

# **Discovery of New Natural Products and Biosynthetic Gene Clusters from *Streptomyces***

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

von

Marc Lukas Stierhof

Saarbrücken

2021

Tag des Kolloquiums:	04.02.2022
Dekan:	Prof. Dr. Jörn Walter
Berichterstatter:	Prof. Dr. Andriy Luzhetskyy Prof. Dr. Rolf Müller
Vorsitz:	Prof. Dr. Uli Kazmaier
Akad. Mitarbeiter:	Dr. Stefan Boettcher

Diese Arbeit entstand unter der Anleitung von Prof. Dr. Andriy Luzhetskyy in der Fachrichtung 8.2, Pharmazeutische Biotechnologie der Naturwissenschaftlich-Technischen Fakultät der Universität des Saarlandes von Februar 2017 bis September 2021.

## Acknowledgments

I am deeply grateful to my doctor father Prof Andriy Luzhetskyy who gave me the opportunity to do my PhD in his group. I am very thankful for his support and guidance during my PhD which engaged me to challenge and complete many sophisticated projects he gave me. I would like to thank the unbeatable tag-team Maksym Myronovskiy and Birgit Rosenkränzer for their support in setting up experiments and sharing their vast knowledge about microbiology, biotechnology, writing papers and much, much more. I also owe my special gratitude to my second supervisor Prof Rolf Müller for giving me advice during the thesis committee meetings and for reviewing my thesis. Furthermore, my deepest thanks go to Josef Zapp who taught me all the secrets about NMR spectroscopy and who infected me with his NMR enthusiasm. I am also thankful for Judith Hoffmann and Lena Keller for hosting the NMR group seminar and for the colleagues from HIPS for a great scientific exchange and sharing warm memories.

I was very lucky to work in a group of many great colleagues who helped me to find solution to any work-related problems and with whom I share many great forever lasting memories. I really enjoyed all the conversations and discussions about all kind of topics including live, death, gossip, politics, and one piece. Thank also to Dominik Wichner with whom I survived my bachelor studies in Regensburg and who motivated me to do my master thesis abroad in Aberdeen.

And last but not least, danke an meine Familie, insbesondere an meine Eltern, für die Unterstützung während meines gesamten Studiums, für die ermutigenden Worte, wenn es mal nicht so gut lief und dafür, dass ihr immer für mich da wart! Danke auch an meine Freunde und meine Freundin Lilli, die mich ermutigt haben an mich zu glauben und die immer bedingungslos für mich da waren.



## Zusammenfassung

Aufgrund des ständigen Aufkommens neuer Krankheiten und Antibiotikaresistenzen werden dringend neue Arzneimittel mit verbesserten Eigenschaften benötigt. Viele der im Handel erhältlichen Medikamente basieren auf Naturstoffen (NS) oder Naturstoffderivaten. Actinobakterien, insbesondere die Gattung *Streptomyces*, sind eine wichtige Quelle für eine Vielzahl von klinisch relevanten NS. Aufgrund der intensiven Ausbeutung von Streptomyceten im 20. Jahrhundert galt ihr Potenzial als erschöpft. Entwicklungen in der instrumentellen Analytik und der Biotechnologie haben jedoch neue verborgene biosynthetische Fähigkeiten von Streptomyceten ans Licht gebracht. In dieser Arbeit wurden durch Dereplikation, Genom-Mining und heterologe Expression mehrere Streptomyceten-Stämme auf die Produktion neuer NS und der zugehörigen biosynthetischen Gencluster (BGC) untersucht. Dies führte zur Entdeckung der Strukturen und BGC von neuen Depsibosamycinen (Peptide) aus *S. aurantiacus* LU19075 und Lipothreninen (Lipo-Aminosäuren) aus *S. aureus* LU18118. *De novo* Strukturaufklärung war ein großer Teil dieser Arbeit und unterstützte die strukturelle Charakterisierung von Cyclofaulknamycin, welches durch einen Target-Genom-Mining-Ansatz entdeckt wurde. Weitere *de novo* Strukturaufklärungen in Zusammenarbeit mit Kollegen lieferten einen Beitrag zur Entdeckung von vielen neuen NS. Die erzielten Ergebnisse leisten einen Beitrag zur Naturstoffforschung und bestätigen Streptomyceten als reiche Quelle für neue NS.

## Abstract

Due to the constant appearance of new diseases and antibiotic resistances new drugs with improved properties are urgently needed. Many of the commercially available drugs are based on natural products (NPs) or natural product derivatives. Actinobacteria, especially the genus *Streptomyces* are an important source of a variety of NPs with clinical relevance. Due to intensive exploitation of *Streptomyces* during the 20<sup>th</sup> century, their potential was considered exhausted. However, developments in analytical chemistry and biotechnology revealed new biosynthetic abilities hidden in *Streptomyces*. Herein, by applying dereplication, genome mining and heterologous expression, several *Streptomyces* strains have been analyzed for the production of new natural products and their corresponding biosynthetic gene clusters (BGCs). This led to the discovery of the structures and BGCs of new depsibosamycins (peptides) from *S. aurantiacus* LU19075 and lipothrenins (lipo-amino acid) from *S. aureus* LU18118. De novo structure elucidation was a large part of this work and supported the structural characterization of cyclofaulknamycin, which was discovered through a target genome mining approach. Further *de novo* structural elucidations in collaboration with colleagues provided a contribution to the discovery of dudomycin, bonsecamin and many more new NPs. The obtained results contribute to the field of natural product research and confirm *Streptomyces* as a rich source of new NPs.

## Publications

Publications created in the course of this work:

**Stierhof M.**, Myronovskyi M., Zapp J., Luzhetskyy A. Discovery and heterologous expression of new cyclic depsibosamycins. *Microorganisms* **2021**, 9(7), 1396

Horbal L., **Stierhof M.**, Paluszczak A., Eckert N., Zapp J., Luzhetskyy A. Cyclofaulknamycins with the rare amino acid D-capreomycinidine isolated from a well-characterized *Streptomyces albus* strain. *Microorganisms* **2021**, 9(8), 1609

Lasch C., **Stierhof M.**, Estévez R.M., Myronovskyi M., Zapp J., Luzhetskyy A. Bonsecamin: A new cyclic pentapeptide discovered through heterologous expression of a cryptic gene cluster. *Microorganisms* **2021**, 9(8), 1640

Lasch C., **Stierhof M.**, Estévez R.M., Myronovskyi M., Zapp J., Luzhetskyy A. Dudomycins: New secondary metabolites produced after heterologous expression of an NRPS cluster from *Streptomyces albus* ssp. *Chlorinus* NRRL B-24108. *Microorganisms* **2020**, 8(11), 1800

Myronovskyi M., Rosenkränzer B., **Stierhof M.**, Petzke L., Seiser T., Luzhetskyy A. Identification and heterologous expression of the albucidine gene cluster from the marine strain *Streptomyces albus* ssp. *Chlorinus* NRRL B-24108. *Microorganisms* **2020**, 8(2), 237

## Conference Contributions

**Stierhof, M.**; Horbal, L.; Paluszczak, A.; Eckert, N.; Myronovskyi, M.; Zapp, J.; Luzhetskyy, A. Hidden rings: Discovery of cyclic NRPS products by moiety target search and dereplication, **12<sup>th</sup> HIPS Symposium, Saarbrücken, Germany, 2021** (poster)

# Table of Contents

Acknowledgements .....	iv
Zusammenfassung .....	v
Abstract .....	vi
Publications and Conference Contributions .....	vii
<b>1 Introduction .....</b>	<b>1</b>
1.1 Natural Product Research .....	1
1.1.1 Historic Overview and Importance .....	1
1.1.2 Microbial Natural Products .....	3
1.2 Biosynthetic Machineries .....	4
1.2.1 Peptide Biosynthesis .....	4
1.2.2 Polyketide and Fatty Acid Biosynthesis .....	6
1.3 Discovery of New Natural Products and the Corresponding Biosynthetic Gene Clusters from Streptomyces .....	8
1.3.1 Methods for Natural Product Discovery in Streptomyces .....	8
1.3.2 De novo Structure Elucidation by NMR .....	11
1.3.2.1 NMR Basics .....	11
1.3.2.2 Challenges in Natural Product NMR .....	14
1.3.3 Elucidation of Natural Product Biosynthesis .....	16
1.3.4 Utilization of New Biosynthetic Gene Clusters and Chemical Scaffolds in Nature Product Synthesis - Chemists vs Microorganism.....	17
1.4 Scope of the Thesis .....	19
References .....	20
<b>2 Publications .....</b>	<b>25</b>
Discovery and Heterologous Production of New Cyclic Depsibosamycins .....	I
Cyclofaulknamycin with the Rare Amino Acid D-capreomycinidine Isolated from a Well- Characterized <i>Streptomyces albus</i> Strain .....	II
Bonsecamin: A New Cyclic Pentapeptide Discovered through Heterologous Expression of a Cryptic Gene Cluster .....	III

Dudomycins: New Secondary Metabolites Produced after Heterologous expression of an Nrps Cluster from <i>Streptomyces albus</i> ssp. <i>Chlorinus</i> Nr1 B-24108 .....	IV
--	----

<b>3 Lipothrenins - N-hydroxylated threonine-hexadecanedioic-acids produced after heterologous expression of a fatty acid biosynthetic gene cluster .....</b>	<b>170</b>
3.1 Introduction .....	170
3.2 Materials and Methods .....	171
3.2.1 General Experimental Procedures .....	171
3.2.2 Metabolite Extraction and Analysis by Mass Spectrometry (MS) .....	171
3.2.3 Nuclear Magnetic Resonance (NMR) Spectroscopy and Optical Rotation (OD) .....	171
3.2.4 General isolation procedure of lipothrenins and NAC-lipothrenins .....	172
3.2.5 Marfey's Analysis .....	172
3.2.6 Isolation and Manipulation of DNA .....	172
3.2.7 Genome Mining and Bioinformatics Analysis .....	173
3.3 Results and Discussion .....	173
3.3.1 Identification and expression of the lipothrenin gene cluster.....	173
3.3.2 Isolation and structure elucidation of lipothrenins .....	174
3.3.3 Identification of the lipothrenin biosynthetic genes .....	177
3.3.3.1 Determination of the minimal set of lipothrenin biosynthesis genes .....	177
3.3.3.2 LitB is involved in hexadecanedioic acid biosynthesis .....	179
3.3.3.3 Stereoselective N-oxygenation .....	180
3.3.3.4 Biosynthetic route to lipothrenins and NAC-lipothrenins .....	181
3.4 Summary .....	183
References .....	184
Supplementary .....	186
<b>4 Summary and Conclusion .....</b>	<b>236</b>
References .....	240

# 1 Introduction

## 1.1 Natural Product Research

### 1.1.1 Historic Overview and Importance

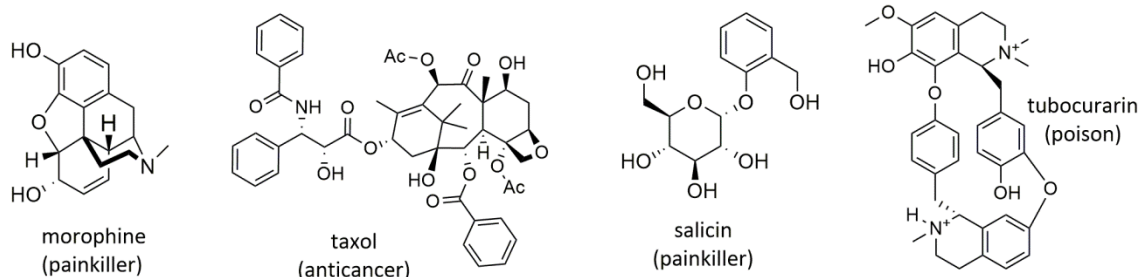
The human species has been withstanding against all imponderables of life since the beginning of its evolution. We developed strategies to defend against predators, developed agriculture, became sedentary and eventually have put us at the top of the food chain. Despite all external threats we have faced and challenged, the human body is susceptible to microbial infections and diseases. Our immune system is able to protect us from light infections but fails in severe diseases such as pest, smallpox and Spanish flue. Millennia ago, the human species experienced and learned, that eating plants has certain biological effects on the body. By trial and error, these effects were systemized and used across many ancient cultures to treat disease, pain or were used for rituals and hunting.<sup>1, 2</sup> The Calabar bean (*Physostigma venenosum*) for example has been used to pass a divine judgement about life and death of persons accused to crimes.<sup>3</sup> It is assumed, in the case of innocence the bean was eaten fast which promoted vomiting and survival while guilty persons died from poisoning by eating the bean slowly. Another example is the use of the seeds from the plant family *Menispermaceae* that contain the hunting poison curare which was sprinkled on arrows to paralyze hit animals (Fig 1a).<sup>4</sup> Plants have formed the bases of traditional medicines in Mesopotamia, Egypt, Greek and China and still play an essential role in nowadays medicine.<sup>5, 6</sup>

The biological effects of plants have long remained a mystery and were connected to a higher power. However, soon the effects were related to active ingredients within the plants. These ingredients, today known as natural products (NPs), are produced by the organism's secondary metabolism. In contrast to primary metabolism, the secondary metabolism is not essential for the growth and reproduction of the organism. Secondary metabolites provide a selective advantage in the constant race of evolution by producing antimicrobials, insecticides, poisons, fragrances, pigments and more. Around 1805 the hunt for biologically active natural products ignited with the isolation of morphine (Fig 1a) from opium by Friedrich Wilhelm Adam Serturner which soon led to the first commercially available painkillers.<sup>7</sup> Since then, NPs from all kinds of sources have been isolated and nature's capability of producing a huge variety of complex structures with a plethora of activities has been revealed. Through the discovery of the antibiotic penicillin (Fig 1b) by Sir Alexander Fleming in 1929, natural product research entered the "golden age" of drug discovery which set focus on microorganisms as source of antibiotics.<sup>8</sup> The "golden age" ended around the 1980s due to the frequent rediscovery of natural products and the rising success

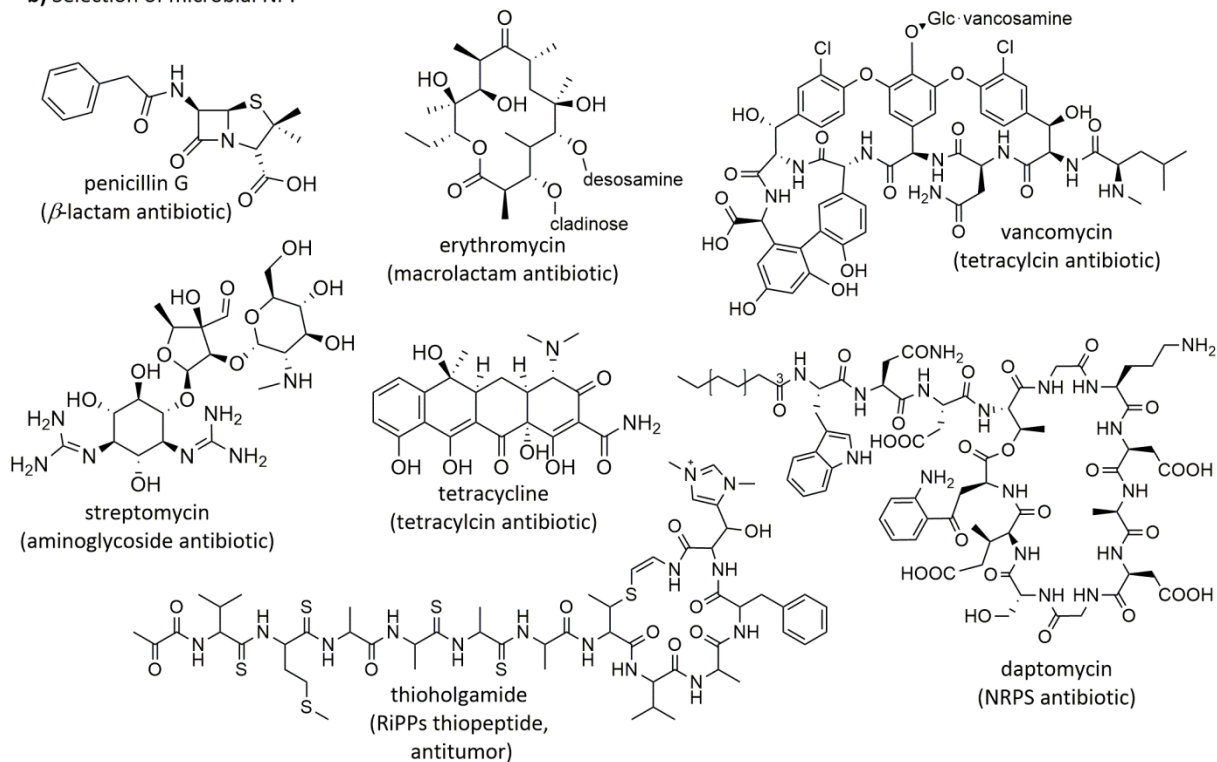
## Introduction

of combinatorial chemistry which enables biological screening of huge chemical libraries. However, this method was moderately successful due to drawbacks of synthetic compounds (Figure 1, c) which are lack of structural diversity, lower bioavailability and more side effects than natural drugs.<sup>9, 10</sup>

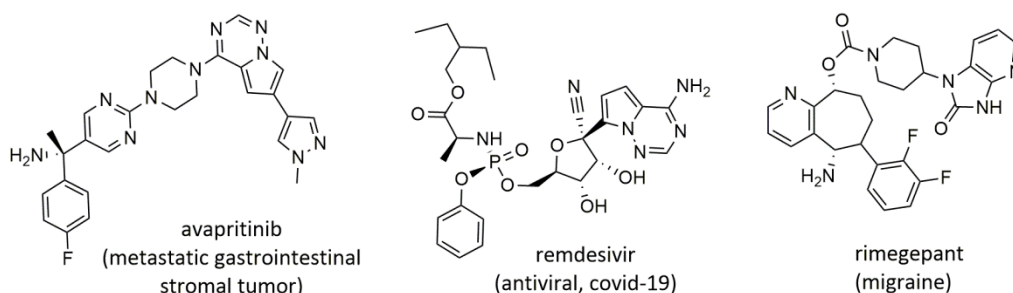
### a) Selection of plant NP:



### b) Selection of microbial NP:



### c) Synthetic drugs (FDA approved 2020):



**Figure 1:** Selection of natural products (NP) a) from plants, b) microorganism and c) synthetic drugs approved by the FDA.

Natural product research was revitalized in the late 20<sup>th</sup> century due to the development of genome sequencing, genetic engineering, mass spectrometry and computer software. Genetic engineering of biosynthetic pathways became an important tool to enable recombinant production of natural products in microbial hosts. Nowadays, the combined potential of recombinant production and chemical synthesis is widely used in the production of semisynthetic compounds with improved biological activities, reduced side-effects and altered mode of action. More than one-third of all drugs approved by the FDA are NPs or NP derivatives.<sup>11</sup> The importance of NPs is undeniable and will continue to play an important role in future drug development to find more effective drugs to fight infectious diseases and to overcome antibiotic resistance

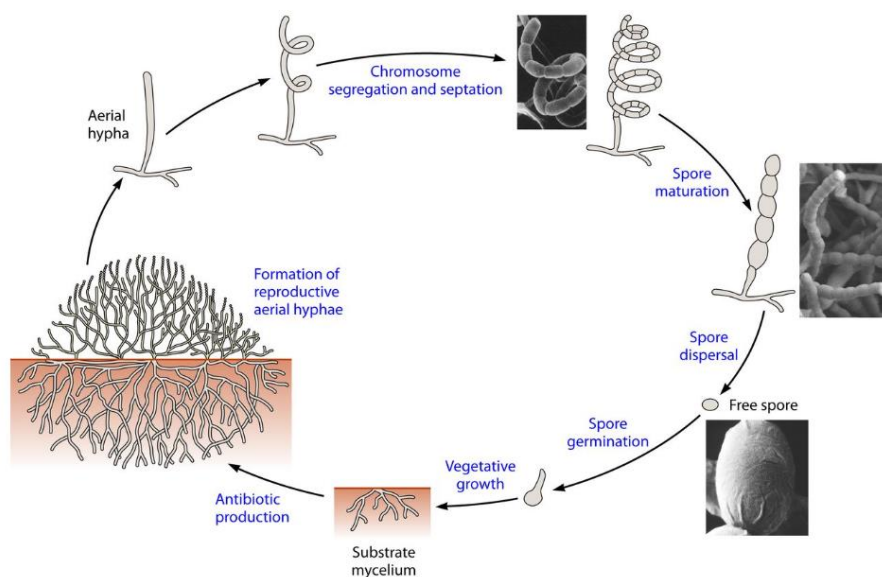
### 1.1.2 Microbial Natural Products

The importance of microbial natural products has become apparent during the “golden age” of drug discovery. Since 1960 microbial NP made up more than half of all NP approved by the FDA while the importance of plants as source of new NPs declined.<sup>12</sup> The rapid success of microbial NPs was induced by their anti-infective activities (Fig 1b) which led to the discovery of highly effective antibiotics including penicillin (1941), streptomycin (1943), erythromycin (1953), tetracycline (1953) and vancomycin (1958).<sup>13-16</sup> As a consequence, from 1940-1980 mortality rates dropped rapidly since infectious diseases became treatable with antibiotics.<sup>13</sup> Actinomycetes largely contributed to the success of microbial NPs by providing more than half of all clinically used classes of antibiotics underlining their importance as source of anti-infective drugs.<sup>17</sup> Actinomycetes are filamentous Gram-positive bacteria with high G+C content in their DNA and can be found in terrestrial or marine environment. They are characterized by the formation of aerial hyphae and their ability to form spores which allow them to survive for an extended period of time in unfavorable conditions (Fig 2).<sup>18</sup>

*Streptomyces* are the largest genus of actinomycetes and a talented producer of complex secondary metabolites which largely contributed to medicine and industry.<sup>19</sup> The large interest in *Streptomyces sp.* was nourished by the constant discovery of fascinating and diverse new natural compounds with a plethora of activities including antimicrobial, immunosuppressant, antitumor, antiparasitic etc.<sup>20</sup> However, the true potential of this genus became apparent after the development of genome sequencing and bioinformatics tools like antiSMASH<sup>21</sup> or PRISM<sup>22</sup>. The first genome sequences of *Streptomyces* strains *S. coelicolor*<sup>23</sup> and *S. avermitilis*<sup>24</sup> revealed a large number of biosynthetic gene clusters (BGCs), exceeding the amount of produced compounds which have previously been discovered. Since then, a major part of NP research has been focusing on the activation of silent BGCs in *Streptomyces* to reveal their full biosynthetic potential.<sup>25, 26</sup> Microbial biosynthetic machineries are as versatile as the compounds they produce; thus, the probability of finding new chemical scaffolds with



unique biological activities is definitely not negligible. Discovery of new compounds and biosynthetic pathways from microorganism paves the way towards more effective drugs and a simplified access through the possibility of recombinant production.



**Figure 2:** Life cycle of sporulating *Actinobacteria*.<sup>18</sup>

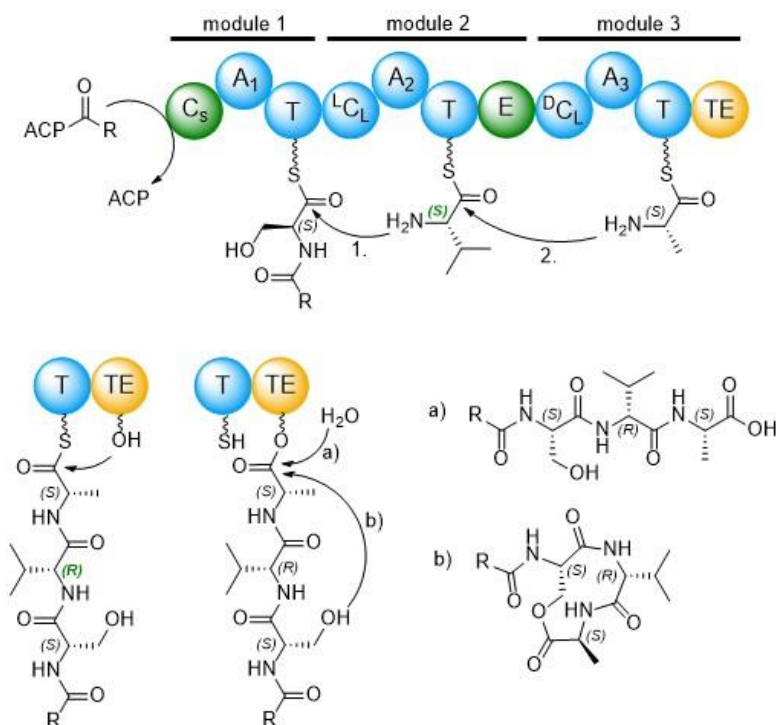
## 1.2 Biosynthetic Machineries

To be able to dissect biosynthetic pathways of new natural products a deep understanding of biosynthetic machineries and of their products is required. The most important compound classes produced by biosynthetic machineries are ribosomally synthesized and posttranslationally modified peptides (RiPPs, e.g. thioholgamide), non-ribosomal peptides (NRP, e.g. daptomycin), polyketides (PKS, e.g. erythromycin), terpenoides (e.g. taxol),  $\beta$ -lactams (penicillins) and aminoglycosides (e.g. streptomycin) (Figure 1). Details about the respective biosynthetic pathways have been described in numerous reviews.<sup>27-32</sup> In the context of this work, the basic principles of peptide, polyketide and fatty acid biosynthesis are briefly summarized.

### 1.2.1 Peptide Biosynthesis

The two biggest classes of peptides are NRPs and RiPPs. The biosynthetic route towards RiPPs starts with the ribosomal synthesis of a precursor peptide consisting of a core (the eventual mature peptide) and a follower peptide which is important for substrate recognition in the followed posttranslational modifications steps. In a final step, the mature peptide is cleaved off by proteolysis from the non-core regions. RiPPs gained a lot of attention due to their potent antibiotic activities and their high

biosynthetic plasticity.<sup>27</sup> They can be divided in more than 20 sub-classes including lasso peptides<sup>33</sup>, lantipeptides<sup>34</sup> and thiopeptides<sup>35</sup>, just to name a few.



**Figure 3:** Illustration of a hypothetical NRPS pathway showing the minimal required domain in blue. A-domains in each module are selective for a specific amino acid: A<sub>1</sub> = serine, A<sub>2</sub> = valine, A<sub>3</sub> = alanine. Module 1 carries an additional starter C domain (C<sub>s</sub>) catalyzing the transfer of an acyl group from an acyl carrier protein (ACP). Module 2 carries an L-L specific C-domain (L<sub>C</sub>L) and an additional E-domain converting (S)-valine into (R)-valine. Module 3 carries a D-L specific C-domain (D<sub>C</sub>L) and TE-domain cleaving off leading a) a linear product by hydrolysis or b) a cyclic product *via* intramolecular reaction with the serine sidechain.

In contrast to RiPPs, non-ribosomal peptides are synthesized by large multi modular mega-synthetases (NRPS). A single module is comprised of a minimal set of domains including adenylation (A) domain, thiolation (T) domain and condensation (C) domain. A-domains select certain amino acids and activate them by a bound ATP co-substrate as aminoacyl-adenylate. The amino acid is then tethered to the HS-pantetheinyl moiety at the T-domain leading an enzyme bound aminoacyl thioester. Peptide bond formation between two amino acids bound to neighboring modules is catalyzed by the condensation (C) domain. C-domains have a high substrate specificity to ensure the correct assembly of the peptide strand. They can be further subdivided into L<sub>C</sub>L (connects two L-amino acids), L<sub>C</sub>D (connects L- and D-amino acids) or starter C-domains (introduction of N-terminal acyl-moieties) (Figure 3).<sup>36</sup> A set of several modules in a row assemble the final peptide strand. In the final termination step, the mature peptide is cleaved off by a thioesterase (TE) domain resulting in a linear (a) or cyclic (b) product (Figure 3). The huge

variety of NRPs is due to modifications during the peptide strand assembly. A-domains are specific for a variety of different substrates and also enable the incorporation of non-proteinogenic amino acids.<sup>37, 38</sup> Further modifications are introduced by additional domains within each module such as epimerase (E), cyclase (Cy), N-methylation (NMT) domains etc.<sup>31</sup> The modular domains are a promising target to perform biosynthetic engineering on NRPS to alter their peptide products.<sup>39</sup>

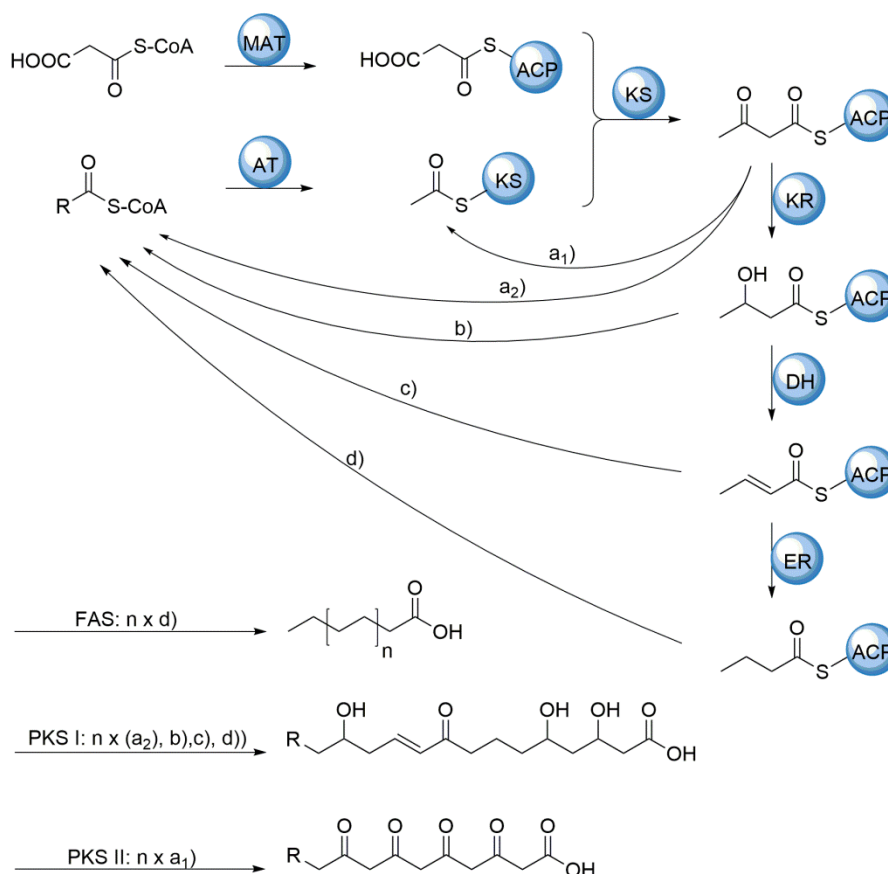
### 1.2.2 Polyketide and Fatty Acid Biosynthesis

Polyketides are a large and diverse NP family with a wide spectrum of activities including antibacterial, antifungal, anticholesterol, antiparasitic, anticancer, and immunosuppressive. The biosynthetic pathways of polyketides share many similarities to saturated fatty acids biosynthesis when comparing precursors, chemistry and architectural design. Polyketide synthases (PKS) and fatty acid synthases (FAS) both use acyl units as building blocks: An acyl starter unit (usually acetyl) is loaded by an acetyl transferase (AT) on the active site cysteine of a keto synthase (KS). Malonyl elongation units are transferred by a malonyl-acetyl transferase (MAT) to a phosphopantetheine of acyl carrier protein (ACP), followed by the decarboxylative condensation with the acyl starter unit catalyzed by KS (Figure 4). PKS typically use acetyl-CoA or propionyl-Co as starter units and malonyl-CoA or methylmalonyl-CoA as monomers for chain elongation, however, more exotic monomers can be also incorporated.<sup>40, 41</sup> PKS can be divided into three sub-types: type I, II and III.

Type I PKS can be further subdivided into iterative (fungal) type I PKS (not reviewed here) and non-iterative (bacterial) type I PKS. Canonical bacterial type I PKS are modular enzymes. Each module is consisting of a minimal set of KS, AT and ACP domains. A module can contain additional ketoreductase (KR), dehydrogenase (DH) and enoylreductase (ER) domains leading different levels of reduction of the carbonyl group in  $\beta$ -position (Figure 4). The mature polyketide is cleaved off by a thioesterase (TE) domain placed in the last module through hydrolytic cleavage or intramolecular reaction. In the last two decades, several non-canonical type I PKS subtypes with unusual arrangements have been identified which illustrates the great structural variability of type I PKS.<sup>42</sup> The variety of different starter and elongation units generates a vast structural diversity and is a valuable target to alter polyketide structures by using pathway engineering.<sup>43</sup>

Type II PKS consist of a minimal set of three enzymes ( $KS_{\alpha}/KS_{\beta}/ACP$ ) which catalyzes polyketide chain formation by iterative decarboxylative condensation using malonyl-CoA extender units (Figure 4). The chain length usually assumes 16, 20 or 24 carbons and is determined by  $KS_{\beta}$ , which is also known as the chain length factor. The polyketide strand is subsequently folded by post PKS tailoring enzymes including ketoreductases, cyclases and aromatases. Further modifications of the folded polyphenol are carried out

by oxygenases, methyltransferases and glycosyltransferases. A deep insight in the enzymatic network has been reviewed by Hertweck *et al.*<sup>44</sup>



**Figure 4:** Overview of the FAS and PKS pathway: MAT (malonyl acyltransferase), ACP (acyl carrier protein), KS (keto synthase), KR (ketoreductase), DH (dehydrogenase), ER (enoylreductase). FAS: complete reduction of keto group; PKS I: modular enzymes with a minimal set of KS, AT, ACP domains and optional KR, DH and ER domains. PKS II: iterative module consisting of KS<sub>α</sub>, KS<sub>β</sub> and ACP.

Type III PKS are self-contained homodimers with two identical KS monomeric domains. Each monomer synthesizes the polyketide strand by iterative priming, extension and cyclisation of the polyketide strand. RppA is the first reported type III PKS from bacteria and a member of five type III PKS subgroups which have been found since the discovery in 1999.<sup>45, 46</sup>

As already indicated above, polyketide synthases are the evolutionary descendant of fatty acid synthases (FAS). Many natural products are modified by fatty acids (FA) e.g. lipopeptides (surfactin). Fatty acid biosynthesis is a potent target for engineering to e.g. increase production titers.<sup>47</sup> Therefore, a precise understanding of FA biosynthesis is necessary. FAS can be sub-divided into the mainly eukaryotic type I FAS and the bacterial type II FAS. FAS I is a single multifunctional enzyme while FAS II synthase

consists of single enzymes which are encoded on separate genes. *Streptomyces* utilise type II FAS to synthesize odd and even numbered linear, *iso* or *anteiso* fatty acids by using acetyl, butyryl, isobutyryl, isovaleryl and anteisovaleryl-CoA as starter units.<sup>48, 49</sup> FAS II share many similarities to PKS II since both are using similar precursors and operate iteratively. However, in FAS the  $\beta$ -keto group is fully reduced after every elongation cycle (Figure 4). Genes coding the enzymes MAT, KSII, KSIII and ACP are clustered within one operon while genes encoding for KR, DH and ER are spread across *Streptomyces* genomes.<sup>48</sup>

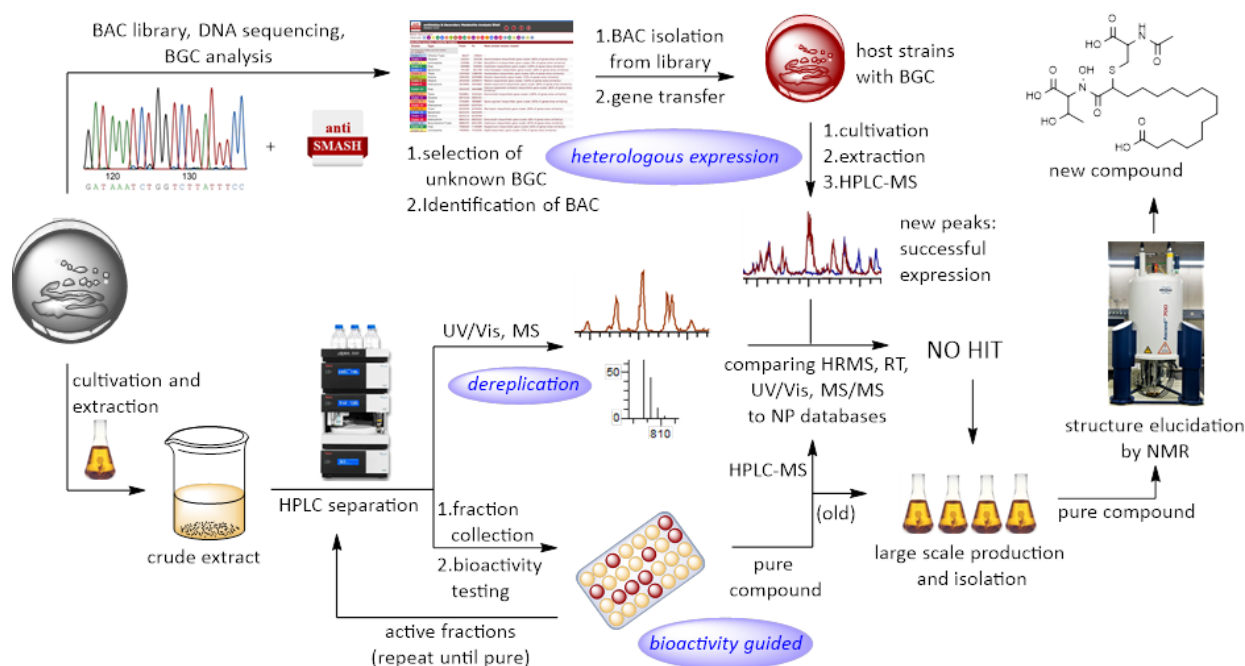
## 1.3 Discovery of New Natural Products and the Corresponding Biosynthetic Gene Clusters from *Streptomyces*

### 1.3.1 Methods for Natural Product Discovery in *Streptomyces*

Since the “golden age” of antibiotic discovery, *Streptomyces* are one of the most important bacterial sources of bioactive natural products. The methods used for mining new bioactive metabolites have changed since then. Bioactivity-guided isolation has been one of the most important methods to identify active compounds from bacterial extracts (Figure 5). An extract of the strain was separated by chromatographic methods and the fractions were screened for a desired activity. The process has been repeated until the pure compound was finally obtained. Various methods were applied to induce the production of new metabolites (one strain many compounds (OSMAC) approach) e.g. through co-culturing, altering cultivation conditions and culture ingredients.<sup>50</sup> This led to the discovery of many new compounds during the “golden age” area, but soon these methods started to suffer from growing rediscovery rates of known natural products. Due to the exchange of biosynthetic gene clusters by horizontal gene transfer similar BGCs can be found across all *Streptomyces* species which explains the high rediscovery rate. The development of dereplication, genome sequencing, genetic engineering and powerful computers for data processing highly improved the NP discovery process and paved the way into a new era of NP research.

