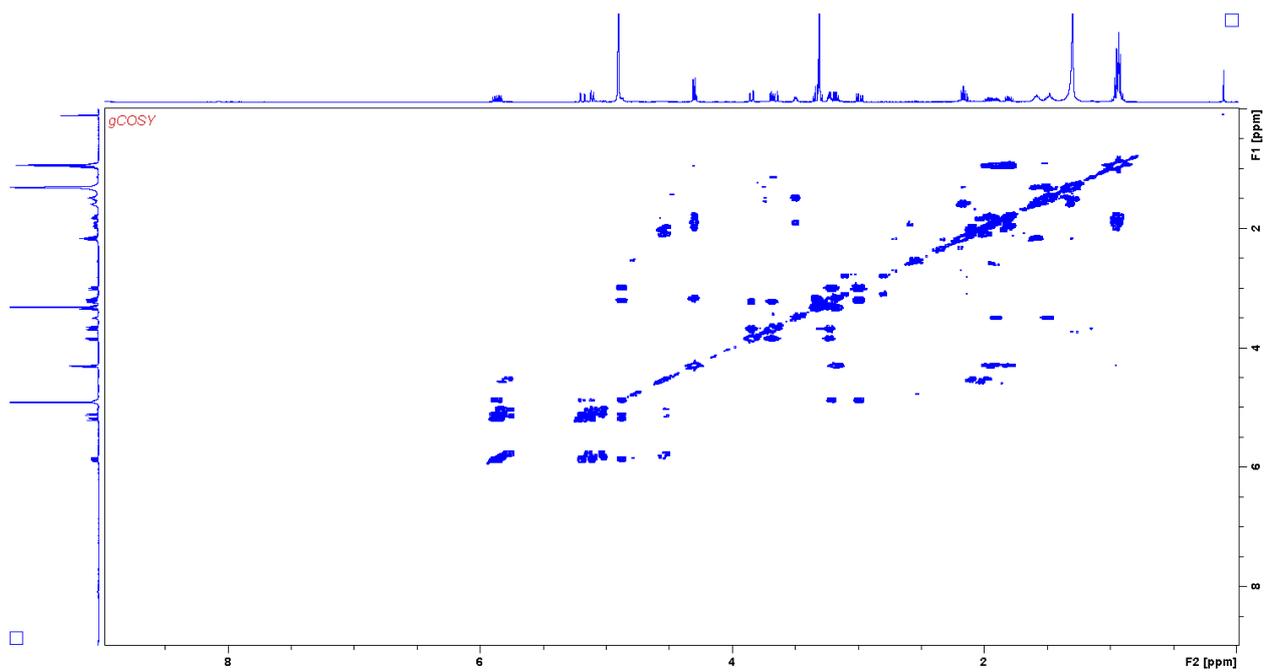
Myxoglucamid 645 was isolated as colorless amorphous solid. *hr*Mass: 645.39551[M+H]<sup>+</sup>; specific optical rotation  $[\alpha]_D^{20}$  -10.4 °g<sup>-1</sup>ml dm<sup>-1</sup>; UV max 197, 241 nm (UV data extracted from LC-MS-UV measurement using a Dionex ultimate 3000 HPLC system equipped with a DAD UV-VIS detector running with H<sub>2</sub>O +0.1 % FA and ACN + 0.1 % FA )



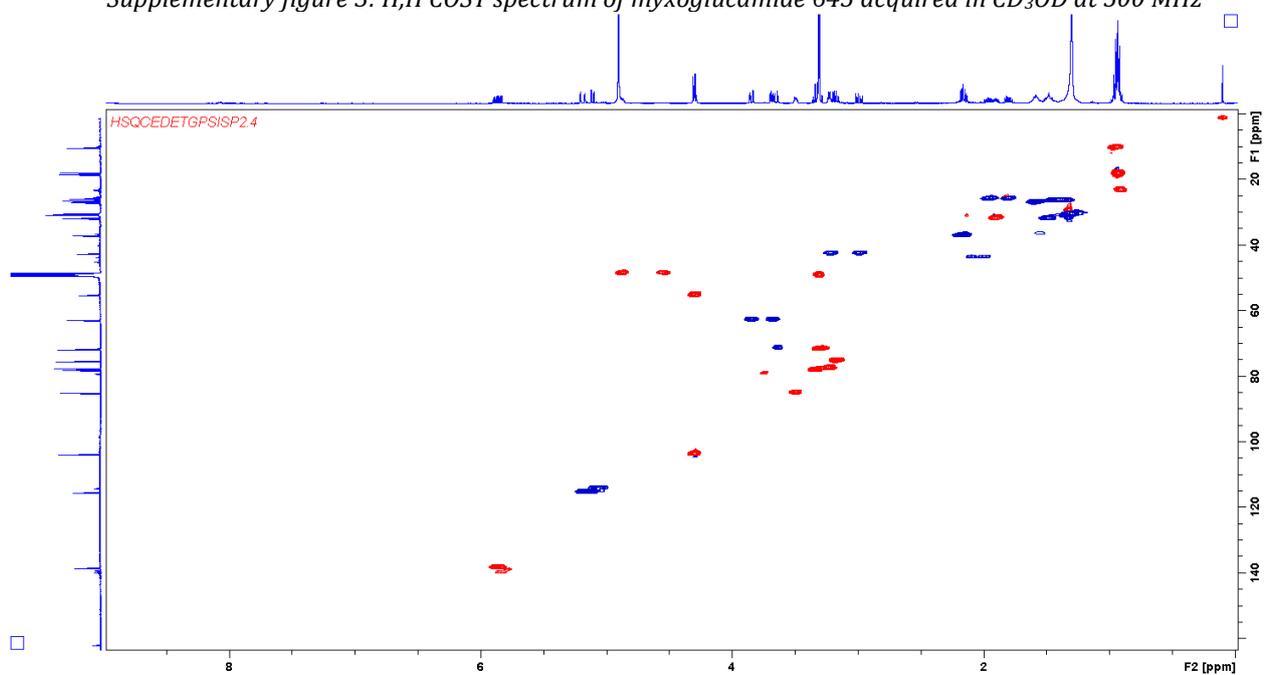
Supplementary figure 1: *hr*MS<sup>2</sup> spectrum and observed fragments of myxoglucamide 645. MS<sup>2</sup> spectrum is extracted from LC-*hr*MS measurement of an extract of MCy 9003 mutant *tn5* NRPS-*t1pks18*.



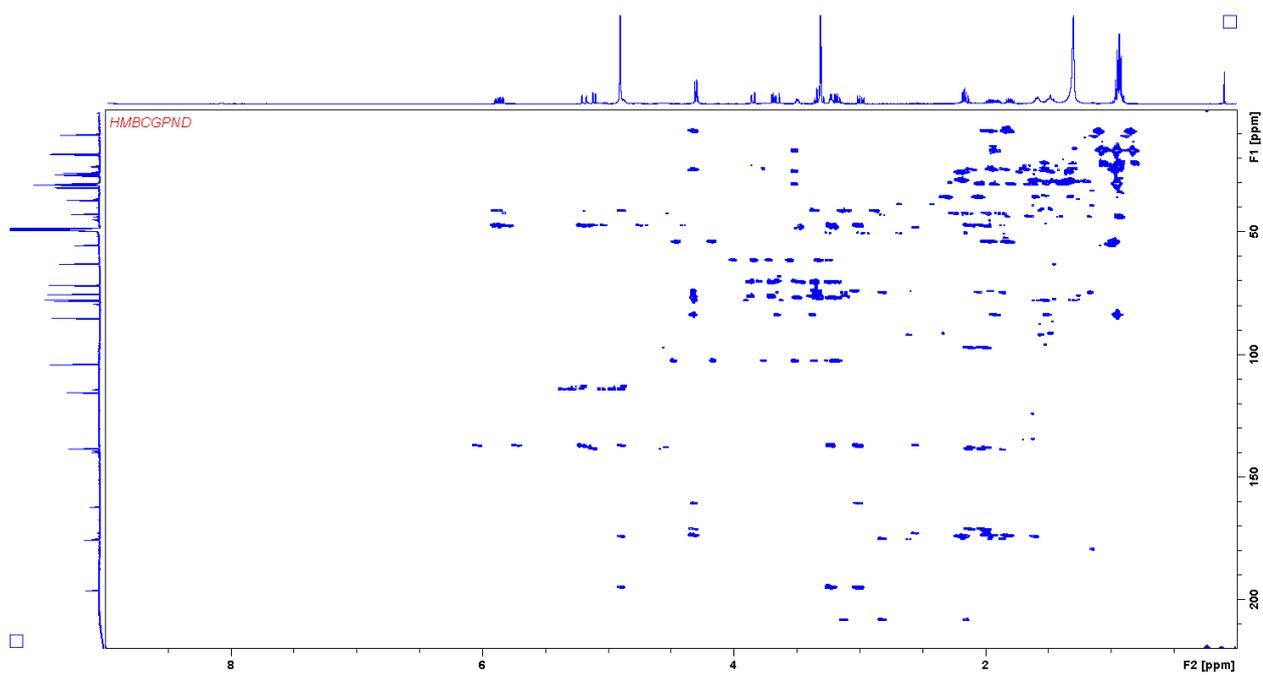
Supplementary figure 2: <sup>1</sup>H spectrum of myxoglucamide 645 acquired in CD<sub>3</sub>OD at 500 MHz



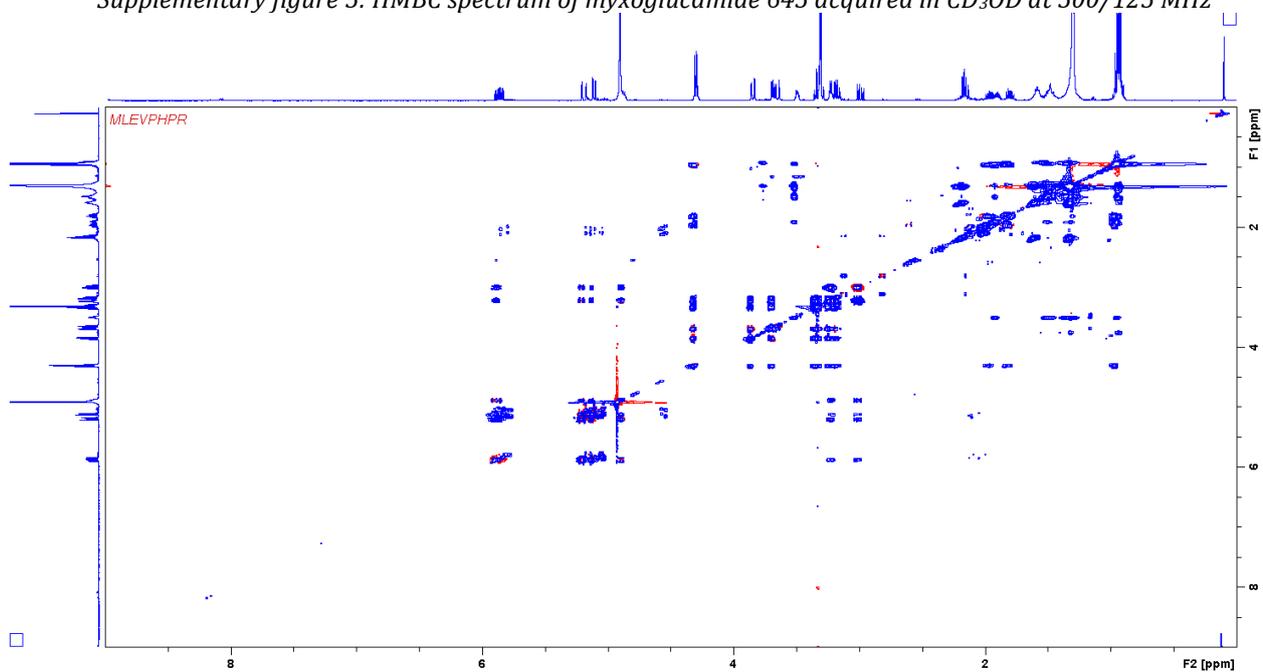
Supplementary figure 3: *H,H* COSY spectrum of myxoglucamide 645 acquired in CD<sub>3</sub>OD at 500 MHz



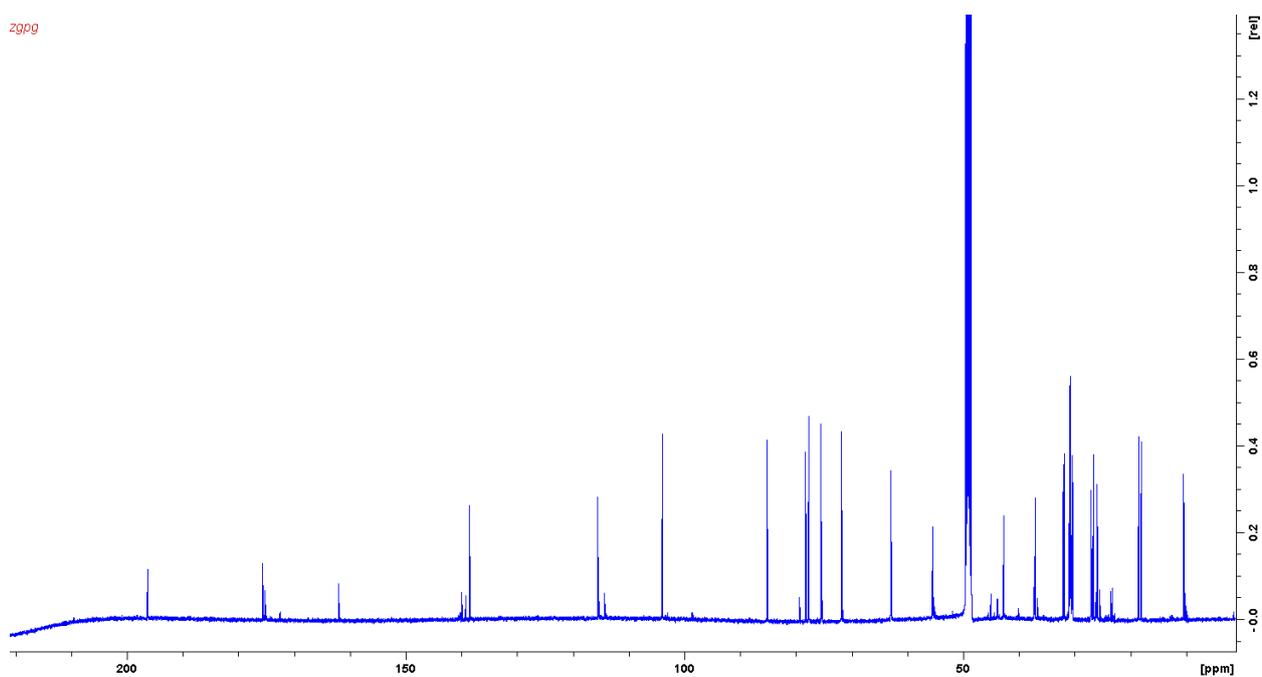
Supplementary figure 4: HSQC spectrum of myxoglucamide 645 acquired in CD<sub>3</sub>OD at 500/125 MHz



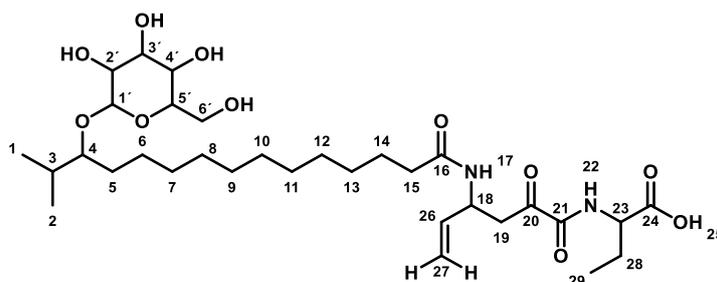
Supplementary figure 5: HMBC spectrum of myxoglucamide 645 acquired in  $CD_3OD$  at 500/125 MHz



Supplementary figure 6: H,H TOCSY spectrum of myxoglucamide 645 acquired in  $CD_3OD$  at 500 MHz



Supplementary figure 7:  $^{13}\text{C}$  spectrum of myxoglucamide 645 acquired in  $\text{CD}_3\text{OD}$  at 125 MHz



Supplementary table 1: Spectroscopic data of myxoglucamide 645 in  $\text{CD}_3\text{OD}$ .

position	$\delta_{\text{C}}^{\text{a}}$	$\delta_{\text{H}}^{\text{b}}$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
1	18.4	0.93 d 7.0	3	2,3
2	18.1	0.94 d 6.6	3	1,3
3	31.6	1.90 m	1,2,4	
4	85.2	3.49 m	3,5	1',1,2,3,4,5
5	31.8	1.49 m	6	
6	26.6	1.36 m	7-13	
7-13	30.3-31.0	1.27-1.32		
14	27.1	1.58 m	7-13,15	
15	37.1	2.16 t	14	14,16
16	175.7			
17	NH			
18	48.6	4.87 m	26	16,19,20,26
19a	42.7	2.99 d,d 16.8, 8.4	19b	18,20,21,26
19b		3.21 d,d 16.8, 5.3	19a	18,20,26
20	196.3			
21	162.0			
22	NH			
23	55.5	4.29 m	28	21,24,28,29

24	175.2			
25	OH			
26	138.6	5.86 d,d,d 5.5, 10.4, 5.5	18,27	18,19
27a	115.5	5.19 dt,1.4,17.2	27b,26	18,26
27b		5.11 dt 1.4, 10.4	27a, 26	18,26
28a	25.9	1.80 m		23,24,29
28b		1.96 m		23,24,29
29	10.5	0.95 t 7.4	23,28	
1'	103.4	4.30 d 7.8	2'	2',4',5'
2'	75.5	3.17 d,d	1',3'	1',4'
3'	78.3	3.32 m		
4'	71.7	3.31 <sup>e</sup>		
5'	77.7	3.22 m	6'a/b, b 5'	4'
6'a	63.0	3.85 d,d 11.6, 2.4	6'b, 5'	4',5'
6'b	63.0	3.68 d,d 11.5, 5.5	6'a,5'	4',5'

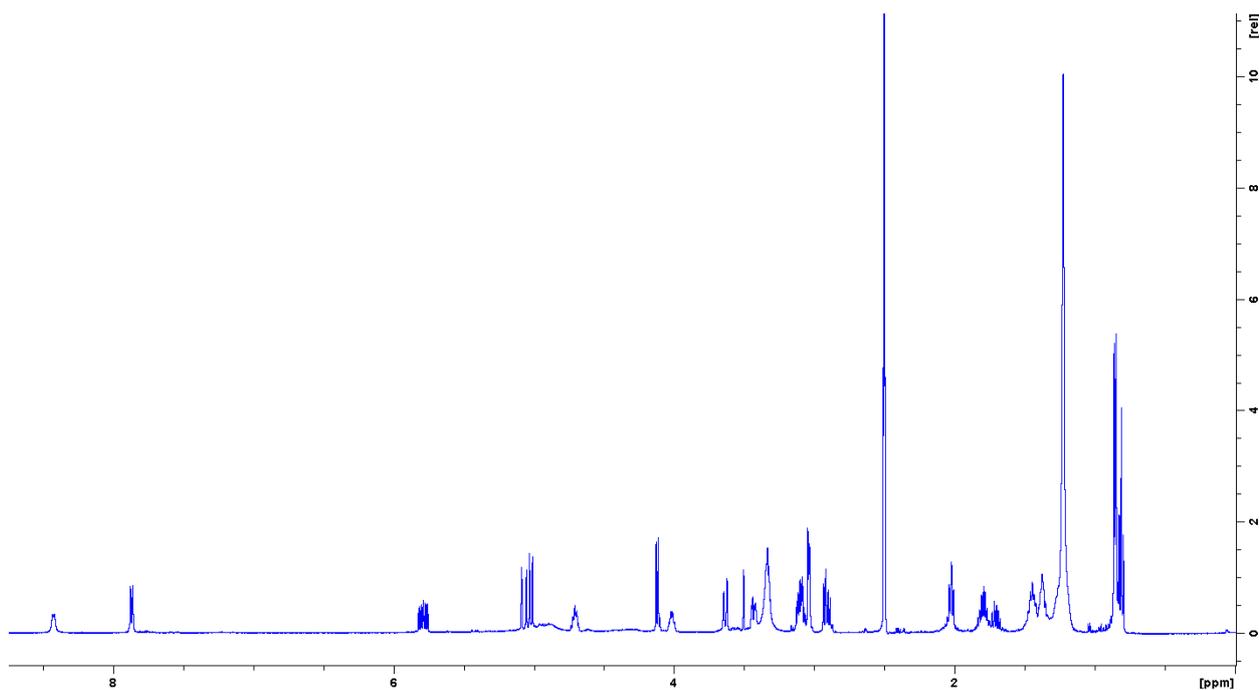
<sup>a</sup> acquired at 125 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

<sup>b</sup> acquired at 500 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

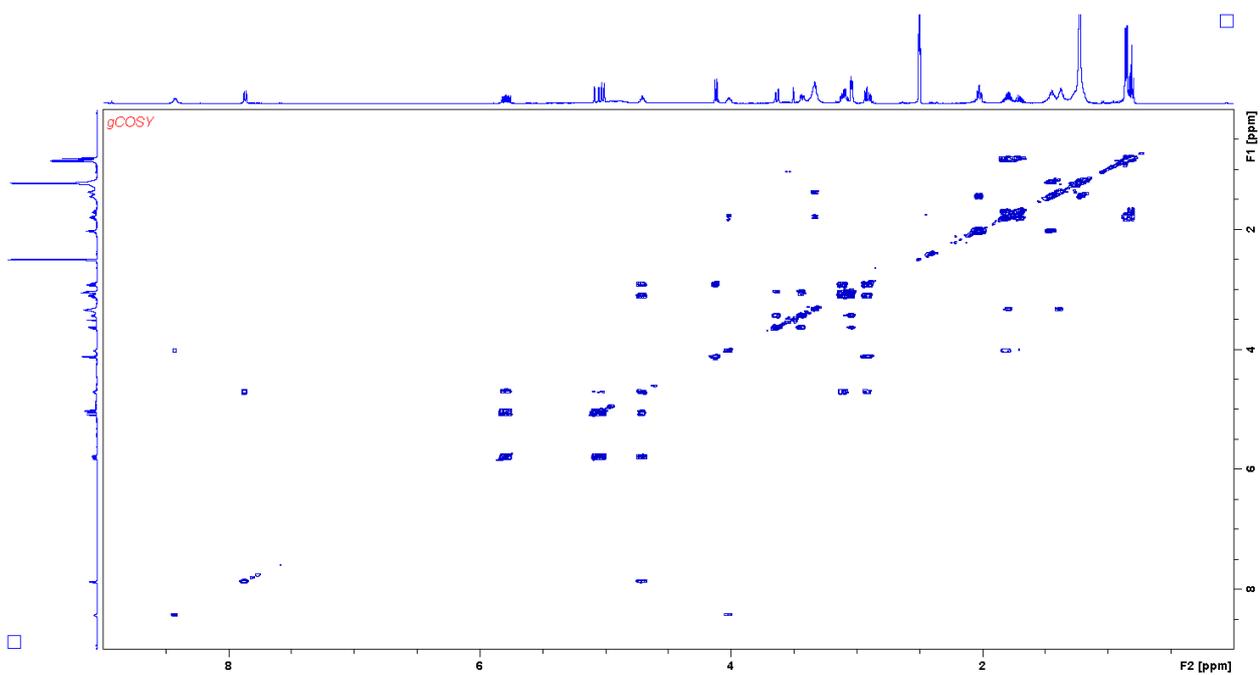
<sup>c</sup> proton showing COSY correlation to indicated proton.

<sup>d</sup> proton showing HMBC correlations to indicated carbons.

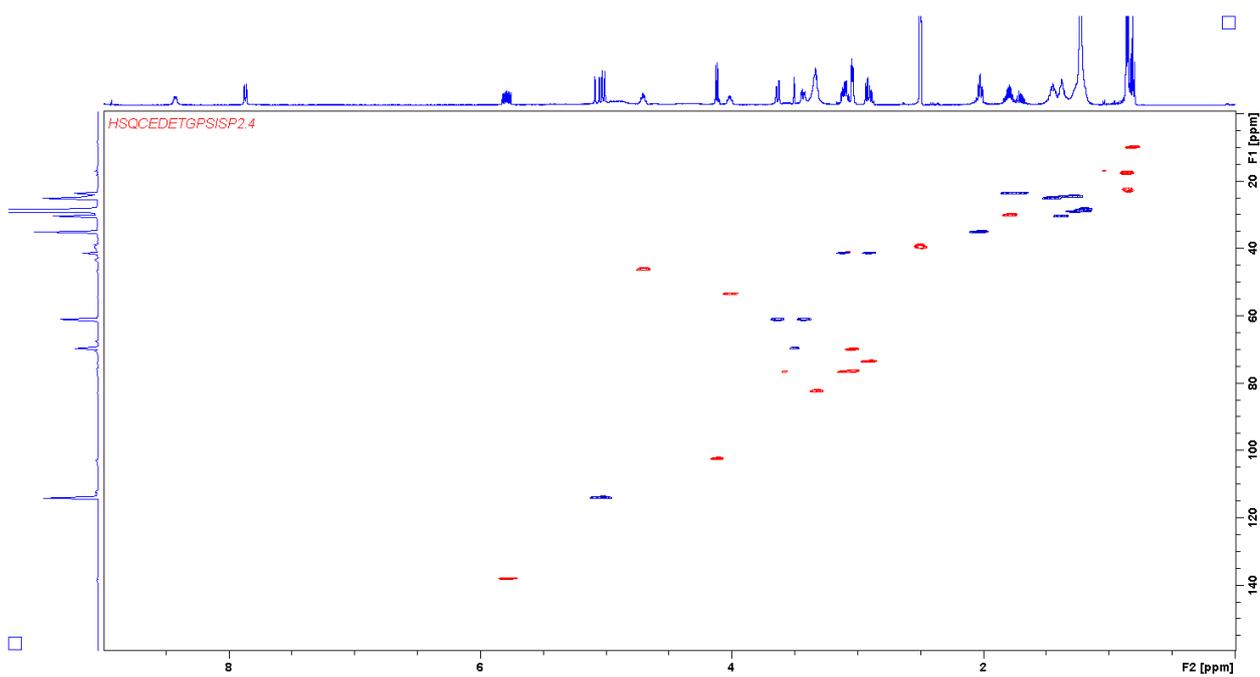
<sup>e</sup> overlap with solvent signal.



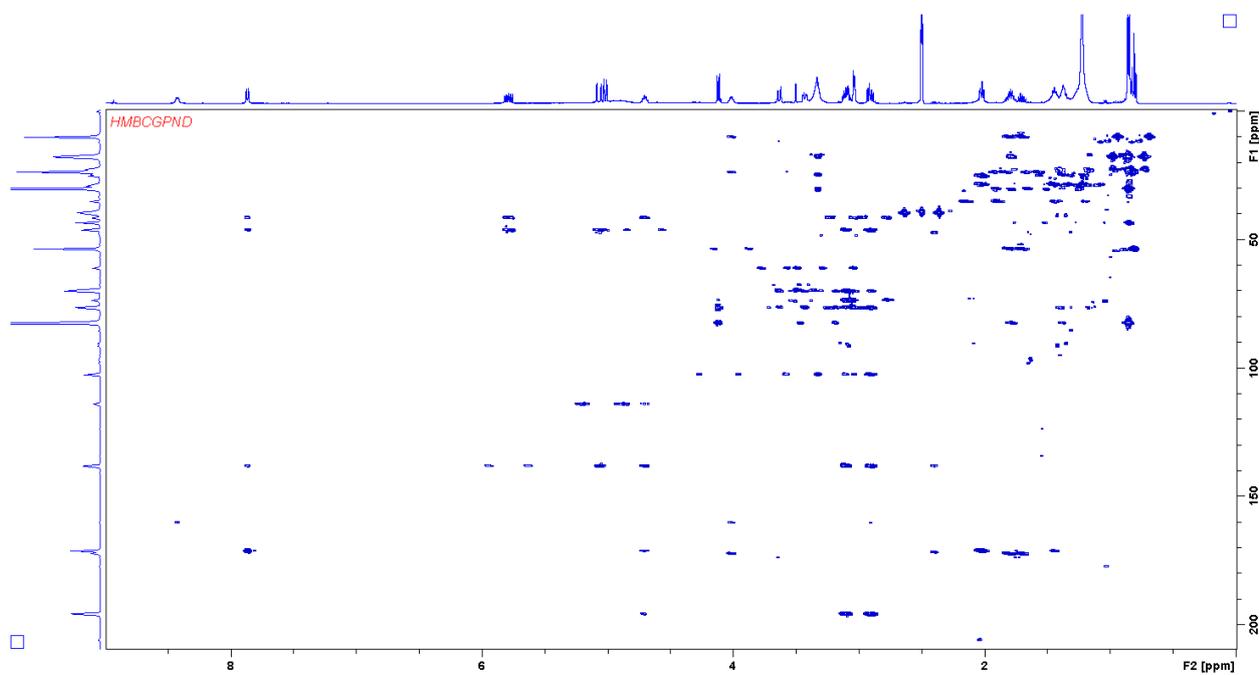
Supplementary figure 8: <sup>1</sup>H spectrum of myxoglucamide 645 acquired in DMSO d<sub>6</sub> at 500 MHz



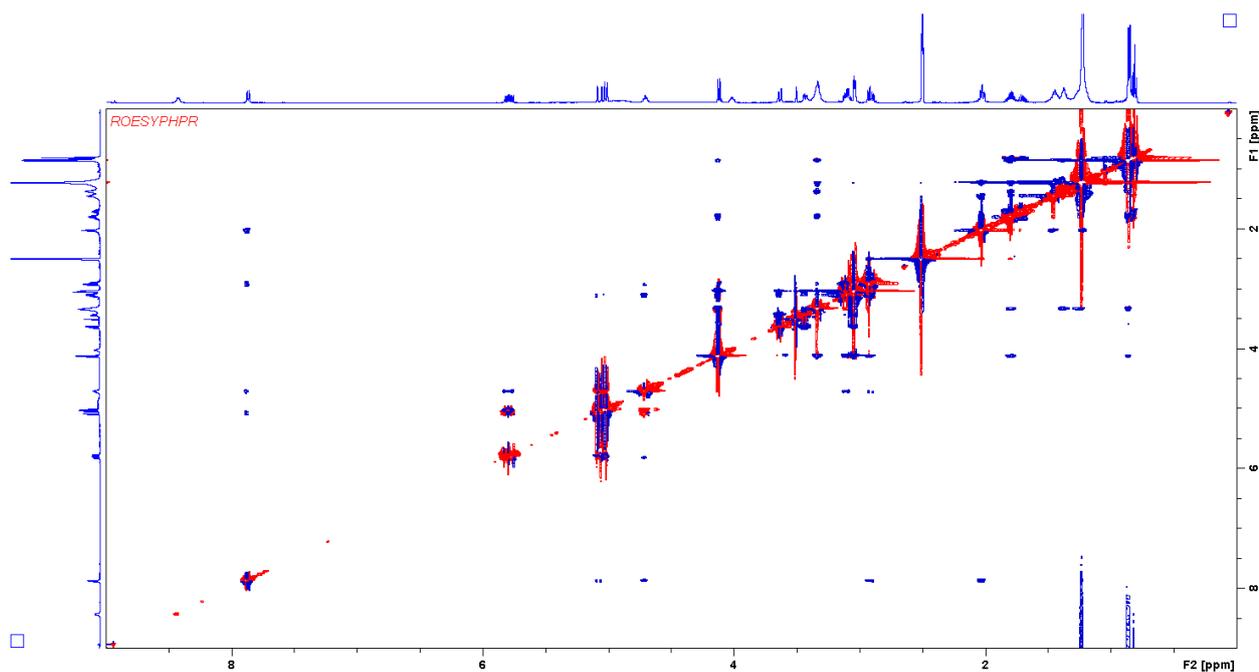
Supplementary figure 9:  $H,H$  COSY spectrum of myxoglucamide 645 acquired in  $DMSO d_6$  at 500 MHz



Supplementary figure 10: HSQC spectrum of myxoglucamide 645 acquired in  $DMSO d_6$  at 500/125 MHz



Supplementary figure 11: HMBC spectrum of myxoglucamide645 acquired in DMSO  $d_6$  at 500/125 MHz



Supplementary figure 12: ROESY spectrum of Myxoglucamide 645 acquired in DMSO  $d_6$  at 500 MHz

Supplementary table 2: Spectroscopic data of myxoglucamide 645 in DMSO  $d_6$ .

position	$\delta_c^a$	$\delta_H^b$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
1	17.6	0.86 d 6.7	3	2,3,4
2	17.6	0.85 d 6.8	3	1,3,4
3	30.0	1.79 m	1,2,4	1,2,4,5
4	82.6	3.33 m	3,5	1',1,2,3,6
5	24.7	1.35 m	5	

6-13	28.2-29.2	1.26-1.19		
14	25.1	1.45 m	15	
15	35.2	2.03 t 7.3	14	14,16
16	171.1			
17		7.87 d 8.1	18	16,18,19,26
18	46.2	4.71 m	17,19a/b, 26	16,19,20,26,27
19a	41.4	2.91 d,d 16.2,7.5	18 19b	18,20,21,26
19b		3.11 d,d 11.0,6.4	18, 19a	18,20,26
20	195.8			
21	160.1			
22		8.42 d 6.3	23	21,23
23	53.5	4.02 m	28a/b	21,24,28,29
24	172.2			
25				
26	138.1	5.79 d,d,d 17.2,10.2,5.3	27,18	18,19
27a	114.0	5.07 d,d,d 17.3,1.5,1.5	18,26,27b	18,19,26
27b		5.02 d,d,d 10.3,1.5,1.5	27a,26,18	18,19,26
28a	23.6	1.80 m	23,29	23,24,29
28b		1.70 m	23,29	23,24,29
29	9.9	0.81 t 7.3	28	23,28
1'	102.5	4.12 d 7.7	2'	2',3',5',4
2'	73.6	2.91 m	1',3'	1',3',4'
3'	76.7	3.11 m	2',4'	4'
4'	70.0	3.05 m	3',5'	
5'	76.4	3.03 m	4',6'	
6'a	61.1	3.64 m	5',6'b	4'
6'b		3.42 m	5',6'a	4'

<sup>a</sup> acquired at 125 MHz and assigned from 2D NMR spectra, referenced to solvent signal DMSO d<sub>6</sub> at δ 39.52 ppm.

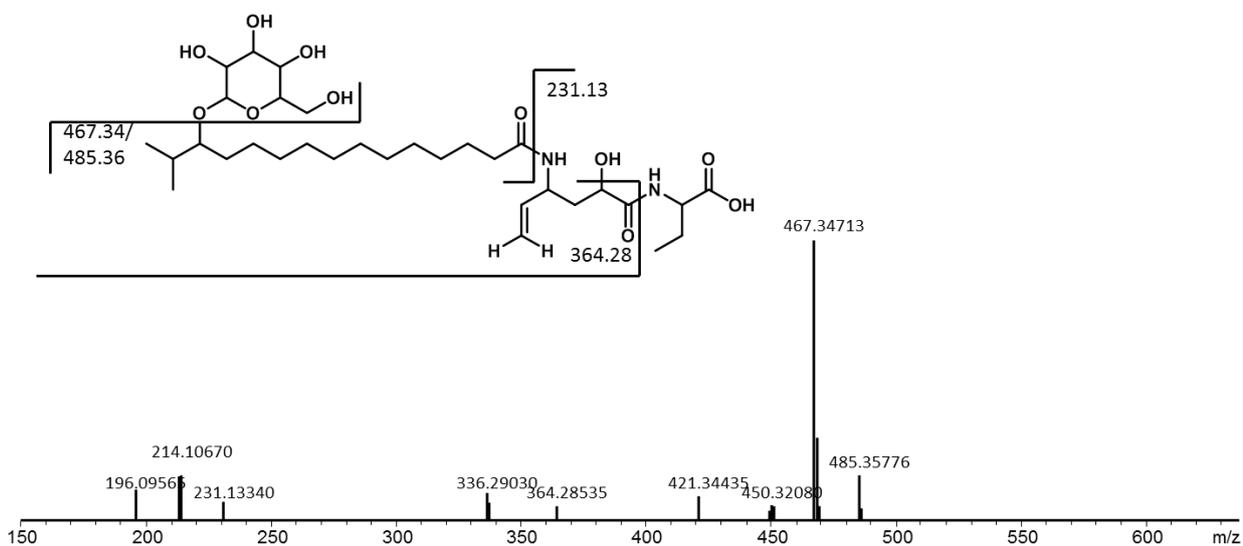
<sup>b</sup> acquired at 500 MHz, referenced to solvent signal DMSO d<sub>6</sub> at δ 2.50 ppm.

<sup>c</sup> proton showing COSY correlation to indicated proton.

<sup>d</sup> proton showing HMBC correlations to indicated carbons.

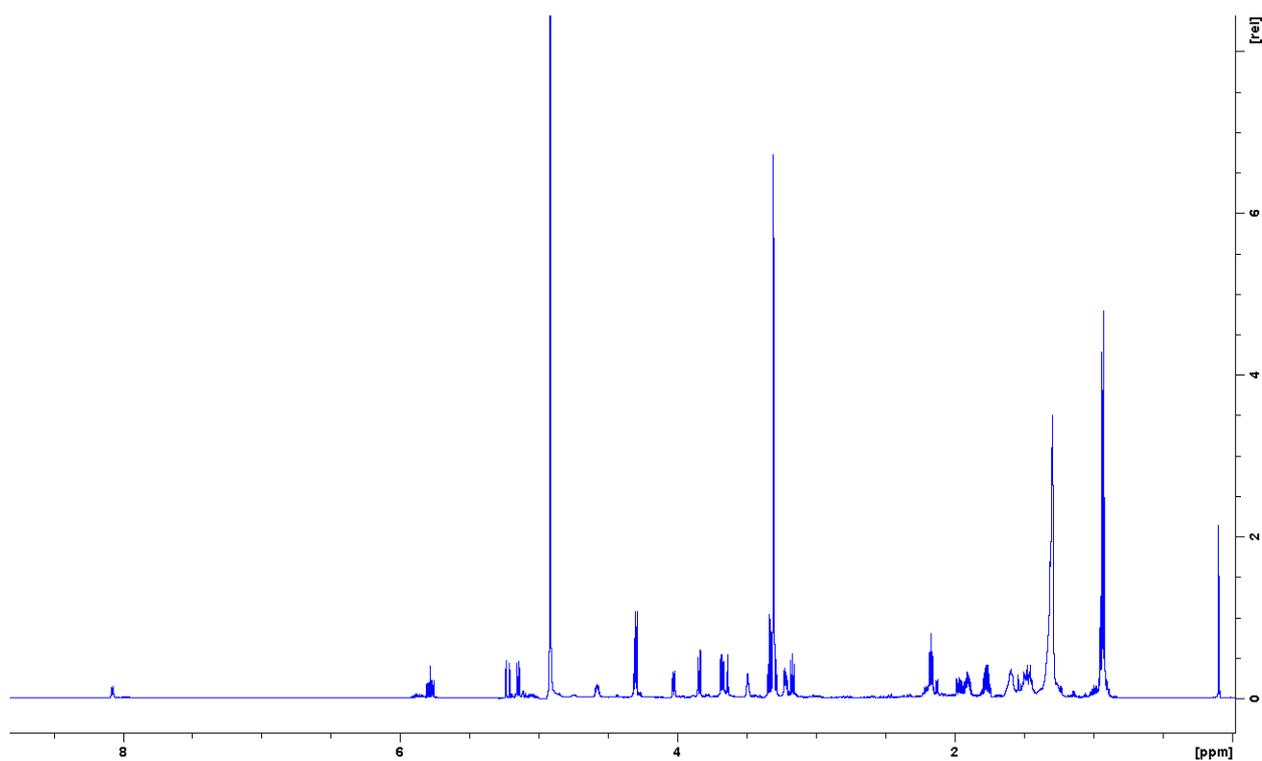
### Analytical data myxoglucamid 647

Myxoglucamid 647 was isolated as colorless amorphous solid. *hr*Mass: 647.41073 [M+H]<sup>+</sup>; specific optical rotation  $[\alpha]_D^{20}$  0.9 °g<sup>-1</sup>ml dm<sup>-1</sup>; UV max: 195 nm (UV data extracted from LC-MS-UV measurement using a Dionex ultimate 3000 HPLC system equipped with a DAD UV-VIS detector running with H<sub>2</sub>O +0.1 % FA and ACN + 0.1 % FA )

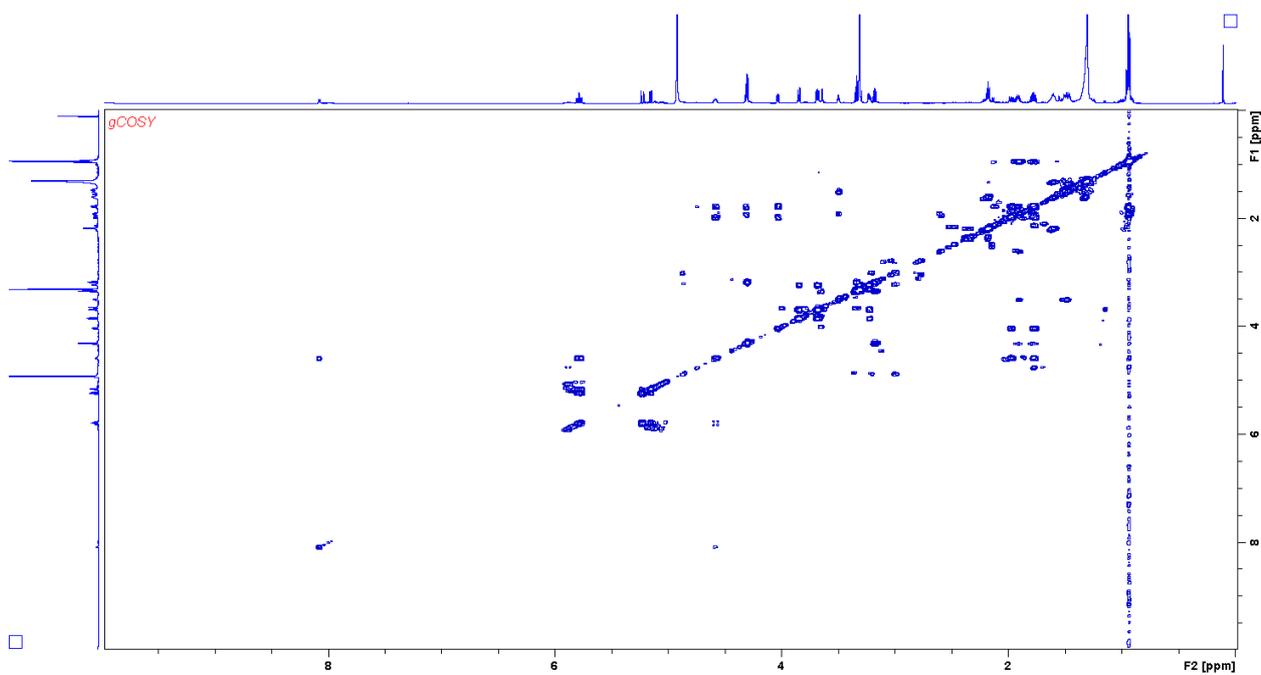


Supplementary figure 13: hrMS<sup>2</sup> spectrum and observed fragments of Myxoglucamide 647. MS<sup>2</sup> spectrum is extracted from LC-hrMS measurement of an extract of *MCy 9003* mutant *tn5 NRPS-t1pks18*.

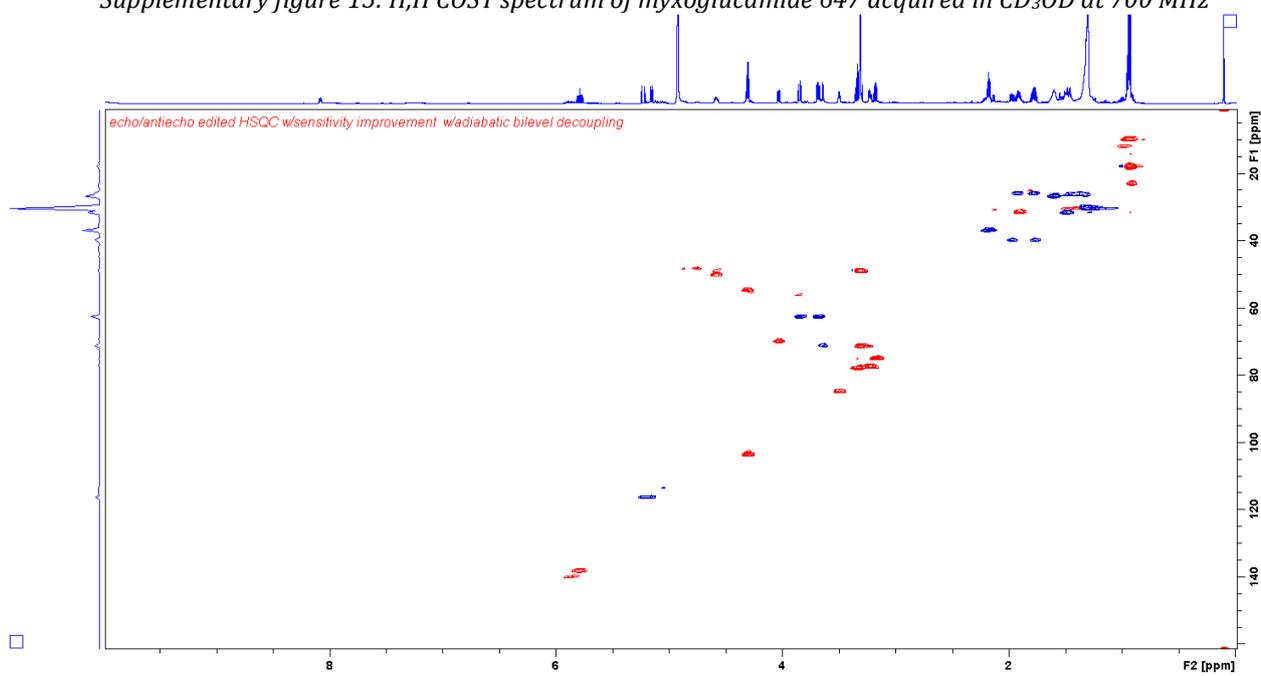
NMR spectra in CD<sub>3</sub>OD were acquired with a Bruker Ascend 700 NMR spectrometer equipped with a 5mm TCI cryoprobe.



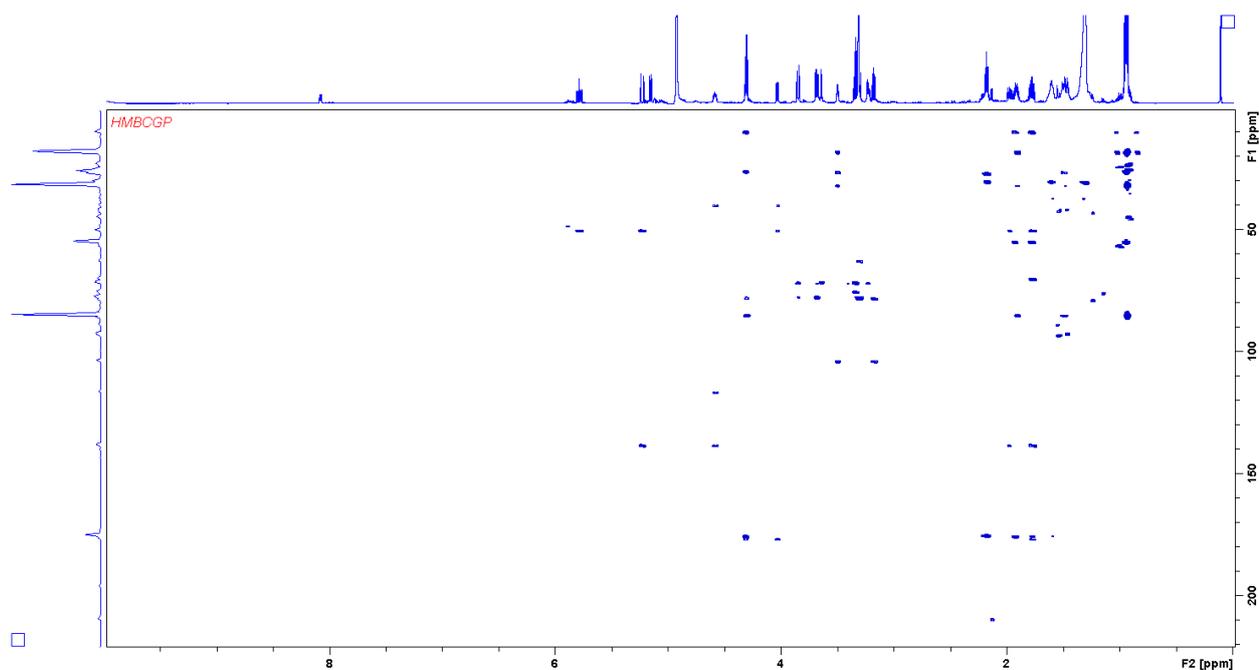
Supplementary figure 14: <sup>1</sup>H spectrum of myxoglucamide 647 acquired in CD<sub>3</sub>OD at 700 MHz



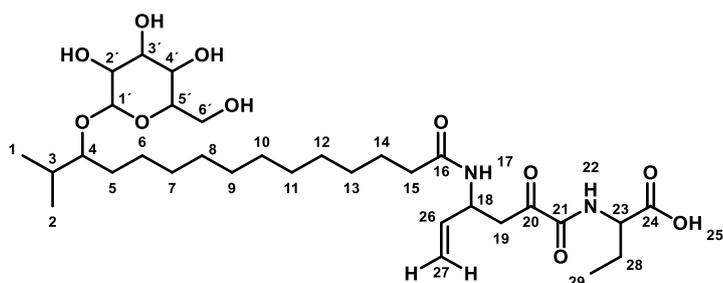
Supplementary figure 15: *H,H* COSY spectrum of myxoglucamide 647 acquired in  $CD_3OD$  at 700 MHz



Supplementary figure 16: HSQC spectrum of myxoglucamide 647 acquired in  $CD_3OD$  at 700/175 MHz



Supplementary figure 17: HMBC spectrum of myxoglucamide 647 acquired in CD<sub>3</sub>OD at 700/175 MHz



Supplementary table 3: Spectroscopic data of myxoglucamide 647 in CD<sub>3</sub>OD.

position	$\delta_c^a$	$\delta_H^b$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
1	18.4	0.93 d 7.0	3	2,3
2	18.1	0.94 d 6.6	3	1,3
3	31.6	1.90 m	1,2,4	
4	85.2	3.49 m	3,5	1',1,2,3,4,5
5	31.8	1.49 m	6	
6	26.6	1.36 m	7-13	
7-13	30.3-31.0	1.27-1.32		
14	27.1	1.58 m	7-13,15	
15	37.1	2.16 t	14	14,16
16	175.7			
17	NH			
18	48.6	4.87 m	26	16,19,20,26
19a	42.7	2.99 d,d 16.8, 8.4	19b	18,20,21,26
19b		3.21 d,d 16.8, 5.3	19a	18,20,26
20	196.3			
21	162.0			
22	NH			
23	55.5	4.29 m	28	21,24,28,29
24	175.2			
25	OH			
26	138.6	5.86 d,d,d 5.5, 10.4, 5.5	18,27	18,19
27a	115.5	5.19 dt,1.4,17.2	27b,26	18,26

27b	115.5	5.11 dt 1.4, 10.4	27a, 26	18,26
28a	25.9	1.80 m		23,24,29
28b		1.96 m		23,24,29
29	10.5	0.95 t 7.4	23,28	
1'	103.4	4.30 d 7.8	2'	2',4',5'
2'	75.5	3.17 d,d	1',3'	1',4'
3'	78.3	3.32 m		
4'	71.7	3.31 <sup>e</sup>		
5'	77.7	3.22 m	6'a/b, b 5'	4'
6'a	63.0	3.85 d,d 11.6, 2.4	6'b, 5'	4',5'
6'b	63.0	3.68 d,d 11.5, 5.5	6'a,5'	4',5'

<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

<sup>b</sup> acquired at 700 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

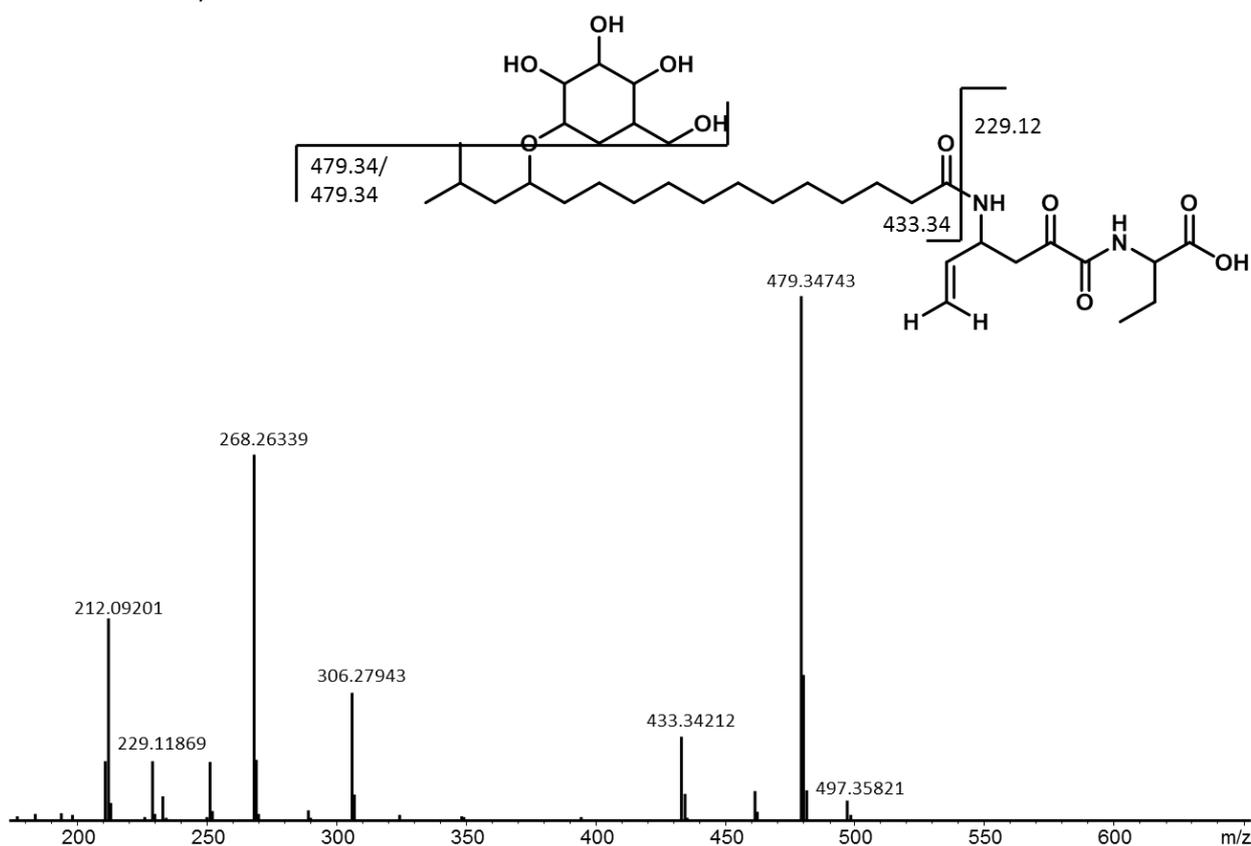
<sup>c</sup> proton showing COSY correlation to indicated proton.

<sup>d</sup> proton showing HMBC correlations to indicated carbons.

<sup>e</sup> overlap with solvent signal.

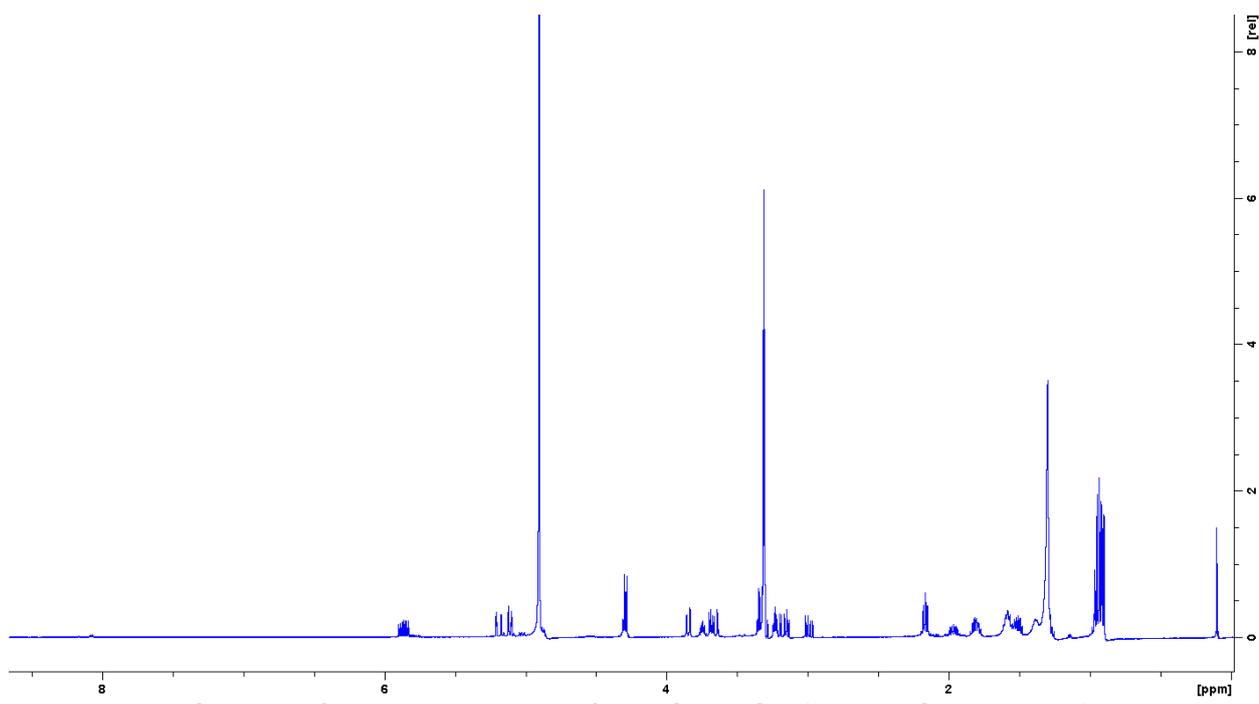
### Analytical data myxoglucamide 659

Myxoglucamide 659 was isolated as colorless amorphous solid. *hr*Mass: 659.41134 [M+H]<sup>+</sup>; specific optical rotation  $[\alpha]_D^{20}$  -3.5 °g<sup>-1</sup>ml dm<sup>-1</sup>; UV max: 193, 241 nm (UV data extracted from LC-MS-UV measurement using a Dionex ultimate 3000 HPLC system equipped with a DAD UV-VIS detector running with H<sub>2</sub>O +0.1 % FA and ACN + 0.1 % FA )

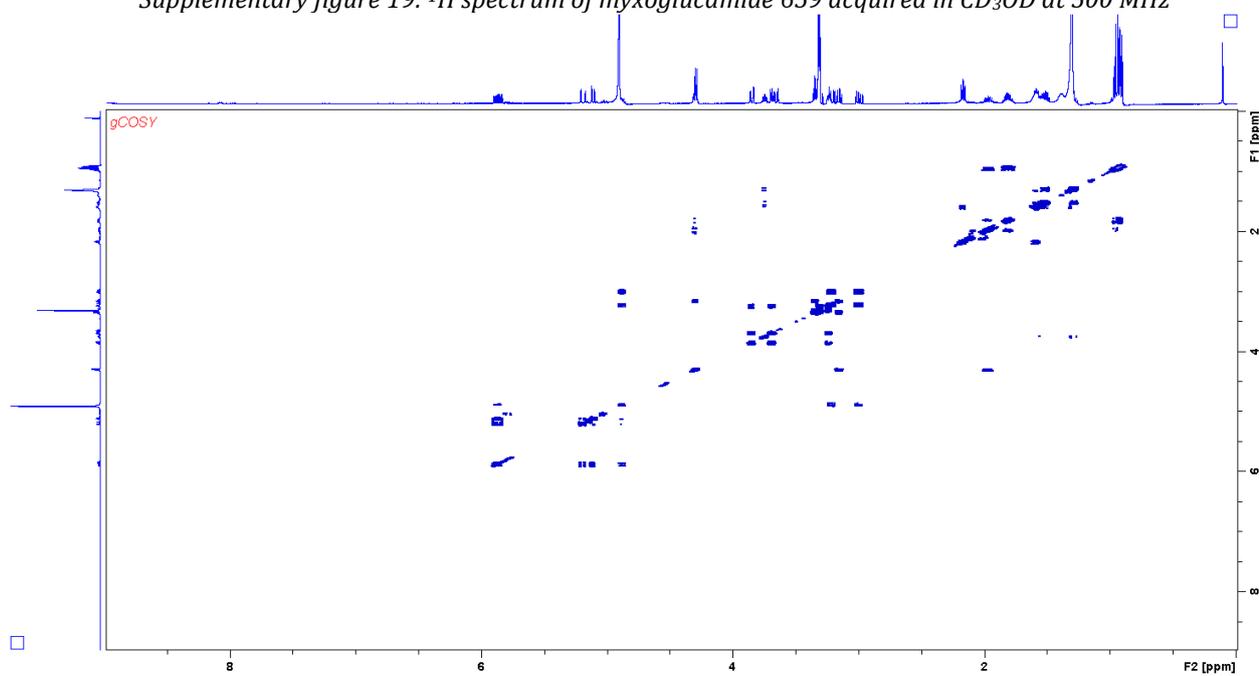


Supplementary figure 18: *hr*MS<sup>2</sup> spectrum and observed fragments of myxoglucamide 659. MS<sup>2</sup> spectrum is extracted from LC-*hr*MS measurement of an extract of *MCy 9003* mutant *tn5 NRPS-t1pks18*.

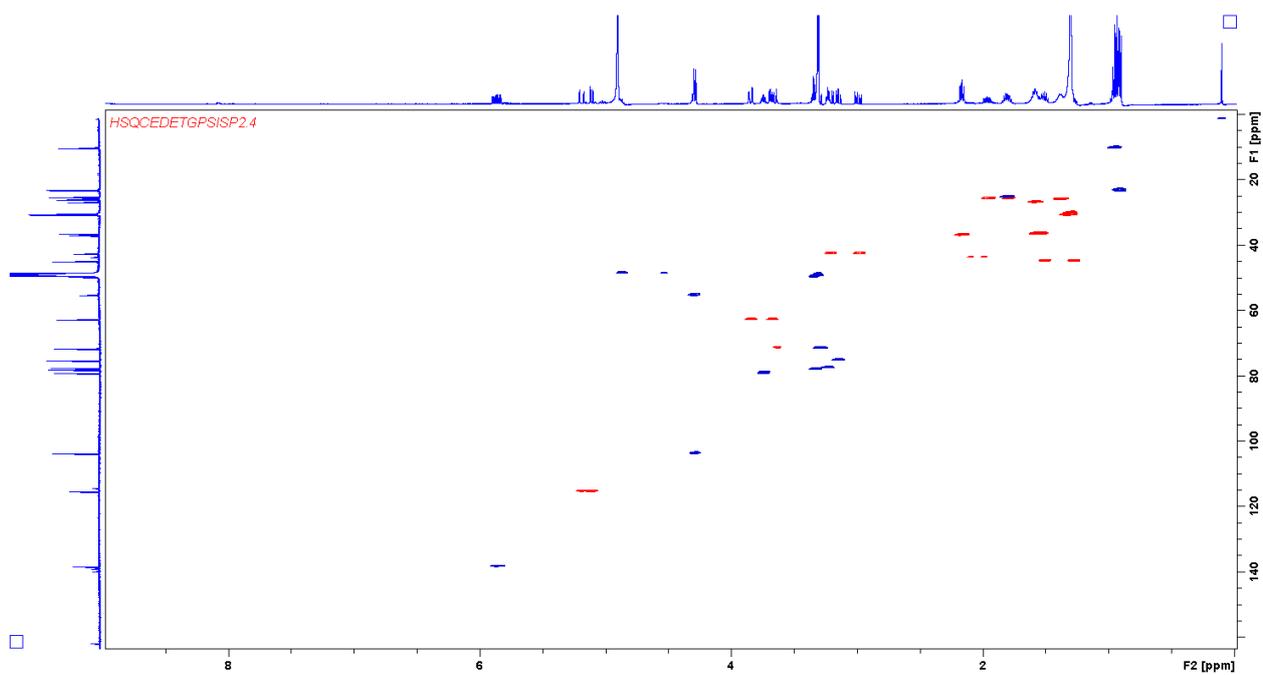
NMR spectra in CD<sub>3</sub>OD were acquired with a UltraShield 500 MHz NMR spectrometer equipped with a 5mm TCI cryoprobe.