The expression “dereplication” includes a variety of techniques to determine the range of metabolites produced by an organism without isolation and structure elucidation (Figure 5). This allows to avoid the rediscovery of known secondary metabolites in an earliest possible stage of the NP discovery process.<sup>51</sup> An analytical separation method coupled to MS detection is the most common procedure to screen for secondary metabolites. Further methods are the direct injection of an extract and analysis by high resolution MS like Fourier-Transform ion cyclotron resonance mass spectrometry (FT-ICR-MS), or dereplication via NMR spectra analysis by using high resolution NMR spectrometry.<sup>52, 53</sup> Bacterial extracts are mostly analyzed by high-performance liquid chromatography electrospray ionization MS

(HPLC-ESI-MS) which provides multiple parameters including high resolution mass, retention time, UV/VIS absorption, MS/MS fragmentation, isotopic pattern and adduct ions to characterize the secondary metabolite profile of a strain. Compounds can be identified by comparisons of the obtained data with one of the many existing natural product data bases.<sup>54</sup>



**Figure 5:** Overview of methods for natural products discovery: activity-guided isolation, dereplication and heterologous expression; BAC = bacterial artificial chromosome; BGC = biosynthetic gene cluster; HPLC = high performance liquid chromatography; MS = mass spectrometry; MS/MS = MS fragmentation analysis; HRMS = high resolution mass spectrometry; RT = retention time; NO HIT = compound not found in database, NP = natural product; NMR = nuclear magnetic resonance spectroscopy.

Dereplication complements many different drug discovery methods including Imaging MS (IMS), high throughput screenings (HTS), or in the omics disciplines.<sup>51</sup> IMS is a method to reveal phenotypes and relevant chemo types of microorganism to understand their behavior in the environment of plants, animals, insects, fungi and other microorganisms. Dereplication of IMS data is particularly useful to discover and understand synergistic effects of antibiotics, find alterations in virulence factor production, explore microbial interactions on a molecular level and to find potent strains for the production of new natural products.<sup>55</sup> HTS is an important tool in the pharmaceutical industry to quickly identify active compound hits by running millions of chemical, genetic or pharmacological tests. HTS of NP libraries is especially challenging due to many bottlenecks in the hit-to-lead development process such as dereplication, isolation/purification and structure elucidation.<sup>56</sup> The key to accelerate hit-to-lead downstream processing is the automation of the rate-determining technologies. Automated

dereplication and computer-assisted structure elucidation will be crucial to restore the economic feasibility of HTS of NP libraries.<sup>57, 58</sup> Additionally, the automated dereplication process can be supported by LC-MS based principle-component-analysis using MS/MS fragmentation to distinguish unique vs. common metabolites which further accelerates the identification process of promising hit compounds.<sup>59</sup> In addition, compound fragmentation in combination with fragmentation prediction tools can accelerate the often difficult *de novo* structure elucidation process of natural products. Dereplication is also required in the genomics-based NP discovery to select potent strains by determining the production of known compounds and to identify new compounds after deletion or activation of silent biosynthetic gene clusters. The combination of dereplication and heterologous expression is a powerful tool for the simultaneous discovery of new compounds and their BGCs.

Since the discovery of the biosynthetic potential of *Streptomyces* strains *S. coelicolor* and *S. avermitilis*, the activation of silent BGCs has become a powerful method to discover new natural products. Expression of silent BGCs is low or absent in standard growth conditions due to transcriptional inhibition, lack of physical or evolutionary stress, lack of precursor supply, weak or absent promoters or mutations etc.<sup>60</sup> Silent BGCs can be activated by a number of approaches: engineering of BGC translation and transcription, manipulation of global regulators or by using e.g. the OSMAC approach.<sup>50, 60-62</sup> Targeted methods are pathway specific manipulations such as introduction of promoters or expression of BGCs in heterologous hosts specialized for secondary metabolites production.<sup>60, 63</sup> Heterologous expression (Figure 5) of silent clusters provides the access to novel chemical entities, high production of known secondary metabolites, and a platform to perform genetic engineering to elucidate NP biosynthetic pathways. The first step towards heterologous expression is the construction and sequencing of a genomic library of cloned BGCs e.g. a BAC-library, cosmid-library etc. BGCs are analyzed using bioinformatics tools and promising candidates are selected based on their similarity to known clusters to assess their potential to produce unknown metabolites. Another important element for the successful expression of metabolites is the introduction of synthetic transcriptional and translational control elements to increase and control metabolite production by using suitable promoters or control elements in the 5'-UTR of a mRNA sequence.<sup>64</sup> The final element is the selection of a suitable and reliable heterologous host. *Streptomyces* are suitable hosts for heterologous expression of secondary metabolites since they are producers of various natural products; hence, they possess an extended network for precursor supply, transporters for excretion, native resistance systems, regulation of enzyme expression and folding.<sup>64</sup> The success rate of the cluster expression can be increased by using optimized heterologous hosts. These optimizations can be archived by insertion of loci for site-specific integration of foreign DNA and by deleting native BGCs to increase precursor supply for the cloned clusters.<sup>65, 66</sup> The expression of new BGCs in the optimized strains *S. albus*  $\Delta$ 14 and *S. lividans*  $\Delta$ 8 and the

discovery of corresponding new compounds have been successfully illustrated with many examples in the literature.<sup>67-71</sup>

### 1.3.2 *De novo* Structure Elucidation by NMR

A major bottleneck in the discovery process of new natural products is structure determination. Amongst other methods including HPLC-MS/MS fragmentation and crystallography, nuclear magnetic resonance (NMR) spectroscopy has become the most important method for structure elucidation of NPs. While UV/VIS and infrared (IR) spectroscopy provide information about functional groups, in NMR spectroscopy the connectivity of atoms and even the spatial arrangement of molecules is shown. Structure elucidation of NPs is especially challenging due to structural complexities and low abundance in nature. NMR fundamentals and special techniques for NPs are summarized in the following course.

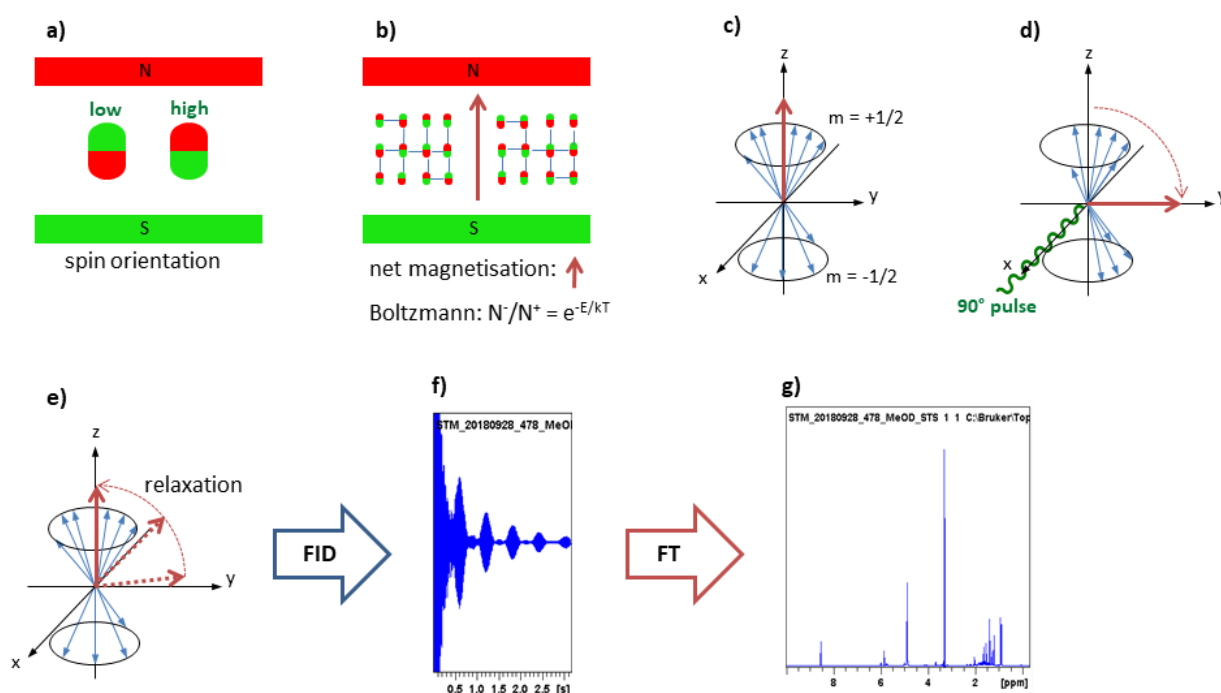
#### 1.3.2.1 *NMR Basics*

Nuclear magnetic resonance occurs when atom nuclei are exposed to a static magnetic field and experience a second oscillating magnetic radio frequency (RF) field. The NMR phenomenon is shown when nuclei possess a certain natural property called spin which comes in multiples of  $\pm 1/2$ . If elements or element isotopes with a spin are naturally abundant in sufficient amounts, NMR experiments can be performed. For *de novo* structure elucidation of natural products  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  nuclei are of special interest since they are the main elements found in NPs. Within the magnetic field, the nuclei experience the Zeeman-Effect which is the energy split of spin levels e.g. for a proton into  $+1/2$  and  $-1/2$ . Another effect is the Larmor precession of the magnetic moment at a distinct Larmor frequency which is dependent on the type of nucleus, the chemical surrounding and the magnetic field strength. If the nuclei experience a photon matching the energy of the Larmor frequency, absorption occurs and the spin is excited between spin levels. This energy transfer is the fundamental phenomena enabling NMR spectroscopy.

To understand the energy absorption-emission process, a nucleus (proton) with a spin can be imagined as a small magnet which can take a high ( $+1/2$ ) and a low energy ( $-1/2$ ) orientation within a magnetic field (Figure 6a). Considering many “magnets” (e.g. all  $^1\text{H}$  within a molecule) in the magnetic field, the lower energy levels slightly outnumber the higher level, a condition described by the Boltzmann statistics (Figure 6b). Within the sample, this results in a net magnetization with precession around the z-axis when it is exposed to the magnetic field (Figure 6c). The occupation difference of both energy levels is very small; hence NMR is a quite insensitive spectroscopic method and requires concentrated samples.



Nowadays NMR signals are obtained by the pulsed Fourier transform experiment (FT-NMR). All spins of a certain nucleus within a sample are excited with a 90° RF-pulse which flips the net magnetization in the x-y plane (Figure 6d). The pulse is containing a range of frequencies broad enough to excite all nuclei within a sample. The frequency range is required since the Larmor frequency of each nucleus within a molecule varies due to the surrounding electrons and is dependent on adjacent electron withdrawing groups promoting shielding or de-shielding from the magnetic field. The absorbed radiation is emitted by the excited spins as a superposition of all excited frequencies. The observable NMR signal is called free induction decay (FID) and is caused by non-equilibrium of the precession after the 90° pulse and decreases over time by an effect called relaxation (Figure 6e, f). Fourier transformation (FT) transfers the time dependent FID into a frequency domain yielding the NMR spectrum (Figure 6g).

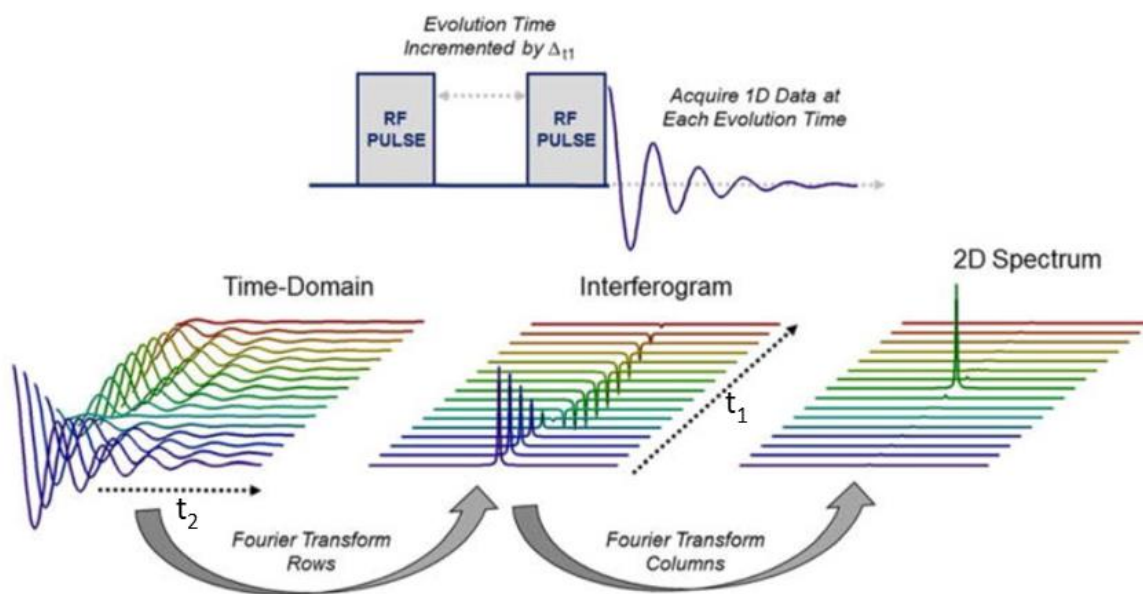


**Figure 6:** NMR spectroscopy principles: a) possible spin orientation in the magnetic field, b) net magnetization of a sample described by the Boltzmann equation, c) net magnetization with Cartesian coordinates, d) net magnetization shifts to x/y plane after 90° pulse, e) net magnetization relaxation, f) time dependent free induction decay and g) Fourier transformation of the FID into the frequency domain.

Shape and position of the peaks provide information about the chemical structure. In dependence of the chemical surrounding of a nucleus within a molecule, the measured Larmor frequencies take distinct values resulting in different chemical shifts. Nuclei within or equal to 3 bond lengths to another nucleus experience the effective field of a neighbored nucleus resulting in scalar spin-spin coupling and peak splitting in the NMR spectrum. The number of coupling nuclei and angle between them can provide

information about their connectivity and spatial orientation. Two or more nuclei in close spatial proximity but without scalar coupling can experience an influence on each other called Nuclear Overhauser Effect (NOE). Permanent RF irradiation at a frequency of a certain nucleus can increase or decrease (depending on the tumbling speed of a molecule) the peak intensity of nuclei within 3-6 Å distance and, thereby, giving information about the 3D structure of a molecule.

For complex NPs the data obtained from 1D NMR experiments is usually insufficient for *de novo* structure elucidation. Additional structural information can be provided by 2D homonuclear ( $^1\text{H}$ - $^1\text{H}$  etc.) or heteronuclear ( $^1\text{H}$ -X etc.) correlations experiments. The 2D spectra are recorded in a direct ( $t_2$ ) and indirect ( $t_1$ ) time domain. The indirect time domain is the evolution time between two pluses and is required at some point in the 2D pulse sequence. A series of FIDs are recorded in dependence of the evolution time  $\Delta t_1$  which is incremented up to a maximum  $t_1$  value. Each FID increment is transformed into the frequency domain  $F_2$  by FT leading an interferogram. The obtained  $F_2$  spectra are Fourier transformed along the  $t_1$  domain resulting in the 2D spectrum (Figure 7).



**Figure 7:** Principle of 2D NMR spectroscopy: a) variation of the evolution time  $\Delta t_1$  between two pulses, b) Fourier transformation (FT) of  $t_2$  increments followed by FT of  $t_1$  increments.<sup>72</sup>

In heteronuclear 2D experiments, correlations of coupling nuclei are obtained through transfer of magnetization by using two pulse sequences in parallel. The most used heteronuclear 2D NMR experiments are the “heteronuclear single quantum correlation” (HSQC) experiment that is showing short range  $^1J_{\text{HX}}$  correlations ( $\text{CH}_\text{x}$  or  $\text{NH}_\text{x}$ ) and the “heteronuclear multiple bond correlation” (HMBC)

experiment that is showing long range  $^2J_{\text{HX}}$  and  $^3J_{\text{HX}}$  correlations. Coupling, chemical shifts and 2D correlation data are very important to determine NP structures. However, in some cases a more specialized experimental setup is required to elucidate challenging NP structures.

### 1.3.2.2 Challenges in Natural Product NMR

The greatest challenge in the structure elucidation process of new NPs is their complexity and low production rate. NMR spectroscopy is an insensitive method; however, advances in software and hardware development enable structure elucidation with sample concentrations in a low mM range.

The constant development of stronger magnets has largely contributed to the field of NMR spectroscopy. The Larmor frequency of a nucleus increases proportionally to the magnetic field strength. Therefore, at higher magnetic fields the RF frequency must be increased in order to match the required Larmor frequency to enable spin transition (required Larmor frequency for  $^1\text{H}$  nuclei: 500 MHz at 11.7 Tesla; 800 MHz at 18.7 Tesla). The increased transition energy results in an improved signals-to-noise (S/N) ratio by a factor of 2 when switching from 500 MHz to 800 MHz NMR machines.<sup>73</sup> The NMR probe is another influencing factor and has to be considered when choosing certain experiments. Broad band probes are optimized for X-nuclei ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ) and provide a better sensitivity for diluted samples when only the X-nucleus is observed (e.g.  $^{13}\text{C}$  NMR). Invers broadband probes are optimized for  $^1\text{H}$  nuclei and provide better sensitivity for  $^1\text{H}$ -X experiments ( $^1\text{H}$ -X-HSQC,  $^1\text{H}$ -X-HMBC, etc.). An enormous sensitivity boost is obtained by cryoprobes.<sup>74</sup> Most of the noise is generated by the electronics and can be reduced by cooling the system using helium or nitrogen leading an S/N increase of a factor 2-3.

A major sensitivity boost is generated by enhanced acquisition methods. The analog FID signal has to be transformed into a digital signal to analyze and process the recorded data. The analog signal is divided into time slots which convert the continuous FID signal into a discrete FID signal by sampling. By using uniform sampling, a regular pattern is acquired resulting in the Nyquist sampling grid. Especially for 2D experiments, this method is rather time-consuming when recording each FID in the indirect time domain  $t_1$ . High signal density of complex NPs and diluted samples require a large number of scans and  $\Delta t_1$  increments to get sufficient resolution and signals strength. Nonuniform sampling (NUS) acquires an irregular pattern with less sampling points (NUS points) which reduces the measurement time by a large factor. The number of  $\Delta t_1$  increments in a 2D experiment is reduced and the missing points are reconstructed by algorithms. The time saved can be reinvested into more scans to enhance S/N of diluted samples. Transformation of NUS spectra in the frequency domain requires special algorithms to reconstruct the compressed spectra. The minimum selection of NUS points is limited by these algorithms and by the number of significant peaks in a spectrum which should match the number of NUS points.<sup>72</sup> Another useful application of NUS is the enhancement of peak resolution. Setting a high

number in the indirect time domain  $\Delta t_1$  on e.g. 2048 and 25% NUS resulting in 512 NUS points will not reduce the measurement time of a standard 2D experiment, but increase resolution by a significant factor.

Resolution is common challenge when observing many nuclei with similar chemical shifts. Prominent examples are methine signals of sugars in poly-glycosylated compounds, carbonyl groups of amide bonds in peptides or multi-compound mixtures. Band selective 2D (HMBC, HSQC) and selective 1D (NOESY, ROESY, COSY, TOCSY) experiments can help to increase resolution of certain areas of a spectrum and simplify signal interpretation.<sup>75</sup> In 2D experiments, a certain area of interest is selected and a band selective pulse excites all frequencies in this area. The number of  $\Delta t_1$  increments is applied on a much narrower range which results in a higher resolution. Selective 1D experiments are particularly useful in crowded spectra with many signals. Selective pulses can be used to exclusively excite one nucleus of interest to isolate and analyze a spin system by COSY/TOCSY experiments or to obtain the spatial surrounding by NOESY/ROESY experiments.

The determination of the absolute configuration of natural products is highly important in the drug development process. Natural products usually contain many stereocenters which are crucial for their biological activity. The determination of the stereochemistry *via* NMR by using the NOE or coupling constants is sometimes not possible. In this case derivatization experiments can be used. Some biosynthetic machineries like NRPSs are able to invert the L-configured  $\alpha$ -CH of proteinogenic amino acids into a D-configuration. Most amino acids have a flexible side chain which prevents the stereochemical assignment by NMR. Therefore, Marfey *et al.* developed a method which is using the chiral derivatization agent L- and D-FDLA ( $N^\alpha$ -(5-Fluoro-2,4-dinitrophenyl)-D/L-leucinamide) to generate amino acid diastereomers which can be separated by standard reversed phase chromatography.<sup>76</sup> In a standard protocol peptides are hydrolyzed and the single amino acids are derivatized by L- or D-FDLA and run against the amino acid references. The  $\alpha$ -CH configuration is determined by comparisons of retention times of the derivatized hydrolysate and the references. To determine chiral alcohols and amines, the Mosher's method which uses S- and R-MTPA (3,3,3-trifluoro-2-methoxy-2-phenylpropanoic acid) as a derivatization agent is a suitable option.<sup>77</sup> The relative position of the proton at the chiral center and the condensed S- and R-MTPA leads to shift alterations of the proton signal correlating with the configuration of the alcohol or amine. The precise absolute configuration is highly important in the further characterization of BGCs and the determination of biosynthetic pathways.

### 1.3.3 Elucidation of Natural Product Biosynthesis

The following methods are usually applied to determine NP biosynthesis: Feeding with labelled precursors, gene knockout experiments, *in vivo/in vitro* studies of enzymes and structural studies of enzymes by x-ray crystallography.<sup>78</sup>

Precursor labelling provides an initial insight into the biosynthesis of complex NP scaffolds and moieties. Furthermore, it can accelerate the genome mining process to identify BGCs when a new NP is identified by dereplication or bioactivity-guided isolation. The producer strain is fed with pathway specific building blocks such as acetate ([D], [<sup>13</sup>C]-labelled) for polyketide or amino acids ([D], [<sup>13</sup>C], [<sup>15</sup>N]-labelled) for peptides etc. Successful incorporation into NPs can be verified by alteration from the original compound mass *via* HPLC-MS or by analyzing signal strengths of <sup>1</sup>H, <sup>13</sup>C or <sup>15</sup>N NMR signals. For instance, by this method Gottardi *et al.* revealed that the antibiotic abyssomicin derived from five acetates, two propionates and unexpectedly one metabolite from the glycolytic pathway.<sup>79</sup>

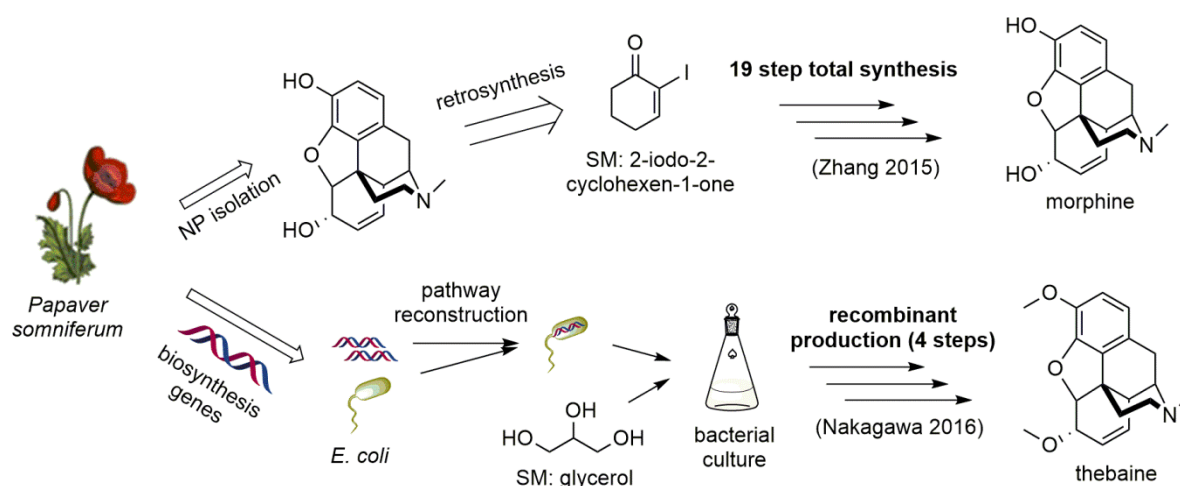
Genetic manipulation in *Streptomyces sp.* to disrupt genes or edit genes to determine their function is typically done by homologous recombination or by DNA double strand break (DSB) repair mechanisms (e.g. CRISPR/Cas9).<sup>80</sup> An alternative way to determine gene functions, especially after successful heterologous expression of a BGC, is  $\lambda$ -Red recombination which allows highly effective substitution of genes in *E. coli*.<sup>81</sup> A resistance cassette, which can be used for clone selection of *Streptomyces* and *E. coli*, is multiplied by PCR amplification using primers with 36-50 bp overhang regions matching the 3' and 5' borders of a chromosomal fragment to be deleted. A BAC is transformed into *E. coli* carrying the  $\lambda$ -Red recombination plasmid. The  $\lambda$ -Red recombinase genes are induced and the PCR product is transformed into *E. coli*. After successful recombination, the construct is isolated and controlled by restriction analysis, PCR or sequencing. The edited BAC is then transferred into the heterologous *Streptomyces* host *via* conjugation and the metabolic profile is compared to the one of the heterologous host with the unmodified BGC. Eventual NP modifications can be monitored by HPLC-MS. Ideally, the results from the deletion experiments will provide information about the biosynthesis. However, interpretation of the data is often difficult and unambiguous. Therefore, *in vitro* studies and x-ray analysis of the isolated proteins or protein complexes need to be performed to precisely understand biosynthesis and enzyme mechanics. *In vitro* reconstruction of enzyme machineries supported by structural biology methods like crystallography can give insight into substrate specificities of enzymes and the exact reaction mechanism which is essential to understand new enzyme mechanics.

### 1.3.4 Utilization of New Biosynthetic Gene Clusters and Chemical Scaffolds in Natural Product Synthesis - Chemists vs Microorganism

The development of a natural product into a drug often fails due to supply issues. Secondary metabolites are often produced in very low quantities, which is enough to perform structure elucidation and bioactivity studies, but too less for preclinical/clinical studies and industrial production.<sup>82</sup>

To solve this problem, one approach is the chemical total synthesis of natural products (Figure 8). The total synthesis of highly complex NP scaffolds is very challenging and requires many steps with an overall low yield which is usually not ideal for upscaling.<sup>83</sup> However, despite their complexity some commercial total synthesis approaches became economically feasible due to the successful increase of total yields into the gram range (e.g. eribulin, 62 step synthesis).<sup>84</sup>

Another approach is the recombinant production of secondary metabolites in microorganisms (Figure 8).<sup>85</sup> In comparison to total synthesis, the production in recombinant microorganisms generates less toxic waste, is more cost efficient, and requires less harsh production conditions and time. The initial step towards recombinant production of natural products in microorganism is the discovery and dissection of biosynthetic gene clusters (BGC) from the natural source. The genetic material from the foreign, genetically distant host usually needs to be engineered in order to achieve successful expression of the BGC in the heterologous host. These adjustments include codon optimization, promotor introduction and improving solubility of the expressed proteins.<sup>82</sup> The engineered BGC can then be transferred in the heterologous host resulting in the recombinant producer organism which is producing the desired compound.



**Figure 8:** Natural product (NP) synthesis methods: Total synthesis (19 steps) of morphine by Zhang *et al.*<sup>86</sup> and stepwise recombinant production of thebaine in *E. coli* by Nakagawa *et al.*<sup>87</sup>

The reconstruction of biosynthetic pathways in recombinant microorganisms for the production of secondary metabolites is a challenging task and sometimes remains just partly successful. Furthermore, generation of NP derivatives is limited by the number of biosynthetic enzymes and flexibility of the biosynthetic machineries. In chemical synthesis on the other hand, structurally diverse NPs derivatives can more easily be generated by using different starting materials and a plethora of reactants which enables the quick access to huge chemical libraries.<sup>88</sup> In combination, both approaches can complete each other in a useful way. NPs are designed by nature; hence they are privileged binding partners for various enzymes and a great foundation for protein inhibitors.<sup>89</sup> Therefore, chemical synthesis of compound libraries for high throughput screening is often based on chemical scaffolds from bioactive NPs.<sup>90</sup> Furthermore, chemical modifications of naturally synthesized NPs can alter their activities and biological targets which is important to overcome antibiotic resistance.<sup>91</sup>

The identification and utilization of enzyme catalysts in chemical synthesis is another benefit resulting from the field of chemical biology. Catalysis in chemical reactions is highly important to increase reaction yields, improve enantioselectivity and to develop new synthesis routes.<sup>92</sup> The increasing necessity of green chemistry has set the focus on finding alternatives for the often toxic catalysts.<sup>93</sup> Since the development of genome editing methods and enzyme engineering, enzyme catalysis has become a valuable alternative for the classical catalysts.<sup>94</sup> Enzymes operate at room temperature and mild conditions while having low toxicity. High specificity of enzymes towards substrates allows late stage modifications which is especially useful in NP semisynthesis.<sup>95</sup> Late stage modifications such as glycosylations are easily executed by enzymes, but are complex to access by synthesis. For instance, C-glycosylation of flavones requires a multistep synthetic route, while the enzymatic reaction with the enzyme UDP glucuronosyltransferase can be done in one step.<sup>96, 97</sup> Enzyme engineering allows to improve limiting factors like enzyme stability, substrate range and products/substrate inhibition and helped to confirm enzyme catalysts as a valuable replacement for the classical chemical catalyst.<sup>98</sup>

The development of recombinant microorganism, semisynthetic drugs and new enzyme catalysts is nourished by the deep understanding about natural product biosynthesis. Discovery of new secondary metabolites and the dissection of biosynthetic pathways are necessary to enrich our set of tools needed to enable recombinant production of natural products.

### 1.4 Scope of the Thesis

The importance of natural products has clearly been demonstrated. Drug development based on natural products has and will decisively contribute to fight antibiotic resistance and diseases. State of the art analytical tools and methods for genetic engineering have opened up a toolbox of new opportunities in the process of finding new natural products. Sustainable and sufficient production of natural products in recombinant microorganisms needs to be a future goal to save resources and to prevent poisoning of our planet. Therefore, extending our knowledge about biosynthetic pathways and to find new biosynthetic machineries and enzymes is crucial to use microorganisms as “biofactories”. *Streptomyces* have become the most important genus of *Actinobacteria* in the production of anti-infectives. The true potential to produce unique NPs has long remained hidden in the genome of *Streptomyces* in the form of silent biosynthetic gene clusters. Herein, by using dereplication and heterologous expression, the biosynthetic potential of the less intensively studies *Streptomyces* strains *S. aurantiacus* LU19075 and *S. aureus* LU1811 was explored. *De novo* structure elucidation of new compounds from *Streptomyces*, the identification of corresponding gene clusters and dissection of biosynthetic pathways were the main objectives of this work.



## References

1. Subbarayappa, B. The roots of ancient medicine: an historical outline. *J. Biosci. (Bangalore)* **2001**, 26 (2), 135-143, doi.org/10.1007/BF02703637.
2. Gurib-Fakim, A. Medicinal plants: traditions of yesterday and drugs of tomorrow. *Molecular aspects of Medicine* **2006**, 27 (1), 1-93, 10.1016/j.mam.2005.07.008.
3. Nelson, L.S.; et al.. *Goldfrank's toxicologic emergencies*; McGraw-Hill Education New York, NY, 2019.
4. Lee, M.R. Curare: the South American arrow poison. *J R Coll Physicians Edinb* **2005**, 35 (1), 83-92.
5. Newman, D.J.; et al.. The influence of natural products upon drug discovery. *Natural product reports* **2000**, 17 (3), 215-234, 10.1039/A902202C.
6. Fabricant, D.S.; et al.. The value of plants used in traditional medicine for drug discovery. *Environ. Health Perspect.* **2001**, 109 (suppl 1), 69-75, doi.org/10.1289/ehp.01109s169.
7. Huxtable, R.J.; et al.. The isolation of morphine—first principles in science and ethics. *Mol. Interventions* **2001**, 1 (4), 189.
8. Fleming, A. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. *British journal of experimental pathology* **1929**, 10 (3), 226.
9. Von Bubnoff, A. Seeking new antibiotics in nature's backyard. *Cell* **2006**, 127 (5), 867-869, 10.1016/j.cell.2006.11.021.
10. Nisar, B.; et al.. Comparison of medicinally important natural products versus synthetic drugs-a short commentary. *Nat. Prod. Chem. Res* **2018**, 6 (2), 308, 10.4172/2329-6836.1000308.
11. Newman, D.J.; et al.. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, 83 (3), 770-803, 10.1021/acs.jnatprod.9b01285.
12. Patridge, E.; et al.. An analysis of FDA-approved drugs: natural products and their derivatives. *Drug Discov. Today* **2016**, 21 (2), 204-207, 10.1016/j.drudis.2015.01.009.
13. Armstrong, G.L.; et al.. Trends in infectious disease mortality in the United States during the 20th century. *Jama* **1999**, 281 (1), 61-66, 10.1001/jama.281.1.61.
14. Bunch, R.L.; et al.. Erythromycin, its salts, and method of preparation. Google Patents: 1953.
15. Conover, L.; et al.. Terramycin. XI. Tetracycline. *J. Am. Chem. Soc.* **1953**, 75 (18), 4622-4623.
16. Anderson, R.; et al.. Symposium: how a drug is born. *Cincinnati Journal of Medicine* **1961**, 42, 49-60.
17. Bérdy, J. Thoughts and facts about antibiotics: where we are now and where we are heading. *The Journal of antibiotics* **2012**, 65 (8), 385-395, doi.org/10.1038/ja.2012.27.
18. Barka, E.A.; et al.. Taxonomy, physiology, and natural products of Actinobacteria. *Microbiol. Mol. Biol. Rev.* **2016**, 80 (1), 1-43, 10.1128/MMBR.00019-15.
19. Hopwood, D.A. *Streptomyces in nature and medicine: the antibiotic makers*; Oxford University Press, 2007.
20. Behal, V. Bioactive products from Streptomyces. *Advances in Applied Microbiology, Vol 47* **2000**, 47, 113-156, 10.1016/S0065-2164(00)47003-6.
21. Blin, K.; et al.. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **2017**, 45 (W1), W36-W41, 10.1093/nar/gkx319.
22. Skinnider, M.A.; et al.. PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Res.* **2017**, 45 (W1), W49-W54, 10.1093/nar/gkx320.
23. Bentley, S.D.; et al.. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3 (2). *Nature* **2002**, 417 (6885), 141-147, doi.org/10.1038/417141a.

24. Ikeda, H.; *et al.* Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat. Biotechnol.* **2003**, *21* (5), 526-531, 10.1038/nbt820.
25. Horbal, L.; *et al.* Secondary metabolites overproduction through transcriptional gene cluster refactoring. *Metab. Eng.* **2018**, *49*, 299-315, 10.1016/j.ymben.2018.09.010v.
26. Piepersberg, W. Pathway engineering in secondary metabolite-producing actinomycetes. *Crit. Rev. Biotechnol.* **1994**, *14* (3), 251-285, doi.org/10.3109/07388554409079835.
27. Hudson, G.A.; *et al.* RiPP antibiotics: biosynthesis and engineering potential. *Curr. Opin. Microbiol.* **2018**, *45*, 61-69, 10.1016/j.mib.2018.02.010.
28. Finking, R.; *et al.* Biosynthesis of nonribosomal peptides. *Annu. Rev. Microbiol.* **2004**, *58*, 453-488, 10.1146/annurev.micro.58.030603.123615.
29. Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **2003**, *7* (2), 285-295, 10.1016/S1367-5931(03)00020-6.
30. Ajikumar, P.K.; *et al.* Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Mol. Pharm.* **2008**, *5* (2), 167-190, 10.1021/mp700151b.
31. Süssmuth, R.D.; *et al.* Nonribosomal peptide synthesis—principles and prospects. *Angewandte Chemie International Edition* **2017**, *56* (14), 3770-3821, 10.1002/anie.201609079.
32. Staunton, J.; *et al.* Polyketide biosynthesis: a millennium review. *Natural product reports* **2001**, *18* (4), 380-416, 10.1039/A909079G.
33. Kodani, S.; *et al.* How to harness biosynthetic gene clusters of lasso peptides. *Journal of Industrial Microbiology & Biotechnology: Official Journal of the Society for Industrial Microbiology and Biotechnology* **2020**, *47* (9-10), 703-714, 10.1007/s10295-020-02292-6.
34. Knerr, P.J.; *et al.* Discovery, biosynthesis, and engineering of lantipeptides. *Annu. Rev. Biochem.* **2012**, *81*, 479-505, 10.1146/annurev-biochem-060110-113521.
35. Bagley, M.C.; *et al.* Thiopeptide antibiotics. *Chemical reviews* **2005**, *105* (2), 685-714, 10.1021/cr0300441.
36. Rausch, C.; *et al.* Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.* **2007**, *7* (1), 78, doi.org/10.1186/1471-2148-7-78.
37. Miethke, M.; *et al.* Inhibition of aryl acid adenylation domains involved in bacterial siderophore synthesis. *The FEBS journal* **2006**, *273* (2), 409-419, 10.1111/j.1742-4658.2005.05077.x.
38. Flissi, A.; *et al.* Norine: update of the nonribosomal peptide resource. *Nucleic Acids Res.* **2020**, *48* (D1), D465-D469, 10.1093/nar/gkz1000.
39. Baltz, R.H. Combinatorial biosynthesis of cyclic lipopeptide antibiotics: a model for synthetic biology to accelerate the evolution of secondary metabolite biosynthetic pathways. *ACS synthetic biology* **2014**, *3* (10), 748-758, 10.1021/sb3000673.
40. Chan, Y.A.; *et al.* Biosynthesis of polyketide synthase extender units. *Natural product reports* **2009**, *26* (1), 90-114, 10.1039/B801658P.
41. Ray, L.; *et al.* Recent advances in the biosynthesis of unusual polyketide synthase substrates. *Natural product reports* **2016**, *33* (2), 150-161, 10.1039/C5NP00112A
42. Chen, H.; *et al.* Iterative polyketide biosynthesis by modular polyketide synthases in bacteria. *Appl. Microbiol. Biotechnol.* **2016**, *100* (2), 541-557, 10.1007/s00253-015-7093-0.
43. Barajas, J.F.; *et al.* Engineered polyketides: synergy between protein and host level engineering. *Synthetic and Systems Biotechnology* **2017**, *2* (3), 147-166, 10.1016/j.synbio.2017.08.005.
44. Hertweck, C.; *et al.* Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Natural product reports* **2007**, *24* (1), 162-190, 10.1039/b507395m.