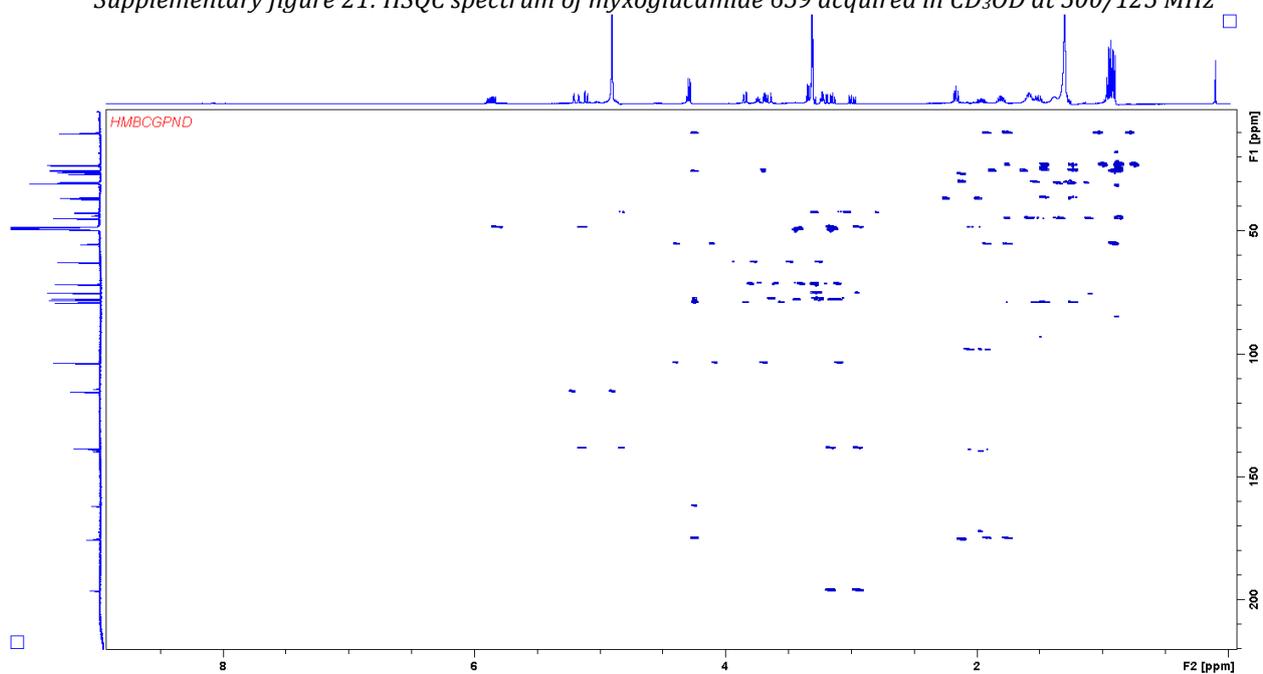
Supplementary figure 19:  $^1\text{H}$  spectrum of myxoglucamide 659 acquired in  $\text{CD}_3\text{OD}$  at 500 MHz



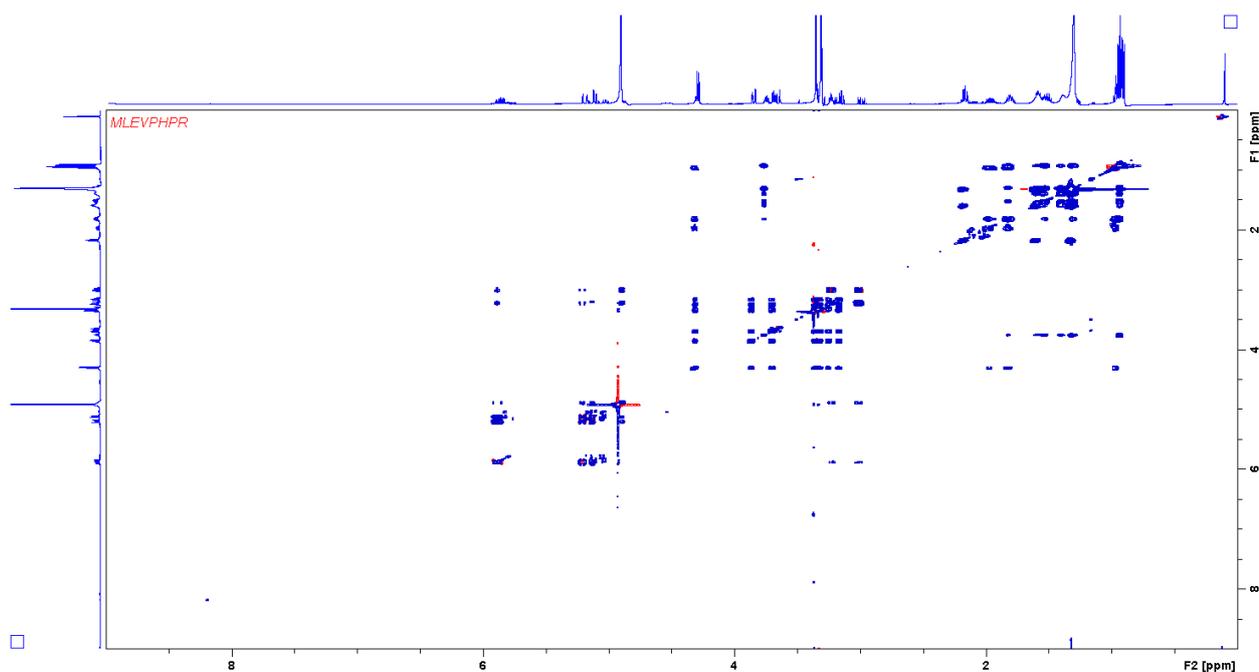
Supplementary figure 20:  $\text{H,H}$  COSY spectrum of myxoglucamide 659 acquired in  $\text{CD}_3\text{OD}$  at 500 MHz



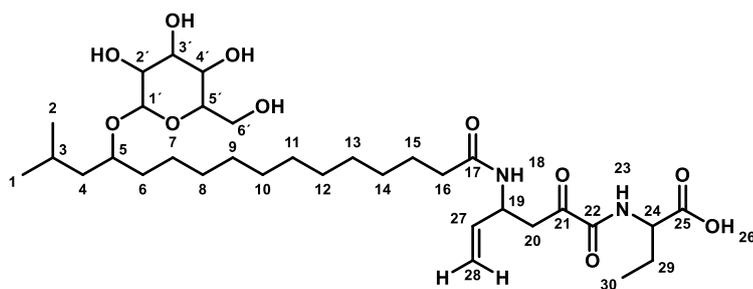
Supplementary figure 21: HSQC spectrum of myxoglucamide 659 acquired in  $CD_3OD$  at 500/125 MHz



Supplementary figure 22: HMBC spectrum of myxoglucamide 659 acquired in  $CD_3OD$  at 500/125 MHz



Supplementary figure 23: TOCSY spectrum of myxoglucamide 659 acquired in CD<sub>3</sub>OD at 500 MHz



Supplementary table 4: Spectroscopic data of Myxoglucamide 659 in CD<sub>3</sub>OD.

position	$\delta_C^a$	$\delta_H^b$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
1	23.5	0.91 d 6.6	3	3,4
2	23.3	0.93 d 6.6	3	3,4
3	25.5	1.81 m	1,2	
4a	45.1	1.28 m	3,4b,5	5,6
4b	45.1	1.52 m	3,4a,5	5,6
5	79.3	3.75 d,d,d 6.1,6.1,6.0,6.0	4,6	1',3,4,7
6a	36.7	1.51 m	5,6b,7	5
6b	36.7	1.60 m	5,6a,7	5
7	26.2	1.38 m	7	
8-14	30.4-31.2	1.28-1.32		
15	27.1	1.58 m		
16	37.1	2.16 t 7.4	15	15, 17
17	175.7			
18	NH			
19	48.7	4.87 c	20, 27	17,20,21,27
20a	42.7	2.99 d,d 16.8,8.5	19,20b	19,21,22
20b	42.7	3.21 d,d 5.4,16.8	19,20a	19,21,22
21	196.3			
22	162.0			
23	NH			

24	55.5	4.29 m	29	22,25,29,30
25	175.3			
26	OH			
27	138.6	5.87 d,d,d 17.1, 10.4, 5.5	19,28	19,20
28a	115.5	5.11 d,t 10.4,1.4	27,28b	19,27
28b	115.5	5.18 d,t 17.0,1.4	27,28a	19,27
29a	25.9	1.80 m	24,30	25
29b	25.9	1.96 m	24,30	25
30	10.5	0.95 t 7.7	29	29,24
1'	103.9	4.29 d 7.8	2'	5,5'
2'	75.4	3.15d,d 8.7,7.8	1',3'	1',3',4'
3'	78.3	3.33 m	2',3'	2',4'
4'	71.8	3.31 <sup>e</sup>		
5'	77.8	3.23 m	4',6'a,6'b	
6'a	62.9	3.84 d,d 2.4,11.9	6'a,5'	4',5'
6'b	62.9	3.84 d,d 5.5,11.6	6'b,5'	4',5'

<sup>a</sup> acquired at 125 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

<sup>b</sup> acquired at 500 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

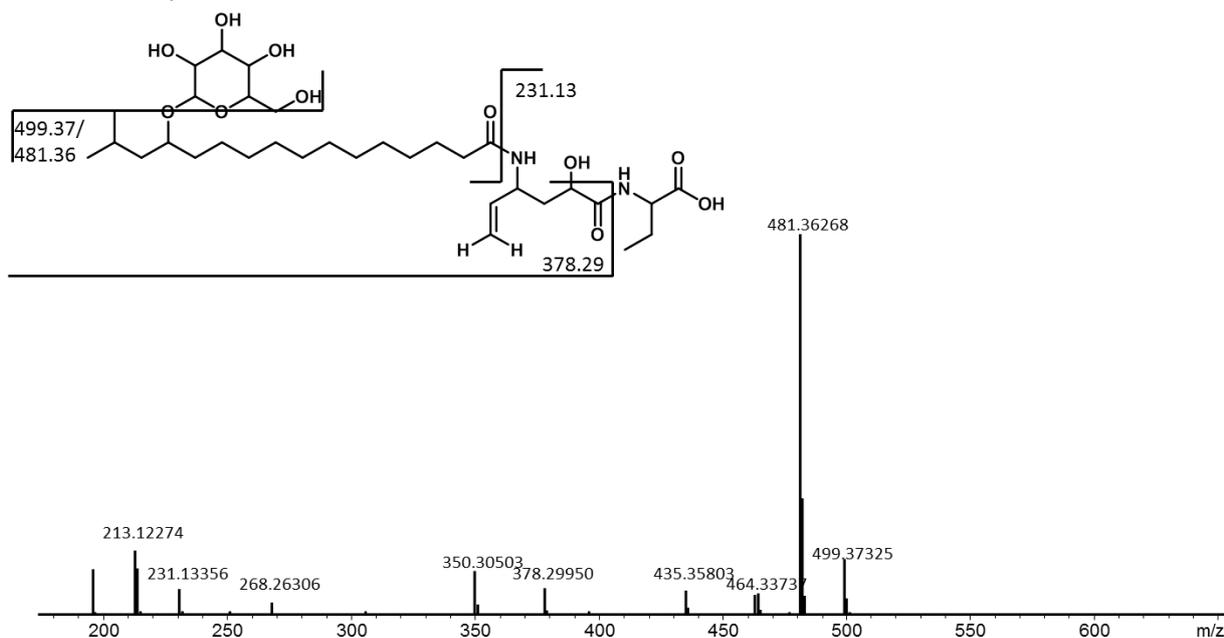
<sup>c</sup> proton showing COSY correlation to indicated proton.

<sup>d</sup> proton showing HMBC correlations to indicated carbons.

<sup>e</sup> overlap with solvent signal.

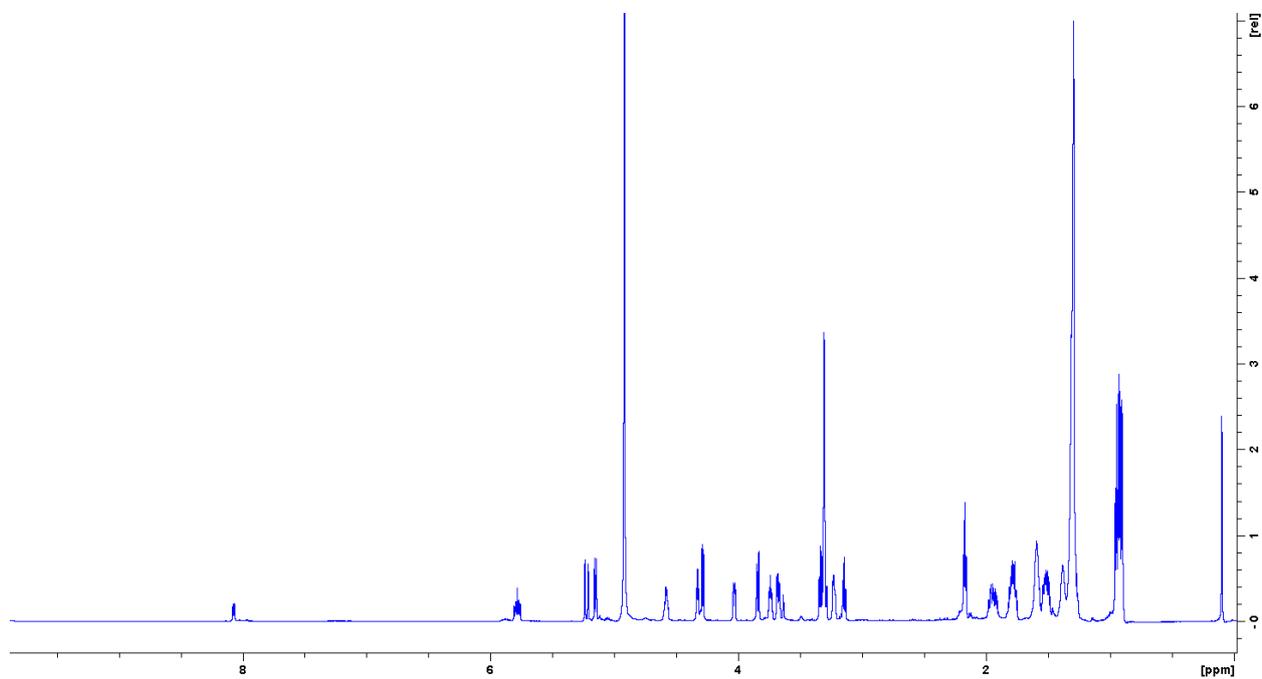
### Analytical data myxoglucamide 661

Myxoglucamide 661 was isolated as colorless amorphous solid. *hr*Mass: 661.42636 [M+H]<sup>+</sup>; specific optical rotation  $[\alpha]_D^{20}$  6.8 °g<sup>-1</sup>ml dm<sup>-1</sup>; UV max: 199 nm (UV data extracted from LC-MS-UV measurement using a Dionex ultimate 3000 HPLC system equipped with a DAD UV-VIS detector running with H<sub>2</sub>O +0.1 % FA and ACN + 0.1 % FA )

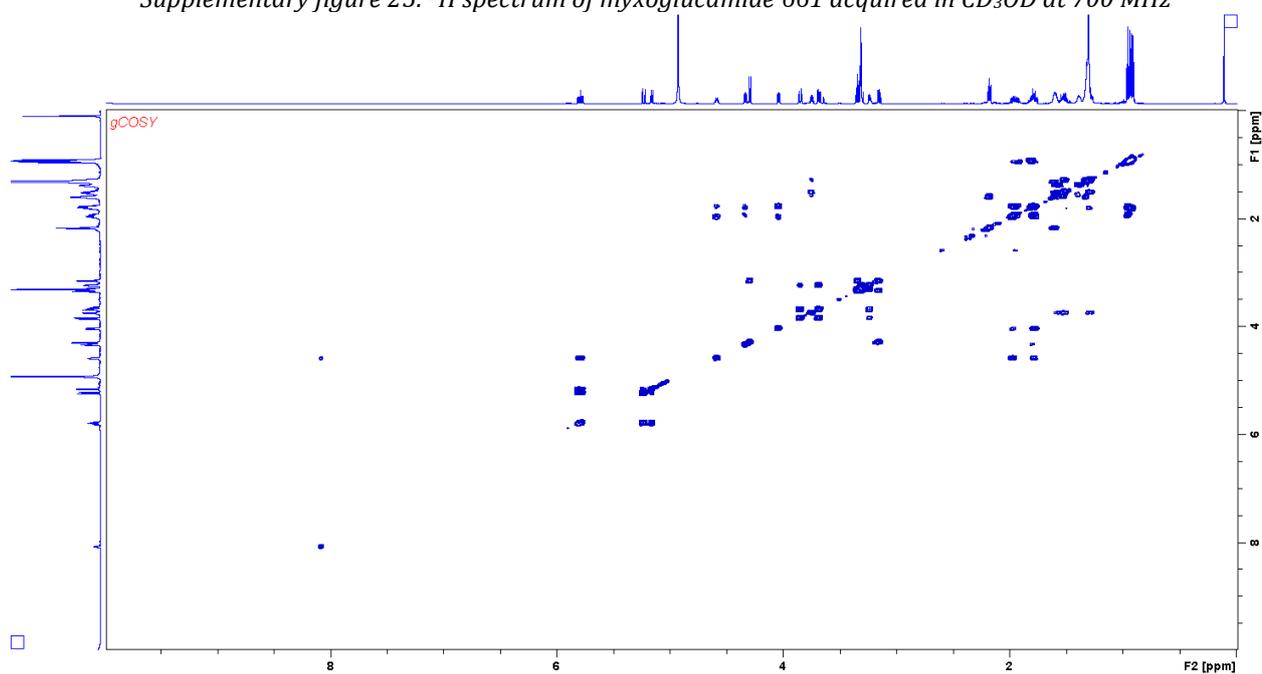


Supplementary figure 24: *hr*MS<sup>2</sup> spectrum and observed fragments of Myxoglucamide 661. MS<sup>2</sup> spectrum is extracted from LC-*hr*MS measurement of an extract of *Mcy* 9003 mutant *tn5 NRPS-t1pks18*.

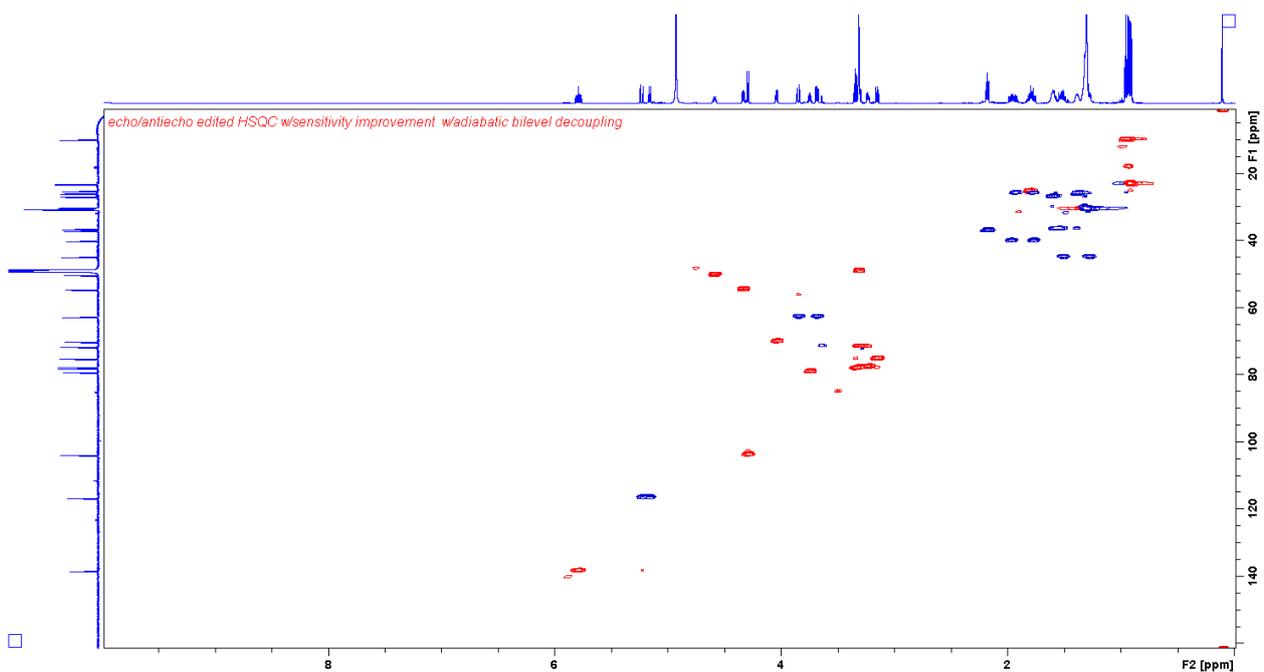
NMR spectra in CD<sub>3</sub>OD were acquired with a Bruker Ascend 700 NMR spectrometer equipped with a 5mm TCI cryoprobe.



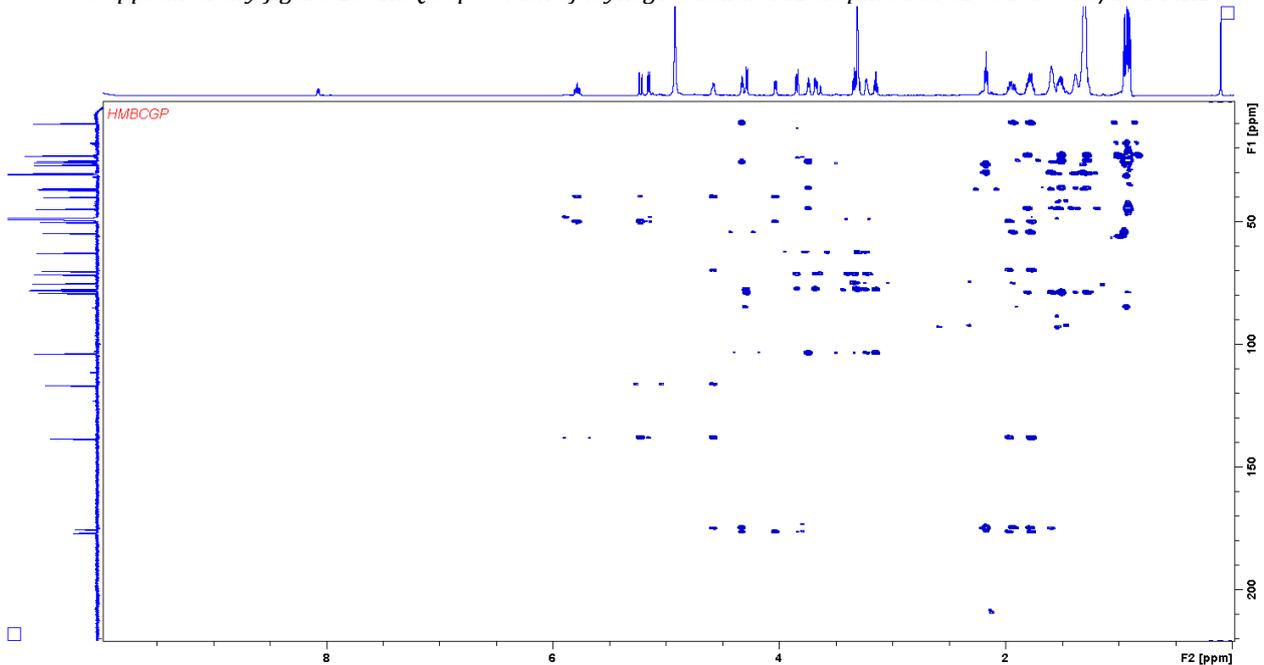
Supplementary figure 25: <sup>1</sup>H spectrum of myxoglucamide 661 acquired in CD<sub>3</sub>OD at 700 MHz



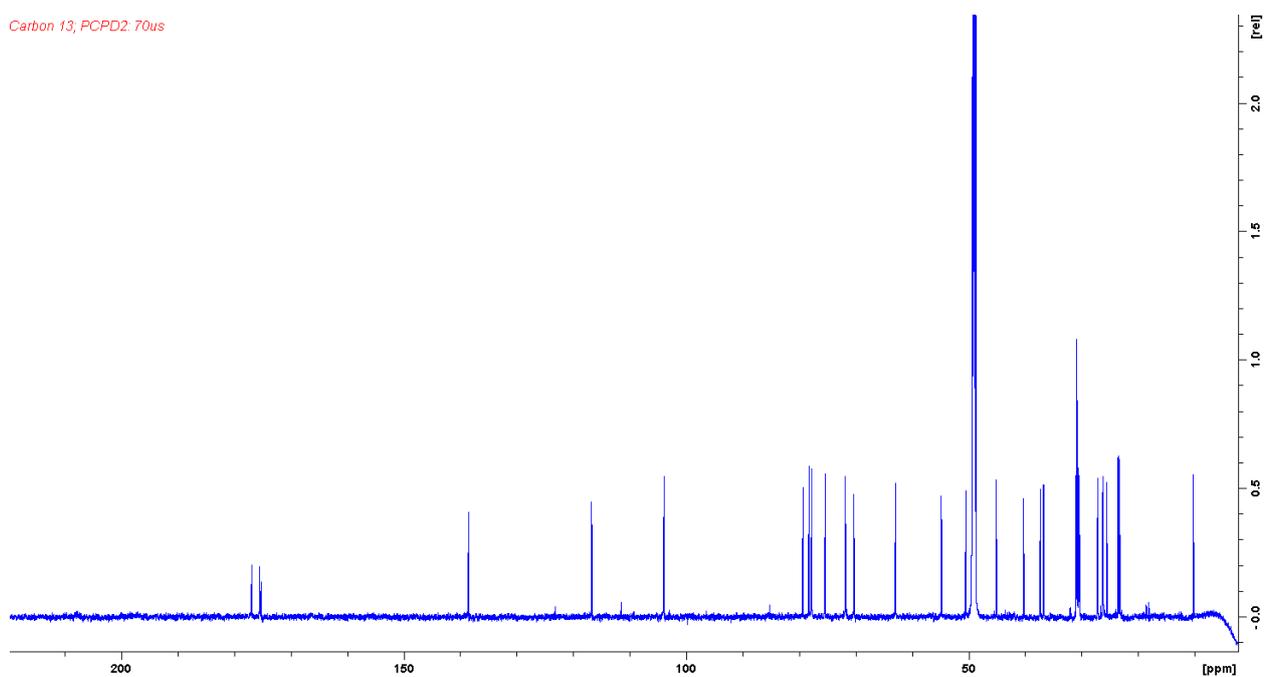
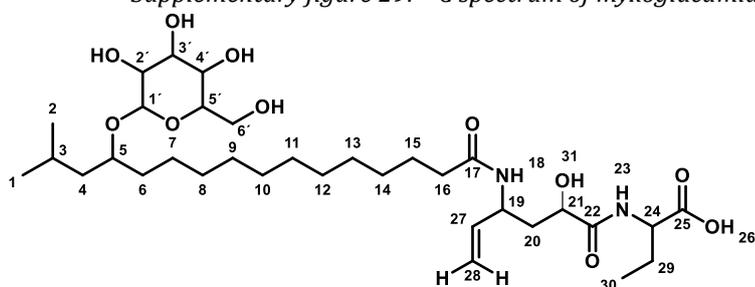
Supplementary figure 26: H,H COSY spectrum of myxoglucamide 661 acquired in CD<sub>3</sub>OD at 700 MHz



Supplementary figure 27: HSQC spectrum of myxoglucamide 661 acquired in CD<sub>3</sub>OD at 700/175 MHz



Supplementary figure 28: HMBC spectrum of myxoglucamide 661 acquired in CD<sub>3</sub>OD at 700/175 MHz

Supplementary figure 29:  $^{13}\text{C}$  spectrum of myxoglucamide 661 acquired in  $\text{CD}_3\text{OD}$  at 175 MHzSupplementary table 5: Spectroscopic data of myxoglucamide 661 in  $\text{CD}_3\text{OD}$ .

position	$\delta_{\text{C}}^{\text{a}}$	$\delta_{\text{H}}^{\text{b}}$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
1	23.3	0.92 d (6.3)	3	2,3
2	23.5	0.91 d (6.6)	3	1,3
3	25.5	1.80 m	1,2 4a/b	1,2,4,5
4a	45.1	1.52 m	3,4b,5	1,2,3,5
4b		1.28 m	3,4a,5	1,2,3,5
5	79.3	3.74 m	4a/b, 6a/b	1',3,4,6,7
6a	36.7	1.59 m	5,6b,7	5
6b		1.53 m	5,6a,7	5
7	26.2	1.38 m		
8-14	30.3-31.0	1.33-1.29 m		
15	27.1	1.59 m		
16	37.2	2.17 t (7.6)	15	15,17
17	175.4			
18	NH			
19	50.4	4.58 m	18,20a/b,27	18,20,27,28
20a	40.2	1.97 m	19,20b,21	19,21,27
20b		1.78 m	19,20a,21	19,21,27
21	70.3	4.04 d,d (9.4,3.6)	20a/b	19,20,22
22	176.9			
23	NH			
24	54.7	4.33 d,d (7.6,5.0)	29 a/b	22,25,29,30
25	175.2			
26	OH			

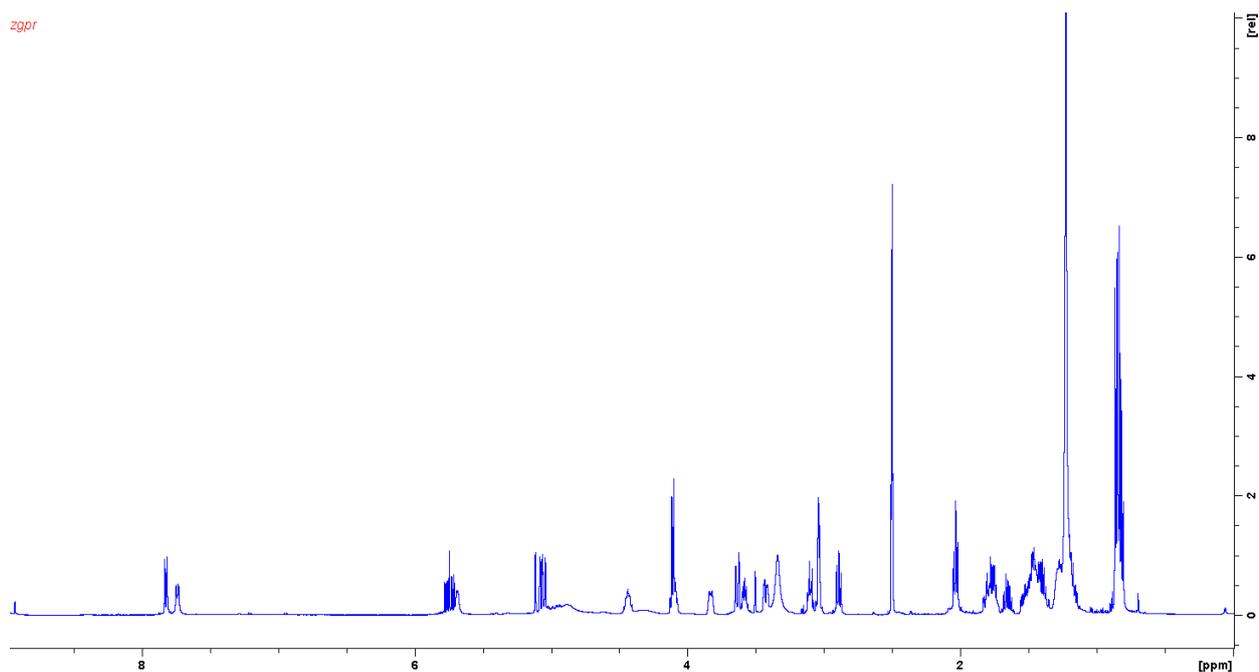
27	138.5	5.79 d,d,d (17.1, 10.3,7.0)	28a/b	19,20,28
28a	116.7	5.23 d,t (1.3,17.1)	27,28b	19,27
28b		5.16 d,t (1.3,10.4)	27,28a	
29a	26.1	1.93 m	24,29b,30	24,30
29b		1.79 m	24,29a,30	24,30
30	10.2	0.95 t (7.6)	29a/b	24,29
1'	103.9	4.29 d (7.9)	2	'5,5'
2'	75.3	3.15 d,d (7.9,8.9)	1',3'	1',3'
3'	78.3	3.33 m	2',4'	1',2',4'
4'	71.8	3.29 m	5'	5,6'
5'	77.8	3.23 d,d,d (9.3,5.4,2.3)	4',6'a/b	1',3',4',6'
6'a	62.9	3.85 d,d (2.6,11.7)	6'b,5'	4',5'
6'b		3.68 d,d (11.8,5.5)	6'a,5'	

<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

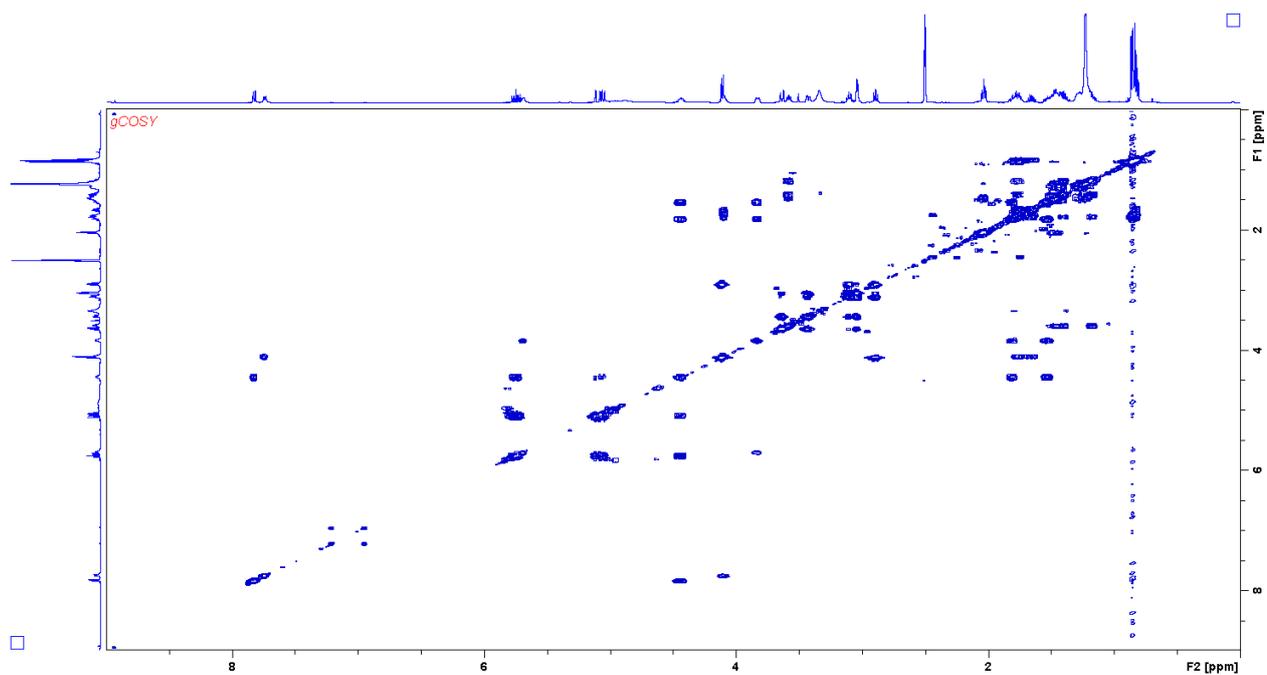
<sup>b</sup> acquired at 125 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

<sup>c</sup> proton showing COSY correlation to indicated proton.

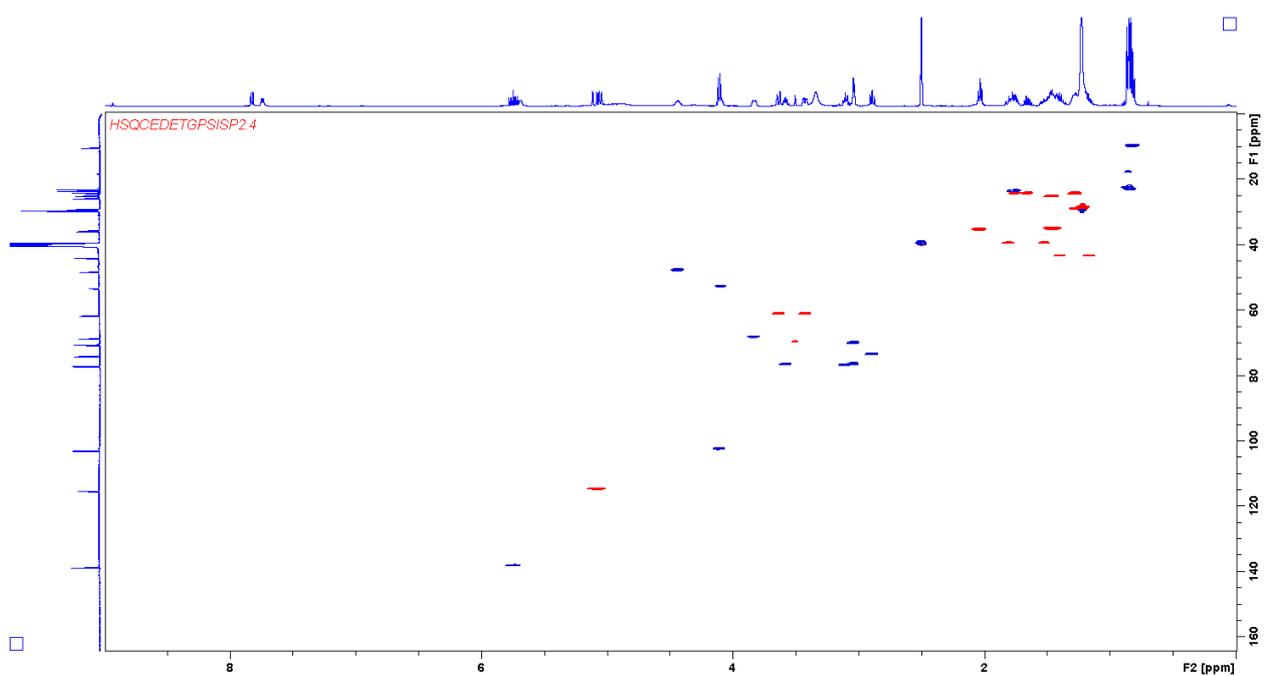
<sup>d</sup> proton showing HMBC correlations to indicated carbons.



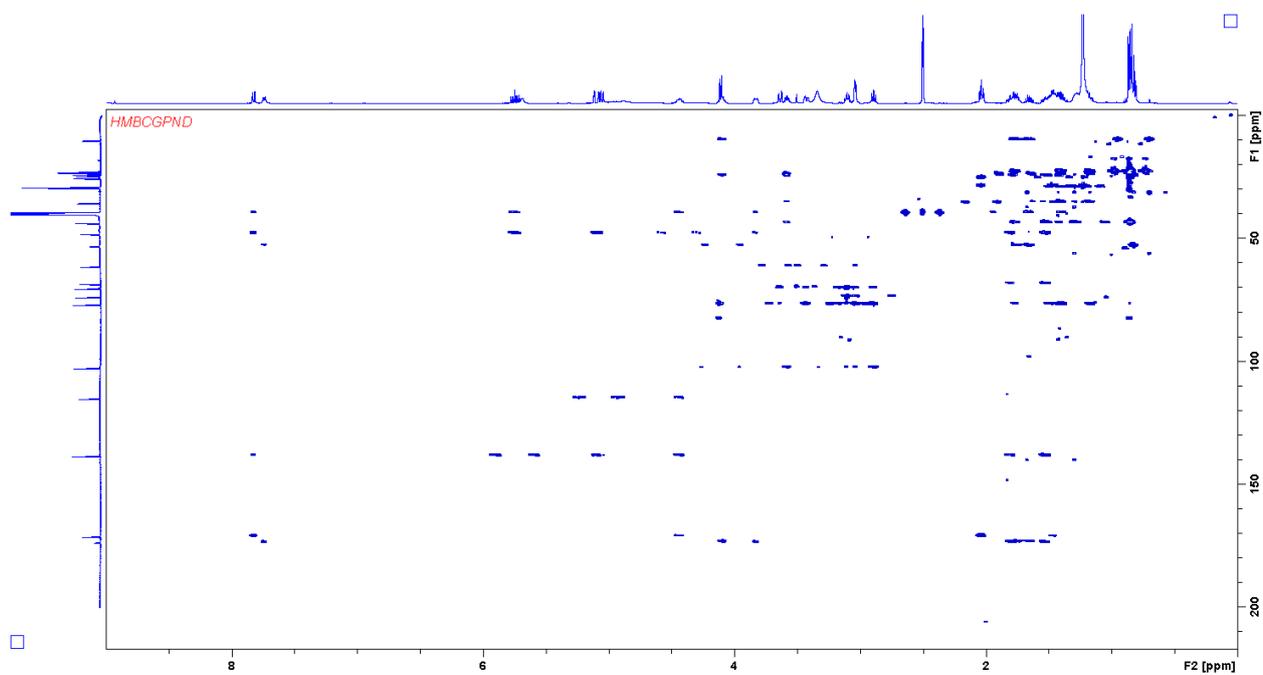
Supplementary figure 30: <sup>1</sup>H spectrum of myxoglucamide 661 acquired in DMSO d<sub>6</sub> at 500 MHz



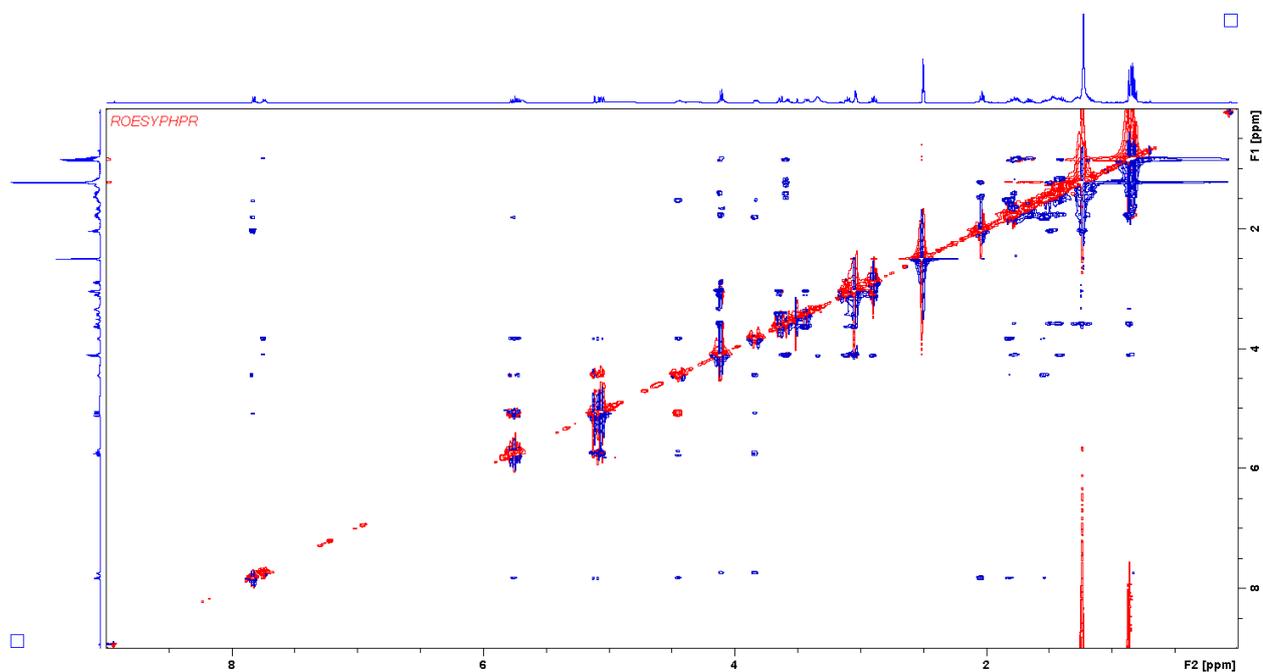
Supplementary figure 31:  $^1\text{H}$ , $^1\text{H}$  COSY spectrum of myxoglucamide 661 acquired in  $\text{DMSO } d_6$  at 500 MHz



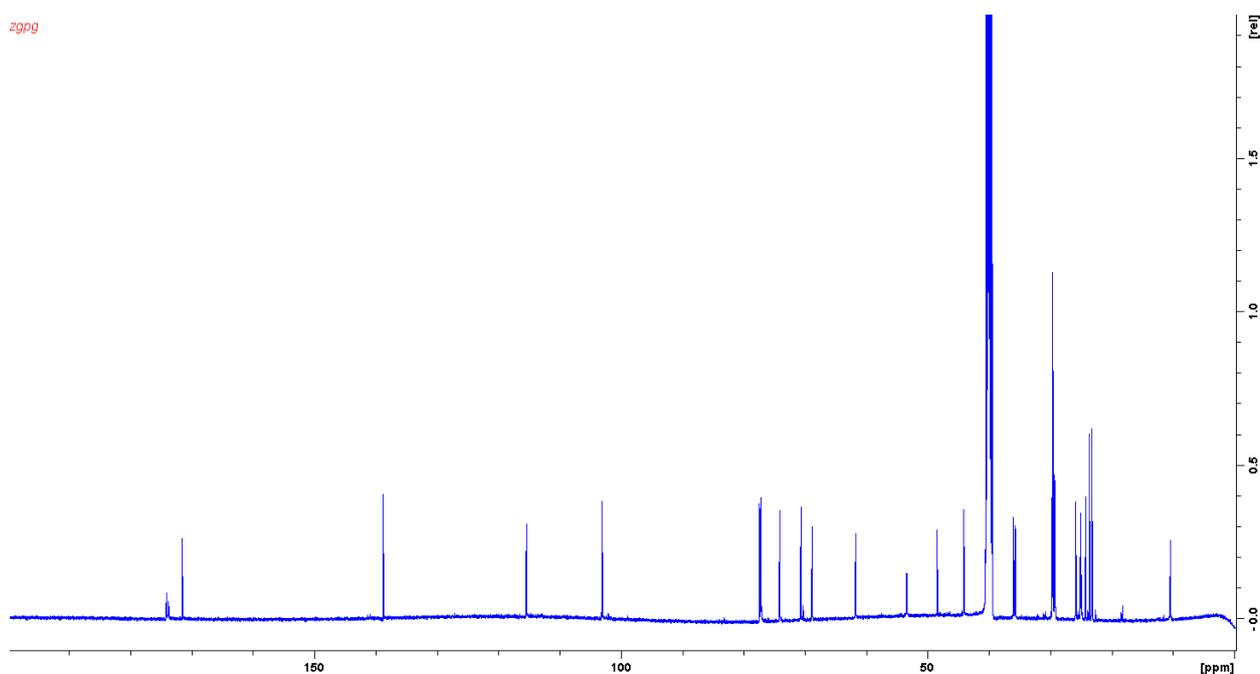
Supplementary figure 32: HSQC spectrum of myxoglucamide 661 acquired in  $\text{DMSO } d_6$  at 500/125 MHz



Supplementary figure 33: HMBC spectrum of myxoglucamide 661 acquired in DMSO  $d_6$  at 500/125 MHz



Supplementary figure 34: ROESY spectrum of myxoglucamide 661 acquired in DMSO  $d_6$  at 500/125 MHz



Supplementary figure 35:  $^{13}\text{C}$  spectrum of myxoglucamide 661 acquired in  $\text{DMSO } d_6$  at 125 MHz

Supplementary table 6: Spectroscopic data of Myxoglucamid 661 in  $\text{DMSO } d_6$ .

position	$\delta_{\text{C}}^{\text{a}}$	$\delta_{\text{H}}^{\text{b}}$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
1	22.4	0.86 d 6.5	3	2,3,4
2	22.9	0.84 d 6.6	3	1,3,4
3	23.4	1.77 m	1, 2, 4a/b	1,2,5
4a	43.4	1.39 m	3,4b,5	1,2,3,5
4b	43.4	1.17 m	4a,5,3	1,2,3,5
5	76.6	3.58 m	4a/b, 6a/b	1',4,6
6a	35.0	1.49 m		
6b	35.0	1.42 m		
7	24.3	1.28 m		
8-14	28.1-29.8	1.17-1.24		
15	25.1	1.46 m	16	17
16	35.3	2.03 t (7.3)	15	15,17
17	170.8			
18	NH	7.82 d (8.4)	19	16,17,19,27
19	47.7	4.44 m	18,20a/b,27	16,17,21,27,28
20a	39.3	1.81 m	19,20b,21	19,21,22,27
20b	39.3	1.52 m	19,20a,21	19,21,22,27
21	68.8	3.83 m	20a/b	19,20,22
22	173.4			
23	NH	7.74 d (7.7)	24	22,23,24,29
24	52.7	4.11 m	23,29a/b	25,29,30
25	173.0			
26	OH			
27	138.1	5.74 d,d,d (17.1,10.4,6.6)	19,28a/b	19,20
28a	114.7	5.09 d,d,d (17.0,1.7,1.7)	28b,27	19,20,27
28b	114.7	5.05 d,d,d (10.4,1.8,1.2)	28a,27	19,20
29a	24.1	1.75 m	24,29a,30	24
29b	24.1	1.65 m	24,29b,30	24
30	9.7	0.82 t (7.1)	29a/b	29

1'	102.4	4.11 d (7.7)	2'	5,5'
2'	73.4	2.89 d,d (8.6,8.0)	1',3'	1',3',4'
3'	76.7	3.10 m	2',4'	1',2',4'
4'	69.9	3.04 m		
5'	76.4	3.04 m		
6'a	61.0	3.63 m		
6'b	61.0	3.43 m	6'a,5'	4'

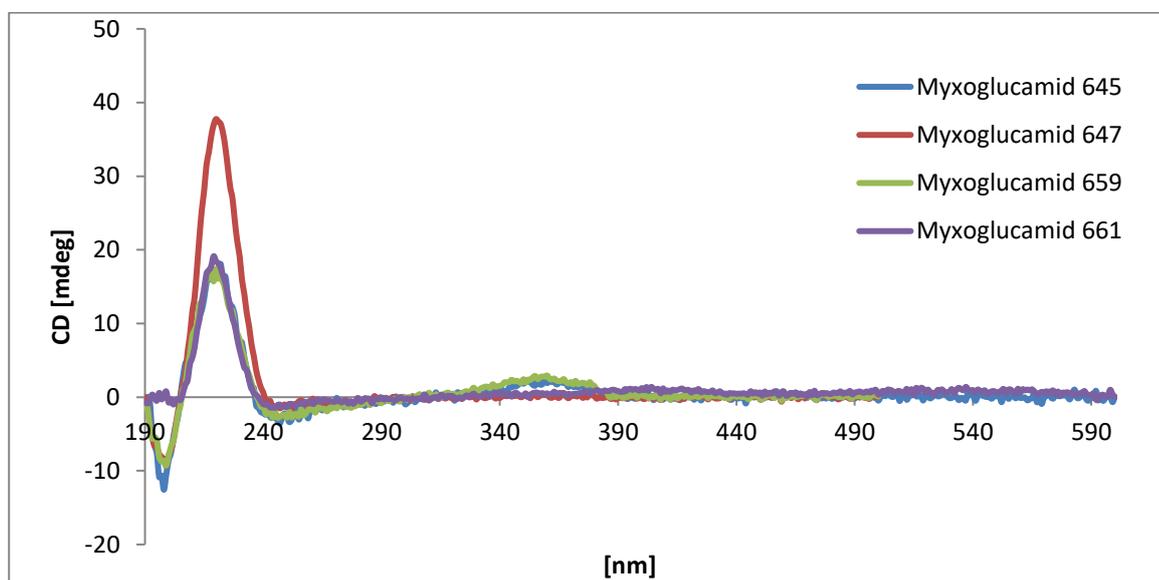
<sup>a</sup> acquired at 125 MHz and assigned from 2D NMR spectra, referenced to solvent signal DMSO d<sub>6</sub> at  $\delta$  39.52 ppm.

<sup>b</sup> acquired at 500 MHz, referenced to solvent signal DMSO d<sub>6</sub> at  $\delta$  2.50 ppm.

<sup>c</sup> proton showing COSY correlation to indicated proton.

<sup>d</sup> proton showing HMBC correlations to indicated carbons.

### Additional spectroscopic data:

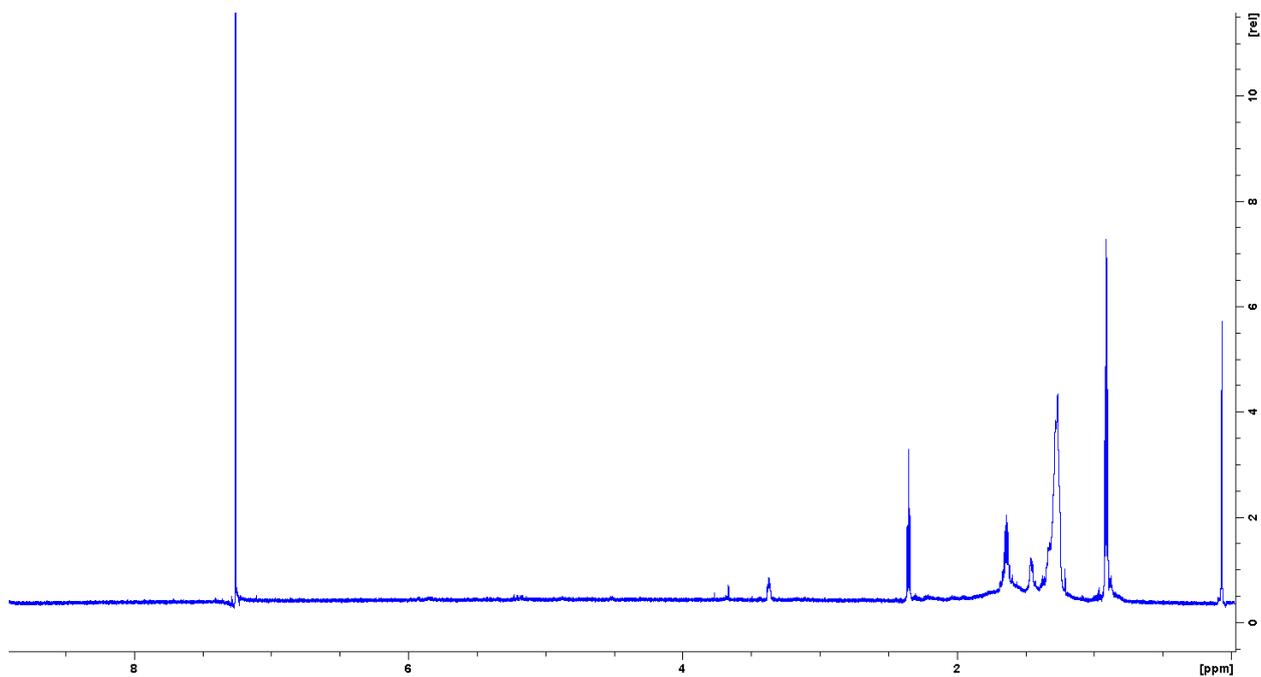


Supplementary figure 36: CD-spectra of myxoglucamide congeners acquired with a Jasco J-1500 CD-spectrometer in MeOH

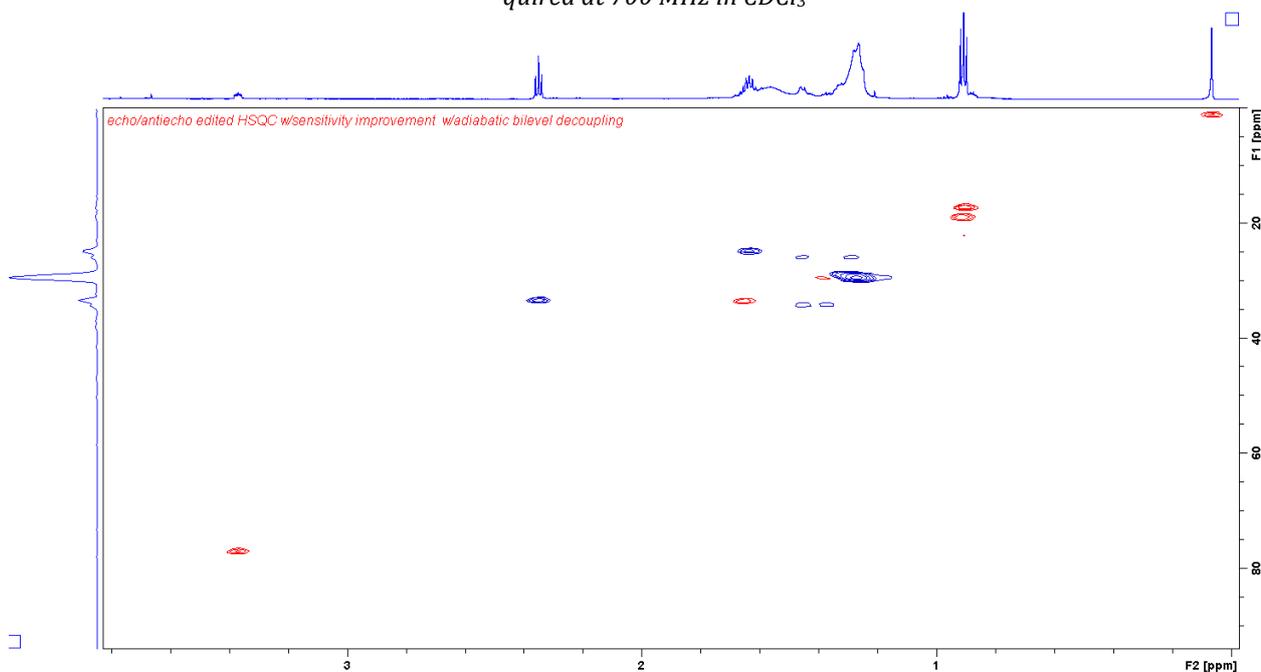
### Stereochemical analysis

#### Acid Hydrolysis of myxoglucamide 645

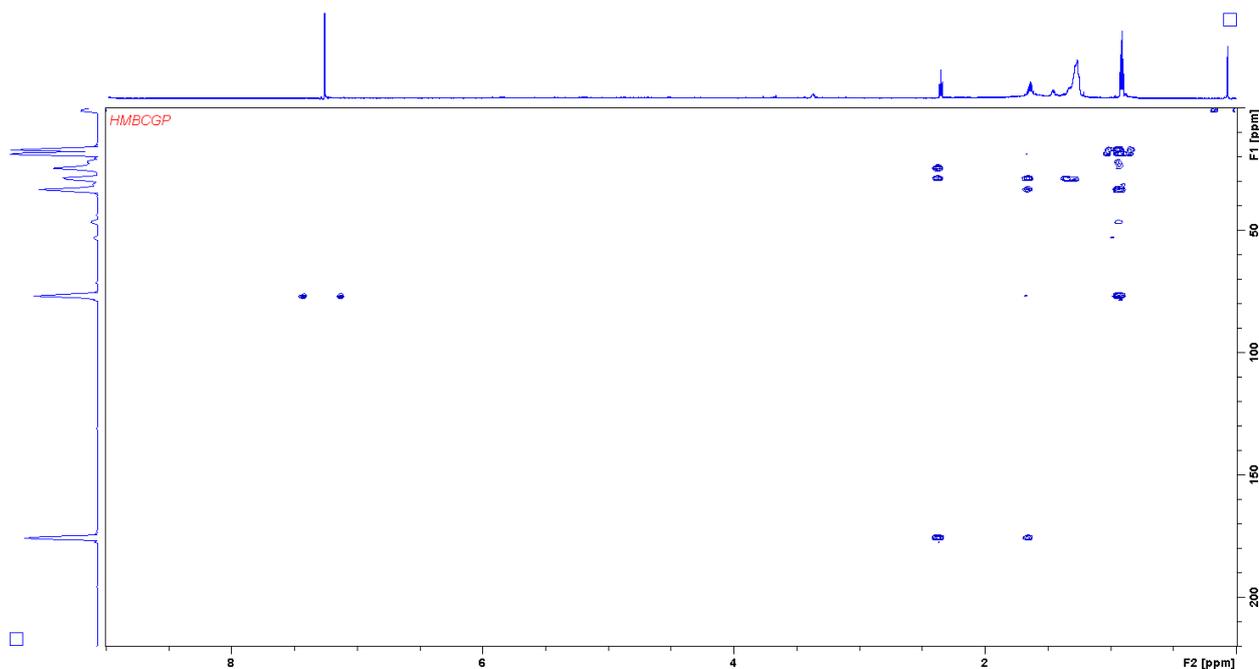
In order to investigate the stereochemistry of the different structural features of the molecule 3.7 mg were hydrolyzed under acid conditions. The hydrolysis mixture was extracted with n-hexane, ethyl-acetate and MeOH in a sequential manner. 0.6 mg of n-hexane extract were obtained. NMR analysis in CDCl<sub>3</sub> showed that the n-Hexane layer contained HMA.



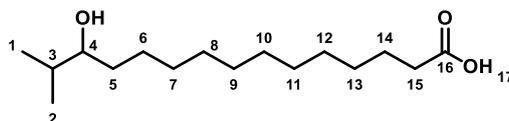
Supplementary figure 37:  $^1\text{H}$  Spectrum of the *n*-hexane extract of the hydrolysis mixture of myxoglucamide 645 acquired at 700 MHz in  $\text{CDCl}_3$



Supplementary figure 38: HSQC spectrum of the *n*-hexane extract of the hydrolysis mixture of myxoglucamide 645 acquired at 700/175 MHz in  $\text{CDCl}_3$



Supplementary figure 39: HMBC spectrum of the n-hexane extract of the hydrolysis mixture of myxoglucamide 645 acquired at 700/175 MHz in CDCl<sub>3</sub>



Supplementary table 7: Spectroscopic data n-hexane extract of the hydrolysis mixture in CDCl<sub>3</sub>

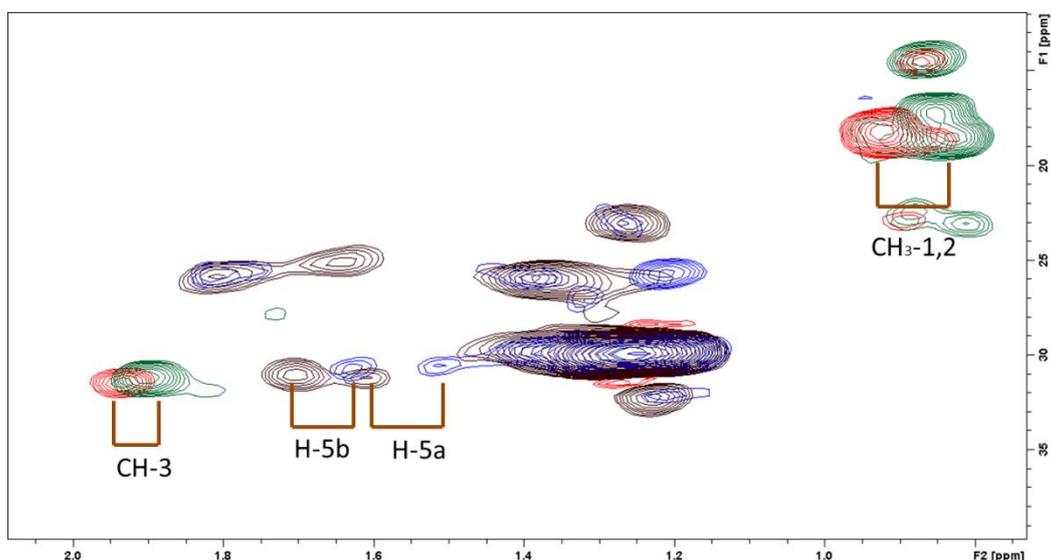
position	$\delta_C^a$	$\delta_H^b$ (J in Hz)
1	17.3	0.92 d
2	19.0	0.91 d
3	33.6	1.64 m
4	77.0	3.37 d,d,d 8.6, 5.3, 3.1
5	34.3	1.45 m
6		1.37 m
7-13	28.7-30.1	1.18-1.34
14	24.8	1.35 m
15	33.6	2.34 t 7.3
16	176.6	

<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CDCl<sub>3</sub> at  $\delta$  77.16 ppm.

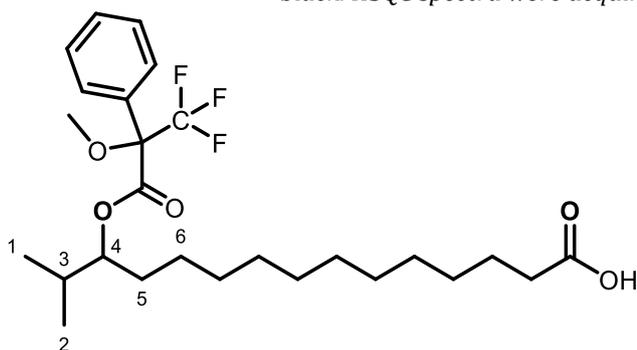
<sup>b</sup> acquired at 700 MHz, referenced to solvent signal CDCl<sub>3</sub> at  $\delta$  7.26 ppm.

### Mosher ester formation and analysis of HMPA

In order to obtain the absolute configuration of the hydroxy group of myxoglucamide 645 a (*S*)-MTPA ester and a (*R*)-MTPA ester of the fatty acid were made. Due to the low amounts, the reaction mixture was directly measured by 2D-NMR.



Supplementary figure 40: HSQC spectra of the (*R*)-Mosher ester of HMPA CH<sub>3</sub> and CH displayed in red, CH<sub>2</sub> displayed in blue overlaid with HSQC spectra of (*S*)-Mosher ester of HMPA CH<sub>3</sub> and CH displayed in green, CH<sub>2</sub> displayed in black. HSQC spectra were acquired at 700/175 MHz in pyridine *d*<sub>5</sub>



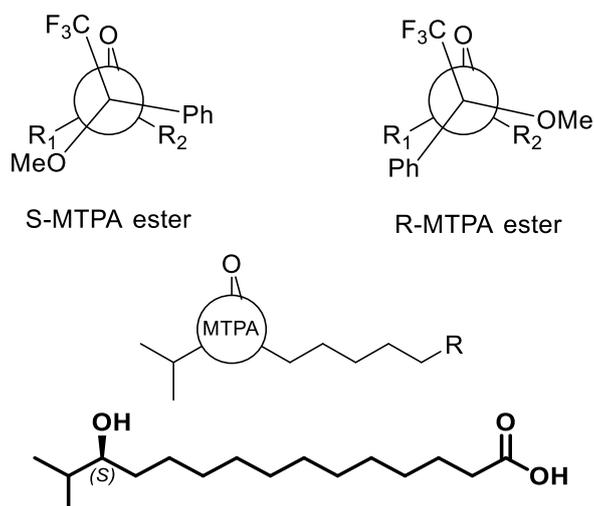
Supplementary table 8: spectroscopic data of both Mosher esters of the HMPA chain in pyridine *d*<sub>5</sub>

Position	(S)-MTPA ester		(R)-MTPA Ester		$\Delta \delta_{(S-R)}$
	$\delta_H^b$	$\delta_C^a$	$\delta_H^b$	$\delta_C^a$	
1	0.926	17.5	0.855	17.2	0.071
2	0.914	18.6	0.829	18.4	0.085
3	1.933	31.3	1.899	31.2	0.034
4	5.163	82.2	5.170	82.1	-0.007
5a	1.628	30.7	1.705	31.2	-0.077
5b	1.506		1.610		-0.104
6	1.397	25.9	1.372	25.9	0.025

<sup>a</sup> acquired at 175 MHz in pyridine *d*<sub>5</sub> and referenced to solvent signal at  $\delta$  123.87

<sup>b</sup> acquired at 700 MHz in pyridine *d*<sub>5</sub> and referenced to solvent signal at  $\delta$  7.22.

Analysis of the Shift difference  $\Delta\delta(S-R)$  yielded a certain difference distribution. An analysis according to Hoyer et al. 2017 [2] suggest an *S*-configuration of hydroxyl group. Positive  $\Delta\delta$  values can be assigned as *R*<sub>1</sub> while negative  $\Delta\delta$  values are assigned as *R*<sub>2</sub>.



Supplementary figure 41: Determination of the absolute configuration of the HMPA chain.

Following this protocol the 13-Hydroxy-14-Methylpentadecenoid acid could be assigned as (*S*)-configured.

### Analysis of the pyranose moiety

The aqueous MeOH extract of 1.2 mg contained the sugar and amino-butyrac acid to a smaller extend as confirmed by NMR analysis in D<sub>2</sub>O. The sugar had identical carbon resonances like a D-glucose standard.

Table 7: Chemical shifts of the β-Glucose in the Methanol extract from the acidic Hydrolysis of AP9003\_645 obtained in D<sub>2</sub>O

position	δ <sub>c</sub> <sup>a</sup>		δ <sub>H</sub> <sup>b</sup> j (Hz)	
	aqueous extract	D -glucose (β-form)	aqueous extract	D -glucose(β-form)
1	98.7	98.7	4.59 d	
2	76.9	76.9	3.19 d,d 9.1,8.0	3.16 8.9, 8.5
3	78.5	78.5	3.42 m	3.40 m
4	72.3	72.3	3.35 m	3.32 m
5	78.7	78.7	3.41 m	3.39 m
6a	63.5	63.5	3.84 d,d 2.3,12.2	3.82 d,d 12.0,2.0
6b			3.66 m	3.65 m

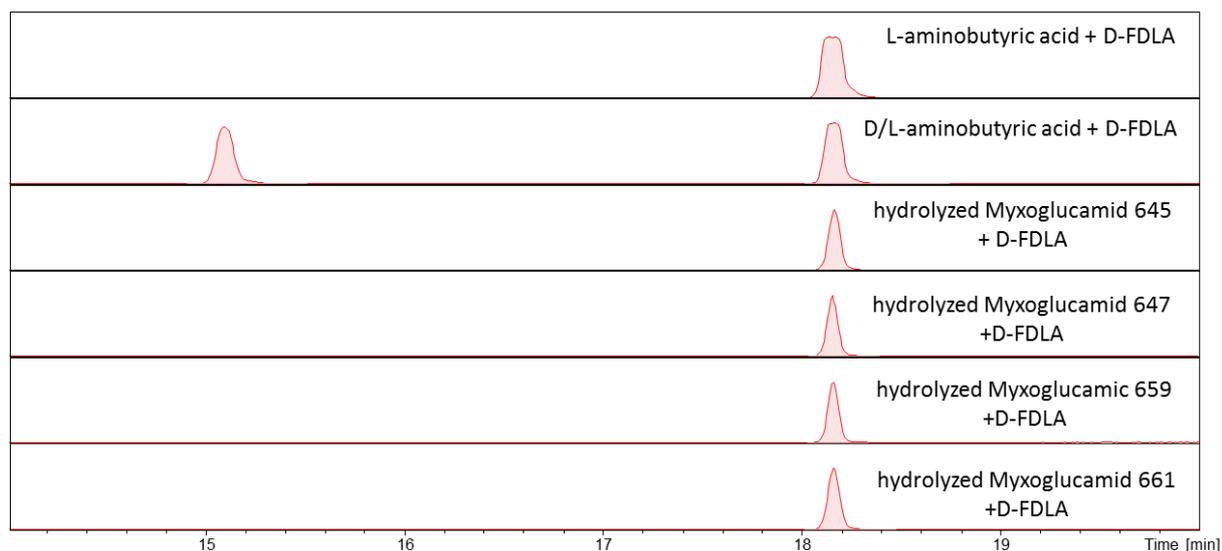
<sup>a</sup> acquired at 175 MHz in D<sub>2</sub>O, referenced to referenced to anomeric proton of β-Glucose at δ 98.66. Reference value obtained Human metabolome

<sup>b</sup> acquired at 700 MHz and referenced to solvent signal at δ 4.79.

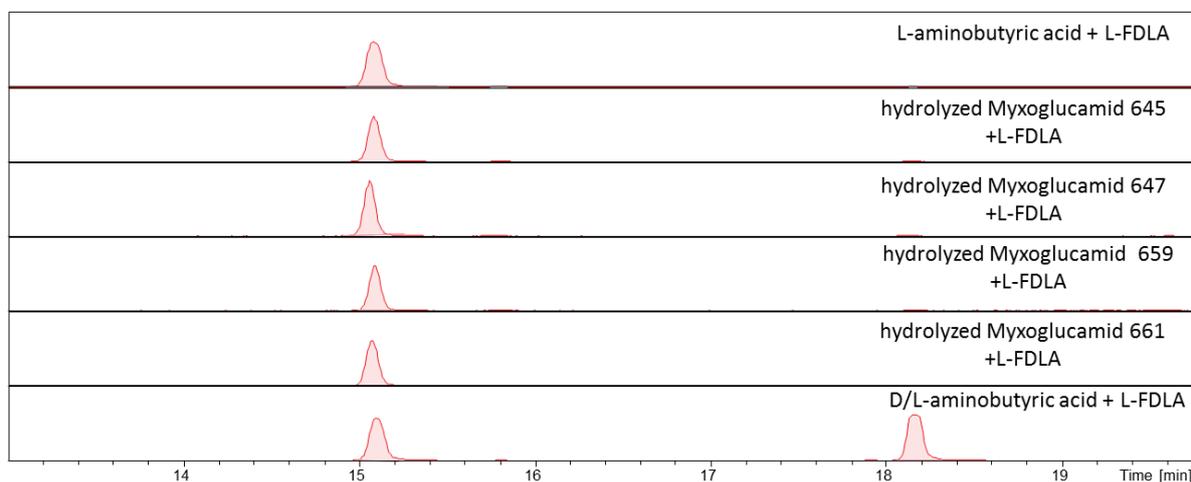
The optical rotation of the MeOH extract was  $[\alpha]_D^{20} 17^\circ \text{ g}^{-1} \text{ ml dm}^{-1}$ . The positive optical rotation indicates that the glucose moiety is D-glucose.

### Determination of C-terminal amino-butyrac acid

The configuration of the C-terminal Amino-butyrac acid was determined for all isolated derivatives by derivatization with D-FDLA and L-FDLA followed by LC-MS analysis.



Supplementary figure 42: Results of Marfey's analysis for amino-butyric acid standards and four isolated myxoglucamid variants after acidic hydrolysis showing the Extracted Ion Chromatogram (EIC)  $398.17 \pm 0.1$  m/z for D-1-Fluoro-2,4-dinitrophenyl-5-Leucine amide (D-FDLA) derivatized amino-butyric.



Supplementary figure 43: Results of Marfey's analysis for amino-butyric acid standards and four isolated myxoglucamid variants after acidic hydrolysis showing the Extracted Ion Chromatogram (EIC)  $398.17 \pm 0.1$  m/z for L-1-Fluoro-2,4-dinitrophenyl-5-Leucine amide (L-FDLA) derivatized amino-butyric.

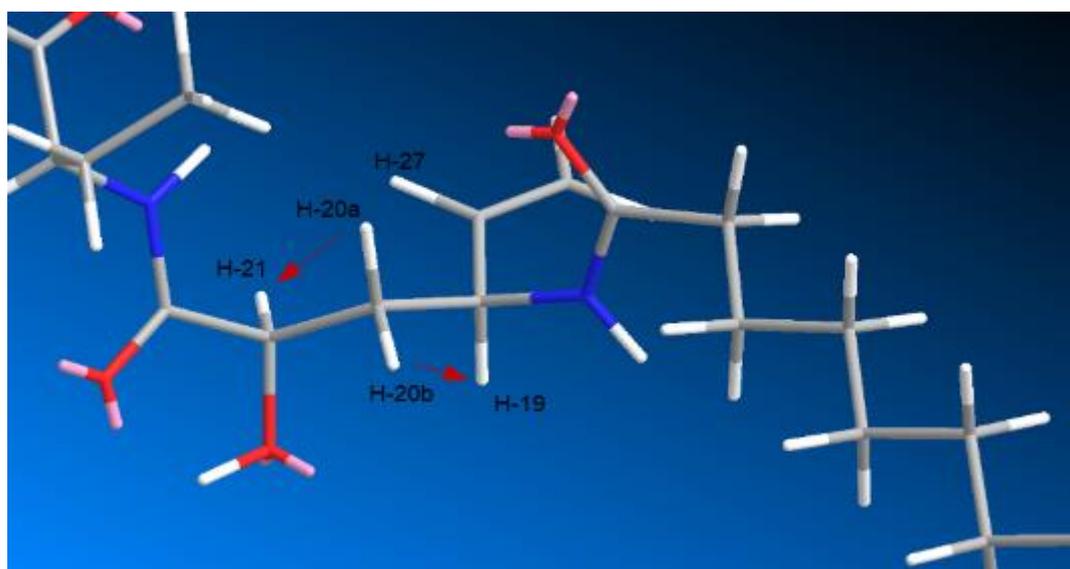
#### Analysis of relative configuration of the 4-amino-2-oxohex-5-enoic acid

In myxoglucamide 661 and 647 the  $\alpha$ -ketone is reduced into an alcohol adding an additional stereo center to the molecule. By a detailed analysis of ROESY correlations the relative configuration of both stereo centers was determined. ROESY crosspeaks were integrated using Bruker Topspin 3.6 software. Using the assumption distance between diastereotopic protons H-20a and H-20b is  $1.76 \text{ \AA}$  [3].

Supplementary table 9: List of integrated crosspeaks for determination of distances in the Hydroxy-Methylpentadecanoic acidin myxoglucamid 661

Integral [rel]	(F2) [ppm]	(F1) [ppm]	Calculated distance [Å]	measured distance [Å]
0.1136	5,7524 H-27	4,4473 H-19	2.529004561	2.79
0.0446	1,8151 H-20a	4,4473 H-19	2.955449376	3.07
0.1063	3,8442 H-21	4,4429 H-19	2.557155415	2.84
0.1617	1,5547 H-20b	4,4429 H-19	2.384483553	2.52
0.0347	1,5536 H-20b	3,8409 H-21	3.081705182	3.09
0.1311	4,4539 H-19	3,8321 H-21	2.469328597	2.84
0.1733	1,8195 H-20a	3,8277 H-21	2.3571085	2.69
0.1718	5,7546 H-27	3,8189 H-21	2.360526106	2.28
0.1041	5,7623 H-27	1,8151 H-20a	2.566084064	2.83
1.0000	1,5438 H-20b	1,8063 H-20a	1.76	1.77
0.0442	4,4528 H-19	1,8019 H-20a	2.959890351	3.07
0.1814	3,8431 H-21	1,8019 H-20a	2.339231054	2.69
0.1741	4,4517 H-19	1,5339 H-20b	2.355299858	2.52
0.9372	1,8107 H-20a	1,5295 H-20b	1.779128382	1.77
0.0235	3,8343 H-21	1,5163 H-20b	3.288526587	3.09

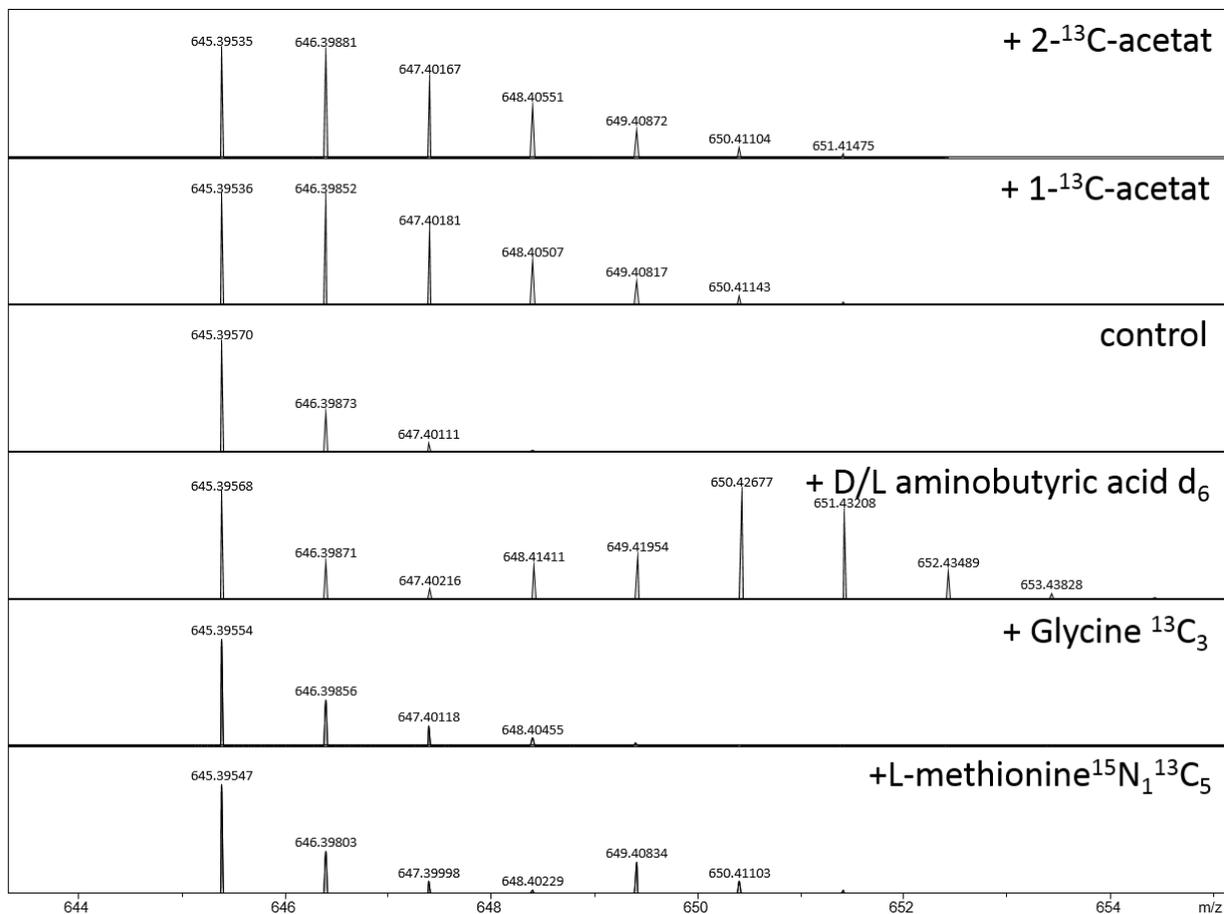
The calculations indicate that with distance around 2.5 Å H-21 and H20a as well as H-20b and H-19 are located in close proximity suggesting H-21 and H-19 to be oriented in opposite directions. A three dimensional model of myxoglucamid 661 was constructed in chemdraw and chem3D software and optimized by molecular dynamics MM2 calculations. The angle of the vinyl group (C-19 and C-27) was manually optimized by using the chem3D dihedral driver to fit the coupling constant of 6.6 Hz. Key distances were compared with the ROESY obtained distances supporting a relative configuration of 19(*S*),21(*R*).



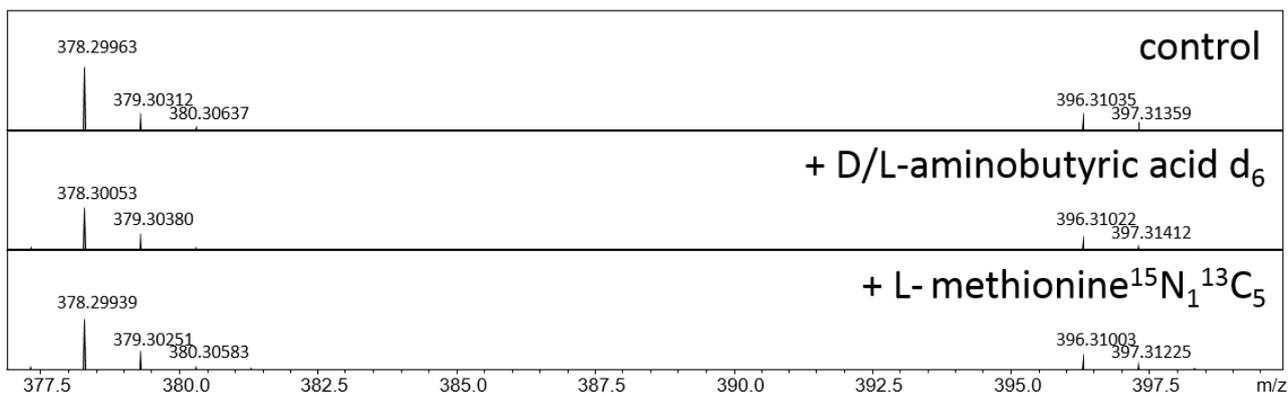
Supplementary figure 45: Showing corresponding part of Myxoglucamide 661 in the stick model displaying key ROESY correlations as red arrows

### Results feeding experiments

In order to create a model for the biosynthesis isotope labelled amino acids and carbon sources were supplemented to cultures of MCy 9003 mutant tn5 NRPS-t1Pk8 18.



Supplementary figure 46: Isotope pattern of myxoglucamide 645 showing single charged mass  $[M+H]^+$  extracted from LC-hrMS from the control culture and cultures supplemented with 2-<sup>13</sup>C-acetate, 1-<sup>13</sup>C-acetate, aminobutyric acid d<sub>6</sub>, <sup>13</sup>C<sub>3</sub>-Glycine and <sup>15</sup>N<sub>1</sub><sup>13</sup>C<sub>5</sub> methionine.



Supplementary figure 47: Isotope pattern of Fragment 378 of myxoglucamide 661 after MS<sup>2</sup> fragmentation of control culture and cultures supplemented with D/L aminobutyric acid d<sub>6</sub> and L-methionine <sup>15</sup>N<sub>1</sub><sup>13</sup>C<sub>5</sub>. MS<sup>2</sup> spectra are extracted from LC-hrMS measurements in autoMS<sup>2</sup> mode.

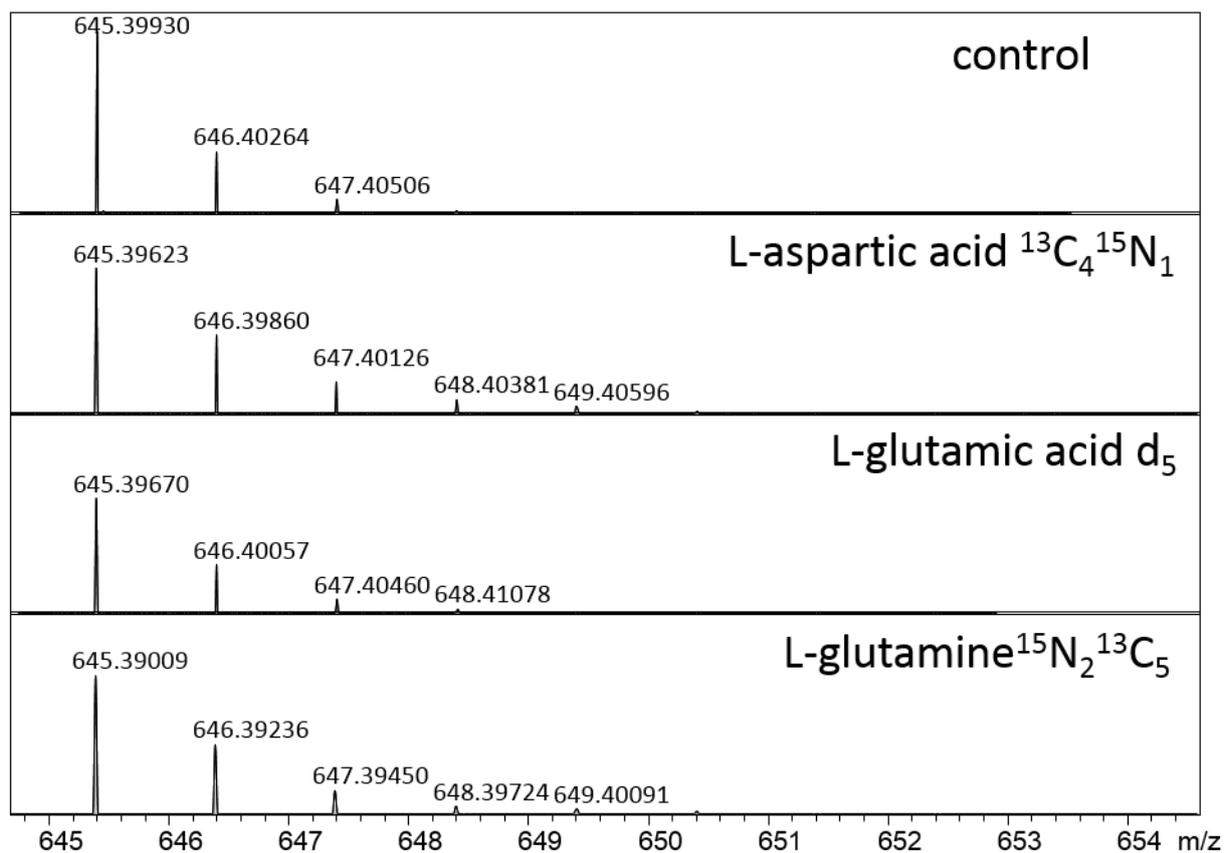
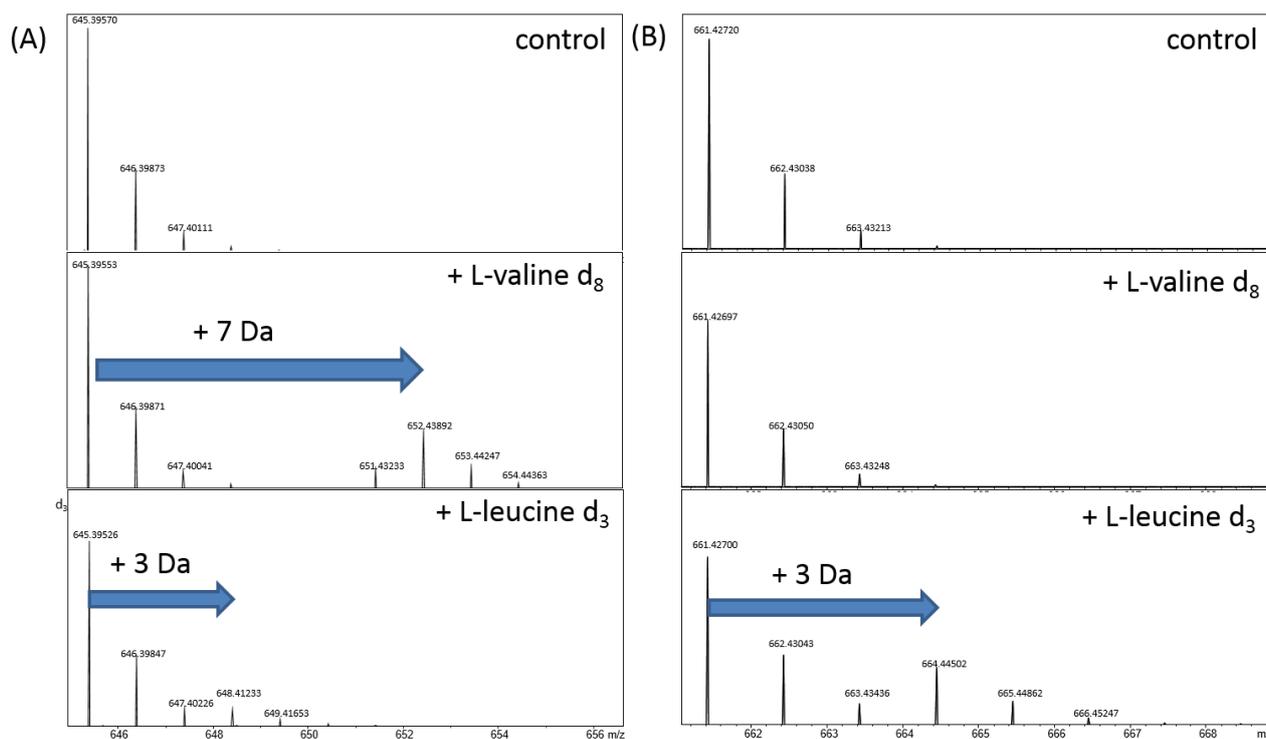
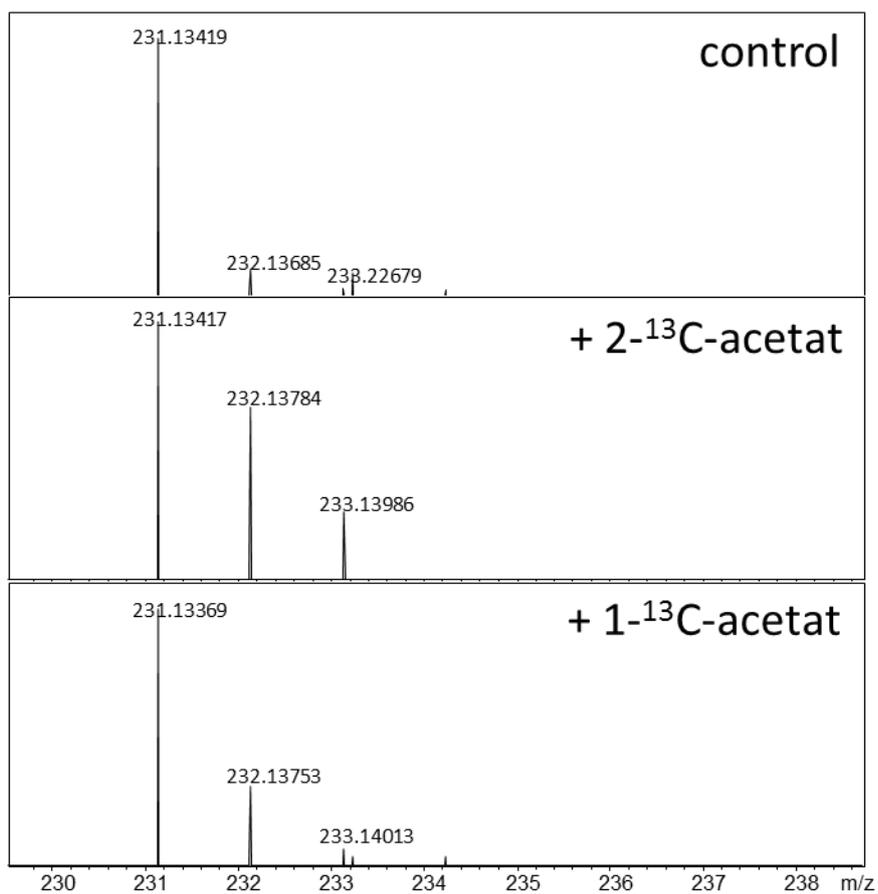


Figure 13: Supplementary figure 48: Isotope pattern of myxoglucamide 645 showing single charged mass [M+H]<sup>+</sup> extracted from LC-hrMS from the control culture and cultures supplemented with L-aspartic acid <sup>13</sup>C<sub>4</sub><sup>15</sup>N<sub>1</sub>, L-glutamic acid d<sub>5</sub> and L-glutamine <sup>15</sup>N<sub>2</sub><sup>13</sup>C<sub>5</sub>



Supplementary figure 49: Isotope pattern of myxoglucamid 645 (A) and Myxoglucamid 661 (B) showing single charged mass  $[M+H]^+$  extracted from LC-hrMS from the control culture and cultures supplemented with valine  $d_8$  and leucine  $d_3$ .

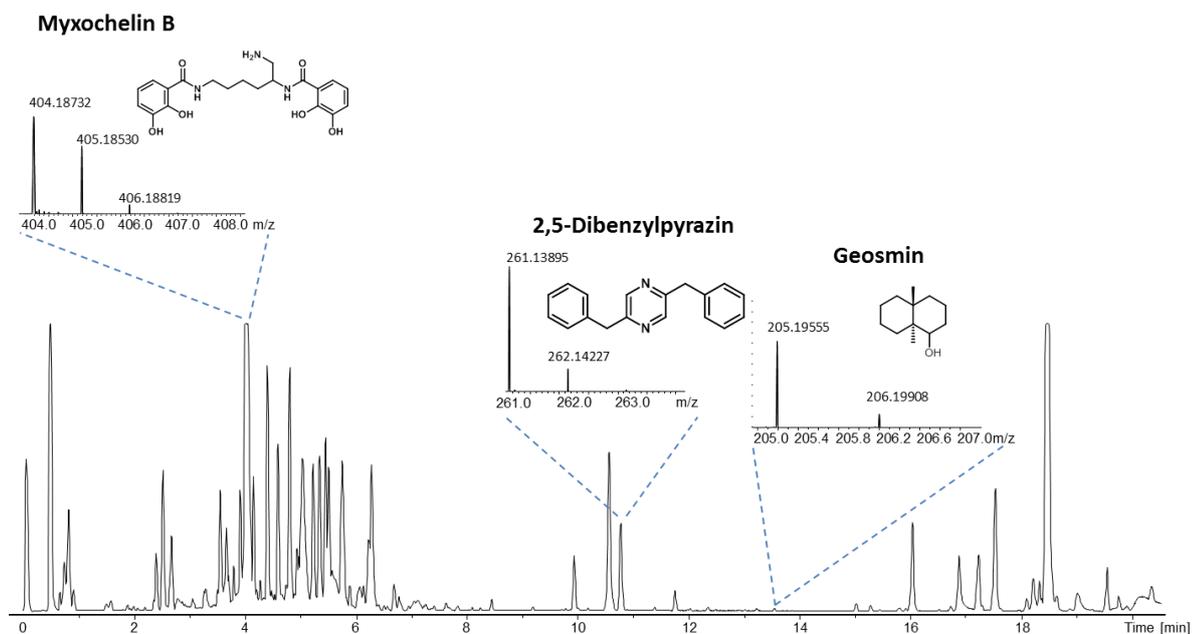
MS<sup>2</sup> fragmentation of myxopentacine 661 provided information about acetate incorporation of the peptide part in the Myxoglucamide compound class.



Supplementary figure 50: Isotope pattern of Fragment 231 of Myxoglucamied 661 after MS<sup>2</sup> fragmentation of control culture and cultures supplemented with 2-<sup>13</sup>C acetate and 1-<sup>13</sup>C acetate. MS<sup>2</sup> spectra are extracted from LC-hrMS measurements in autoMS<sup>2</sup> mode.

#### Metabolites produced by MCy 9003 WT

The crude extract of MCy 9003 was screened for known metabolites using the *inhouse* database "Myxobase". The hits were displayed in the following supplementary figure.



Supplementary figure 51: LC-hrMS measurement of a crude extract of MCy 9003. High resolution Mass spectra of identified metabolites are displayed according to their peaks. In contrast to other metabolites geosmin mass peak shows not the protonated molecule but the  $[M+Na]^+$  ion.

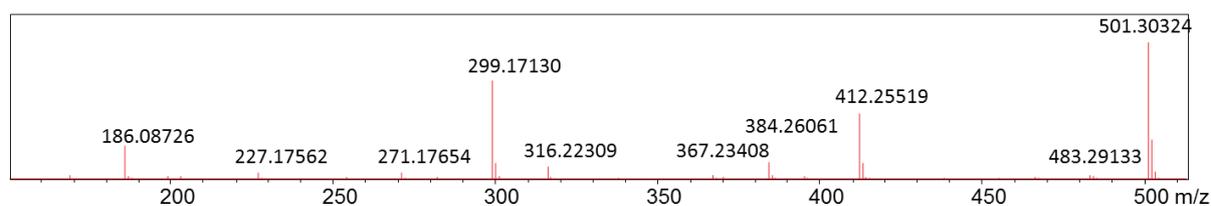
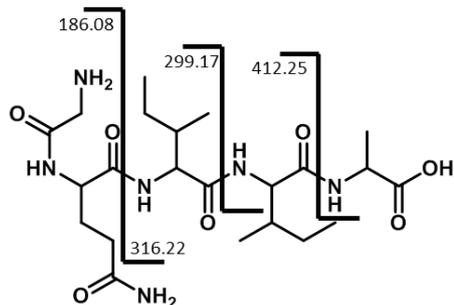
### Natural products derived from other BGC activation

Beside the myxoglucamides, the selection criteria were fulfilled for two other BGCs. This resulted in the discover of peptides 501 and 487, aswell as in the discovery of peptide 472.

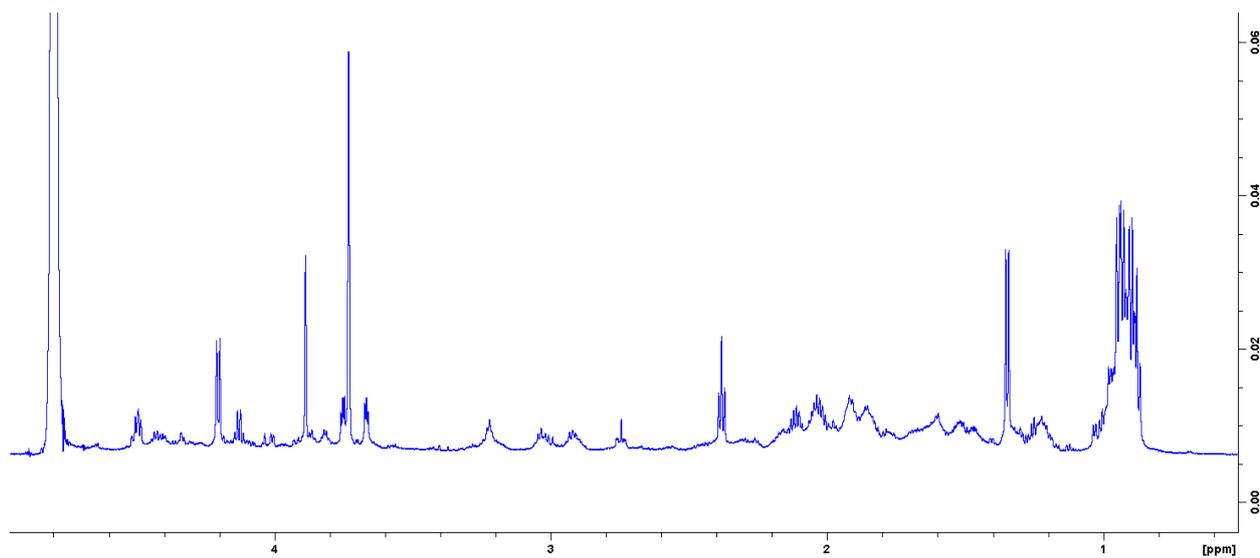
### Structure Peptides 501, 487

#### Peptide 501

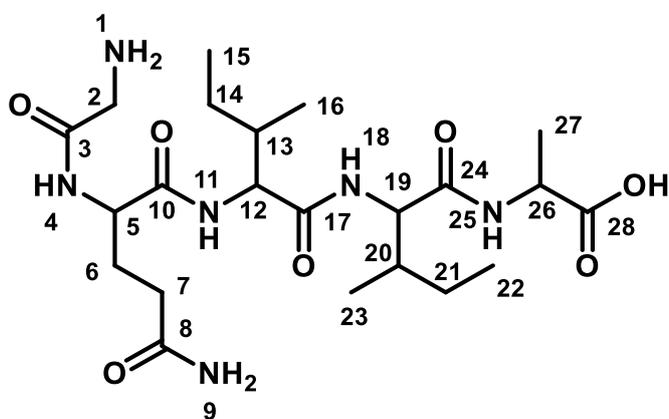
The structure was elucidated in combination of NMR Data and MS<sup>2</sup> fragmentation data.



Supplementary figure 52: hrMS<sup>2</sup> spectrum and observed fragments of peptide 501. MS<sup>2</sup> spectrum is extracted from LC-hrMS measurement of an extract of MCy 9003 mutant tn5 NRPS-t1pks33.



Supplementary figure 53:  $^1\text{H-NMR}$  spectra of Peptide 501 acquired in  $\text{D}_2\text{O}$  at 700 MH



Supplementary table 10: Spectroscopic data of Peptide 501 in  $\text{D}_2\text{O}$ .

position	$\delta_{\text{C}}^{\text{a}}$	$\delta_{\text{H}}^{\text{b}}$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
1	NH			
2	40.3	3.89 s		
3	166.9			
4	NH			
5	52.8	4.49 d,d (8.2, 8.2)	6a, b	3,6,7,10
6a	27.0	2.12 m	5, 6b	5,7,8,10
6b	27.0	2.03 m	5, 6a	5,7,8,10
7	30.9	2.38 t (7.4)	6a,b	5,6,8
8	177.6			
9	NH			
10	173.1			
11	NH			
12	58.4	4.20 d (8.4)	13	10,13,14,16,17
13	35.7	1.92 m	12,16	
14a	24.4	1.51 m	14b, 15	
14b	24.4	1.21 m	14a, 15	
15	9.9	0.89 m		
16	14.6	0.95-0.92	13	12,13, 14
17	172.0			

18	N-H			
19	58.4	4.20 d (8.45)	20	17, 20, 21, 23, 24
20	35.5	1.92 m	20, 23	
21a	24.4	1.51 m	21b, 20	
21b	24.4	1.21 m	21a, 20	
22	9.9	0.89 m	21	20, 21
23	14.6	0.95-0.92	20	19, 20, 21
24	172.0			
25	N-H			
26	51.5	4.13 q (7.25)	27	24, 27, 28
27	17.3	1.35 d (7.3)	26	26, 28
28	179.6			

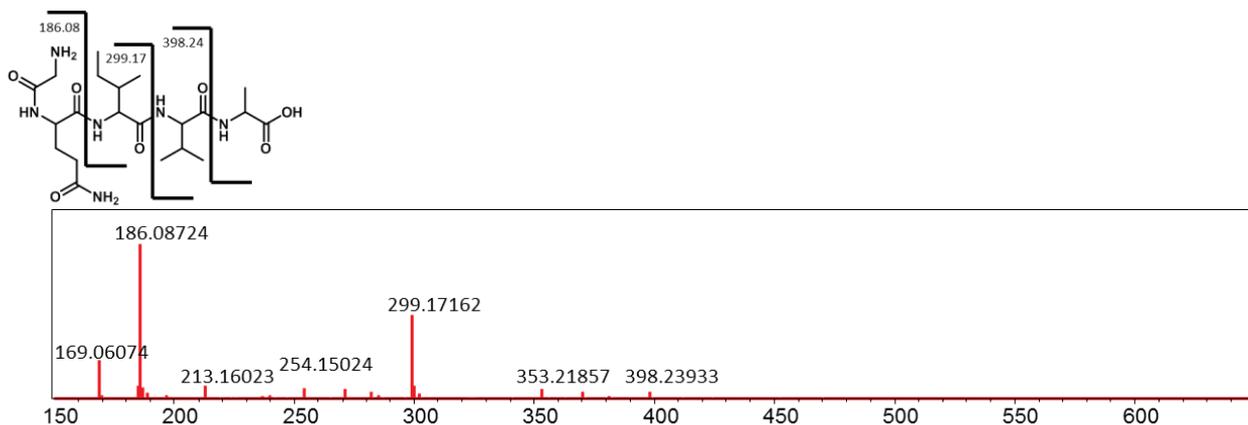
<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra.

<sup>b</sup> acquired at 700 MHz.

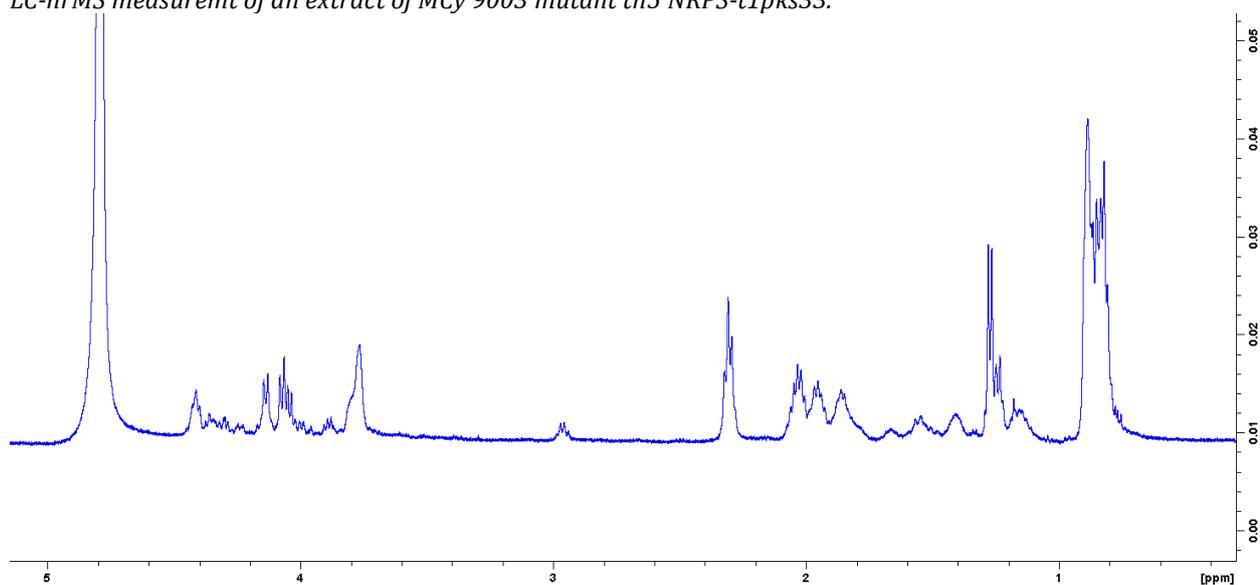
<sup>c</sup> proton showing COSY correlation to indicated proton.

<sup>d</sup> proton showing HMBC correlations to indicated carbons.

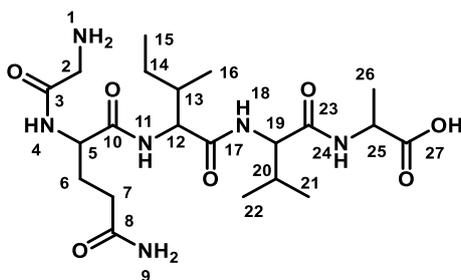
### Peptide 487



Supplementary figure 54: hrMS<sup>2</sup> spectrum and observed fragments of peptide 501. MS<sup>2</sup> spectrum is extracted from LC-hrMS measurement of an extract of MCy 9003 mutant tn5 NRPS-t1pks33.



Supplementary figure 55: Supplementary figure 56: <sup>1</sup>H-NMR spectra of Peptide 501 acquired in D<sub>2</sub>O at 500 MHz



Supplementary table 11: Supplementary table 12: Spectroscopic data of Peptide 487 in D<sub>2</sub>O.

position	$\delta_C^a$	$\delta_H^b$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
1	NH			
2	40.3	3.79 s		3
3	167.2			
4	NH			
5	52.9	4.41 m		
6a	26.8	2.05 m	5,6b,7	
6b	26.8	1.95 m	5,6a,7	
7	30.8	2.31 d,d (6.9,6.9)	6	5,6,9
8	177.8			
9	NH			
10	Signal missing			
11	NH			
12	58.5	4.13 d (8.3)	13	13,17
13	35.6	1.85 m	12,14b	
14a	24.3	1.41 m	14b	
14b	24.3	1.15 m	14a	
15	9.6	0.82 m		
16	14.6	0.85 m		
17	173.4			
18	N-H			
19	59.3	4.05 m	20	20,21,22,23
20	29.9	1.92 m		
21	17.6	0.88 m	20	19,20,22
22	17.6	0.88 m	20	19,20,21
23	171.9			
24	N-H			
25	51.0	4.04 m		
26	17.2	1.27 d (6.9)		27
27	179.8			

<sup>a</sup> acquired at 125 MHz and assigned from 2D NMR spectra.

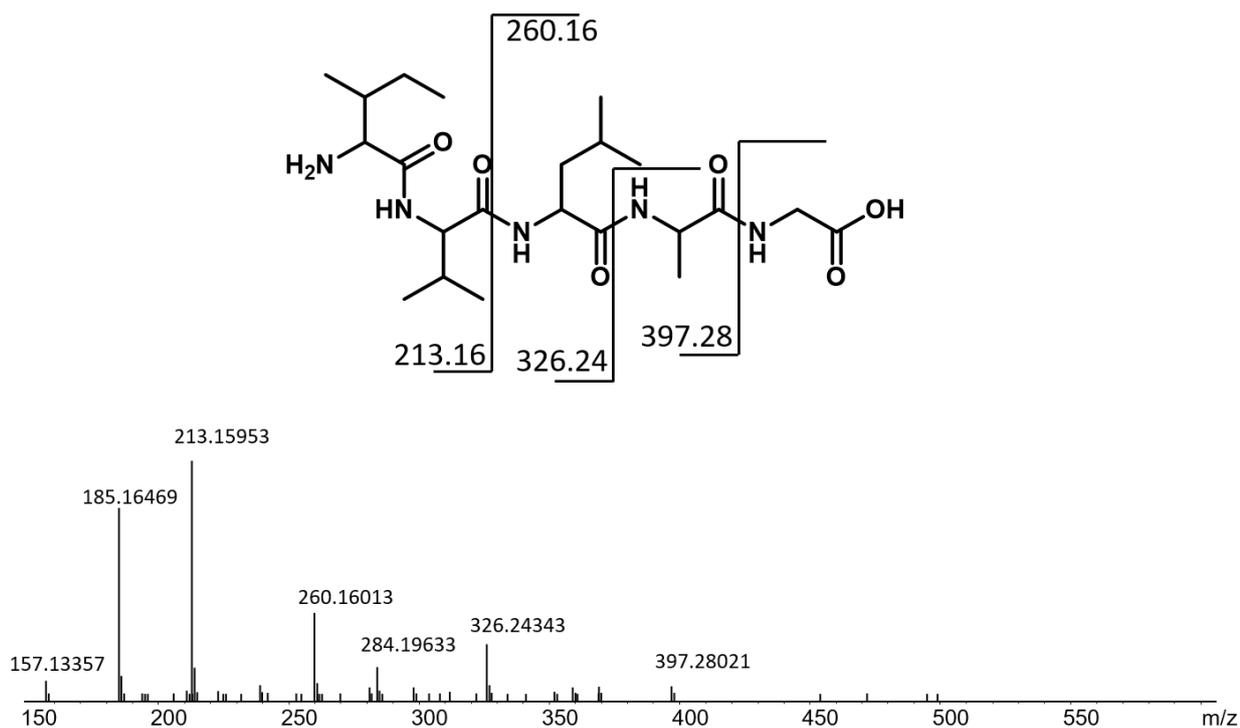
<sup>b</sup> acquired at 500 MHz.

<sup>c</sup> proton showing COSY correlation to indicated proton.

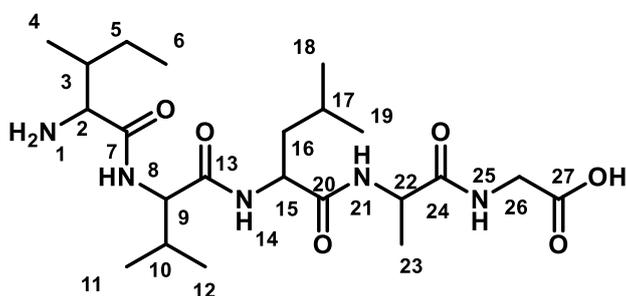
<sup>d</sup> proton showing HMBC correlations to indicated carbons.

### **Structure Peptide 472**

The structure of peptide 472 was elucidated in combination of NMR Data and MS<sup>2</sup> fragmentation data.



Supplementary figure 57: hrMS<sup>2</sup> spectrum and observed fragments of peptide 501. MS<sup>2</sup> spectrum is extracted from LC-hrMS measurement of an extract of MCy 9003 mutant tn5 NRPS-t1pks34B.



Supplementary table 13: Spectroscopic data of Peptide 472 in CD<sub>3</sub>OD

position	$\delta_c^a$	$\delta_H^b$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
1	NH			
2	58.9	3.84 d (5.9)	3	3,4,5,7
3	37.8	1.93 m	2,4,5a	2,4,6
4	15.1	1.06 d (7.0)	3	3,5
5a	25.0	1.58 m	3,5b,6	
5b		1.23 m	5a,6	3,4,6
6	11.1	0.97	5a/b	3
7	171.0			
8	NH			
9	60.7	4.19 d (5.9)	10	7,10,11,12,13
10	31.6	2.16 m	9,11,12	3
11	19.8	1.00 d (7.0)	3	9, 11,12,13
12	18.0	0.96 m	3	
13	173.6			

14	NH			
15	53.5	4.25 d,d (5.4, 9.5)	3a/b	16,17,20
16a	40.2	1.66 m		15,17,18,19, 20
16b		1.70		
17	25.6	1.66	18,19	
18	23.1	0.97 m	17	
19	21.7	0.93 d (6.6)	17	
20	173.6			
21	NH			
22	52.4	4.45 q (7.2)	23	20,23,24
23	17.4	1.39 d (17.4)	22	22,24
24	174.2			
25	NH			
26a	44.1	3.91 d (16.9)	26b	24,27
26b		3.54 d (17.3)	26a	24,27
27	176.4			

<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

<sup>b</sup> acquired at 700 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

<sup>c</sup> proton showing COSY correlation to indicated proton.

<sup>d</sup> proton showing HMBC correlations to indicated carbons.

Structure elucidation of peptide 619, 691, 717 and 743 did not succeeded.

## References

- [1] T. R. Hoye, C. S. Jeffrey, F. Shao, *Nat. Protoc.* **2007**, 2, 2451.
- [2] C. P. Butts, C. R. Jones, E. C. Towers, J. L. Flynn, L. Appleby, N. J. Barron, *Organic & biomolecular chemistry* **2011**, 9, 177.

## Chapter 4 Myxopentacin

### 4.1 Discovery of the Myxopentacin family

The search for new natural products and optimization of production plays a major role in natural products research. Methods such as the overexpression of a biosynthetic gene cluster (BGC), the usage of chemical elicitors or co-cultivation were proven to be effective in achieving these goals<sup>[1]</sup>. All these approaches address the regulatory mechanisms of BGCs in order to unlock them from natural restrictions. For myxobacteria, some regulators for secondary metabolite production are described, like the ChiR regulator<sup>[2]</sup> or HsfA<sup>[3]</sup>, but the majority of regulators of secondary metabolite production in myxobacteria remains elusive. In general, by insertion of artificial promoters in front of the genetic operons within the BGC, limiting regulatory effects are expected to be removed and the metabolic product of the cluster can simply be identified by increased production compared to the wildtype. However, for many BGC “activations” in this study, no product was found but some systemic changes in the metabolome were registered. As described in chapter 3, the genome mining approach by overexpressing BGCs resulted in the discovery of a novel group of glycolipopeptides. In addition to this discovery, we observed some unspecific effects of the implemented genome manipulations. Detailed analysis of LC-hrMS chromatograms of WT extracts and extracts from different activation mutants (Chapter 3) revealed a group of 12 mass trace peaks that exhibited an increased peak area of 100 % or more in case of three mutants. In case of other activation mutants, these peaks showed variation of up to 100 %. All corresponding masses had two features in common: The double charged ion of all identified masses was the most intense ion type and MS<sup>2</sup> spectra of the masses shared a fragment with the mass of 223.14 [M+H]<sup>+</sup>. Furthermore, in all MS<sup>2</sup> spectra, a characteristic mass difference of 111.069 Da between some fragments indicated structural similarities between the identified masses.

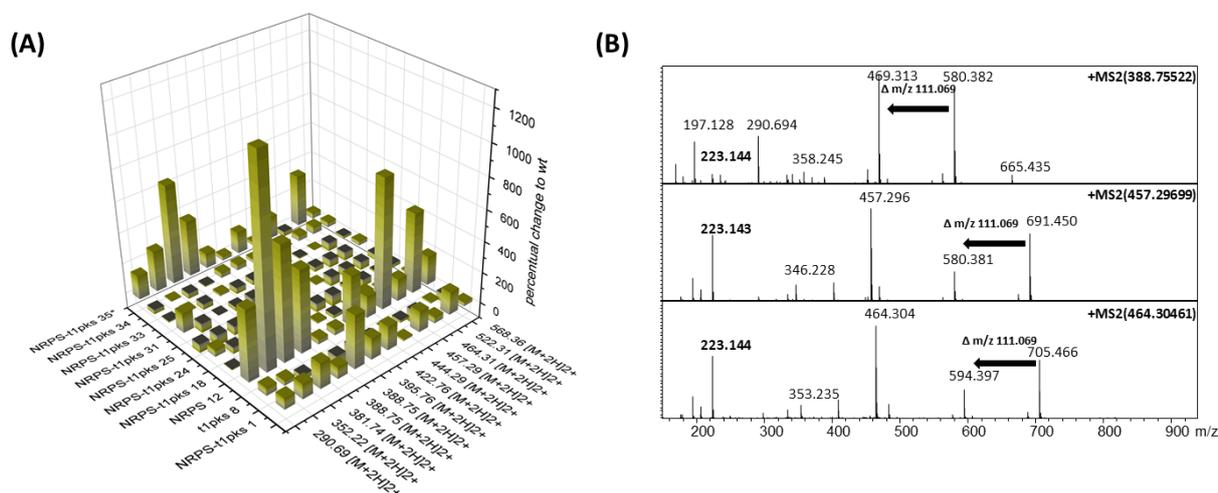


Figure 1: (A) Percentual change of peak areas of identified masses in extracts of the activation mutants compared to the wildtype extracts. \*In case of NRPS-t1pks 35, a vanillate promotor was used instead of a *tn5* promotor. The difference refers to mutant cultures supplemented with vanillate and mutant cultures without vanillate. (B) Three representative MS<sup>2</sup> fragmentation spectra showing the common fragment 223.14 m/z and the common difference between the fragments of 111.07 Da. All spectra originate from the fragmentation of the double charged mass [M+2H]<sup>2+</sup> ion and were extracted from LC-hrMS measurements of extracts of MCy 9003.

Since increased production of the same group of masses occurred after different BGC activations, it did not seem likely that the observed masses were the direct metabolic products of these clusters, but rather were an unspecific effect of the activation. Dereplication using our *inhouse* database and the DNP did not yield any

known compounds. Therefore, we decided to purify and elucidate the structures of the respective compounds. Five out of the twelve identified compounds were produced in a high enough yield for isolation. Three compounds with the smaller masses (myxopentacin 761, myxopentacin 776A and myxopentacin 776B) were isolated from a 2.1 l culture of tn5 NRPS12 mutant, cultivated for 12 days in CyS media with XAD-16 adsorber resin, while the two compounds with the larger masses were isolated from a 2 l culture of MCy 9003 WT cultivated for 12 days in RG5 media with XAD-16 adsorber resin.

The mutant tn5 NRPS12 was created in the context which is described in chapter 3. In case of both cultivations, cells and resin were separated from the supernatant and extracted using a mixture of acetone and MeOH. The molecules were isolated from the organic extract using a mixture of size exclusion chromatography and semipreparative HPLC. The structure of the isolated molecules were determined by 2D NMR experiments. All isolated compounds contained several cyclopentyl rings and therefore, we named the compound class myxopentacins. All characterized myxopentacin derivatives differ significantly in length and composition of the peptide chain.

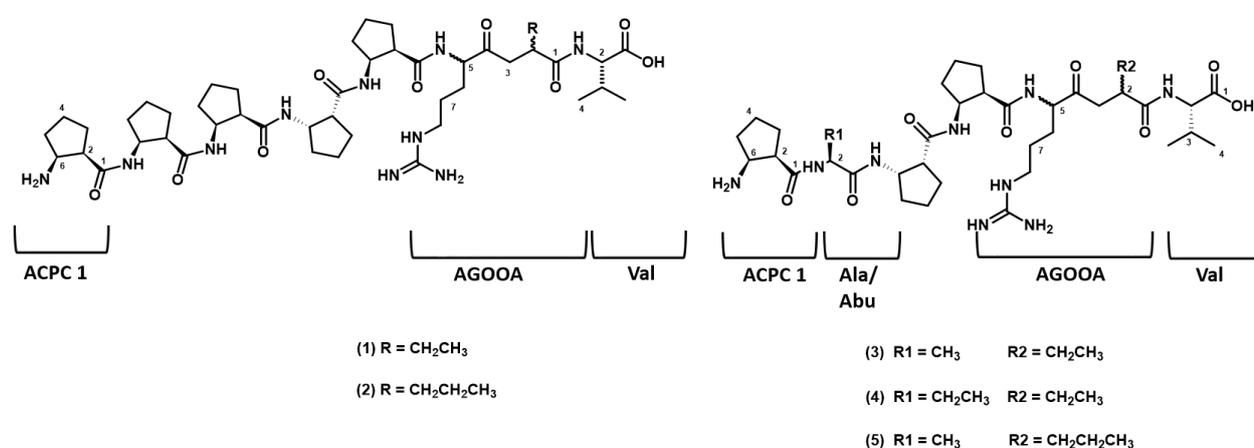


Figure 2: Structures of Myxopentacin 913 (1), 927 (2), 762 (3), 776a (4), 776b (5)

## 4.2 Structure elucidation of the Myxopentacins

### Myxopentacin 913

Myxopentacin 913 (1) was isolated as a white solid. The high resolution mass spectrum showed an intense signal for the double charged ion  $[M+2H]^{2+}$  at 457.297  $m/z$  and a weak signal for the single charged ion  $[M+H]^+$  at 913.586  $m/z$ . Using data analysis software, the sum formula of C<sub>46</sub>H<sub>78</sub>N<sub>10</sub>O<sub>9</sub> was determined for the double charged ion (calc. 457.29711 measured 457.29713  $\Delta$  = 0.04 ppm)

The proton spectrum of 1 in CD<sub>3</sub>OD exhibited a group of overlapping signals between  $\delta_H$  4.44-4.40 and  $\delta_H$  4.32-4.26, indicating the presence of peptidic  $\alpha$ -protons. In addition, three isolated signals were found at  $\delta_H$  4.36 (H-6<sub>ACPC5</sub>),  $\delta_H$  4.09 (H-2<sub>Val</sub>) and  $\delta_H$  3.67 (H-6<sub>ACPC1</sub>), also characteristic for peptidic  $\alpha$ -protons. Furthermore, a signal at  $\delta_H$  3.17 (H<sub>2</sub>-8<sub>AGOOA</sub>) suggested the presence of two methylene protons with an adjacent nitrogen due to the indicative downfield shift. At  $\delta_H$  2.90 (H-2<sub>ACPC5</sub>), a methine proton was observed and next to this signal, a group of overlapping signals between  $\delta_H$  2.80 and 2.84 were also considered to be methine protons based on the shift values. Between  $\delta_H$  2.09-1.50, a large group of overlapping signals was found most likely representing methylene protons. Moreover, a signal characteristic for a methyl group was identified at  $\delta_H$  0.96 (H<sub>3</sub>-2'<sub>AGOOA</sub>) together with a signal implying the presence of two chemically equivalent methyl groups at  $\delta_H$  0.93 (H<sub>3</sub>-4,5<sub>Val</sub>).

Examination of the HSQC spectrum revealed three additional signals characteristic for  $\alpha$ -positions of amino acids at  $\delta_H$  4.42,  $\delta_C$  53.3 (pos-6<sub>ACPC2</sub>),  $\delta_H$  4.29,  $\delta_C$  53.7-54.0 (pos-6<sub>ACPC3+4</sub>), and  $\delta_H$  4.43,  $\delta_C$  58.5 (pos-5<sub>AGOOA</sub>). The signal of the second  $\alpha$ -positions was more intense than the other signals suggesting two  $\alpha$ -positions with

identical shift values. In addition, a diastereotopic methylene group with proton signals at  $\delta_{\text{H}}$  3.06 and 2.57 (H-3a/b<sub>AGOOA</sub>) and a carbon resonance at  $\delta_{\text{C}}$  42.7 (C-3<sub>AGOOA</sub>) was found. The overlapping proton signals between  $\delta_{\text{H}}$  2.80-2.84 belonged to two methine signals in the HSQC spectrum,  $\delta_{\text{H}}$  2.82,  $\delta_{\text{C}}$  46.6 (pos-2<sub>ACPC1</sub>) and  $\delta_{\text{H}}$  2.82,  $\delta_{\text{C}}$  49.1 (pos-2<sub>ACPC2+3+4</sub>). The intensity of the second signal suggested it to be derived from methine groups with identical shift values. The characteristic shift values of these methine groups imply an adjacent nitrogen. Furthermore, two additional methine groups were found in the HSQC spectrum:  $\delta_{\text{H}}$  2.08,  $\delta_{\text{C}}$  32.3 (pos-3<sub>Val</sub>) and  $\delta_{\text{H}}$  2.78,  $\delta_{\text{C}}$  44.2 (pos-2<sub>AGOOA</sub>). Examination of the overlapping area between  $\delta_{\text{H}}$  2.09-1.50 showed several groups of signals that belong to chemically very similar diastereotopic methylenes. This finding suggested the repeated occurrence of the same moiety, which was later identified as 1,2-amino-cyclopentyl carboxylic acid moiety (ACPC). The first group of methylenes has proton signals at  $\delta_{\text{H}}$  1.8 and 2.0 (H-3a/b<sub>ACPC</sub>) and carbon resonances between 28 and 29 ppm (C-3<sub>ACPC</sub>). The second group shows proton signals at  $\delta_{\text{H}}$  1.8 and 1.9 (H-4a/b<sub>ACPC</sub>) and carbon signals between 22 and 23 ppm (C-4<sub>ACPC</sub>). The third group shows proton signals at  $\delta_{\text{H}}$  1.6 and 1.9 (H-5a/b<sub>ACPC</sub>) and carbon resonances between 31.5 and 33 ppm (C-5<sub>ACPC</sub>). Within these strongly overlapping diastereotopic methylene groups, three additional methylenes were identified at  $\delta_{\text{Ha}}$  1.57,  $\delta_{\text{Hb}}$  1.85,  $\delta_{\text{C}}$  27.7 (pos-6a/b<sub>AGOOA</sub>),  $\delta_{\text{Ha}}$  1.48,  $\delta_{\text{Hb}}$  1.57,  $\delta_{\text{C}}$  25.3 (pos-7a/b<sub>AGOOA</sub>) and  $\delta_{\text{Ha}}$  1.64,  $\delta_{\text{Hb}}$  1.51,  $\delta_{\text{C}}$  26.8 (pos-6a/b<sub>AGOOA</sub>).

As mentioned previously, this data suggested that the compound contains an unknown amino acid moiety that occurs multiple times in the structure. To determine the nature of this amino acid, COSY correlations from the isolated  $\alpha$ -proton signal of H-6<sub>ACPC5</sub> were investigated. The COSY showed crosspeaks with the isolated resonance of H-2<sub>ACPC5</sub> and resonances of the overlapping diastereotopic methylene protons H-5a/b<sub>ACPC</sub>. As indicated by further observations in the COSY spectrum, H-2<sub>ACPC5</sub> shows crosspeaks with the overlapping diastereotopic methylene protons H-5a/b<sub>ACPC</sub>, indicating both positions are neighboring. In addition, from the isolated  $\alpha$ -proton signal H-6<sub>ACPC</sub>, HMBC correlations to carbon resonances of overlapping methylene carbons C-3<sub>ACPC</sub>, C-4<sub>ACPC</sub>, C-5<sub>ACPC</sub> can be observed. In turn, the neighboring proton H-2<sub>ACPC5</sub> showed correlations to the same resonances. Moreover, from H-2<sub>ACPC5</sub>, an HMBC correlation to an isolated carbonyl resonance at  $\delta_{\text{C}}$  175.1 (C-1<sub>ACPC5</sub>) was found. The observed correlations and chemical shifts identify the unknown amino acid as 1,2-amino-cyclopentyl carboxylic acid moiety (ACPC). Examination of COSY and HMBC correlations of the other  $\alpha$ -positions indicated  $\delta_{\text{H}}$  4.42,  $\delta_{\text{C}}$  53.3 (pos-6<sub>ACPC1</sub>),  $\delta_{\text{H}}$  3.67,  $\delta_{\text{C}}$  54.6 (H-6<sub>ACPC1</sub>) and  $\delta_{\text{H}}$  4.29,  $\delta_{\text{C}}$  53.7-54.0 (pos-6<sub>ACPC3+4</sub>) to be part of ACPC amino acids as well.

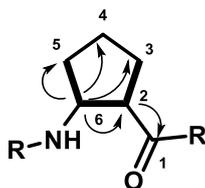


Figure 3: Key COSY (bold line) and HMBC (arrow) correlations of ACPC moiety

The amino acid with the  $\alpha$ -position at  $\delta_{\text{H}}$  4.43,  $\delta_{\text{C}}$  58.5 (pos-5<sub>AGOOA</sub>) was deduced using COSY and HMBC correlations. Examination of COSY crosspeaks revealed H-5<sub>AGOOA</sub> to be part of a spin system consisting of two sets of diastereotopic methylene protons H-6a/b<sub>AGOOA</sub> and H-7a/b<sub>AGOOA</sub> as well as the methylene protons H<sub>2</sub>-8<sub>AGOOA</sub>. HMBC correlations from the methylene protons H<sub>2</sub>-8<sub>AGOOA</sub> to a characteristic carbon resonance at  $\delta_{\text{C}}$  158.3 (C-9<sub>AGOOA</sub>) indicate a guanidino-group, which can also be found in arginine. Unlike the  $\alpha$ -position in arginine, the pseudo- $\alpha$ -position 5<sub>AGOOA</sub> is located next to a keto group, as demonstrated by HMBC correlations to a carbonyl resonance at  $\delta_{\text{C}}$  208.2 (C-4<sub>AGOOA</sub>). The diastereotopic methylene protons H-3/b<sub>AGOOA</sub> also show HMBC correlations to the same carbonyl carbon, indicating the proximity of this methylene group to C-4<sub>AGOOA</sub>.

Moreover, H-3a/b<sub>AGOOA</sub> show COSY correlations to H-2<sub>AGOOA</sub>, which exhibits further correlations to an ethyl group, consisting of two diastereotopic protons H-1'a/b<sub>AGOOA</sub> and the three methyl protons H<sub>3</sub>-2'<sub>AGOOA</sub>. The two diastereotopic protons H-1'a/b<sub>AGOOA</sub> show an HMBC correlation to a carbonyl resonance at  $\delta$  176.4 (C-1<sub>AGOOA</sub>). Regarding the previously described connectivity, this carbonyl group can only be located next to pos-2<sub>AGOOA</sub>. The described chemical shifts and correlations indicate this moiety to be 5-amino-2-ethyl-8-guanidino-4-oxooctanoic acid (AGOOA). The remaining unassigned  $\alpha$ -proton H-2<sub>Val</sub> was determined to be part of a valine residue based on characteristic COSY correlations to the methine proton H-3<sub>Val</sub> and further COSY correlations from H-3<sub>Val</sub> to two sets of methyl protons (H<sub>3</sub>-4,5<sub>Val</sub>).

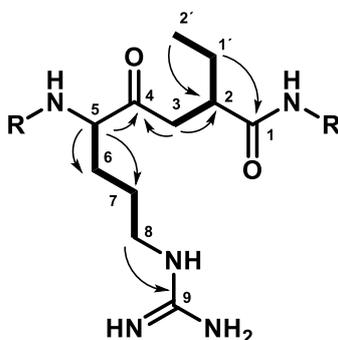


Figure 4: Key COSY (bold line) and HMBC (arrow) correlations of AGOOA moiety in myxopentacin variants (1),(3),(4)

Connectivity between the amino acids was established by examination of HMBC correlations. H-2<sub>Val</sub> shows an HMBC correlation to the carbonyl group of the AGOOA moiety (C-1<sub>AGOOA</sub>), indicating both amino acids are linked via an amide bond. The AGOOA moiety is further linked by an amide bond to an ACPC residue as shown by an HMBC correlation to the corresponding carbon resonance of the ACPC (C-1<sub>ACPC5</sub>). As indicated by HMBC correlations from the  $\alpha$ -protons of the ACPC rings to carbonyl resonances of neighboring ACPC rings and MS<sup>2</sup> fragmentation data, the ACPC rings form a chain consisting of five ACPC, which are connected via amide bonds.