45. Katsuyama, Y.; *et al.* Type III polyketide synthases in microorganisms. *Methods Enzymol.* **2012**, *515*, 359-377, 10.1016/B978-0-12-394290-6.00017-3.
46. Yu, D.; *et al.* Type III polyketide synthases in natural product biosynthesis. *IUBMB life* **2012**, *64* (4), 285-295, 10.1002/iub.1005.
47. Wu, Q.; *et al.* Systematically engineering the biosynthesis of a green biosurfactant surfactin by *Bacillus subtilis* 168. *Metab. Eng.* **2019**, *52*, 87-97, 10.1016/j.ymben.2018.11.004.
48. Gago, G.; *et al.* Fatty acid biosynthesis in actinomycetes. *FEMS Microbiol. Rev.* **2011**, *35* (3), 475-497, 10.1111/j.1574-6976.2010.00259.x.
49. Kaneda, T. Iso-and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. *Microbiol. Mol. Biol. Rev.* **1991**, *55* (2), 288-302.
50. Wei, H.; *et al.* OSMAC (one strain many compounds) approach in the research of microbial metabolites--a review. *Wei sheng wu xue bao= Acta microbiologica Sinica* **2010**, *50* (6), 701-709.
51. Gaudêncio, S.P.; *et al.* Dereplication: racing to speed up the natural products discovery process. *Natural product reports* **2015**, *32* (6), 779-810, 10.1039/C4NP00134F
52. Han, F.; *et al.* Comparative study of direct injection analysis and liquid chromatography mass spectrometry for identification of chemical constituents in Kudiezi injection by FT-ICR MS. *International Journal of Mass Spectrometry* **2016**, *405*, 32-38, 10.1016/j.ijms.2016.05.016.
53. Williams, R.B.; *et al.* Dereplication of natural products using minimal NMR data inputs. *Organic & biomolecular chemistry* **2015**, *13* (39), 9957-9962, 10.1039/C5OB01713K
54. Sorokina, M.; *et al.* On the redundancy of natural products public databases and where to find data in 2020-a review on natural products databases. **2019**, 10.20944/preprints201912.0332.v1.
55. Shih, C.-J.; *et al.* Bringing microbial interactions to light using imaging mass spectrometry. *Natural product reports* **2014**, *31* (6), 739-755, 10.1039/C3NP70091G
56. Koehn, F.E. High impact technologies for natural products screening. *Natural Compounds as Drugs Volume I* **2008**, 175-210, 10.1007/978-3-7643-8117-2\_5.
57. Elyashberg, M.; *et al.* Computer Assisted Structure Elucidation (CASE): Current and future perspectives. *Magnetic Resonance in Chemistry* **2021**, *59* (7), 669-690, 10.1002/mrc.5115.
58. Bertrand, S.; *et al.* Successes and pitfalls in automated dereplication strategy using liquid chromatography coupled to mass spectrometry data: A CASMI 2016 experience. *Phytochemistry letters* **2017**, *21*, 297-305, 10.1016/j.phytol.2016.12.025.
59. Chanana, S.; *et al.* Natural product discovery using planes of principal component analysis in R (PoPCAR). *Metabolites* **2017**, *7* (3), 34, 10.3390/metabo7030034.
60. Rutledge, P.J.; *et al.* Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nature reviews microbiology* **2015**, *13* (8), 509-523, 10.1038/nrmicro3496.
61. Wei, J.; *et al.* Regulation of antibiotic biosynthesis in actinomycetes: perspectives and challenges. *Synthetic and systems biotechnology* **2018**, *3* (4), 229-235, 10.1016/j.synbio.2018.10.005.
62. Niu, G.; *et al.* Specialised metabolites regulating antibiotic biosynthesis in *Streptomyces* spp. *FEMS Microbiol. Rev.* **2016**, *40* (4), 554-573, 10.1093/femsre/fuw012.
63. Myronovskyi, M.; *et al.* Native and engineered promoters in natural product discovery. *Natural product reports* **2016**, *33* (8), 1006-1019, 10.1039/C6NP00002A
64. Myronovskyi, M.; *et al.* Heterologous production of small molecules in the optimized *Streptomyces* hosts. *Natural product reports* **2019**, *36* (9), 1281-1294, 10.1039/C9NP00023B.
65. Ahmed, Y.; *et al.* Engineering of *Streptomyces lividans* for heterologous expression of secondary metabolite gene clusters. *Microbial cell factories* **2020**, *19* (1), 1-16, 10.1186/s12934-020-1277-8.

66. Myronovskyi, M.; *et al.* Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. *Metab. Eng.* **2018**, *49*, 316-324, 10.1016/j.ymben.2018.09.004.
67. Lasch, C.; *et al.* Dudomycins: New Secondary Metabolites Produced after Heterologous Expression of an Nrps Cluster from *Streptomyces albus* ssp. *Chlorinus* Nr1 B-24108. *Microorganisms* **2020**, *8* (11), 1800, 10.3390/microorganisms8111800.
68. Gummerlich, N.; *et al.* Targeted Genome Mining—From Compound Discovery to Biosynthetic Pathway Elucidation. *Microorganisms* **2020**, *8* (12), 2034, 10.3390/microorganisms8122034.
69. Lasch, C.; *et al.* Loseolamycins: A Group of New Bioactive Alkylresorcinols Produced after Heterologous Expression of a Type III PKS from *Micromonospora endolithica*. *Molecules* **2020**, *25* (20), 4594, 10.3390/molecules25204594.
70. Rodríguez Estévez, M.; *et al.* Benzantronic acid, a novel metabolite from *Streptomyces albus* Del14 expressing the nybomycin gene cluster. *Frontiers in chemistry* **2020**, *7*, 896, 10.3389/fchem.2019.00896.
71. Shuai, H.; *et al.* Identification of a biosynthetic gene cluster responsible for the production of a new pyrrolopyrimidine natural product—Huimycin. *Biomolecules* **2020**, *10* (7), 1074, 10.3390/biom10071074.
72. Delaglio, F.; *et al.* Non-uniform sampling for all: more NMR spectral quality, less measurement time. *American pharmaceutical review* **2017**, *20* (4).
73. Webb, A. Increasing the sensitivity of magnetic resonance spectroscopy and imaging. *Analytical chemistry* **2012**, *84* (1), 9-16, 10.1021/ac201500v.
74. Styles, P.; *et al.* The first cryoprobe—Some recollections. *Journal of magnetic resonance (San Diego, Calif.: 1997)* **2011**, *213* (2), 355-356, 10.1016/j.jmr.2011.08.013
75. Alexandersson, E.; *et al.* Band-selective NMR experiments for suppression of unwanted signals in complex mixtures. *RSC Advances* **2020**, *10* (54), 32511-32515, 10.1039/D0RA06828D.
76. Fujii, K.; *et al.* Further application of advanced Marfey's method for determination of absolute configuration of primary amino compound. *Tetrahedron Lett.* **1998**, *39* (17), 2579-2582, 10.1016/S0040-4039(98)00273-1.
77. Kusumi, T.; *et al.* The modified Mosher's method and the sulfoximine method. *Bulletin of the Chemical Society of Japan* **2006**, *79* (7), 965-980, 10.1246/bcsj.79.965.
78. Kishimoto, S.; *et al.* Elucidation of biosynthetic pathways of natural products. *The Chemical Record* **2017**, *17* (11), 1095-1108, 10.1002/tcr.201700015.
79. Gottardi, E.M.; *et al.* Abyssomicin biosynthesis: formation of an unusual polyketide, antibiotic-feeding studies and genetic analysis. *ChemBioChem* **2011**, *12* (9), 1401, 10.1002/cbic.201100172.
80. Cobb, R.E.; *et al.* High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS synthetic biology* **2015**, *4* (6), 723-728, 10.1074/jbc.M113.510040.
81. Gust, B.; *et al.* Lambda red-mediated genetic manipulation of antibiotic-producing *Streptomyces*. *Adv. Appl. Microbiol.* **2004**, *54*, 107-128, 10.1016/S0065-2164(04)54004-2.
82. Dziggel, C.; *et al.* Tools of pathway reconstruction and production of economically relevant plant secondary metabolites in recombinant microorganisms. *Biotechnology journal* **2017**, *12* (1), 1600145, 10.1002/biot.201600145.
83. Wong, H.N. *Efficiency in natural product total synthesis*; John Wiley & Sons, 2018.
84. Melvin, J.Y.; *et al.* From micrograms to grams: scale-up synthesis of eribulin mesylate. *Natural product reports* **2013**, *30* (9), 1158-1164, 10.1039/C3NP70051H.

85. Pham, J.V.; *et al.* A review of the microbial production of bioactive natural products and biologics. *Frontiers in microbiology* **2019**, *10*, 1404, doi.org/10.3389/fmicb.2019.01404.
86. Li, Q.; *et al.* Total synthesis of codeine. *Chemistry—A European Journal* **2015**, *21* (46), 16379-16382, 10.1002/chem.201503594.
87. Nakagawa, A.; *et al.* Total biosynthesis of opiates by stepwise fermentation using engineered *Escherichia coli*. *Nature communications* **2016**, *7* (1), 1-8, 10.1038/ncomms10390.
88. Liu, R.; *et al.* Combinatorial chemistry in drug discovery. *Curr. Opin. Chem. Biol.* **2017**, *38*, 117-126, 10.1016/j.cbpa.2017.03.017.
89. Balamurugan, R.; *et al.* Design of compound libraries based on natural product scaffolds and protein structure similarity clustering (PSSC). *Molecular BioSystems* **2005**, *1* (1), 36-45, 10.1039/B503623B
90. Barnes, E.C.; *et al.* The use of isolated natural products as scaffolds for the generation of chemically diverse screening libraries for drug discovery. *Natural product reports* **2016**, *33* (3), 372-381, 10.1039/C5NP00121H.
91. Rolinson, G.; *et al.* Semisynthetic penicillins. *Advances in Pharmacology* **1973**, *11*, 151-220, 10.1016/S1054-3589(08)60458-5.
92. Derouane, E.G. Catalysis in the 21st Century, Lessons from the Past, Challenges for the Future. *Cattech* **2001**, *5* (4), 214-225, 10.1023/A:1014036813345.
93. Erythropel, H.C.; *et al.* The Green ChemisTREE: 20 years after taking root with the 12 principles. *Green chemistry* **2018**, *20* (9), 1929-1961, 10.1039/C8GC00482J.
94. Koeller, K.M.; *et al.* Enzymes for chemical synthesis. *Nature* **2001**, *409* (6817), 232-240, 10.1038/35051706.
95. Li, F.; *et al.* Enzymatic CH functionalizations for natural product synthesis. *Curr. Opin. Chem. Biol.* **2019**, *49*, 25-32, 10.1016/j.cbpa.2018.09.004.
96. Ferreyra, M.L.F.; *et al.* Identification of a bifunctional maize C-and O-glucosyltransferase. *J. Biol. Chem.* **2013**, *288* (44), 31678-31688.
97. Jesus, A.R.; *et al.* Exploiting the therapeutic potential of 8- $\beta$ -d-glucopyranosylgenistein: synthesis, antidiabetic activity, and molecular interaction with islet amyloid polypeptide and amyloid  $\beta$ -peptide (1–42). *J. Med. Chem.* **2014**, *57* (22), 9463-9472, 10.1021/jm501069h.
98. Liu, D.-M.; *et al.* Recent advances in nano-carrier immobilized enzymes and their applications. *Process Biochem.* **2020**, *92*, 464-475, 10.1016/j.procbio.2020.02.005.

## 2 Publications

### I

#### **Discovery and Heterologous Expression of New Cyclic Depsibosamycins**

Marc Stierhof, Maksym Myronovskiy, Josef Zapp and Andriy Luzhetskyy

*Microorganisms* **2021**, 9(7), 1396

DOI: 10.3390/microorganisms9071396

Published online 28<sup>nd</sup> June 2021



## Article

# Discovery and Heterologous Production of New Cyclic Depsibosamycins

Marc Stierhof <sup>1</sup>, Maksym Myronovskiy <sup>1</sup>, Josef Zapp <sup>2</sup> and Andriy Luzhetskyy <sup>1,3,\*</sup>

<sup>1</sup> Department of Pharmaceutical Biotechnology, Saarland University, 66123 Saarbruecken, Germany; marc.stierhof@uni-saarland.de (M.S.); maksym.myronovskiy@uni-saarland.de (M.M.)

<sup>2</sup> Department of Pharmaceutical Biology, Saarland University, 66123 Saarbruecken, Germany; j.zapp@mx.uni-saarland.de

<sup>3</sup> AMEG Department, Helmholtz Institute for Pharmaceutical Research Saarland, 66123 Saarbrücken, Germany

\* Correspondence: a.luzhetskyy@mx.uni-saarland.de; Tel.: +49-681-3027-0200

**Abstract:** *Streptomyces* are producers of valuable secondary metabolites with unique scaffolds that perform a plethora of biological functions. Nonribosomal peptides are of special interest due to their variety and complexity. They are synthesized by nonribosomal peptide synthetases, large biosynthetic machineries that are encoded in the genome of many *Streptomyces* species. The identification of new peptides and the corresponding biosynthetic gene clusters is of major interest since knowledge can be used to facilitate combinatorial biosynthesis and chemical semisynthesis of natural products. The recently discovered bosamycins are linear octapeptides with an interesting 5-OMe tyrosine moiety and various modifications at the N-terminus. In this study, the new cyclic depsibosamycins B, C, and D from *Streptomyces aurantiacus* LU19075 were discovered. In comparison to the linear bosamycins B, C, and D, which were also produced by the strain, the cyclic depsibosamycins showed a side-chain-to-tail lactonization of serine and glycine, leading to a ring of four amino acids. In silico identification and heterologous expression of the depsibosamycin (*dbm*) gene cluster indicated that the cyclic peptides, rather than the linear derivatives, are the main products of the cluster.

**Keywords:** bosamycin; *Streptomyces*; heterologous expression; NRPS; cyclic peptide



**Citation:** Stierhof, M.; Myronovskiy, M.; Zapp, J.; Luzhetskyy, A. Discovery and Heterologous Production of New Cyclic Depsibosamycins. *Microorganisms* **2021**, *9*, 1396. <https://doi.org/10.3390/microorganisms9071396>

Academic Editors:  
Stéphane Cociancich and  
Valérie Leclère

Received: 3 June 2021  
Accepted: 24 June 2021  
Published: 28 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

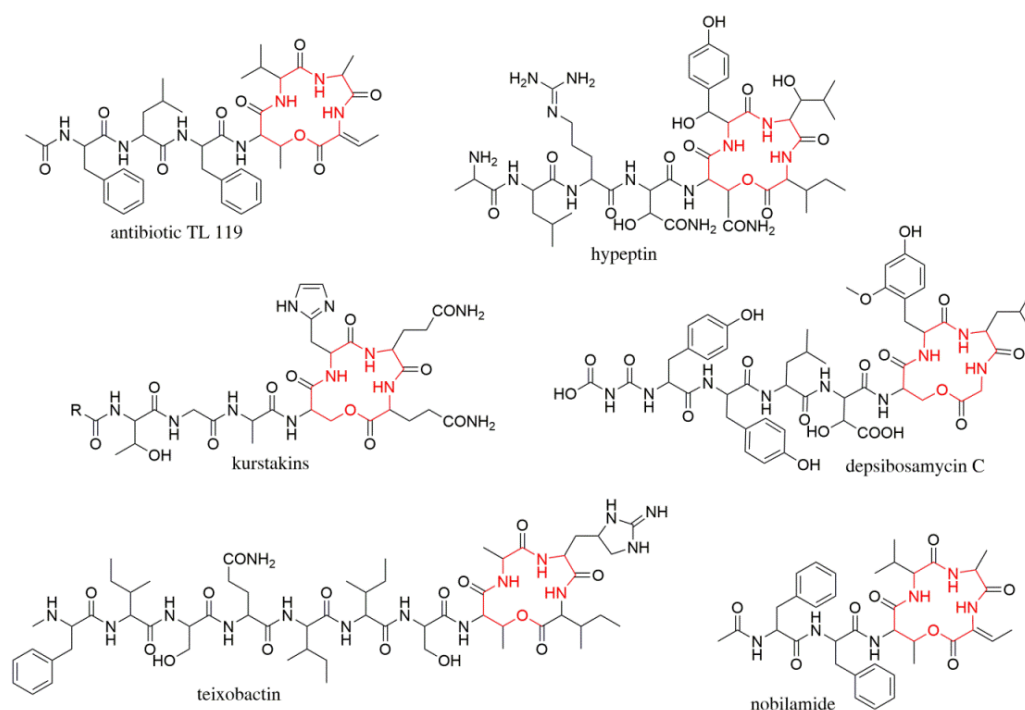
## 1. Introduction

*Streptomyces* are known to produce valuable secondary metabolites with a plethora of unique structural classes [1,2]. These compounds often express a variety of biological functions, which are the product of centuries of evolutionary design to develop a selective advantage for the strain [3]. The enormous structural complexity of natural products from *Streptomyces*, which are generated by enzymatic biosynthetic machinery, can be recreated only with difficulty by chemical synthesis. Therefore, a detailed understanding of the biosynthetic origin of natural products is of major interest to facilitate combinatorial biosynthesis and semisynthetic approaches, which are important tools in drug development.

Nonribosomal peptide synthetases (NRPSs) generate highly diverse peptides and are a coveted target for biosynthetic engineering [4,5]. Nonribosomal peptides (NRPs) are synthesized by modular enzymes. Individual modules consist of a minimal set of three domains: the adenylation domain (A), thiolation domain (T), and condensation domain (C). Biosynthesis is initiated by the A-domain, which specifically activates natural or modified amino acids. The activated amino acid is tethered to the pantetheinyl arm of the T-domain. The condensation reaction with another amino acid, selected by an adjacent module, is specifically catalyzed by the C-domain. Each module can contain additional domains to generate D-amino acids, N-methylated peptide bonds, heterocyclic amino acids, etc. [6]. The mature peptide is cleaved from the NRPS assembly line by the thioesterase domain (TE), resulting in a linear or cyclic peptide [6,7].

The recent development of gene sequencing, bioinformatics, and genomic tools has enabled the targeted identification, expression, or activation of unique biosynthetic gene clusters (BGCs), which is a major leap forward in the discovery of new natural products [8]. By targeting novel NRPS architectures, Xu et al. recently discovered the bosamycin gene cluster (*bsm* cluster), encoding the production of new linear octapeptides named bosamycins. The NRPS assembly line contains eight modules leading to the following peptide sequence: L-Tyr, D-Tyr, L-Leu, L-erythro- $\beta$ -OH-Asp, L-Ser, 5-MeO-D-Tyr, L-Leu, and Gly. The 5-OH tyrosine moiety is generated by a stand-alone NRPS with a unique P450 domain, but no C-domain is present in the module. The octapeptide chain is further modified by the two separate genes encoding an O-methyltransferase and a dioxygenase, leading to 5-OMe tyrosine and erythro- $\beta$ -OH asparagine. Bosamycin E showed slight inhibitory activity towards SHP2 (Src homology 2-containing protein tyrosine phosphatase 2), while other derivatives had no biological activity [9].

Herein, we describe the discovery of the new cyclic depsibosamycins B, C, and D. The compounds were identified by metabolic profiling of *Streptomyces aurantiacus* LU19075 and characterized by MS/MS fragmentation and NMR spectroscopy. Structure elucidation studies revealed a side-chain-to-tail lactonization of serine and glycine within the structure depsibosamycin. This led to a ring including four amino acids, a feature that has also been described for hypeptin, teixobactin, and others (Figure 1). In this paper, we postulate that the cyclic depsibosamycins, rather than the linear ones, are the main products of the BGC. This was supported by the results of the heterologous expression experiments and comparisons of the production rate.



**Figure 1.** Structure comparison of compounds with a similar side-chain-to-tail lactone ring (red) formed by an intramolecular reaction of  $\beta$ -hydroxy amino acids and the C-terminus.

## 2. Materials and Methods

### 2.1. General Procedures

All strains and bacterial artificial chromosomes (BACs) used in this work are listed in Table S1. *Escherichia coli* strains were cultured in LB medium [10]. *Streptomyces* strains were grown on soya flour mannitol agar (MS agar) [11] or in liquid tryptic soy broth (TSB; Sigma-Aldrich, St. Louis, MO, USA) for cultivation. Liquid DNPM medium (40 g/L dextrin, 7.5 g/L soytone, 5 g/L baking yeast, and 21 g/L MOPS, pH 6.8) was used for



secondary metabolite expression. The antibiotics kanamycin, apramycin, and nalidixic acid were supplemented when required.

## 2.2. Metabolite Extraction and Analysis by Mass Spectrometry (MS)

*S. aurantiacus* LU19075 was grown in 20 mL of TSB in a 100 mL baffled flask for one day, and 1 mL of the culture was inoculated into 100 mL of DNPM production medium in a 500 mL baffled flask. Cultures were grown for 6 days at 28 °C and 180 rpm in an Infors multitron shaker (Infors AG, Basel, Switzerland). Metabolites were extracted from the supernatant with butanol, concentrated to dryness, and dissolved in 1 mL methanol (MeOH). One microliter was separated on a Dionex Ultimate 3000 UPLC system (Thermo Fisher Scientific, Waltham, MA, USA) equipped with an ACQUITY UPLC BEH C18 1.7  $\mu$ m column (30, 50, or 100 mm, Waters Corporation, Milford, MA, USA) using a linear gradient of 5–95 vol% aqueous acetonitrile (ACN) with 0.1% formic acid (FA) at a flow rate of 0.6 mL/min and a column oven temperature of 45 °C. Sample analysis was carried out on a coupled PDA detector followed by an amaZon speed (Bruker, Billerica, MA, USA) for production control. High-resolution masses were obtained from either an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) or a MaXis high-resolution LC-QTOF system (Bruker, Billerica, MA, USA) using positive ionization mode and mass range detection of  $m/z$  200 to 2000. Data analysis was performed using Compass Data Analysis v. 4.1 (Bruker) and Xcalibur v. 3.0 (Thermo Fisher Scientific).

## 2.3. Nuclear Magnetic Resonance (NMR) Spectroscopy and Optical Rotation (OD)

The NMR spectra of  $\beta$ -OH-Asp and bosamycin D were recorded on a Bruker Avance, UltraShield 500 MHz (Bruker, BioSpin GmbH, Rheinstetten, Germany) equipped with a 5 mm BBO probe at 298 K. All other NMR spectra were acquired on a Bruker Avance III, Ascent 700 MHz spectrometer at 298 K equipped with a 5 mm TCI cryoprobe. The chemical shifts ( $\delta$ ) were reported in parts per million (ppm) relative to tetramethylsilane (TMS). As solvents, deuterated CD<sub>3</sub>OD ( $\delta_H$  3.31 ppm,  $\delta_C$  49.15 ppm), DMSO-*d*<sub>6</sub> ( $\delta_H$  2.50 ppm,  $\delta_C$  39.51 ppm), and D<sub>2</sub>O ( $\delta_H$  4.75 ppm) from Deutero (Kastellaun, Germany) were used. Edited-HSQC, HSQC-TOCSY, HMBC, <sup>1</sup>H-<sup>1</sup>H COSY, ROESY, NOESY, and N-HSQC spectra were recorded using standard pulse programs from TOPSPIN v.3.6 software. Optical rotations were measured using a Perkin Elmer Polarimeter Model 241 (Überlingen, Deutschland).

## 2.4. Isolation of Bosamycin B–D and Depsibosamycin C

*S. aurantiacus* LU19075 was grown in 10 L of DNPM production medium and extracted with butanol. The dry crude extract was dissolved in 100 mL of methanol. The isolation process was guided by LC-MS to identify fractions. The extract was purified on an Isolera<sup>TM</sup> One flash purification system equipped with a SNAP Ultra C<sub>18</sub> 400 g (Biotage, Uppsala, Sweden) using a gradient of 5–30 vol% aqueous methanol for 3 column volumes (CV) followed by 30–95 vol% aqueous methanol for 1.5 CV at a flow rate of 100 mL/min and UV detection at 210 and 280 nm. Further impurities were removed by size exclusion chromatography (SEC; stationary phase: Sephadex-LH20) with isocratic elution using 100% methanol.

The crude material was dissolved in methanol, and reversed-phase HPLC was carried out on a Waters Autopurification System (Waters Corporation, Milford, MA, USA) equipped with a SQD2-MS-Detector and equipped with a preparative VP 250/21 NUCLEODUR C<sub>18</sub> HTec 5  $\mu$ m column (MACHERY NAGEL, Düren, Germany) using a gradient of 55–75 vol% aqueous methanol with 0.1% FA for 12 min at a flow rate of 20 mL/min.

The final purification step was carried out on an Agilent Infinity 1100 series reversed-phase HPLC system equipped with a Synergi<sup>TM</sup> 4  $\mu$ m Fusion-RP C<sub>18</sub> 80 Å 250 × 10 mm column (Phenomenex, Torrance, CA, USA) using a gradient of 31–33 vol% aqueous ACN with 0.1% FA, a flow rate of 4 mL/min, and an oven temperature of 20 °C. The compounds were detected with a PDA detector at 210 and 280 nm at RT = 10.06 min (bosamycin B), 10.29 min (bosamycin C), 10.69 min (bosamycin D), and 11.74 min (depsibosamycin C).

#### 2.4.1. Bosamycin B

White powder; 1.6 mg;  $[\alpha]_D^{20}$ -38 (c 0.06, MeOH); UV (37% ACN in H<sub>2</sub>O + 0.1% FA)  $\lambda_{\max}$  (log  $\epsilon$ ) 200 (4.25), 276 (2.93) nm; <sup>1</sup>H and <sup>13</sup>C NMR data, see Table S3; ESI-TOF-MS  $m/z$  1082.4650 [M + H]<sup>+</sup> (calc. for C<sub>50</sub>H<sub>68</sub>N<sub>9</sub>O<sub>18</sub> 1082.4677), see Figure S1.

#### 2.4.2. Bosamycin C

White powder; 6.8 mg;  $[\alpha]_D^{20}$ -3 (c 0.27, DMSO); UV (37% ACN in H<sub>2</sub>O + 0.1% FA)  $\lambda_{\max}$  (log  $\epsilon$ ) 200 (3.82), 276 (2.62) nm; <sup>1</sup>H and <sup>13</sup>C NMR data, see Table S4; ESI-TOF-MS  $m/z$  1126.4595 [M + H]<sup>+</sup> (calc. for C<sub>51</sub>H<sub>68</sub>N<sub>9</sub>O<sub>20</sub> 1126.4575), see Figure S1.

#### 2.4.3. Bosamycin D

White powder; 9.1 mg;  $[\alpha]_D^{20}$ -31 (c 0.21, MeOH); UV (37% ACN in H<sub>2</sub>O + 0.1% FA)  $\lambda_{\max}$  (log  $\epsilon$ ) 200 (3.96), 276 (2.95) nm; <sup>1</sup>H and <sup>13</sup>C NMR data, see Table S5; ESI-TOF-MS  $m/z$  1081.4718 [M + H]<sup>+</sup> (calc. for C<sub>51</sub>H<sub>69</sub>N<sub>8</sub>O<sub>18</sub> 1081.4724), see Figure S1.

#### 2.4.4. Depsibosamycin C

White powder; 4.8 mg;  $[\alpha]_D^{20}$ -13 (c 0.24, DMSO); UV (42% ACN in H<sub>2</sub>O + 0.1% FA)  $\lambda_{\max}$  (log  $\epsilon$ ) 200 (3.95), 276 (2.65) nm; <sup>1</sup>H and <sup>13</sup>C NMR data, see Table S2; ESI-TOF-MS  $m/z$  1108.4496 [M + H]<sup>+</sup> (calc. for C<sub>51</sub>H<sub>66</sub>N<sub>9</sub>O<sub>19</sub> 1108.4470), see Figure S1.

### 2.5. Marfey's Analysis

Bosamycin C was hydrolyzed in 100  $\mu$ L 6 N HCl at 110 °C for 1 h. While cooling, the sample was dried for 15 min under nitrogen and dissolved in 110 mL water, and 50  $\mu$ L of each sample was transferred into 1.5 mL Eppendorf tubes. To the hydrolysate, 20  $\mu$ L of 1 N NaHCO<sub>3</sub> and 20  $\mu$ L of 1% L-FDLA (*N*<sup>α</sup>-(5-Fluoro-2,4-dinitrophenyl)-L-leucinamide) or D-FDLA in acetone were added. The amino acid standards were prepared the same way using only L-FDLA. The reaction mixtures were incubated at 40 °C for 90 min at 700 rpm and subsequently quenched with 2 N HCl to stop the reaction. The samples were diluted with 300  $\mu$ L ACN, and 1  $\mu$ L aliquots were analyzed by a MaXis high-resolution LC-QTOF system using aqueous ACN with 0.1 vol% FA and an adjusted gradient of 5–10 vol% in 2 min, 10–25 vol% in 13 min, 25–50 vol% in 7 min, and 50–95 vol% in 2 min. Sample detection was carried out at 340 nm.  $\beta$ -Hydroxyaspartic acid diastereomers were synthesized from *cis*- and *trans*-epoxysuccinic acid according to the literature [12].

#### 2.5.1. DL-Erythro-Hydroxyaspartic Acid

White crystals; yield 24%; <sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O, pH = 4)  $\delta$ : 4.47 (d, *J* = 3.2, 1H), 4.17 (d, *J* = 3.2, 1H) (Figure S37); ESI-TOF-MS  $m/z$  150.0398 [M + H]<sup>+</sup> (calc. for C<sub>4</sub>H<sub>8</sub>NO<sub>5</sub> 150.0403).

#### 2.5.2. DL-Threo-Hydroxyaspartic Acid

White crystals, yield 42%, <sup>1</sup>H NMR (500 MHz, D<sub>2</sub>O, pH = 12)  $\delta$ : 4.35 (d, *J* = 2.2, 1H), 3.60 (d, *J* = 2.2, 1H) (Figure S38); ESI-TOF-MS  $m/z$  150.0398 [M + H]<sup>+</sup> (calc. for C<sub>4</sub>H<sub>8</sub>NO<sub>5</sub> 150.0403).

### 2.6. Isolation and Manipulation of DNA

BAC extraction from a *S. aurantiacus* LU19075 constructed genomic library (Intact Genomics, St. Louis, MO, USA), DNA manipulation, *E. coli* transformation, and *E. coli*/Streptomyces intergeneric conjugation were performed according to standard protocols [10,11,13,14]. Plasmid DNA was purified with the BACMAX™ DNA Purification Kit (Lucigen, Middleton, WI, USA). Restriction endonucleases were used according to the manufacturer's recommendations (New England Biolabs, Ipswich, MA, USA).

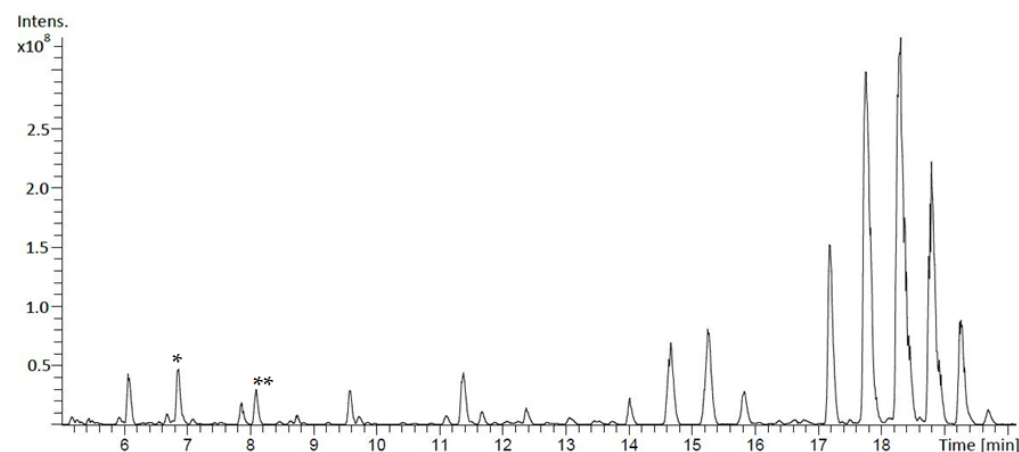
## 2.7. Genome Mining and Bioinformatics Analysis

An *S. aurantiacus* LU19075 genome BAC library using the pSMART vector with apramycin resistance gene was constructed and sequenced by Intact Genomics (Missouri, St. Louis, USA). The genome was screened for secondary metabolite biosynthetic gene clusters using the antiSMASH online tool with loose settings (Available online: <https://antismash.secondarymetabolites.org/#!/start>) [15] (accessed on 3 June 2021). Analysis of genetic data was performed using Geneious 11.0.3 [16]. The DNA sequence of the *dbm* gene cluster from LU19075 was deposited into GenBank under the accession number MW740414.

## 3. Results and Discussion

### 3.1. Identification and Structural Elucidation of Depsibosamycins B, C, and D

We screened several strains from our strain library for the production of new compounds by comparing UV/VIS spectra, retention times, MS/MS fragmentation patterns, and HRMS data with those from natural product databases, such as “Dictionary of Natural Products” (DNP) [17]. By this, *Streptomyces aurantiacus* LU19075 caught our attention due to the production of a large quantity of compounds that were identified as well described members of the nactin family (Figure 2) [18]. Furthermore, six unknown compounds eluted in two groups at RT = 6.8 min and RT = 8.1 min; each group contained three compounds with slightly different retention times. (Figure 2). All compounds showed similar UV/VIS spectra, MS/MS fragmentation patterns, and no hits in natural product databases. The first group of compounds at RT = 6.8 min showed  $[M + H]^+$  molecular ions of  $m/z$  1081.5, 1082.5, and 1126.5 Da. The compounds were isolated, and their structures were determined by NMR spectroscopy (Figures S11–S33), MS/MS fragmentation (Figure S38–S40), and Marfey’s method to assign the stereochemistry (Figures S34–S37). This resulted in the assignment of the previously reported bosamycins B, C, and D. The second group of compounds eluting at RT = 8.1 min showed  $[M + H]^+$  molecular ions  $[M + H]^+$  with  $m/z$  1063.5, 1064.5, and 1108.4 Da. The compound with  $m/z$  1108.4 was the only compound that could be isolated to perform structure elucidation.



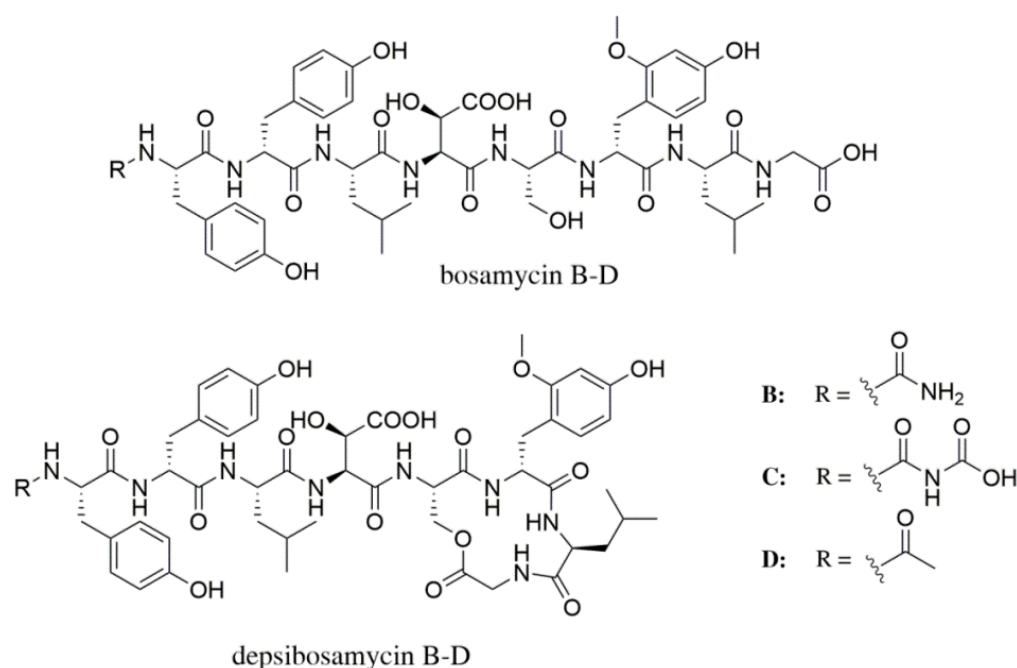
**Figure 2.** LC-MS chromatogram of the butanol extract of *Streptomyces aurantiacus* LU19075 cultivated in DNPM medium showing the peaks of bosamycins B, C, and D (\*) and depsibosamycins B, C, and D (\*\*). In addition to that, nactins (17–20 min) and nactin degradation products (9–16 min) were identified.

*S. aurantiacus* LU19075 was cultivated in DNPM, and the metabolites were extracted with butanol. The compound with  $m/z$  1108.4 was further purified by reversed-phase flash purification and size exclusion chromatography. The final HPLC purification step led to the successful isolation of the compound. The structure was determined by extensive analysis of 1D and 2D NMR (Figures S2–S10).

The precise mass of the compound was determined to be  $m/z$  1108.4496 ( $[M + H]^+$ ), indicating a water loss compared to bosamycin C. The molecular formula was calculated to

be  $C_{51}H_{66}N_9O_{19}$  with 24 degrees of unsaturation; thus, we expected an additional double bond or ring formation. The structure assignment by NMR revealed a peptide highly similar to bosamycin C but with chemical shift differences observed in the edited-HSQC spectra (Figure S10). The differences were narrowed down to the Ser- $\beta$ -CH<sub>2</sub> signal showing a low field shift and the Gly-CH<sub>2</sub> signal, which showed the emergence of diastereotopic proton signals. Therefore, we suggested ring formation between serine and glycine through condensation as the final step in the biosynthetic pathway. Due to ambiguous HMBC correlations to prove the connection between serine and glycine, further support was required to confirm the cyclic structure. To obtain further evidence, we analyzed the OH-proton signals after exhaustively freeze-drying the sample. OH-proton signals were observed for all tyrosines ( $\delta_H$  9.11–9.16), as well as the  $\beta$ -OH-Asp ( $\delta_H$  6.11) signals. The only proton signal missing was from OH-serine, which indicates that serine is condensed with the C-terminal carboxyl group (Figures S5 and S7). Due to structural similarities to bosamycin C and side-chain-to-tail lactonization, the new compound was named depsibosamycin C. The proposed cyclic structure of depsibosamycin C was confirmed by MS/MS analysis (Figure S41). The fragmentation pattern showed a-ion fragments resulting from the ring structure and y-ions with a difference of 18 Da corresponding to water loss. The b-ion fragments b<sub>1</sub>–b<sub>4</sub> were the same since depsibosamycin C and bosamycin C share the same carboxyl carbamoyl moiety at the N-terminus.

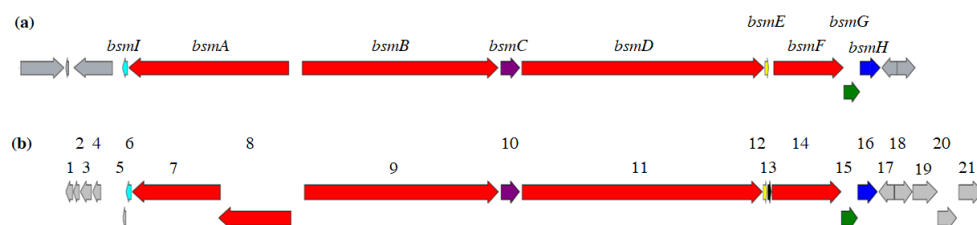
The other two compounds showed precise  $[M + H]^+$  molecular ions of  $m/z$  1064.4598 and 1063.4603 Da and very similar retention times compared to depsibosamycin C. However, the production rate was 20-fold lower; thus, characterisation by MS/MS was used to identify the structures (Figures S42 and S43). The compounds showed the same y-ion pattern as depsibosamycin C, indicating structural similarities. Differences were observed in the b-ion fragments, indicating different N-terminal caps. The b-ion fragments b<sub>1</sub>–b<sub>4</sub> of compound  $m/z$  1064.5 correspond to those of bosamycin B, while the b-ion fragments b<sub>1</sub>–b<sub>4</sub> of compound  $m/z$  1063.5 correspond to those of bosamycin D. In addition, both new compounds are different by 18 Da from their corresponding bosamycins. In conclusion, both new compounds are cyclic variants of linear bosamycins B and D; thus, they were named depsibosamycins B (1064.5) and D (1063.4) (Figure 3).



**Figure 3.** Structures of bosamycins B, C, and D and depsibosamycins B, C, and D.

### 3.2. Heterologous Expression of the Depsibosamycin (*dbm*) Gene Cluster

A provided sequenced *S. aurantiacus* LU19075 genomic BAC library was analyzed with the genome mining software antiSMASH in loose settings to identify potential BGC candidates for depsibosamycins. Annotated NRPS clusters in the genome were analyzed in silico regarding their similarity to depsibosamycins by comparing the adenylation domain composition and the stand-alone NRPS module, which is involved in the biosynthesis of 5-OMe tyrosine [19–22]. One of the NRPS clusters showed an adenylation domain composition that corresponds the depsibosamycin amino acid sequence and the unique stand-alone NRPS unit (*dbmF*), which is expected to produce the 5-OMe tyrosine moiety. Alignment of the *bsm* cluster (GenBank: MN509472) and the identified NRPS cluster from *S. aurantiacus* LU19075, herein called the *dbm* cluster (GenBank: MW740414), showed 99.2% similarity within *dbmI*–*dbmH* (Figure 4). Gene *bsmA* was split into two separate genes, *dbmA*<sub>1</sub> and *dbmA*<sub>2</sub>; however, the domain organisation remained the same, implying similar function. The *dbm* cluster was chosen as a promising candidate for encoding the biosynthetic pathway of depsibosamycins.



**Figure 4.** Alignment of the bosamycin (*bsm*) cluster (a) and the depsibosamycin (*dbm*) cluster (b) revealed 99.2% similarity.

To prove the production of depsibosamycins, the *dbm* cluster was heterologously expressed in *Streptomyces lividans* TK24 [23]. BAC I7, covering the entire *dbm* cluster (Table 1), was identified and isolated from a *S. aurantiacus* LU19075 genomic library.

**Table 1.** Genes encoded within the chromosomal fragment cloned in BAC I7 and annotation of the *bsm* and *dbm* genes.

Gene #	<i>dbm</i> Genes <sup>a</sup>	Putative Function	<i>bsm</i> Genes <sup>b</sup>
1	<i>orf</i> – 5	Hypothetical protein	
2	<i>orf</i> – 4	Hypothetical protein	
3	<i>orf</i> – 3	Hypothetical protein	
4	<i>orf</i> – 2	Hypothetical protein	
5	<i>orf</i> – 1	Dehydrogenase	
6	<i>dbmI</i>	Acyl carrier protein	<i>bsmI</i>
7	<i>dbmA</i> <sub>1</sub>	NRPS (C-A-PCP-E)	<i>bsmA</i>
8	<i>dbmA</i> <sub>2</sub>	NRPS (C-A-PCP)	
9	<i>dbmB</i>	NRPS (C-A-PCP-C-A-PCP-C-A-PCP)	<i>bsmB</i>
10	<i>dbmC</i>	Dioxygenase	<i>bsmC</i>
11	<i>dbmD</i>	NRPS (C-A-PCP-E-C-A-PCP-C-A-PCP-TE)	<i>bsmD</i>
12	<i>dbmE</i>	MbtH family protein	<i>bsmE</i>
13		Hypothetical protein	
14	<i>dbmF</i>	NRPS (P450-A-PCP)	<i>bsmF</i>
15	<i>dbmG</i>	Alpha/beta hydrolase	<i>bsmG</i>
16	<i>dbmH</i>	O-methyltransferase	<i>bsmH</i>
17	<i>orf</i> + 1	ABC transporter	
18	<i>orf</i> + 2	ABC transporter	
19	<i>orf</i> + 3	Hypothetical protein	
20	<i>orf</i> + 4	ABC transporter	
21	<i>orf</i> + 5	Hypothetical protein	

<sup>a</sup> GenBank accession number MW740414 <sup>b</sup> GenBank accession number MN509472.

The BAC was transferred into *Streptomyces lividans* TK24 by biparental conjugation. BAC I7 containing clones were selected from apramycin containing MS plates leading the exconjugant strain *S. lividans* I7. The exconjugants were cultivated in DNPM production medium and extracted with n-butanol. LC-MS revealed successful heterologous expression in *S. lividans* (Figure S47). The production of depsibosamycin D could be readily observed in the extract of the strain, while depsibosamycins B and C were absent. Bosamycin D was produced only in trace amounts. In addition to the known depsibosamycins, two new compounds with  $m/z$  1021.4523 [M + H] and 1107.4506 [M + H] were identified. Of these two, compound  $m/z$  1021.4523 [M + H] was produced in the highest amounts. The structures of the new compounds were characterized by MS/MS fragmentation. The y-ion patterns correspond to those previously observed for depsibosamycin B, C, and D. Differences were seen in the b-ion fragments  $b_1$ – $b_4$ , indicating altered N-terminal moieties (Figures S44 and S45). The major compound with  $m/z$  1021.4523 [M + H] was characterized as depsibosamycin N, the cyclic version of bosamycin N previously discovered by heterologous expression [9]. Comparison of the molecular formula of the compound with  $m/z$  1107.4506 [M + H] ( $C_{52}H_{66}N_8O_{19}$ ) and of depsibosamycin C ( $C_{51}H_{66}N_9O_{10}$ ) revealed one more carbon and one fewer nitrogen. The compound did not show similarity to any of the known bosamycins regarding the N-terminal cap; hence, it was named depsibosamycin O. The calculated molecular formula of the remaining fragment  $b_1$  ( $C_{12}H_{12}NO_5^+$ ) indicates an N-malonyl cap at the tyrosine residue (Figure S45).

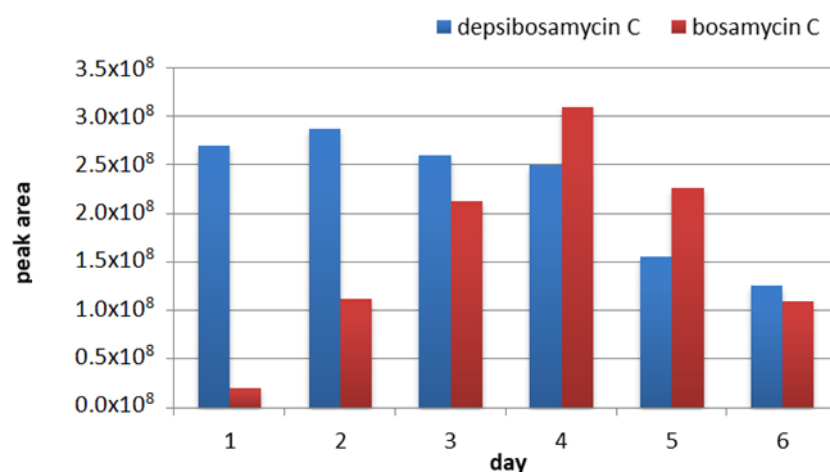
The spectrum of depsibosamycins produced in the heterologous strain is in accordance with previously published data. Depsibosamycins B and C are two members of the depsibosamycin family carrying carbamoyl and carboxyl carbamoyl N-terminal modifications, respectively. As the genes required for the attachment of these moieties are located outside the *dbm* gene cluster, the production of depsibosamycins B and C in heterologous hosts is not expected. Depsibosamycin N is the smallest member of the depsibosamycin family without an N-terminal modification. It is produced in the highest amounts, while the modified depsibosamycins D and O are the minor products. The observed depsibosamycins D and O carry acetyl and malonyl moieties at their N-termini, respectively. Attachment of the acetyl moiety was proposed to be catalyzed by an unidentified enzyme in the natural bosamycin producer [9]. It is likely that the enzymatic activities required for the attachment of the acetyl and malonyl residues of depsibosamycins D and O are encoded by unidentified genes within the genome of the heterologous host *S. lividans* TK24.

### 3.3. Determination of the Direct Biosynthetic Product of the *dbm* Gene Cluster

Both the linear bosamycins and the cyclic depsibosamycins have been identified in the strain *S. aurantiacus* LU19074, while only linear derivatives were reported in a previous study [9]. High similarity of the *bsm* and *dbm* (Figure 4) clusters raises the question of which of the compound families, bosamycins or depsibosamycins, are the direct biosynthetic products. To investigate this, the time course of bosamycin and depsibosamycin production was studied. For this purpose, the *S. aurantiacus* strain was cultivated in DNPM for six days, and the production was screened by LC-MS every day (Figure 5). Cyclic depsibosamycin C was already detected in high amounts after 1 day, while its linear form bosamycin C was barely detectable at this time point. In the course of the following five days, the concentration of depsibosamycin C in the culture did not increase significantly. In contrast, the concentration of bosamycin C increased over the course of cultivation and peaked on the fourth day before it declined again. The obtained data indicate that linear bosamycin C originates from the linearization of cyclic depsibosamycin C. This is in accordance with the heterologous expression results. In the recombinant strain *S. lividans* I7, linear bosamycin D was detected in very low amounts, while cyclic depsibosamycin D was produced in high quantities (Figure S47). Furthermore, depsibosamycins N and O did not appear in their linear forms. The cyclisation of the nascent polypeptide chain is often catalyzed by the thioesterase (TE) domain; thus, we took a closer look at the TE domains of *dbmD* and *bsmD* to determine structural differences. Alterations in the geometry of the TE binding



pocket could have catalyzed the formation of the cyclic product by enabling intramolecular esterification [6,7]. However, alignment of the TE domains of *bsmD* and *dbmD* revealed 98% identity (Figure S46), suggesting that altered catalytic activity is unlikely. Furthermore, the sequence analysis of the genes flanking the *dbm* cluster (*orf* − 5–*orf* − 1 and *orf* + 1–*orf* + 5) did not reveal any genes that could catalyze the cyclization reaction (Table 1).



**Figure 5.** Depsibosamycin C and bosamycin C production in *S. aurantiacus* LU19075 over the course of 6 days.

This suggests that the TE domain of *DbmD* is responsible for catalyzing the cyclization of the mature linear peptide rather than hydrolytic cleavage. The poor stability of the cyclic depsibosamycins might have prevented their previous discovery in the *Streptomyces* sp. 120454 strain.

#### 4. Conclusions

In this study, we describe the discovery of the new cyclic depsibosamycins B, C, and D from the strain *S. aurantiacus* LU19075. The compounds consist of the same amino acids as the previously discovered bosamycins B, C, and D but differ by the side-chain-to-tail lactonization of serine and glycine. By in silico analysis, we identified a depsibosamycin (*dbm*) gene cluster. Heterologous expression of the *dbm* cluster revealed the production of cyclic depsibosamycin D and the two new depsibosamycins N and O. Linear bosamycin D was detected only in small quantities. Since the cyclic depsibosamycins have not been reported previously, the origin of cyclic depsibosamycins and linear bosamycins was investigated. By comparing the production rates of linear bosamycin C and cyclic depsibosamycin C, we found that the linear bosamycin likely originates from cyclic depsibosamycin. Furthermore, we compared the two biosynthetic gene clusters, in particular, the TE domains. No significant difference could be detected, suggesting the same product for both gene clusters. Therefore, we postulate that the linear bosamycins might be a product of enzymatic or physicochemical degradation.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/microorganisms9071396/s1>, Figure S1: LC-MS spectra of isolated bosamycin B–D and depsibosamycin C, Figures S2–S6: <sup>1</sup>H, <sup>13</sup>C, HSQC, COSY, and HMBC NMR data of depsibosamycin C, Figure S7: Extra dry HMBC spectrum of depsibosamycin C, Figures S8 and S9: N-HSQC and ROESY spectrum of depsibosamycin C, Figure S10: Overlapping edited HSQC spectra of depsibosamycin C and bosamycin C, Figures S11–S17: <sup>1</sup>H, <sup>13</sup>C, HSQC, COSY, HMBC, N-HSQC, and ROESY NMR data of bosamycin B, Figures S18–S27: <sup>1</sup>H, <sup>13</sup>C, HSQC, COSY, HSQC-TOCSY, HMBC, N-HSQC, and ROESY NMR data of bosamycin C, Figures S28–S33: <sup>1</sup>H, <sup>13</sup>C, HSQC, COSY, HMBC, and NOESY NMR data of bosamycin D, Figure S34: Marfey's method: MS spectra of bosamycin C, Figure S35: Structure of bosamycin C with stereochemistry, Figure S36: <sup>1</sup>H NMR spectrum of DL-erythro-β-hydroxyaspartic acid, Figure S37: <sup>1</sup>H NMR spectrum of DL-threo-β-hydroxyaspartic acid, Figures

S38–S40: MS/MS spectra of bosamycin B–D, Figures S41–S43: MS/MS spectra of depsibosamycin B–D, Figure S44: MS/MS spectra of depsibosamycin N, Figure S45: MS/MS spectra of depsibosamycin O, Figure S46: TE domain alignment of the depsibosamycin cluster (Query) and the bosamycin cluster (Sbjct), Figure S47: LC-MS chromatograms of *S. lividans* I7 with the dbm gene cluster, Table S1: Strains, BACs, plasmids, and primers used in this work, Table S2: NMR data for depsibosamycin C, Table S3: NMR data for bosamycin B, Table S4: NMR data for bosamycin C, Table S5: NMR data for bosamycin D.

**Author Contributions:** The experiments were designed and evaluated by M.S., M.M., and A.L. and the practical work performed by M.S. NMR experiments were set up, carried out, and evaluated by M.S. and the data reviewed by J.Z. The manuscript was drafted by M.S. and M.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We thank Helmholtz Institute for Pharmaceutical Research Saarland, Saarbruecken, Germany (HIPS) for support of NMR measurements. The strain *S. aurantiacus* LU19075 was provided by BASF SE Ludwigshafen, Germany.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Barka, E.A.; Vatsa, P.; Sanchez, L.; Gaveau-Vaillant, N.; Jacquard, C.; Klenk, H.-P.; Clément, C.; Ouhdouch, Y.; van Wezel, G.P. Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol. Mol. Biol. Rev.* **2016**, *80*, 1–43. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Chen, Y.; Ntai, I.; Ju, K.-S.; Unger, M.; Zamdborg, L.; Robinson, S.J.; Doroghazi, J.R.; Labeda, D.P.; Metcalf, W.W.; Kelleher, N.L. A Proteomic Survey of Nonribosomal Peptide and Polyketide Biosynthesis in Actinobacteria. *J. Proteome Res.* **2011**, *11*, 85–94. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Běhal, V. Bioactive products from streptomyces. *Adv. Virus Res.* **2000**, *47*, 113–156. [\[CrossRef\]](#)
4. Strieker, M.; Tanović, A.; Marahiel, M.A. Nonribosomal peptide synthetases: Structures and dynamics. *Curr. Opin. Struct. Biol.* **2010**, *20*, 234–240. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Bozhüyük, K.A.J.; Linck, A.; Tietze, A.; Kranz, J.; Wesche, F.; Nowak, S.; Fleischhacker, F.; Shi, Y.-N.; Grün, P.; Bode, H.B. Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nat. Chem.* **2019**, *11*, 653–661. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Süßmuth, R.D.; Mainz, A. Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew. Chem. Int. Ed.* **2017**, *56*, 3770–3821. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Finking, R.; Marahiel, M.A. Biosynthesis of Nonribosomal Peptides. *Annu. Rev. Microbiol.* **2004**, *58*, 453–488. [\[CrossRef\]](#)
8. Myronovskyi, M.; Luzhetskyy, A. Heterologous production of small molecules in the optimized Streptomyces hosts. *Nat. Prod. Rep.* **2019**, *36*, 1281–1294. [\[CrossRef\]](#)
9. Hill, R.A.; Sutherland, A. Hot off the Press. *Nat. Prod. Rep.* **2020**, *37*, 1294–1299. [\[CrossRef\]](#)
10. Green, J.M.R.; Sambrook, J. *Molecular Cloning: A Laboratory Manual*, 4th ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2012.
11. Kieser, T.; Bibb, M.J.; Buttner, M.J.; Chater, K.F.; Hopwood, D.A. *Practical Streptomyces Genetics. A Laboratory Manual*; John Innes Foundation: Norwich, UK, 2000.
12. Robertson, A.W.; McCarville, N.G.; MacIntyre, L.W.; Correa, H.; Haltli, B.; Marchbank, D.H.; Kerr, R.G. Isolation of Imaqobactin, an Amphiphilic Siderophore from the Arctic Marine Bacterium *Variovorax* Species RKJM285. *J. Nat. Prod.* **2018**, *81*, 858–865. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Rebets, Y.; Kormanec, J.; Luzhetskyy, A.; Bernaerts, K.; Anné, J. *Metagenomics: Methods and Protocols*; Streit, W., Daniel, R., Eds.; Springer: New York, NY, USA, 2017; pp. 99–144.
14. Flett, F.; Mersinias, V.; Smith, C.P. High efficiency intergeneric conjugal transfer of plasmid DNA from Escherichia coli to methyl DNA-restricting streptomycetes. *FEMS Microbiol. Lett.* **1997**, *155*, 223–229. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Medema, M.H.; Blin, K.; Cimermanic, P.; De Jager, V.; Zakrzewski, P.; Fischbach, M.; Weber, T.; Takano, E.; Breitling, R. Antismash: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **2011**, *39*, W339–W346. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [\[CrossRef\]](#) [\[PubMed\]](#)



17. Whittle, M.; Willett, P.; Klaffke, W.; van Noort, P. Evaluation of Similarity Measures for Searching the Dictionary of Natural Products Database. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 449–457. [[CrossRef](#)] [[PubMed](#)]
18. Smith, L.L. An additional source of macrotetrolide antibiotics. *J. Antibiot.* **1975**, *28*, 1000–1003. [[CrossRef](#)] [[PubMed](#)]
19. Challis, G.; Ravel, J.; Townsend, C.A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **2000**, *7*, 211–224. [[CrossRef](#)]
20. Prieto, C.; García-Estrada, C.; Lorenzana, D.; Martín, J.F. NRPSp: Non-ribosomal peptide synthase substrate predictor. *Bioinformatics* **2012**, *28*, 426–427. [[CrossRef](#)] [[PubMed](#)]
21. Rausch, C. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res.* **2005**, *33*, 5799–5808. [[CrossRef](#)] [[PubMed](#)]
22. Röttig, M.; Medema, M.H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. NRPSpredictor2—A web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **2011**, *39*, W362–W367. [[CrossRef](#)] [[PubMed](#)]
23. Rückert, C.; Albersmeier, A.; Busche, T.; Jaenicke, S.; Winkler, A.; Friðjónsson, Ó.H.; Hreggviðsson, G.Ó.; Lambert, C.; Badcock, D.; Bernaerts, K.; et al. Complete genome sequence of *Streptomyces lividans* TK24. *J. Biotechnol.* **2015**, *199*, 21–22. [[CrossRef](#)]

# Discovery and gene cluster of depsibosamycins – cyclic bosamycins from *Streptomyces aurantiacus* LU19075

Marc Stierhof<sup>1</sup>, Maksym Myronovskiy<sup>1</sup>, Josef Zapp<sup>2</sup> and Andriy Luzhetskyy<sup>1,3,\*</sup>

<sup>1</sup> Department of Pharmaceutical Biotechnology, Saarland University, 66123 Saarbruecken, Germany; marc.stierhof@uni-saarland.de (M.S.); maksym.myronovskiy@uni-saarland.de (M.M.)

<sup>2</sup> Department of Pharmaceutical Biology, Saarland University, 66123 Saarbruecken, Germany; j.zapp@mx.uni-saarland.de (J.Z.)

<sup>3</sup> AMEG Department, Helmholtz Institute for Pharmaceutical Research Saarland

\* Correspondence: a.luzhetskyy@mx.uni-saarland.de; Tel.: +49-681-302-70200 (A.L.)

**Supplementary.**

1. Strains, BACs, Plasmids and Primers Used in this Work. ....	4
2. High Resolution Masse of the Isolated Bosamycin B-D and Depsibosamycin C. ....	5
3. NMR Data of the Isolated Bosamycin B-D and Depsibosamycin C. ....	6
4. Stereochemical Assignment by Marfey's Method. ....	31
5. MS/MS Fragmentation Data. ....	34

### *Table of Figures*

Figure S1: LC-MS spectra of isolated bosamycin B-D (1-3) and depsibosamycin C (4). ....	5
Figure S2: <sup>1</sup> H NMR spectrum (700 MHz, DMSO- <i>d</i> <sub>6</sub> ) of depsibosamycin C. ....	8
Figure S3: <sup>13</sup> C NMR spectrum (700 MHz, DMSO- <i>d</i> <sub>6</sub> ) of depsibosamycin C. ....	8
Figure S4: Edited HSQC spectrum (700 MHz, 50% NUS, DMSO- <i>d</i> <sub>6</sub> ) of depsibosamycin C. ....	9
Figure S5: Extra dry COSY spectrum (700 MHz, 50% NUS, DMSO- <i>d</i> <sub>6</sub> ) of depsibosamycin C showing the serine 28-CH and 28-OH correlation (red circle). ....	9
Figure S6: HMBC spectrum (700 MHz, 50% NUS, DMSO- <i>d</i> <sub>6</sub> ) of depsibosamycin C. ....	10
Figure S7: Extra dry HMBC spectrum (700 MHz, 50% NUS, DMSO- <i>d</i> <sub>6</sub> /TFA) of depsibosamycin C showing the correlation of 59-NH and C-60 (red circle) and the tyrosine-OH correlations (green circle). ....	10
Figure S8: N-HSQC spectrum (700 MHz, 25% NUS, DMSO- <i>d</i> <sub>6</sub> /TFA) of depsibosamycin C. ....	11
Figure S9: ROESY spectrum (700 MHz, DMSO- <i>d</i> <sub>6</sub> /TFA) of depsibosamycin C. ....	11
Figure S10: Overlapping edited HSQC spectra (700 MHz, DMSO- <i>d</i> <sub>6</sub> /TFA) of depsibosamycin C and bosamycin C. ....	12
Figure S11: <sup>1</sup> H NMR spectrum (700 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin B. ....	15
Figure S12: <sup>13</sup> C NMR spectrum (500 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin B. ....	15
Figure S13: Edited HSQC spectrum (500 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin B. ....	16
Figure S14: COSY spectrum (700 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin B. ....	16
Figure S15: HMBC spectrum (700 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin B. ....	17
Figure S16: N-HSQC spectrum (700 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin B. ....	17
Figure S17: ROESY spectrum (700 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin B. ....	18
Figure S18: <sup>1</sup> H NMR spectrum (700 MHz, DMSO- <i>d</i> <sub>6</sub> /TFA) of bosamycin C. ....	21
Figure S19: <sup>1</sup> H NMR spectrum (700 MHz, MeOD) of bosamycin C. ....	21
Figure S20: <sup>1</sup> H NMR spectrum (700 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin C. ....	22
Figure S21: <sup>13</sup> C NMR spectrum (700 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin C. ....	22
Figure S22: HSQC spectrum (700 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin C. ....	23
Figure S23: COSY spectrum (700 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin C. ....	23
Figure S24: HSQC-TOCSY spectrum (700 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin C. ....	24
Figure S25: HMBC spectrum (700 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin C. ....	24
Figure S26: N-HSQC spectrum (700 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin C. ....	25
Figure S27: ROESY spectrum (700 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin C. ....	25
Figure S28: <sup>1</sup> H NMR spectrum (500 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin D. ....	28
Figure S29: <sup>13</sup> C NMR spectrum (500 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin D. ....	28

Figure S30: Edited HSQC spectrum (500 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin D. ....	29
Figure S31: COSY spectrum (500 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin D. ....	29
Figure S32: HMBC spectrum (500 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin D. ....	30
Figure S33: NOESY spectrum (500 MHz, DMSO- <i>d</i> <sub>6</sub> ) of bosamycin D. ....	30
Figure S34: Marfey's method: MS spectra of bosamycin C derivatized with D-/L-FDLA (bottom) and the reference amino acids (AA) derivatized with D-/L-FDLA (top). ....	31
Figure S35: Structure of bosamycin C with stereochemistry determined by Marfey's method. ....	31
Figure S36: <sup>1</sup> H NMR spectrum (500 MHz, D <sub>2</sub> O) of DL- <i>erythro</i> -β-hydroxyaspartic acid. ....	32
<b>Figure S37:</b> <sup>1</sup> H NMR spectrum (500 MHz, D <sub>2</sub> O) of DL- <i>threo</i> -β-hydroxyaspartic acid. ....	32
Figure S38: MS/MS analysis of bosamycin B. ....	33
Figure S39: MS/MS analysis of bosamycin C. ....	33
Figure S40: MS/MS analysis of bosamycin D. ....	34
Figure S41: MS/MS analysis of depsibosamycin C. ....	34
Figure S42: MS/MS of depsibosamycin B. ....	35
Figure S43: MS/MS analysis of depsibosamycin D. ....	35
Figure S44: MS/MS of depsibosamycin N, produced by <i>S. lividans</i> I7. ....	36
Figure S45: MS/MS analysis and the calculated masses of the expected fragments of depsibosamycin O produced by <i>S. lividans</i> I7. ....	36
Figure S46: TE domain alignment of the depsibosamycin cluster (Query) and the bosamycin cluster (Sbjct). ....	37
Figure S47: LC-MS chromatograms of <i>S. lividans</i> I7 with the dbm gene cluster showing (a) the extracted masses of depsibosamycin D, N and O (b) the extracted mass of bosamycin D, (c) the base peak chromatograms (bpc) of <i>S. lividans</i> I7 and (d) the bpc of the empty host <i>S. lividans</i> TK24 ....	37

#### Table of Tables

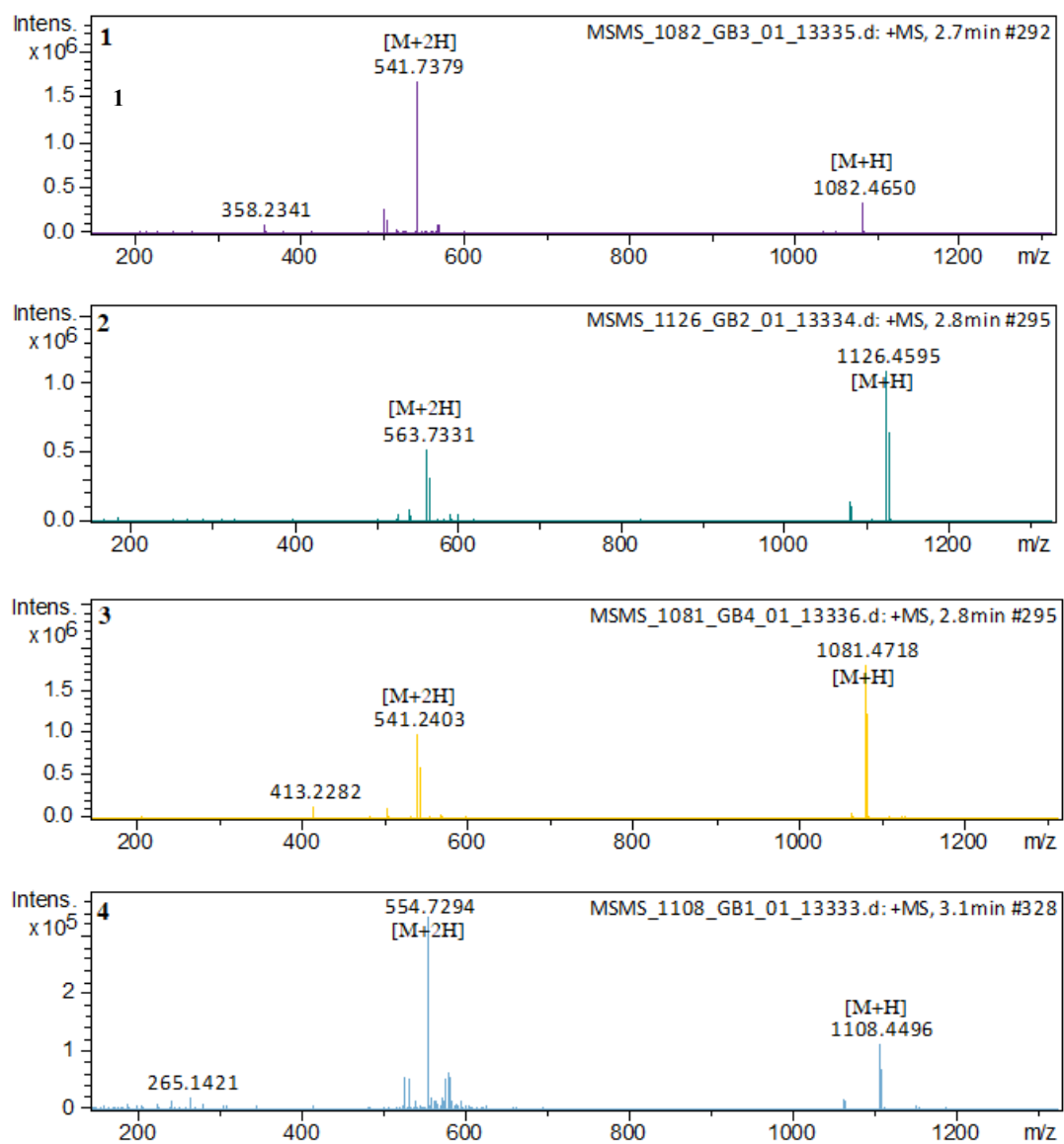
Table S1: Strains, BACs, plasmids and primers used in this work. ....	4
Table S2: NMR data (700 MHz, DMSO- <i>d</i> <sub>6</sub> ) for depsibosamycin C. ....	6
Table S3: NMR data (700 MHz, DMSO- <i>d</i> <sub>6</sub> ) for bosamycin B. ....	13
Table S4: NMR data (700 MHz, DMSO- <i>d</i> <sub>6</sub> +TFA) for bosamycin C. ....	19
Table S5: NMR data (500 MHz, DMSO- <i>d</i> <sub>6</sub> ) for bosamycin D. ....	26

## 1. Strains, BACs, Plasmids and Primers Used in this Work

**Table S1:** Strains, BACs, plasmids and primers used in this work.

Material	Purpose
<b>Bacterial strains</b>	
<i>Streptomyces aurantiacus</i> LU19075	wild type strain [BASF]
<i>Streptomyces lividans</i> TK24	heterologous host[1]
<i>Escherichia coli</i> GB05 RedCC	cloning host [Helmholtz Institute für Pharmaceutische Forschung (HIPS)]
<i>Escherichia coli</i> ET12567 pUB307	alternate host intergeneric conjugation[2]
<b>BACs</b>	
I7	Heterologous expression of depsibosamycin cluster
<b>Plasmids</b>	
pSMART	Lucigen (USA)

## 2. High Resolution Masse of the Isolated Bosamycin B-D and Depsibosamycin C



**Figure S1:** LC-MS spectra of isolated bosamycin B-D (1-3) and depsibosamycin C (4).

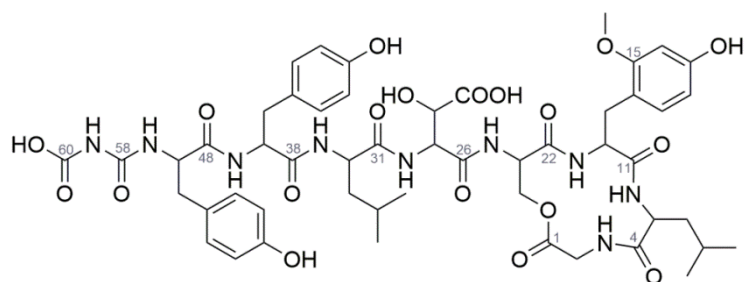
### 3. NMR Data of the Isolated Bosamycin B-D and Depsibosamycin C

**Table S2:** NMR data (700 MHz, DMSO-*d*6) for depsibosamycin C.

unit	$\delta_H$ , multiplicity, (J in Hz)	$\delta_C$	$\delta_N$
Gly			
1 – COOH		168.7	
2 – CH $_{\alpha,1}$	3.98, dd (16.6, 6.1)	41.6	
CH $_{\alpha,2}$	3.73, m		
3 – NH	7.56, t (6.1)		105.2
Leu			
4 – CO		171.2	
5 – CH $_{\alpha}$	4.29, m	49.6	
6 – CH $_{\beta,1}$	1.42, ovl*	37.5	
CH $_{\beta,2}$	1.34, m		
7 – CH $_{\gamma}$	1.26, m	23.8	
8 – CH $_{\delta,1}$	0.76, d (5.7)	22.8	
9 – CH $_{\delta,2}$	0.70, d (5.5)	21.5	
10 – NH	7.97, d (7.8)		125.5
o-MeO-Tyr			
11 – CO		171.0	
12 – CH $_{\alpha}$	4.50, ovl*	52.4	
13 – CH $_{\beta,1}$	2.91, m	29.2	
CH $_{\beta,2}$	2.64, ovl*		
14		115.2	
15		158.0	
16	6.33, bs	98.7	
17		157.3	
17 – OH	9.16, bs		
18	6.17, d (7.33)	106.4	
19	6.81, d (7.6)	130.9	
20	3.71, s	55.1	
21 – NH	7.66, d (8.5)		122.8
Ser			
22 – CO		168.5	
23 – CH $_{\alpha}$	4.52, ovl*	51.8	
24 – CH $_{\beta,1}$	4.11, d (9.9)	65.7	
CH $_{\beta,2}$	4.21, d (8.5)		
25 – NH	8.27, ovl*		116.6
$\beta$ -OH-Asp			
26 – CO		168.9	
27 – CH $_{\alpha}$	4.69, t (7.2)	55.2	
28 – CH $_{\beta}$	4.16, d (6.6)	70.9	
28 – OH	6.11, bs		
29 – COOH		172.6	
30 – NH	8.19, d (7.5)		111.3

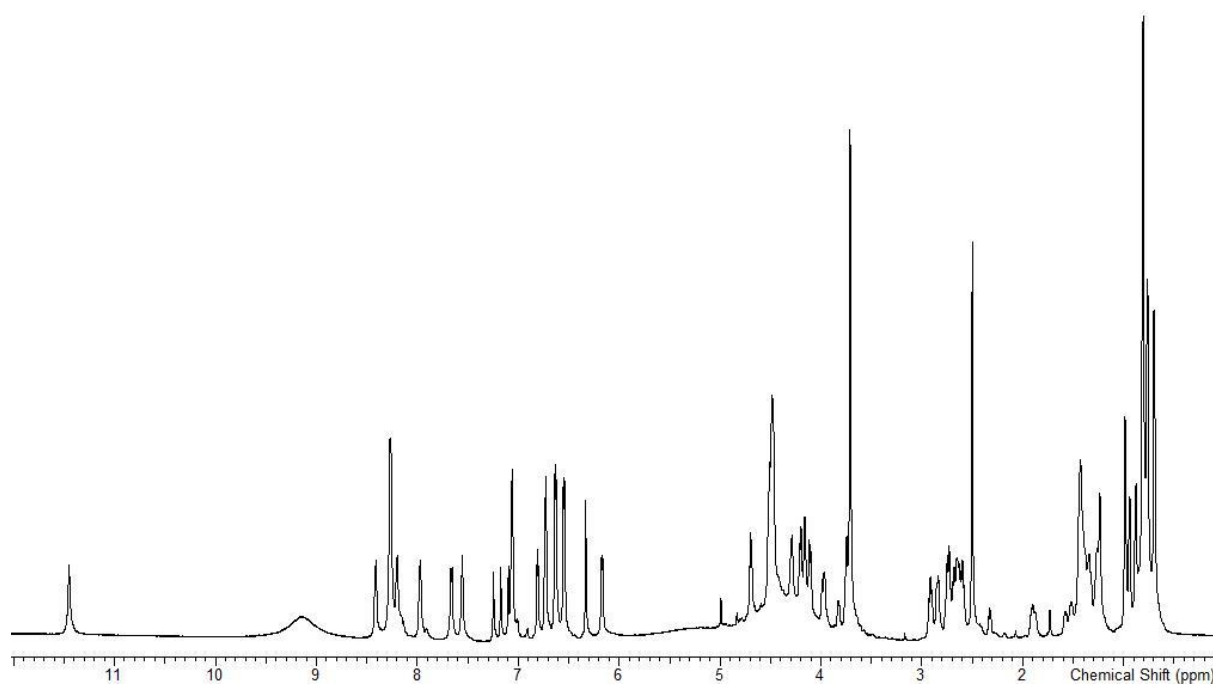
Leu		
31 – CO		172.3
32 – CH <sub>α</sub>	4.48, ovl*	50.9
33 – CH <sub>β</sub>	1.43, ovl*	41.0
34 – CH <sub>γ</sub>	1.38, m	24.0
35 – CH <sub>δ,1</sub>	0.81, ovl*	23.2
36 – CH <sub>δ,2</sub>	0.80, ovl*	21.2
37 – NH	8.26, ovl*	119.8
Tyr		
38 – CO		171.2
39 – CH <sub>α</sub>	4.48, ovl*	54.7
40 – CH <sub>β,1</sub>	2.84, m	37.3
CH <sub>β,2</sub>	2.67, ovl*	
41		127.2
42, 46	7.06, d (7.2)	130.2
43, 45	6.63, d (7.3)	114.8
44		156.0
44 – OH	9.11, bs	
47 – NH	8.41, d (6.1)	118.6
Tyr		
48 – CO		170.1
49 – CH <sub>α</sub>	4.47, ovl*	53.9
50 – CH <sub>β,1</sub>	2.60, ovl*	36.6
CH <sub>β,2</sub>	2.73, m	
51		127.0
52, 56	6.73, d (7.2)	130.0
53, 55	6.55, d (7.3)	114.8
54		155.8
54 – OH	9.11, bs	
57 – NH	8.26, ovl*	114.7
FCA		
58 – CO		158.7
59 – NH	11.45, bs	169.2
60 – COOH		156.2