### Myxopentacin 927

Myxopentactin 927 (**2**) was isolated as a white solid. The high resolution mass spectrum showed an intense signal for the double charged ion  $[M+2H]^{2+}$  at 464.305  $m/z$  and a weak signal for the single charged ion  $[M+H]^+$  at 927.605  $m/z$ . Using data analysis software the sum formula of C<sub>47</sub>H<sub>80</sub>N<sub>10</sub>O<sub>9</sub> was determined for the double charged ion (calc. 464.30488 measured 464.30494  $\Delta$  = 0.13 ppm)

The mass difference of 14 Da to **1** indicates an additional methylene group in **2**. Indeed, the <sup>1</sup>H spectrum of **2** showed great similarities with the spectrum of **1**. The methyl group of the AGOOA moiety (H<sub>3</sub>-3'<sub>AGOOA</sub>) was shifted slightly upfield to  $\delta_H$  0.92. Examination of the HSQC spectrum revealed an additional methylene group with a proton resonance at  $\delta_H$  1.37 (H<sub>2</sub>-2'<sub>AGOOA</sub>) and a corresponding carbon resonance at  $\delta_C$  21.2. Analysis of COSY correlations indicated the presence of a propyl group consisting of H<sub>3</sub>-3'<sub>AGOOA</sub>, H<sub>2</sub>-2'<sub>AGOOA</sub> and H-1'a/b<sub>AGOOA</sub> instead of the ethyl group, which was found in **1**.

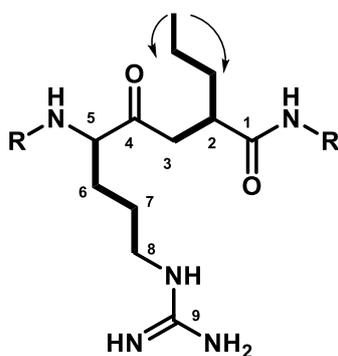


Figure 5: Key COSY (bold line) and HMBC (arrow) correlations of AGOOA moiety in myxopentacin 927(2)

MS<sup>2</sup> fragmentation data resemble the one of **1**, implying five ACPC in a sequence to be present in **2**.

Table 1: Spectroscopic data of myxopentacin 913 and 927

moiety	Pos.	Myxopentacin 913		Pos.	Myxopentacin 927	
		$\delta^{13}\text{C}^a$	$\delta^1\text{H}^b$ (J in Hz)		$\delta^{13}\text{C}^c$	$\delta^1\text{H}^d$ (J in Hz)
ACPC 1	1	174.2		1	174.1	
	2	46.6	2.82 m	2	46.6	2.82 m
	3a	29.7	1.88 m	3a	29.7	1.88 m
	3b	29.7	2.03 m	3b	29.7	2.03 m
	4a	22.4	1.90 m	4a	22.4	1.89 m
	4b	22.4	1.74 m	4b	22.4	1.73 m
	5a	31.6	2.09 m	5a	31.6	2.09 m
	5b	31.6	1.80 m	5b	31.6	1.79 m
6	54.6	3.67 m	6	54.5	3.67 m	
ACPC 2	1	174.6-175.0		1	174.0-175.0	
	2	49.1	2.82 m	2	49.2	2.82 m
	3a	28.5	1.97 m	3a	28.5	1.97 m
	3b	28.5	1.81 m	3b	28.5	1.81 m
	4a	23.5	1.88 m	4a	23.5	1.88 m
	4b	23.5	1.54 m	4b	23.5	1.54 m
	5a	33.0	1.95 m	5a	33.0	1.95 m
	5b	33.0	1.73 m	5b	33.0	1.73 m
6	53.5	4.42 m	6	53.5	4.42 m	
ACPC 3+4	1	174.6-175.1		1	174.6-175.0	
	2	49.1	2.82 m	2	49.1	2.83 m
	3a	28.5	1.96 m	3a	28.5	1.93 m
	3b	28.5	1.80 m	3b	28.5	1.80 m
	4a	23.5	1.86 m	4a	23.5	1.86 m
	4b	23.5	1.54 m	4b	23.5	1.54 m
	5a	32.9	1.90 m	5a	33.0	1.88 m
	5b	32.9	1.63 m	5b	33.0	1.67 m
6	53.7-54.0	4.29 m	6	53.7	4.29 m	
ACPC 5	1	175.3		1	175.4	
	2	49.1	2.90 d,d,d (7.9,7.7,7.6)	2	49.1	2.91 d,d (15.5,7.9)
	3a	28.2	1.94 m	3a	28.2	1.94 m
	3b	28.2	1.86 m	3b	28.2	1.86 m
	4a	23.2	1.85 m	4a	23.2	1.57 m
	4b	23.2	1.57 m	4b	23.2	1.85 m
	5a	33.2	1.89 m	5a	33.2	1.89 m
	5b	33.2	1.64 m	5b	33.2	1.64 m
6	54.0	4.36 d,d,d(6.7,6.8,6.5)	6	53.7	4.36 m	
AGOOA	1	176.4		1	176.4	
	2	44.2	2.78 m	2	42.7	2.84 m
	3a	42.7	3.06 d,d (18.4,10.8)	3a	43.0	3.07 d,d (18.5,11.1)
	3b	42.7	2.57 d,d (18.3,3.6)	3b	43.0	2.58 d,d (18.1,3.7)
	4	208.2		4	208.2	
	5	58.5	4.43 m	5	58.5	4.43 m
	6a	27.7	1.85 m	6a	27.7	1.85 m
	6b	27.7	1.57 m	6b	27.7	1.57 m
	7a	25.3	1.48 m	7a	25.3	1.48 m
	7b	25.3	1.57 m	7b	25.3	1.57 m
	8	41.7	3.17 d,d (6.1,6.7)	8	41.7	3.17 d (6.7)
	9	158.3		9	158.3	
	1'a	26.8	1.64 m	1'a	36.0	1.59 m
1'b	26.8	1.51 m	1'b	36.0	1.42 m	

	2'	11.8	0.96 t (7.3)	2'	21.2	1.38 m
				3'	14.1	0.92 m
	1	178.3		1	178.3	
	2	61.4	4.09 d (5.6)	2	61.4	4.09 d (5.8)
<b>Val</b>	3	32.3	2.08 m	3	32.3	2.08 m
	4	18.4	0.93 d (6.9)	4	18.4	0.93 m
	5	19.9	0.93 d (6.9)	5	19.9	0.93 m

<sup>a</sup> acquired at 125 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

<sup>b</sup> acquired at 500 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

<sup>c</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm. <sup>d</sup> acquired at 700 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

### Structure elucidation of Myxopentacin 762

Myxopentactin 762 (**3**) was isolated as a white solid. The high resolution mass spectrum showed an intense signal for the double charged ion  $[M+2H]^{2+}$  at 381.747  $m/z$  and a weak signal for the single charged ion  $[M+H]^+$  at 762.487  $m/z$ . Using data analysis software, the sum formula of C<sub>37</sub>H<sub>65</sub>N<sub>9</sub>O<sub>8</sub> was determined for the double charged ion (calc. 381.74778 measured 381.74726  $\Delta$  = 1.36 ppm)

The <sup>1</sup>H spectrum of **3** resembled the spectra of the other myxopentacin derivatives. The only major difference was a signal of methyl protons at  $\delta_H$  1.32 (H<sub>3</sub>-3<sub>Ala</sub>) and an additional signal at  $\delta_H$  4.32 (H-2<sub>Ala</sub>) characteristic for an amino acid  $\alpha$ -proton. Using COSY correlations together with the chemical shift values, this amino acid was assigned as alanine. HMBC correlations from H-6<sub>ACPC2</sub> to the alanine carbonyl resonance  $\delta_C$  174.0 (C-1<sub>Ala</sub>) indicated that the alanine is located in the ACPC chain. Using MS<sup>2</sup> fragmentation data, the amino acid sequence was established as ACPC1-Ala-ACPC2-ACPC3-AGOOA-Val.

### Structure elucidation of Myxopentacin 776a

Myxopentactin 776a (**4**) was isolated as a white solid. The high resolution mass spectrum showed an intense signal for the double charged ion  $[M+2H]^{2+}$  at 388.755  $m/z$  and a weak signal for the single charged ion  $[M+H]^+$  at 776.502  $m/z$ . Using data analysis software, the sum formula of C<sub>38</sub>H<sub>67</sub>N<sub>9</sub>O<sub>8</sub> was determined for the double charged ion (calc. 388.75508 measured 388.75522  $\Delta$  = 0.36 ppm)

The <sup>1</sup>H spectrum of **4** resembled spectra the other myxopentacin derivatives, especially the spectrum of **3**. The identified differences included a signal characteristic for an  $\alpha$ -proton at  $\delta_H$  4.16 (H-2<sub>Abu</sub>) and a signal from methyl protons at  $\delta_H$  0.94 (H<sub>3</sub>-3<sub>Abu</sub>). Inspection of the HSQC spectrum revealed two additional diastereotopic protons at  $\delta_H$  1.78 and 1.65 (H-3a/b<sub>Abu</sub>) and a corresponding carbon resonance of  $\delta_C$  25.9 (C-3<sub>Abu</sub>). Analysis of COSY correlations revealed that these new signals are derived from aminobutyric acid (Abu). HMBC correlations from H-2<sub>Abu</sub> to C-1<sub>ACPC1</sub> indicate that the Abu moiety replaces the Ala moiety in myxopentacin 762. In addition, MS<sup>2</sup> fragmentation pattern confirms this finding.

### Structure elucidation of Myxopentacin 776b

Myxopentactin 776b (**5**) was isolated as a white solid. The high resolution mass spectrum showed an intense signal for the double charged ion  $[M+2H]^{2+}$  at 388.756  $m/z$  and a weak signal for the single charged ion  $[M+H]^+$  at 776.704  $m/z$ . Using data analysis software, the sum formula of C<sub>38</sub>H<sub>67</sub>N<sub>9</sub>O<sub>8</sub> was determined for the double charged ion (calc. 388.75508 measured 388.75649  $\Delta$  = 3.63 ppm)

The <sup>1</sup>H spectrum of **5** strongly resembled the spectra of other myxopentacin derivatives. Similarities with the spectra of **2** indicate an identical AGOOA moiety as in **2** with a propyl residue. The signal at  $\delta_H$  1.32 (H<sub>3</sub>-3<sub>Ala</sub>) indicates the presence of an alanine residue as in **3**. The MS<sup>2</sup> fragmentation pattern confirms the amino acid sequence as ACPC1-Ala-ACPC2-ACPC3-AGOOA(propyl)-Val.

Table 2: Spectroscopic data of myxopentacin 762, 776a and 776b

Myxopentacin 762	Myxopentacin 776a	Myxopentacin 776b
------------------	-------------------	-------------------

moiety	Pos.	$\delta^{13}\text{C}^a$	$\delta^1\text{H}^b$ (J in Hz)	Pos.	$\delta^{13}\text{C}^a$	$\delta^1\text{H}^b$ (J in Hz)	Pos.	$\delta^{13}\text{C}^a$	$\delta^1\text{H}^b$ (J in Hz)
ACPC 1	1	174.3		1	174.9		1	174.3	
	2	46.7	2.93 m	2	46.8	2.94 m	2	46.7	2.93 m
	3a	29.1	1.97 m	3a	29.3	1.92 m	3a	29.1	1.97 m
	3b	29.1	2.17 m	3b	29.3	2.16 m	3b	29.1	2.17 m
	4a	22.5	1.93 m	4a	22.6	1.92 m	4a	22.5	1.93 m
	4b	22.5	1.76 m	4b	22.6	1.75 m	4b	22.5	1.76 m
	5a	31.5	2.12 m	5a	31.7	2.11 m	5a	31.5	2.12 m
	5b	31.5	1.85 m	5b	31.7	1.83 m	5b	31.5	1.85 m
6	54.6	3.72 m	6	54.7	3.72 m	6	54.6	3.72 m	
Ala/ Abu	1	174.0		1	173.3		1	174.0	
	2	49.0	4.32 m	2	55.8	4.16 d,d (8.7,5.2)	2	49.0	4.32 m
	3	18.0	1.32 d (7.2)	3a	25.9	1.78 m	3	18.0	1.32 d (7.2)
				3b	25.9	1.65 m			
			4	10.6	0.94 m				
ACPC 2	1	174.6		1	174.6		1	174.6	
	2	49.1	2.84 m	2	48.4	2.83 d,d,d (7.8,7.8,7.8)	2	49.1	2.84 m
	3a	28.7	1.93 m	3a	28.9	1.93 m	3a	28.7	1.93 m
	3b	28.7	1.83 m	3b	28.9	1.83 m	3b	28.7	1.83 m
	4a	23.3	1.85 m	4a	23.4	1.86 m	4a	23.3	1.85 m
	4b	23.3	1.59 m	4b	23.4	1.60 m	4b	23.3	1.59 m
	5a	33.1	1.96 m	5a	33.3	1.86 m	5a	33.1	1.96 m
	5b	33.1	1.67 m	5b	33.3	1.60 m	5b	33.1	1.67 m
6	53.4	4.38 m	6	53.7	4.37 d,d,d (6.9,6.9,6.9)	6	53.4	4.38 m	
ACPC 3	1	175.3		1	175.3		1	175.3	
	2	48.9	2.92 m	2	49.2	2.93 m	2	48.9	2.92 m
	3a	28.6	1.95 m	3a	28.5	1.96 m	3a	28.6	1.95 m
	3b	28.6	1.86 m	3b	28.5	1.87 m	3b	28.6	1.86 m
	4a	23.2	1.85 m	4a	23.2	1.85 m	4a	23.2	1.85 m
	4b	23.2	1.59 m	4b	23.2	1.58 m	4b	23.2	1.59 m
	5a	32.9	1.92 m	5a	32.8	1.91 m	5a	32.9	1.92 m
	5b	32.9	1.70 m	5b	32.8	1.70 m	5b	32.9	1.70 m
6	53.6	4.34 m	6	53.8	4.33 d,d,d (7.0,7.0,7.0)	6	53.6	4.34 m	
AGOOA	1	176.4		1	176.4		1	176.4	
	2	44.3	2.78 m	2	44.2	2.77 m	2	42.5	2.84 m
	3a	42.8	3.06 d,d (10.9,18.2)	3a	42.8	3.07 d,d (18.5,11.0)	3a	43.1	3.06 d,d (18.7, 10.9)
	3b	42.8	2.59 d,d (3.9,18.3)	3b	42.8	2.59 d,d (18.3,3.5)	3b	43.1	2.58 d,d (18.3, 3.4)
	4	208.3		4	208.4		4	208.3	
	5	58.6	4.42 m	5	58.6	4.41 d,d (6.4,5.1)	5	58.6	4.41 m
	6a	27.8	1.86 m	6a	27.8	1.86 m	6a	27.8	1.86 m
	6b	27.8	1.59 m	6b	27.8	1.59 m	6b	27.8	1.59 m
	7a	25.4	1.51 m	7a	25.8	1.58 m	7a	25.4	1.51 m
	7b	25.4	1.57 m	7b	25.8	1.50 m	7b	25.4	1.57 m
	8	41.6	3.17 t (6.5)	8	41.8	3.18 d,d (6.3,6.3)	8	41.6	3.17 t (6.5)
	9	158.7		9	158.4		9	158.7	
	1'a	26.8	1.64 m	1'a	26.8	1.64 m	1'a	36.0	1.59 m
1'b	26.8	1.52 m	1'b	26.8	1.52 m	1'b		1.43 m	
2'	11.8	0.96 t (7.4)	2'	11.7	0.96 m	2'	21.0	1.37 m	
3'						3'	14.0	0.92 m	
Val	1	178.0		1	178.3		1	178.0	
	2	61.1	4.09 d (6.0)	2	61.4	4.09 d(5.6)	2	61.1	4.09 d (6.0)
	3	32.3	2.08 m	3	32.5	2.08 m	3	32.3	2.08 m
	4	18.5	0.93 d (6.8)	4	19.9	0.93 d (4.3)	4	18.5	0.93 d (6.8)
	5	19.9	0.93 d (6.8)	5	18.4	0.92 d (4.6)	5	19.9	0.93 d (6.8)

<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

<sup>b</sup> acquired at 700 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

### 4.2.1 Stereochemistry

The absolute configuration of all five myxopentacin derivatives was determined by acid hydrolysis followed by derivatization with L- and D-1-fluoro-2,4-dinitrophenyl-5-Leucine amide (FDLA) in combination with chromatographic analysis of the reaction mixture<sup>[4]</sup>. As standards, an amino acid mix containing L-alanine, L-valine and L-aminobutyric acid were used. For the ACPC 1*S*,2*S*-, 1*R*,2*R*- and 1*S*,2*R*-configured standards together with a racemic mixture of 1*S*,2*R* and 1*R*,2*S*-ACPC were used for derivatization.

Table 3: Results of Marfey's analysis of myxopentacin derivatives compared to the respective standards

type	m/z	Myxpentacin 762		Myxopentacin 776a		Myxopentacin 776b		Myxopentacin 913		Myxopentacin 927		Standard	
		L-FDLA	D-FDLA	L-FDLA	D-FDLA	L-FDLA	D-FDLA	L-FDLA	D-FDLA	L-FDLA	D-FDLA	L-FDLA	D-FDLA
1R,2S ACPC	424.18	14.94	18.46	14.95	18.46	14.96	18.46	14.96	18.46	14.96	18.46	14.98	18.45
1R,2R ACPC	424.18	x	x	x	x	x	x	x	x	x	x	19.39	17.28
1S,2S ACPC	424.18	x	x	x	x	x	x	x	x	x	x	17.30	19.36
1S,2R ACPC	424.18	x	x	x	x	x	x	x	x	x	x	18.44	14.99
L-Abu	398.16	x	x	15.28	18.33	x	x	x	x	x	x	15.28	18.36
L-Ala	384.16	14.23	16.71	x	x	14.21	16.70	x	x	x	x	14.20	16.69
L-Val	412.19	16.22	19.69	16.22	19.68	16.20	19.67	16.08	19.58	16.22	19.68	16.21	19.65

A comparison of the retention times of the derivatized amino acids in the hydrolysis products and the respective standards revealed that all natural amino acids including amino butyric acid are L-configured. The ACPC moiety is 1R,2S-configured. During acid hydrolysis, isomerization of the ACPC unit occurred. Therefore, the duration was shortened to avoid inconclusive results. Due to a missing standard, the configuration of the AGOOA moiety could not be determined by application of Marfey's method. In addition, ROE correlations from the neighboring L-valine and ACPC5 were not sufficient for the determination of the configuration.

### 4.3 Biosynthesis

The unique structural features of the myxopentacin derivatives raised the question of the biosynthetic origin. In a first approach, cluster 1 and 12 were analyzed since promoter insertion in front of these clusters increased myxopentacin production. Unfortunately, inactivation mutants of both clusters were still able to produce all myxopentacin variants. In a second approach, a "retro-biosynthetic approach" was used to search for a potential BGC based on the myxopentacin structures: The scaffold suggests that the molecules are derived from a NRPS-PKS hybrid containing an arginine-specific module, a polyketide module and a module specific for valine incorporation. However, none of the BGCs predicted by antiSMASH<sup>[5]</sup> resembled the expected biosynthetic logic.

The obvious conclusion was that the biosynthetic logic behind this unique scaffold does not match the classical logic of NRPS/PKS machinery. Therefore, we intended to gain further insights into the biosynthesis of the ACPC and the AGOOA moiety.

#### 4.3.1 ACPC moiety

The first point of interest was the origin of the ACPC moiety. In the beginning, it was investigated if ACPC is present in the crude extract like other canonical amino acids. To do so, Fmoc derivatization reagent was added to the crude extract to derivatize amino acids being present in the crude extract and simplify their detection by LC-MS<sup>[6]</sup>. Indeed, using a racemic mixture of 1R,2S/1R,2S ACPC derivatized with Fmoc, free 1R,2S ACPC was detected in the crude extract (figure 6).

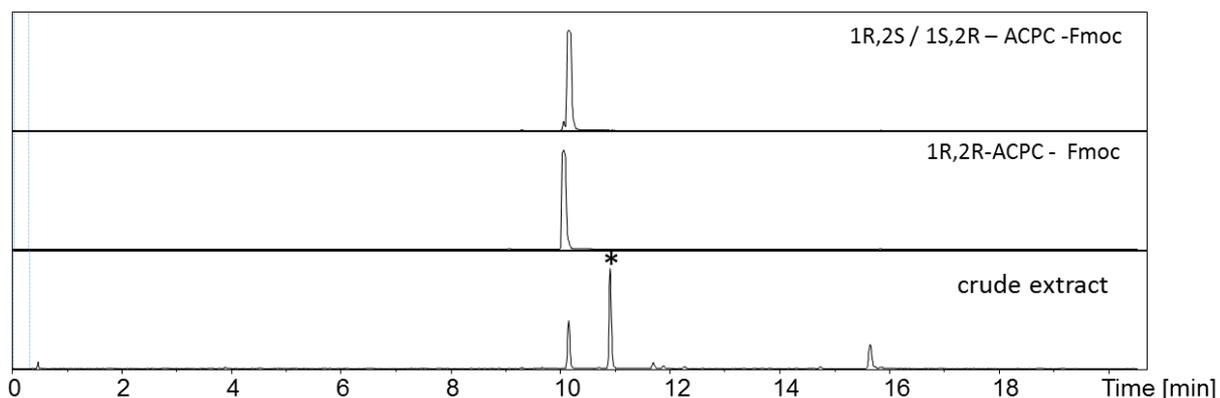


Figure 6: LC-MS chromatogram showing the EIC  $352.14 \pm 0.1$  for racemic standard of 1R,2S/1S,2R ACPC derivatized with Fmoc, the 1R,2R ACPC standard derivatized with Fmoc and the crude extract derivatized with Fmoc. The marked peak is pipercolic acid which has the same mass.

The presence of ACPC in the crude extract raised the question, if it is utilized as a building block and incorporated into the myxopentacin molecule, or if ACPC is formed in the myxopentacin scaffold and released by a hydrolytic cleavage. A feeding experiment would be able to answer this question. Due to a lack of isotope labelled ACPC, a reverse feeding experiment was conducted: Cultures of MCy 9003 were grown in a medium with a  $^{13}\text{C}$  labelled carbon source and fed with a racemic mixture of 1R,2S/ 1S,2R ACPC.

Extracts of the cultures were measured by LC-*hr*MS and the isotope pattern the produced myxopentacin derivatives were examined. Cultures supplemented with ACPC showed a shifted isotope pattern compared to cultures that were grown without the addition of ACPC (Figure 7). The shift of the isotope pattern clearly shows incorporation of ACPC into the myxopentacin molecules. Together with the presence of ACPC in the crude extract, the result implies that ACPC is a building block in the myxopentacin biosynthesis.

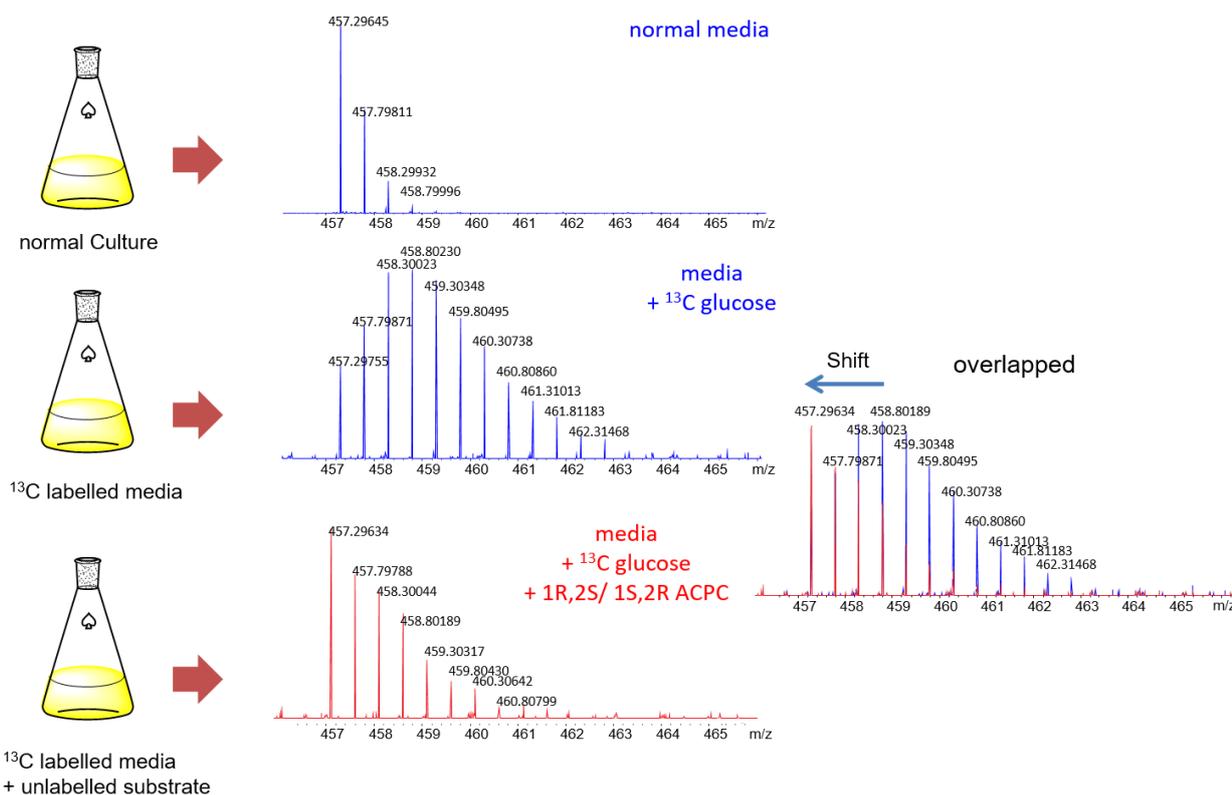
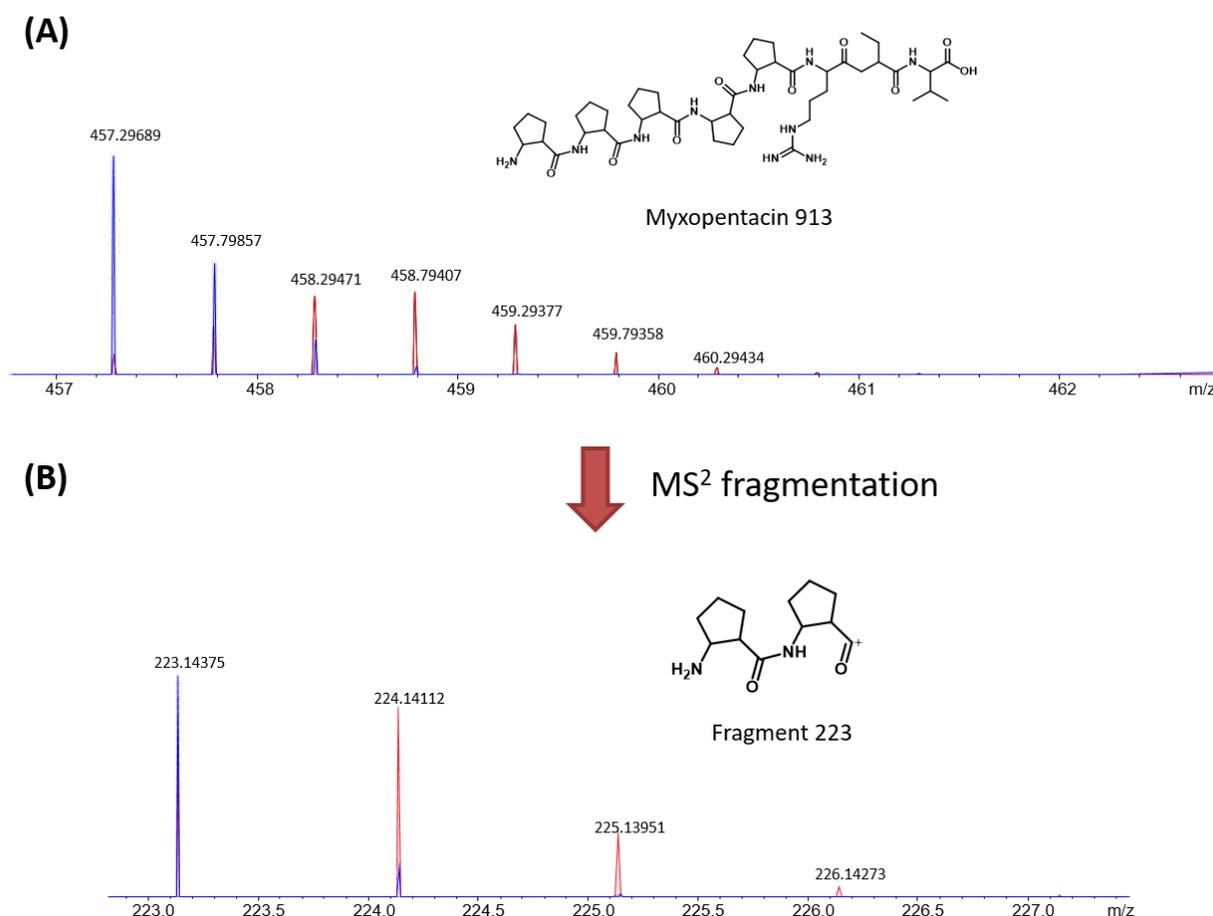


Figure 7: (A) Expectation reverse feeding. (B) Isotope pattern of myxopentacin 913 (**1**) of reverse feeding experiment showing double charged mass  $[M+2H]^{2+}$  extracted from LC-hrMS from the control culture and a culture using  $^{13}C$  Glucose supplemented with racemic *cis* configured ACPC and a control culture with  $^{13}C$  Glucose.

The origin of the ACPC unit was further investigated: The number of carbons suggested ACPC to originate from lysine that underwent a cyclization of yet unknown mechanism. Therefore, a feeding experiment with lysine was performed using  $^{15}N_2^{13}C_6$  isotope labelled L-Lysine. In case of incorporation of the full lysine carbon skeleton and one nitrogen, a mass shift of 35 Da in the isotope pattern would be expected. The LC-hrMS of the extracts of the cultures, however, showed only a small shift of 4 Da in the isotope pattern compared to the control cultures.



*Figure 8: (A) Comparison of the isotope pattern of myxopentacin 913 (**1**) in the control culture (blue signals) and a culture supplemented with isotope labelled lysine (red signals). The double charged mass  $[M+2H]^{2+}$  is shown extracted from LC-hrMS. (B) Isotope pattern of Fragment 223.14  $m/z$  after MS<sup>2</sup> fragmentation is displayed.*

The small shift difference indicates an enrichment of <sup>15</sup>N or <sup>13</sup>C isotopes in the myxopentacin molecule. MS<sup>2</sup> fragmentation of myxopentacin 913 (**1**) shows that one isotope atom from lysine is incorporated into the ACPC ring. To determine whether a carbon or a nitrogen atom in the ACPC is originating from lysine, an isotope analysis using direct infusion Fourier transform ion cyclotron resonance (DI-FTICR) was conducted. Due to the mass defect, each isotope has its specific mass. FT-ICR measurements provide a sufficient resolution to differ between each isotope. The crude extract of the culture fed with <sup>15</sup>N<sub>2</sub><sup>13</sup>C<sub>6</sub> lysine was directly injected into FT-ICR. The precursor mass of myxopentacin 913 (**1**) was selected and fragmented. The first isotope peak of fragment 223.14  $m/z$  was investigated. In lower resolution mass spectra, the <sup>15</sup>N peak and the <sup>13</sup>C merge into one peak while in FT-ICR measurements, both isotope peaks can be distinguished from each other.

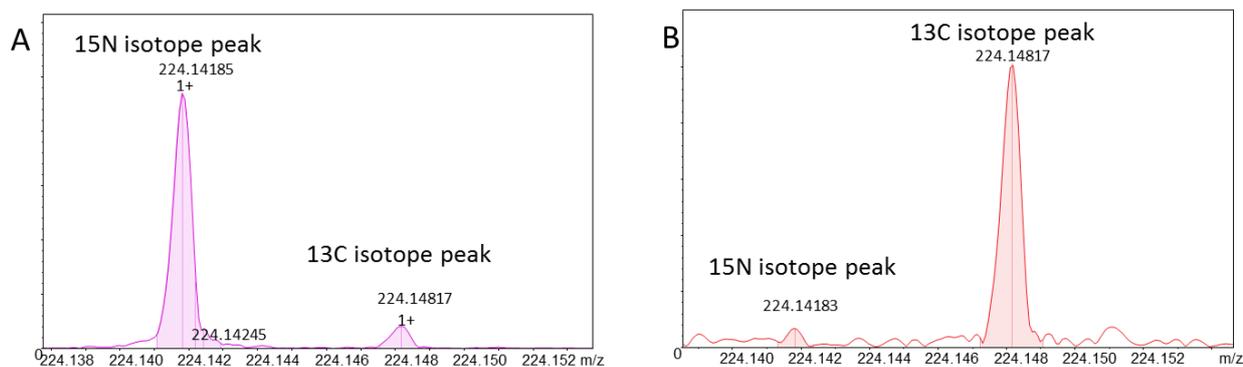
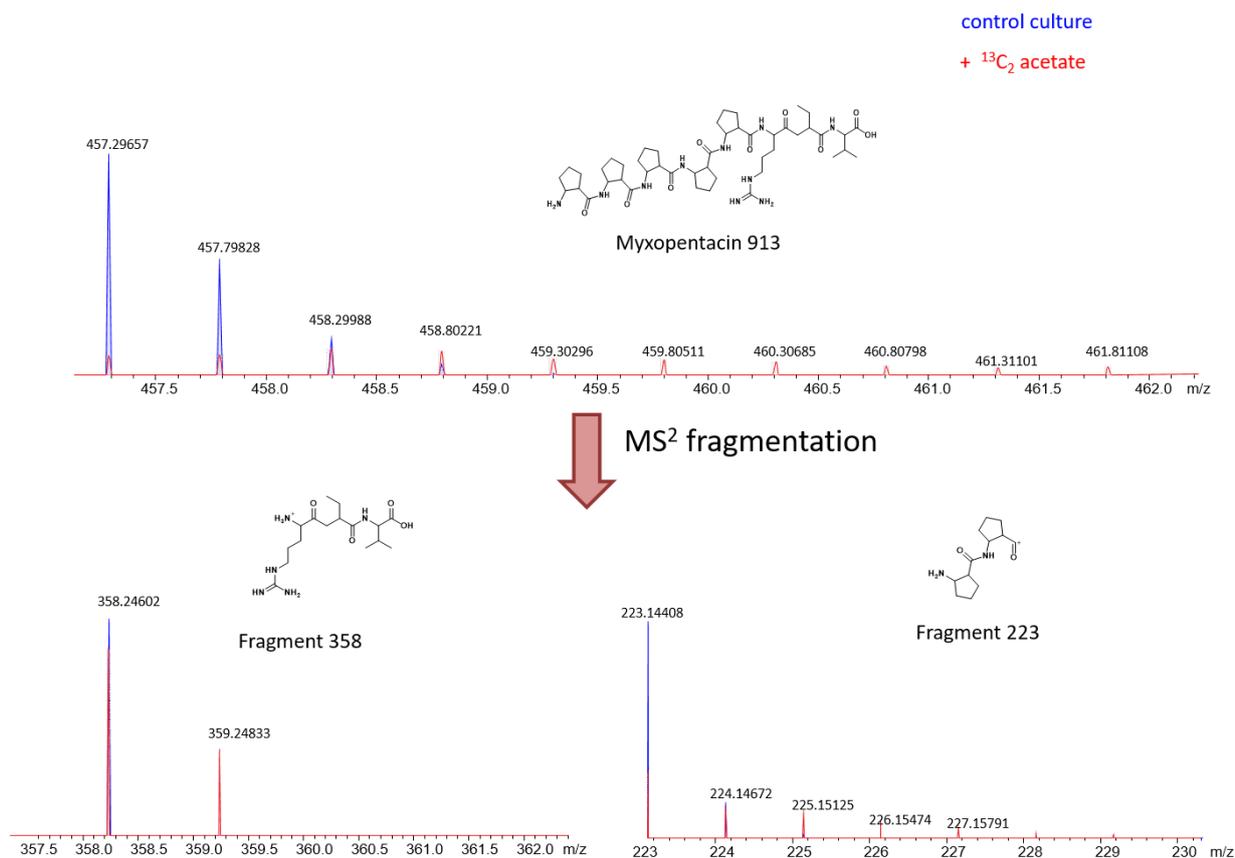


Figure 9: FT-ICR mass spectrum of fragment 223.14 from myxopentacin 913 showing lysine fed culture (A) and control culture (B) Data are obtained by direct injection and collection of the mass 457. 29  $[M+2H]^{2+}$  followed by fragmentation.

Comparison of the FTICR measurement of the lysine fed culture and the control culture showed an increase of the  $^{15}\text{N}$  isotope peak. This result suggests that lysine can function as nitrogen donor in the ACPC biosynthesis. The lysine scaffold itself is not incorporated in ACPC biosynthesis. For cultures supplemented with fully  $^{15}\text{N}$  and  $^{13}\text{C}$  labelled asparagine and glutamine, no change in the isotope pattern of fragment 223.14  $m/z$  was observed, suggesting lysine to be the so far the exclusive nitrogen donor in the ACPC biosynthesis. In cultures supplemented with  $^{13}\text{C}$  labelled acetate, the isotope pattern of myxopentacin 913 (**1**) showed a higher number of incorporated acetate units. The same accounts for cultures fed with 1- $^{13}\text{C}$  and 2- $^{13}\text{C}$  acetate. MS<sup>2</sup> fragmentation revealed that fragment 223.14  $m/z$ , which is a dimer of two ACPC units, showed a high degree of acetate incorporation (figure 10b), indicating an origin from an acetate dependent pathway.



*Figure 10: Isotope pattern of myxopentacin 913 showing double charged mass  $[M+2H]^{2+}$  extracted from LC-hrMS from the control culture (blue signals) and a culture supplemented with  $^{13}\text{C}_2$  labelled acetate (red signals). Furthermore isotope pattern of MS<sup>2</sup> fragments 358 and 223 are displayed.*

In order to ascertain exact positions of acetate incorporation, 100  $\mu\text{g}$  of Myxopentacin 913 (**1**) were purified from a culture supplemented with 1- $^{13}\text{C}$  acetate. A  $^{13}\text{C}$  NMR spectrum of the labelled molecule was acquired in Methanol  $d_4$ . Due to the low abundance of the  $^{13}\text{C}$  isotope, all visible signals except the solvent signal at  $\delta$  49.0 and the formic acid impurity at  $\delta$  170.1 were considered to be derived from incorporated 1- $^{13}\text{C}$  acetate. Based on existing assignments, the acetate incorporation pattern was determined.

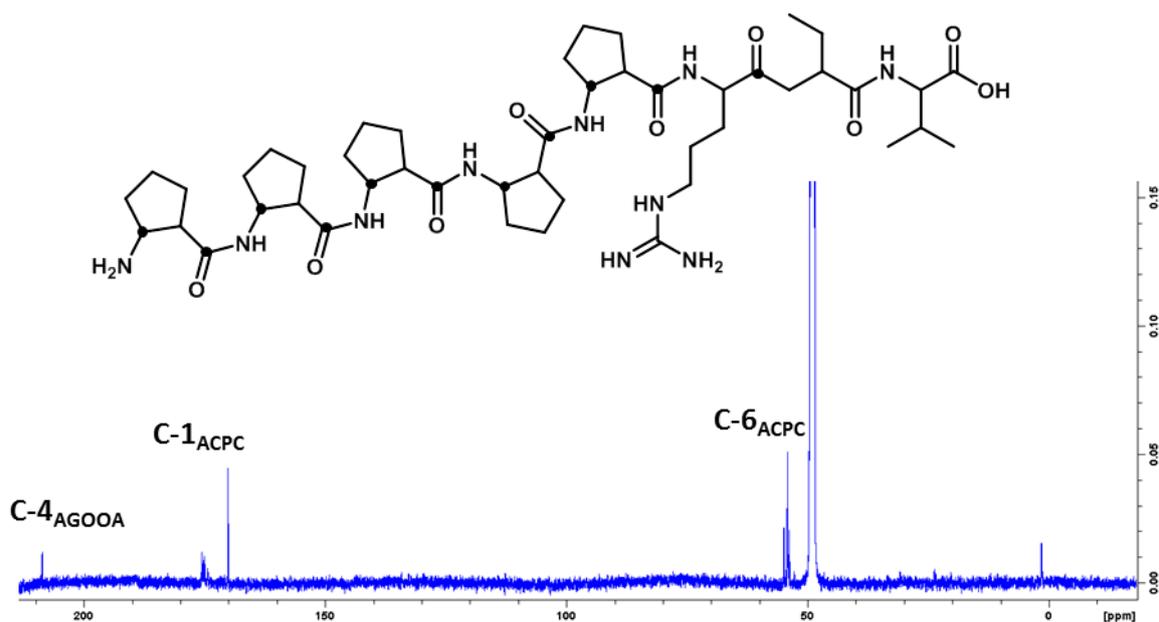


Figure 11:  $^{13}\text{C}$  NMR spectra of 100  $\mu\text{g}$  Myxopentacin 913(1) isolated from culture supplement 1- $^{13}\text{C}$  acetate.

Two  $^{13}\text{C}$  isotopes are incorporated in each ACPC moiety, one at the carboxylic acid C-1<sub>ACPC</sub> and one at the amine-carrying position C-6<sub>ACPC</sub>. These results are in agreement with previous feeding experiments, demonstrating the acetate incorporation into the ACPC moiety. Together with the knowledge about the lysine-derived nitrogen, these results imply that the ACPC is derived from an acetate based pathway, followed by a transamination reaction.

#### 4.3.1 AGOOA moiety

In the next step, the biosynthesis of the AGOOA moiety was investigated. Based on the AGOOA structure, it was hypothesized that the moiety has a polyketidic origin besides the arginine. However, feeding experiments using  $^{13}\text{C}$  labelled acetate could not confirm this hypothesis. In case of a polyketidic origin of the AGOOA moiety, a total number of five  $^{13}\text{C}$  atoms was expected. In a culture supplemented with  $^{13}\text{C}_2$  acetate MS<sup>2</sup> fragment 358.24  $m/z$  was observed (figure 10C). This fragment contains the AGOOA moiety and the C-terminal valine. Indeed, the isotope pattern showed only a small shift compared to a control, indicating a low degree of acetate incorporation. This finding is also in agreement with findings from the  $^{13}\text{C}$  NMR spectrum of 1- $^{13}\text{C}$  labelled myxopentacin 913 (1) (figure 11), where only the keto group carbon C-4<sub>AGOOA</sub> of the AGOOA moiety is clearly acetate derived. Arginine biosynthesis cannot explain this isotope incorporation pattern<sup>[7]</sup>. Interestingly, myxopentacin 913 (1) produced by a culture supplemented with  $^{13}\text{C}_6^{15}\text{N}_4$  arginine showed an isotope pattern shifted by 9 Da, while 10 Da would be expected for a completely incorporated arginine. These findings suggest a replacement of the arginine carboxyl group by the acetate carboxyl group. Since no further acetate incorporation into the AGOOA moiety was found, a different origin of the ethyl group was considered. Structural similarities suggested 2-oxobutyrate (6) as potential precursor. Cultures supplemented with D/L aminobutyric acid- $d_6$  showed an isotope pattern shifted by 5 Da compared to the control culture. By a deamination reaction, aminobutyric acid is converted to 2-oxobutyrate (6) reducing the number of deuterium atoms by one as observed in this experiment. Therefore, it can be concluded that the AGOOA moiety consists of arginine, one acetate unit and one 2-oxobutyrate molecule.

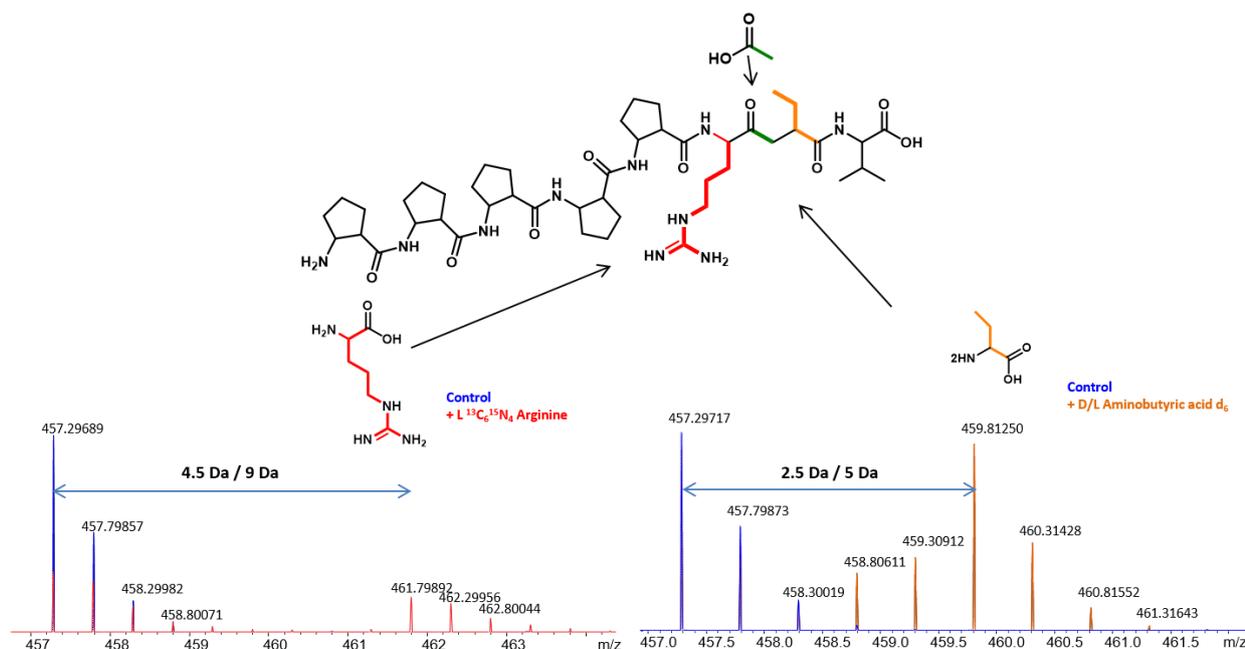


Figure 12: Isotope pattern of Myxopentacin 913 (**1**) showing double charged mass  $[M+2H]^{2+}$  extracted from LC-hrMS from the control culture (blue signals), from a culture supplemented with  $^{13}C_6$   $^{15}N_4$  arginine (red signals) and from a culture supplemented with aminobutyric acid  $d_6$ .

### 4.3.3 Identification and analysis of the myxopentacin gene cluster

In order to identify a candidate myxopentacin gene cluster, we attempted to search for structurally related molecules and their published biosynthesis. The ACPC moiety can be found in the nucleotide antibiotic amipurimycin<sup>[8]</sup> and in cispentacin<sup>[9]</sup>. In addition, the natural products arphamenines A and B<sup>[10][11]</sup> as well as the ketomemcins<sup>[12]</sup> contain a carbonylmethylene scaffold similar to the AGOAA moiety.

In the course of this work the biosynthesis of the nucleotide antibiotic amipurimycin was published.<sup>[13][14]</sup> The authors proposed a pathway for the biosynthesis of the ACPC moiety, a structural moiety that is shared by amipurimycin and the myxopentacins. Homologs of all proposed ACPC biosynthesis genes *apmA1-7* from one published amipurimycin pathway<sup>[13]</sup> were identified in the genome of the myxopentacin producer MCy 9003 and found clustered between 10.808.415 and 10.815.420 bp (figure 13A). Based on the identified homologs and the  $^{13}C$  incorporation pattern in the ACPC moiety of the myxopentacin molecules, which resembles the one published for amipurimycin, it is assumed that the biosynthesis of ACPC occurs in a similar manner in myxopentacin biosynthesis: Starting from  $\alpha$ -ketoglutarate (**7**), succinate semialdehyde (**8**) is generated by decarboxylation by a 2-oxoglutarate decarboxylase. Since no enzyme homologue was identified in close proximity, it is assumed that succinate semialdehyde is provided by primary metabolism like proclaimed for the amipurimycin biosynthesis. In the next step, MxpD, an acyl CoA ligase, forms a succinate semialdehyde CoA ester. In a Claisen condensation type reaction catalyzed by MxpF the CoA ester forms a C-6 chain (**9**) with ACP (MxpJ) bound malonate. MxpE a coronofac acid synthetase homologue is expected to catalyze the cyclization into 2-carboxy-3-hydroxycyclopentanone (**10**). MxpG, a  $\beta$ -hydroxyacyl-ACP dehydratase, reduces the Hydroxyl group into a double bond. The release from the ACP is catalyzed by MxpH a thioesterase. An aminotransferase (MxpI) transforms the ketone into an amine via a PLP dependent reaction. MxpI can be classified as acetylornithine aminotransferase based on sequence homology. The same observation was made in the amipurimycin biosynthesis<sup>[13,14]</sup>. An enoyl reductase that would be necessary to catalyze the last step to form the final ACPC was not identified like in the amipurimycin biosynthesis published by Kang et al.<sup>[13]</sup>. A second published amipurimycin biosynthesis by Romo et al.<sup>[14]</sup> also proposed an ACPC biosynthesis,

which is identical except one step. Here an enolyreductase (AmcA) was found to catalyze the last reduction step. Homologues of AmcA were not identified in close proximity of the other genes.

Regarding ketomemycin, the biosynthesis was published, including the BGC and a characterization of involved enzymes<sup>[15]</sup>. Homologs to the ketomemycin biosynthetic genes *ktmA-ktmC* were identified in the genome of MCy 9003 in close proximity to the *apmA* homologs between 10.815.062 and 10.817.779 bp.

Therefore, biosynthesis of the AGOOA moiety is expected to occur in a similar manner. In a first step, oxobutyrate (**6**) condenses with malonyl-CoA catalyzed by MxpK to form a 2-ethyl-2-hydroxysuccinic CoA ester (**13**). The formed CoA ester is reduced by MxpN a dehydratase into 2-ethylfumaric CoA. In the following reaction an second C-C bond is formed between the C<sub>α</sub> from arginine and the carbonyl group of 2-ethylfumaric CoA ester. The corresponding enzyme homologue of KtmB from Ketomemycin biosynthesis is MxpM. This enzyme contains a PLP binding site and can be classified as amino acid C<sub>α</sub>-acyltransferase, suggesting a PLP dependent Claisen type reaction like described in 2-Amino-3-Hydroxycyclopent-2-Enone biosynthesis<sup>[16]</sup>. During this reaction, the carbonyl group of arginine is lost as indicated by feeding experiments (figure 11 and 12). Genes encoding for an enzyme homologue encoded by *ktmF* to reduce the double bond could not be identified within this genomic region.

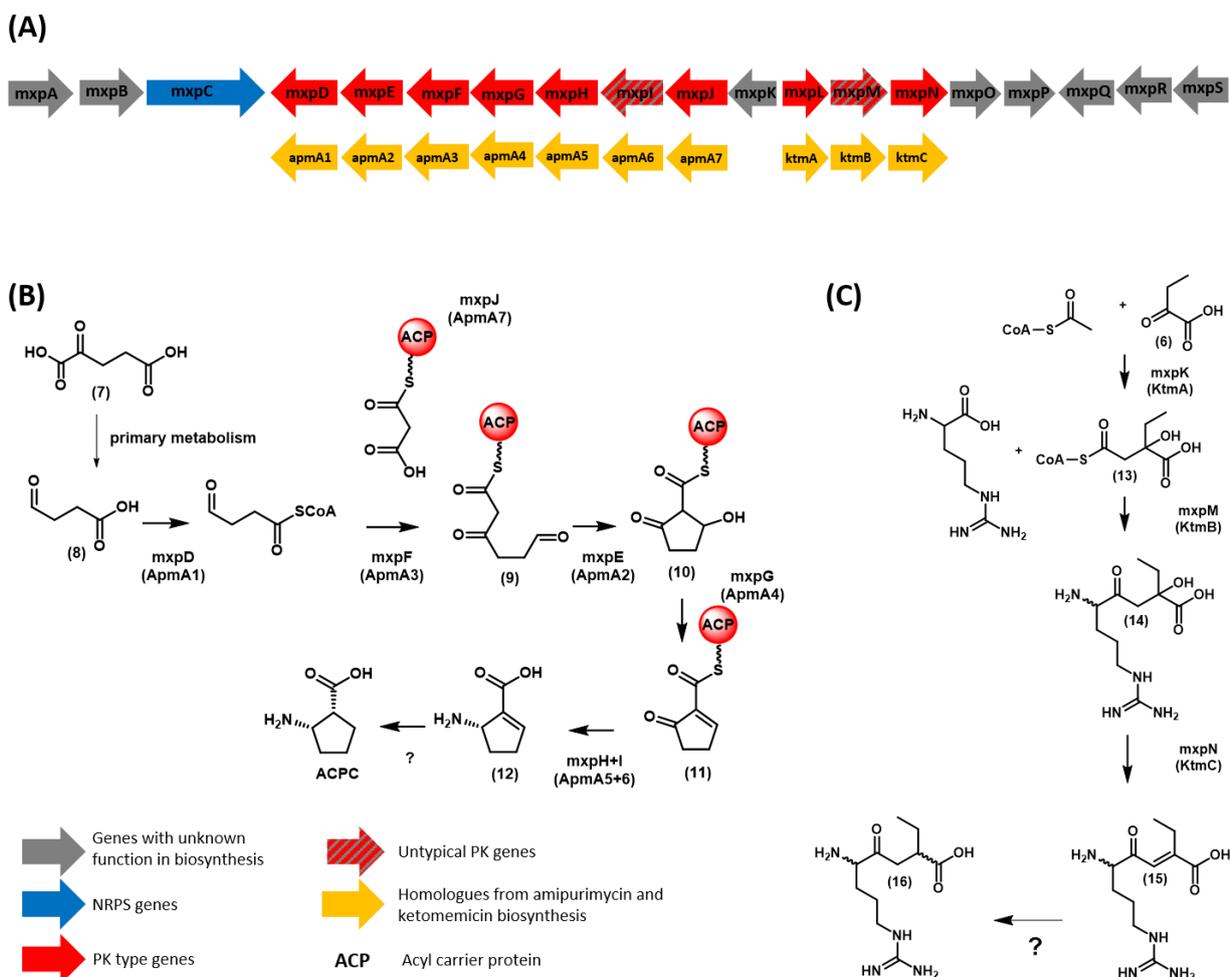


Figure 13: (A) Identified genes involved in myxopentacin biosynthesis with homologues from amipurimycin biosynthesis and ketomemycin biosynthesis. (B) Biosynthesis of ACPC with corresponding genes and homologues from amipurimycin biosynthesis (C) Biosynthesis of AGOOA moiety with corresponding genes and homologues from ketomemycin biosynthesis.

Biosynthetic genes for ACPC and AGOOA biosynthesis are oriented in different directions and appear to have their own promoter. Since biosynthetic genes are commonly clustered, genes in close proximity were examined for their potential role in the biosynthesis of the myxopentacins. Found 9.7 kbp upstream from ACPC biosynthetic genes, a promoter region is located in front of three putative genes. Therefore, these genes were investigated in closer detail. The same accounts for the six genes downstream of the AGOOA biosynthetic genes.

*Table 4: putative genes of the myxopentacin biosynthetic gene Cluster*

Gene	Size <sup>a</sup>	Protein homolog <sup>b</sup>	Proposed function	Catalytic domains
<i>mxpA (+)</i>	283	RYZ40139.1 (98/83)	endonuclease	
<i>mxpB (+)</i>	226	WP_120596930.1 (100/81)	unknown	
<i>mxpC (+)</i>	2579	WP_084609656.1 (99/79)	NRPS	C,A,PCP C,A,PCP C
<i>mxpD (-)</i>	469	WP_015348087.1 (100/81)	Acyl-CoA -ligase [Myxococcus stipitatus]	
<i>mxpE (-)</i>	194	WP_073559287.1 (96/76)	coronafacic acid synthetase	
<i>mxpF (-)</i>	374	WP_073559288.1 (98/83)	beta-ketoacylsynthase	
<i>mxpG (-)</i>	174	WP_120205330.1 (100/79)	coronafacic acid dehydratase/ $\beta$ -hydroxyacyl-ACP dehydratase	
<i>mxpH (-)</i>	148	WP_043392327.1 (89/77)	4-hydroxybenzoyl-CoA thioesterase	
<i>mxpI (-)</i>	465	WP_073559291.1 (100/83)	aminotransferase class III-fold pyridoxal phosphate-dependent enzyme	
<i>mxpJ (-)</i>	89	WP_043392332.1 (100/79)	acyl carrier protein	
<i>mxpK (-)</i>	153	AKJ06841.1 (98/83)	YbaK/prolyl-tRNA synthetase associated region	
<i>mxpK (+)</i>	282	WP_120625560.1 (100/77)	CoA ester lyase	
<i>mxpM (+)</i>	410	WP_002634949.1 (100/78)	class I/II-fold PLP -dependent enzyme/ amino acid C $\alpha$ -acyltransferases	
<i>mxpN (+)</i>	178	WP_073559295.1 (96/89)	MaoC family dehydratase	
<i>mxpO (+)</i>	529	WP_120546462.1 (100/89)	MBL fold metallo-hydrolase	
<i>mxpP (+)</i>	534	WP_120601942.1 (100/81)	MBL fold metallo-hydrolase/ $\alpha\beta$ -Hydroxylase	
<i>mxpQ (-)</i>	293	WP_120557531.1 (89/65)	CPBP family intramembrane metalloprotease	
<i>mxpR (-)</i>	272	WP_120557532.1 (100/67)	aspartyl/asparaginyl $\beta$ -hydroxylase domain-containing protein	
<i>mxpS (-)</i>	243	WP_120619789.1 (97/70)	FkbM family methyltransferase	

<sup>a</sup> Sizes are given in amino acids; <sup>b</sup> Accession numbers and percentage of identity/similarity are given in parentheses.

MxpC is a non-ribosomal peptide synthetase containing domains to perform the enzymatic setup for two NRPS-type elongation steps. The second adenylation domain is specific for Valine according to the NRPS predictor2 algorithm<sup>[17]</sup>. Since all discovered myxopentacin derivatives contain a C-terminal valine and the AGOOA moiety is bound to the valine via a peptide bond, it was hypothesized that MxpC partly assembles the myxopentacins in a NRPS-type fashion: The AGOOA moiety is recognized as an amino acid building block by the first A domain and C domain in MxpC, adenylated and attached to the PCP domain. The timing of the attachment of the characteristic ACPC residue cannot be concluded based on the existing results. In a further

elongation step, L-valine is condensed to the AGOOA. A TE domain was not identified inside MxpC. Instead, a third C-domain is located where a TE domain would be expected. Therefore it is assumed, that release from the assembly chain is conducted by the C-domain such as expected in the crocacin biosynthesis<sup>[18]</sup>.

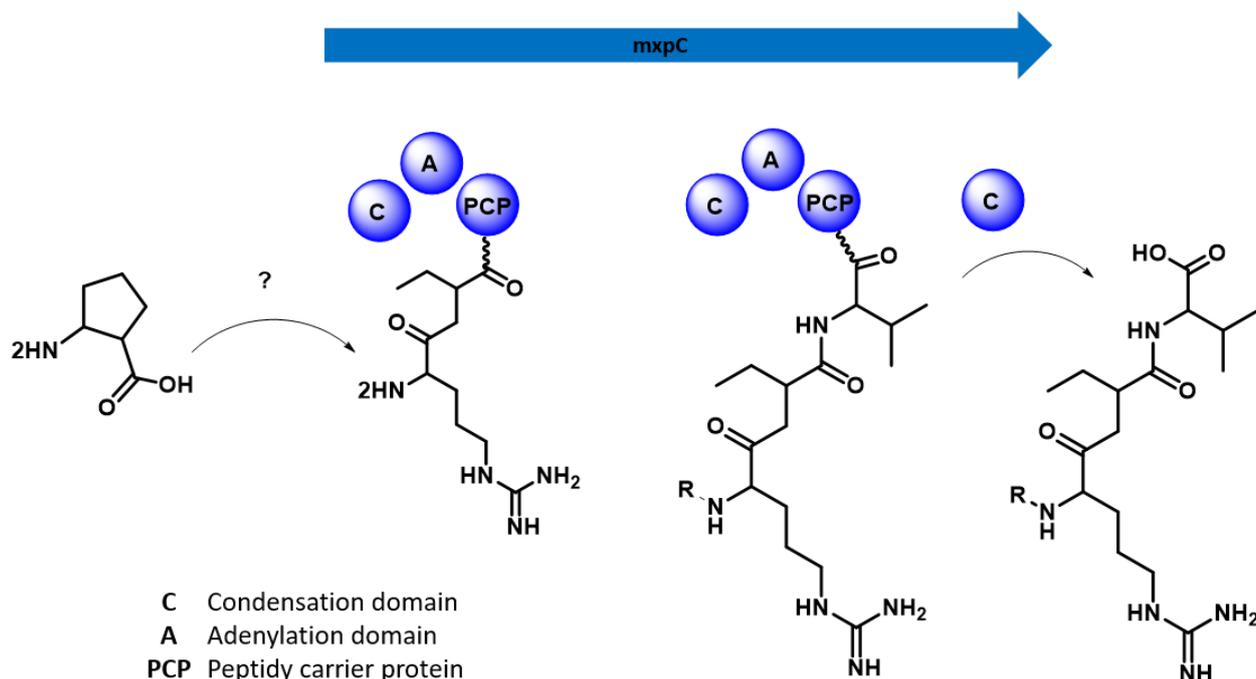


Figure 14: NRPS type step of myxopentacin biosynthesis conducted by MxpC

In order to verify the importance of MxpC in myxopentacin biosynthesis, the gene *mxpC* was inactivated by single cross over based homologous recombination, targeting the first A-domain. The inactivation mutant was created analogously to the procedure described for the activation mutants in chapter 3. The only difference here was the usage of a pCR2.1 vector with the sequence of the first A-domain of *mxpC* as homology target. Inactivation mutant was cultivated in parallel with the wildtype strain in triplicates and extracted. Crude organic extracts were measured by LC-*hr*MS and analyzed for the presence myxopentacin derivatives. No traces of the previously described myxopentacin derivatives that can clearly be differed from background noise were found in the mutant extracts. Therefore, it can be stated that MxpC is part of the myxopentacin assembly line.

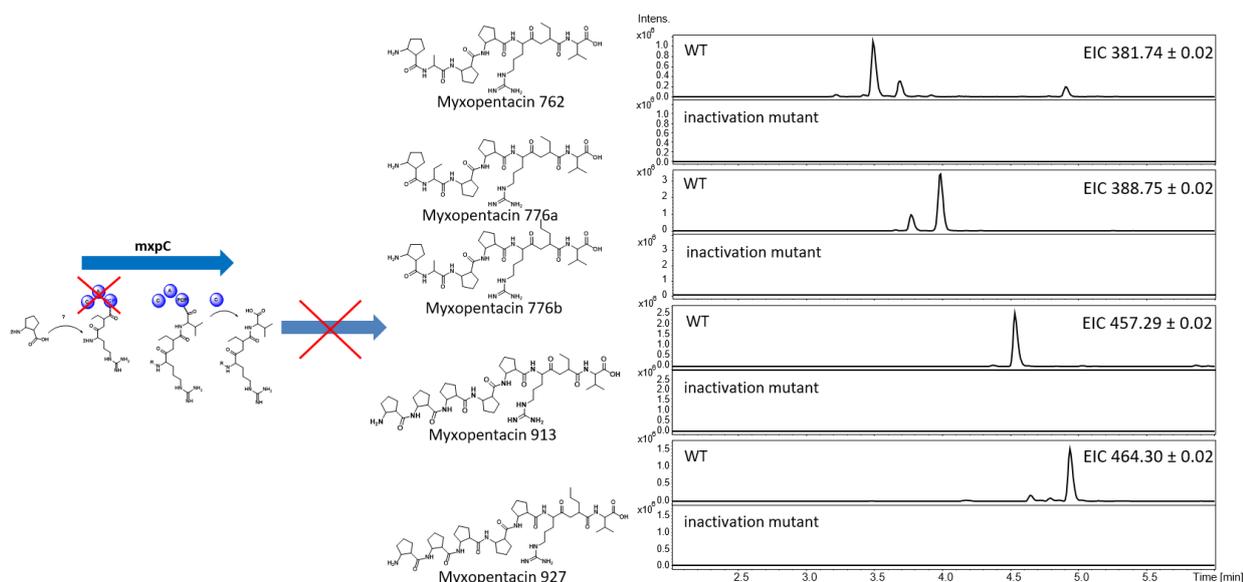


Figure 15: Result of inactivation of the myxopentacin biosynthetic gene *mxpC* by single cross over based homologous recombination disruption. Production of all five characterized myxopentacin variants shown as the Extracted Ion Chromatogram (EIC) of their respective double charged ions  $[M+2H]^{2+}$  is depleted compared to the WT.

#### 4.4 Identification of additional myxopentacin variants

The structural diversity within the discovered myxopentacins implies a high number of possible myxopentacin derivatives beside the seven additional potential ones that were listed in figure 1. In fact, the chromatogram of the crude extract of the inactivation mutant showed major differences compared to the WT extracts (figure 16A). To simplify the search for unknown myxopentacin derivatives, a principal component analysis was conducted using Metaboscape software in order to identify all peaks that disappeared in the inactivation mutant. The analysis yielded 1266 buckets that disappeared in the inactivation mutant (figure 16B). Considering a myxopentacin containing two ACPC moieties to be the smallest variant with a mass of 579 Da, the number of buckets is reduced to 957. Since every identified bucket is not necessarily a myxopentacin variant,  $MS^2$  fragmentation spectra were used as criteria to pick potential myxopentacin variants out of the bucket pool. Therefore, a single precursor list (SPL) was generated from the identified buckets and  $hrMS^2$  spectra were acquired. The spectra were analyzed using the GNPS clustering tool to arrange  $MS^2$  spectra based on their similarity<sup>[19]</sup>. To do so, the identical  $MS^2$  spectra and spectra with the same precursor ion were merged into one single spectrum, these merged spectra were compared with each other and a cosine score is generated, as criteria for the similarity. For visualization as a molecular network each merged spectra is displayed as a node and connected to other nodes via an edge if the cosine score is above a selected threshold value.<sup>[20]</sup> By selecting a threshold cosine score 0.2, two large Clusters were the result of the networking analysis (figure 16C).



Cluster 1 is the largest cluster consisting of 94 nodes. Five certain nodes inside this cluster represent each isolated myxopentacin derivative, suggesting that all other nodes represent so far unknown myxopentacin molecules. Cluster 2 is smaller and consists of 30 nodes. A manual investigation of MS<sup>2</sup> spectra of the corresponding precursor ions showed that the characteristic 223.14 *m/z* fragment was present in these spectra. In addition, fragments of higher masses showed mass differences of 111.069 Da, which is also characteristic for myxopentacin type molecules. The chosen cosine score of 0.2 as threshold value was low in comparison to the recommended value of 0.7<sup>[20]</sup>. However, using higher values resulted in the isolated myxopentacin molecules, which are obviously related but are not present in the same cluster. An even lower value would merge both clusters into one bigger Cluster. Assuming each node in both clusters represents a different myxopentacin molecule, the minimum number of different myxopentacin molecules detected would be 124.