\*signal overlap

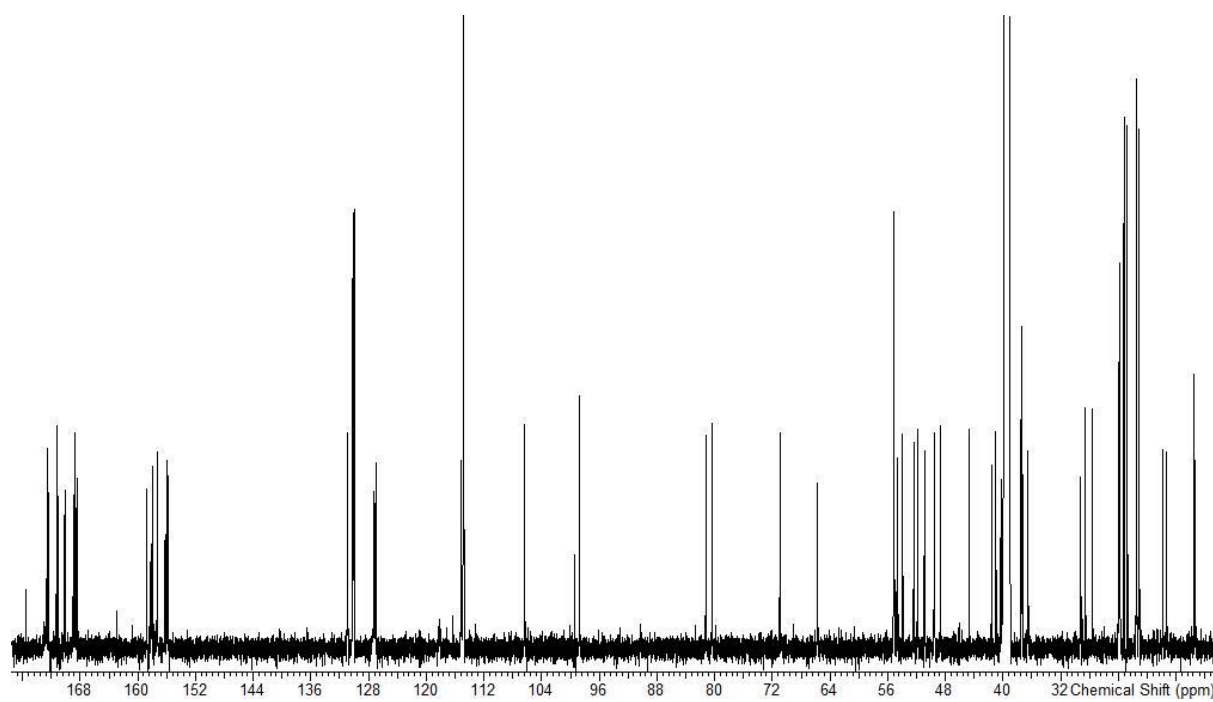


depsibosamycin C

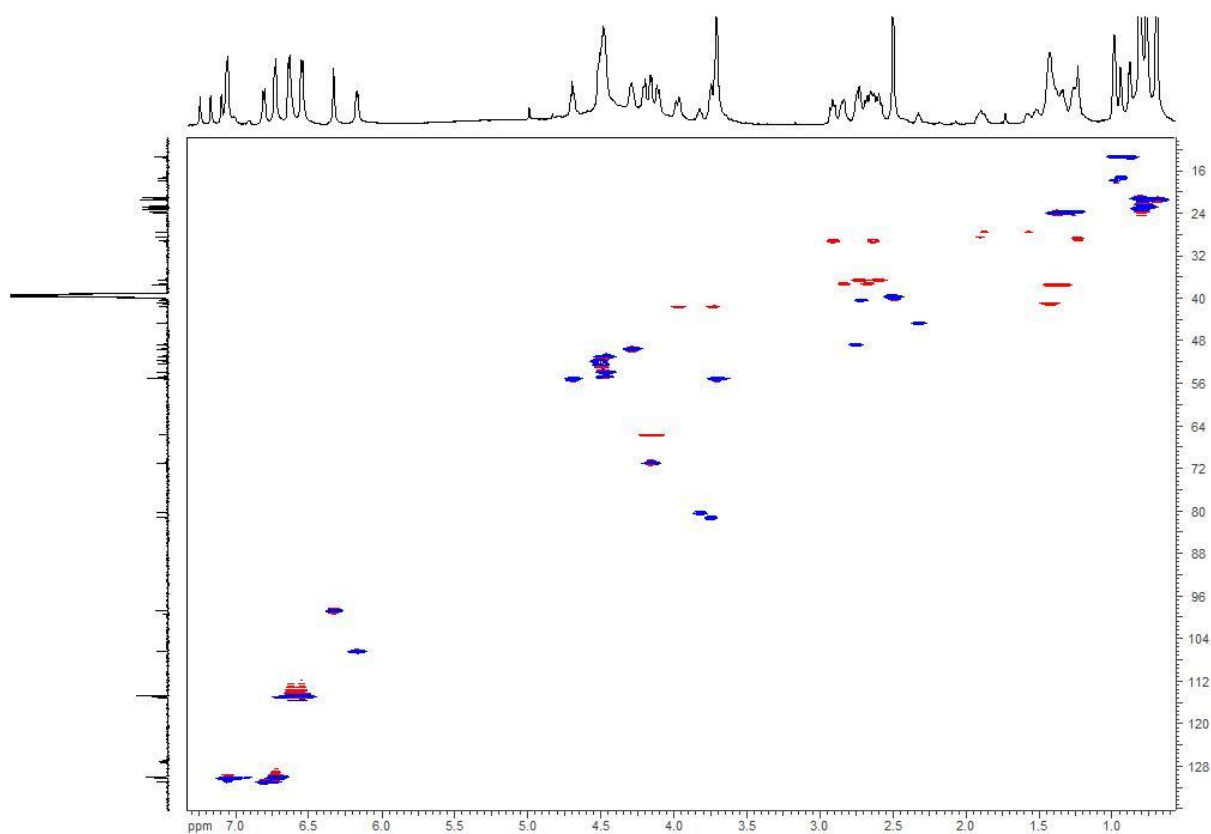




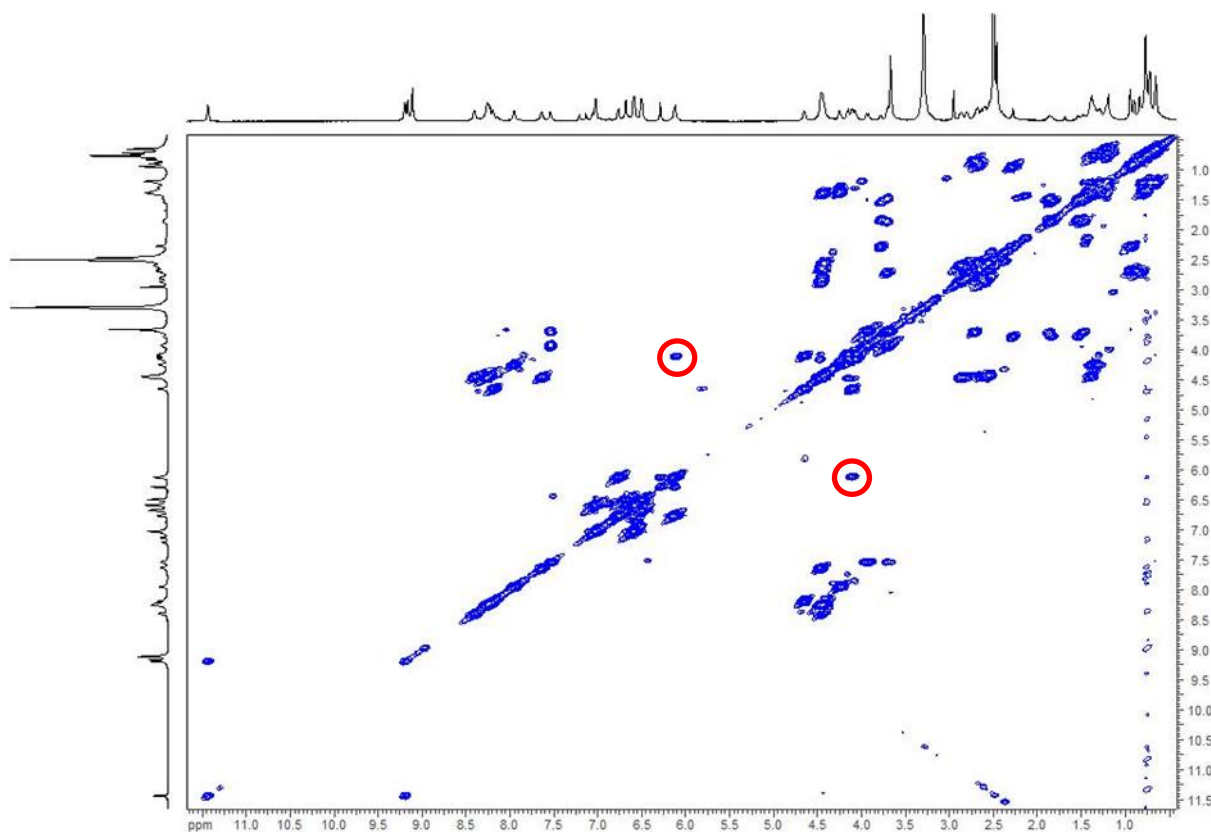
**Figure S2:**  $^1\text{H}$  NMR spectrum (700 MHz,  $\text{DMSO-}d_6$ ) of depsibosamycin C.



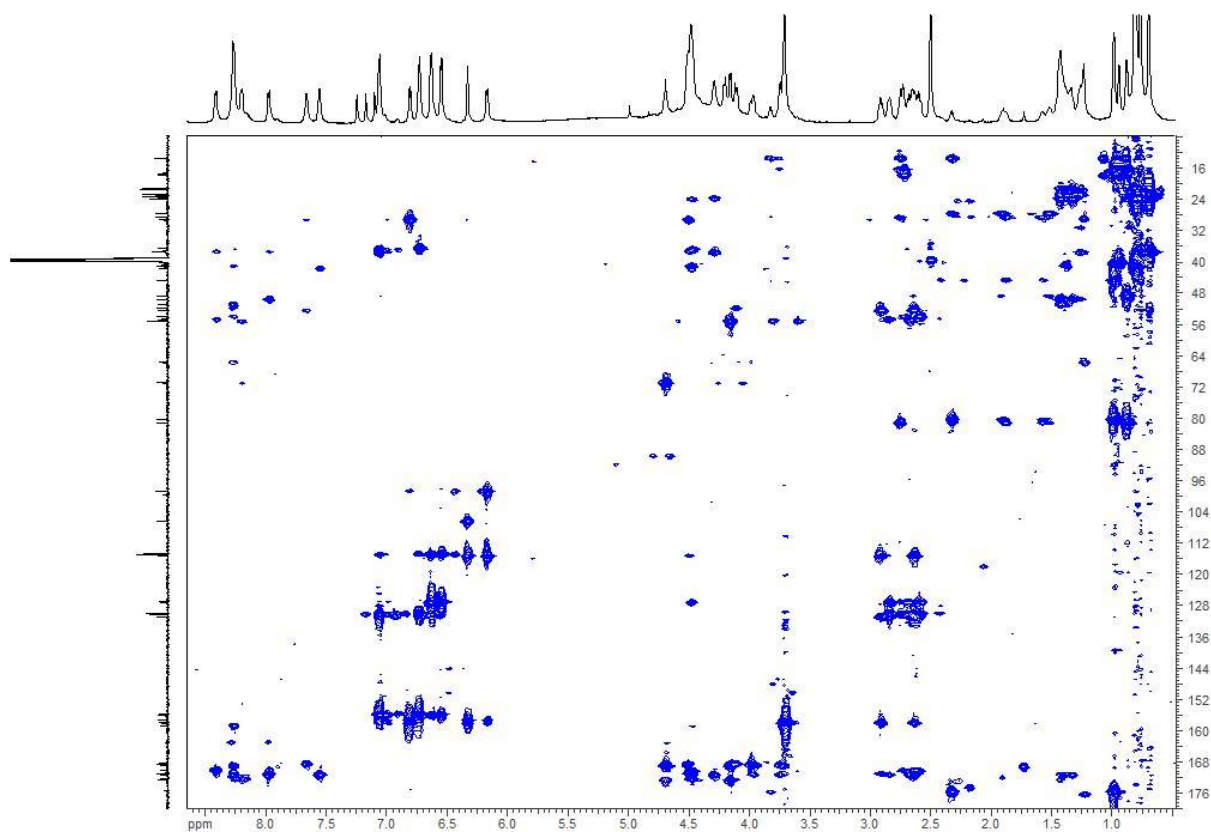
**Figure S3:**  $^{13}\text{C}$  NMR spectrum (700 MHz,  $\text{DMSO-}d_6$ ) of depsibosamycin C.



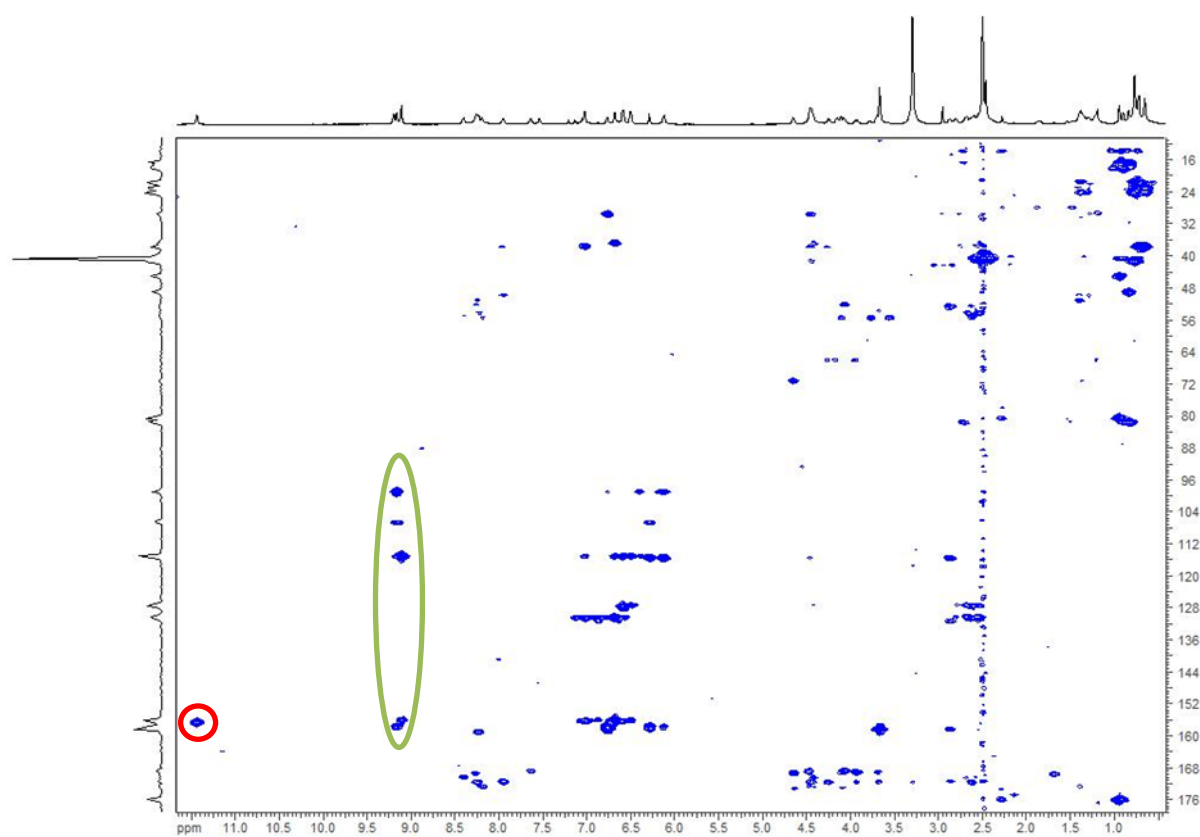
**Figure S4:** Edited HSQC spectrum (700 MHz, 50% NUS, DMSO-*d*<sub>6</sub>) of depsibosamycin C.



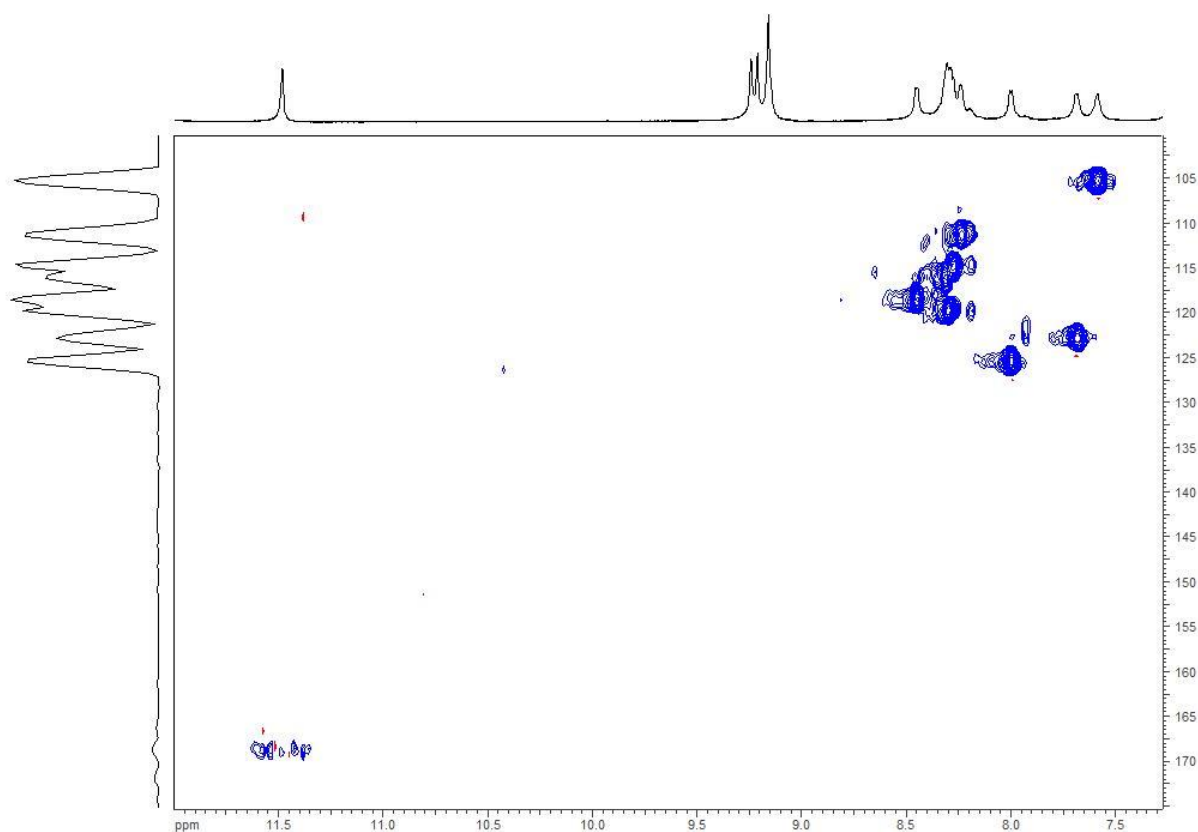
**Figure S5:** Extra dry COSY spectrum (700 MHz, 50% NUS, DMSO-*d*<sub>6</sub>) of depsibosamycin C showing the serine 28-CH and 28-OH correlation (red circle).



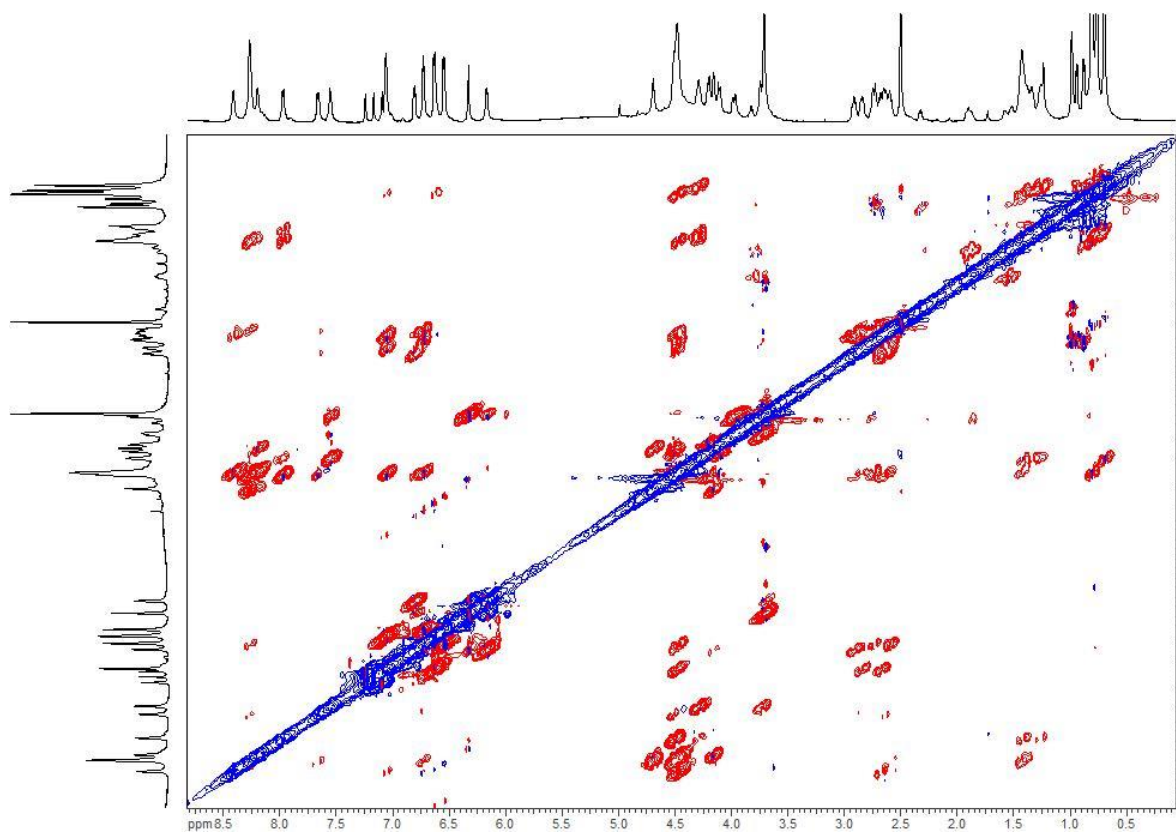
**Figure S6:** HMBC spectrum (700 MHz, 50% NUS, DMSO-*d*<sub>6</sub>) of depsibosamycin C.



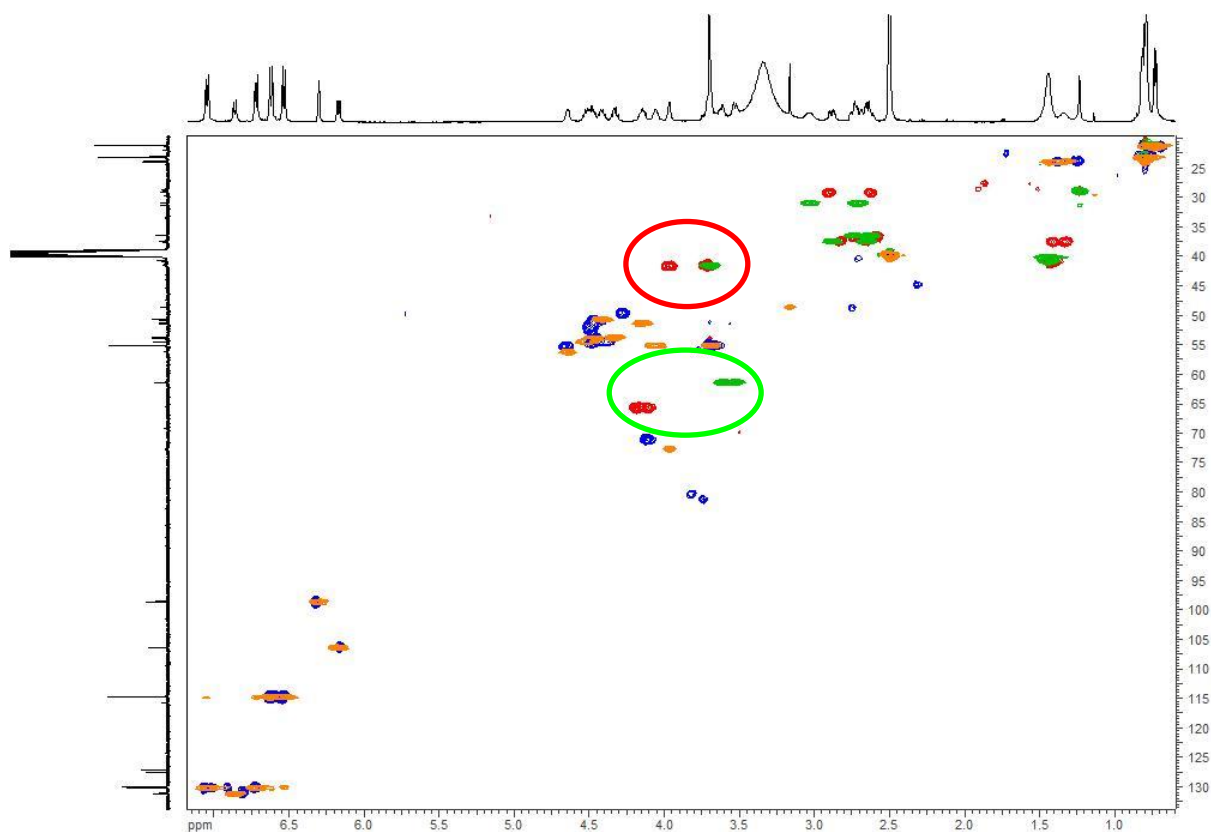
**Figure S7:** Extra dry HMBC spectrum (700 MHz, 50% NUS, DMSO-*d*<sub>6</sub>/TFA) of depsibosamycin C showing the correlation of 59-NH and C-60 (red circle) and the tyrosine-OH correlations (green circle).



**Figure S8:** N-HSQC spectrum (700 MHz, 25% NUS, DMSO-*d*<sub>6</sub>/TFA) of depsibosamycin C.



**Figure S9:** ROESY spectrum (700 MHz, DMSO-*d*<sub>6</sub>/TFA) of depsibosamycin C.



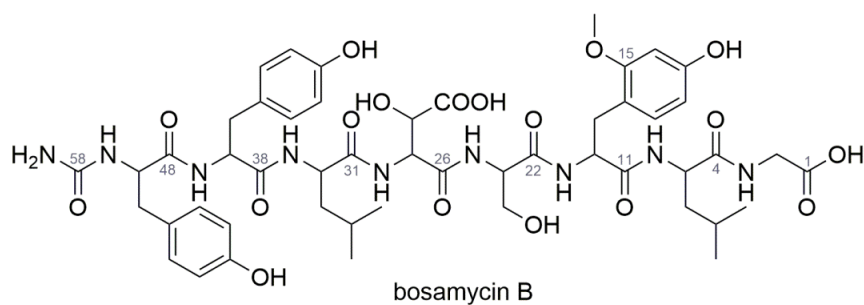
**Figure S10:** Overlapping edited HSQC spectra (700 MHz, DMSO-*d*<sub>6</sub>/TFA) of depsibosamycin C and bosamycin C. Methylene/methine signals are depicted as red/blue (dep C) and green/orange (bos C). The methylene signals Gly-CH<sub>2</sub> (red circle) and Ser-CH<sub>2</sub> (green circle) of depsibosamycin A showed significant peak splitting (Gly) and a shift difference (Ser) indicating an esterification between Ser-OH and Gly-COOH.

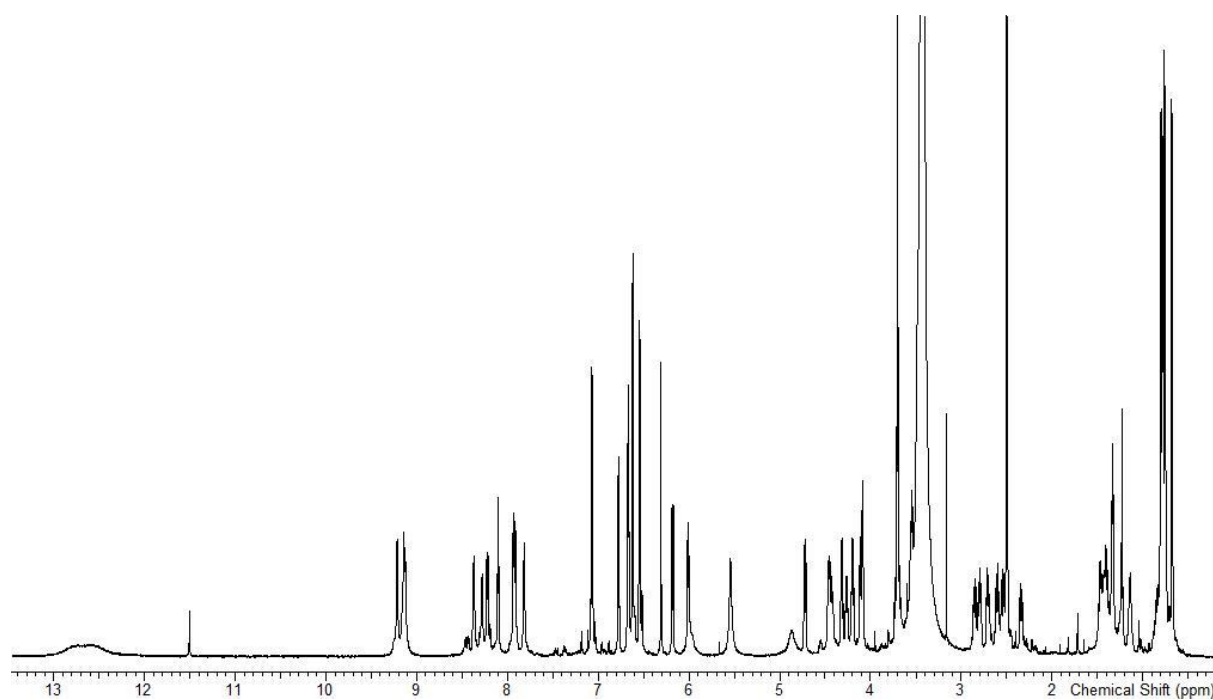
**Table S3:** NMR data (700 MHz, DMSO-*d*6) for bosamycin B.

unit	$\delta_{\text{H}}$ , multiplicity, (J in Hz)	$\delta_{\text{C}}$	$\delta_{\text{N}}$
Gly			
1 – COOH		171.7	
2 – CH $_{\alpha}$	3.71, m	41.2	
3 – NH	8.10, t (5.8)		105.2
Leu			
4 – CO		173.2	
5 – CH $_{\alpha}$	4.11, m	51.4	
6 – CH $_{\beta}$	1.34, m	40.8	
7 – CH $_{\gamma}$	1.14, m	24.4	
8 – CH $_{\delta,1}$	0.75, d (6.5)	23.7	
9 – CH $_{\delta,2}$	0.68, d (6.5)	22.1	
10 – NH	7.91, d (8.3)		121.3
o-MeO-Tyr			
11 – CO		171.4	
12 – CH $_{\alpha}$	4.32, m	54.4	
13 – CH $_{\beta,1}$	2.71, dd (13.5, 7.3)	32.2	
CH $_{\beta,2}$	2.79, dd (13.8, 8.3)		
14		115.5	
15		158.8	
16	6.31, d (2.0)	99.2	
17		158.2	
17 – OH	9.21, bs		
18	6.18, dd (8.1, 2.1)	107.1	
19	6.78, d (8.3)	131.7	
20	3.70, s	55.7	
21 – NH	7.93, d (6.8)		121.3
Ser			
22 – CO		170.5	
23 – CH $_{\alpha}$	4.20, m	56.1	
24 – CH $_{\beta,1}$	3.49, overl*	62.3	
CH $_{\beta,2}$	3.55, dd (10.5, 5.4)		
25 – NH	7.82, d (7.3)		116.3
$\beta$ -OH-Asp			
26 – CO		169.4	
27 – CH $_{\alpha}$	4.72, dd (8.7, 6.0)	55.7	
28 – CH $_{\beta}$	4.08, d (5.8)	71.7	
29 – COOH		173.5	
30 – NH	8.37, d (8.7)		112.5
Leu			
31 – CO		172.9	
32 – CH $_{\alpha}$	4.42, m	51.3	
33 – CH $_{\beta}$	1.45, m	41.5	
34 – CH $_{\gamma}$	1.40, m	24.7	

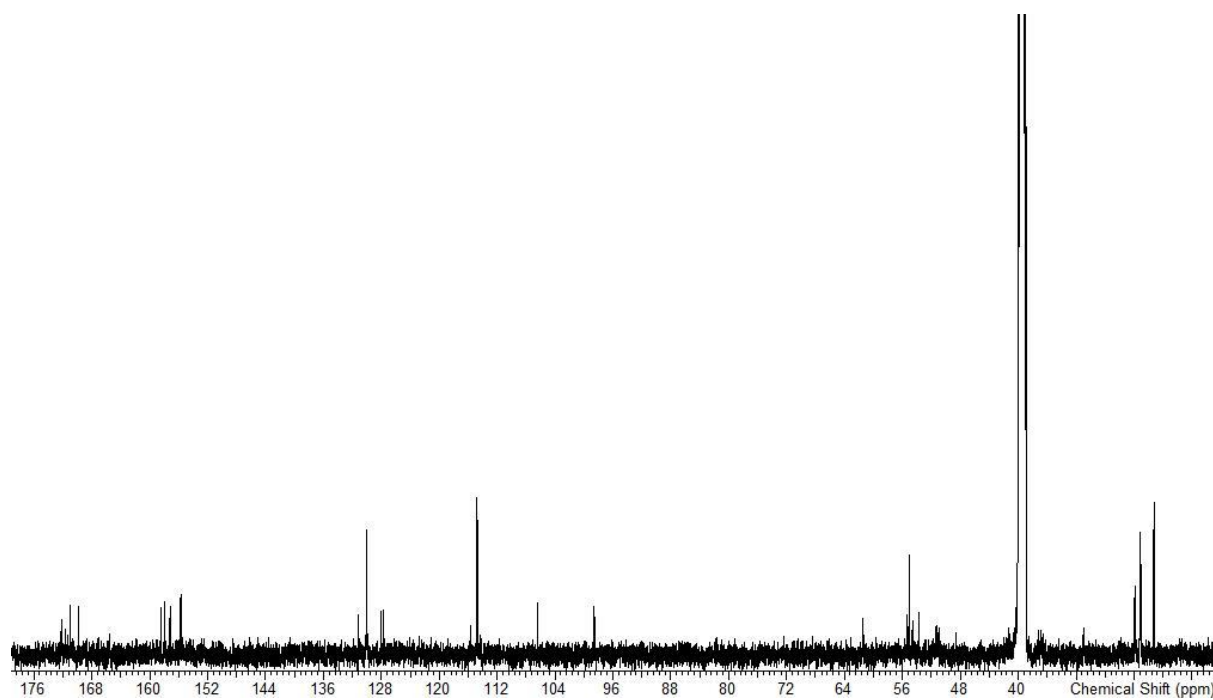
35 – CH <sub>δ,1</sub>	0.80, d (6.1)	24.1	
36 – CH <sub>δ,2</sub>	0.78, d (6.1)	21.7	
37 – NH	8.22, d (8.3)		119.5
Tyr			
38 – CO		171.9	
39 – CH <sub>α</sub>	4.45, m	55.2	
40 – CH <sub>β,1</sub>	2.60, dd (13.4, 9.8)	37.9	
CH <sub>β,2</sub>	2.85, dd (13.5, 4.4)		
41		128.7	
42, 46	7.07, d (8.5)	131.0	
43, 45	6.62, d (8.5)	115.5	
44		156.6	
44 – OH	9.14, bs		
47 – NH	8.28, d (8.0)		118.5
Tyr			
48 – CO		172.5	
49 – CH <sub>α</sub>	4.26, m	54.9	
50 – CH <sub>β,1</sub>	2.34, dd (13.6, 8.2)	38.3	
CH <sub>β,2</sub>	2.55, dd (14.1, 4.0)		
51		128.3	
52, 56	6.67, d (8.3)	130.7	
53, 55	6.55, d (8.3)	115.4	
54		156.2	
54 – OH	9.12, bs		
57 – NH	6.01, d(8.2)		87.2
FCA			
58 – CO		159.0	
59 – NH <sub>2</sub>	5.54, bs		76.1