## 4.5 Bioactivity

The isolated myxopentacin variants were tested against a broad array of fungal (*M. hiemalis*, *C. albicans*) and bacterial (*S. aureus*, *B. subtilis*, *E. coli* WT, *M. luteus*) test strains, but did not show any activity in concentrations up to 64 µg/ml. Tested against the KB3.1 cell line, the majority of myxopentacin derivatives showed a cytotoxic activity.

Table 5: Cytotoxic activity of myxopentacin variants

Cell line	Myxopentacin 762 [µg/ml]	Myxopentacin 776a [µg/ml]	Myxopentacin 776b [µg/ml]	Myxopentacin 913 [µg/ml]	Myxopentacin 927 [µg/ml]
IC <sub>50</sub> KB3.1	19.1	3.1	6.1	56.4	>111.1

## 4.6 Conclusions

Here, a novel compound class was discovered by investigation of unspecific results from BGC activation experiments. Five representatives of this compound class were isolated and the structures were determined using multidimensional NMR experiments. A unique feature of this compound class are the amino-cyclopentylcarboxylic acid (ACPC) rings and a carbonylmethylene moiety. The absolute configuration was partially determined. For the ACPC ring, an 1*R*,2*S*-configuration was determined, while the C-terminal valine, the aminobutyric acid and the alanine residue were all found to be L-configured.

Using various feeding experiments, the ACPC moiety was determined to be of an unusual polyketidic origin, including a transamination reaction with lysine as untypical nitrogen donor. In addition, the 5-amino-2-ethyl-8-guanidino-4-oxooctanoic acid (AGOOA) moiety was found to be derived from arginine, an acetate unit and oxobutyrate. By searching for homologues of genes from biosynthetic pathways of amipurimycin and ketomemycin published in 2019, which exhibit similar structural moieties, a candidate BGC was identified. Inactivation of one module within this BGC confirmed the link between the BGC and myxopentacin production. Moreover, with the inactivation mutant in hand, additional myxopentacin derivatives were identified. The isolated myxopentacin molecules did not show any antimicrobial activity against the test strains, but some myxopentacin derivatives exhibited moderate cytotoxic activity.

## 4.7 References

- [1] K. Ochi, T. Hosaka, *Appl. Microbiol. Biotechnol.* **2013**, *97*, 87.
- [2] S. Rachid, K. Gerth, I. Kochems, R. Müller, *Mol. Microbiol.* **2007**, *63*, 1783.
- [3] C. Volz, C. Kegler, R. Müller, *Chem. Biol.* **2012**, *19*, 1447.
- [4] R. Bhushan, H. Bruckner, *Amino Acids* **2004**, *27*, 231.
- [5] K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema, T. Weber, *Nucleic Acids Res.* **2019**, W81-W87.
- [6] A. Fabiani, A. Versari, G. P. Parpinello, M. Castellari, S. Galassi, *Journal of chromatographic science* **2002**, *40*, 14.
- [7] M. Caldara, G. Dupont, F. Leroy, A. Goldbeter, L. D. Vuyst, R. Cunin, *J. Biol. Chem.* **2008**, *283*, 6347.
- [8] T. Goto, Y. Toya, T. Ohgi, T. Kondo, *Tetrahedron Letters* **1982**, *23*, 1271.
- [9] M. Konishi, M. Nishio, K. Saitoh, T. Miyaki, T. Oki, H. Kawaguchi, *The Journal of antibiotics* **1989**, *42*, 1749.
- [10] S. Ohuchi, A. Okuyama, H. Naganawa, T. Aoyagi, H. Umezawa, *J Antibiot* **1984**, *37*, 518.
- [11] H. Umezawa, T. Aoyagi, S. Ohuchi, A. Okuyama, H. Suda, T. Takita, M. Hamada, T. Takeuchi, *J Antibiot* **1983**, *36*, 1572.
- [12] Y. Ogasawara, J. Kawata, M. Noike, Y. Satoh, K. Furihata, T. Dairi, *ACS chemical biology* **2016**, *11*, 1686.
- [13] W.-J. Kang, H.-X. Pan, S. Wang, B. Yu, H. Hua, G.-L. Tang, *Organic letters* **2019**.
- [14] A. J. Romo, T. Shiraishi, H. Ikeuchi, G.-M. Lin, Y. Geng, Y.-H. Lee, P. H. Liem, T. Ma, Y. Ogasawara, K. Shinya et al., *J. Am. Chem. Soc.* **2019**.
- [15] J. Kawata, T. Naoe, Y. Ogasawara, T. Dairi, *Angew. Chem. Int. Ed.* **2017**.
- [16] W. Zhang, M. L. Bolla, D. Kahne, C. T. Walsh, *J. Am. Chem. Soc.* **2010**, *132*, 6402.
- [17] M. Röttig, M. H. Medema, K. Blin, T. Weber, C. Rausch, O. Kohlbacher, *Nucleic Acids Res.* **2011**, *39*, W362-W367.
- [18] S. Müller, S. Rachid, T. Hoffmann, F. Surup, C. Volz, N. Zaburannyi, R. Müller, *Chem. Biol.* **2014**, *21*, 855.
- [19] M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapon, T. Luzzatto-Knaan et al., *Nat. Biotechnol.* **2016**, *34*, 828.
- [20] R. A. Quinn, L.-F. Nothias, O. Vining, M. Meehan, E. Esquenazi, P. C. Dorrestein, *Trends in Pharmacological Sciences* **2017**, *38*, 143.

## 4.8 Supplementary information

### 4.8.1 Experimental procedures

#### Strains

strains	suborder	family	genus	source
MCy 9003	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	Isolated by Dr. Ronald Garcia from a soil sample from Philippines
MCy 9003 mutant tn5 NRPS-t1pks 1	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant tn5 t1pks 8	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant tn5 NRPS 12	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant tn5 NRPS-t1pks 18	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant tn5 NRPS-t1pks 24	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant tn5 NRPS-t1pks 25	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant tn5 NRPS-t1pks 31	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant tn5 NRPS-t1pks 33	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant tn5 NRPS-t1pks 34	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant van NRPS-t1pks 35	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant inactivation NRPS-t1pks 1	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant inactivation NRPS 12	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant van NRPS 12	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar
MCy 9003 mutant inactivation Myxopentacin Cluster	<i>Cystobacterineae</i>	<i>Cystobacteraceae</i>	<i>Archangium</i>	created by Maja Remškar

#### Media

##### CyS-agar

Casitone	0.25 %
Yeast extract	0.1 %
Soluble starch	0.1 %
HEPES	10 mM
CaCl <sub>2</sub> · 2H <sub>2</sub> O	0.1 %
Agar	1.5 %

The ingredients were dissolved in MiliQ H<sub>2</sub>O. The pH was adjusted to 7.0 with KOH.

#### VY/2 agar

Baker's Yeast	0.5 %
MgSO <sub>4</sub> · 7H <sub>2</sub> O	0.1 %
CaCl <sub>2</sub> · 2H <sub>2</sub> O	0.5 %
HEPES	10 mM
Agar	1.5 %

The ingredients were dissolved in MiliQ H<sub>2</sub>O. The pH was adjusted to 7.2 with KOH.

#### VY/2-media

Baker's Yeast	0.5 %
MgSO <sub>4</sub> · 7H <sub>2</sub> O	0.1 %
CaCl <sub>2</sub> · 2H <sub>2</sub> O	0.5 %
HEPES	10 mM

The ingredients were dissolved in MiliQ H<sub>2</sub>O. The pH was adjusted to 7.2 with KOH.

#### RG5 media

Soy peptone	0.05 %
Soytone	0.05 %
Soy meal	0.2 %
Cornsteep solids	0.1 %
Yeast extract	0.05 %
Soluble starch	0.8 %
Baker's yeast	0.5 %
Gluten from wheat	0.2 %
MgSO <sub>4</sub> · 7H <sub>2</sub> O	0.1 %
CaCl <sub>2</sub> · 2H <sub>2</sub> O	0.1 %
HEPES	25 mM

The ingredients were dissolved in MiliQ H<sub>2</sub>O. The pH was adjusted to 7.2 with KOH.

#### CyS media

Casitone	0.25 %
Yeast extract	0.1 %
Soluble starch	0.1 %
HEPES	10 mM
CaCl <sub>2</sub> · 2H <sub>2</sub> O	0.1 %

The ingredients were dissolved in MiliQ H<sub>2</sub>O. The pH was adjusted to 7.2 with KOH.

#### <sup>13</sup>C-Glucose media

MgSO <sub>4</sub>	0.1 %
CaCl <sub>2</sub>	0.05 %
HEPES	10 mM
Yeast	0.25 %
L-Alanine	0.044%
L-Arginine	0.105 %
L-Valine	0.58 %
<sup>13</sup> C-Glucose	0.5 %

The ingredients were dissolved in MiliQ H<sub>2</sub>O. The pH was adjusted to 7.15 with KOH

## General analytic procedures

### Standard LC-MS analysis

## **amazon**

The amaZon is an Ion trap MS/MS mass spectrometer manufactured by Bruker Daltonics coupled with a Dionex HPLC from Thermo Scientific. This machine was used for standard analytics using following template gradient programs:

### **6 min gradient:**

The following reversed phase HPLC methods were used for the mass spectrometry coupled analysis: Column: ACQUITY BEH 50 x 2.1 mm, 1.7  $\mu\text{m}$ , 130  $\text{\AA}$ , a flow rate of 0.6 ml/ min and a column temperature of 45  $^{\circ}\text{C}$ . Following gradient was applied with  $\text{H}_2\text{O} + 0.1\%$  FA as eluent A and  $\text{ACN} + 0.1\%$  FA as eluent B. 0-1 min: 5 % Eluent B; 1-7 min: linear increase of eluent B from 5 to 95 %; 7-8.5 min: 95 % eluent B; 8.5-9 min: linear decrease of eluent B from 95 to 5 % B; 9-11.5 min: re-equilibration with 5 % eluent B. The detection was usually performed in the positive ESI MS mode, using a scan range from 200-2000 m/z.

### **9 min gradient:**

The following reversed phase HPLC methods were used for the mass spectrometry coupled analysis: Column: ACQUITY BEH 50 x 2.1 mm, 1.7  $\mu\text{m}$ , 130  $\text{\AA}$ , a flow rate of 0.6 ml/ min and a column temperature of 45  $^{\circ}\text{C}$ . Following gradient was applied with  $\text{H}_2\text{O} + 0.1\%$  FA as eluent A and  $\text{ACN} + 0.1\%$  FA as eluent B. 0-0.5 min: 5 % Eluent B; 0.5-9.5 min: linear increase of eluent B to 95 %; 9.5-10.5 min: 95 % eluent B; 10.5-11 min: linear decrease of eluent B; 11-12.5 min: re-equilibration with 5 % eluent B. The detection was usually performed in the positive ESI MS mode, using a scan range from 200-2000 m/z.

### **18 min gradient:**

The following reversed phase HPLC methods were used for the mass spectrometry coupled analysis: Column: ACQUITY BEH 50 x 2.1 mm, 1.7  $\mu\text{m}$ , 130  $\text{\AA}$ , a flow rate of 0.6 ml/ min and a column temperature of 45  $^{\circ}\text{C}$ . Following gradient was applied with  $\text{H}_2\text{O} + 0.1\%$  FA as eluent A and  $\text{ACN} + 0.1\%$  FA as eluent B. 0-0.5 min: 5 % Eluent B; 0.5-18.5 min: linear increase of eluent B to 95 %; 18.5-20.5 min: 95 % eluent B; 20.5-20.8 min: linear decrease of eluent B; 20.8-22.5 min: re-equilibration with 5 % eluent B. The detection was usually performed in the positive ESI MS mode, using a scan range from 200-2000 m/z.

## **maXis 4G**

The maXis 4G is a high resolution TOF mass spectrometer manufactured by Bruker Daltonics coupled with a Dionex HPLC from Thermo Scientific. This machine was used for high resolution LC-MS measurements. Following template method was used:

Column: ACQUITY BEH 100 x 2.1 mm 1.7  $\mu\text{m}$  130  $\text{\AA}$ , a flow rate of 0.6 ml/ min and a column temperature of 45  $^{\circ}\text{C}$ . Following gradient was applied with  $\text{H}_2\text{O} + 0.1\%$  FA as eluent A and  $\text{ACN} + 0.1\%$  FA as eluent B. 0-0.5 min: 5 % Eluent B; 0.5-18.5 min: linear increase of eluent B to 95 %; 18.5-20.5 min: 95 % eluent B; 20.5-20.8 min: linear decrease of eluent B; 20.8-22.5 min: re-equilibration with 5 % eluent B. The detection was usually performed in the positive MS mode or  $\text{MS}^2$  mode, using a scan range from 150-2500 m/z.

## **FTICR solariX**

The solariX is a very high resolution mass spectrometer. Measurements were carried out by direct infusion (DI) in positive ESI mode.

## **Spectroscopic methods**

NMR were acquired with a Bruker Ascend 700 NMR spectrometer equipped with a 5mm TCI cryoprobe and a Bruker UltraShield 500 NMR spectrometer equipped with a 5mm TCI cryoprobe. Optical rotation was measured using a Jasco polarimeter. CD spectra were acquired using a Jasco J-1500 CD spectrometer.

#### **Cultivation of MCy 9003 in liquid media**

MCy 9003 was cultivated in a flask at 30 °C with 180 rpm on a shaker. The cultures were either inoculated from a liquid pre-culture or from a piece of agar from an already grown agar culture by cutting out the agar piece under sterile conditions and dropping it into the flask with media. For extract preparation the absorber resin XAD-16 was added to the culture one day after inoculation. The cultures were incubated between 9 and 12 days.

#### **Cultivation of MCy 9003 on Agar**

The agar plate was inoculated with a piece from a grown agar plate or a drop of liquid media and cultivated at 30 °C. For Agar plates, VY/2-Agar or CyS-Agar was used.

#### **Cultivation of mutants of MCy 9003**

MCy 9003 mutants were cultivated in a baffled 300 ml flask at 30 °C with 180 rpm on a shaker in 100 ml CyS media with 75 µg/ml kanamycin. The cultures were either inoculated from a liquid pre-culture or from a piece of agar from an already grown agar culture by cutting out the agar piece under sterile conditions and dropping it into the flask with media. For extract preparation the absorber resin XAD-16 was added to the culture one day after inoculation. The cultures were incubated between 9 and 12 days.

#### **Genome sequencing**

Genomic DNA was isolated by a standard chloroform/phenol extraction protocol and was performed by Dr. Ronald Garcia and Irene Cochems. The genomic DNA was submitted to pacBio Inc. for sequencing. The raw sequencing data from PacBio sequencing were assembled from by Nestor Zaburanyi.

#### **Biological testing**

The testing of all crude extracts, Sephadex fractions, HPLC fractions and pure compounds were performed by Viktoria Schmitt, Irene Cochems and Stefanie Schmidt according to existing protocols.<sup>[1]</sup>

#### **Preparation of crude extract from MCy 9003**

XAD and cells were frozen at -20 °C and lyophilized. After lyophilization, metabolites were extracted with 50 ml MeOH/acetone 1:1 under constant shaking for 2.5 h. The supernatant was filtrated into a round flask and the solvent was evaporated under reduced pressure and the residue was dissolved in MeOH and transferred into a 4 ml Vial and stored at 20 °C.

#### **Isolation of Myxopentacin 913 and Myxopentacin 927**

Myxopentacin 913 and myxopentacin 927 were purified from a cultivation of WT MCy 9003. The strain was cultivated (20 x 100 ml) in RG5 medium in 300 ml flasks for 12 days. For harvesting, the cultures were poured through a 100 µm sieve in order to separate the XAD-16 and the cells were separated from the media. The filter cake was transferred into a flask and extracted four times with 100 ml acetone. In the next step, the acetone was filtrated into a round bottom flask using a funnel filled with glass wool. The solvent was evaporated under reduced pressure. In order to remove lipophilic compounds, the extract was dissolved in 100 ml MeOH with 10 % H<sub>2</sub>O and extracted twice with 100 ml n-Hexane. The target compounds were contained in the MeOH-layer, which was subsequently dried and dissolved in MeOH. The extract was fractionated using size exclusion chromatography with Sephadex LH-20 resin and MeOH as eluent. A flow rate of 30 drops/ min

was applied and 600 drops were collected for each fraction. An aliquot of every fraction was taken and measured using LC-MS to screen for the target masses. The target masses compounds, which were identified in fraction 7-10, were pooled, dried and dissolved in 1 ml MeOH. For further purification, HPLC with the following setup was used:

Dionex Ultimate 3000 coupled with Bruker High capacity trap mass spectrometer (HCT); Column: XSelect CSH (PFP) 250 x 10 mm, 4  $\mu$ m; flowrate 4 ml/min under HILIC conditions and column temperature 40 °C with H<sub>2</sub>O as eluent A, ACN as eluent B and 100 mM Ammonium formiate (AmFo) in dH<sub>2</sub>O as eluent C. The following gradient was applied: 0-1 min 97 % eluent B; 1-21 min linear decrease of eluent B to 47 %; 21-25 min eluent B at 47 %, 25-26 min linear increase of eluent B to 97 %, 26-30 min re-equilibration with 97 % eluent B. Eluent C was constantly at 3% over the whole gradient. The target masses were collected in one fraction from 13.1-13.7 min. The collected fraction was dried and dissolved in 1 ml MeOH. For final purification, HPLC with following setup was used:

Dionex Ultimate 3000 coupled with Bruker High capacity trap mass spectrometer (HCT); Column: Jupiter Proteo 250 x 10 mm, 4  $\mu$ m; flowrate 5 ml/min and column temperature 40 °C with H<sub>2</sub>O as eluent A, ACN as eluent B and 100 mM AmFo in dH<sub>2</sub>O as eluent C. The following gradient was applied: 0-1 min linear increase of eluent B from 5 % to 15 % and eluent constant at 3 %; 1-21 min linear increase of eluent B from 15 % to 25 % and eluent C constant at 3 %; 21-22 min linear increase of eluent B from 25 % to 95 % and eluent C constant at 3 %; 22-27 min 95 % eluent B and 3 % eluent C; 27-28 min linear decrease of eluent B from 95 % to 5 % and eluent C constant at 3 %; 28-31 min re-equilibration with 5 % eluent B and 3 % eluent C. Myxopentacin 913 was collected from 17.0-17.5 min and myxopentacin 927 was collected from 20.24-20.83 min. The collected fraction were dried and dissolved in MeOH.

#### **Isolation of Myxopentacin 762, Myxopentacin 776a and Myxopentacin 776b**

A 2.1 L culture of MCy 9003 mutant tn5 NRPS 12 in CyS media (21 x 100 ml) medium was centrifuged with 14,000 g for 15 min at 4 °C. The supernatant was removed and the cells and the XAD-16 were lyophilized for 6 h. After lyophilization, they were extracted 3 x with 400 ml Aceton/MeOH 1:1. The solvents were poured through a funnel with glass wool and the solvent was evaporated. The dried extract was re-dissolved in 200 ml Hexane and 300 ml MeOH and transferred into a separatory funnel. The layers were separated and the liquid/liquid extraction was repeated with 300 ml Hexane. The MeOH layer was dried and re-dissolved in 200 ml H<sub>2</sub>O and 200 ml ethyl acetate. The layers were separated in a separatory funnel and the aqueous layer was re-extracted with 200 ml ethyl acetate. The myxopentacin variants were present in the aqueous layer. The aqueous layer was dried and dissolved in 4 ml MeOH.

For further purification, size exclusion chromatography was performed with the aqueous fraction using Sephadex LH-20 resin and MeOH as eluent. A flow rate of 12 drops/ min was applied and 600 drops were collected for each fraction. An aliquot of every fraction was taken and measured using LC-MS to screen for the target masses. Fractions containing the target masses were evaporated to a volume of 1 ml. For further purification, HPLC with following setup was used:

Dionex Ultimate 3000 coupled with Bruker High capacity trap mass spectrometer (HCT); Column: Synergi polar RP 250 x 10 mm, 4  $\mu$ m; flowrate 5 ml/min and column temperature 40 °C with H<sub>2</sub>O + 0.1 % FA as eluent A and ACN + 0.1% FA as eluent B. The following gradient was applied: 0-2 min 5 % eluent B; 2-3 min linear increase of eluent B to 20 %; 3-33 min linear increase of eluent B to 30 %; 33-33 min linear increase of eluent B to 95 %; 34-37 min 95 % eluent B; 37-38 min linear decrease of eluent B to 5 %; 38-41 min re-equilibration with 5 % eluent B. Myxopentacin 762 was collected from 16.3-17.8 min; Myxopentacin 776a was collected from 20.1-22.0 min and Myxopentacin 776b was collected from 23.3-25.7 min. The collected fractions were dried and dissolved in 4 ml MeOH.

### **Marfey's Analysis**

Between 50 and 200 µg of the molecule were placed in a 2 ml glass vial. 100 µl 6N HCl were added and the sample was heated to 110 °C for 45 min under nitrogen atmosphere. In a next step, the sample was dried and dissolved in 110 µl H<sub>2</sub>O. 50 µl of the sample were transferred into two 1.5 ml plastic tubes. 20 µl 1N NaHCO<sub>3</sub> with pH 9 were added to each tube followed by addition of 1 % Marfey's reagent in acetone (D-FDLA and L-FDLA). The reaction mix was incubated for 2 h under constant shaking with 700 rpm. The reaction was stopped by addition of 10 µl 2 N HCl. After the reaction was stopped 300 µl ACN were added and the sample was centrifuged and an aliquot of the supernatant was measured by *hr*-LC-MS analysis using maXis4G analytical setup. Standards were handled in the similar manner skipping the hydrolysis with HCl. Following gradient was used: Column: ACQUITY BEH 100 x 2.1 mm 1.7 µm 130 A flowrate 0.6 ml/min and column temperature 45 °C with H<sub>2</sub>O + 0.1 % FA as eluent A and ACN + 0.1% FA as eluent B. % Eluent B; 0-1 min: linear increase of eluent B to 10 %; 1-15 min: linear increase of eluent B to 35 %; 15-22 min: linear increase of eluent B to 55 %; 15-22 min: linear increase of eluent B to 55 %; The detection was performed in positive MS mode.

### **Detection of Aminocyclopentyl-carboxylic acid in crude extract**

Aminocyclopentyl-acarboxylic acid (ACPC) was indirectly detected as Fmoc derivative. For Fmoc derivatization, 5 µl of ACPC standard (racemic mixture) or 20 µl of crude extract were mixed with 50 µl borate buffer (pH 7.9) and 100 µl of 3 mM Fmoc in acetone. The reaction mixture was centrifuged and the supernatant was analyzed by LC-MS.

### **Reverse feeding experiment**

The reverse feeding was conducted with WT MCy 9003 in 10 ml <sup>13</sup>C-glucose medium containing <sup>13</sup>C-labelled glucose as carbon source. The feeding culture was inoculated from a pre-culture also cultivated in <sup>13</sup>C-glucose medium. In order to observe incorporation of ACPC, 10 µl of a racemic mixture of 1R,2S and 1S,2R ACPC in a concentration of 0.25 M were added to the cultures two times per day over 4 days. Five days after inoculation, 0.5 ml XAD-16 suspension were added to the cultures. 10 days after inoculation, the cultures were harvested by centrifugation. The supernatant was discarded and the cells and the XAD were frozen and lyophilized before extraction as previously described. The dried extract was dissolved in 1 ml MeOH.

### **Feeding experiments**

The feeding study was performed with WT MCy 9003 in Cys medium in 10 ml cultures in 50 ml flasks. Following isotope labelled substrates were used:

Isotope labelled substrates	concentration	Solvent for application
L-valine <i>d</i> <sub>8</sub>	0.25 M	MeOH/dH <sub>2</sub> O 1:1
<sup>13</sup> C <sub>2</sub> -acetat	0.5 M	
1- <sup>13</sup> C-acetat	0.5 M	
2- <sup>13</sup> C-acetat	0.5 M	
L-arginine <sup>15</sup> N <sub>4</sub> <sup>13</sup> C <sub>6</sub>	0.25 M	
L-lysine <sup>15</sup> N <sub>2</sub> <sup>13</sup> C <sub>6</sub>	0.25 M	
L-methionine <sup>13</sup> C-methyl	0.25 M	
L/D aminobutyric acid <i>d</i> <sub>6</sub>	0.25 M	
L-asparagine <sup>15</sup> N <sub>2</sub> <sup>13</sup> C <sub>4</sub>	0.25 M	
L-glutamine <sup>15</sup> N <sub>2</sub> <sup>13</sup> C <sub>5</sub>	0.25 M	

To test substrate incorporation, 10 µl of the labelled substrate-solution was added to the cultures two times over two days. 24 h after last addition of L-valine-solution 400 µl of XAD-16 suspension were added. 9 days

after inoculation the cultures were harvested by centrifugation. The supernatant was discarded and the cells and the XAD were frozen and lyophilized before extraction as previously described. The dried extract was dissolved in 1 ml MeOH.

### **Preparation and purification of 1-<sup>13</sup>C-labelled Myxopentacin 913**

To produce 1-<sup>13</sup>C-acetate labelled myxopentacin 913, a culture (20 x 100 ml) of MCy 9003 mutant tn5 nrps12 in CyS media with 75 µg/ml kanamycin was started. For four days, 500 µl of a 0.5 M 1-<sup>13</sup>C acetate solution were added. Four days after inoculation, 2 ml of a XAD-16 suspension were added to the cultures. The culture was harvested by centrifugation. The supernatant was discarded and the cells and XAD were frozen at -20° C for storage. For extraction, the cells and XAD were melted and extracted three times with 400 ml MeOH. The solvent was dried and the extract was dissolved in MeOH. The extract was fractionated using size exclusion chromatography with Sephadex LH-20 resin and MeOH as eluent. A flow rate of 30 drops/ min was applied and 500 drops were collected for each fraction. An aliquot of every fraction was taken and measured using LC-MS to screen for the target masses. Fractions containing the target mass were pooled, dried and dissolved in 1 ml MeOH. For further purification HPLC with following setup was used:

Dionex Ultimate 3000 coupled with Bruker High capacity trap mass spectrometer (HCT); Column: Xselect CSH130 250 x 10 mm, 5 µm; flowrate 5 ml/min and column temperature 40 °C with H<sub>2</sub>O + 0.1 % FA as eluent A and ACN + 0.1% FA as eluent B. The following gradient was applied: 0-0.3 min linear increase of eluent B from 5 % to 13 %; 0.3-20.3 min linear increase of eluent B to 18 %; 20.3-21 min linear increase of eluent B to 95 %; 21-24 min 95 % eluent B; 24-25 min linear decrease of eluent B to 5 %; 25-29 min re-equilibration with 5 % eluent B. 100 µg Isotope labelled myxopentacin 913 was collected from 5.8-7.8 min. The collected fraction was dried and dissolved in MeOH.

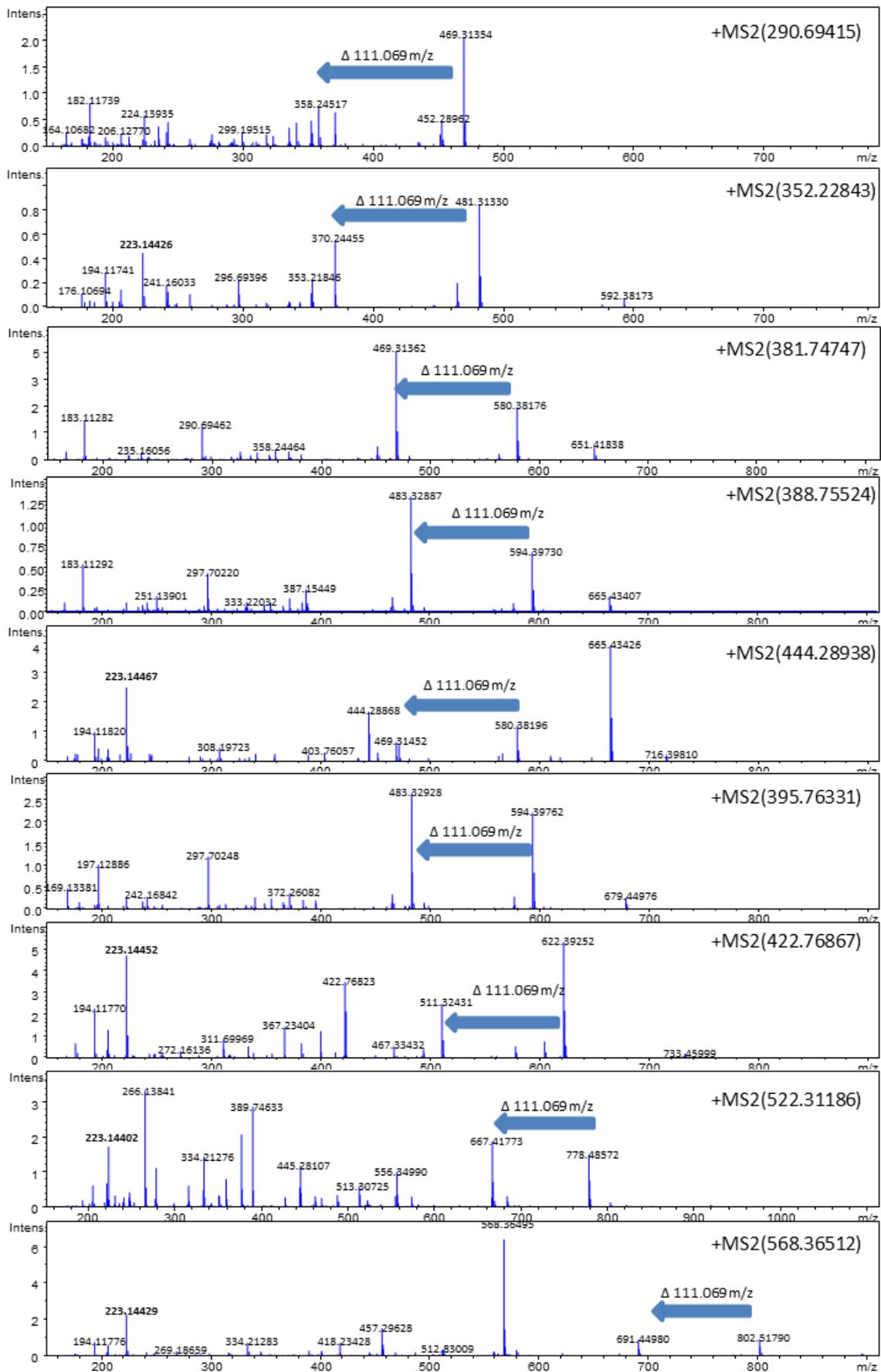
### **Identification of putative Myxopentacin variants**

In order to search for novel myxopentacin variants, a statistical approach was chosen. 3 cultures of WT MCy 9003 and the myxopentacin inactivation mutant were cultivated in CyS medium in 100ml scale. The cultures were extracted as described previously and measured by *hr* LC-MS. Using Metaboscape, a bucket list was created for each *hr* LC-MS dataset, followed by a principal component analysis to differentiate between buckets present in the WT and missing in the myxopentacin inactivation mutant. Using the buckets missing in the inactivation mutant, a scheduled precursor list (SPL) was created. Based on the SPL list a *hr* LC-MS<sup>2</sup> dataset was acquired and used for molecular networking using the GNPS client server. Following parameters were used:

The data were clustered with MS-Cluster with a parent mass tolerance of 1.0 Da and a MS<sup>2</sup> fragment ion tolerance of 0.1 Da to create consensus spectra. A network was then created, where edges were filtered to have a cosine score above 0.2 and more than 3 matched peaks. Further edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes.

#### **4.8.2 Supplementary information/ additional experimental data**

Observation of LC-*hr*MS chromatograms of extracts from mutants of MCy 9003 and WT draw the attention to a group of masses that showed increased peak areas in most mutants. The percentual change compared to the Peak in WT is displayed in supplementary table 1. The masses have similar MS<sup>2</sup> spectra displayed in figure 1 and supplementary figure 1. A list of the masses and their respective retention time from measurements on the maXis4G analytical system using an 18 minute gradient is given in supplementary table 2.



Supplementary figure 1: hrMS<sup>2</sup> spectra of identified masses that showed an increased area of more than 100 % compared to their WT counterpart. All spectra originate from the fragmentation of the double charged mass  $[M+2H]^{2+}$  ion and were extracted from LC-hrMS measurements of extracts of MCy 9003. The characteristic mass difference is marked with an arrow.

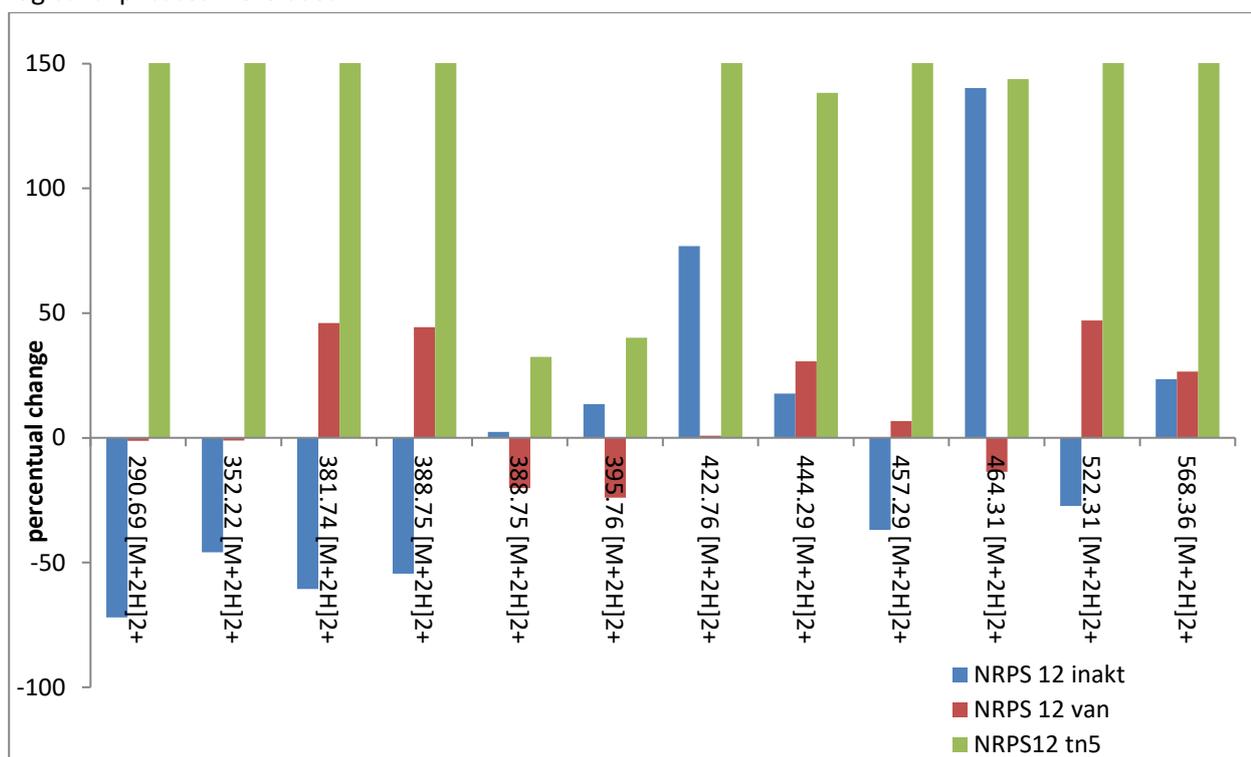
Supplementary table 1: Percentual change of peak areas of identified masses in extracts of the activation mutants compared to the wildtype extracts. Peak areas of Extracted Ion chromatograms (EICs)  $\pm$  0.1 from listed double charged ions were determined by Bruker Quant Analysis. For each sample type biological triplicate were measured and the average value was calculated. The average peak area of the corresponding wildtype extract was subtracted from the average peak area of the mutant extract. The difference was divided by the average peak area of corresponding wild type extract and multiplied by 100 to give the percentual change.

NRPS-t1pks 35	NRPS-tpks 34	NRPS-t1pks 33	NRPS-t1pks 31	NRPS 25	NRPS 24	NRPS-t1pks 18	NRPS 12	t1pks 8	NRPS 1	
153.6567486	-28.2187304	-8.664597515	125.0739173	-55.29688412	45.72694902	-5.206929486	452.1031201	50.43250845	71.56276337	290.69 [M+2H] <sup>2+</sup>
262.6822801	6.617513588	10.03884572	-59.23247583	-67.37226848	-42.40769801	45.807343	1299.92739	32.25601554	95.34987425	352.22 [M+2H] <sup>2+</sup>
641.3930992	-35.12978509	-37.11902949	-1.036212733	-56.63541922	-18.56068249	-83.15118422	727.7163506	-31.1221597	160.6674856	381.74 [M+2H] <sup>2+</sup>
361.3672723	-8.899842878	14.5052085	-26.49230258	-78.42178849	-64.97865142	-31.88575853	525.4748999	-40.78478607	73.56572231	388.75 [M+2H] <sup>2+</sup>
96.73050125	35.06176162	-40.20407149	-81.79165582	-78.34394119	-91.42628422	-58.30994084	32.45115468	-80.32283067	339.6280817	388.75 [M+2H] <sup>2+</sup>
27.50943227	38.03051973	-32.63807826	-77.1605057	-38.22878863	-94.61285897	10.79407466	40.16197266	-89.04773252	135.8668306	395.76 [M+2H] <sup>2+</sup>
140.5282039	23.07558423	-0.456579864	-47.4437168	-76.39093385	-55.45890904	-22.71755177	326.882231	-92.53571458	153.3536008	422.76 [M+2H] <sup>2+</sup>
23.88587384	3.67141176	0.568501319	-73.19198128	8.887554778	-72.29125682	-4.195197155	138.3388837	2.415190293	11.42445377	444.29 [M+2H] <sup>2+</sup>
146.222334	3.329389825	-8.261144417	-31.68480298	-60.72073865	-18.32429827	36.70556914	832.6238016	-15.40076414	133.4647953	457.29 [M+2H] <sup>2+</sup>
-32.98401602	37.56683057	2.6978828	-53.58682153	-55.16306629	-46.35139243	9.7258665034	143.858538	-87.69602224	41.19811436	464.31 [M+2H] <sup>2+</sup>
341.3139751	51.92820592	-56.42373688	-56.1639526	-53.65134974	-12.67172944	6.040583285	518.8807194	10.94308914	145.6087298	522.31 [M+2H] <sup>2+</sup>
37.10518712	35.67951979	-18.56484863	27.77287775	-36.55882884	-23.99907161	11.99777903	187.5742932	-55.62165285	24.70219252	568.36 [M+2H] <sup>2+</sup>

Supplementary table 2: Retention time, single charged and double charged mass of Myxopentacin variants identified by comparison of LC-hrMS data from activation mutant cultures and WT cultures

Rt [min]	[M+H] <sup>+</sup>	[M+2H] <sup>2+</sup>
2.96	580.38	290.69
3.88	703.45	352.22
3.57	762.487	381.75
3.84	776.50	388.75
4.05	776.49	388.75
4.28	790.51	395.76
4.46	844.54	422.77
4.22	887.57	444.29
4.62	913.57	457.29
5.04	927.61	464.31
6.51	1043.61	522.31
5.36	1135.72	568.36

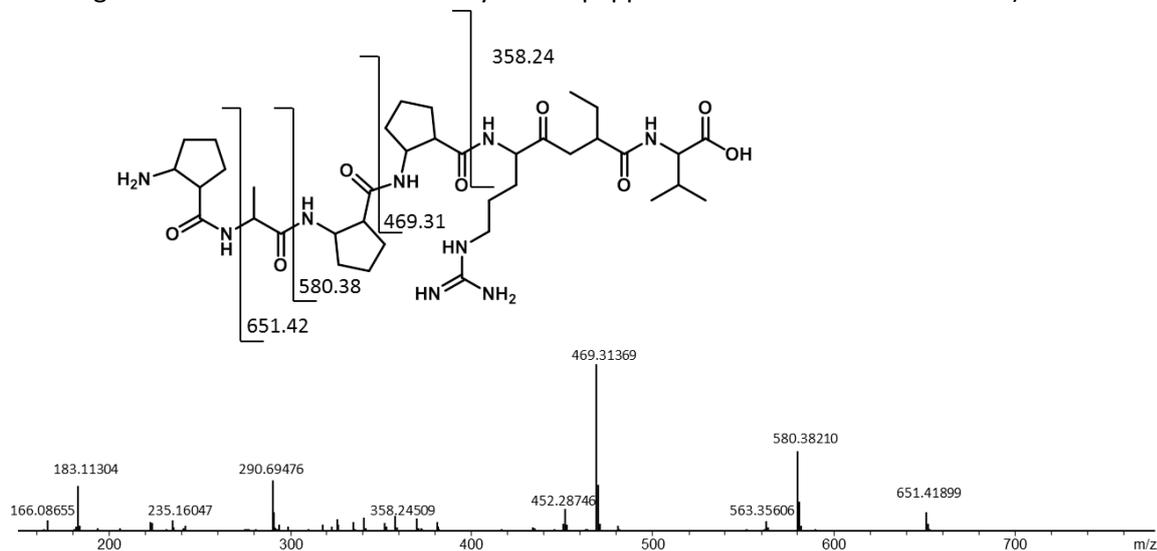
The influence of BGC NRPS 12 on production of the group of masses was investigated: Peak areas of the group of masses was determined in mutants having a vanilate promoter in front of Cluster 12 and mutants where Cluster 12 is inactivated. For mutants containing the vanilate promoter percentual changes between cultures treated with 1 mM vanilate and untreated cultures was determined. For the inactivation mutant the percentual change of the inactivation mutant to the WT mutant was determined. For the investigation biological triplicates were used.



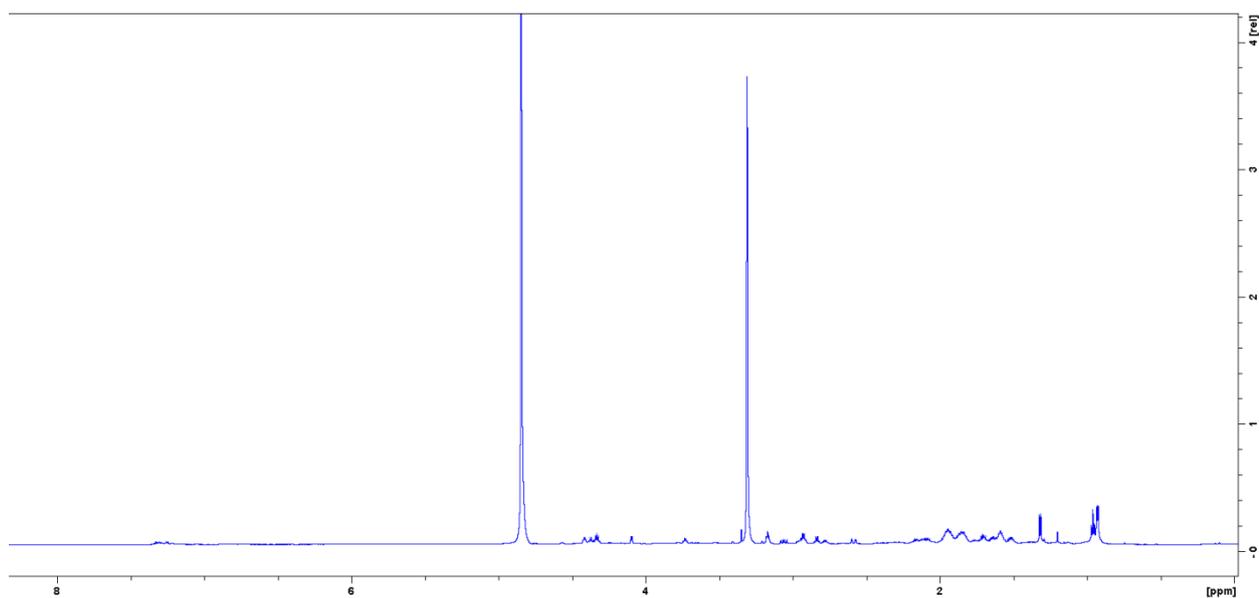
Supplementary figure 2: Percentual change of peak areas of identified masses showing the change in mutant tn5 NRPS 12 compared to the WT (green bars), the change in mutant van NRPS 12 between cultures treated with vanilate and untreated cultures (red bars) and the change between the inactivation mutant of NRPS 12 and WT cultures.

### Analytical data Myxopentacin 762

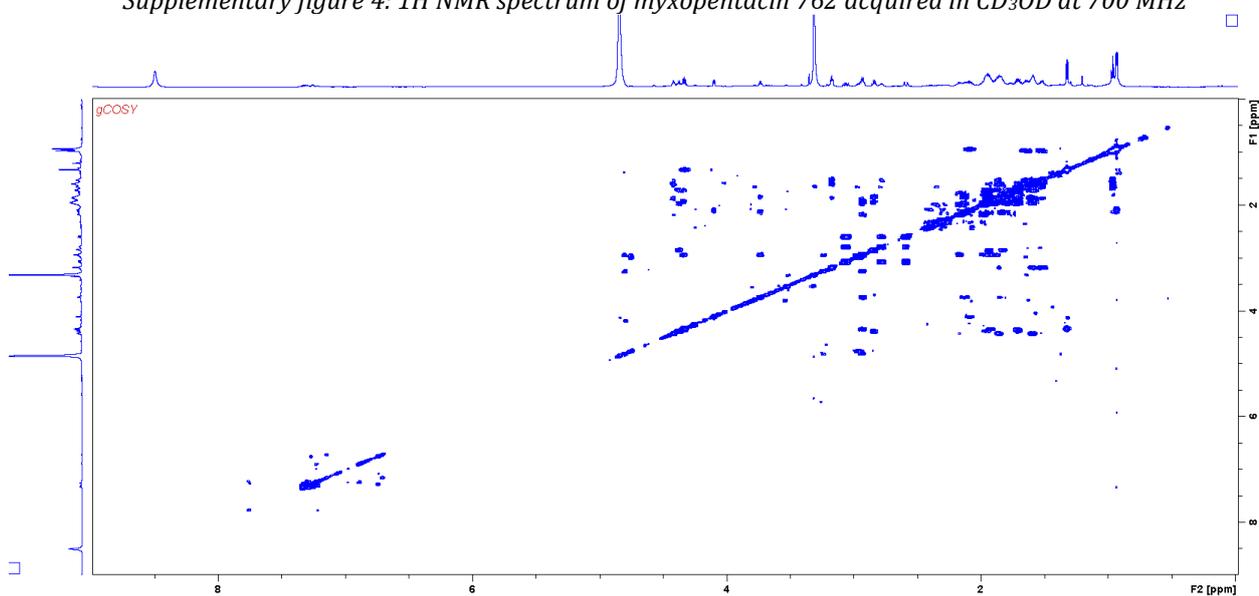
Myxopentacin 762 was isolated as white solid. *hr*-Masses: double charged ion  $[M+2H]^{2+}$ : 381.74726 m/z; single charged ion  $[M+H]^+$ : 762.49203 m/z. UV max < 200 nm (UV data extracted from LC-MS-UV measurement using a Dionex ultimate 3000 HPLC system equipped with a DAD UV-VIS detector)



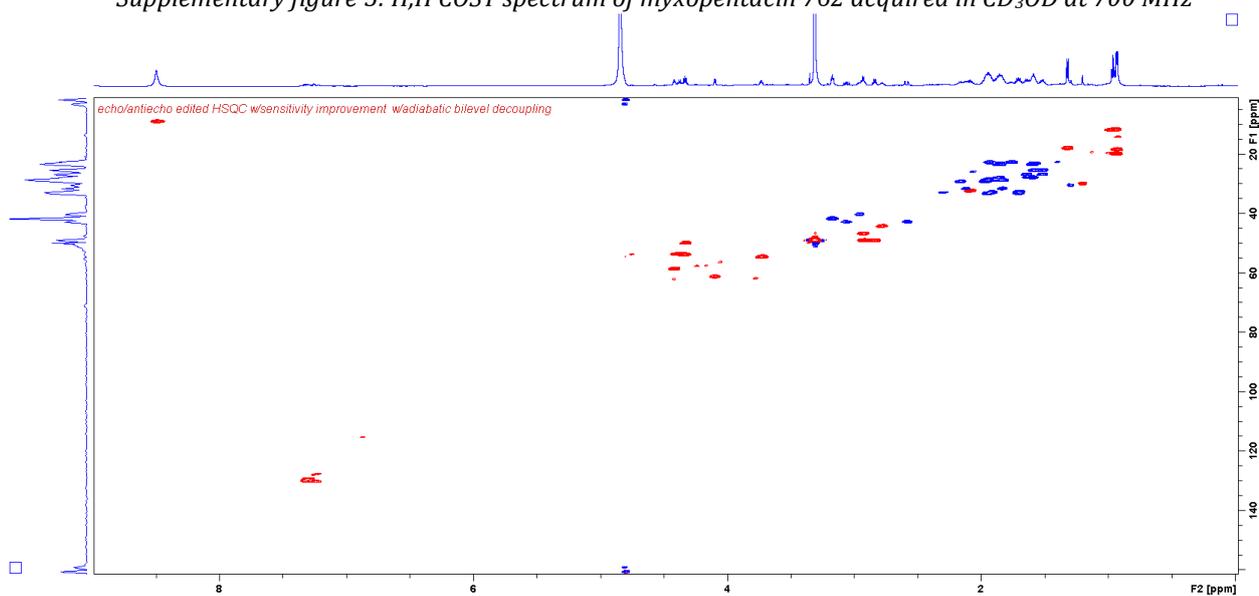
Supplementary figure 3: HR MS<sup>2</sup> spectrum and observed fragments of myxopentacin 762. MS<sup>2</sup> spectrum is extracted from LC-*hr*MS measurement of an extract of MCy 9003.



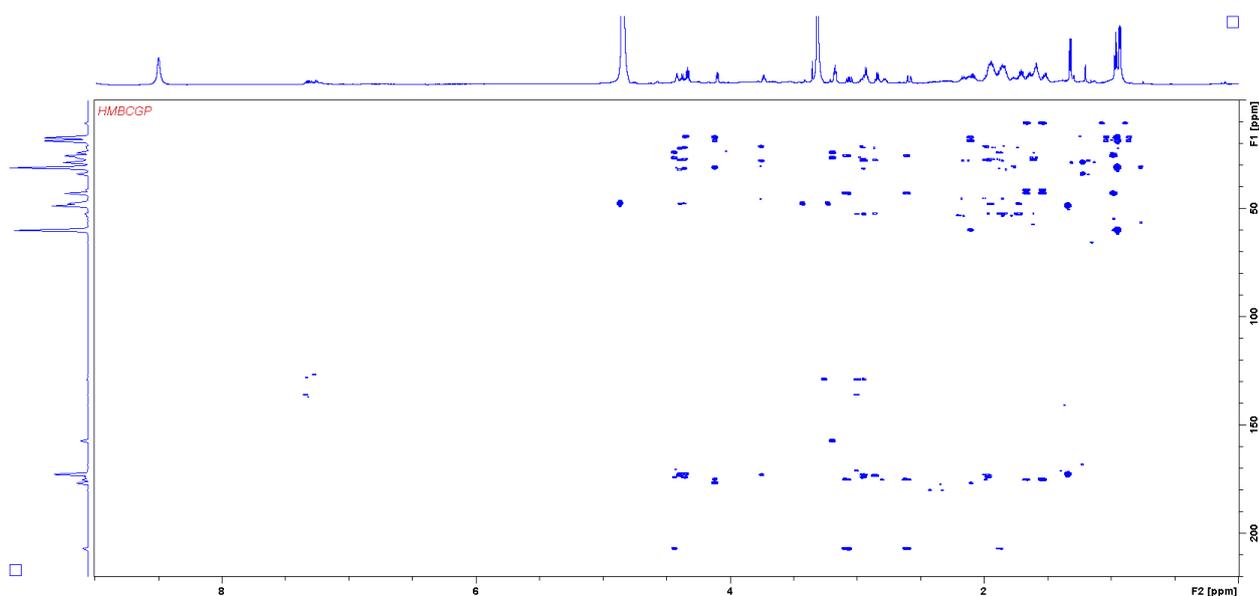
Supplementary figure 4:  $^1\text{H}$  NMR spectrum of myxopentacin 762 acquired in  $\text{CD}_3\text{OD}$  at 700 MHz



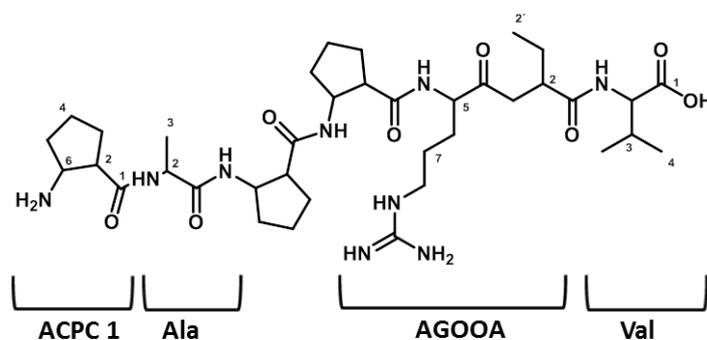
Supplementary figure 5:  $\text{H,H}$  COSY spectrum of myxopentacin 762 acquired in  $\text{CD}_3\text{OD}$  at 700 MHz



Supplementary figure 6: HSQC spectrum of myxopentacin 762 acquired in  $\text{CD}_3\text{OD}$  at 700/175 MHz



Supplementary figure 7: HMBC spectrum of myxopentacin 762 acquired in CD<sub>3</sub>OD at 700/175 MHz



Supplementary figure 8: Structure of Myxopentacin 762 with assigned residues.

Supplementary table 3: Spectroscopic data of Myxopentacin 762 in CD<sub>3</sub>OD.

moiety	position	$\delta^{13}\text{C}^{\text{a}}$	$\delta^1\text{H}^{\text{b}}$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
ACPC 1	1	174.3			
	2	46.7	2.93 m	3,6	1,3,4,5,6
	3a	29.1	1.97 m		
	3b	29.1	2.17 m		
	4a	22.5	1.93 m		
	4b	22.5	1.76 m		
	5a	31.5	2.12 m		
	5b	31.5	1.85 m		
	6	54.6	3.72 m		2,5
Ala	1	174.0			
	2	49.0	4.32 m	3	1,3
	3	18.0	1.32 d (7.2)	2	1,2
ACPC 2	1	174.6			
	2	49.1	2.84 m	3,6	1,3,4,5,6
	3a	28.7	1.93 m		
	3b	28.7	1.83 m		

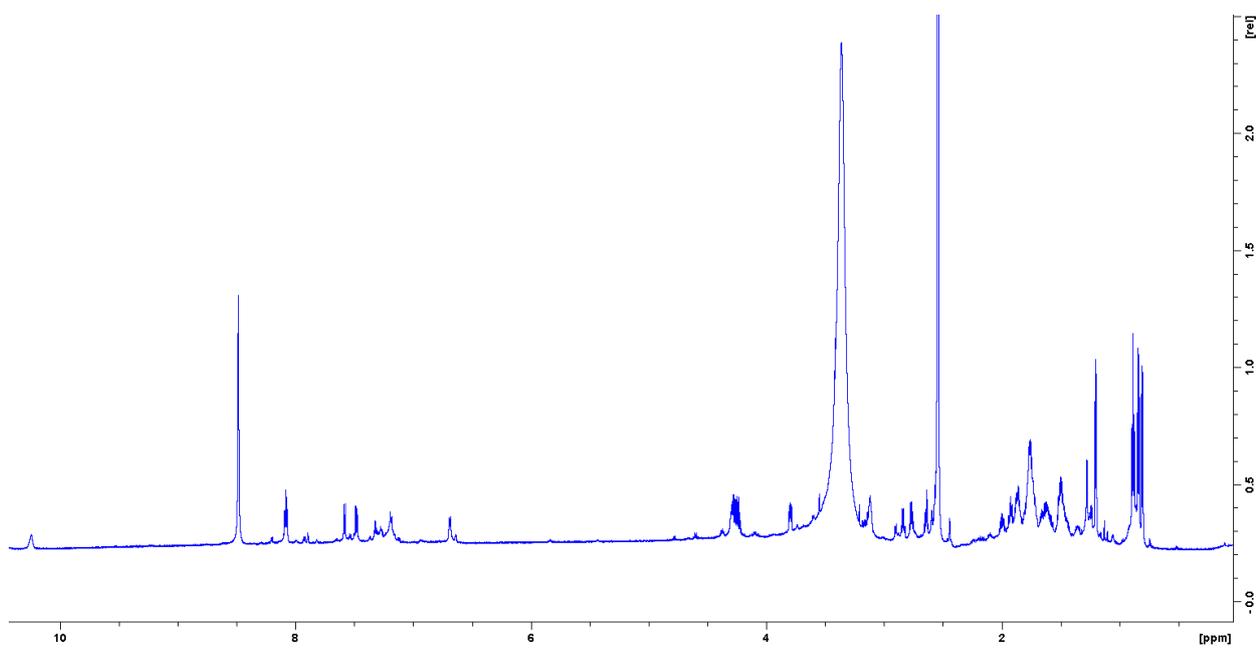
	4a	23.3	1.85 m		
	4b	23.3	1.59 m		
	5a	33.1	1.96 m		
	5b	33.1	1.67 m		
	6	53.4	4.38 m	2,5	1,2,3,4,5,1 <sub>Ala</sub>
<b>ACPC 3</b>	1	175.3			
	2	48.9	2.92 m	3,6	1,3,4,5,6
	3a	28.6	1.95 m		
	3b	28.6	1.86 m		
	4a	23.2	1.85 m		
	4b	23.2	1.59 m		
	5a	32.9	1.92 m	5	
	5b	32.9	1.70 m	5	
	6	53.6	4.34 m	2,5	1,2,3,4,5
<b>AGOOA</b>	1	176.4			
	2	44.3	2.78 m	3,1'	1,3,1'
	3a	42.8	3.06 d,d (10.9,18.2)	2,3b	2,4,1'
	3b	42.8	2.59 d,d (3.95,18.3)	2,3a	2,4,1'
	4	208.3			
	5	58.6	4.42 m	6	4,6,7,1 <sub>ACPC 3</sub>
	6a	27.8	1.86 m	5,6b,7	5,6
	6b	27.8	1.59 m	5,6a,7	5,6
	7a	25.4	1.51 m	6,7b,8	6
	7b	25.4	1.57 m	6,7a,8	6
	8	41.6	3.17 t (6.5)	7	6,7,9
	9	158.7			
	1'a	26.8	1.64 m	2,1'b,2'	2,3,2'
1'b	26.8	1.52 m	2,1'a,2'	2,3,2'	
2'	11.8	0.96 t (7.4)	1'	2,1'	
<b>Val</b>	1	178.0			
	2	61.1	4.09 d (6.0)	3	1,3,4,5,1 <sub>AGOOA</sub>
	3	32.3	2.08 m	2	1,2,4,5
	4	18.5	0.93 d (6.8)	3	2,3,5
	5	19.9	0.93 d (6.8)	3	2,3,4

<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

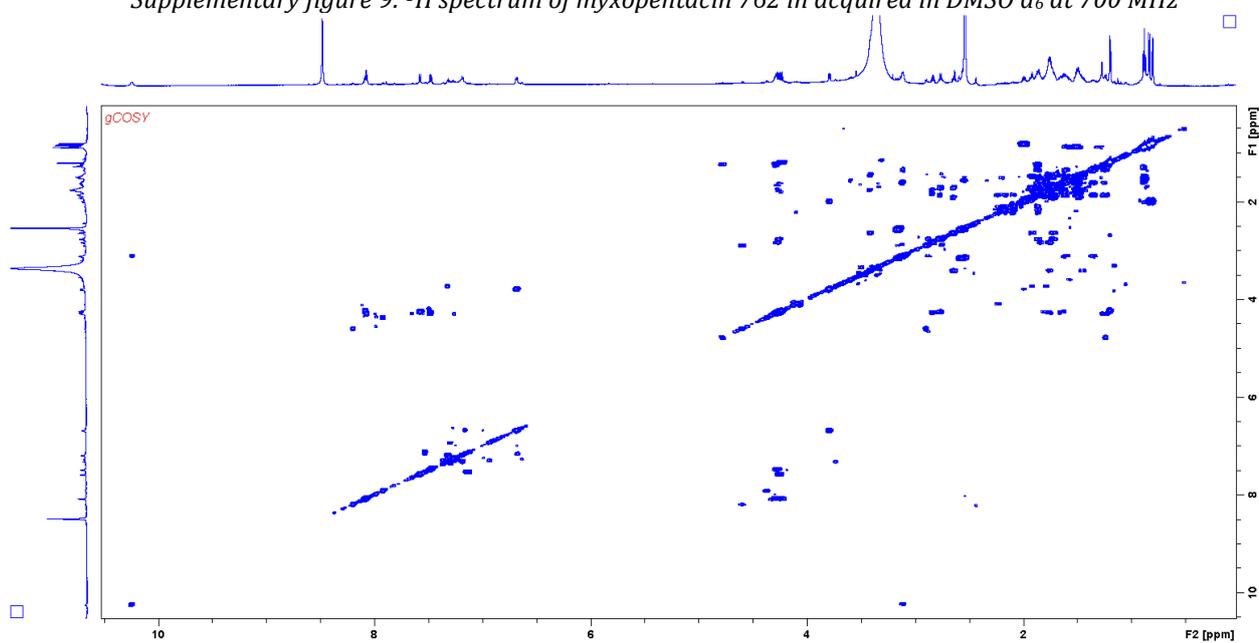
<sup>b</sup> acquired at 700 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

<sup>c</sup> proton showing COSY correlation to indicated proton.

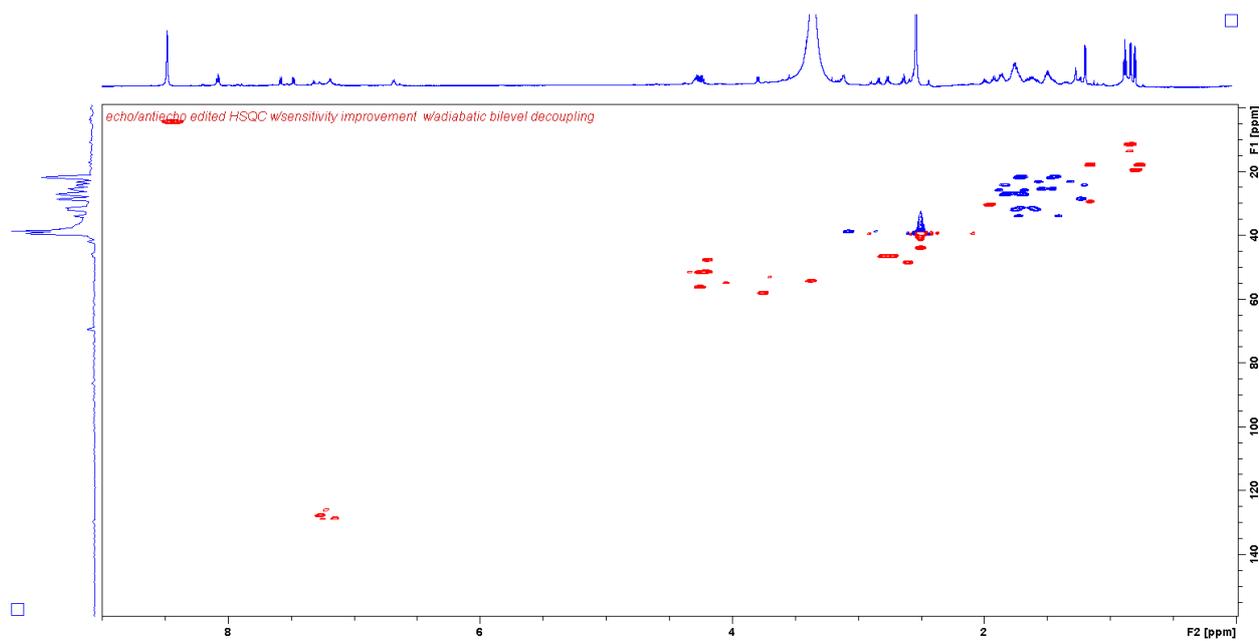
<sup>d</sup> proton showing HMBC correlations to indicated carbons.



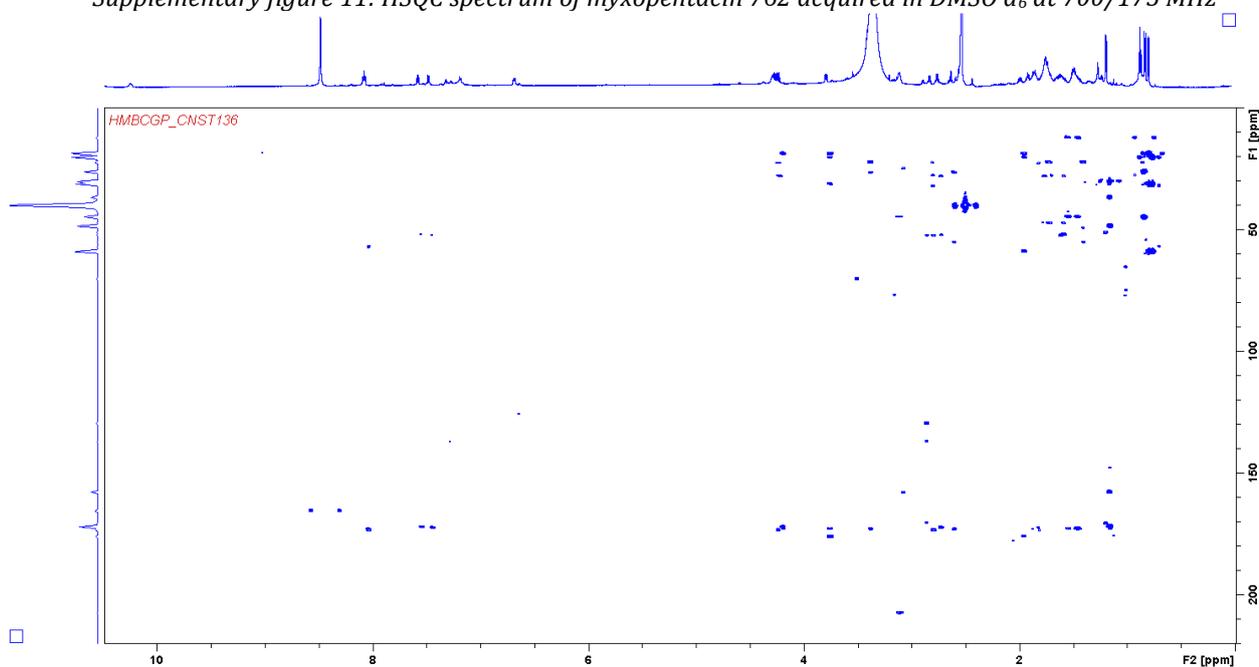
Supplementary figure 9:  $^1\text{H}$  spectrum of myxopentacin 762 in acquired in  $\text{DMSO } d_6$  at 700 MHz



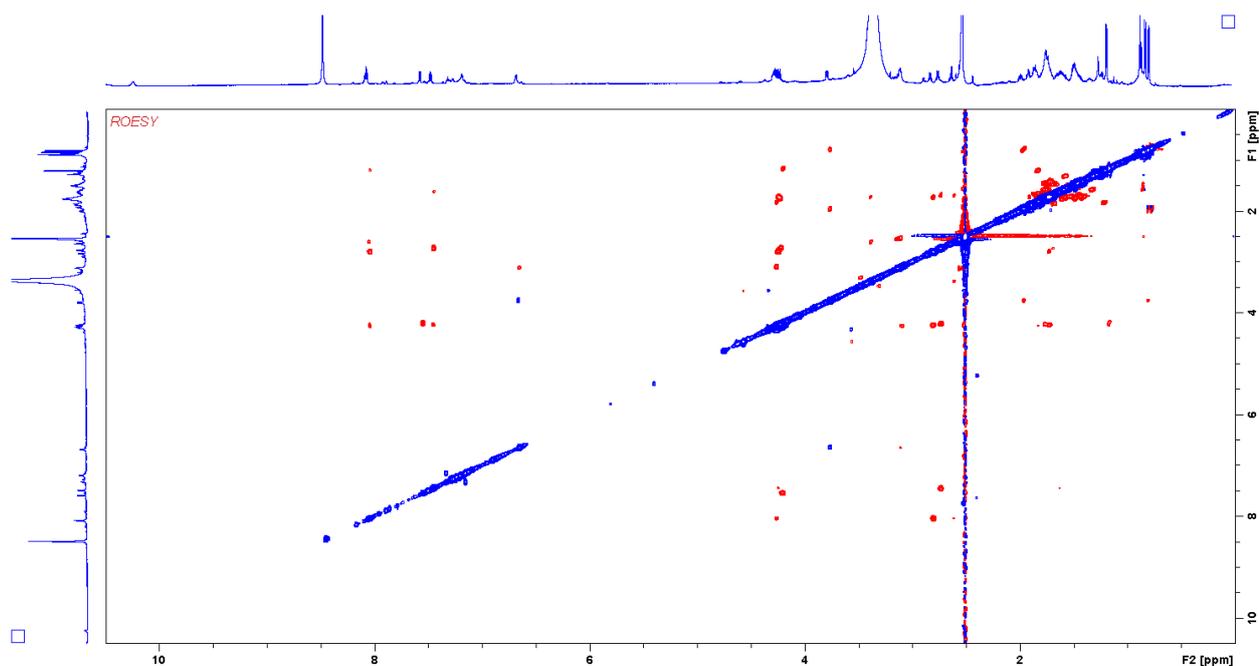
Supplementary figure 10:  $\text{H,H}$  COSY spectrum of myxopentacin 762 acquired in  $\text{DMSO } d_6$  at 700 MHz



Supplementary figure 11: HSQC spectrum of myxopentacin 762 acquired in DMSO  $d_6$  at 700/175 MHz



Supplementary figure 12: HMBC spectrum of myxopentacin 762 acquired in DMSO  $d_6$  at 700/175 MHz



Supplementary figure 13: ROESY spectrum of myxopentacin 762 acquired in DMSO  $d_6$  at 700 MHz

Supplementary table 4: Spectroscopic data of Myxopentacin 762 in DMSO  $d_6$ .

moiety	position	$\delta^{13}\text{C}^a$	$\delta^1\text{H}^b$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
ACPC 1	1	172.2			
	2	48.6	2.60 m	3,6	1,3,4,5,6
	3a	25.7	1.68 m		
	3b	25.7	1.88 m		
	4a	21.4	1.44 m		
	4b	21.4	1.72 m		
	5a	34.0	1.42 m		
	5b	34.0	1.72 m		
	6	54.3	3.37 m	2,5	1,2,3,4,5
	6-NH		n.d.		
Ala	1	171.2			
	2	47.7	4.19 m	3,2-NH	1,3
	2-NH		8.04 d (7.9)	2	2,3
	3	18.0	1.16 d (7.1)	2	1,2
ACPC 2	1	171.6			
	2	46.5	2.73 d,t (7.7,8.0)	3,6	1,3,4,5,6
	3a	27.3	1.68 m		
	3b	27.3	1.83 m		
	4a	21.8	1.45 m		
	4b	21.8	1.73 m		
	5a	32.0	1.57 m		
	5b	32.0	1.73 m		
	6	51.4	4.22 m	2,5,6-NH	2,3,4,5
	6-NH		7.54 d (8.2)	6	6,1 <sub>Ala</sub>
ACPC 3	1	172.6			
	2	46.6	2.79 d,t (7.3,7.5)	3,6	1,3,4,5,6
	3a	27.0	1.72 m	2	
	3b	27.0	1.81 m	2	

	4a	21.7	1.45 m		
	4b	21.7	1.72 m		
	5a	31.4	1.70 m		
	5b	31.4	1.62 m		
	6	51.5	4.25 m	2,5,6-NH	2,3,4,5,1 <sub>ACPC 2</sub>
	6-NH		7.44 d (8.2)	6	6, 1 <sub>ACPC 2</sub>
	1	172.1			
	2	44.1	2.50 <sup>e</sup>	3b,1'	1,3
	3	41.9	2.55 <sup>e</sup>		2,4
		41.9	3.14 m	3a,1'	2,4
	4	206.8			
	5	56.3	4.25 m	6,5-NH	7, 1 <sub>ACPC 3</sub>
	5-NH		8.03 d (7.5)	5	5, 1 <sub>ACPC 3</sub>
<b>AGOOA</b>	6a	24.3	1.83 m	5,6b,7	
	6b	24.3	1.29 m	5,6a,7	
	7a	23.3	1.56 m	6,7b,8	
	7b	23.3	1.32 m	6,7a,8	
	8	38.9	3.08 m	7	6,7,9
	8-NH		10.21 s		
	9	156.8			
	1'a	25.4	1.58 m	1'b,2'	2,2'
1'b	25.4	1.48 m	1'a,2'	2,2'	
	2'	11.6	0.84 t (7.5)	1'	2,1'
	1	175.2			
<b>Val</b>	2	58.3	3.76 d,d (8.2,5.0)	2-NH,3	1,3,4,5,1 <sub>AGOOA</sub>
	2-NH		6.65 d (6.3)	2	
	3	30.3	1.96	2,4,5	1,2,4,5
	4	19.5	0.79 d (6.8)	3	2,3,5
	5	18.1	0.76 d (6.8)	3	2,3,4

<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal DMSO d<sub>6</sub> at  $\delta$  39.52 ppm.

<sup>b</sup> acquired at 700 MHz, referenced to solvent signal DMSO d<sub>6</sub> at  $\delta$  2.54 ppm.

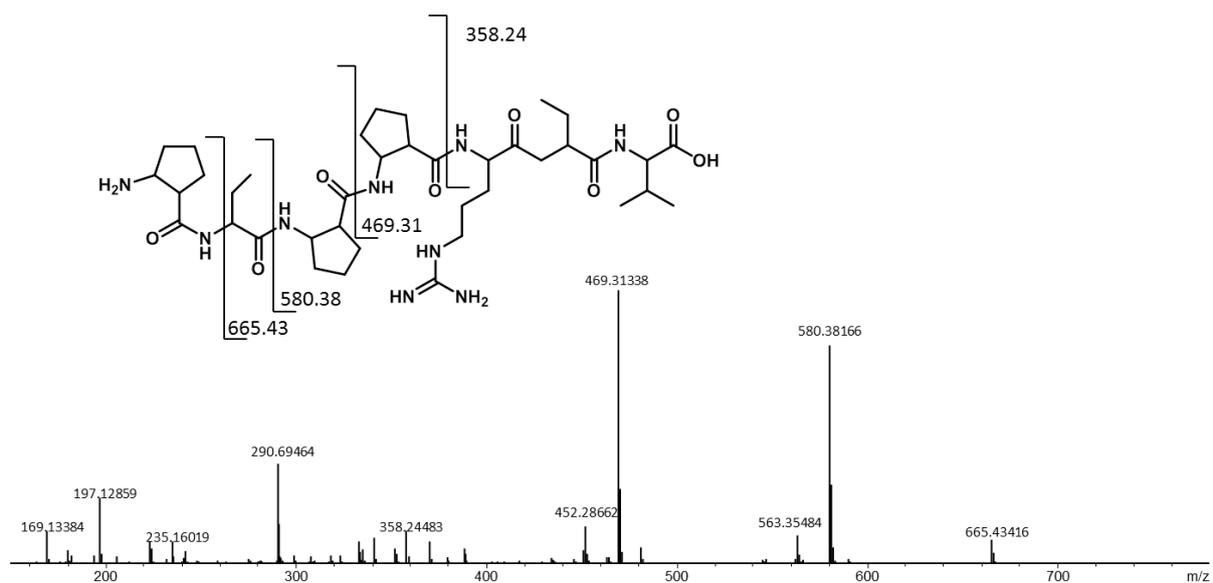
<sup>c</sup> proton showing COSY correlation to indicated proton.

<sup>d</sup> proton showing HMBC correlations to indicated carbons.

<sup>e</sup> overlap with solvent signal

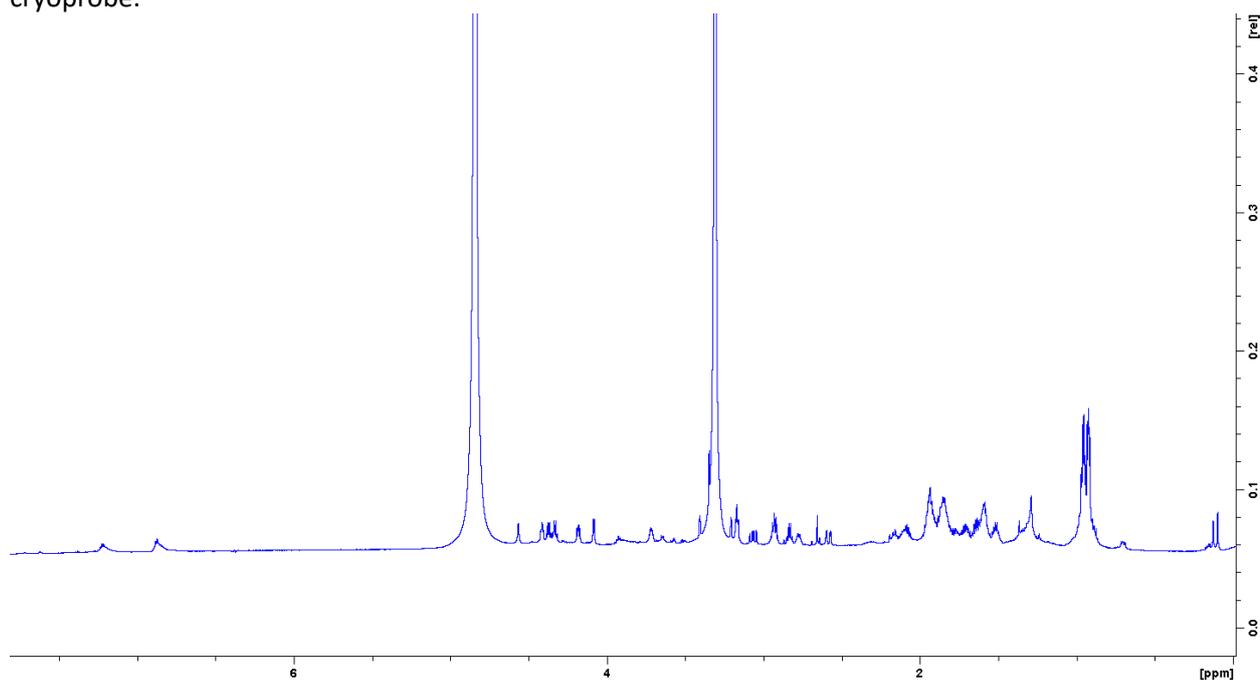
### Analytical data Myxopentacin 776a

Myxopentacin 776a was isolated as white solid. *hr*-Masses: double charged ion [M+2H]<sup>2+</sup>: 388.75522 m/z; single charged ion [M+H]<sup>+</sup>: 776.50760 m/z. UV max < 200 nm (UV data extracted from LC-MS-UV measurement using a Dionex ultimate 3000 HPLC system equipped with a DAAD UV-VIS detector)

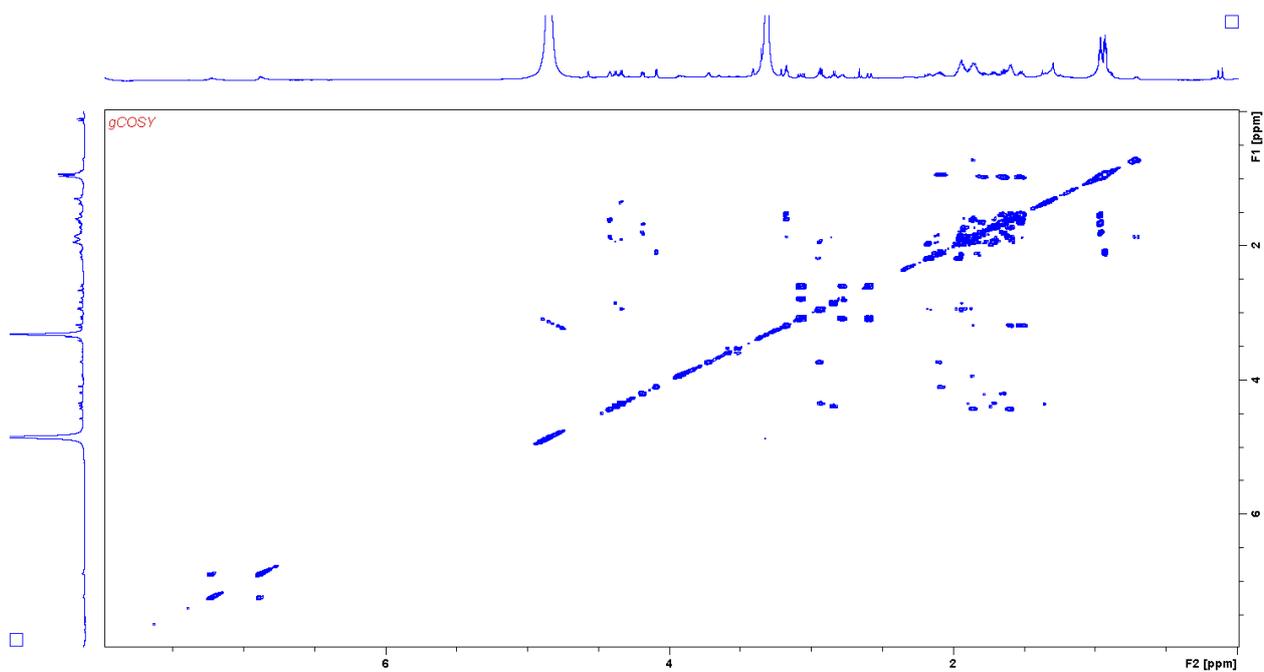


Supplementary figure 14: HR MS<sup>2</sup> spectrum and observed Fragments of myxopentacin 776a. MS<sup>2</sup> spectrum is extracted from LChr-MS measurement of an extract of MCy 9003.

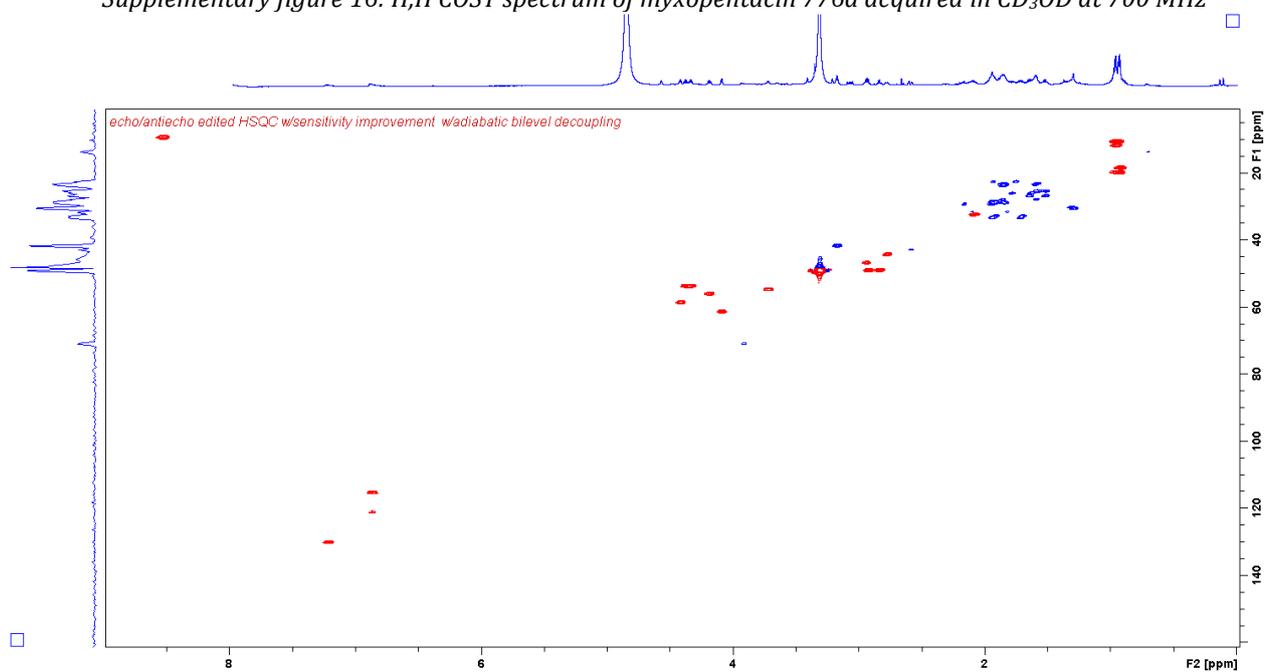
NMR spectra were acquired with a Bruker Ascend 700 NMR spectrometer equipped with a with a 5 mm TCI cryoprobe.



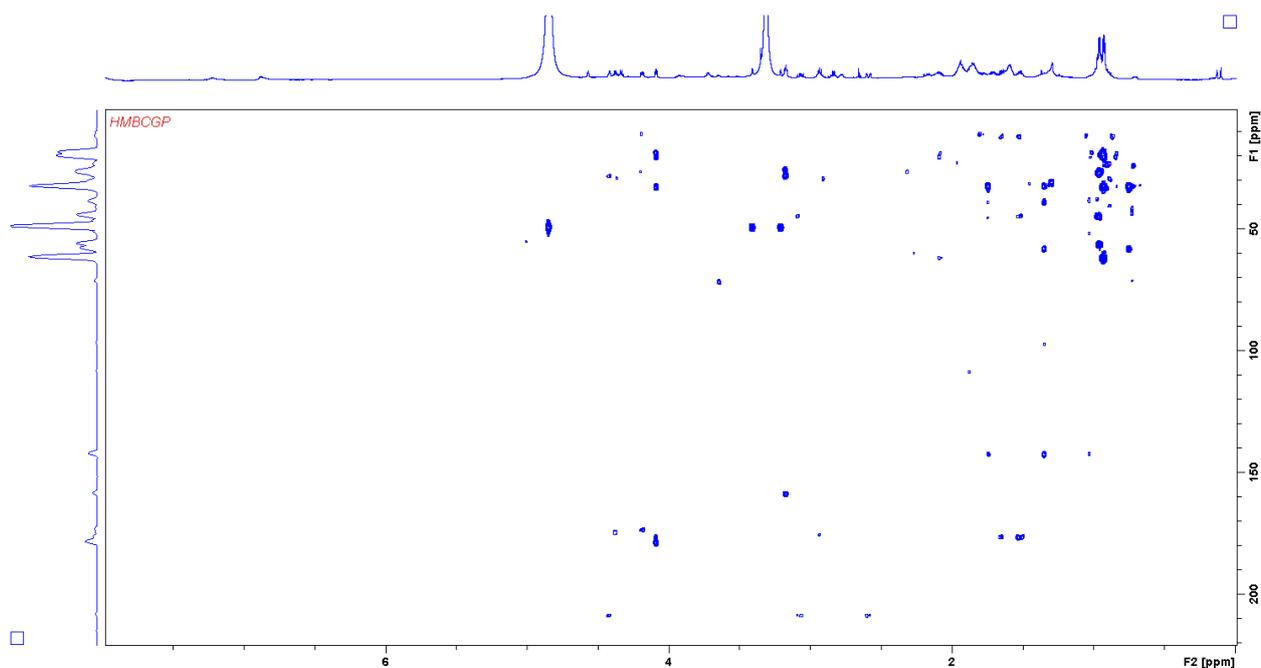
Supplementary figure 15: <sup>1</sup>H spectrum of myxopentacin 776a acquired in CD<sub>3</sub>OD at 700 MHz



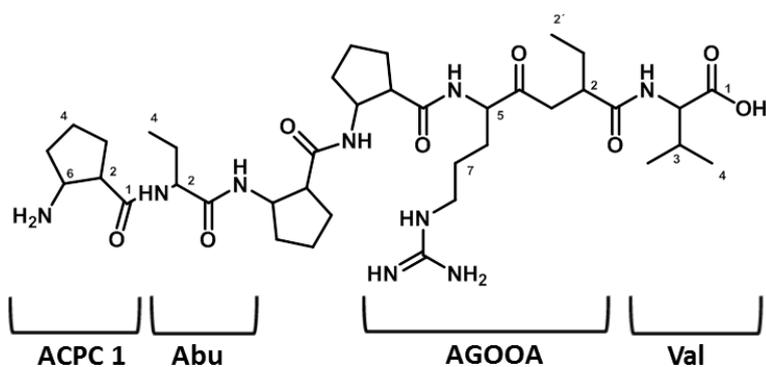
Supplementary figure 16:  $H,H$  COSY spectrum of myxopentacin 776a acquired in  $CD_3OD$  at 700 MHz



Supplementary figure 17: HSCQ spectrum of myxopentacin 776a acquired in  $CD_3OD$  at 700/175 MHz



Supplementary figure 18: HMBC spectrum of myxopentacin 776a acquired in CD<sub>3</sub>OD at 700/175 MHz



Supplementary figure 19: Structure of Myxopentacin 776a with assigned residues.

Supplementary table 5: Spectroscopic data of Myxopentacin 776a in CD<sub>3</sub>OD.

moiety	position	$\delta^{13}\text{C}^{\text{a}}$	$\delta^1\text{H}^{\text{b}}$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
ACPC 1	1	174.9			
	2	46.8	2.94 m	3,6	6,3
	3a	29.3	1.92 m		
	3b	29.3	2.16 m		
	4a	22.6	1.75 m		
	4b	22.6	1.92 m		
	5a	31.7	1.83 m		
	5b	31.7	2.11 m		

	6	54.7	3.72 m	2,5	3,4
<b>Abu</b>	1	173.3			
	2	55.8	4.16 d,d (8.7,5.2)	3	1,3,4,1 <sub>ACPC 1</sub>
	3a	25.9	1.78 m	2,3b,4	1,2,4
	3b	25.9	1.65 m	2,3a,4	1,2,4
	4	10.6	0.94 m		2,3
<b>ACPC 2</b>	1	174.6			
	2	48.4	2.83 d,d,d (7.8,7.8,7.8)	3,6	1,3
	3a	28.9	1.83 m		
	3b	28.9	1.93 m		
	4a	23.4	1.60 m		
	4b	23.4	1.86 m		
	5a	33.3	1.60 m		
	5b	33.3	1.86 m		
	6	53.7	4.37 d,d,d (6.9,6.9,6.9)	2,5	2,3,4,5,1 <sub>Abu</sub>
<b>ACPC 3</b>	1	175.3			
	2	49.2	2.93 m	3,6	1,3,6
	3a	28.5	1.87 m		
	3b	28.5	1.96 m		
	4a	23.2	1.58 m		
	4b	23.2	1.85 m		
	5a	32.8	1.70 m		
	5b	32.8	1.91 m		
	6	53.8	4.33 d,d,d (7.0,7.0,7.0)	2,5	4,5,1 <sub>ACPC 2</sub>
<b>AGOOA</b>	1	176.4			
	2	44.2	2.77 m	3	
	3a	42.8	3.07 d,d (18.5,11.0)	2,3b	4,1'
	3b	42.8	2.59 d,d (18.3,3.5)	2,3a	4,1'
	4	208.4			
	5	58.6	4.41 d,d (6.4,5.1)	6	4,6,7,1 <sub>ACPC 3</sub>
	6a	27.8	1.86 m		
	6b	27.8	1.86 m		
	7a	25.8	1.58 m		
	7b	25.8	1.50 m		
	8	41.8	3.18 d,d (6.3,6.3)	7	6,7,9
	9	158.4			
	1'a	26.8	1.64 m	2,2'	2,3,2'
	1'b	26.8	1.52 m	2,2'	2,3,2'
2'	11.7	0.96 m	1'		
<b>Val</b>	1	178.3			
	2	61.4	4.09 d(5.6)	3	1,3,1 <sub>AGOOA</sub>
	3	32.5	2.08 m	2,4,5	2,4,5
	4	19.9	0.93 d (4.3)		
	5	18.4	0.92 d (4.6)		

<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

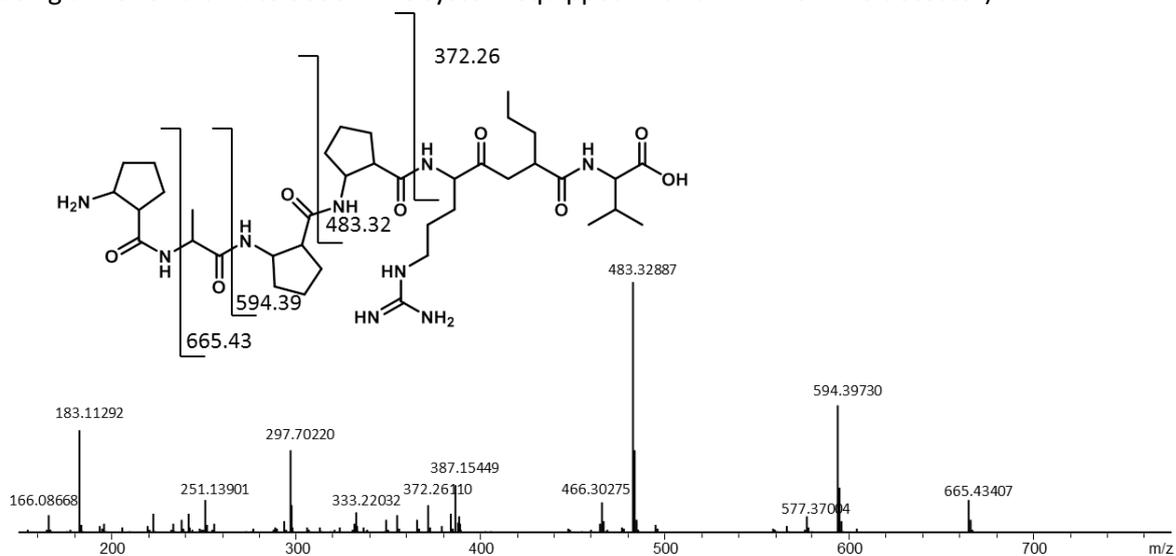
<sup>b</sup> acquired at 700 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

<sup>c</sup> proton showing COSY correlation to indicated proton.

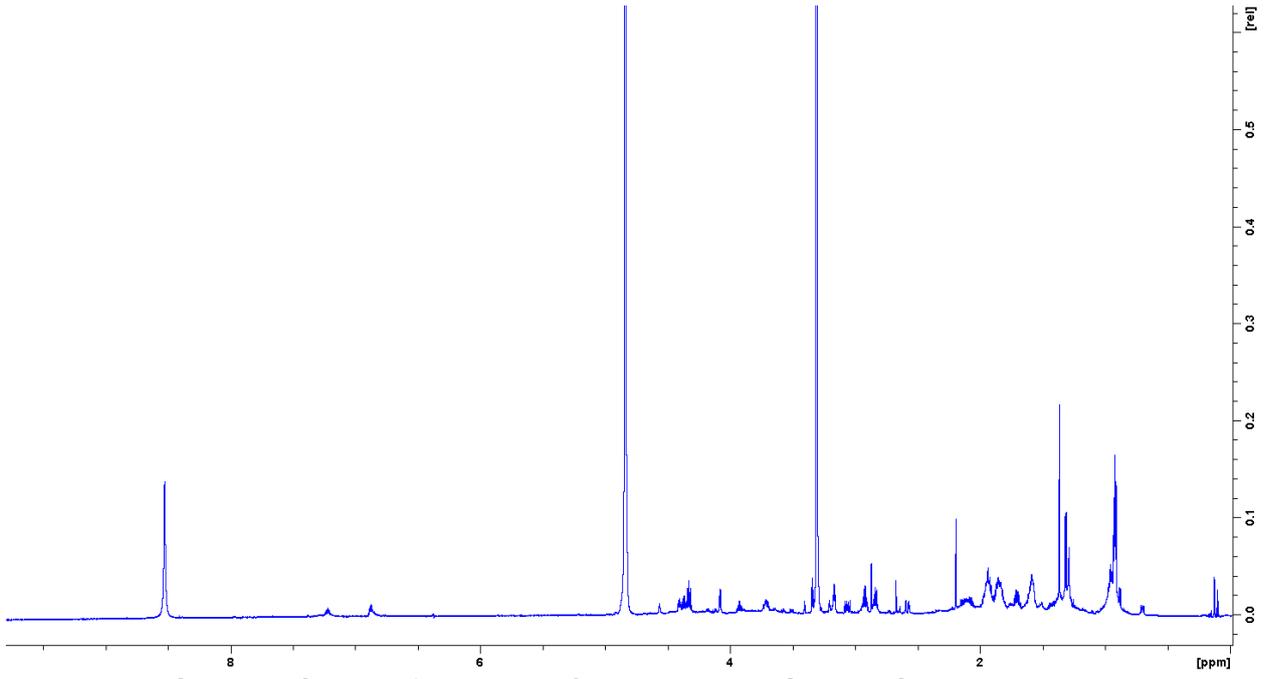
<sup>d</sup> proton showing HMBC correlations to indicated carbons.

### Analytical data Myxopentacin 776b

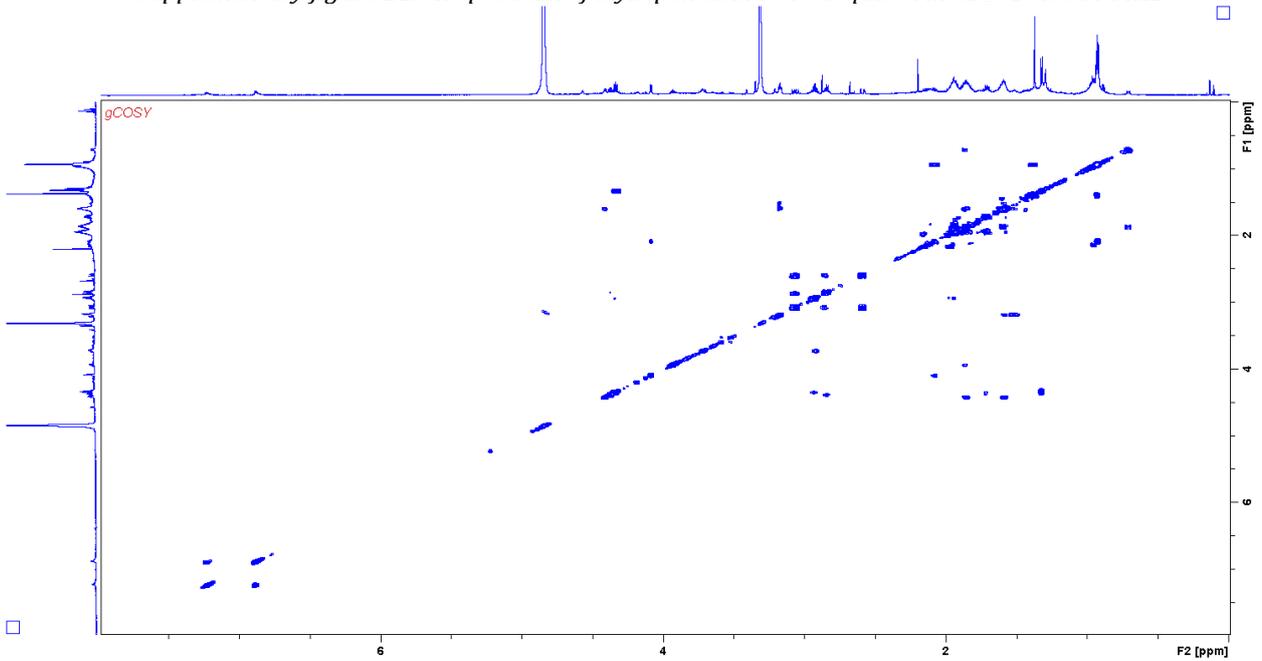
Myxopentacin 776b was isolated as white solid. *hr*-Masses: double charged ion  $[M+2H]^{2+}$ : 388.75649 m/z; single charged ion  $[M+H]^+$ : 776.50760 m/z. UV max < 200 nm (UV data extracted from LC-MS-UV measurement using a Dionex ultimate 3000 HPLC system equipped with a DAAD UV-VIS detector)



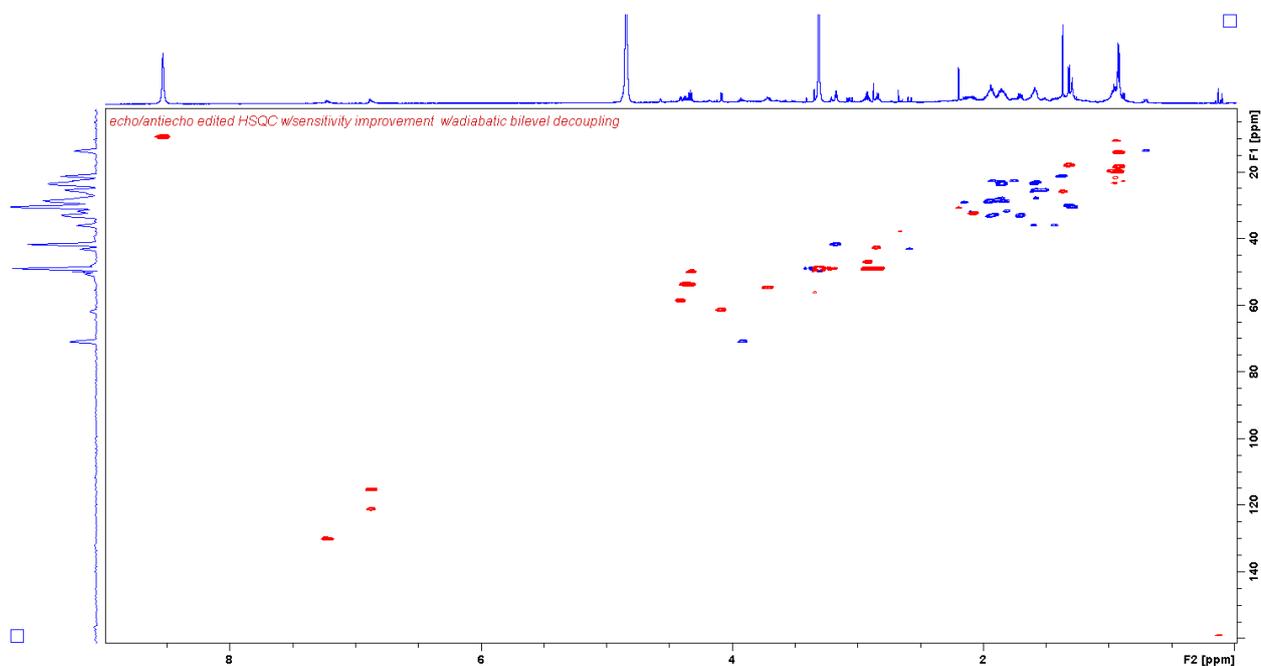
Supplementary figure 20: HR MS<sup>2</sup> spectrum and observed fragments of Myxopentacin 776b. MS<sup>2</sup> spectrum is extracted from LC-*hr*MS measurement of an extract of MCy 9003.



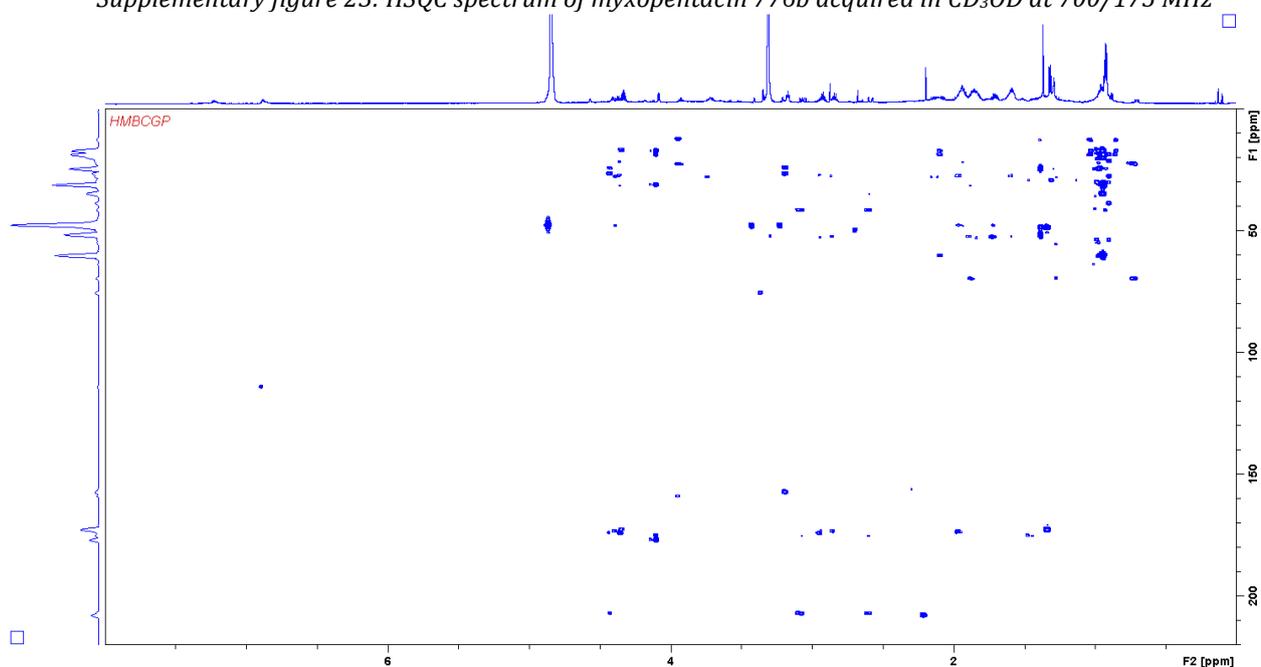
Supplementary figure 21:  $^1\text{H}$  spectrum of myxopentacin 776b acquired in  $\text{CD}_3\text{OD}$  at 700 MHz



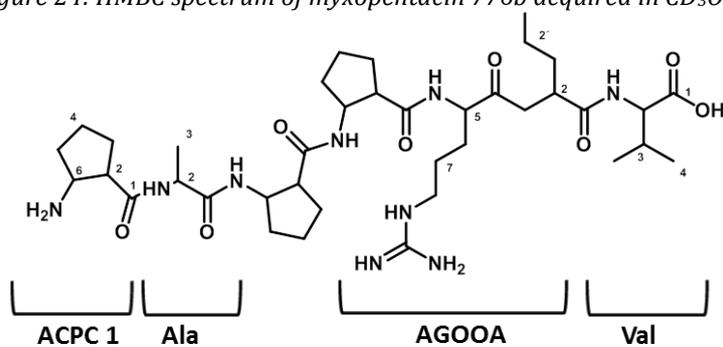
Supplementary figure 22:  $\text{H,H}$  COSY spectrum of myxopentacin 776b acquired in  $\text{CD}_3\text{OD}$  at 700 MHz



Supplementary figure 23: HSQC spectrum of myxopentacin 776b acquired in CD<sub>3</sub>OD at 700/175 MHz



Supplementary figure 24: HMBC spectrum of myxopentacin 776b acquired in CD<sub>3</sub>OD at 700/175 MHz



Supplementary figure 25: Structure of myxopentacin 776b with assigned residues.

Supplementary table 6: Spectroscopic data of Myxopentacin 776b in CD<sub>3</sub>OD.

moiety	position	$\delta^{13}\text{C}^a$	$\delta^1\text{H}^b$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
ACPC 1	1	174.3			

	2	46.7	2.93 m	3,6	1,3,4,5,6
	3a	29.1	1.97 m		
	3b	29.1	2.17 m		
	4a	22.5	1.93 m		
	4b	22.5	1.76 m		
	5a	31.5	2.12 m		
	5b	31.5	1.85 m		
	6	54.6	3.72 m	2,5	1,2,3,4,5
<b>Ala</b>	1	174.0			
	2	49.0	4.32 m	3	1,3
	3	18.0	1.32 d (7.2)	2	1,2
<b>ACPC 2</b>	1	174.6			
	2	49.1	2.84 m	3,6	1,3,4,5,6
	3a	28.7	1.93 m		
	3b	28.7	1.83 m		
	4a	23.3	1.85 m		
	4b	23.3	1.59 m		
	5a	33.1	1.96 m		
	5b	33.1	1.67 m		
	6	53.4	4.38 m	2,5	2,3,4,5
<b>ACPC 3</b>	1	175.3			
	2	48.9	2.92 m	3,6	1,3,4,5,6
	3a	28.6	1.95 m		
	3b	28.6	1.86 m		
	4a	23.2	1.85 m		
	4b	23.2	1.59 m		
	5a	32.9	1.92 m		
	5b	32.9	1.70 m		
	6	53.6	4.34 m	2,5	1,2,3,4,5
<b>AGOOA</b>	1	176.4			
	2	42.5	2.84 m	4,1'	4
	3a	43.1	3.06 d,d (18.7, 10.9)	2	1,2,4,1'
	3b	43.1	2.58 d,d (18.3, 3.4)	2	1,2,4,1'
	4	208.3			
	5	58.6	4.41 m	6	4,6,7,1 <sub>ACPC 1</sub>
	6a	27.8	1.86 m	5,6b,7	5,7
	6b	27.8	1.59 m	5,6a,7	5,7
	7a	25.4	1.51 m	6,7b,8	5
	7b	25.4	1.57 m	6,7a,8	5
	8	41.6	3.17 t (6.5)	7	6,7,9
	9	158.7			
	1'a	36.0	1.59 m		
	1'b		1.43 m		
	2'	21.0	1.37 m	1',3'	
3'	14.0	0.92 m	2'	1',2'	
<b>Val</b>	1	178.0			
	2	61.1	4.09 d (6.0)	3	1,3,4,5,1 <sub>AGOOA</sub>
	3	32.3	2.08 m	2	1,2,4,5
	4	18.5	0.93 d (6.8)	3	2,3,5
	5	19.9	0.93 d (6.8)	3	2,3,4

<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

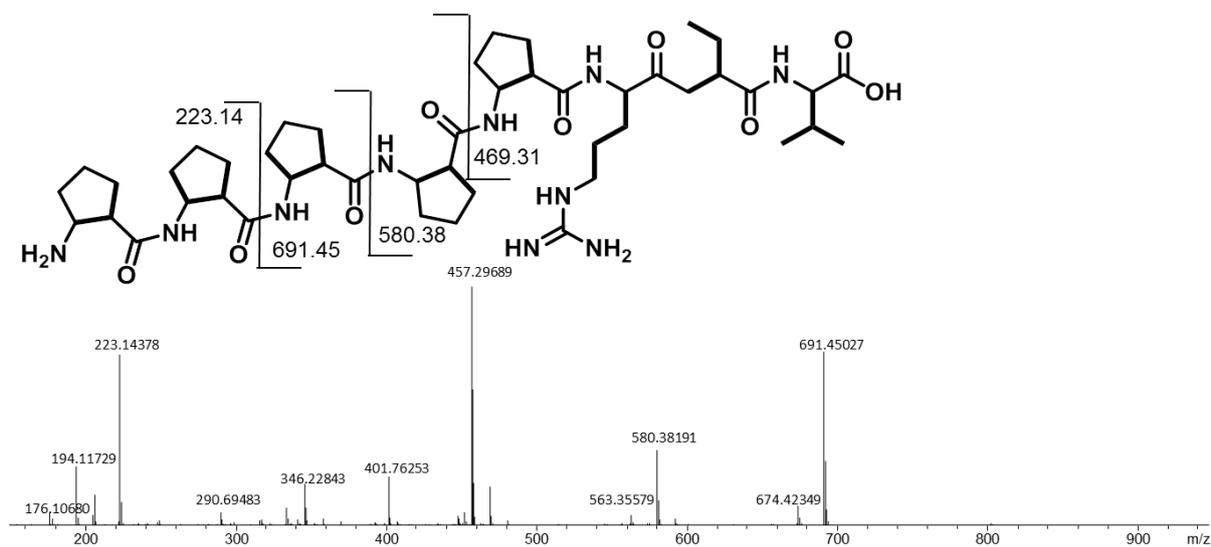
<sup>b</sup> acquired at 700 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

<sup>c</sup> proton showing COSY correlation to indicated proton.

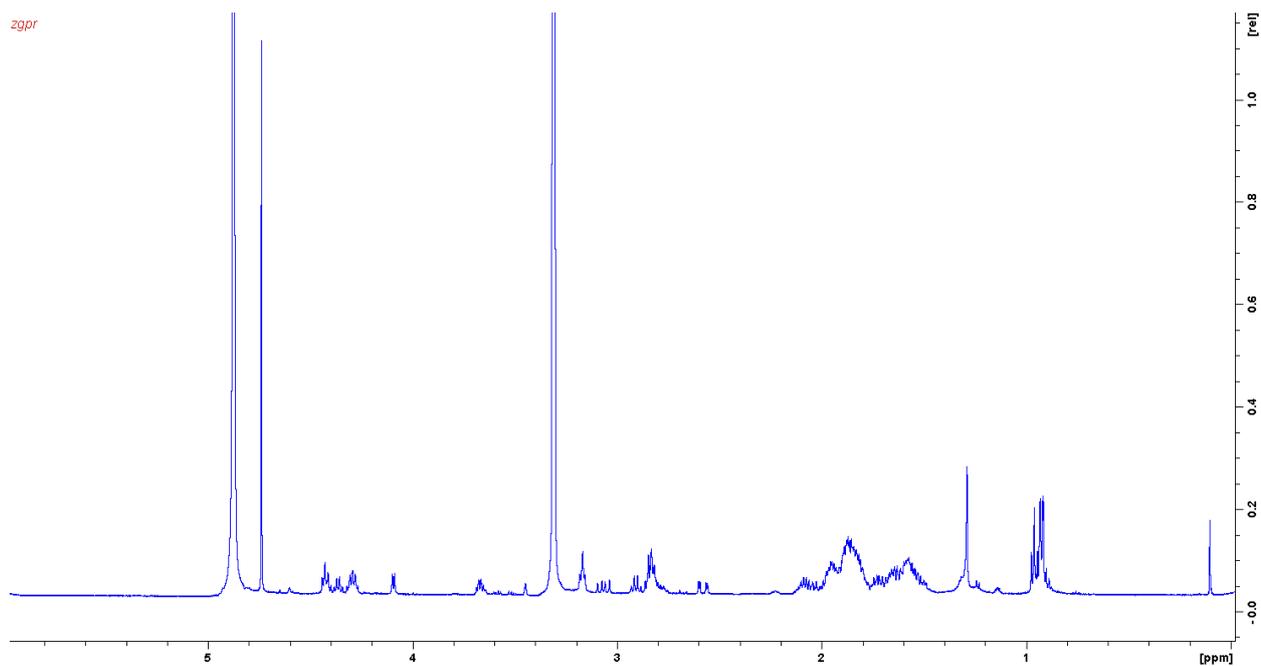
<sup>d</sup> proton showing HMBC correlations to indicated carbons.

### Analytical data Myxopentacin 913

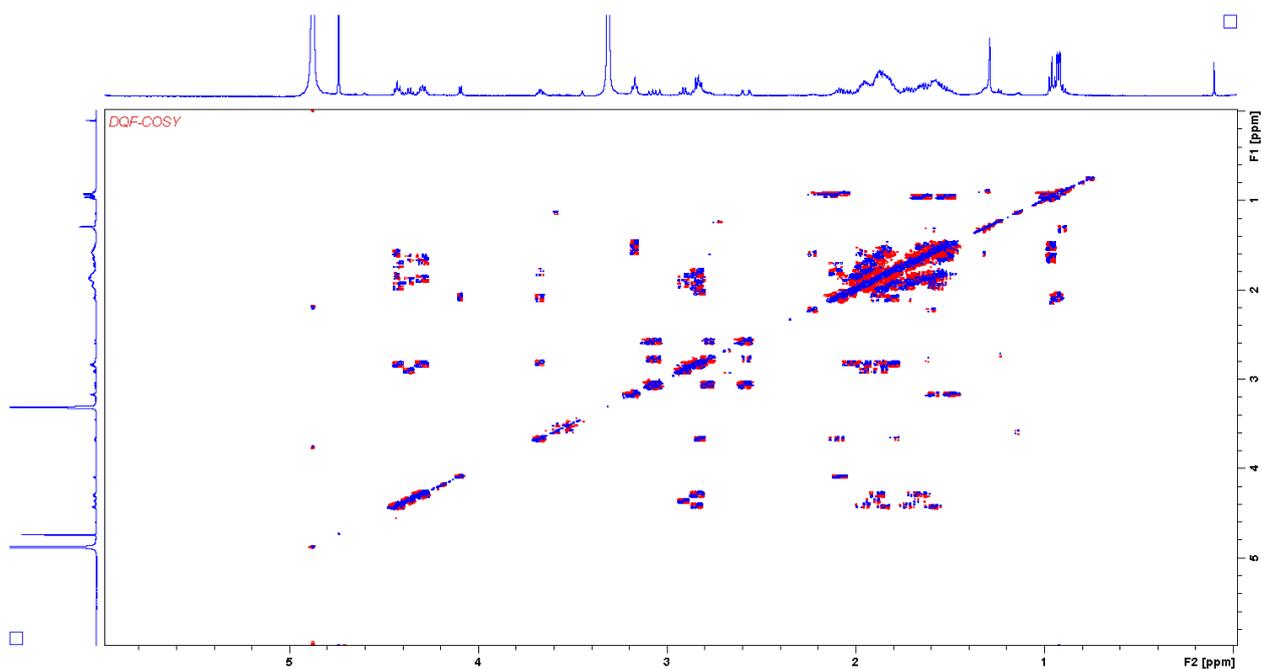
Myxopentacin 913 was isolated as white solid. *hr*-Masses: double charged ion  $[M+2H]^{2+}$ : 457.29713 m/z; single charged ion  $[M+H]^+$ : 913.59322 m/z. UV max < 200 nm (UV data extracted from LC-MS-UV measurement using a Dionex Ultimate 3000 HPLC system equipped with a DAAD UV-VIS detector)



Supplementary figure 26: HR MS<sup>2</sup> spectrum and observed fragments of myxopentacin 913. MS<sup>2</sup> spectrum is extracted from LC-*hr*MS measurement of an extract of MCy 9003.



Supplementary figure 27:  $^1\text{H}$  spectrum of myxopentacin 913 acquired in  $\text{CD}_3\text{OD}$  at 500 MHz



Supplementary figure 28:  $\text{H,H DQF COSY}$  spectrum of myxopentacin 913 acquired in  $\text{CD}_3\text{OD}$  at 500 MHz



Supplementary table 7: Spectroscopic data of Myxopentacin 913 in CD<sub>3</sub>OD.

moiety	position	$\delta^{13}\text{C}^a$	$\delta^1\text{H}^b$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
ACPC 1	1	174.2			
	2	46.6	2.82 m	3,6	1,3
	3a	29.7	1.88 m		
	3b	29.7	2.03 m		
	4a	22.4	1.90 m		
	4b	22.4	1.74 m		
	5a	31.6	2.09 m		
	5b	31.6	1.80m		
	6	54.6	3.67 m	2,5	3,4
ACPC 2	1	174.6-175.0			
	2	49.1	2.82 m	3,6	3,4,5
	3a	28.5	1.81 m		
	3b	28.5	1.97 m		
	4a	23.5	1.54 m		
	4b	23.5	1.88 m		
	5a	33.0	1.95 m		
	5b	33.0	1.73 m		
	6	53.5	4.42 m	2,5	2,3,4,5
ACPC 3	1	174.6-175.1			
	2	49.1	2.82 m	3,6	1,3,4,5,6
	3a	28.5	1.80 m		
	3b	28.5	1.96 m		
	4a	23.5	1.54 m		
	4b	23.5	1.86 m		
	5a	32.9	1.63 m		
	5b	32.9	1.90 m		
	6	53.7-54.0	4.29 m	2,5	2,3,4,5,1 <sub>ACPC 2</sub>
ACPC 4	1	174.6-175.1			
	2	49.1	2.82 m	3,6	1,3,4,5,6
	3a	28.5	1.80 m		
	3b	28.5	1.96 m		
	4a	23.5	1.54 m		
	4b	23.5	1.86 m		
	5a	32.9	1.63 m		
	5b	32.9	1.90 m		
	6	53.7-54.0	4.29 m	2,5	2,3,4,5,1 <sub>ACPC 3</sub>
ACPC 5	1	175.3			
	2	49.1	2.90 d,d,d (7.9,7.7,7.6)	3,6	1,3,4,5,6
	3a	28.2	1.86 m		
	3b	28.2	1.94 m		
	4a	23.2	1.57 m		
	4b	23.2	1.85 m		
	5a	33.2	1.64 m		
	5b	33.2	1.89 m		
	6	54.0	4.36 d,d,d(6.7,6.8,6.5)	2,5	1,2,3,4,5,6,1 <sub>ACPC 4</sub>
AGOOA	1	176.4			
	2	44.2	2.78 m	1',3	
	3a	42.7	2.57 d,d (18.3,3.6)	2,3b	1,2,4
	3b	42.7	3.06 d,d (18.4,10.8)	2,3a	1,2,4
	4	208.2			

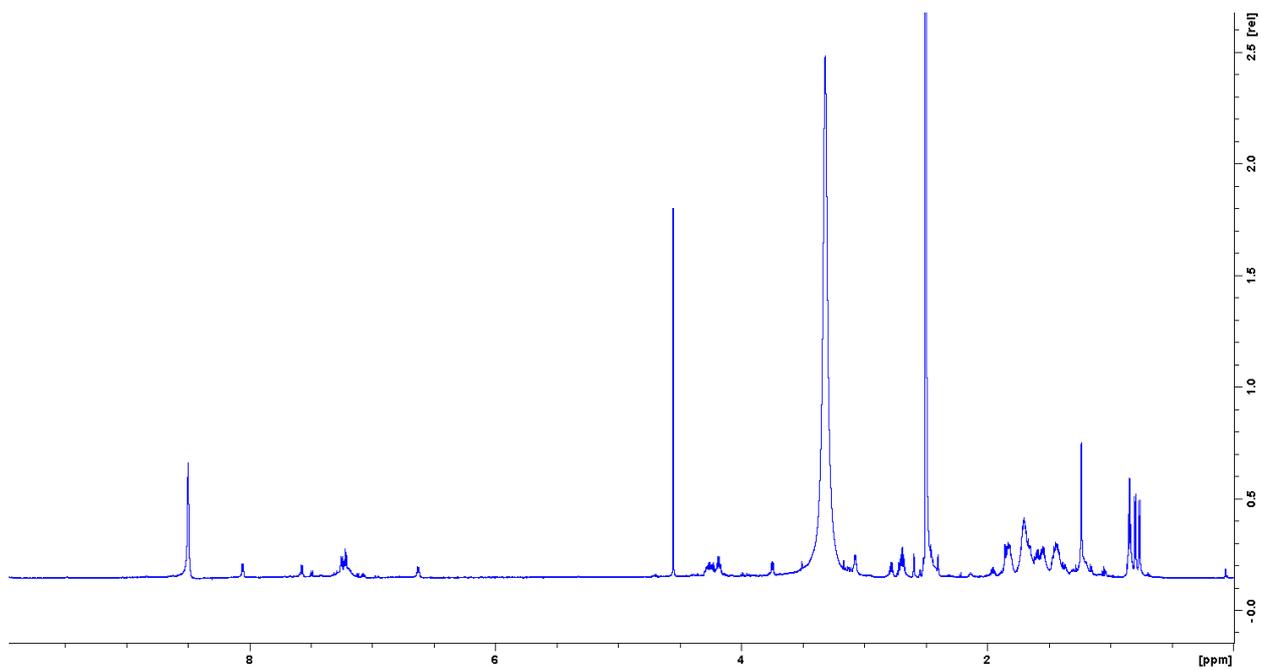
	5	58.5	4.43 m	6	4,6,7,1 <sub>ACPC</sub> 5
	6a	27.7	1.57 m	5,6b,7	
	6b	27.7	1.85 m	5,6a,7	
	7a	25.3	1.48 m	6,7b,8	
	7b	25.3	1.57 m	6,7a,8	
	8	41.7	3.17 d,d (6.1,6.7)	7	6,7,9
	9	158.3			
	1'a	26.8	1.51 m	2,1'b,2'	1
	1'b	26.8	1.64 m	2,1'a,2'	1
	2'	11.8	0.96 t (7.3)	1'	2,1'
	1	178.3			
	2	61.4	4.09 d (5.6)	3	1,3,4,5,1 <sub>AGOOA</sub>
<b>Val</b>	3	32.3	2.08 m	2,4,5	
	4	18.4	0.93 d (6.9)	3	2,3,5
	5	19.9	0.93 d (6.9)	3	2,3,4

<sup>a</sup> acquired at 125 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

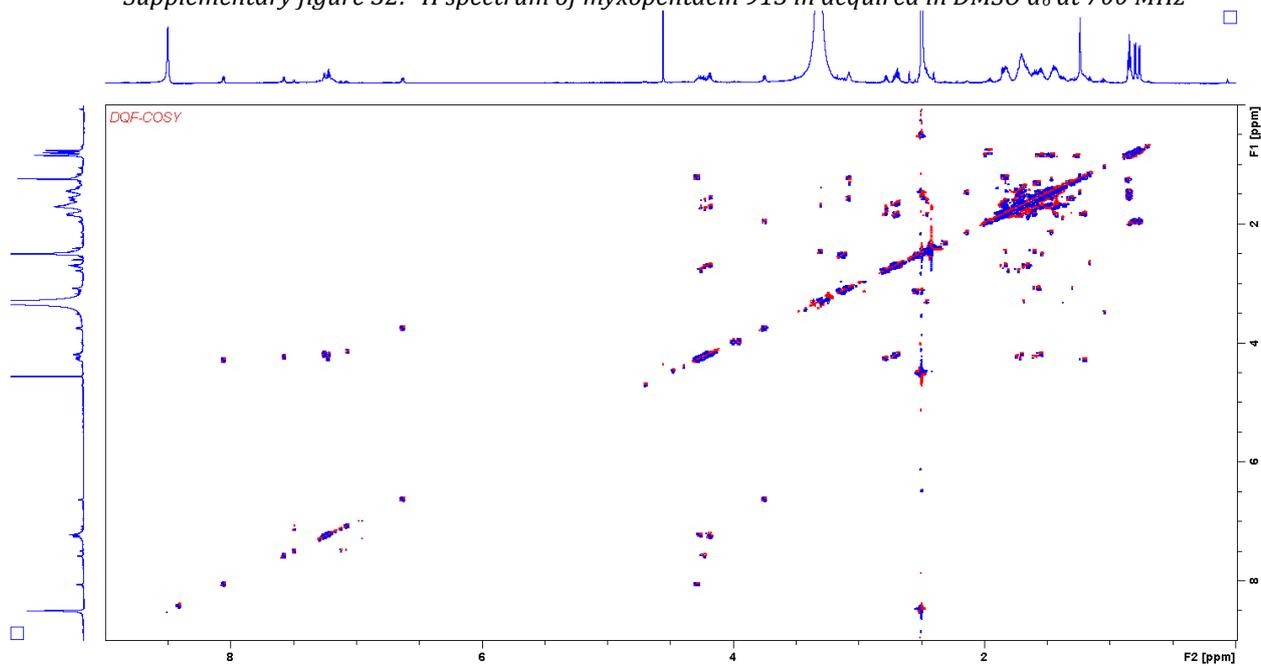
<sup>b</sup> acquired at 500 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

<sup>c</sup> proton showing COSY correlation to indicated proton.

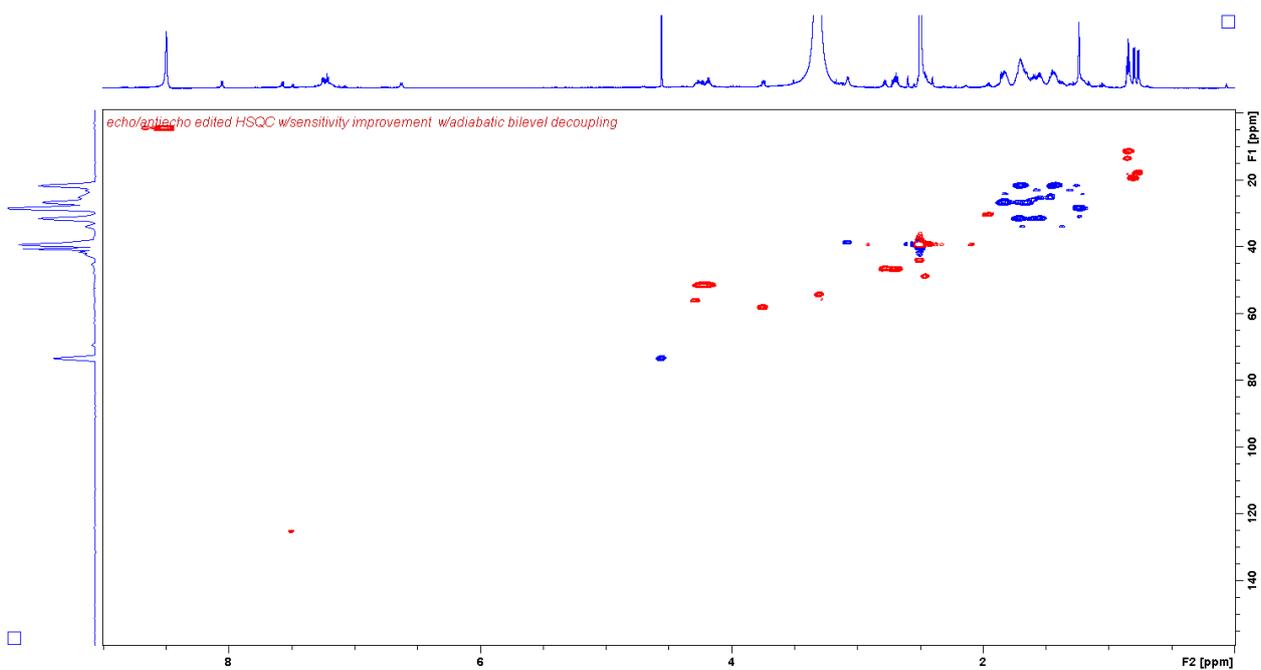
<sup>d</sup> proton showing HMBC correlations to indicated carbons.