\*signal overlapping



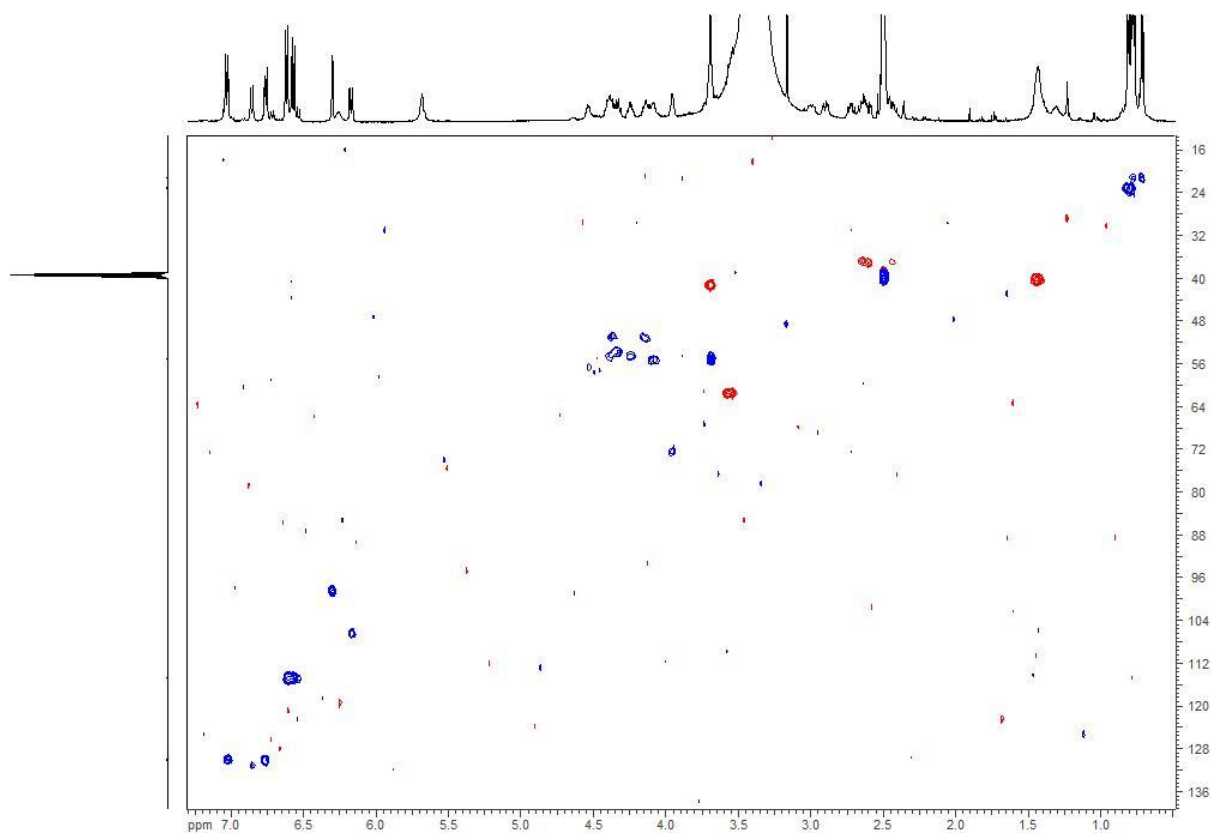


**Figure S11:**  $^1\text{H}$  NMR spectrum (700 MHz,  $\text{DMSO-}d_6$ ) of bosamycin B.

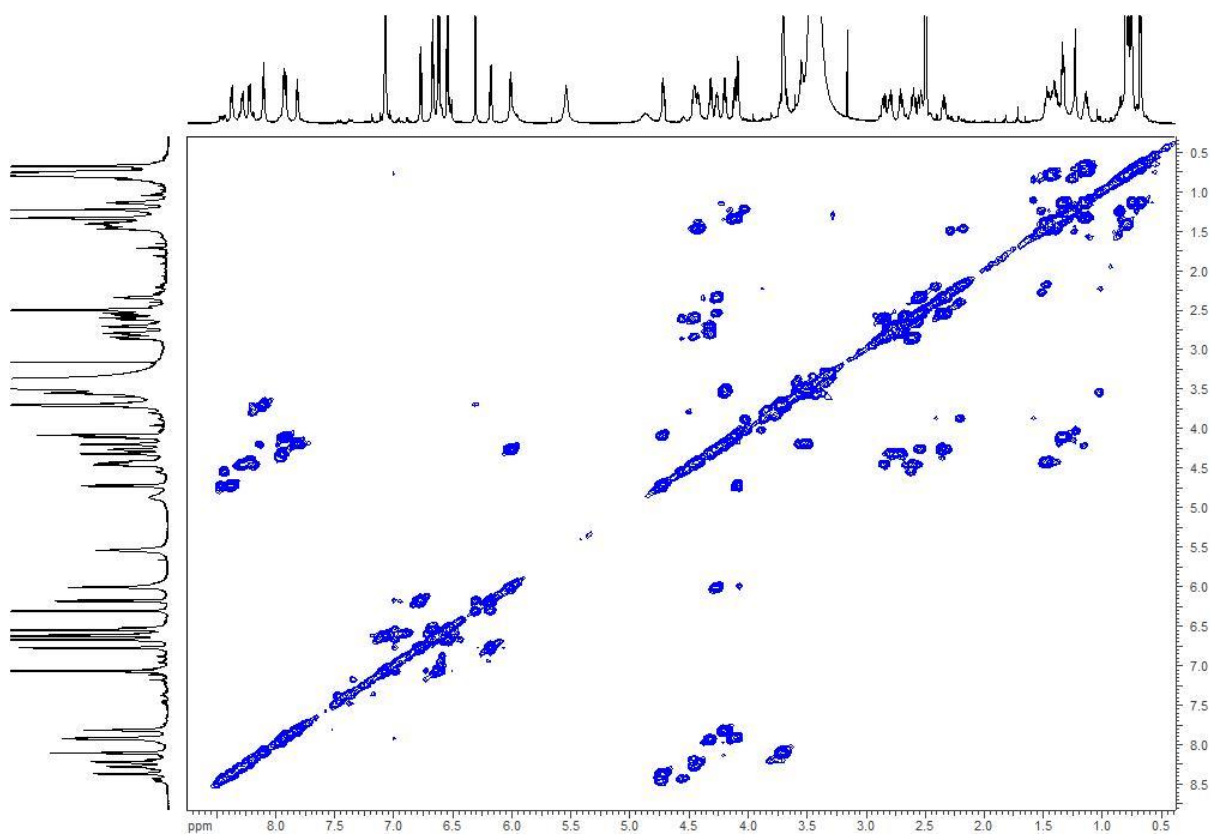


**Figure S12:**  $^{13}\text{C}$  NMR spectrum (500 MHz,  $\text{DMSO-}d_6$ ) of bosamycin B.

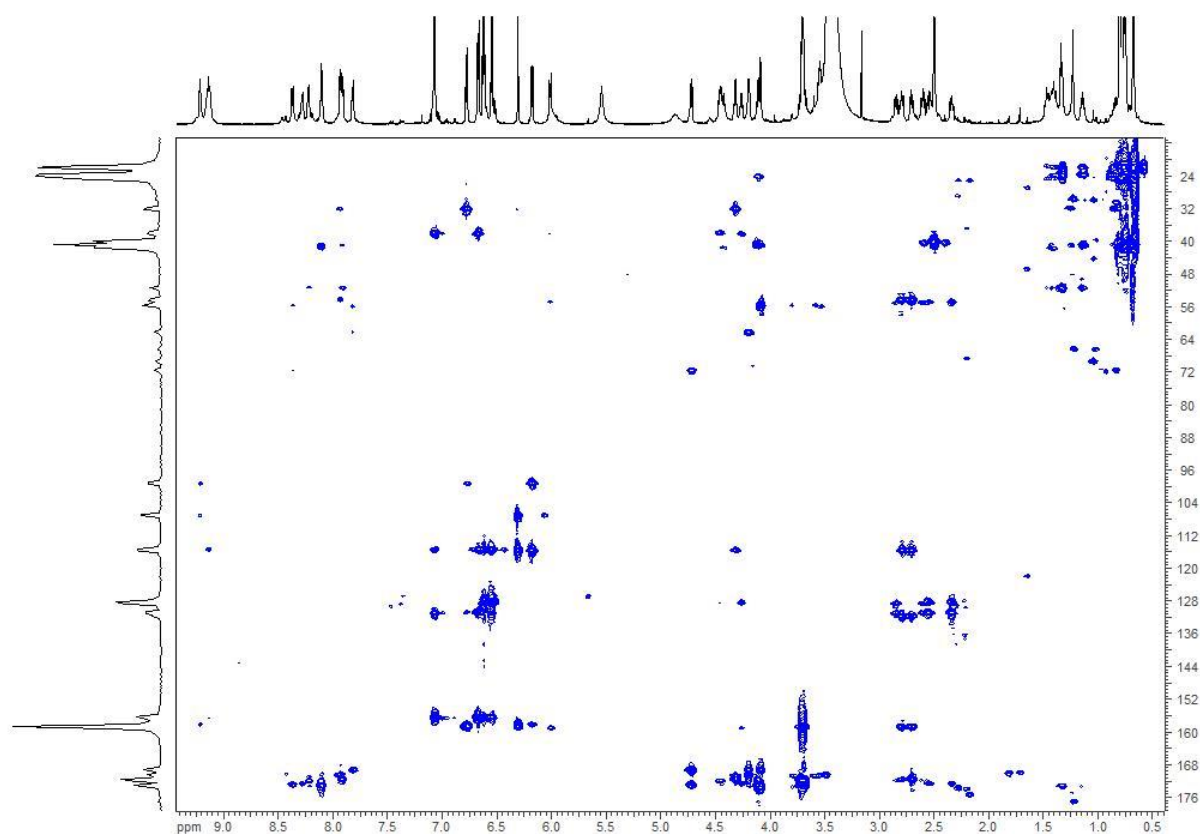




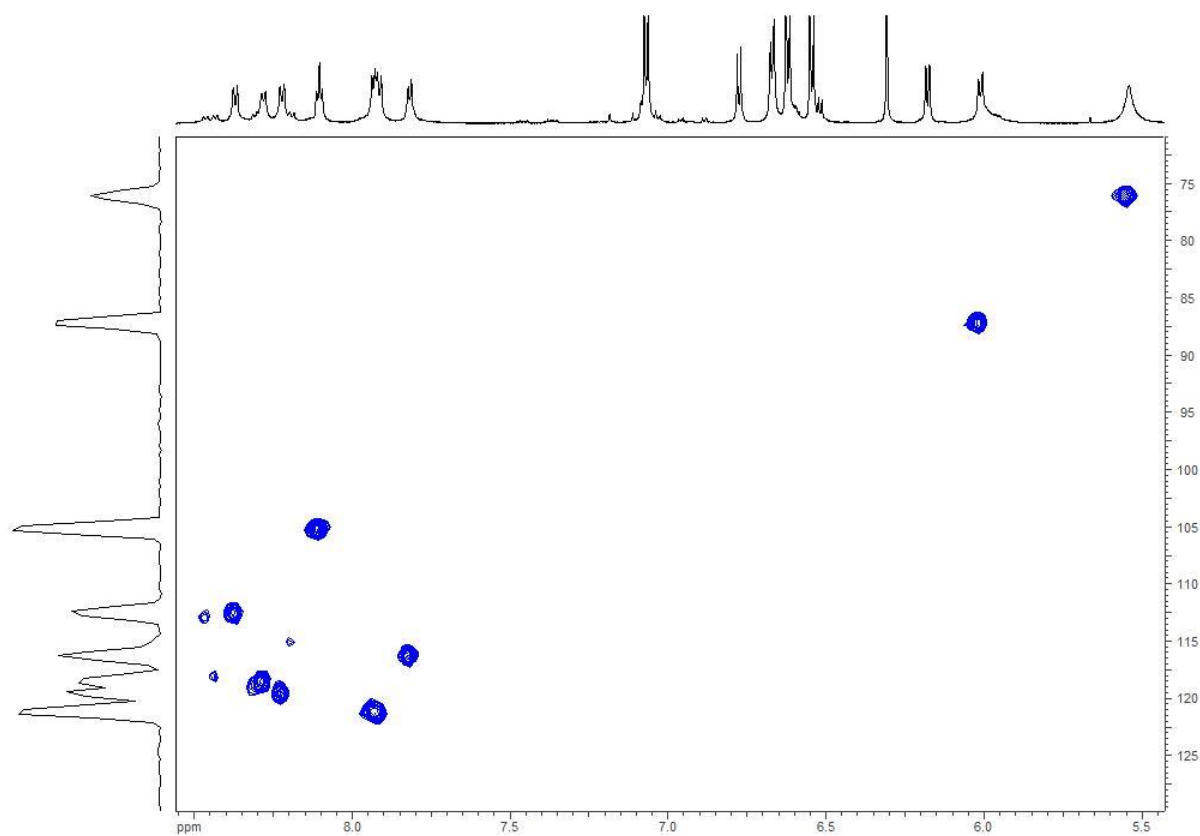
**Figure S13:** Edited HSQC spectrum (500 MHz,  $\text{DMSO-}d_6$ ) of bosamycin B.



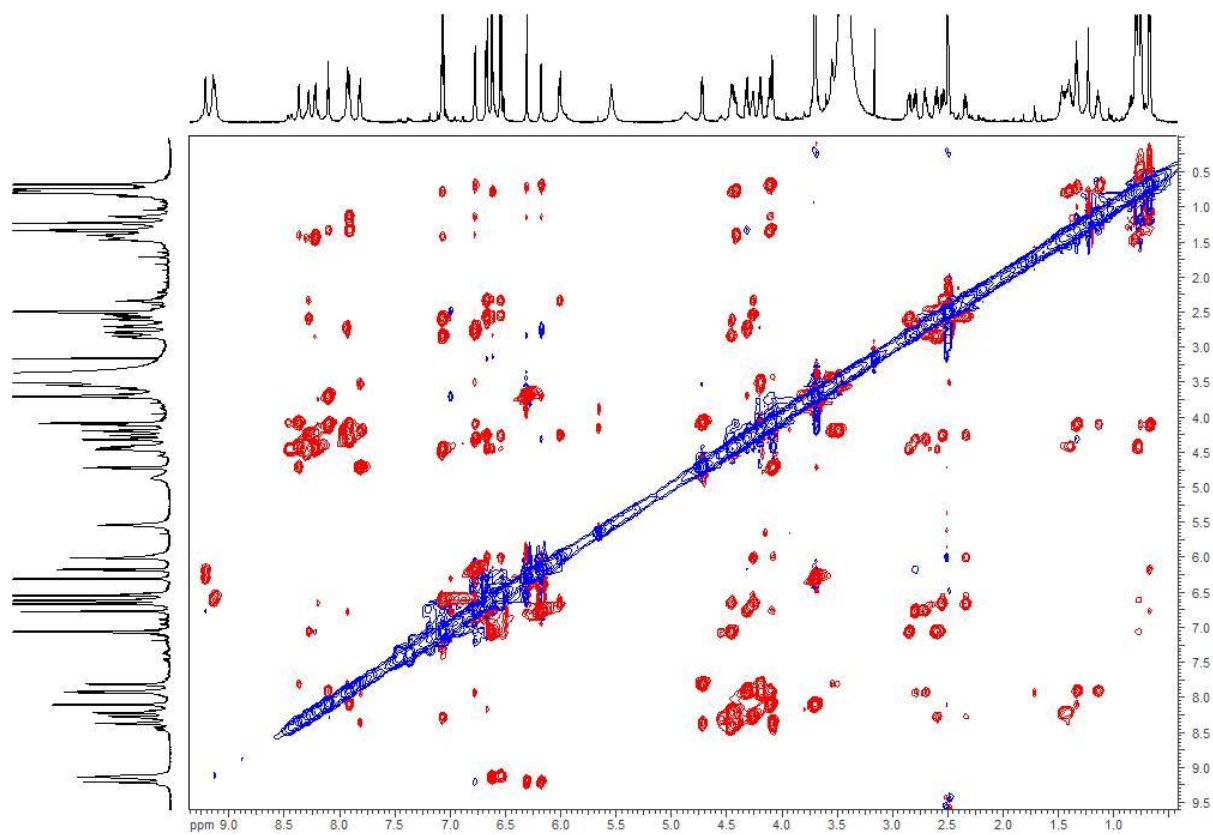
**Figure S14:** COSY spectrum (700 MHz,  $\text{DMSO-}d_6$ ) of bosamycin B.



**Figure S15:** HMBC spectrum (700 MHz, DMSO-*d*<sub>6</sub>) of bosamycin B.



**Figure S16:** N-HSQC spectrum (700 MHz, DMSO-*d*<sub>6</sub>) of bosamycin B.



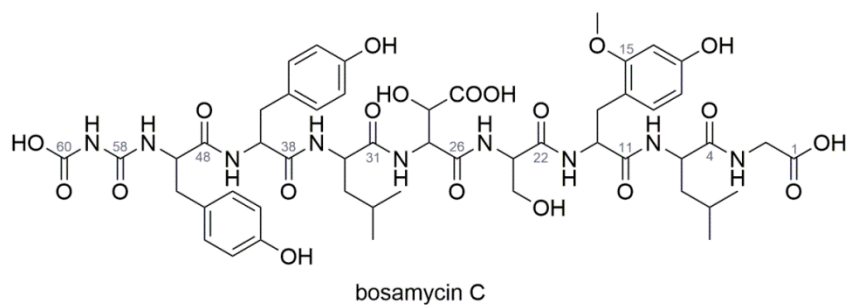
**Figure S17:** ROESY spectrum (700 MHz, DMSO-*d*<sub>6</sub>) of bosamycin B.

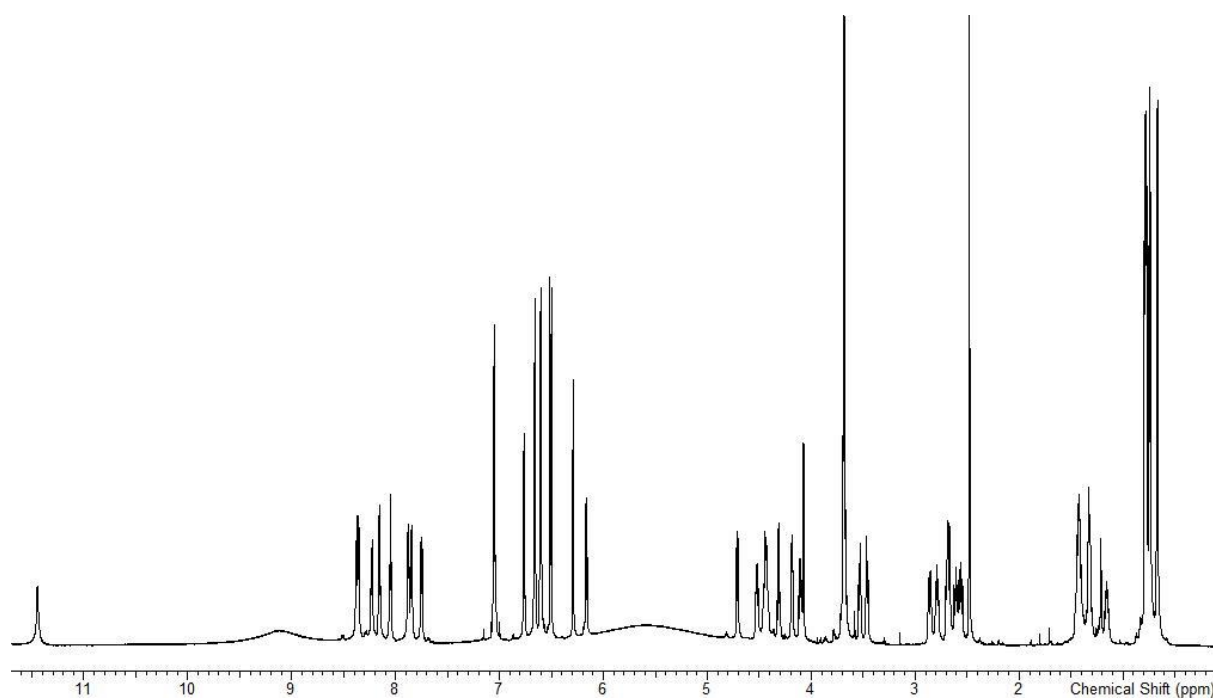
**Table S4:** NMR data (700 MHz, DMSO-*d*<sub>6</sub>+TFA) for bosamycin C.

unit	$\delta_{\text{H}}$ , multiplicity, (J in Hz)	$\delta_{\text{C}}$	$\delta_{\text{N}}$
Gly			
1 – COOH		171.4	
2 – CH $_{\alpha}$	3.70, ovl*	41.5	
3 – NH	8.53, bs		105.0
Leu			
4 – CO		172.4	
5 – CH $_{\alpha}$	4.15, dt (5.2, 9.2)	51.4	
6 – CH $_{\beta}$	1.44, m	40.1	
7 – CH $_{\gamma}$	1.34, m	23.8	
8 – CH $_{\delta,1}$	0.80, ovl*	23.1	
9 – CH $_{\delta,2}$	0.73, d (6.4)	21.2	
10 – NH	7.90, d (8.5)		120.4
o-MeO-Tyr			
11 – CO		171.0	
12 – CH $_{\alpha}$	4.33, pq	53.7	
13 – CH $_{\beta,1}$	3.03, m	31.0	
CH $_{\beta,2}$	2.72, ovl*		
14		115.8	
15		158.0	
16	6.30, d (2.2)	98.6	
17		157.2	
17 – OH	9.15, bs		
18	6.17, dd (2.2, 8.1)	106.4	
19	6.87, d (8.2)	131.2	
20	3.70, s	55.1	
21 – NH	8.15, bs		120.4
Ser			
22 – CO		169.9	
23 – CH $_{\alpha}$	4.05, m	55.2	
24 – CH $_{\beta,1}$	3.62, dd (4.4, 10.2)	61.4	
CH $_{\beta,2}$	3.52, dd (3.7, 10.5)		
25 – NH	7.68, d (6.5)		115.5
$\beta$ -OH-Asp			
26 – CO		168.9	
27 – CH $_{\alpha}$	4.64, dd (3.6, 7.5)	56.2	
28 – CH $_{\beta}$	3.97, d (3.0)	72.7	
29 – COOH		174.5	
30 – NH	8.11, bs		112.6
Leu			
31 – CO		171.8	
32 – CH $_{\alpha}$	4.42, m	50.7	
33 – CH $_{\beta}$	1.44, m	40.8	

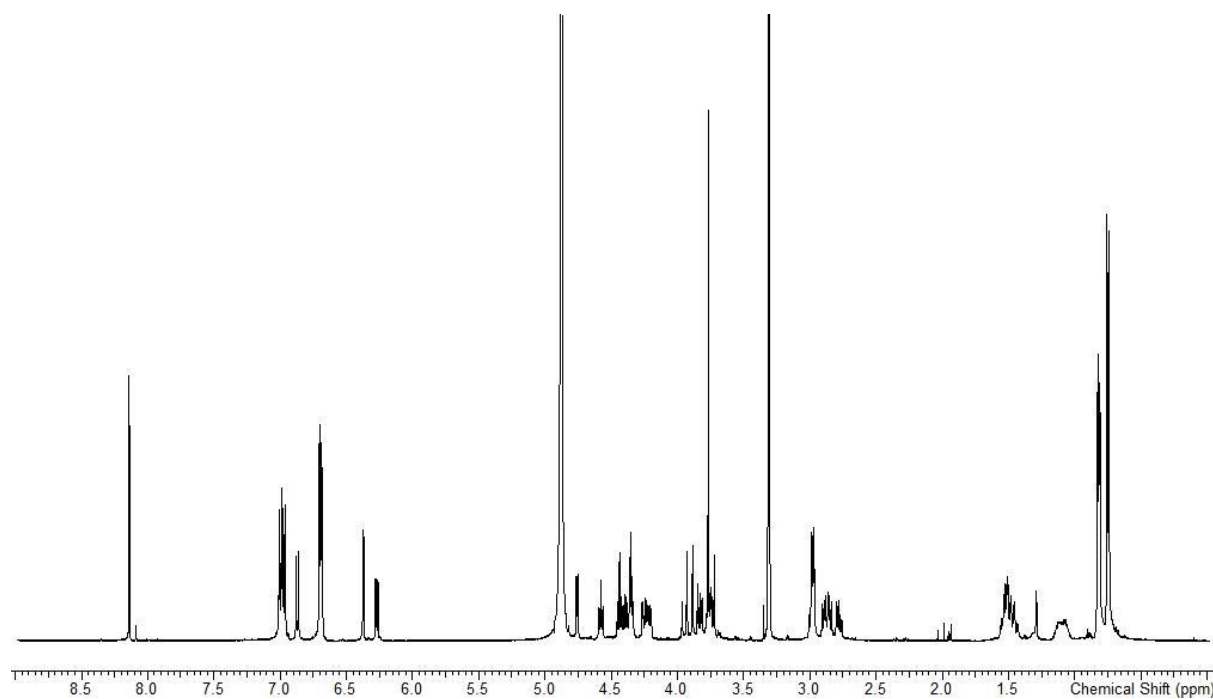
34 – CH <sub>γ</sub>	1.43, ovl*	24.0	
35 – CH <sub>δ,1</sub>	0.82, d (5.9)	23.3	
36 – CH <sub>δ,2</sub>	0.78, ovl*	21.2	
37 – NH	8.23, d (7.9)		119.3
Tyr			
38 – CO		170.7	
39 – CH <sub>α</sub>	4.52, dt (5.1, 8.9)	54.5	
40 – CH <sub>β,1</sub>	2.88, dd (5.3, 13.8)	37.5	
CH <sub>β,2</sub>	2.66, ovl*		
41		127.6	
42, 46	7.05 (8.4)	130.2	
43, 45	6.62, d (8.4)	114.8	
44		155.9	
44 – OH	9.11, bs		
47 – NH	8.42, d( 7.34)		117.9
Tyr			
48 – CO		169.7	
49 – CH <sub>α</sub>	4.48, dt (4.5, 8.7)	54.0	
50 – CH <sub>β,1</sub>	2.73, ovl*	36.5	
CH <sub>β,2</sub>	2.64, ovl*		
51		127.2	
52, 56	6.72, d (8.4)	130.1	
53, 55	6.54, d (8.4)	114.8	
54		155.8	
54 – OH	9.11, bs		
57 – NH	8.27, d (8.7)		114.9
FCA			
58 – CO		158.8	
59 – NH	11.51, s		168.2
60 – COOH		156.3	

\*signal overlapping

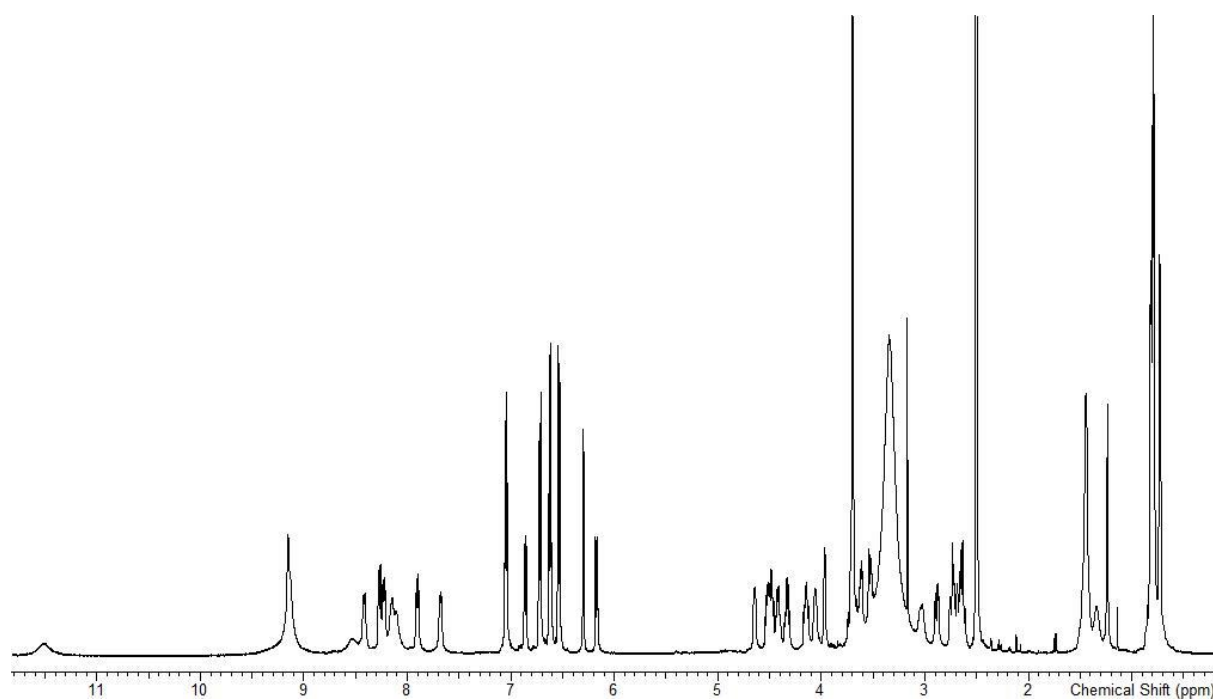




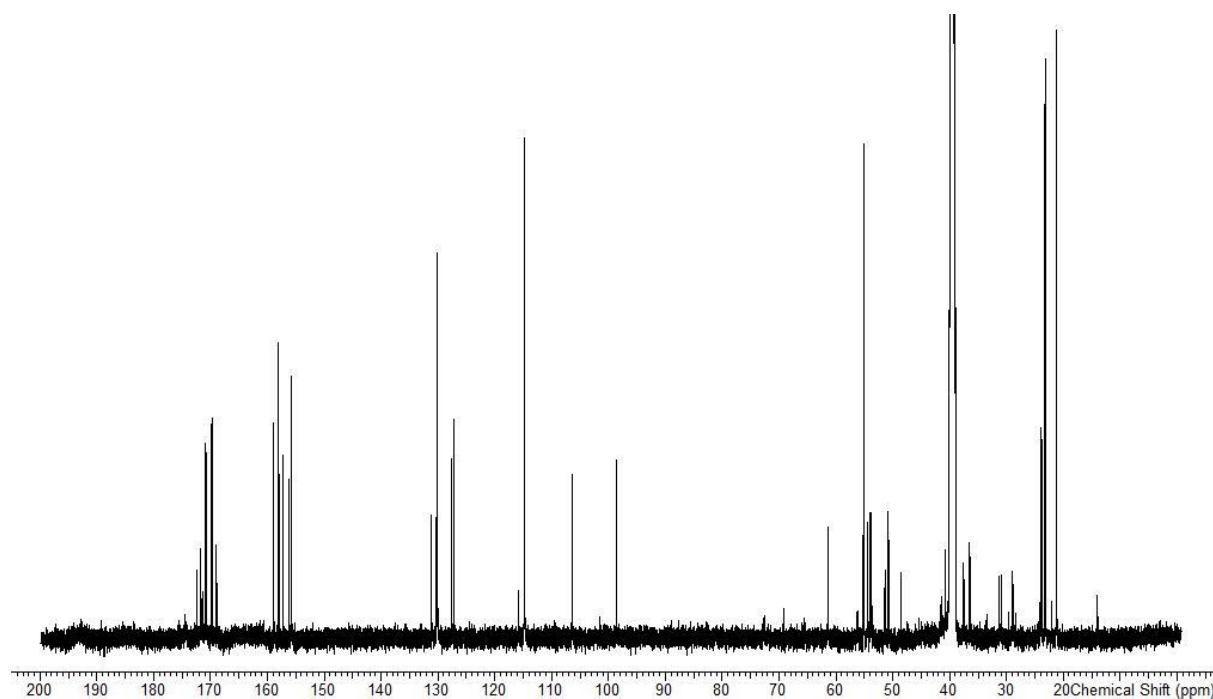
**Figure S18:**  $^1\text{H}$  NMR spectrum (700 MHz,  $\text{DMSO-}d_6/\text{TFA}$ ) of bosamycin C.



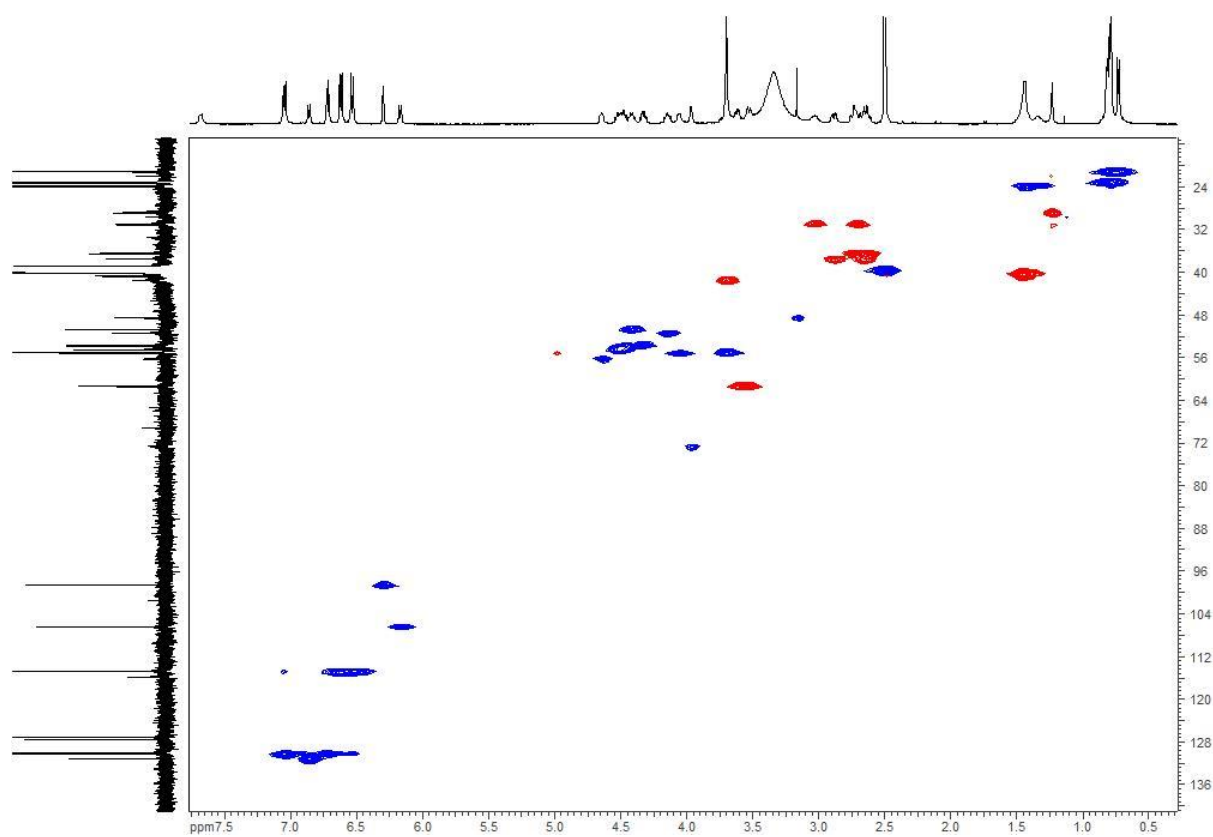
**Figure S19:**  $^1\text{H}$  NMR spectrum (700 MHz, MeOD) of bosamycin C.



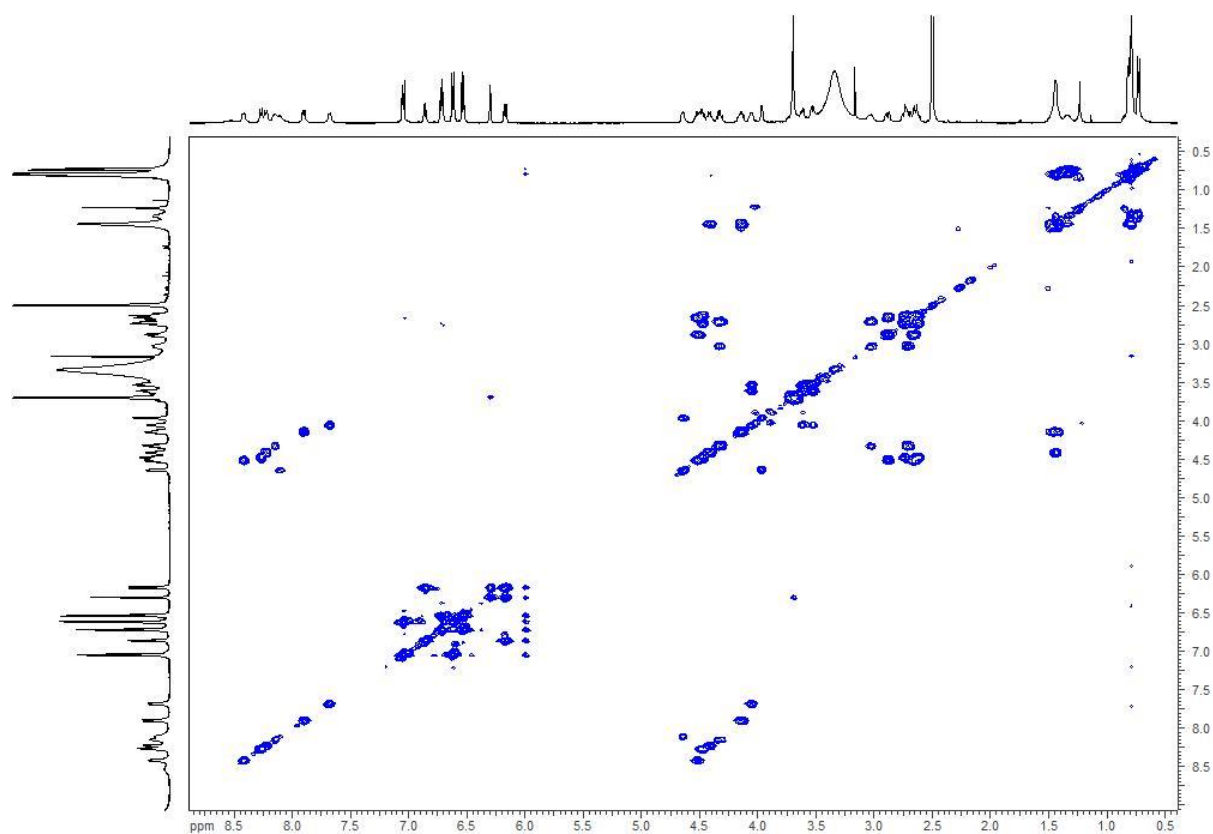
**Figure S20:**  $^1\text{H}$  NMR spectrum (700 MHz,  $\text{DMSO-}d_6$ ) of bosamycin C.



**Figure S21:**  $^{13}\text{C}$  NMR spectrum (700 MHz,  $\text{DMSO-}d_6$ ) of bosamycin C.

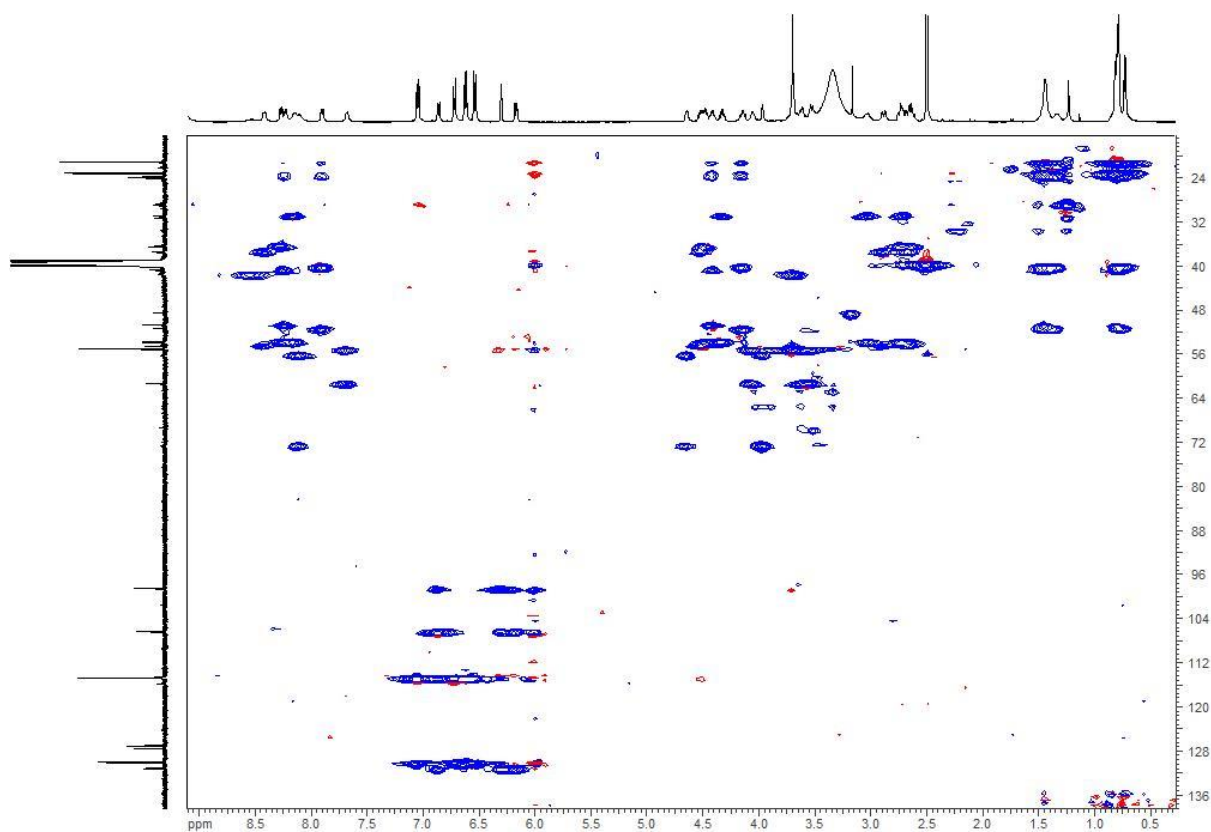


**Figure S22:** HSQC spectrum (700 MHz, DMSO-*d*<sub>6</sub>) of bosamycin C.

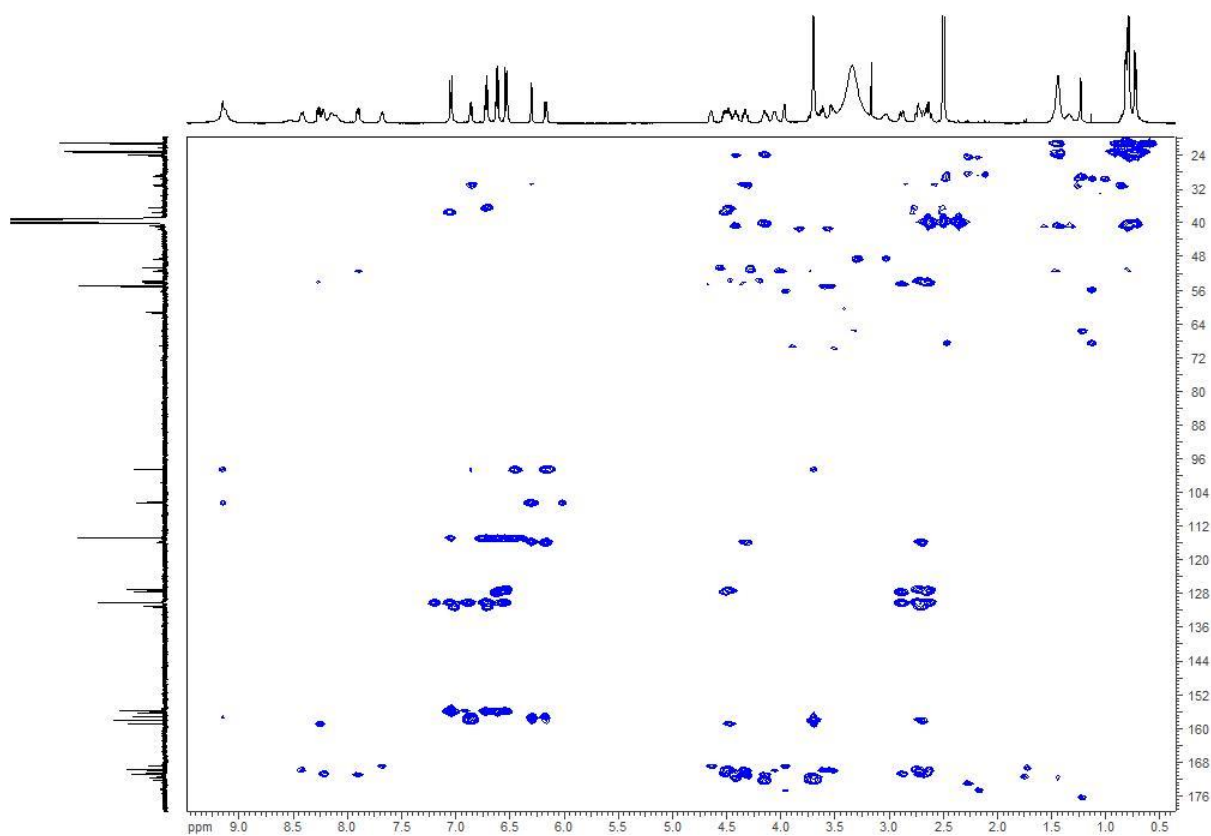


**Figure S23:** COSY spectrum (700 MHz, DMSO-*d*<sub>6</sub>) of bosamycin C.

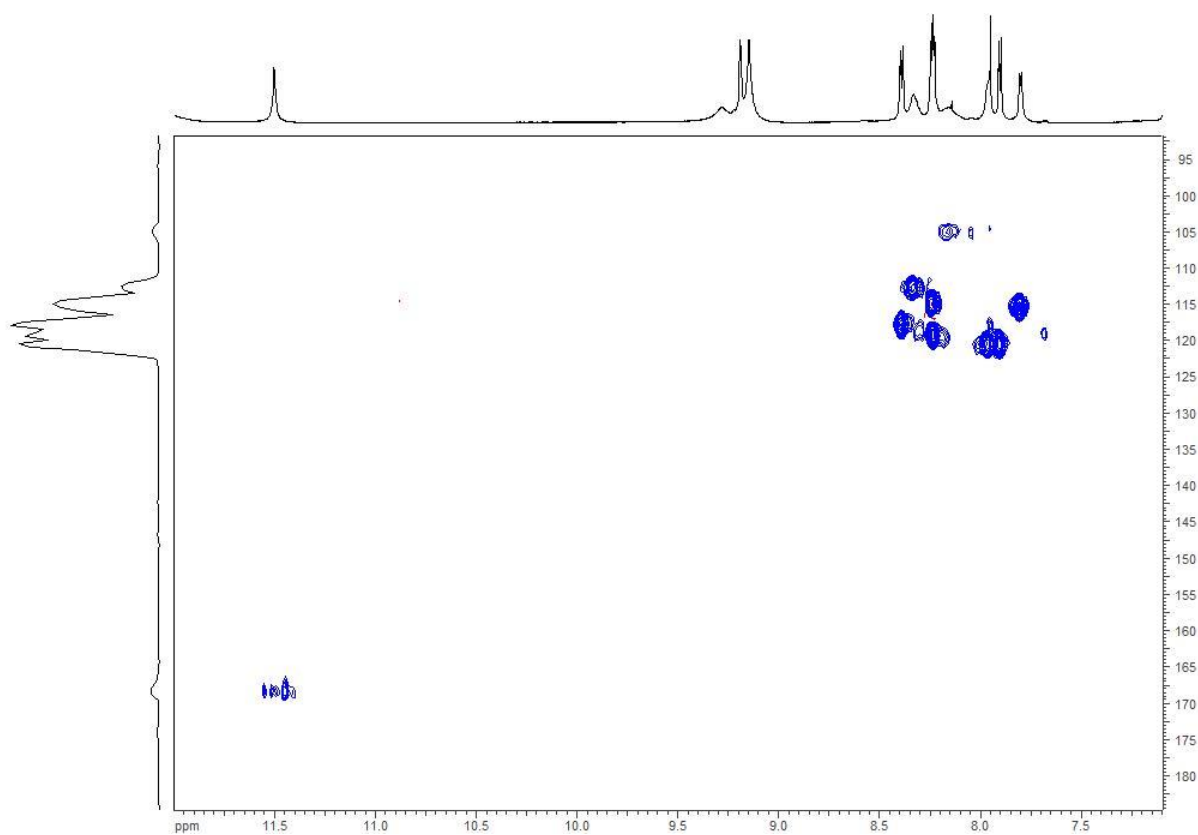




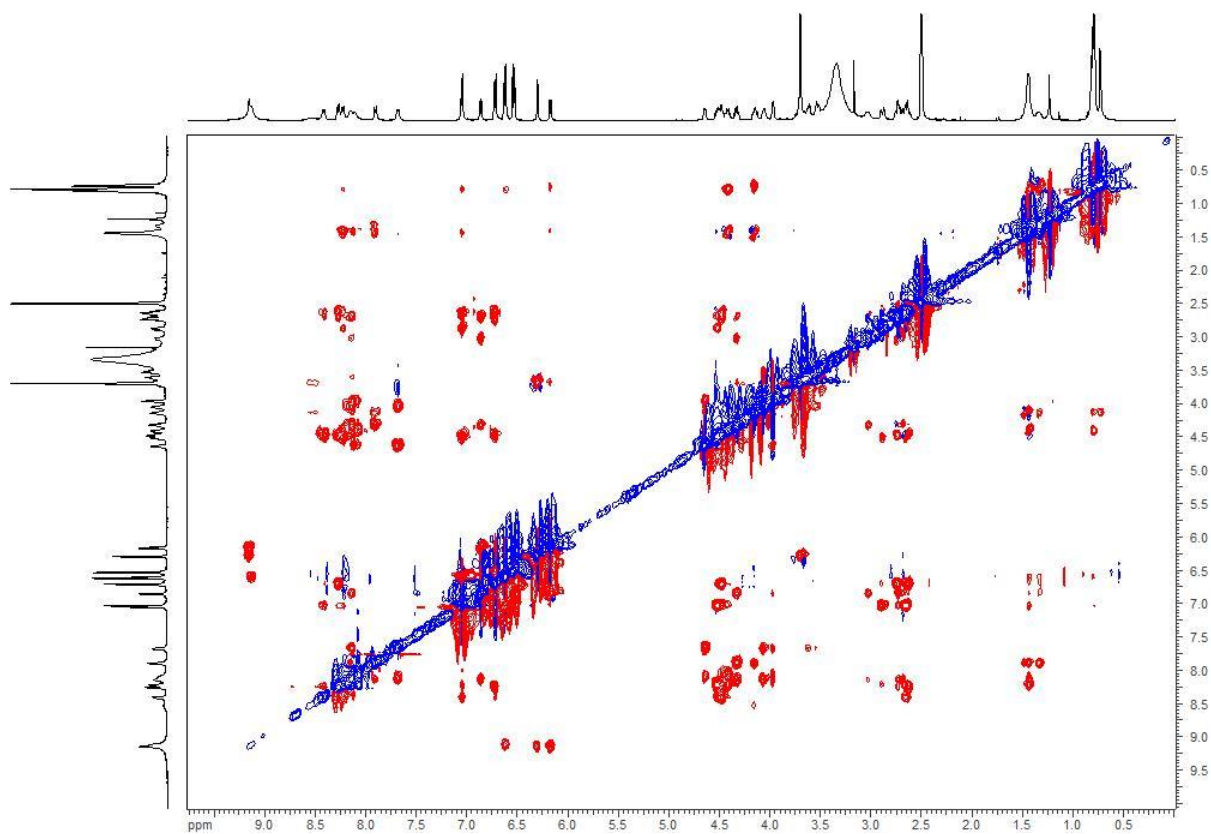
**Figure S24:** HSQC-TOCSY spectrum (700 MHz, DMSO-*d*<sub>6</sub>) of bosamycin C.



**Figure S25:** HMBC spectrum (700 MHz, DMSO-*d*<sub>6</sub>) of bosamycin C.



**Figure S26:** N-HSQC spectrum (700 MHz,  $\text{DMSO-}d_6$ ) of bosamycin C.



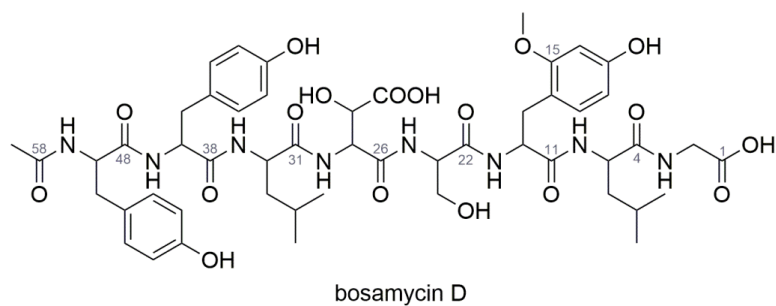
**Figure S27:** ROESY spectrum (700 MHz,  $\text{DMSO-}d_6$ ) of bosamycin C.

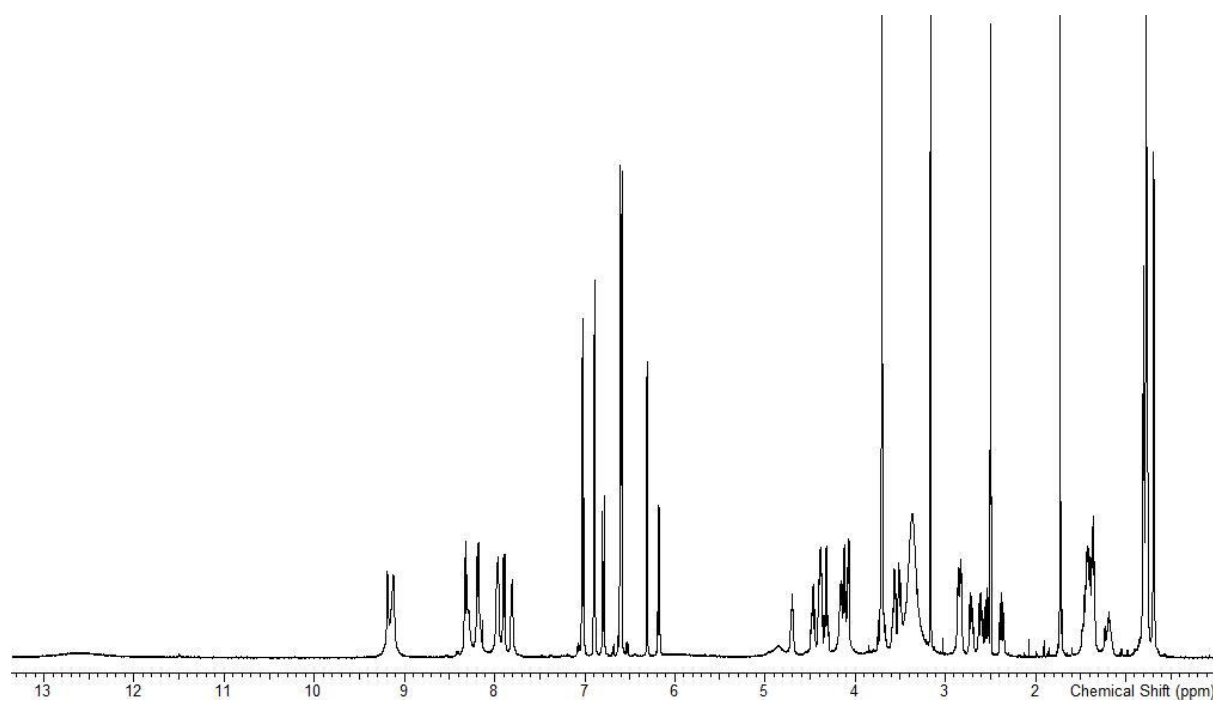
**Table S5:** NMR data (500 MHz, DMSO-*d*6) for bosamycin D.

unit	$\delta_{\text{H}}$ , multiplicity, (J in Hz)	$\delta_{\text{C}}$
Gly		
1 – COOH		171.1
2 – CH $_{\alpha}$	3.71, ovl*	40.8
3 – NH	8.17, ovl*	
Leu		
4 – CO		172.4
5 – CH $_{\alpha}$	4.12, ovl*	50.9
6 – CH $_{\beta}$	1.36, t (7.3)	40.3
7 – CH $_{\gamma}$	1.20, m	23.7
8 – CH $_{\delta,1}$	0.76, d (5.0)	23.2
9 – CH $_{\delta,2}$	0.69, d (6.5)	21.3
10 – NH	7.90, d (8.2)	
o-MeO-Tyr		
11 – CO		170.9
12 – CH $_{\alpha}$	4.32, dt (7.2, 7.6)	53.7
13 – CH $_{\beta,1}$	2.84, ovl*	31.4
CH $_{\beta,2}$	2.71, dd (13.2, 7.5)	
14		115.2
15		158.1
16	6.31, d (2.2)	98.6
17	-	157.4
17 – OH	9.18, bs	
18	6.18, dd (8.0, 2.12)	106.4
19	6.80, d (8.2)	131.0
20	3.70, s	55.1
21 – NH	7.96, bs	
Ser		
22 – CO		169.8
23 – CH $_{\alpha}$	4.16, ovl*	55.3
24 – CH $_{\beta,1}$	3.57, dd (11.0, 5.6)	61.6
CH $_{\beta,2}$	3.51 dd (10.5, 4.1)	
25 – NH	7.81, d (7.6)	
$\beta$ -OH-Asp		
26 – CO		168.7
27 – CH $_{\alpha}$	4.70, t (6.0)	55.3
28 – CH $_{\beta}$	4.08, d (5.6)	71.3
29 – COOH		173.0
30 – NH	8.29, m	
Leu		
31 – CO		172.1
32 – CH $_{\alpha}$	4.39, ovl*	50.8
33 – CH $_{\beta}$	1.43, ovl*	40.7
34 – CH $_{\gamma}$	1.40, ovl	24.0

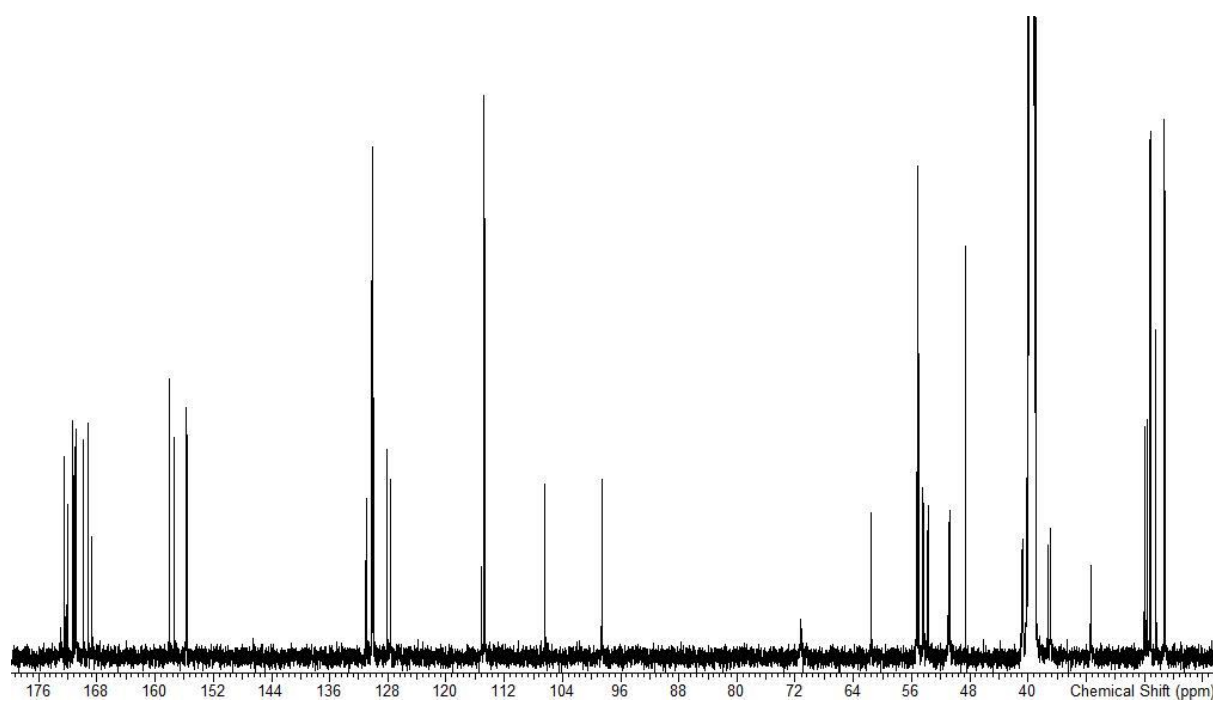
35 – CH <sub>δ,1</sub>	0.80, d (5.9)	23.3
36 – CH <sub>δ,2</sub>	0.77, d (3.6)	21.2
37 – NH	8.18, d (8.2)	
Tyr		
38 – CO		171.1
39 – CH <sub>α</sub>	4.47, m	54.3
40 – CH <sub>β,1</sub>	2.84, dd (13.6, 5.2)	37.3
CH <sub>β,2</sub>	2.61 dd (13.6, 3.8)	
41		127.6
42, 46	7.02, d (8.5)	130.2
43, 45	6.59, d (3.8)	114.7
44		155.8
44 – OH	9.12, bs	
47 – NH	8.32, d (8.2)	
Tyr		
48 – CO		171.3
49 – CH <sub>α</sub>	4.38, ovl	54.5
50 – CH <sub>β,1</sub>	2.55, dd (13.9, 4.0)	36.9
CH <sub>β,2</sub>	2.37, dd (13.6, 9.9)	
51		128.1
52, 56	6.89, d (8.6)	130.0
53, 55	6.60, d (3.8)	114.8
54		155.7
54 – OH	9.12, bs	
57 – NH	7.97, bs	
Acetyl		
58 – CO		169.3
59	1.72, s	22.4

\*signal overlapping

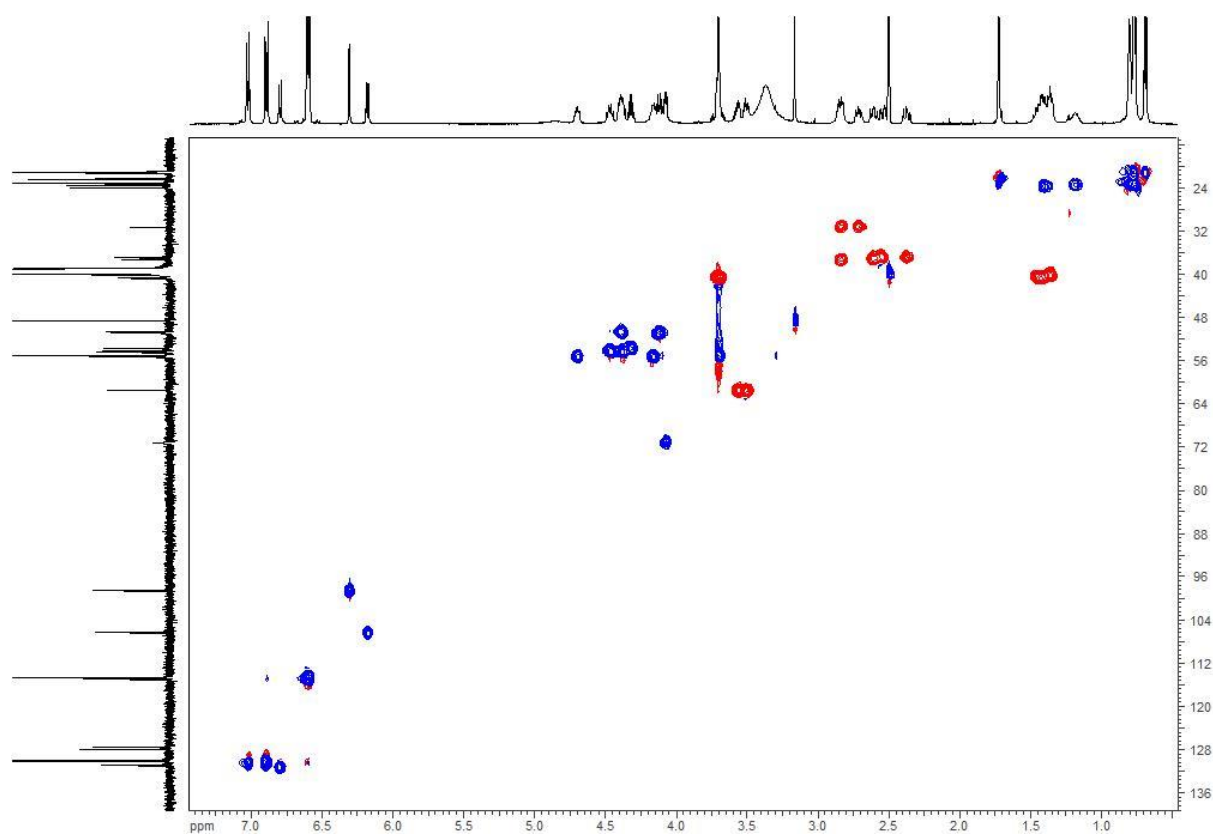




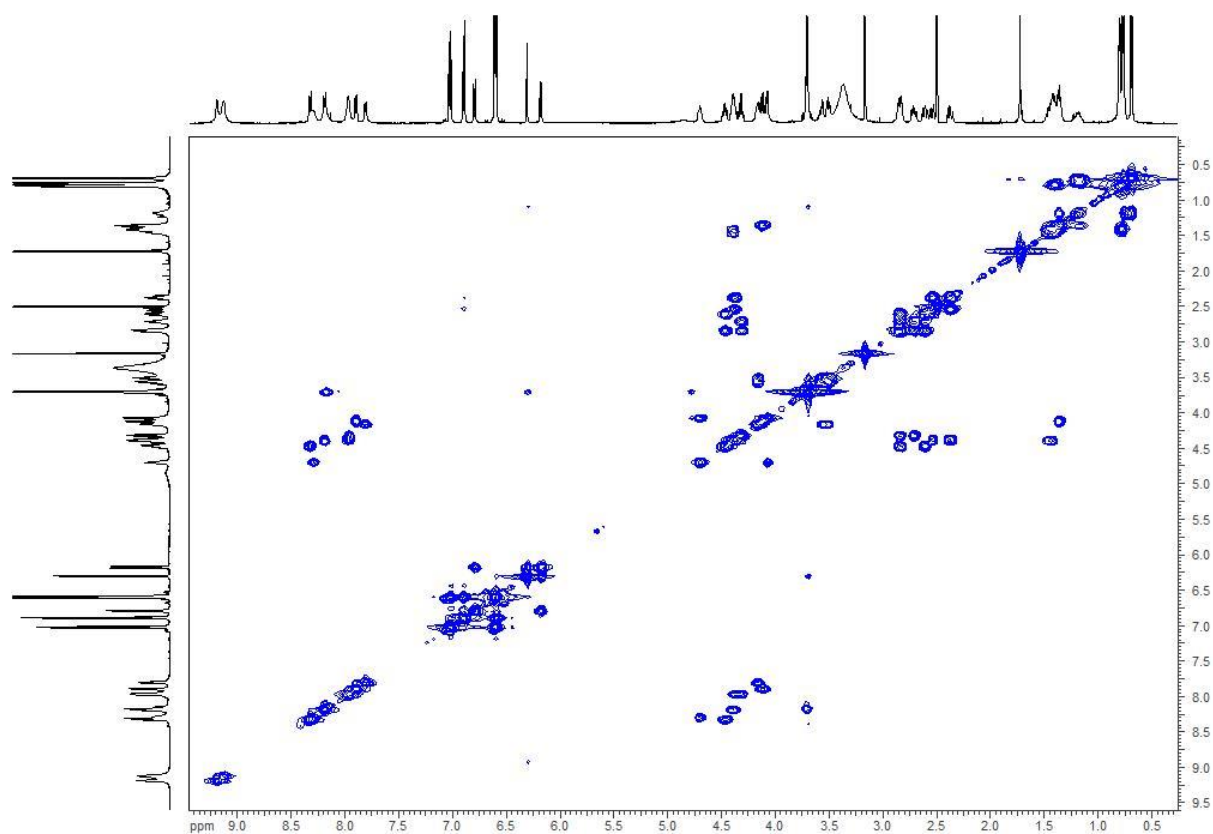
**Figure S28:**  $^1\text{H}$  NMR spectrum (500 MHz,  $\text{DMSO-}d_6$ ) of bosamycin D.



**Figure S29:**  $^{13}\text{C}$  NMR spectrum (500 MHz,  $\text{DMSO-}d_6$ ) of bosamycin D.



**Figure S30:** Edited HSQC spectrum (500 MHz, DMSO-*d*<sub>6</sub>) of bosamycin D.



**Figure S31:** COSY spectrum (500 MHz, DMSO-*d*<sub>6</sub>) of bosamycin D.

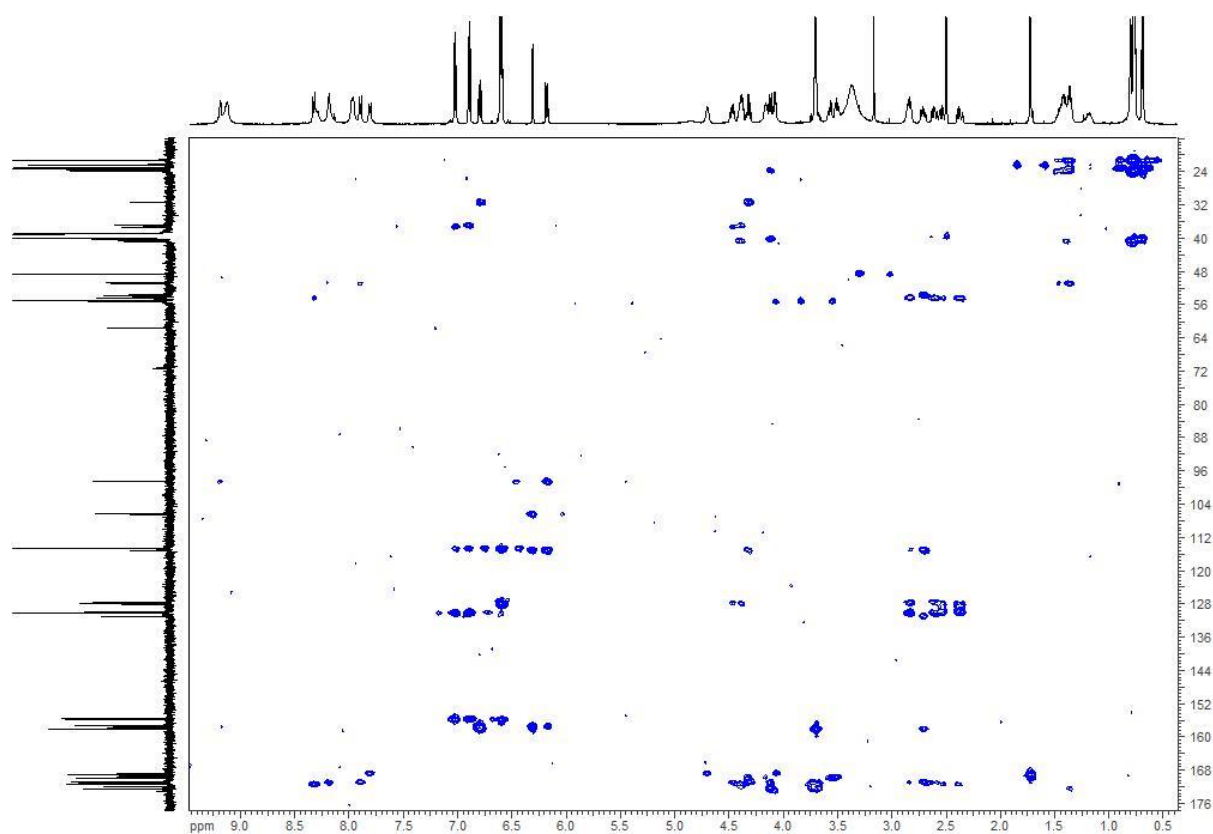


Figure S32: HMBC spectrum (500 MHz, DMSO-*d*<sub>6</sub>) of bosamycin D.

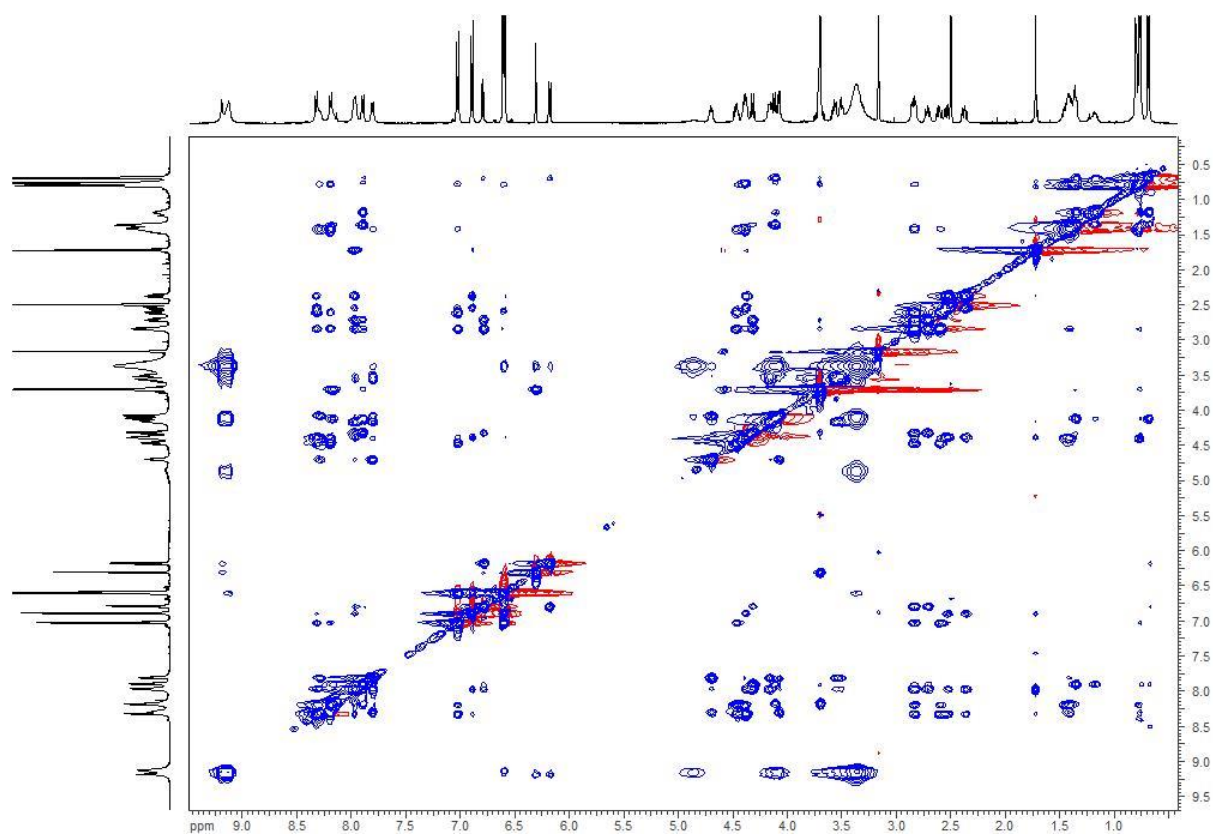
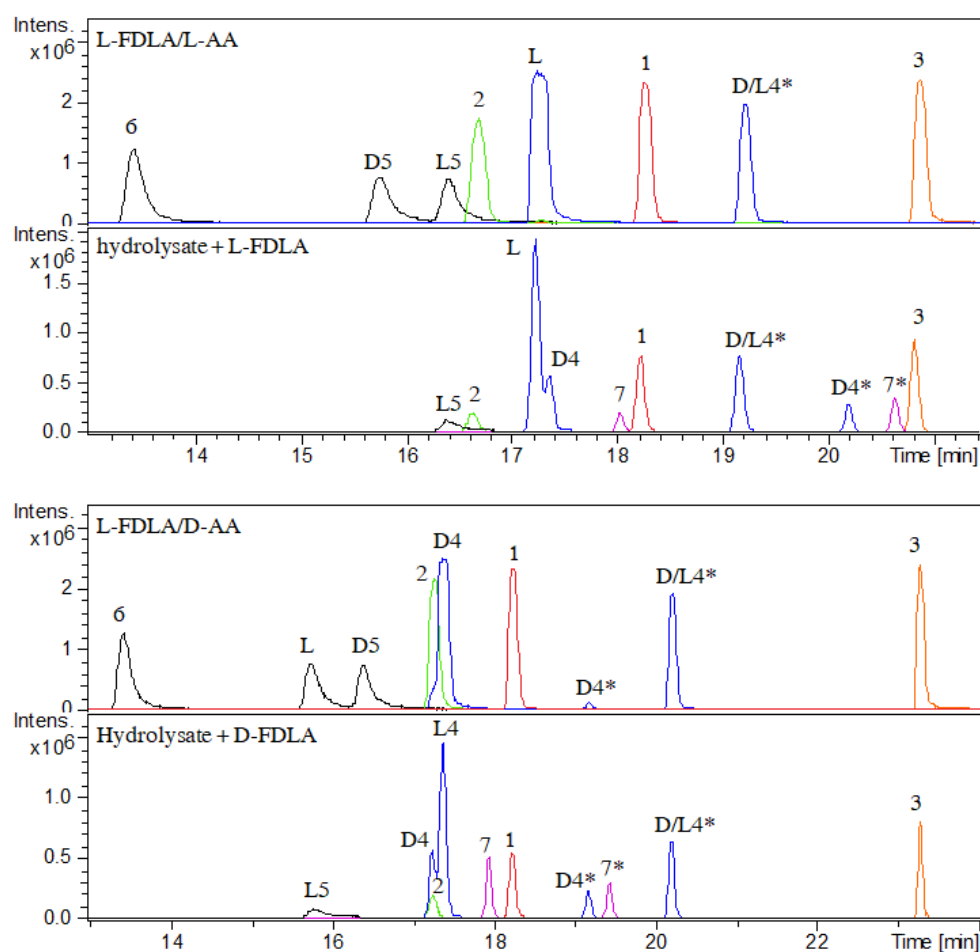
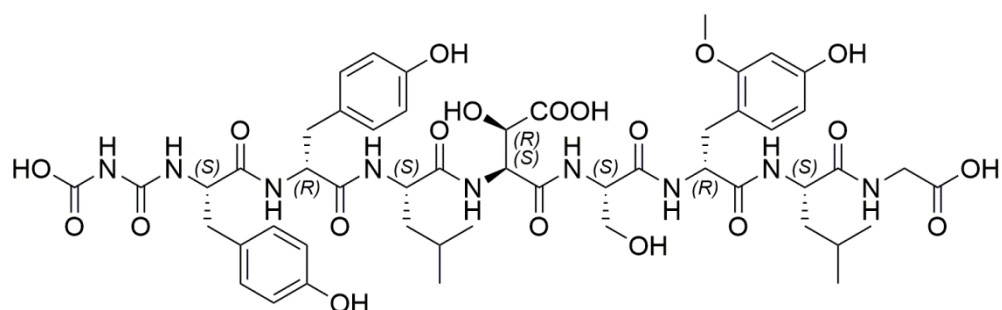


Figure S33: NOESY spectrum (500 MHz, DMSO-*d*<sub>6</sub>) of bosamycin D.