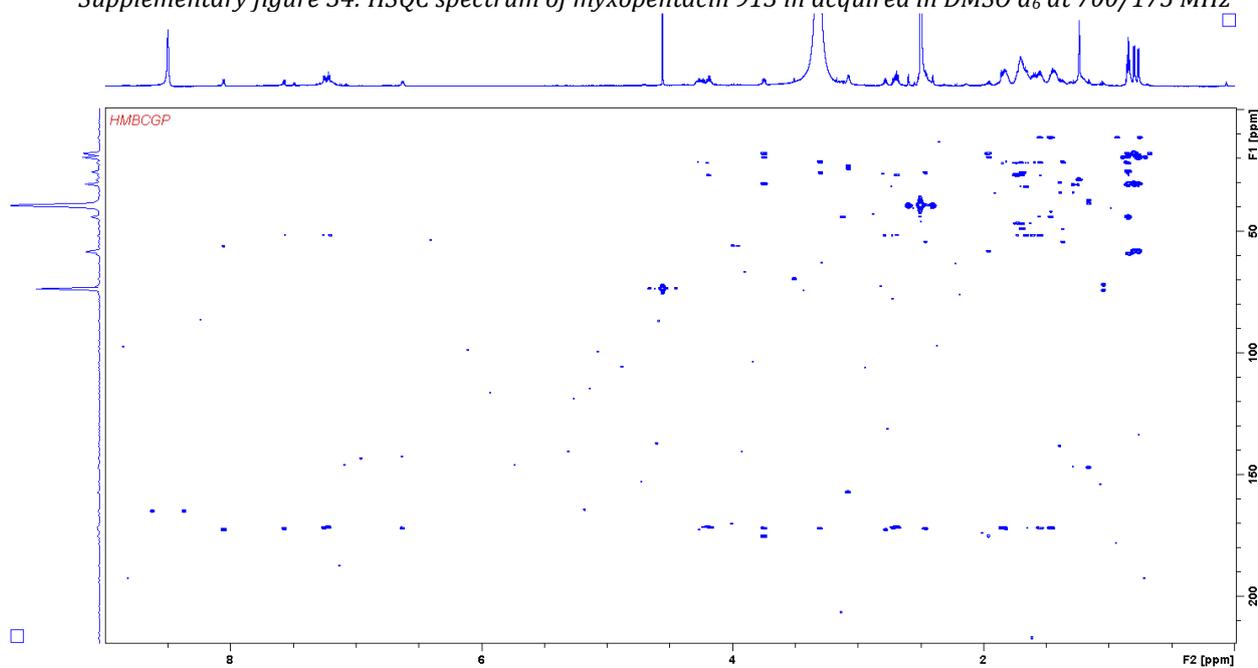
Supplementary figure 32:  $^1\text{H}$  spectrum of myxopentacin 913 in acquired in  $\text{DMSO } d_6$  at 700 MHz



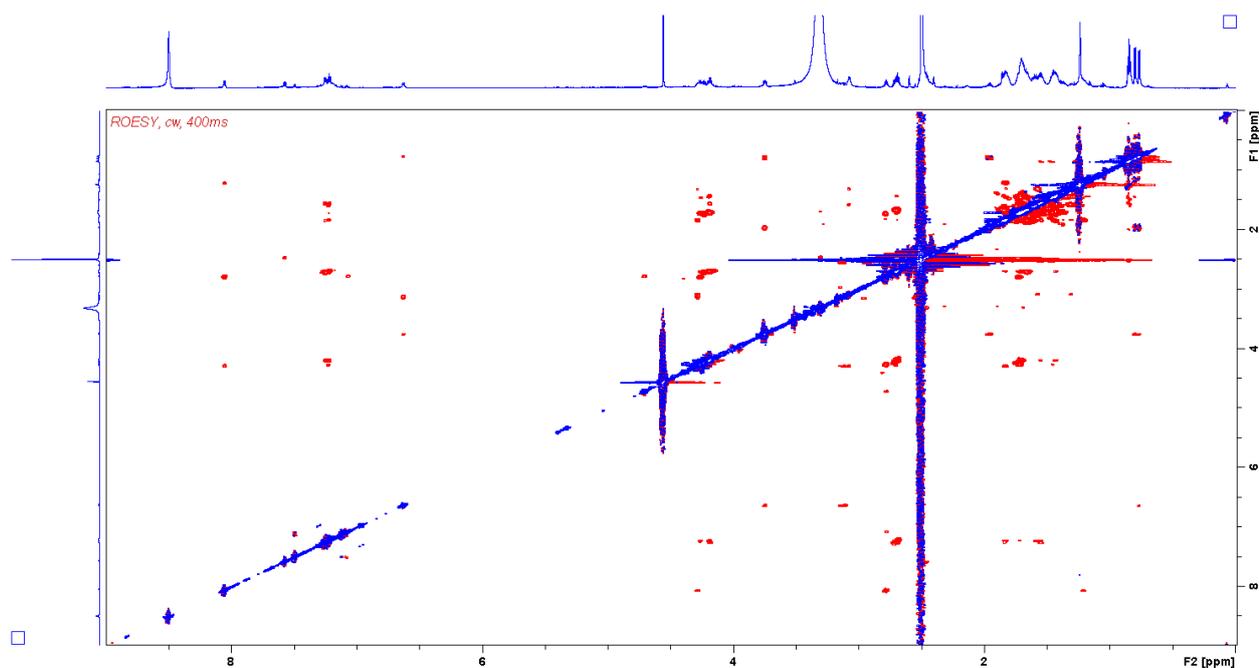
Supplementary figure 33:  $\text{H,H DQF COSY}$  spectrum of myxopentacin 913 in acquired in  $\text{DMSO } d_6$  at 700 MHz



Supplementary figure 34: HSQC spectrum of myxopentacin 913 in acquired in DMSO  $d_6$  at 700/175 MHz



Supplementary figure 35: HSQC spectrum of myxopentacin 913 in acquired in DMSO  $d_6$  at 700/175 MHz



Supplementary figure 36: ROESY spectrum of Myxopentacin 913 in acquired in DMSO  $d_6$  at 700 MHz

Supplementary table 8: Spectroscopic data of Myxopentacin 913 in DMSO  $d_6$ .

moiety	position	$\delta^{13}\text{C}^a$	$\delta^1\text{H}^b$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
ACPC 1	1	172.4			
	2	48.9	2.46 <sup>e</sup>	3,6	1,3,4
	3a	26.1	1.82 m		
	3b	26.1	1.61 m		
	4a	21.7	1.69 m		
	4b	21.7	1.43 m		
	5a	34.1	1.69 m		
	5b	34.1	1.36 m		
	6	54.6	3.31 <sup>e</sup>	2,5	1,3,4
	6-NH		n.d.		
ACPC 2	1	171.8			
	2	46.5	2.71 m	3,5	1,3,6
	3a	26.7	1.84 m		
	3b	26.7	1.67 m		
	4a	21.9	1.70 m		
	4b	21.9	1.44 m		
	5a	31.7	1.72 m		
	5b	31.7	1.56 m		
	6	51.6	4.22 m	2,5,6-NH	2,3,5
	6-NH		7.57 d (7.8)	6	6,1 <sub>ACPC 1</sub>
ACPC 3	1	171.8			
	2	46.5	2.71 m	3,6	1,3,5,6
	3a	26.7	1.84 m		
	3b	26.7	1.67 m		
	4a	21.9	1.70 m		
	4b	21.9	1.44 m		
	5a	31.7	1.72 m		
	5b	31.7	1.56 m		
	6	51.6	4.22 m	2,5,6-NH	5,2,1 <sub>ACPC 2</sub>
	6-NH		7.22 m	6	6,1 <sub>ACPC 2</sub>
ACPC 4	1	171.7			

	2	46.6	2.70 m	3,6	1,2,3
	3a	26.7	1.84 m		
	3b	26.7	1.67 m		
	4a	21.9	1.70 m		
	4b	21.9	1.44 m		
	5a	31.7	1.72 m		
	5b	31.7	1.56 m		
	6	51.5	4.19 m	5,2,6-NH	2,3,4,1 <sub>ACPC 3</sub>
	6-NH		7.22 m	6	6,1 <sub>ACPC 3</sub>
<b>ACPC 5</b>	1	172.7			
	2	46.6	2.78 d,d,d (7.6)	3,6	1,5
	3a	27.1	1.82 m		
	3b	27.1	1.71 m		
	4a	21.7	1.45 m		
	4b	21.7	1.72 m		
	5a	31.7	1.70 m		
	5b	31.7	1.62 m		
	6	51.7	4.26 m	2,5,6-NH	4,5,1 <sub>ACPC 4</sub>
	6-NH		7.24 m	6	6,1 <sub>ACPC 4</sub>
<b>AGOOA</b>	1	172.2			
	2	44.2	2.50 <sup>e</sup>	3a,1'	1
	3a	41.9	3.12	3b,2	1,4
	3b	41.9	2.55	3a,2	
	4	206.8			
	5	56.2	4.29 m	6,5-NH	6,1 <sub>ACPC 5</sub>
	5-NH		8.06 d (7.05)	5	5,1 <sub>ACPC 5</sub>
	6a	24.4	1.83 m		
	6b	24.4	1.21 m		
	7a	23.1	1.58 m		
	7b	23.1	1.31 m		
	8	38.9	3.08 m	7	6,7,9
	9	157.3			
	1'a	25.6	1.46 m		
1'b	25.6	1.55 m			
	2'	11.5	0.84 t (7.4)	1'	2,1'
<b>Val</b>	1	175.4			
	2	58.3	3.75 d,d (7.9,4.7)	3,2-NH	1,3,4,5,1 <sub>AGOOA</sub>
	2-NH		6.62 d (8.2)	2	1 <sub>AGOOA</sub>
	3	30.5	1.96 m	2,3,4,5	1,2,4,5
	4	17.9	0.76 d (6.8)	3	2,3,5
	5	19.6	0.79 d (6.7)	3	2,3,4

<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal DMSO d<sub>6</sub> at  $\delta$  39.52 ppm.

<sup>b</sup> acquired at 700 MHz, referenced to solvent signal DMSO d<sub>6</sub> at  $\delta$  2.54 ppm.

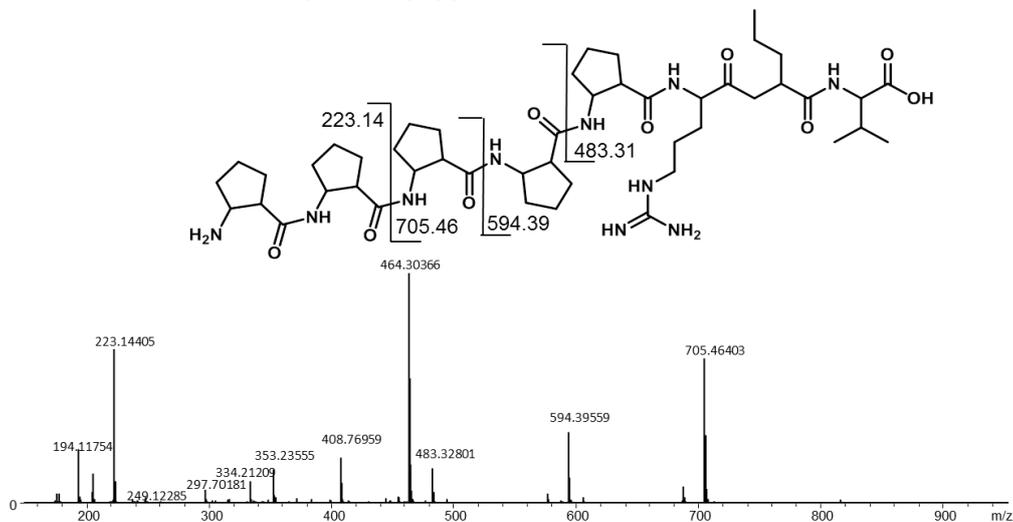
<sup>c</sup> proton showing COSY correlation to indicated proton.

<sup>d</sup> proton showing HMBC correlations to indicated carbons.

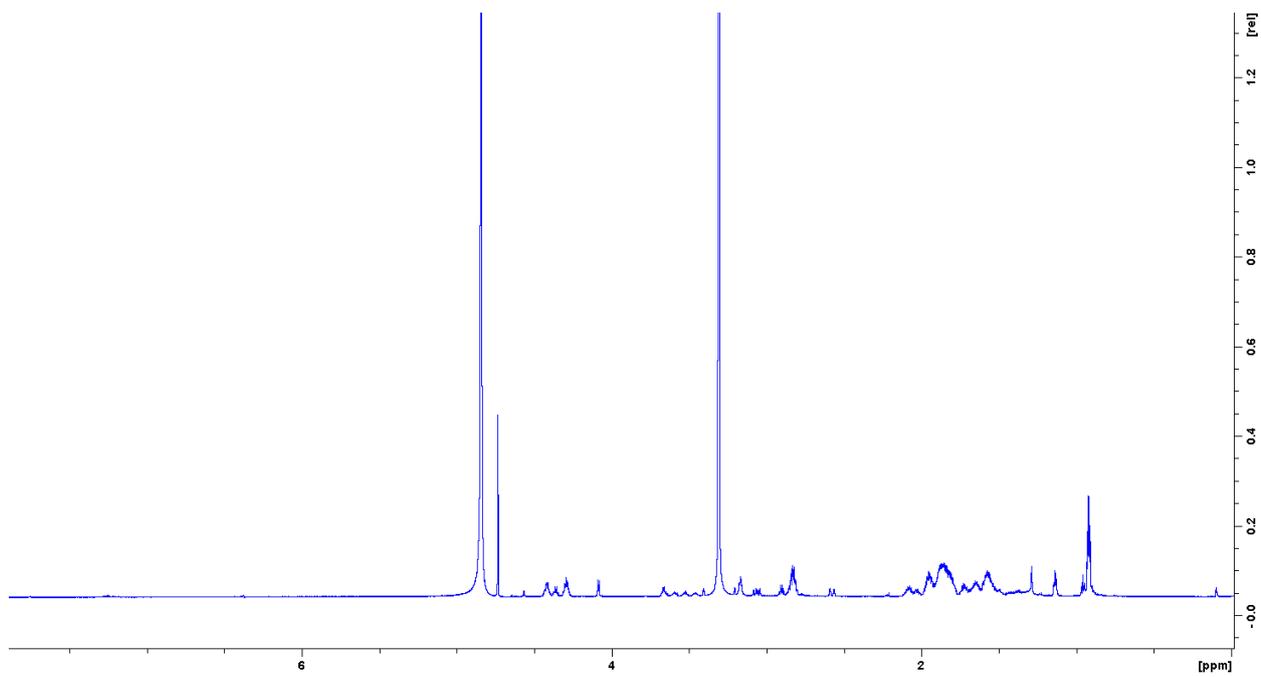
<sup>e</sup> overlap with solvent signal

### Analytical data Myxopentacin 927

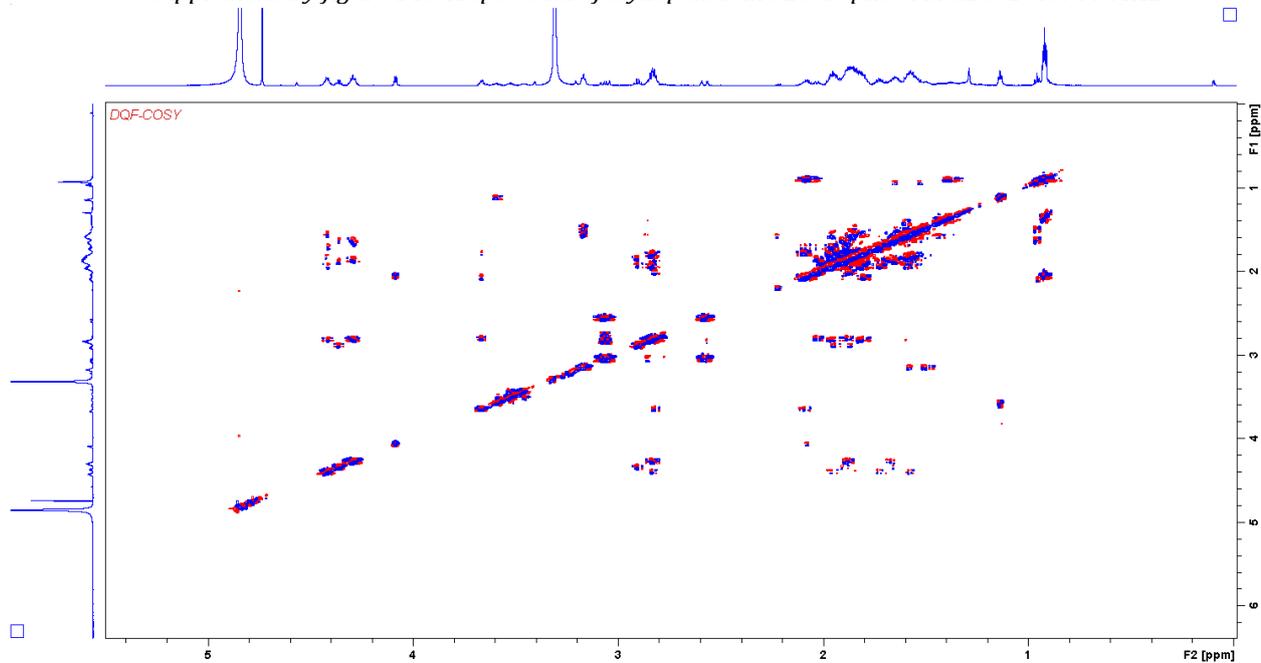
Myxopentacin 927 was isolated as white solid. *hr*-Masses: double charged ion  $[M+2H]^{2+}$ : 464.30494 m/z; single charged ion  $[M+H]^+$ : 927.60913 m/z. UV max < 200 nm (UV data extracted from LC-MS-UV measurement using a Dionex Ultimate 3000 HPLC system equipped with a DAAD UV-VIS detector)



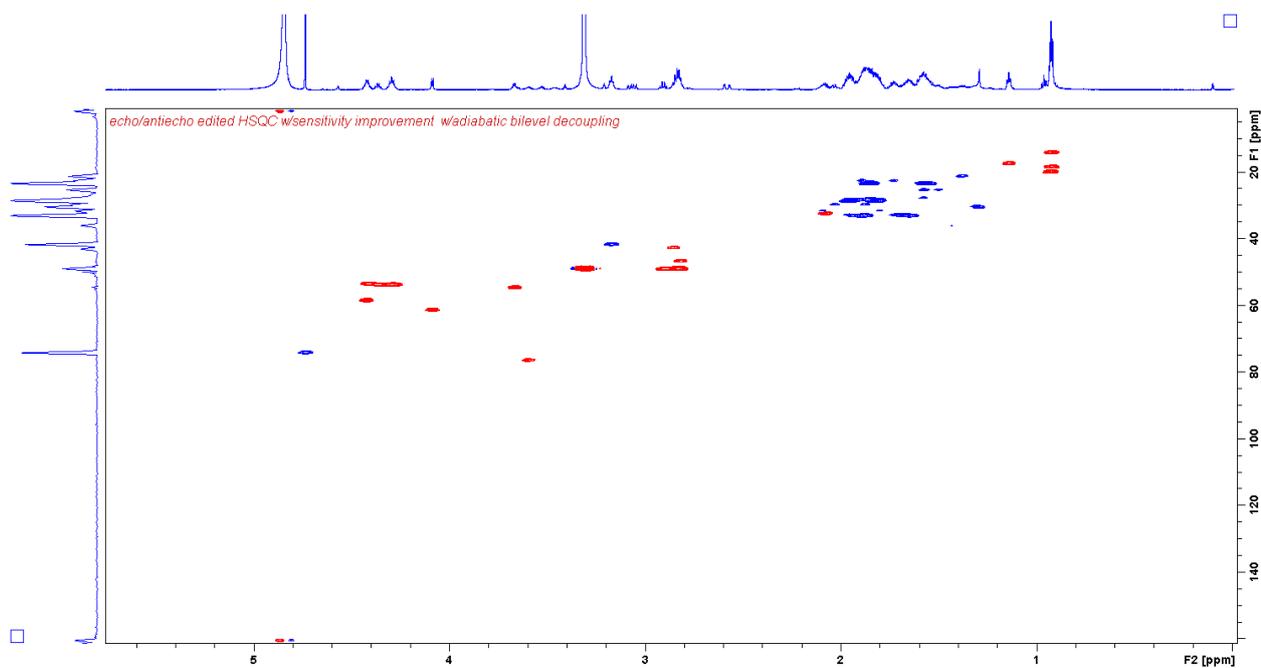
Supplementary figure 37: HR MS<sup>2</sup> spectrum and observed fragments of myxopentacin 927. MS<sup>2</sup> spectrum is extracted from LC-*hr*MS measurement of an extract of MCy 9003.



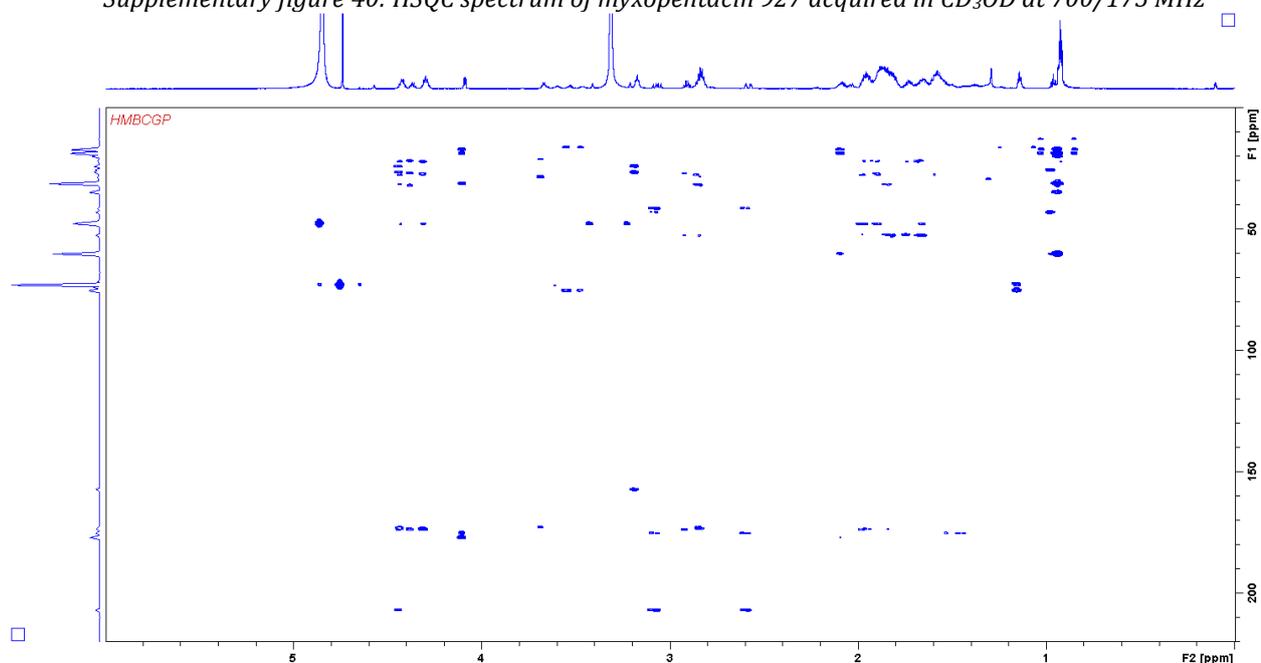
Supplementary figure 38:  $^1\text{H}$  spectrum of myxopentacin 927 acquired in  $\text{CD}_3\text{OD}$  at 700 MHz



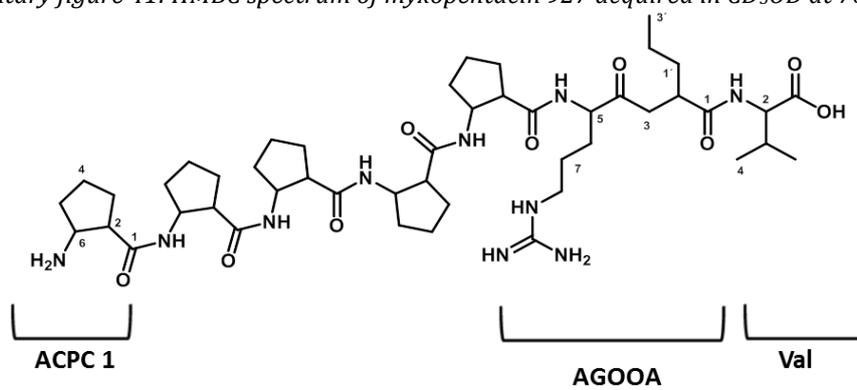
Supplementary figure 39:  $\text{H,H}$  DQF COSY spectrum of myxopentacin 927 acquired in  $\text{CD}_3\text{OD}$  at 700 MHz



Supplementary figure 40: HSQC spectrum of myxopentacin 927 acquired in CD<sub>3</sub>OD at 700/175 MHz



Supplementary figure 41: HMBC spectrum of myxopentacin 927 acquired in CD<sub>3</sub>OD at 700/175 MHz



Supplementary figure 42: Structure of Myxopentacin 27 with assigned residues.

Supplementary table 9: Spectroscopic data of Myxopentacin 927 in CD<sub>3</sub>OD.

moiety	position	$\delta^{13}\text{C}^{\text{a}}$	$\delta^1\text{H}^{\text{b}}$ (J in Hz)	COSY <sup>c</sup>	HMBC <sup>d</sup>
ACPC 1	1	174.1			
	2	46.6	2.82 m	3,6	1,3
	3a	29.7	1.88 m		
	3b	29.7	2.03 m		
	4a	22.4	1.89 m		
	4b	22.4	1.73 m		
	5a	31.6	2.09 m		
	5b	31.6	1.79 m		
	6	54.5	3.67 m	2,5	1,3,4,5
ACPC 2	1	174.0-175.0			
	2	49.2	2.82 m	3,6	1,3,4,5,6
	3a	28.5	1.97 m		
	3b	28.5	1.81 m		
	4a	23.5	1.88 m		
	4b	23.5	1.54 m		
	5a	33.0	1.95 m		
	5b	33.0	1.73 m		
	6	53.5	4.42 m	2,5	2,3,4,5
ACPC 3	1	174.6-175.0			
	2	49.1	2.83 m	3,6	1,3,4,5,6
	3a	28.5	1.93 m		
	3b	28.5	1.80 m		
	4a	23.5	1.86 m		
	4b	23.5	1.54 m		
	5a	33.0	1.88 m		
	5b	33.0	1.67 m		
	6	53.7	4.29 m	2,5	2,3,4,5
ACPC 4	1	174.6-175			
	2	49.1	2.83 m	3,6	1,3,6
	3a	28.5	1.80 m	2	
	3b	28.5	1.93 m		
	4a	23.5	1.54 m		
	4b	23.5	1.86 m		
	5a	33.0	1.88 m		
	5b	33.0	1.67 m		
	6	53.7	4.29 m	2,5	2,3,4,5
ACPC 5	1	175.4			
	2	49.1	2.91 d,d (15.5,7.9)	3,6	1,3,4
	3a	28.2	1.94 m		
	3b	28.2	1.86 m		
	4a	23.2	1.57 m		
	4b	23.2	1.85 m		
	5a	33.2	1.89 m		
	5b	33.2	1.64 m		
	6	53.7	4.36 m	2,5	3,4,5,1 <sub>ACPC4</sub>
AGOOA	1	176.4			
	2	42.7	2.84 m	1'	
	3a	43.0	3.07 d,d (18.5,11.1)	2,3b	1,4
	3b	43.0	2.58 d,d (18.1,3.7)	2,3a	1,4
	4	208.2			
	5	58.5	4.43 m	6	4,6,7,1 <sub>ACPC4</sub>

	6a	27.7	1.85 m		
	6b	27.7	1.57 m		
	7a	25.3	1.48 m		
	7b	25.3	1.57 m		
	8	41.7	3.17 d (6.7)	7	6,7,9
	9	158.3			
	1'a	36.0	1.59 m	2,2'	
	1'b	36.0	1.42 m	2,2'	
	2'	21.2	1.38 m	1',3'	
	3'	14.1	0.92 m	2'	1',2'
	1	178.3			
	2	61.4	4.09 d (5.8)	3	1,3,4,5,1 <sub>AGOOA</sub>
<b>Val</b>	3	32.3	2.08 m	2,4,5	
	4	18.4	0.93 m	3	3
	5	19.9	0.93 m	3	3

<sup>a</sup> acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.15 ppm.

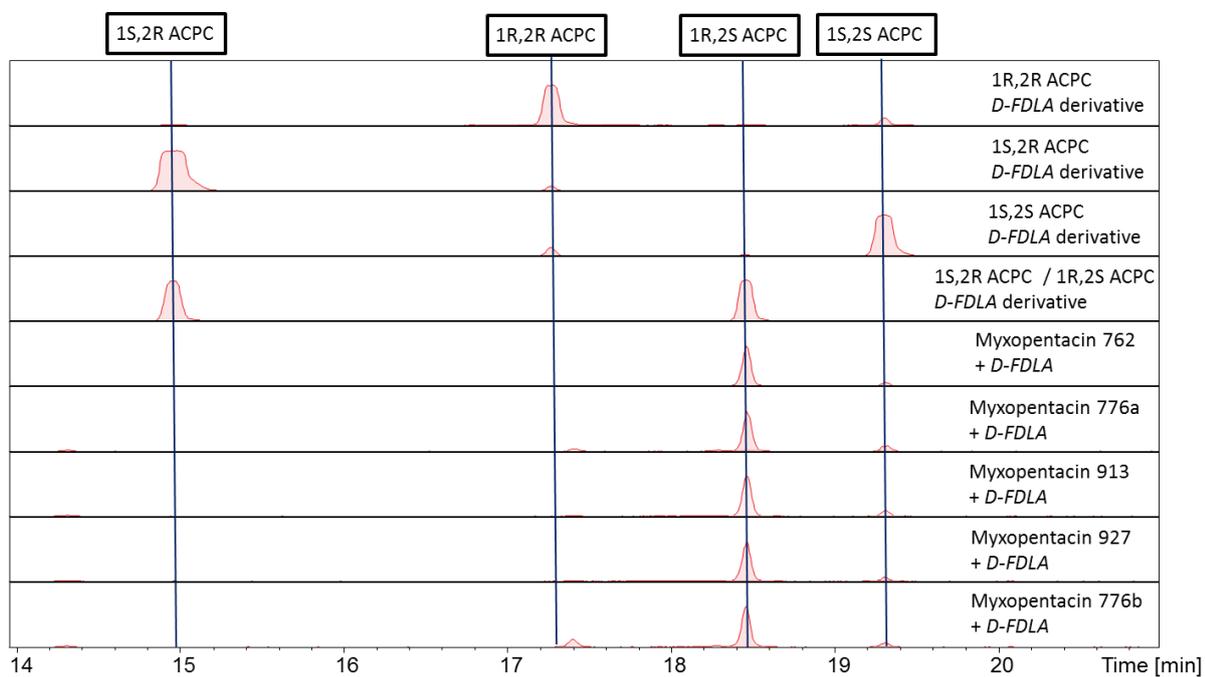
<sup>b</sup> acquired at 700 MHz, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  3.31 ppm.

<sup>c</sup> proton showing COSY correlation to indicated proton.

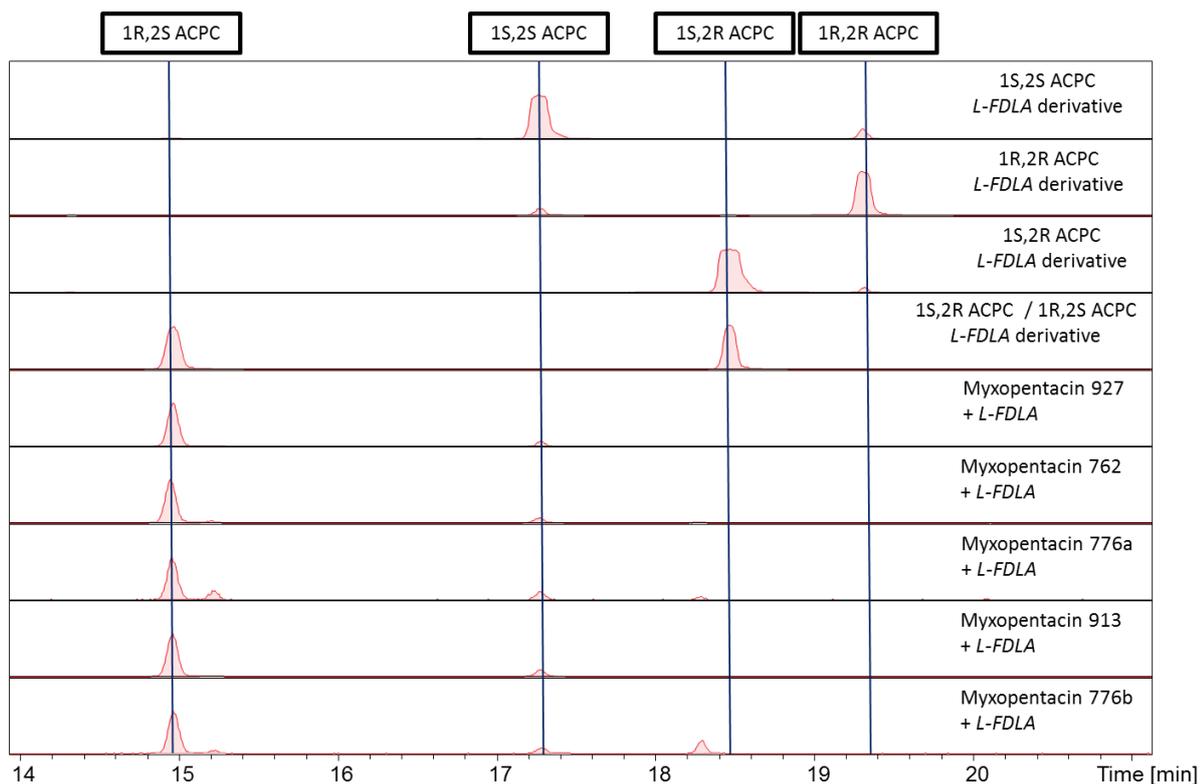
<sup>d</sup> proton showing HMBC correlations to indicated carbons.

## Results Marfey's analysis

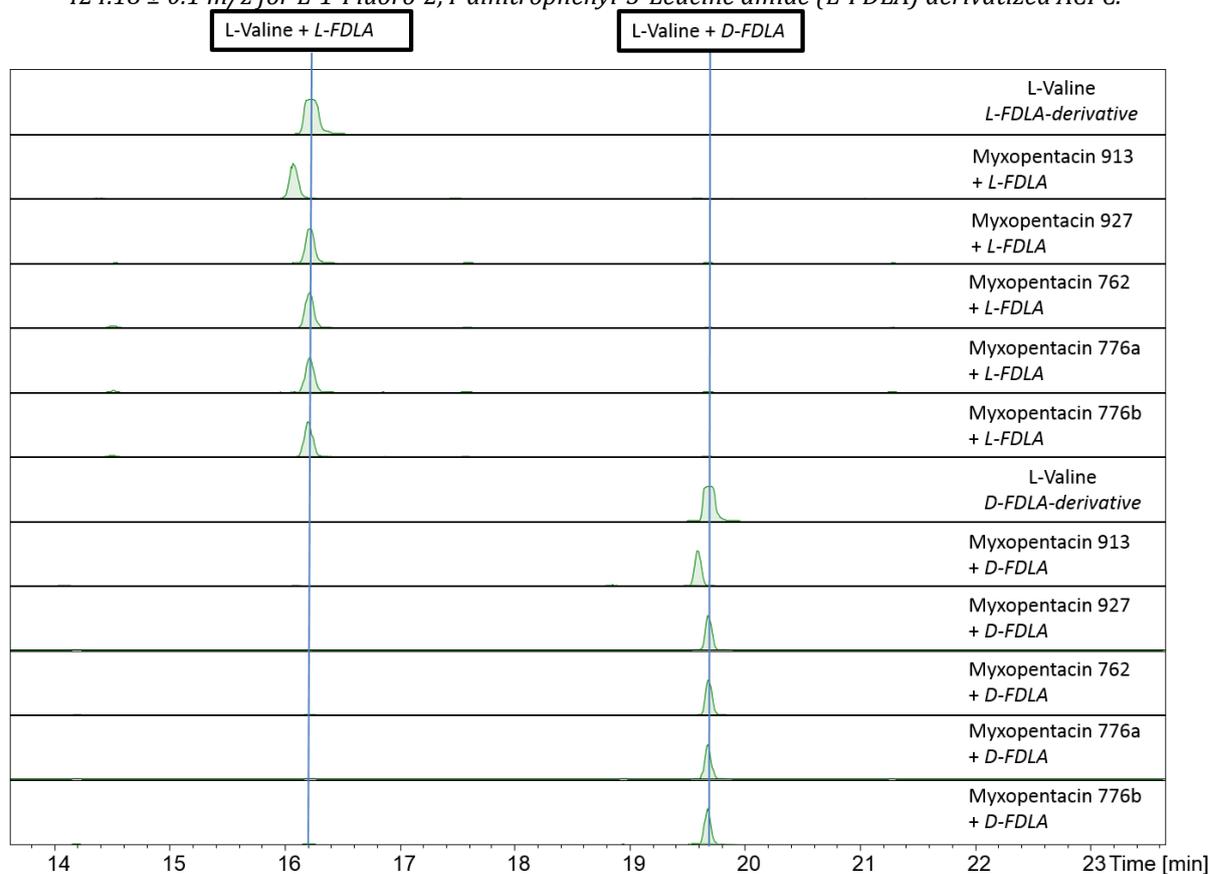
Assignment of the configuration of the amino acids was done using Marfey's method. Hydrolysis mixtures of all isolated Myxopentacin variants were derivatized Marfey's reagent. The assignment of the configuration was conducted by comparison of the retention time with the respective amino acid standart.



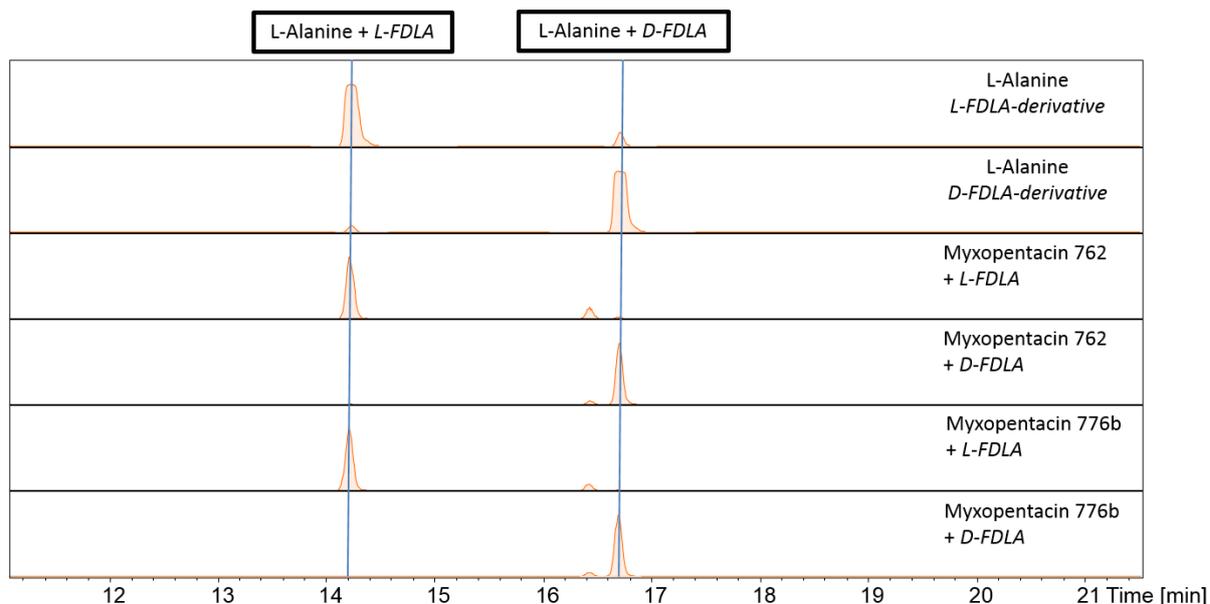
Supplementary figure 43: Results of Marfey's Analysis for 1,2 Amino-cyclopentyl-carboxylic acid (ACPC) standards and five isolated myxopentacin variants after acidic hydrolysis showing the Extracted Ion Chromatogram (EIC)  $424.18 \pm 0.1$  m/z for D-1-Fluoro-2,4-dinitrophenyl-5-Leucine amide (D-FDLA) derivatized ACPC.



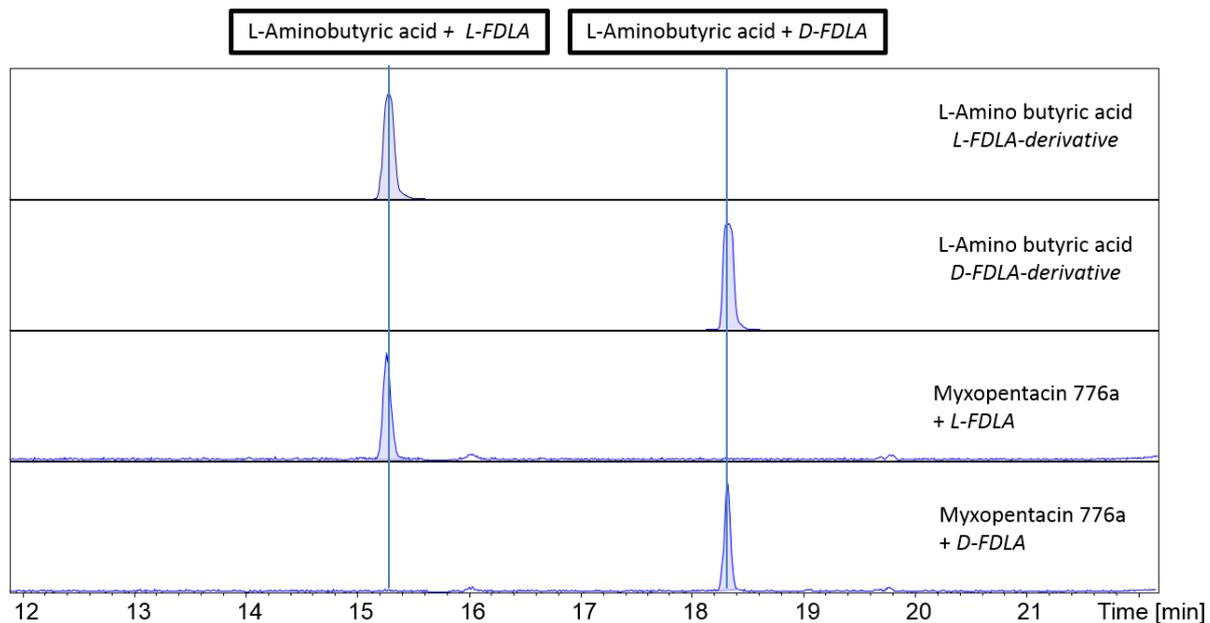
Supplementary figure 44: Results of Marfey's Analysis for 1,2 Amino-cyclopentyl-carboxylic acid (ACPC) standards and five isolated myxopentacin variants after acidic hydrolysis showing the Extracted Ion Chromatogram (EIC)  $424.18 \pm 0.1$  m/z for L-1-Fluoro-2,4-dinitrophenyl-5-Leucine amide (L-FDLA) derivatized ACPC.



Supplementary figure 45: Results of Marfey's Analysis for L-valine standard and five isolated myxopentacin variants after acidic hydrolysis showing the Extracted Ion Chromatogram (EIC)  $412.19 \pm 0.1$  m/z for D/L-1-Fluoro-2,4-dinitrophenyl-5-Leucine amide (D/L-FDLA) derivatized valine.



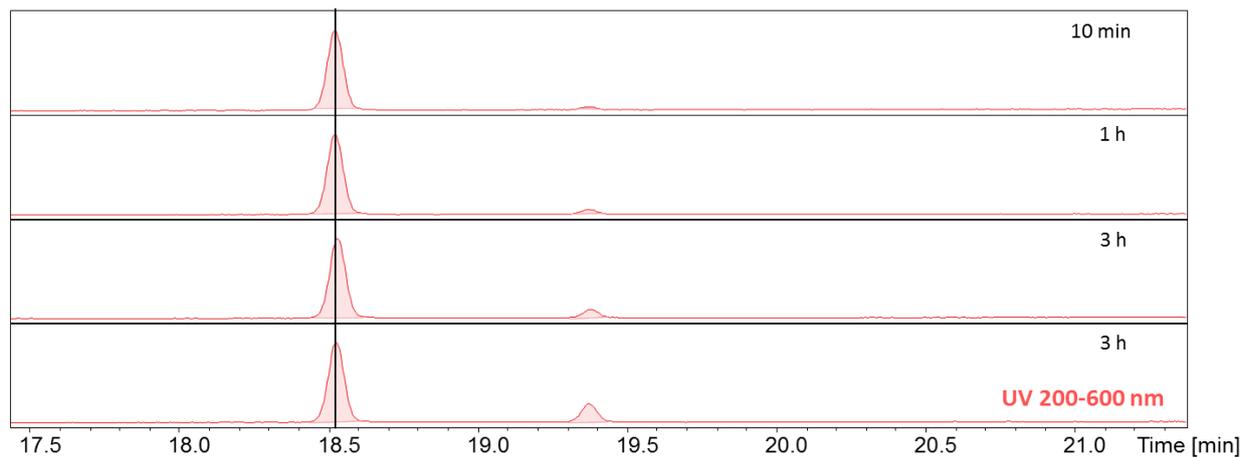
Supplementary figure 46: Results of Marfey's Analysis for L-alanine standard and two isolated myxopentacin variants after acidic hydrolysis showing the Extracted Ion Chromatogram (EIC)  $384.16 \pm 0.1$  m/z for D/L-1-Fluoro-2,4-dinitrophenyl-5-Leucine amide (D/L-FDLA) derivatized alanine.



Supplementary figure 47: Results of Marfey's Analysis for L-amino-butyric acid standard and myxopentacin 776a after acidic hydrolysis showing the Extracted Ion Chromatogram (EIC)  $398.16 \pm 0.1$  m/z for D/L-1-Fluoro-2,4-dinitrophenyl-5-Leucine amide (D/L-FDLA) derivatized amino-butyric acid.

Stability of 1S,2R ACPC as standard in 6 M HCl at 100°C (conditions of acidic hydrolysis in the Marfey's method) was tested. After longer exposure to these conditions, isomerization was observed.

1S,2R ACPC + L-FDLA.

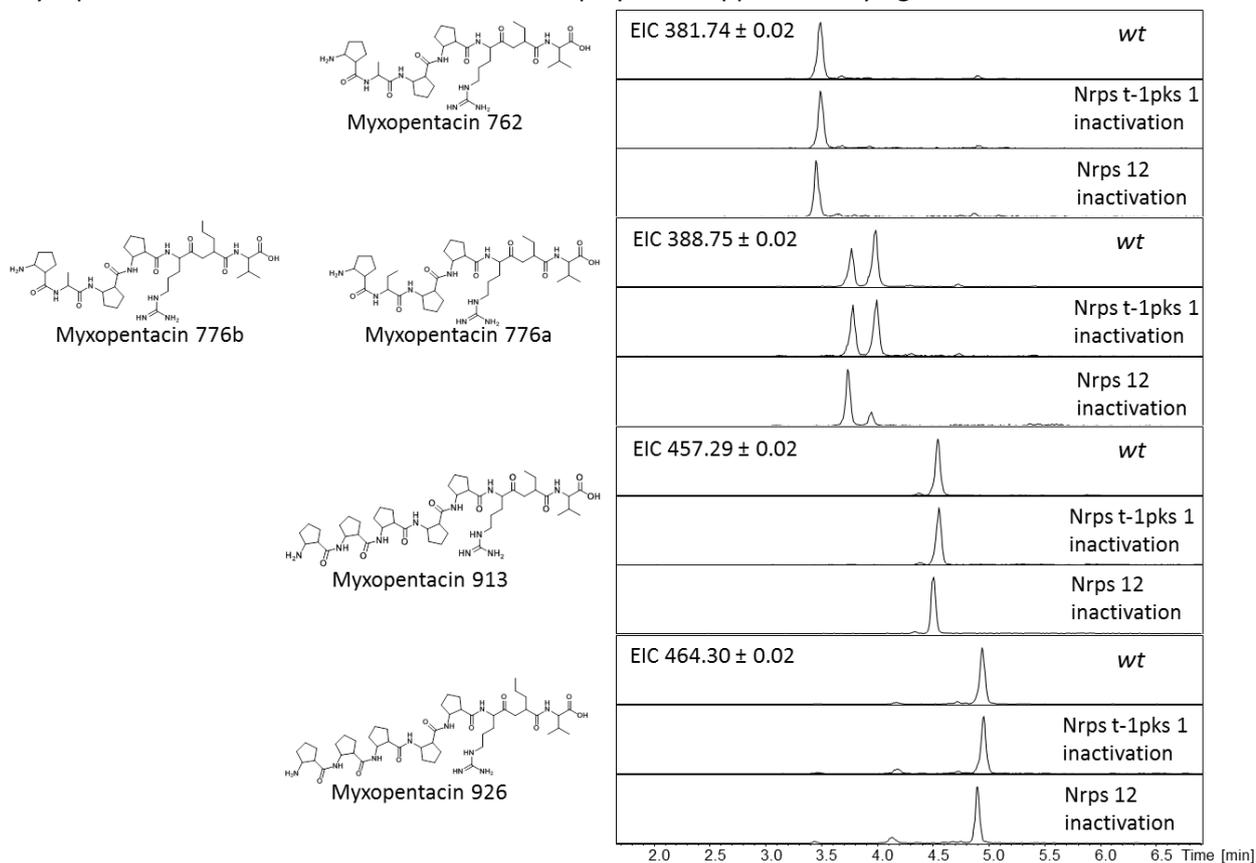


Supplementary figure 48: Acidic hydrolysis of 1S,2R ACPC under acidic conditions. LC-UV chromatogram showing the 1S,2R ACPC peak and a second peak peak at 19.4 min with greater intensity after longer hydrolysis time.

## Biosynthesis

### Results inactivation of putative myxopentacin BGC

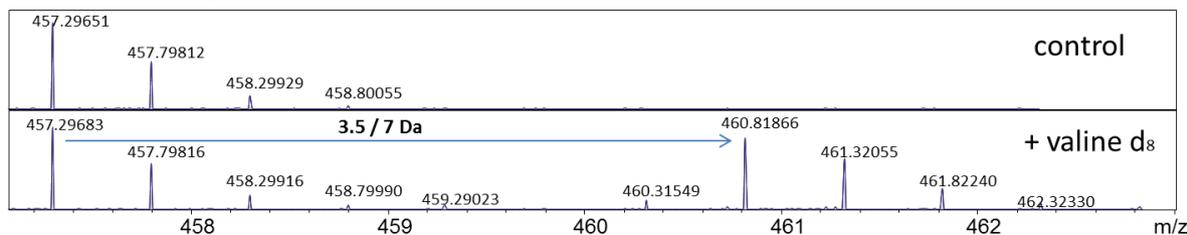
MCy 9003 inactivation mutants of Cluster NRPS-t1pks 1 and NRPS 12 were investigated for their capability to produce the isolated myxopentacin variants. In extracts of both inactivation mutants all isolated Myxopentacin variants could be detected as displayed in supplementary figure 47.



Supplementary figure 49: Results of inactivation of Cluster 1 and Cluster 12.

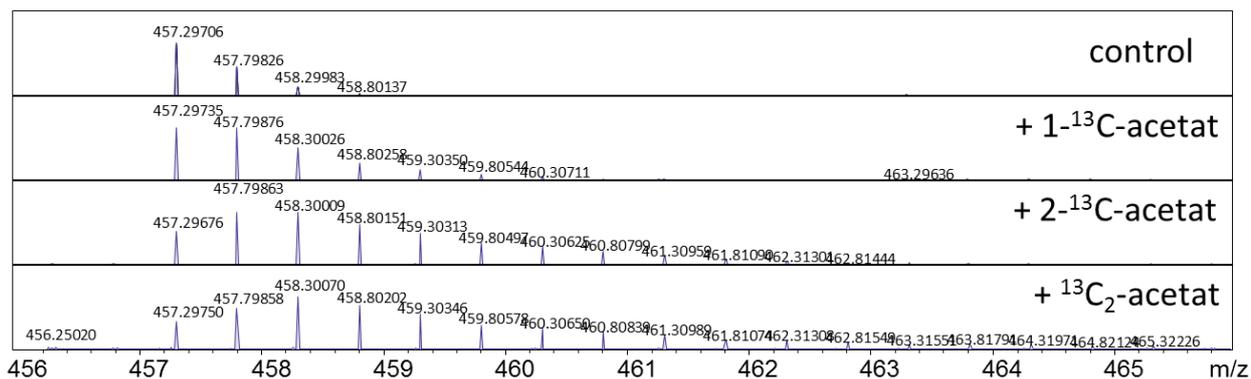
### Results feeding experiments

In order to test incorporation of valine, valine  $d_8$  was added to the culture.

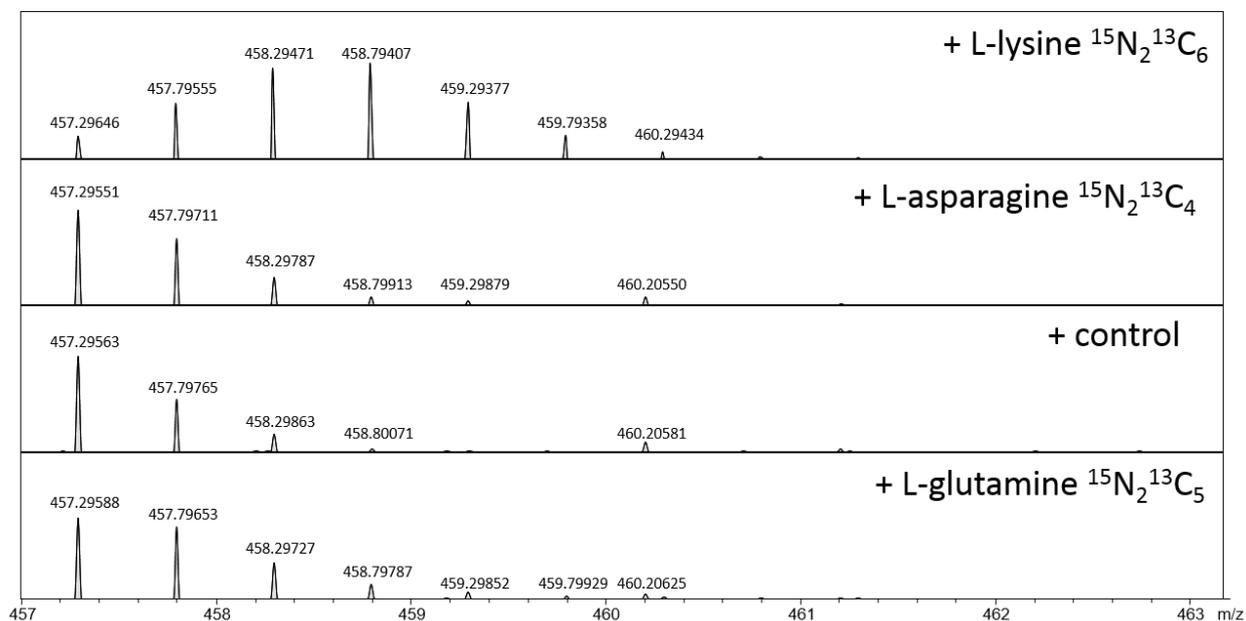


Supplementary figure 50: Isotope pattern of myxopentacin 913 showing double charged mass  $[M+2H]^{2+}$  extracted from LC-hrMS from the control culture and a culture supplemented with valine  $d_8$ .

Incorporation of acetate was investigated by feeding differently labelled acetate to cultures of MCy 9003.

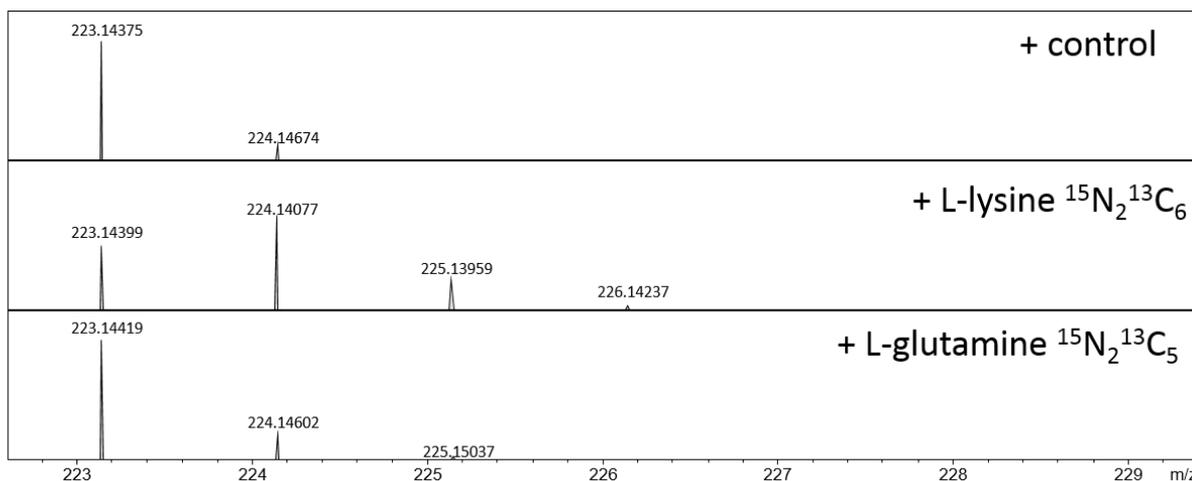


Supplementary figure 51: Isotope pattern of myxopentacin 913 showing double charged mass  $[M+2H]^{2+}$  extracted from LC-hrMS from the control culture and cultures supplemented with 1-<sup>13</sup>C acetate, 2-<sup>13</sup>C acetate and <sup>13</sup>C<sub>2</sub>-acetate.



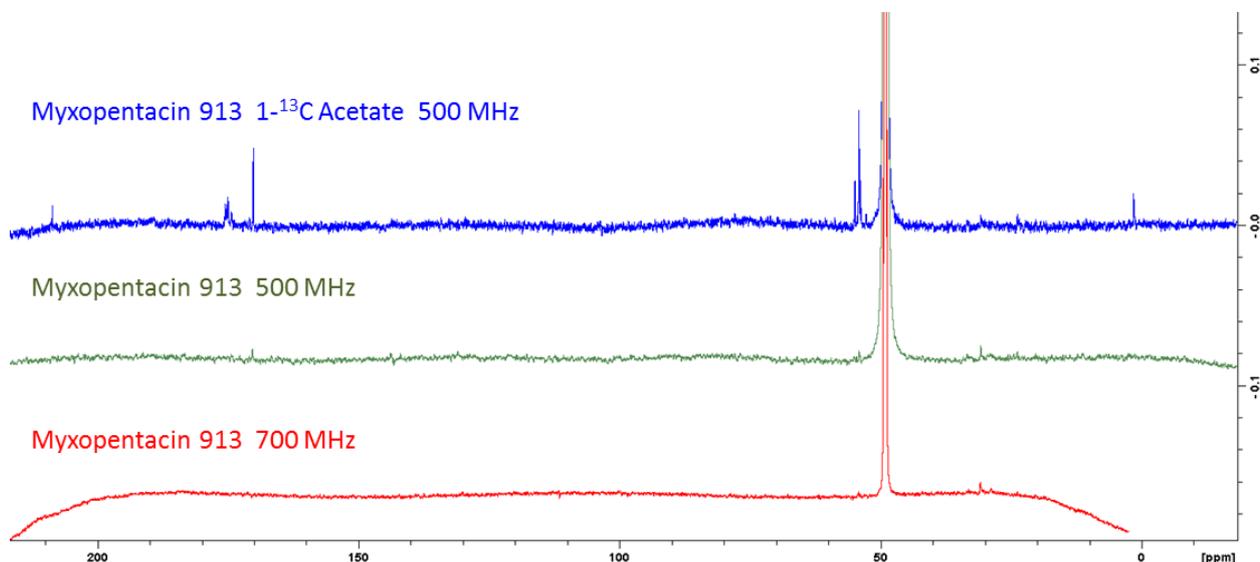
Supplementary figure 52: Isotope pattern of myxopentacin 913 showing double charged mass  $[M+2H]^{2+}$  extracted from LC-hrMS from the control culture and cultures supplemented with <sup>15</sup>N<sub>2</sub><sup>13</sup>C<sub>6</sub> lysine, <sup>15</sup>N<sub>2</sub><sup>13</sup>C<sub>4</sub> asparagine and <sup>15</sup>N<sub>2</sub><sup>13</sup>C<sub>5</sub> glutamine.

Incorporation into ACPC moiety was closer investigated by MS<sup>2</sup> fragmentation. Isotope pattern of fragment 223.14 m/z which is a dimer of two ACPC units was closer observed.



Supplementary figure 53: Isotope pattern of Fragment 223 after MS<sup>2</sup> fragmentation of control culture and cultures supplemented with <sup>15</sup>N<sub>2</sub><sup>13</sup>C<sub>6</sub> lysine and <sup>15</sup>N<sub>2</sub><sup>13</sup>C<sub>5</sub> glutamine. MS<sup>2</sup> spectra are extracted from LC-hrMS measurements in autoMS<sup>2</sup> mode.

In order to further investigate myxopentacin biosynthesis, myxopentacin 913 was isolated from a cultivation batch supplemented with 1-<sup>13</sup>C acetate. To determine the acetate incorporation pattern in the 1-<sup>13</sup>C acetate labelled Myxopentacin 913, a <sup>13</sup>C NMR dataset was acquired using a Bruker 500 Neo NMR spectrometer. For comparison, unlabelled myxopentacin 913 was measured on the same device and on a Bruker Ascend 700 NMR spectrometer. Spectra are displayed in supplementary figure 54. Assignment of <sup>13</sup>C signals is listed in supplementary table 10.



Supplementary figure 54: <sup>13</sup>C spectrum of myxopentacin 913 isolated from a culture supplemented with 1-<sup>13</sup>C labelled acetate acquired at 500 MHz (blue); <sup>13</sup>C spectrum of myxopentacin 913 acquired at 500 MHz and <sup>13</sup>C spectrum of myxopentacin 913 acquired at 700 MHz. All spectra were acquired in CD<sub>3</sub>OD.

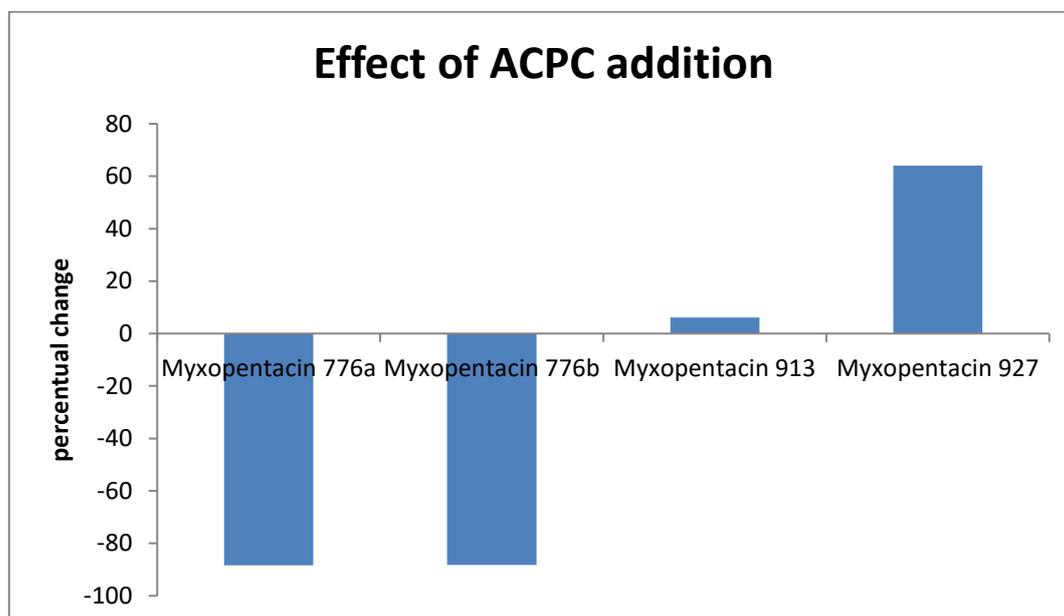
Supplementary table 10:  $^{13}\text{C}$  signals derived of 1- $^{13}\text{C}$  acetate labeled Myxopentacin 913

$^{13}\text{C}$ signal <sup>1</sup>	Assigned position
208.6	C-2 <sub>AG00A</sub>
175.4	C-1 <sub>ACPC 5</sub>
175.2	C-1 <sub>ACPC 4</sub>
175.0	C-1 <sub>ACPC 3</sub> OR C-1 <sub>ACPC 2</sub>
175.0	C-1 <sub>ACPC 3</sub> OR C-1 <sub>ACPC 2</sub>
174.3	C-1 <sub>ACPC 1</sub>
54.8	C-6 <sub>ACPC 1</sub>
54.1	C-6 <sub>ACPC 4</sub> OR C-6 <sub>ACPC 4</sub>
54.1	C-6 <sub>ACPC 4</sub> OR C-6 <sub>ACPC 4</sub>
53.8	C-6 <sub>ACPC 3</sub>
53.6	C-6 <sub>ACPC-2</sub>

<sup>1</sup> acquired at 125 MHz in CD<sub>3</sub>OD, referenced to solvent signal CD<sub>3</sub>OD at  $\delta$  49.00 ppm

### Influence ACPC addition to the culture

To test effects of ACPC on myxopentacin production, MCy 9003 tn5 NRPS12 mutant was cultivated in 100 ml CyS media. 100  $\mu\text{l}$  of a 0.4 M solution of a racemic mixture of (1R,2S) and (1S,2R) ACPC were added to the cultures. The cultures were cultivated for 12 days. After cultivation, the cultures were extracted and the extracts were measured using LC-*hr*MS. The peaks of myxopentacin variants 776a, 776b, 913 and 926 were integrated based on their corresponding mass trace. The Peak areas were compared and the percentual change compared to the control was calculated.



Supplementary figure 55: Percentual change of production of myxopentacin variants after addition of (1R,2S/1S,2R) ACPC to a culture of MCy 9003 mutant tn5 NRPS12

### New putative myxopentacin variants

New myxopentacin variants were identified by a statistical analysis using MCy 9003 WT and myxopentacin inactivation mutant using Bruker metaboscope. Buckets not present in the inactivation mutant were identified; a scheduled precursor (SPL) list was created using these buckets to acquire MS<sup>2</sup> spectra of all buckets by one LC-*hr*MS measurement. The MS<sup>2</sup> spectra were analysed using GNPS Clustering tool. Two big clusters

were created. In Cluster one precursor ions of the isolated myxopentacin variants were contained. Therefore all other precursor ions were considered as putative myxopentacin variants. MS<sup>2</sup> spectra of the precursor ions listed in cluster 2 also showed characteristic features of myxopentacin variants. All listed precursor ions of cluster 1 were also considered as putative myxopentacin variants. Precursor ions and respective retention times referred to maxis4G analytical system using an 18 minute gradient are listed in supplementary table 11.

Supplementary table 11: List of precursor ions of MS<sup>2</sup> fragmentation spectra that form a Cluster. Precursor ions of characterized myxopentacin (marked in red) variants were identified based on the hrMass and the retention.

Cluster 1		Cluster 2	
precursor ion mass	rt [min]	precursor ion mass	rt [min]
290.70	3.80	581.34	5.86
290.70	2.85	587.35	4.74
297.70	3.21	592.38	5.04
297.71	3.38	597.37	7.07
304.71	3.85	601.36	5.29
321.70	4.14	613.36	5.61
327.70	4.56	632.41	5.50
329.70	3.22	634.42	6.36
332.22	2.97	647.42	4.80
339.22	3.84	648.44	6.71
339.22	3.38	660.44	6.08
343.74	3.34	662.45	6.71
346.23	4.12	677.38	5.42
346.33	3.28	682.42	6.98
353.24	3.94	684.40	5.23
361.24	3.70	692.41	6.27
367.73	2.58	696.44	6.82
374.74	4.52	698.42	5.56
374.74	3.17	708.44	7.27
374.74	3.16	710.42	6.33
377.24	4.65	712.43	5.69
381.74	4.91	719.48	6.05
<b>381.75</b>	<b>3.69</b>	733.49	6.39
381.77	2.98	745.49	6.87
383.24	5.12	767.48	6.85
385.24	3.78	771.51	6.49
<b>388.76</b>	<b>3.96</b>	785.52	7.68
388.84	3.33	793.49	7.29
393.35	6.00	807.51	7.54
395.77	4.24	809.49	6.23
398.75	5.21		
402.24	6.65		
411.14	6.75		
412.76	4.62		
414.77	4.47		
415.78	4.83		
418.25	6.40		
419.77	4.86		
420.77	5.70		
425.26	6.93		
427.78	6.29		
430.27	4.90		
437.53	4.22		
440.56	5.42		
441.79	6.79		
443.26	5.79		
444.30	3.99		
446.96	6.83		
449.10	6.64		
450.29	5.52		
451.30	4.54		
455.86	6.03		
<b>457.30</b>	<b>4.53</b>		
458.78	6.81		
460.78	6.01		
463.29	4.88		
<b>464.31</b>	<b>4.93</b>		
466.49	7.15		
468.29	4.98		
472.11	6.07		
474.29	5.06		
475.30	5.16		
480.80	7.19		
483.31	6.57		
488.31	5.47		
496.07	5.34		
498.81	5.55		
500.32	6.67		
504.71	6.31		
505.82	5.97		
509.31	6.59		
510.81	5.79		

511.77	7.12			
513.31	7.24			
514.32	7.64			
516.32	6.35			
519.84	5.36			
522.26	7.66			
525.62	5.07			
528.33	8.28			
532.32	7.84			
543.84	5.89			
554.35	5.83			
555.84	6.89			
559.87	7.82			
561.36	6.20			
568.37	5.28			
571.72	7.26			
574.87	5.00			
577.57	7.38			
599.37	5.69			
607.38	5.41			
623.90	5.67			

## References

- [1] a) S. Hüttel, G. Testolin, J. Herrmann, T. Planke, F. Gille, M. Moreno, M. Stadler, M. Brönstrup, A. Kirschning, R. Müller, *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 12760; b) J. Herrmann, S. Hüttel, R. Müller, *ChemBioChem* **2013**, *14*, 1573.

## 5. Discussion

In this work, three different compound classes were discovered using two different mining strategies. Each compound family has unique structural features to establish them as noncanonical natural products. Interestingly each compound class was discovered in a different way, emphasizing that there is no “ultimate strategy” to natural products discovery. Moreover, in the case of the myxoquaterines and the myxopentacin compound class, the outcome was different from what was expected at the beginning of the discovery process. The structural features of the discovered molecules turned the elucidation of the biosynthesis into a challenging task. Only a few literature examples were found that described similar scaffolds and could be used as reference. The initial goal of this work was the identification of novel scaffolds that might even be useful as leads for drug development. Considering the question if this goal was achieved, the answer is clearly yes. For the myxoquaterines even drug development can be envisaged, since their cytotoxicity marks them as promising scaffolds for future anti-cancer drugs. Regarding the myxopentacin family and the myxoglucamides, new scaffolds were obtained. The moderate cytotoxic activity does not designate these molecules as anti-cancer leads, but it cannot be excluded, that both compound classes have more specific bioactivity that was not discovered so far. In addition, the unique structural features might serve as an inspiration for chemists synthesizing novel bioactive compounds.

### ***5.1 Evaluation of activity-guided compound isolation vs. genome mining***

Many publications in recent time, consider genome mining as the best method to find new natural products and proclaim that classical cultivation methods are exhausted<sup>[1]</sup>. Considering the results summarized in this work, it should better be stated that genome mining is an alternative to the classical cultivation with activity guided compound isolation. Especially when working with underexplored bacterial species or a new and taxonomically significantly different strain, the classical approach might be the better choice under the premise the organism shows reasonable production titers of compounds. In the following, the reasons that lead to this conclusion will be explained:

The main argument against the classical cultivation approaches is that they are often associated with the rediscovery of natural products. Of course, rediscovery is a serious problem, but there are options to avoid this. A proper dereplication is crucial. But not using the same type of strains and switching to more taxonomic distant strains instead is the most valid option. For example, recent studies showed a correlation between taxonomic and metabolic diversity. Certain groups of natural products are preferably found only in certain genus types<sup>[2]</sup>. Therefore, working with novel strains that are taxonomically different from what was described so far offers a higher chance of finding novel compound classes.

The discovery of the presented myxoquaterines is a good example. The producer strain MSr 11954 belongs to novel genus of the myxobacterial *Sorangineae* suborder. So far, the myxoquaterines were not identified in other types of myxobacteria or even other types of bacteria. Even though, working with a novel, yet not explored genus, is as well no protection from rediscovery of known compounds. Metabolic profiles are still overlapping to certain extent with profiles with other myxobacterial genera and metabolites from other types of bacteria can also be found, although with a low percentage. Consequently, dereplication is still recommended. For example in MSr 11954 not all produced metabolites were unknown. By using a LC-*hr*MS based dereplication, the antifungal aminopyrrolnitrin firstly reported from *Pseudomonas* species was identified in the organic extract of MSr 11954<sup>[3]</sup>.

A big effort of the classical cultivation in combination with the activity guided compound isolation is its simplicity. Cultivation of a strain requires usually between a few days and few weeks. A standard fractionation to identify bioactive needs a few additional days including the biotesting of the fractions. In contrast, the required genetic manipulations of a strain for genome mining approaches are way more labor- and time-intensive. Like with every strategy, there are certain drawbacks. Typical drawbacks of the classical cultivation in combination with the activity guided compound isolation are:

1. Especially when working with novel strains, poor knowledge about optimal cultivation conditions can lead to poor growth and low production titers. Also, the metabolome of the investigated strain can change under laboratory conditions in a way that initial production profiles are not reproducible<sup>[4]</sup>. For example, while working with MCy 9003 an initial anti gram-negative bioactivity of a crude organic extract was not reproducible (unpublished results).
2. In practice, it is often observed that molecules identified by the standard fractionation only showed weak or no bioactivity against the indicator strain when isolated. The same observation was made with the myxoquaterines. Fractions from the standard fractionation that contained the myxoquaterines showed activity against the indicator strain *M. hiemalis*. However, the bioactivity of the isolated myxoquaterines against *M. hiemalis* was rather weak (MIC: 64 µg/ml). The reasons for this can be diverse. In the standard fractionation workflow, the concentration of a compound in a fraction is unknown. A highly concentrated compound will result in an active fraction, even though the actual MIC is high. Regarding the myxoquaterines, the LC-*hr*MS chromatogram shows the myxoquaterines to be the most intense peaks, suggesting also high myxoquaterine concentrations in the collected fractions. Other reasons for the lack of bioactivity of isolated compounds can be synergistic effects of additional molecules being present in the collected fraction or that simply the wrong molecule was identified.
3. Another drawback of activity guided compound isolation is that the focus is only set on one molecule or a group of molecules with certain bioactivity. Other molecules with no activity or other bioactivities that are not covered in the screening and are likely to fall through the cracks. Metabolic approaches focusing more on structural novelties may also consider these molecules.<sup>[5]</sup>

In contrast to the classical cultivation with an activity guided compound isolation, the genome mining approach is much more complex. This approach requires the bacterial genome which is analysed using bioinformatics tools like antiSMASH<sup>[6]</sup>. Usually the output of these analyses indicates a way higher number of potential natural products, based on the number of identified biosynthetic gene clusters (BGCs), than actually observed in the metabolome. Therefore, in the context of genome mining, it is often emphasized that the observable metabolome of a microorganism at least under laboratory conditions is only the tip of the iceberg. Taking the myxobacterial model strain DK1622 as an example, 9 natural products were found being produced by this strain while the genome analysis yielded 18 BGCs<sup>[7]</sup>. The antiSMASH development team tried to estimate the joint biosynthetic potential of thousands of strains: 152,106 BGCs out of 24,776 full and draft genomes were identified using the antiSMASH tool<sup>[8]</sup>. A number over 150 000 is enormous and leads to hopes to find the same number of new natural products. However, can more than 150 thousand new natural products be expected? The answer is clearly no. To explain this answer, a few aspects need to be considered regarding how this number is generated and how the natural product connected with the BGC can be accessed.

The BGCs were identified via a computer algorithm based on sequence homologies to existing BGCs. This approach is associated with two main problems that influence the type and actual number of expectable natural products. The first problem is the missing experimental proof that the BGC can be functional. It cannot be excluded that the BGC is simply not functional due to mutations or because it is in an evolutionary state. Therefore, out of these 150 000 predicted BGCs, some will most likely not be able to yield a natural product. The second problem and often-mentioned drawback of genome mining approaches is that bioinformatics algorithms can only detect sequences of known biosynthetic genes and therefore actual metabolites of predicted BGCs will not have any new structural features. This statement has legitimate reasons, but in practice, the unknown plays a smaller role for recognition than expected. Most BGCs based on their type follow a stringent logic. Unknown genes inside BGCs are usually outnumbered by known genes, which enables detection by toolkits like antiSMASH based on homology and the organization within BGC. Therefore, the myxoquaterine and the myxoglucamide BGC were annotated by antiSMASH despite some unknown biosynthetic genes. However, not every biosynthesis pathway will follow such NRPS collinearity principle/rule which will cause difficulty to correlate the BGC with the compounds. For example, the myxopentacin BGC, which mostly consists of biosynthetic genes for its building blocks and one small ORF encoding a small NRPS (MxpC), which can conduct two elongation steps. Here, core biosynthetic genes were annotated but due to lack of stringent logic, only MxpC was detected as NRPS module but assigned to be part of the neighboring BGC. In addition, when core biosynthetic genes are distributed over the whole genome, a detection is not possible by conventional genome mining algorithms. To conclude this issue, novelty of a BGC is a smaller problem for the detection algorithms, but it is strongly implied that there are some BGCs in addition to the 150 000 predicted ones.

Beside the genome mining algorithms, another factor influences the number of expectable novel natural products: Many strains share BGCs with high identity, which reduces the actual number of different BGC and therefore the number of different natural products to a certain extent. The fundamental question is which minimum percentage of amino acid sequence identity of two given BGCs is necessary to produce the same natural product. Conversely, the question arises how to determine the sequence identity threshold of two given BGCs, where both biosynthetic pathways start to generate different natural products. A good example to consider is the discovery of the myxoquaterines. The two myxoquaterin BGCs in MSr 11954 and MSr 11367 show despite a different organization a high degree of identity and produce both myxoquaterine variants 450 and 535. Nevertheless, the one additional prenyltransferase Mxql in the MSr 11954 BGC leading to production of myxoquaterin 669 and 671, which are practically different compounds. On the other hand, the geosmin BGC which can be found in different species<sup>[9]</sup> yielded only geosmin in MCy 9003 as confirmed by LC-hrMS. Therefore a clear answer cannot be given but underlines the importance of dereplication since working on BGCs with great homologies to published or known BGC increases the risk of rediscovery like for examples with the alkylpyrones which were discovered in myxobacteria after being previously described in *Streptomyces* sp.<sup>[10]</sup>. Additionally, the number of already discovered microbial natural products is estimated to be around 70.000<sup>[11]</sup>. For many of these natural products, no corresponding BGCs are described which will definitely result in rediscoveries when a genome mining approach is conducted.

The biggest limitation of genome mining is the connection between the genome and the metabolome. The correlation of a predicted BGC to an actual metabolite typically requires either a genetic manipulation of the strain or heterologous expression. Both techniques are extremely labor-intensive and there is no warranty of positive output. For example during the genome mining approach in MCy 9003, promoter activations were done with 10 BGC, but only three mutants showed a positive result. This success rate of 30 % can definitely not be generalized but reflects the drawbacks of the genome mining based Cluster activation approach quite

well. In addition, heterologous expression is often associated with an unsuccessful result. Reasons for these failures can be explained as follows:

1. The regulation of BGC expression is most likely a more complex procedure. Transcription of a BGC does not mean there will be a metabolite. For example in the myxobacterial model strain DK1622 mRNA of 12 BGC were found to be present while only 9 natural products were identified<sup>[12]</sup>, which implies additional restrictions.
2. Some BGCs consist of several operons. Therefore, a simple promoter insertion is not sufficient and overexpression of these BGCs require complex manipulation strategies.
3. Heterologous expression especially when a bacterial strain cannot be genetically manipulated or when working in a culture-independent approach is a promising approach<sup>[13]</sup>. Drawbacks also exist here. The success can be limited by the selection of the host and a non-functional cluster will not give a metabolite in a heterologous host as well. Availability of a required precursor in the heterologous host is also a limiting factor. In addition, if a natural product is toxic for its heterologous host, this approach is also bound to fail.

For a final evaluation of the discovery strategies, it is the question of whether the other method (genome mining or classical cultivation with activity-guided isolation) would have led to the discovery of the compounds under the same circumstances. For the myxoquaterines the answer is clearly no. So far, it was not possible to genetically manipulate the strain MSr 11954 due to multi-resistance against every antibiotic used as selection markers. In addition, heterologous expression of the myxoquaterin BGC did not succeed putatively due to toxicity towards the heterologous host (personal communication Dr. Yunkun Liu). Considering the myxoglucamides, only the genome mining approach would have led to the discovery since no production was observed in WT. For the myxopentacin molecules, the discovery was an unexpected side effect of the BGC activations. An activity-guided isolation would not have yielded in the discovery due to missing activity against commonly used indicator strains. But using metabolic approaches like for example used in the discovery of the fulvuthiazene<sup>[5]</sup> molecules would sooner or later have led to the discovery. Since the myxopentacin BGC was not detected as such, a genome mining approach for example by a cluster activation would not have led to discovery of the compound family. It needs to be mentioned that compounds derived from the genome mining approach did not showed strong bioactivity in contrast to the myxoquaterines from the activity-guided isolation.

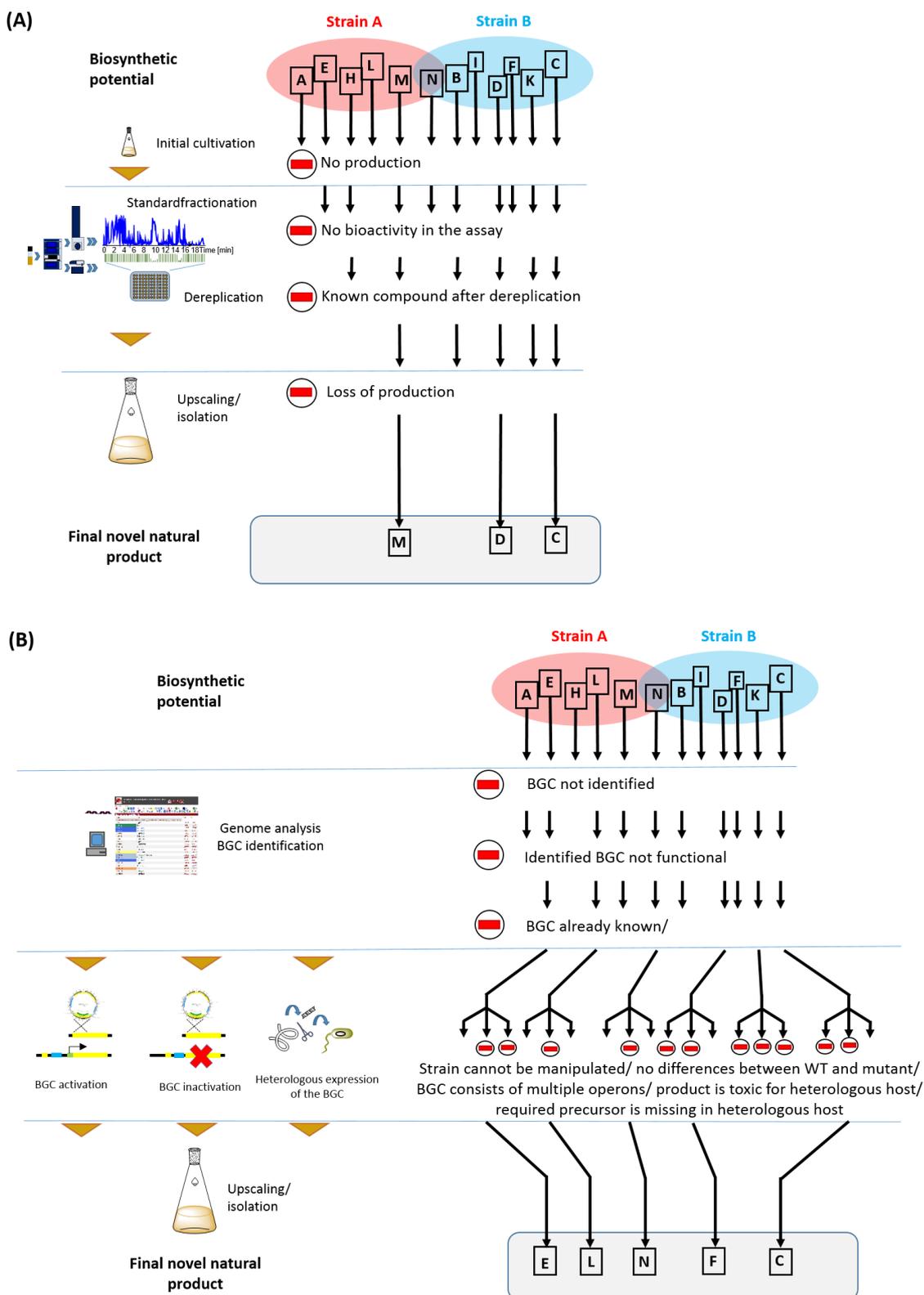


Figure 1: Comparison of natural product discovery strategies. (A) Activity-guided compound isolation with typical workflow and limitations reducing the biosynthetic potential to the final output. (B) Genome-mining based discovery strategy, showing the general workflow and typical drawback and limitations.

To summarize, the success of a strategy is limited by the respective circumstances. Since also in natural product biology circumstances are different not one single strategy will work for everything. The art lies in recognition of the given circumstances and the choice of the right strategy.

## 5.2 Considerations regarding Biosynthesis

### 5.2.1 Elucidation of the biosynthesis – many questions – few answers

With more and more available genomic information and easy accessible bioinformatics tools like antiSMASH, the development of a biosynthesis hypothesis for a newly discovered natural product is becoming more and more convenient. The process usually begins with the identification of the corresponding BGC. In a second step, the hypothesis is refined, which involves the linkage of every catalytic step to a certain catalytic domain. In the past, the identification of a BGC was a laborious and time-consuming task, often involving the creation of thousands of transposon clones or generation of large cosmid libraries. But in the last ten years, such approaches were more and more replaced by *in-silico* approaches. Tools like antiSMASH can annotate most of the BGCs in a bacterial genome. In addition, the catalytic domains within the BGC are annotated. In order to identify the right BGC of a novel natural product, a retro-biosynthetic analysis is the best strategy. In a retro biosynthetic analysis, it is concluded, from which building blocks a natural product is made of and which catalytic reactions are necessary in which sequential order, giving an expected BGC. The expected BGC is compared with every annotated BGC from the bioinformatics analysis. If the expected BGC shows great similarities with one annotated BGC, this BGC can be considered as a candidate cluster for further analysis.<sup>[6,14]</sup> Discrepancies between expected BGC and annotated BGC like skipped modules or iteratively acting modules, can complicate the identification of correctly annotated BGC<sup>[15,16]</sup>.

A newer feature of antiSMASH is a direct reference of a published BGC with the greatest homology. Using this feature a retro-biosynthetic analysis of the myxoquaterines was not necessary, since one annotated BGC in MSr 11954 showed 35 % identity with the published DKxanthen BGC<sup>[16,17]</sup>. Due to the structural similarity of the myxoquaterines to the DKxanthenes and since no other annotated cluster showed such a high similarity, this annotated cluster was considered as a candidate cluster.

For the myxoglucamides, identification of the corresponding BGC was obsolete, since a genetic manipulation of the BGC led to the discovery of this compound class. In addition, a retro-biosynthetic analysis was in agreement with the BGC.

Regarding the myxopentacin BGC, the identification was a more difficult task. Retro-biosynthetic analysis suggested an NRPS type assembly line with at least five NRPS modules with the same amino acid specificity, one NRPS module with arginine specificity followed by a PKS module and a C-terminal valine were expected. The additional existence of myxopentacin derivatives with alanine or amino butyric acid as second amino acid were slightly in contrast to this expected assembly line. One single NRPS module with a very broad substrate specificity was regarded unlikely. A second assembly line with an alanine specific module and a potential crosstalk between these two assembly lines could make this hypothesis still valid<sup>[18]</sup>. Nonetheless, such assembly lines were not found among the annotated BGCs.

Finally, search for similar structural scaffolds in described natural products and their biosynthetic genes resulted in the identification of the myxopentacin BGC. The identified cluster contained only one NRPS type gene (*mxgC*), that encodes for two NRPS modules. This significant difference explains why the retro-biosynthetic approach did not succeed.

The refinement process of the biosynthesis hypotheses involves usually the assignment of each catalytic domain to a certain catalytic step in the biosynthesis. In the case of a biosynthetic pathway, that follows strictly the well-described textbook logic of NRPS or PKS assembly lines, this process is straightforward. But unfortu-

nately or better said luckily, nature does not always follow textbook logic. In these cases, the domain functions, which were automatically assigned by the antiSMASH tool, need to be considered with care. Other more precise structure-based search tools like Pfam<sup>[19]</sup> or HHpred<sup>[20]</sup> can provide a more accurate prediction of the function. Apart from starting with the domain function to explain a biosynthetic step, also the final structure itself can lead to the biosynthesis. This analysis principle can be summarized by the German proverb “*Der Apfel fällt nicht weit vom Stamm*”. Similar structural features in natural products often originate from similar biosynthetic pathways. For an analysis here, it is crucial to have cases of other natural products with similar scaffolds and a well-described biosynthesis. The deduced gene products which contain domains share homology to such that were described to catalyze the biosynthetic steps of interest can be used for the search of gene homologues in the actual BGC. If homologues are found it can be assumed, that biosynthesis takes place in a similar manner.

To support results from the *in-silico* analysis, feeding experiments using stable isotope labelled precursors like amino acids or acetate are a valuable tool. As the outcome of these experiments, it can be said if a certain precursor is incorporated into a molecule and utilized as a building block and therefore if a hypothesis is valid or not. In addition, certain mass shifts in the isotopic pattern of a specific natural product as results from feeding experiments can give insights into the biosynthetic mechanisms, like for example decarboxylation steps. By a detailed bioinformatic analysis, using the described steps and feeding experiments a biosynthesis hypothesis was developed for every described compound class. Since all three compound classes showed major deviations from textbook logic, the task was not straightforward.

Regarding the myxoquaterine biosynthesis, the biggest difficulty was to explain the chain release. The missing TE domain and required reductive steps to form the precursor myxoquaterine 450. The first idea was a reductive release like found in myxochelin or myxalamid A or saframycin biosynthesis<sup>[21]</sup>. However, no enzymatic homologues were found and the catalytic domains present in the myxoquaterine BGC were not sufficient to explain the formation of myxoquaterine 450. A valid explanation was found with the described biosynthetic pathways of fumonisins and prodigiosin<sup>[22]</sup>. One domain in MxqO, which catalyzes the last polyketide elongation step, contains a PLP binding site. Therefore, it was annotated as aminotransferase domain by antiSMASH, but is rather a PLP dependent amino acid C-acyltransferase domain and catalyzes the chain release by addition of serine. Another difficulty was the assignment of the prenyltransferase. Initially, the prenyltransferase was expected to be located outside the BGC, since this was reported for other terpene- PKS-NRPS hybrid molecules<sup>[23]</sup> and no domain inside the BGC was annotated as prenyltransferase by antiSMASH. But analysis of domains, that had no homologues from the DKxanthen biosynthesis, using the HHpred tool showed that one catalytic domain (MxqI) has great structural homology to aromatic prenyltransferase domains. An experimental prove can be achieved by heterologous production and isolation of recombinant MxqI for an *in vitro* reconstitution of this reaction using unprenylated myxoquaterine 535 or myxoquaterine 450 as substrates. In addition, in case of a confirmed prenylation reaction, it would be interesting to see if MxqI is able to prenylate molecules from the DKxanthen compound family since they have an identical prenylation site.

Development of a biosynthetic hypothesis for the myxoglucamides was associated with other difficulties. The biosynthesis is initiated by loading of an iso-fatty acid chain from primary metabolism onto the ACP domain in MxgI. In a further step, the fatty acid chain is hydroxylated in  $\omega$ -position. So far, the enzyme conducting the hydroxylation could not be determined. Potential candidates are the KR/aldoreductase domain in MxgF, a cytochrome p450 (MxgG), a monooxygenase domain in MxgH and a Condensation/TauD like domain<sup>[24]</sup> in MxgI. Search for molecules with similar scaffolds yielded the ieodoglucamides where neither biosynthesis nor genome is published<sup>[25]</sup>. A second natural product class containing a hydroxylated fatty acid chain are the

malyngamides. But here the fatty acid chain is not derived from primary metabolism and the hydroxylation the result of missing reduction step.<sup>[26]</sup> Therefore, a prediction of involved enzymes based on a similar biosynthetic pathway was not possible. In addition, the formation of the vinylic moiety in the 4-amino-2-oxo-hex-5-enoic acid block remains elusive. The structure of the BGC suggests the 4-amino-2-oxo-hex-5-enoic acid block to be derived from one amino acid that is elongated by a polyketide extender. The type of amino acid is still unknown. The A- and C-domain in module 2 (MxgJ) show L-glutamic acid specificity according to the antiSMASH analysis. A first ratio was to test the incorporation of stable isotope labelled L-glutamine, L-glutamic acid and L-aspartic acid, since these amino acids were reported precursors or building blocks of natural products containing a vinylic moiety<sup>[27]</sup>. However, no incorporation was observed, indicating a different biosynthetic pathway. Without a confirmed precursor of the vinylic moiety, a prediction of a biosynthetic pathway only based on predicted domain functions is not possible. One option to consider is direct incorporation of vinylglycine, analogous to the incorporation of ACPC like seen in myxopentacin biosynthesis. Vinylglycine itself is a reported fungal metabolite that implies existence for a yet unknown pathway<sup>[28]</sup>.

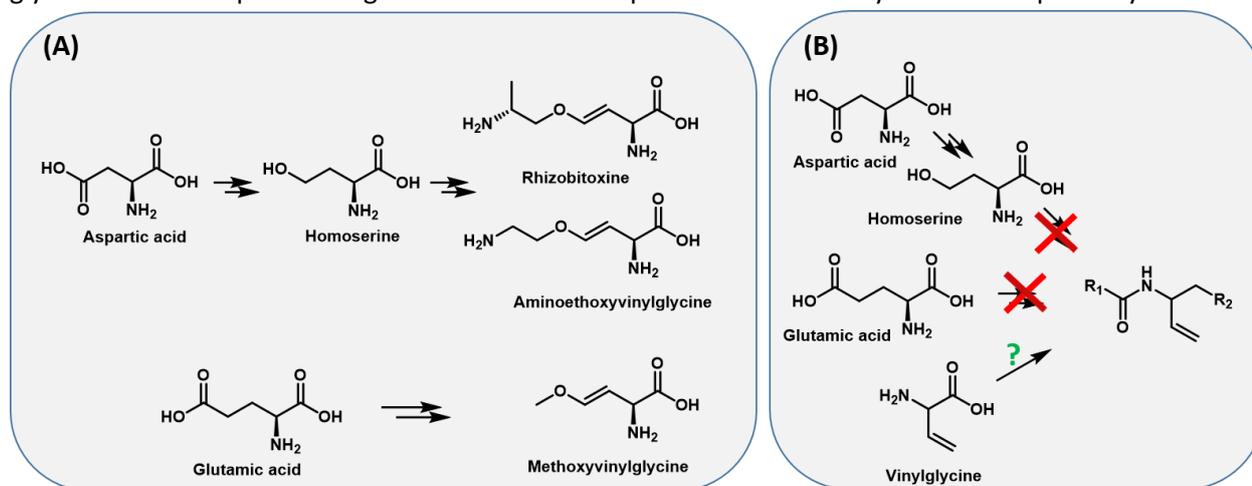


Figure 2: (A): Known precursors (homoserine which is derived from aspartic acid and glutamic acid) of natural products with a vinylic moiety. (B) Expected precursors of the vinylic moiety (glutamic acid and aspartic acid) that were not incorporated in feeding experiments and vinylglycine as potential building block.

Regarding the myxopentacin biosynthesis, the main and remaining difficulty lies in the explanation of the ACPC chain formation. ACPC and AGOOA block biosynthesis can be explained by homologues of biosynthetic enzymes from amipurimycin biosynthesis<sup>[29,30]</sup> and ketomemicin biosynthesis<sup>[31]</sup>. In addition, results from feeding experiments are in agreement with the proposed pathways. But no indications were found for ACPC-chain formation. The only catalytic domains that were expected to form amide bonds are in MxpC and are considered to conduct the two NRPS elongations steps and termination of the biosynthesis. Therefore, involvement of MxpC in ACPC chain formation is regarded as less likely.

Therefore, further options of ACPC chain formation were considered: Other enzymes known to catalyze amide bond formation are GRASP ligases<sup>[32]</sup>. For example in amipurimycin biosynthesis the linkage of ACPC onto the C-9 sugar in is mediated by a GRASP ligase<sup>[29]</sup>. Nevertheless, homologue genes encoding such enzymes were not found in close proximity to ACPC biosynthetic genes. Downstream of ketomemicin biosynthetic homologues *mvpK-N* four metalloprotease/  $\beta$ -hydroxylase type genes can be found. Enzymes encoded by these genes usually have oxidative function ( $\beta$ -hydroxylases) or cleave amide bonds (metalloproteases). *In vitro* studies could show that proteases can also form amide bonds<sup>[32]</sup>. A role in the ACPC chain formation cannot be excluded since no biosynthetic function could be assigned to these enzymes. In addition, a second NRPS type assembly line as mentioned previously cannot be excluded.

Inactivation of one or two  $\beta$ -hydroxylase or metalloprotease genes via creation of single crossover mutants followed by investigation of the metabolome of the mutants can provide an answer towards the relevance of these genes for the myxopentacin biosynthesis.

The highly repetitive character of the myxopentacin molecules implies an iterative mechanism. In contrast to polyketides and fatty acids, iterative acting biosynthetic pathways for peptides are scarcely found. Rare examples are the formation of the polyamino acids like  $\epsilon$ -poly(L-lysine) and poly(L-diaminopropionic acid) as well as the biosynthesis of distamycin and congocidine (figure 1). These natural products originate from untypical NRPS type pathways. The poly(L-diaminopropionic acid) synthase which shares great similarity with  $\epsilon$ -poly(L-lysine) consists of one A-domain, one PCP domain and three C-domains embedded into seven transmembrane domains and act iteratively.<sup>[33]</sup> In contrast to the poly amino acids, distamycin contains only three repetitive blocks of amino-pyrrole-carboxylic acid (congocidine contains two repetitive blocks of amino-pyrrole-carboxylic acid). The amino-pyrrole-carboxylic acid polymerization is expected to be catalyzed iteratively by one standard NRPS module (A-domain,PCP-domain and C-domain) and one free standing PCP and one free standing C-domain.<sup>[34]</sup> Remarkable for both biosynthetic pathways is the strong deviation from textbook logic. Regarding MxpC, the proposed function in chapter 4 reflects textbook logic best. A different, probably iterative function cannot be excluded since NRPS does not always act straight linear, like mentioned. The missing TE domain in MxpC, like in poly-amino acid and distamycin biosynthesis, can be an indication that some other biosynthetic steps than proposed are conducted by MxpC. Nevertheless, based on *in-silico* data without any experimental data towards substrate specificity of A-domains in MxpC, the proposed biosynthetic pathway in chapter 4 seems to be the most likely one. To prove if the biosynthesis follows the expected textbook logic or if MxpC catalyzes iteratively the ACPC chain formation, expression and isolation of both A-domain followed by testing of their specificity towards L-valine, the AGOOA moiety or ACPC would be the next logic step.

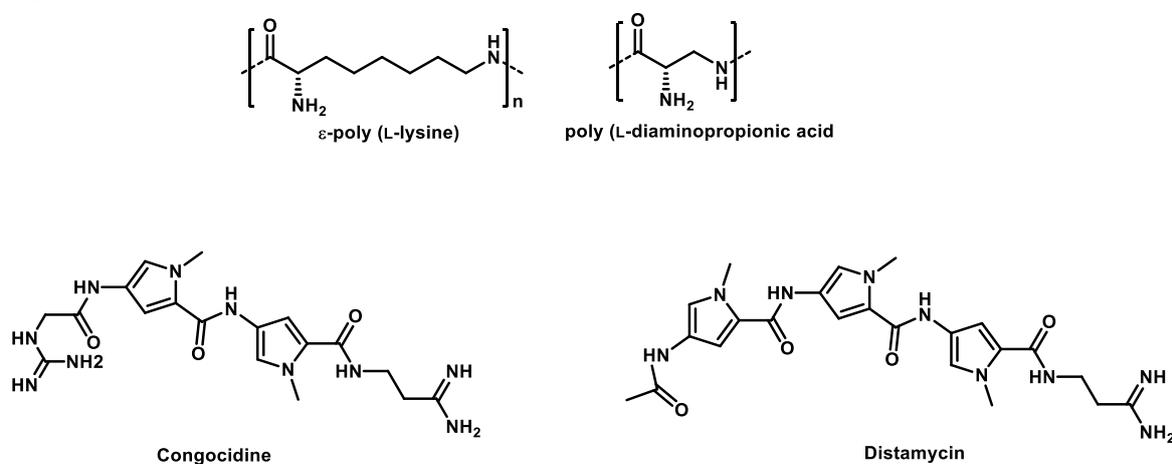


Figure 3: Structures of poly( $\epsilon$ -L-lysine), poly(L-diaminopropionic acid), congocidine and distamycin

## 5.2.1 PLP catalyzed condensation-considerations regarding the resulting stereochemistry

In two of the three identified compound families, a C-C bond is formed in a PLP-dependent manner instead of the typical PKS type Claisen condensation. In most biochemical examples, PLP is the major cofactor in transamination reactions especially in amino acid metabolism. Also in natural product biosynthesis the major role of PLP is found to be transamination reactions during polyketide biosynthesis or epimerization of amino acids in NRPS biosynthesis. Besides these reactions also Claisen type condensation reactions were reported

in the biosynthesis of the natural products ketomemycin, prodigiosin, fumonisin, saxitoxine and also in the sphingosine biosynthesis<sup>[35,36]</sup>. Enzymes conducting these reactions are classified as PLP-dependent aldolases<sup>[37]</sup>. The PLP (as cofactor) is bound to a lysine residue inside the enzyme as an internal aldimine. In order to react, PLP needs to bind to its amino acid substrate forming an external aldimine. A protonation turns the aldimine complex into resonance stabilized quinonoid. From this state on a broad range of reactions are possible<sup>[38]</sup>. For the Claisen condensation, a temporary negative charge at the Schiff's base can perform a nucleophilic attack on a carbonyl group of a CoA ester. Like in a PKS type Claisen reaction the carboxyl group is the leaving group. Protonation of the Schiff base restores the external aldimine and re-implements a stereocenter at the amine (figure 2).

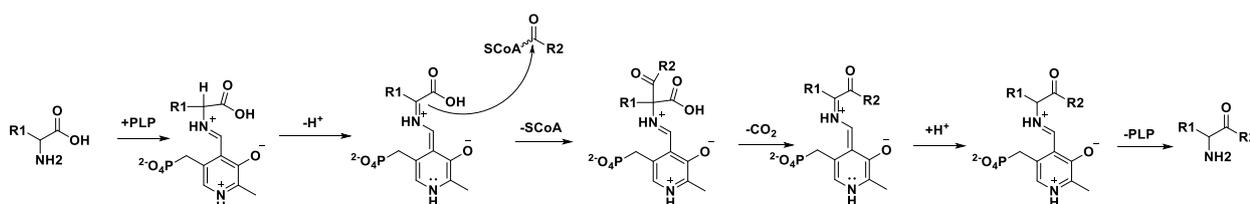


Figure 4: Procedure of PLP dependent claisen condensation.

Based on *in-silico* data the configuration of stereocenters can be predicted for most NRPs and PKS<sup>[39,40]</sup>. The question that lies here: Can based on comparison to described natural products conclusions be drawn towards the stereochemistry of the arginine residue in the myxopentacin compound family and the trimethylammonium/amine in the myxoquaterine family? Regarding literature examples with a defined stereocenter there is only fumonisin, D-sphingosine and an intermediate from saxitoxin biosynthesis.

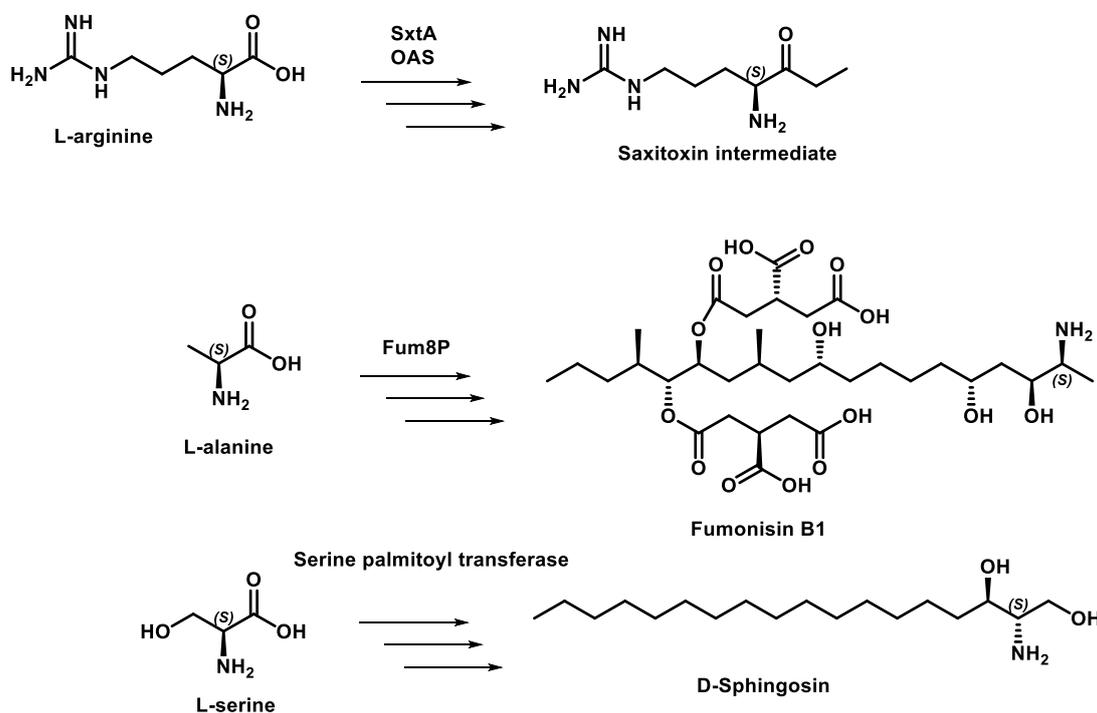
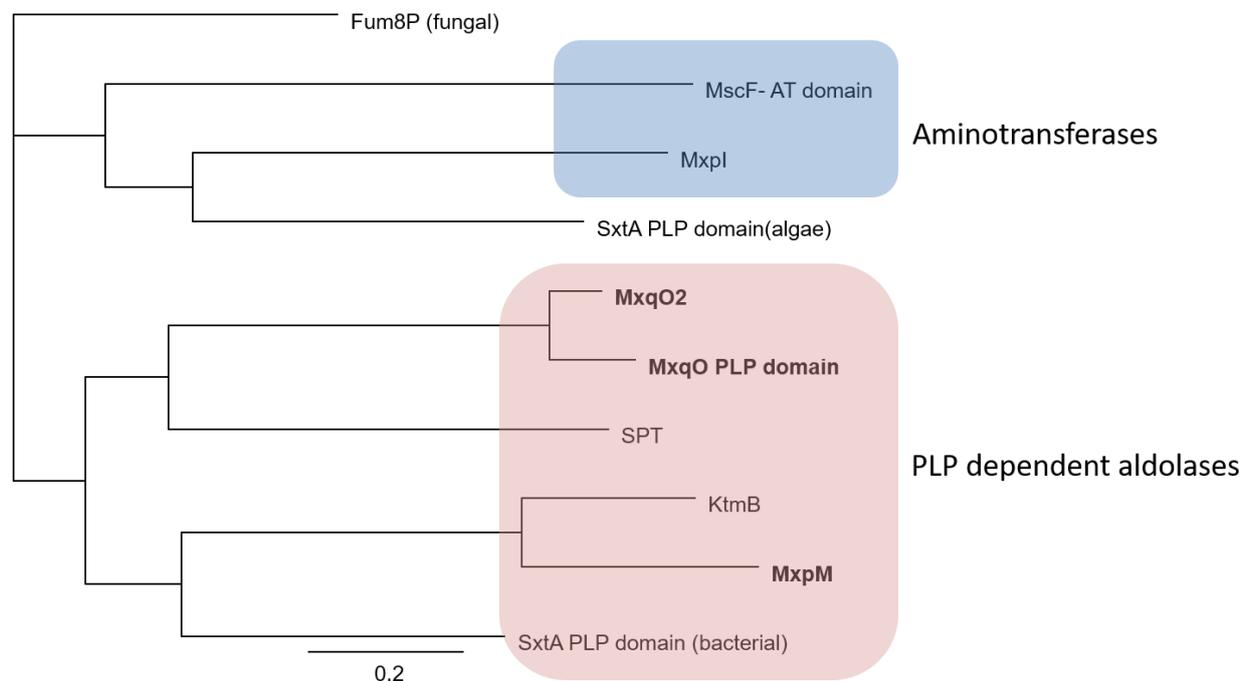


Figure 5: Amino acid substrates and the final products showing the stereo centers and the respective enzymes conducting the PLP dependent C-C bond formation<sup>[22,36,41]</sup>.

Interestingly in the three presented examples, the stereocenters remain preserved even though the stereo information is lost during the enzymatic catalysis. For ketomemycin two different stereoisomers with *R*- and *S*-configuration at the amine were reported. In addition, the *in vitro* reconstitution of corresponding reaction showed the formation of two stereoisomers.<sup>[31]</sup> An amino acid sequence comparison (figure 4) of mentioned

enzymes showed that both myxobacterial PLP-dependent C-acyltransferases MxqI and MxqO are distinct from the myxobacterial aminotransferases MscF and MxpI. However, eukaryotic PLP dependent aldolases are closer to the myxobacterial bacterial aminotransferases.



*Figure 6: Phylogenetic analysis of protein sequences of PLP-dependent acyltransferases. Two myxobacterial aminotransferases (MxpI from the myxopentacin biosynthesis and the AMT domain of MscF from the microsclerdermin biosynthesis) were also included for comparison. From the myxoquaterin BGC MxqO variants from both strains (MSr 11954 and MSr 11367) were included.*

Regarding the stereocenter in myxoquaterine resulting from MxqI: Both domains from MSr 11954 and MSr 11367 are closely related to the serine palmitoyltransferase SPT from *E.coli*. However, for a trustworthy prediction regarding the stereocenter from the serine to predict whether it is recovered like by SPT in sphingosine biosynthesis, this information is not sufficient. A comparison on the structural level may provide the additional data needed. MxpM from the myxopentacin BGC shows great similarity with KtmB from the ketomemycin BGC. Therefore, a similar mechanism is indicated and the formation of both stereoisomers of the AGOOA block is likely. However, only one main peak can be seen for the isolated myxopentacin derivatives. This implies that only one putative AGOOA isomer is preferably utilized in myxopentacin biosynthesis. So far, the free AGOOA block was not detected in the crude extract. To summarize, no prediction can be given for the stereoselectivity of MxpM. In general for valid predictions a way bigger dataset like used for the C-domain specificity prediction is necessary<sup>[39]</sup>.

### 5.3 Function of natural products

If talking of natural products as secondary metabolites, it is automatically assumed that they have a beneficial function for the producer organism. The rationale of isolating natural products is that we hope they also have a beneficial function for humans such as providing new lead scaffolds for the development of pharmaceutical drugs or crop protection agents<sup>[42]</sup>. As anticipated, the biological function of many natural products are not associated directly with the need of humans. Actually, there is a huge discrepancy between what is beneficial

for us and what is beneficial for a natural product-producing microorganism, since the function of natural products is more diverse than killing the local competitors of the producer organism. For example natural products also play a vital role in cell-cell communication, iron uptake and organization of the sporulation.<sup>[43]</sup>

Regarding the presented molecules in this work, it is crucial to understand the function of the natural product for the producer strain, but also to evaluate if the molecule can be beneficial for us. Considering the bioactivity, the myxoquaterines are the most interesting compound class presented here, since they are highly cytotoxic and exhibit a moderate fungicidal and anti-gram positive bioactivity. The most characteristic feature of this compound class is the quaternary ammonium ion. Among bacteria-derived natural products, these molecules are rare exceptions. Another example of bacteria-derived molecules with a quaternary ammonium ion is the amino acid ergothioneine<sup>[44]</sup>. In addition, some natural products with a quaternary ammonium ion were isolated from fungi or plants. Examples are muscarine and aeruginascine from fungi and the plant derived tubocurarine. All these molecules have in common that they interfere with neuron signaling and are considered as neuro toxins.<sup>[45]</sup> Beside the mentioned natural products, large groups of synthetic molecules with quaternary ammonium exist.

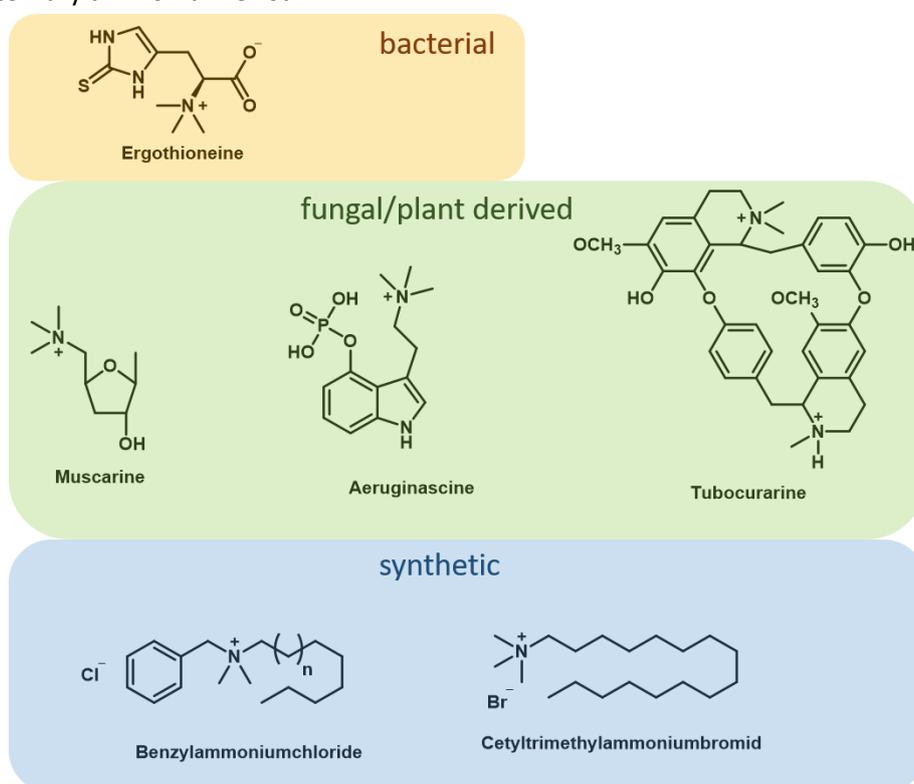


Figure 7: Examples of bacterial, plant and fungi derived molecules as well as two synthetic molecules benzylammoniumchloride and cetyltrimethylammoniumbromide, commonly used as disinfectants.

Most of these synthetic quaternary ammonium compounds (QACs) act as cationic surfactants and have broad antimicrobial activity. The antimicrobial activity is mediated through intercalation with the cell membrane leading to membrane disruption. Therefore they are mainly used as surface disinfectants.<sup>[46]</sup> A similar function is also expected for the myxoquaterine compound family suggesting these molecules are the first bacterial derived disinfectants. Considering the natural soil environment filled with other competitors, production of this compound class could give the producer strains a huge advantage. Myxoquaterin 450, which has no quaternary ammonium ion, does not show any antibacterial activity, which supports the hypothesis that the positive charge of the quaternary ammonium ion is essential for the antibacterial activity. This finding is also in agreement with a study showing that synthetic chemical addition of a quaternary ammonium ion to

natural products increased or led to a gain of antibacterial activity<sup>[47]</sup>. Since a moderate anti-fungal activity is retained in myxoquaterine 450, a different mode of action against fungi is implied.

The antimicrobial effect of cationic surfactants through cell wall disruption is highly unspecific. Production of molecules from the myxoquaterine compound family requires self-resistance. A described resistance mechanism towards quaternary ammonium compounds are QAC transporter like QacE, which is found in many QAC resistant bacteria<sup>[46]</sup>. But so far, none of these bacteria were described to produce QACs. Within the myxoquaterine BGC no homologues of QacE were identified, but the cluster contains three transport/efflux pump proteins MxqJ, MxqK and MxqL. Since the producer strains needs to tolerate high concentrations of its own toxin, which targets the cell wall, efflux mechanisms might not be sufficient and a different self-resistance mechanism may exist.

The high cytotoxic activity also marks this natural product class as beneficial for us in terms of a lead scaffold for a potential anti-cancer drug. Initially, the cytotoxicity was also considered to be mediated by the trimethyl-ammonium ion since this structural feature was considered as major difference between myxoquaterin 535 to the Dkxanthenes. In contrast to the myxoquaterines, the DKxanthenes do not show antimicrobial or cytotoxic activity. But cytotoxicity of quaternary ammonium compounds is usually in the range between 10-100  $\mu\text{M}$ <sup>[48]</sup>, while the determined  $\text{IC}_{50}$  values of the myxoquaterines are significantly lower. In addition, the cytotoxicity in myxoquaterine 450 is retained even though this molecule does not contain the trimethyl-ammonium group, what implies that the cytotoxicity is mediated through a different mode action than cell wall disruption as described for quaternary ammonium compounds.

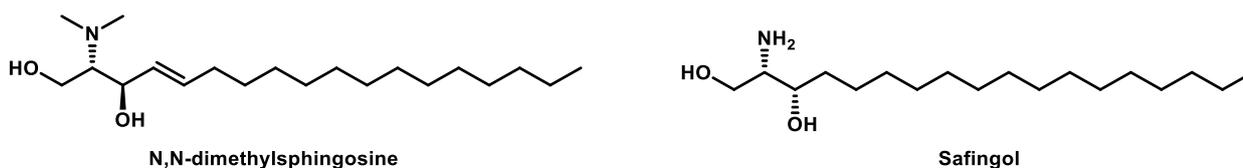


Figure 8: Kinase inhibitors with structural similarities to the myxoquaterin compound family.

The C-terminal part of myxoquaterine 450 strongly resembles the kinase inhibitors N,N-dimethylsphingosine<sup>[49]</sup> and safingol, which is currently tested in class I clinical trials<sup>[50]</sup>. Safingol is described to inhibit protein kinase C and sphingosine kinase 1 and to induce to cell death<sup>[51]</sup>. For the myxoquaterines, especially myxoquaterine 450, a similar mode of action is likely to occur. Therefore, future plans include the testing of the myxoquaterines for their ability to inhibit crucial kinases *in vitro*.

For the myxopentacin molecules only a moderate cytotoxic activity was detected. Nonetheless, the structure is intriguing and contains scaffolds parts that were associated with interesting bioactivity. But without concrete indications considerations towards a beneficial function for the producer strains and usability as a lead scaffold for a future drug are highly speculative.

Thinking of the fact that after certain genetic manipulations production of most myxopentacin derivatives dramatically increased, a possible connection between DNA integration and myxopentacin production cannot be excluded. Integration of foreign DNA into genomic DNA, as it was the case with the promoter insertions, might be considered as phage attack and will lead to a response by the bacteria. Common bacterial answers are restriction enzymes or activation of the CRISPR/Cas system. But recently, also some natural products were described to have anti phage activity<sup>[52]</sup>. A similar effect cannot be excluded here. But experimental observations so far do not support this hypothesis. If myxopentacin production would be a direct answer to the insertion of foreign DNA, increased myxopentacin would be expected after every genetic manipulation, which was not the case. Beside the speculative function as anti-phage system, it is possible that the myxopentacin molecules simply function as ACPC storage scaffold. ACPC, also is known as cispentacin,

has antibiotic effects against fungi<sup>[53]</sup>. Studies have shown that cispentacin interferes with the amino acid metabolism and prolyl-tRNA in *C. albicans*<sup>[54]</sup>. Interestingly the next gene following the genes for ACPC biosynthesis genes has a significant homology to an YbaK prolyl tRNA associated region. YbaK proteins were described to deacylate tRNA<sup>[55]</sup>. An enzyme function like this can be interpreted as some kind of self-resistance mechanism at least towards ACPC, which was detected in the crude extract. Interestingly a similar protein was not found within the amipurimycin BGC. Since no antifungal bioactivity was found for the myxopentacin molecules, it can be assumed that the ACPC moiety loses its antifungal properties if linked in a peptide chain. Considering the cispentacin (ACPC) mode of action to function as anti-metabolite, it is not a surprise, that peptide bound cispentacin (ACPC) shows no antifungal activity.<sup>[54]</sup> Interestingly this observation wasn't made for amipurimycin where ACPC is also linked to a nucleoside via a peptide bond<sup>[56]</sup>. Considering this cytotoxic activity in combination with the carbonylmethylene moiety, a potential pharmacological use can be hypothesized: A very similar structure can be found in the arphamenines which are known aminopeptidase B inhibitors<sup>[57]</sup>. Aminopeptidase B is a mammalian endopeptidase expected to be involved in various cellular processes including signaling in inflammatory cascades<sup>[58]</sup>. For the myxopentacin molecules a similar function as described for arphamenine is thinkable since both molecules mimic a peptide bond. To test the myxopentacin molecules for their ability to inhibit aminopeptidase B, an assay like described by Pham et al.<sup>[59]</sup> would be an option.

The myxoglucamides represent a new class of glycolipopeptides. Many glycolipopeptide type scaffolds are pharmacologically used for immunomodulation. They are under study for example as an adjuvant for vaccines or recently also for immunotherapy against cancer.<sup>[60]</sup> In general, glycosylation pattern on small molecules, proteins or whole cells play an important function in the recognition process for the immune cells. A prime example for such a recognition process is the binding of lipo-polysaccharides from the cell wall of gram-negative bacteria to the toll-like receptor 4 on immune cells, which promotes a strong immune response often leading to a septic shock. But among the glycolipopeptides with a pharmacological relevance, natural products are rare. Examples are occidiofungin isolated from *Burkholderia contaminans*, hassallidins first described from the cyanobacterium *Hassallia sp.*, both with antifungal bioactivity and the ieodoglucomides from *Bacillus licheniformis* which provide a broad array antimicrobial and a moderate cytotoxic bioactivity (figure 7)<sup>[25,61]</sup>. From these glycolipopeptides the ieodoglucomides resemble most the myxoglucamides. Two of the isolated myxoglucamides (myxoglucamide 647 and 661) also exhibited moderate cytotoxic bioactivity. In contrast to the ieodoglucomides, no antimicrobial bioactivity was found in the biotesting. Nonetheless, using a more diverse test panel including assays regarding immunomodulatory functions may show some interesting bioactivity.

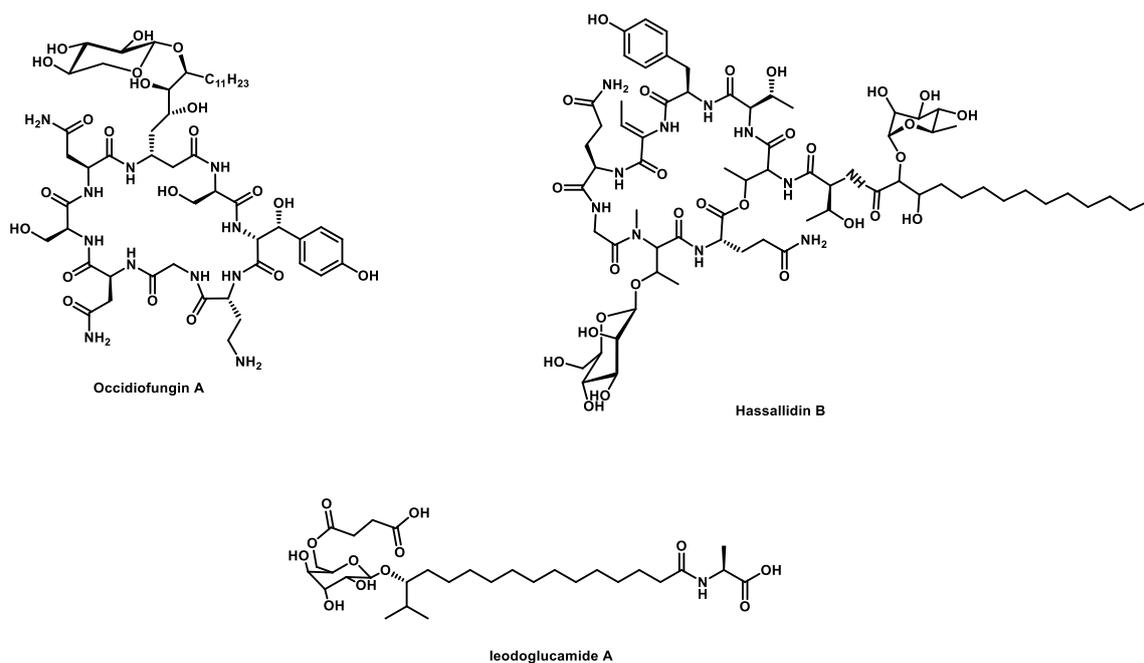


Figure 9: Structures of the Glycolipopeptides occidiofungin A, hassallidin B and ieodolgucomide A

## 5.6 Final words

Beside their social lifestyle, myxobacteria are excellent factories for novel natural products. The identified natural products in this thesis show an extreme diversity even though they all originate from NRPS-PKS type assembly lines. It is fascinating, that structural features, usually found among different bacterial species, can also be found in molecules isolated from myxobacteria. Could it be that the majority of bacterial biosynthetic diversity can be found in the order of Myxococcales?

## 5.7 References

- [1] P. R. Jensen, K. L. Chavarria, W. Fenical, B. S. Moore, N. Ziemert, *J. Ind. Microbiol. Biotechnol.* **2014**, *41*, 203.
- [2] T. Hoffmann, D. Krug, N. Bozkurt, S. Duddela, R. Jansen, R. Garcia, K. Gerth, H. Steinmetz, R. Müller, *Nat. Commun.* **2018**, *9*, 803.
- [3] K. Arima, H. Imanaka, M. Kousaka, A. Fukuta, G. Tamura, *Agric. Biol. Chem.* **1964**, *28*, 575.
- [4] K. Ochi, T. Hosaka, *Appl. Microbiol. Biotechnol.* **2013**, *97*, 87.
- [5] F. Panter, D. Krug, R. Müller, *ACS Chem. Biol.* **2019**, *14*, 88.
- [6] K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema, T. Weber, *Nucleic Acids Res.* **2019**, W81-W87.
- [7] C. D. Bader, F. Panter, R. Müller, *Biotechnol Adv* **2019**, submitted, 107480.
- [8] K. Blin, V. Pascal Andreu, E. L. C. de Los Santos, F. Del Carratore, S. Y. Lee, M. H. Medema, T. Weber, *Nucleic acids research* **2018**.
- [9] a) M. L. Micallef, P. M. D'Agostino, D. Sharma, R. Viswanathan, M. C. Moffitt, *BMC genomics* **2015**, *16*, 669; b) D. E. Cane, X. He, S. Kobayashi, S. Ōmura, H. Ikeda, *J Antibiot* **2006**, *59*, 471.
- [10] a) M. Funabashi, N. Funata, S. Horinouchi, *J. Biol. Chem.* **2008**, *283*, 13983; b) J. J. Hug, F. Panter, D. Krug, R. Müller, *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 319.
- [11] J. Bérdy, *The Journal of antibiotics* **2012**, *65*, 385.

- [12] H. B. Bode, M. W. Ring, G. Schwär, M. O. Altmeyer, C. Kegler, I. R. Jose, M. Singer, R. Müller, *ChemBioChem* **2009**, *10*, 128.
- [13] B. M. Hover, S.-H. Kim, M. Katz, Z. Charlop-Powers, J. G. Owen, M. A. Ternei, J. Maniko, A. B. Estrela, H. Molina, S. Park et al., *Nat. Microbiol.* **2018**, *3*, 415.
- [14] a) S. Khater, S. Anand, D. Mohanty, *Synthetic and Systems Biotechnology* **2016**, *1*, 80; b) A. Greule, S. Zhang, T. Paululat, A. Bechthold, *J. Vis. Exp.* **2017**.
- [15] S. C. Wenzel, B. Kunze, G. Höfle, B. Silakowski, M. Scharfe, H. Blöcker, R. Müller, *ChemBioChem* **2005**, *6*, 375.
- [16] P. Meiser, K. J. Weissman, H. B. Bode, D. Krug, J. S. Dickschat, A. Sandmann, R. Müller, *Chem. Biol.* **2008**, *15*, 771.
- [17] P. Meiser, H. B. Bode, R. Müller, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19128.
- [18] M. Hahn, T. Stachelhaus, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 275.
- [19] A. Bateman, E. Birney, R. Durbin, S. R. Eddy, K. L. Howe, E. L. Sonnhammer, *Nucleic Acids Res.* **2000**, *28*, 263.
- [20] L. Zimmermann, A. Stephens, S.-Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A. N. Lupas, V. Alva, *Journal of molecular biology* **2018**, *430*, 2237.
- [21] L. Du, L. Lou, *Nat. Prod. Rep.* **2010**, *27*, 255.
- [22] R. Gerber, L. Lou, L. Du, *J. Am. Chem. Soc.* **2009**.
- [23] M. Kopp, H. Irschik, K. Gemperlein, K. Buntin, P. Meiser, K. J. Weissman, H. B. Bode, R. Müller, *Mol. Biosyst.* **2011**, *7*, 1549.
- [24] E. Eichhorn, J. R. van der Ploeg, M. A. Kertesz, T. Leisinger, *Journal of Biological Chemistry* **1997**, *272*, 23031.
- [25] F. S. Tareq, J. H. Kim, M. A. Lee, H.-S. Lee, Y.-J. Lee, J. S. Lee, H. J. Shin, *Organic letters* **2012**, *14*, 1464.
- [26] N. A. Moss, T. Leao, M. Rankin, T. M. McCullough, P. Qu, A. Korobeynikov, J. L. Smith, L. Gerwick, W. H. Gerwick, *ACS Chem. Biol.* **2018**.
- [27] a) J. B. Patteson, Z. D. Dunn, B. Li, *Angew. Chem. Int. Ed.* **2018**; b) T. Yasuta, S. Okazaki, H. Mitsui, K. Yuhashi, H. Ezura, K. Minamisawa, *Applied and environmental microbiology* **2001**, *67*, 4999; c) M. Fernández, Y. Cuadrado, J. F. Aparicio, J. F. Martín, *Microbiology* **2004**, *150*, 1467.
- [28] G. Dardenne, J. Casimir, M. Marlier, P. O. Larsen, *Phytochemistry* **1974**, *13*, 1897.
- [29] W.-J. Kang, H.-X. Pan, S. Wang, B. Yu, H. Hua, G.-L. Tang, *Organic letters* **2019**.
- [30] A. J. Romo, T. Shiraishi, H. Ikeuchi, G.-M. Lin, Y. Geng, Y.-H. Lee, P. H. Liem, T. Ma, Y. Ogasawara, K. Shin-Ya et al., *J. Am. Chem. Soc.* **2019**.
- [31] J. Kawata, T. Naoe, Y. Ogasawara, T. Dairi, *Angew. Chem. Int. Ed.* **2017**.
- [32] A. Goswami, S. G. van Lanen, *Molecular bioSystems* **2014**, *11*, 338.
- [33] Z. Xu, Z. Sun, S. Li, Z. Xu, C. Cao, Z. Xu, X. Feng, H. Xu, *Scientific Reports* **2015**, *5*, 17400.
- [34] C. Hao, S. Huang, Z. Deng, C. Zhao, Y. Yu, *PLoS ONE* **2014**, *9*, e99077.
- [35] Y.-L. Du, K. S. Ryan, *Nat. Prod. Rep.* **2019**, *36*, 430.
- [36] K. Hanada, *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **2003**, *1632*, 16.
- [37] J. Liang, Q. Han, Y. Tan, H. Ding, J. Li, *Front. Mol. Biosci.* **2019**, *6*, 4.
- [38] M. D. Toney, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2011**, *1814*, 1407.
- [39] C. Rausch, I. Hoof, T. Weber, W. Wohlleben, D. H. Huson, *BMC Evol. Biol.* **2007**, *7*, 78.
- [40] S. A. Bonnett, J. R. Whicher, K. Papireddy, G. Florova, J. L. Smith, K. A. Reynolds, *Chem. Biol.* **2013**, *20*, 772.
- [41] S. Tsuchiya, Y. Cho, K. Konoki, K. Nagasawa, Y. Oshima, M. Yotsu-Yamashita, *Scientific Reports* **2016**, *6*, 20340.
- [42] F. E. Dayan, C. L. Cantrell, S. O. Duke, *Bioorg Med Chem* **2009**, *17*, 4022.
- [43] M. F. Traxler, R. Kolter, *Nat. Prod. Rep.* **2015**, *32*, 956.
- [44] F. P. Seebeck, *J. Am. Chem. Soc.* **2010**, *132*, 6632.
- [45] a) J. A. Homer, J. Sperry, *Journal of natural products* **2017**, *80*, 2178; b) R. F. Ochillo, C. S. Tsai, M. H. Tsai, *Br. J. Pharmacol.* **1981**, *72*, 225; c) A. A. Selyanko, V. A. Derkach, D. A. Dé Kurennyi, V. I. Skok, *Neurophysiology* **1988**, *20*, 493.
- [46] M. C. Jennings, K. P. C. Minbiole, W. M. Wuest, *ACS infectious diseases* **2015**, *1*, 288.

- [47] M. D. Joyce, M. C. Jennings, C. N. Santiago, M. H. Fletcher, W. M. Wuest, K. P. Minbiole, *J Antibiot (Tokyo)* **2016**, *69*, 344.
- [48] H. Nagamune, T. Maeda, K. Ohkura, K. Yamamoto, M. Nakajima, H. Kourai, *Toxicology in Vitro* **2000**, *14*, 139.
- [49] L. C. Edsall, J. R. van Brocklyn, O. Cuvillier, B. Kleuser, S. Spiegel, *Biochemistry* **1998**, *37*, 12892.
- [50] M. A. Dickson, R. D. Carvajal, A. H. Merrill, M. Gonen, L. M. Cane, G. K. Schwartz, *Clin. Cancer Res.* **2011**, *17*, 2484.
- [51] a) J. Coward, G. Ambrosini, E. Musi, J.-P. Truman, A. Haimovitz-Friedman, J. C. Allegood, E. Wang, A. H. Merrill, G. K. Schwartz, *Autophagy* **2009**, *5*, 184; b) M. Hamada, H. Kameyama, S. Iwai, Y. Yura, *Cell Death Discov.*, *3*, 1.
- [52] S. Kronheim, M. Daniel-Ivad, Z. Duan, S. Hwang, A. I. Wong, I. Mantel, J. R. Nodwell, K. L. Maxwell, *Nature* **2018**, *1*.
- [53] M. Konishi, M. Nishio, K. Saitoh, T. Miyaki, T. Oki, H. Kawaguchi, *The Journal of antibiotics* **1989**, *42*, 1749.
- [54] J. O. Capobianco, D. Zakula, M. L. Coen, R. C. Goldman, *Biochemical and Biophysical Research Communications* **1993**, *190*, 1037.
- [55] B. Ruan, D. Söll, *Journal of Biological Chemistry* **2005**, *280*, 25887.
- [56] T. Goto, Y. Toya, T. Ohgi, T. Kondo, *Tetrahedron Letters* **1982**, *23*, 1271.
- [57] H. Umezawa, T. Aoyagi, S. Ohuchi, A. Okuyama, H. Suda, T. Takita, M. Hamada, T. Takeuchi, *J Antibiot* **1983**, *36*, 1572.
- [58] L. Orning, J. K. Gierse, F. A. Fitzpatrick, *Journal of Biological Chemistry* **1994**, *269*, 11269.
- [59] V.-L. Pham, M.-S. Cadel, C. Gouzy-Darmon, C. Hanquez, M. C. Beinfeld, P. Nicolas, C. Etchebest, T. Foulon, *BMC Biochem* **2007**, *8*, 21.
- [60] I. Bettahi, G. Dasgupta, O. Renaudet, A. A. Chentoufi, X. Zhang, D. Carpenter, S. Yoon, P. Dumy, L. BenMohamed, *Cancer immunology, immunotherapy : CII* **2009**, *58*, 187.
- [61] a) S. E. Lu, J. Novak, F. W. Austin, G. Gu, D. Ellis, M. Kirk, S. Wilson-Stanford, M. Tonelli, L. Smith, *Biochemistry* **2009**; b) J. Vestola, T. K. Shishido, J. Jokela, D. P. Fewer, O. Aitio, P. Permi, M. Wahlsten, H. Wang, L. Rouhiainen, K. Sivonen, *PNAS* **2014**, *111*, E1909-17.