#### 4. Stereochemical Assignment by Marfey's Method

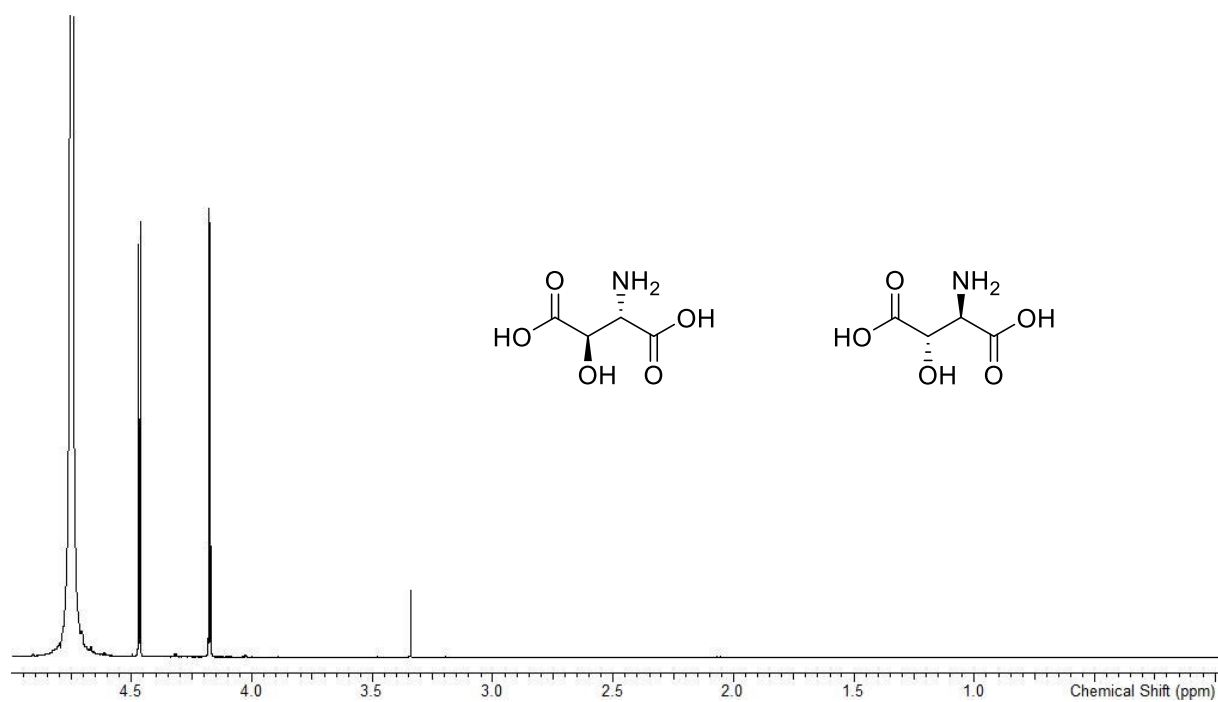


**Figure S34:** Marfey's method: MS spectra of bosamycin C derivatized with D-/L-FDLA (bottom) and the reference amino acids (AA) derivatized with D-/L-FDLA (top). The annotation is as follows: Gly (1), L-Ser (2), L-Leu (3), L-Tyr (L4, L4\*), D-Tyr (D4, D4\*), L-erythro-2-OH-Asp (L5), D-erythro-2-OH-Asp (D5), DL-threo-2-OH-Asp (6) and *o*-MeO-D-Tyr (7, 7\*). L4\*, D4\* and 7\* represent AA+2xFDLA. The exact location of D- and L-tyrosine was achieved later on by comparison of the biosynthetic gene cluster domains. 5-MeO Tyr was identified as D-conformer due to a shorter retention time when derivatized with D-FDLA

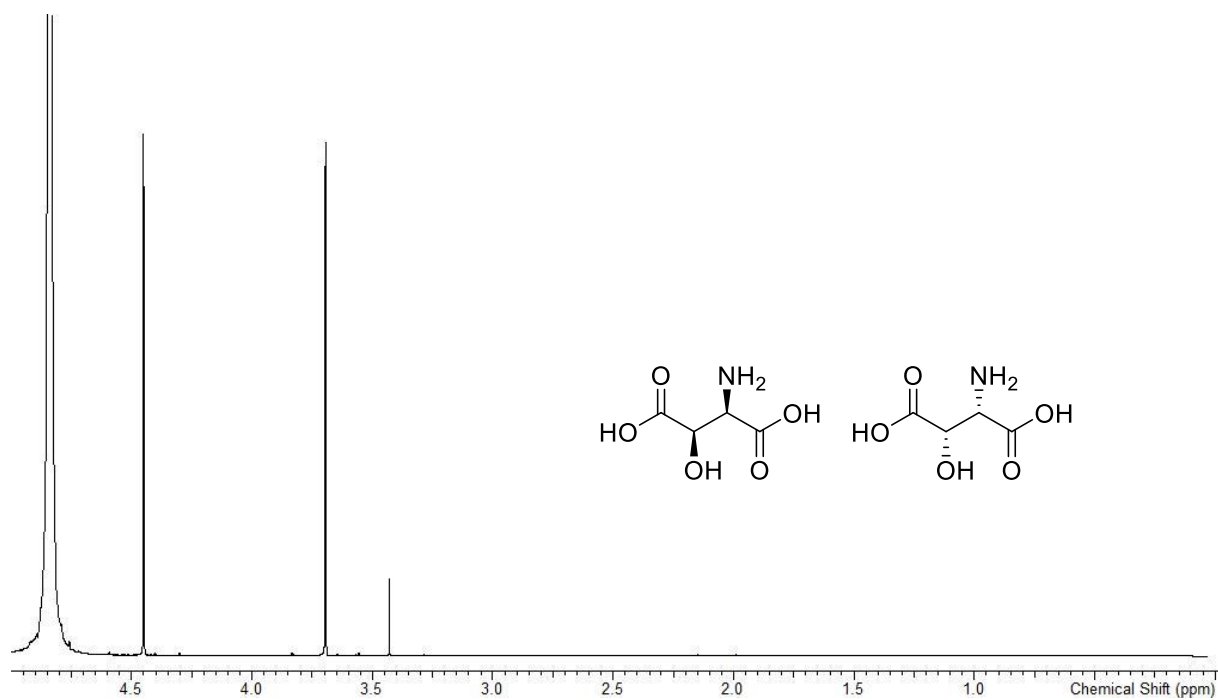


**Figure S35:** Structure of bosamycin C with stereochemistry determined by Marfey's method. The stereo chemical assignment of bosamycin C was in agreement with the data reported in the literature [3].





**Figure S36:**  $^1\text{H}$  NMR spectrum (500 MHz,  $\text{D}_2\text{O}$ ) of DL-erythro- $\beta$ -hydroxyaspartic acid.



**Figure S37:**  $^1\text{H}$  NMR spectrum (500 MHz,  $\text{D}_2\text{O}$ ) of DL-threo- $\beta$ -hydroxyaspartic acid.

## 5. MS/MS Fragmentation Data

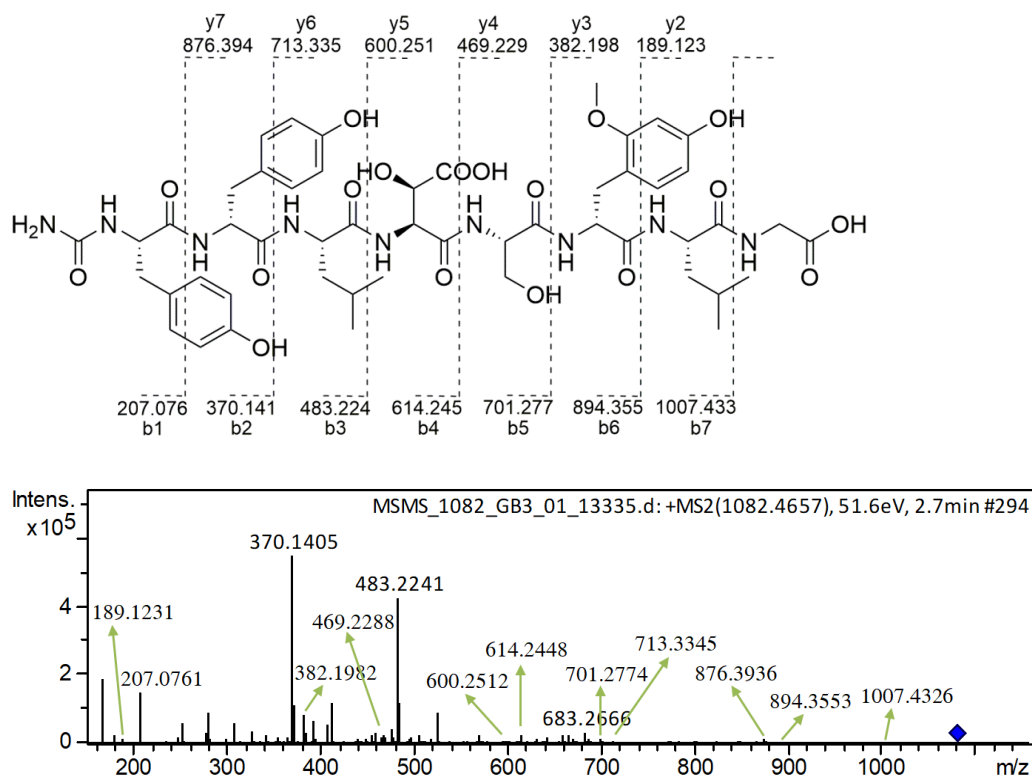


Figure S38: MS/MS analysis of bosamycin B.

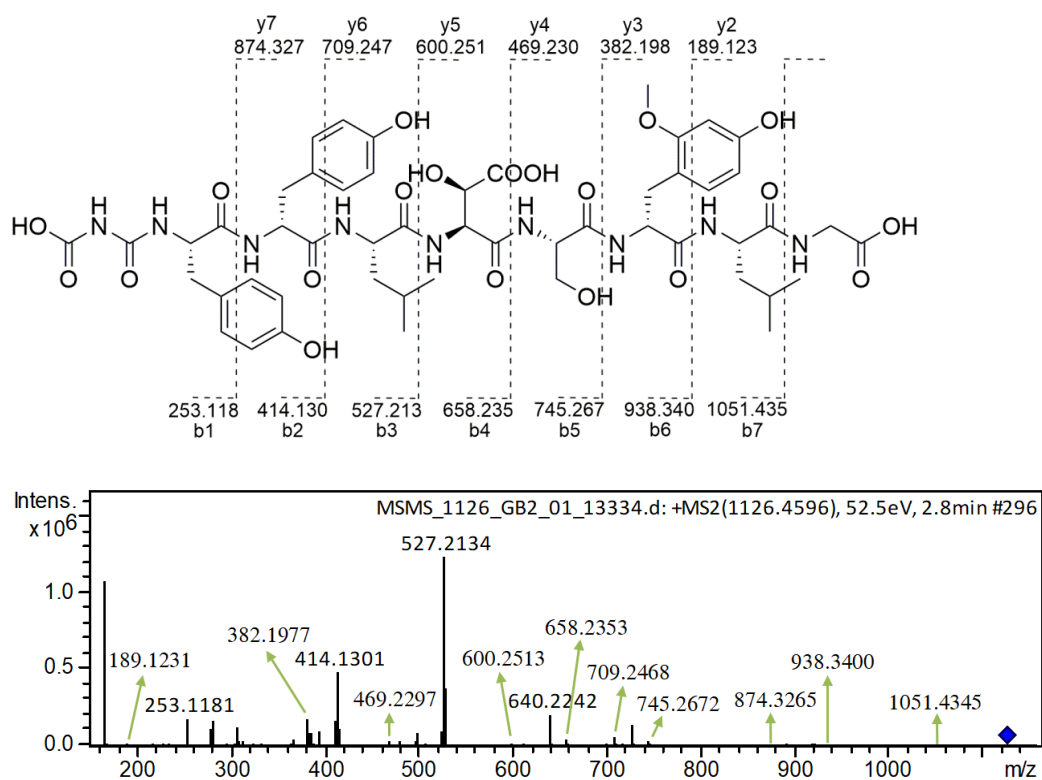
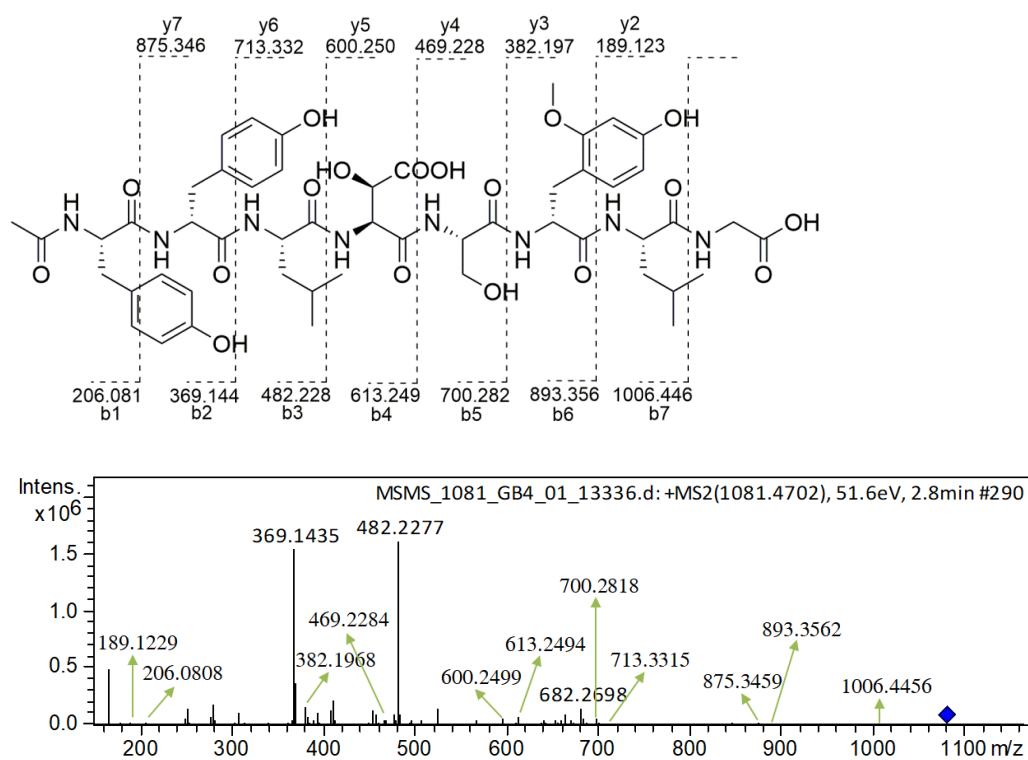
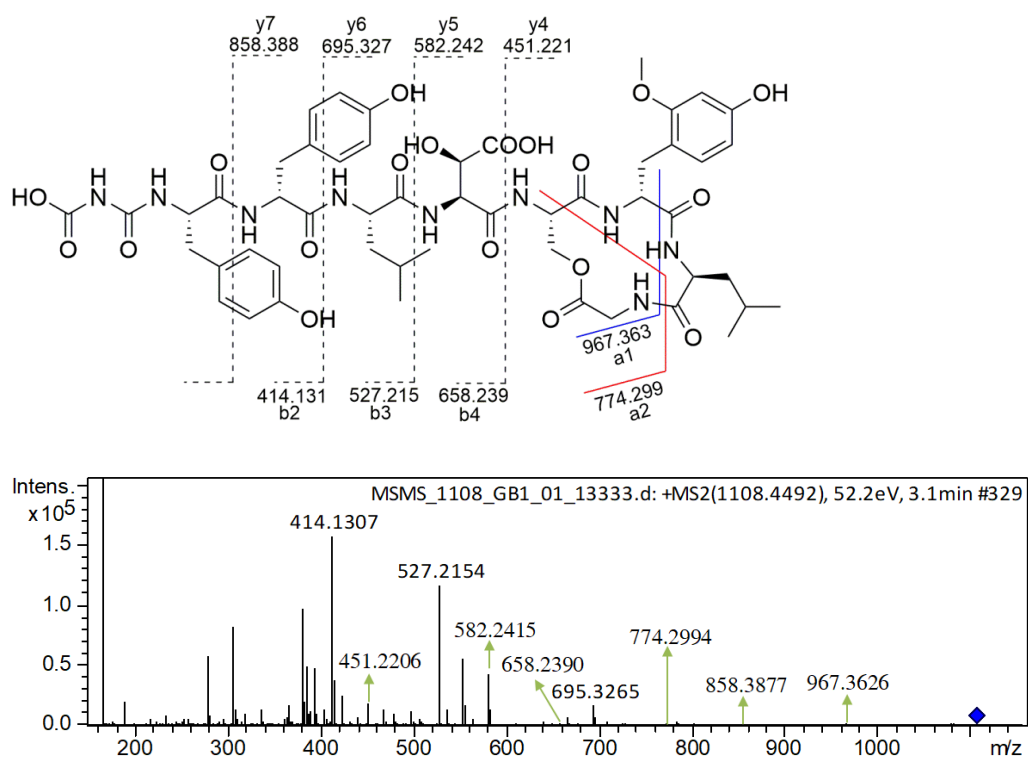


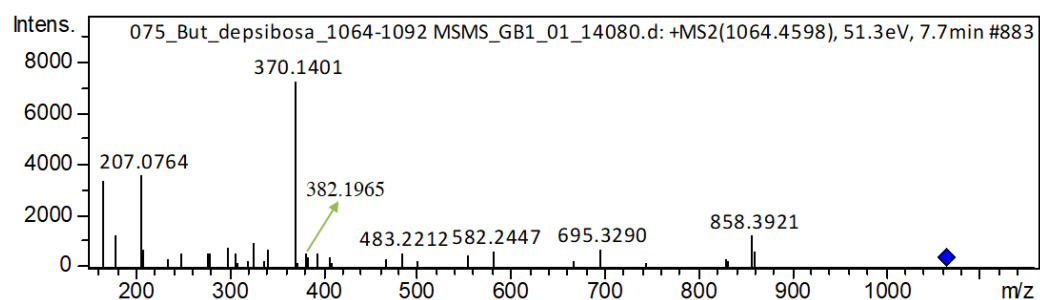
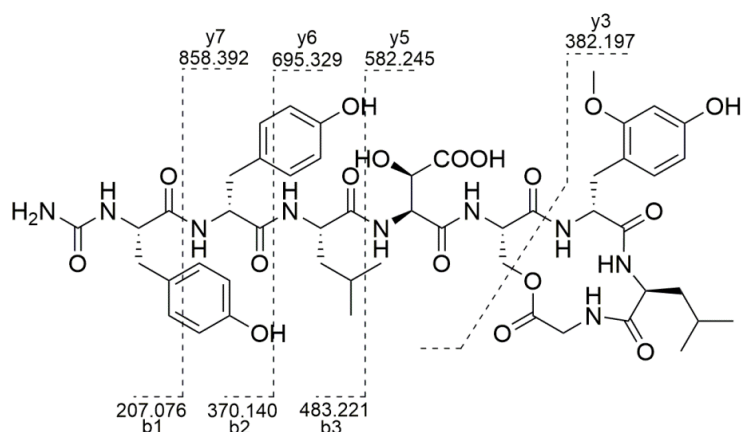
Figure S39: MS/MS analysis of bosamycin C.



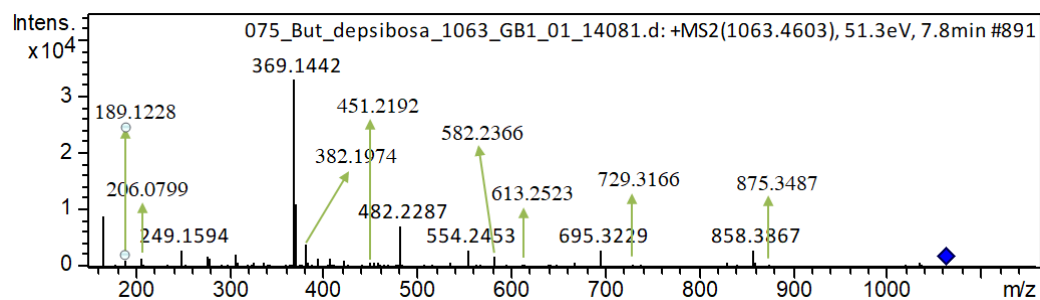
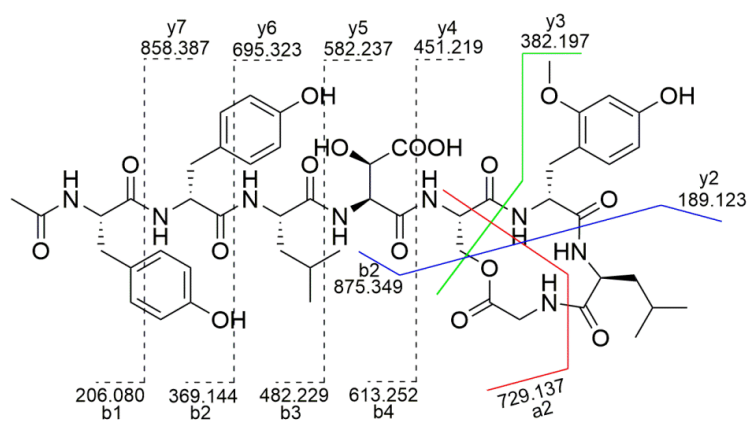
**Figure S40:** MS/MS analysis of bosamycin D.



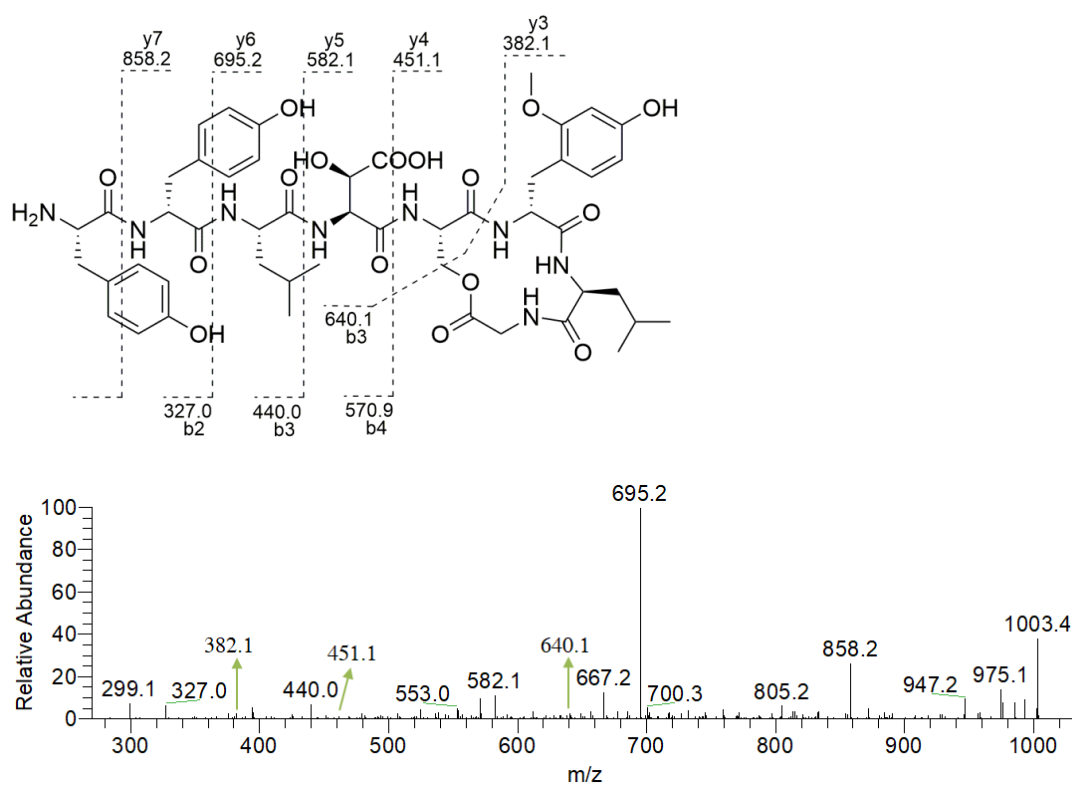
**Figure S41:** MS/MS analysis of depsibosamycin C.



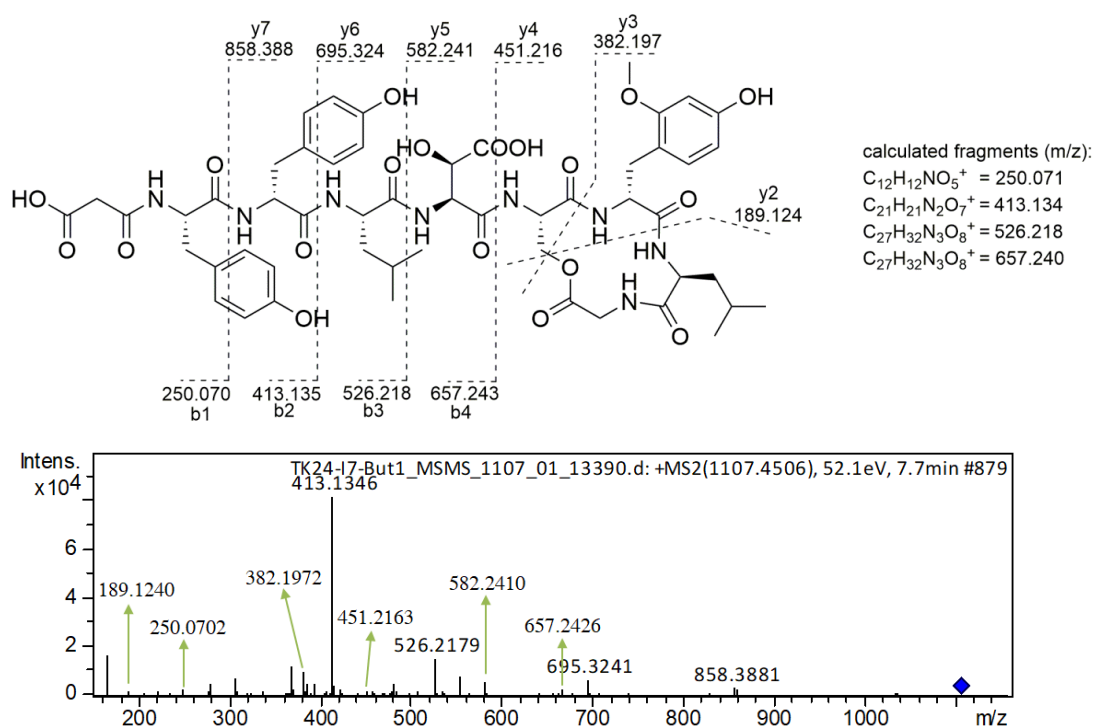
**Figure S42:** MS/MS of depsibosamycin B.



**Figure S43:** MS/MS analysis of depsibosamycin D.



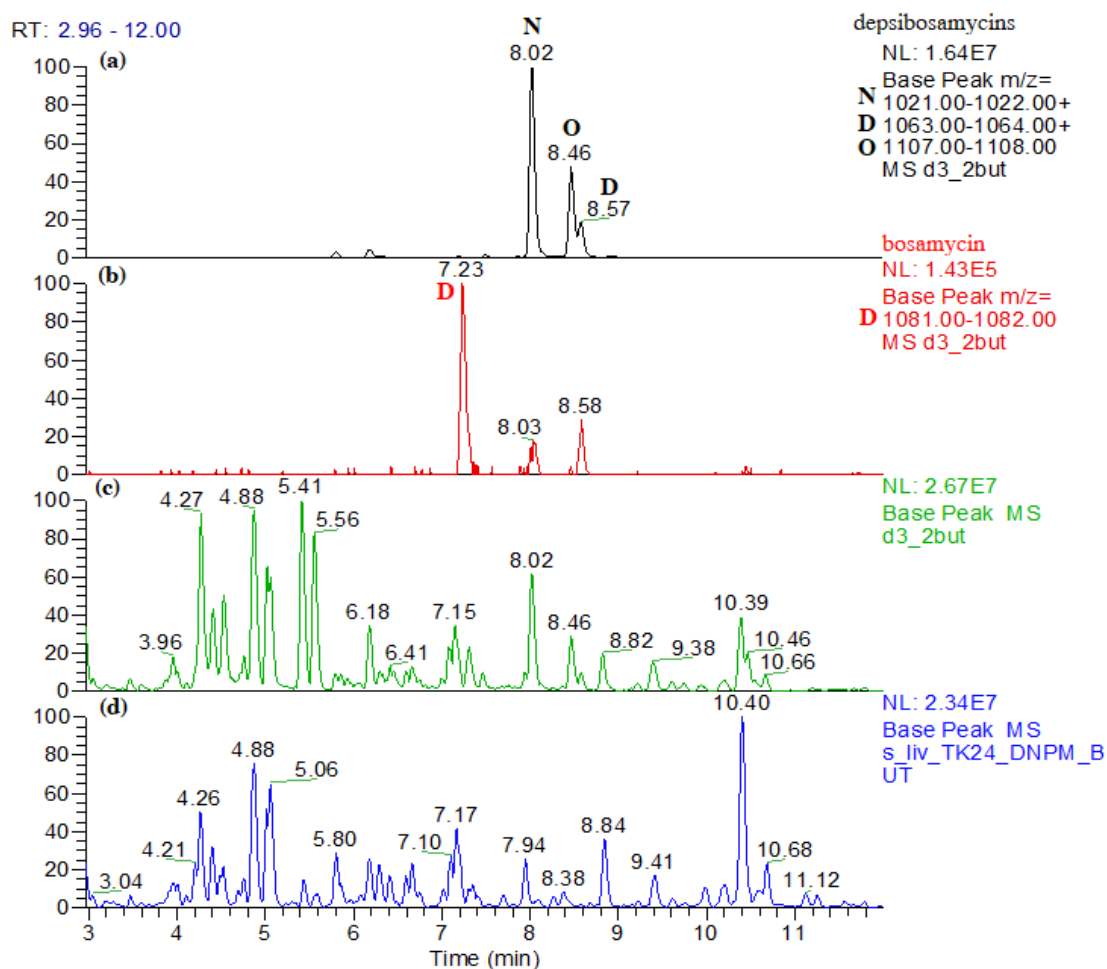
**Figure S44:** MS/MS of depsibosamycin N, produced by *S. lividans* I7.



**Figure S45:** MS/MS analysis and the calculated masses of the expected fragments of depsibosamycin O produced by *S. lividans* I7.

Score	Expect	Method	Identities	Positives	Gaps
484 bits(1247)	5e-154	Compositional matrix adjust.	235/239(98%)	236/239(98%)	0/239(0%)
Query 1		PIFCVHPAGGHSWGYHYLKNHLPFGHPLYGLQARTLLQGAEPPATLAAMAADYVEQLRSV			60
		PIFCVHPAGGHSWGYHYLKNHLPFGHPLYGLQARTLLQGAEPPATLAAMAADYVEQLRSV			
Sbjct 3676		PIFCVHPAGGHSWGYHYLKNHLPFGHPLYGLQARTLLQGAEPPATLAAMAADYVEQLRSV			3735
Query 61		QPHGPYHIVGWSFGGLVAFEMATRLQAQGEVALLALLDAFPSAATEEGDEAVPADEDSM			120
		QPHGPYHIVGWSFGGLVAFEMATRLQAQGE VALLALLDAFPSAA EEGDEAVPADEDSM			
Sbjct 3736		QPHGPYHIVGWSFGGLVAFEMATRLQAQGEVALLALLDAFPSAAAEEGDEAVPADEDSM			3795
Query 121		LRMIAANAGYDPKDITTVGADEPLTPSALSEFFQQVGAVMANLTADDLRTFAHTLRHNAH			180
		LRMIAANAGYDPKDITTVGADEPLTPSALSEFFQQVGAVMANLTADDLRTFAHTLRHNAH			
Sbjct 3796		LRMIAANAGYDPKDITTVGADEPLTPSALSEFFQQVGAVMANLTADDLRTFAHTLRHNAH			3855
Query 181		IGAAFTPGVFRGDVLLLTAAALGRDGAETAPEKGRHAWTAHVTGRIDEHRVPCRHHDDLML			239
		IGAAFTPGVFRGDVLLLTAAALGRDGAETAPEKGR AWTAHVTGRIDEHR+PCRHHDDLML			
Sbjct 3856		IGAAFTPGVFRGDVLLLTAAALGRDGAETAPEKGRRAWTAHVTGRIDEHRLPCRHHDDLML			3914

**Figure S46:** TE domain alignment of the depsibosamycin cluster (Query) and the bosamycin cluster (Sbjct).



**Figure S47:** LC-MS chromatograms of *S. lividans* I7 with the dbm gene cluster showing (a) the extracted masses of depsibosamycin D, N and O (b) the extracted mass of bosamycin D, (c) the base peak chromatograms (bpc) of *S. lividans* I7 and (d) the bpc of the empty host *S. lividans* TK24.

## References

1. Rückert, C.; Albersmeier, A.; Busche, T.; Jaenicke, S.; Winkler, A.; Friðjónsson, Ó.H.; Hreggviðsson, G.Ó.; Lambert, C.; Badcock, D.; Bernaerts, K.; et al. Complete genome sequence of *Streptomyces lividans* TK24. *J. Biotechnol.* **2015**, *199*, 21-22, 10.1016/j.jbiotec.2015.02.004.
2. Flett, F.; Mersinias, V.; Smith, C.P. High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. *FEMS Microbiol. Lett.* **1997**, *155* (2), 223-229, 10.1111/j.1574-6968.1997.tb13882.x.
3. Xu, Z.F.; Bo, S.T.; Wang, M.J.; Shi, J.; Jiao, R.H.; Sun, Y.; Xu, Q.; Tan, R.X.; Ge, H.M. Discovery and biosynthesis of bosamycins from *Streptomyces* sp. 120454. *Chemical Science* **2020**, *11* (34), 9237-9245, 10.1039/D0NP90037K.

II

**Cyclofaulknamycin with the Rare Amino Acid D-Capreomycidine  
Isolated from a Well-Characterized *Streptomyces albus* Strain**

Liliya Horbal, Marc Stierhof, Anja Paluszczak, Nikolas Eckert, Josef Zapp and Andriy Luzhetskyy

*Microorganisms* **2021**, *9*(8), 1609

DOI: 10.3390/microorganisms9081609

Published online 28<sup>nd</sup> July 2021



## Article

# Cyclofaulknamycin with the Rare Amino Acid D-capreomycin Isolated from a Well-Characterized *Streptomyces albus* Strain

Liliya Horbal <sup>1,†</sup>, Marc Stierhof <sup>1,†</sup>, Anja Paluszczak <sup>1</sup>, Nikolas Eckert <sup>2</sup>, Josef Zapp <sup>3</sup> and Andriy Luzhetskyy <sup>1,3,\*</sup>

<sup>1</sup> Department of Pharmaceutical Biotechnology, Saarland University, 66123 Saarbrücken, Germany; lihorbal@gmail.com (L.H.); marc.stierhof@uni-saarland.de (M.S.); anja.paluszczak@uni-saarland.de (A.P.)

<sup>2</sup> AMEG Department, Helmholtz Institute for Pharmaceutical Research Saarland, 66123 Saarbrücken, Germany; nikolas.eckert@uni-saarland.de

<sup>3</sup> Department of Pharmaceutical Biology, Saarland University, 66123 Saarbrücken, Germany; j.zapp@mx.uni-saarland.de

\* Correspondence: a.luzhetskyy@mx.uni-saarland.de; Tel.: +49-681-302-70200

† Both authors contributed equally.



**Citation:** Horbal, L.; Stierhof, M.; Paluszczak, A.; Eckert, N.; Zapp, J.; Luzhetskyy, A. Cyclofaulknamycin with the Rare Amino Acid D-capreomycin Isolated from a Well-Characterized *Streptomyces albus* Strain. *Microorganisms* **2021**, *9*, 1609. <https://doi.org/10.3390/microorganisms9081609>

Academic Editors:  
Stéphane Cociancich and  
Valérie Leclère

Received: 14 June 2021  
Accepted: 22 July 2021  
Published: 28 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Targeted genome mining is an efficient method of biosynthetic gene cluster prioritization within constantly growing genome databases. Using two capreomycin biosynthesis genes, alpha-ketoglutarate-dependent arginine beta-hydroxylase and pyridoxal-phosphate-dependent aminotransferase, we identified two types of clusters: one type containing both genes involved in the biosynthesis of the abovementioned moiety, and other clusters including only arginine hydroxylase. Detailed analysis of one of the clusters, the *flk* cluster from *Streptomyces albus*, led to the identification of a cyclic peptide that contains a rare D-capreomycin moiety for the first time. The absence of the pyridoxal-phosphate-dependent aminotransferase gene in the *flk* cluster is compensated by the *XNR\_1347* gene in the *S. albus* genome, whose product is responsible for biosynthesis of the abovementioned nonproteinogenic amino acid. Herein, we report the structure of cyclofaulknamycin and the characteristics of its biosynthetic gene cluster, biosynthesis and bioactivity profile.

**Keywords:** D-capreomycin; cyclofaulknamycin; cyclopeptide; *Streptomyces*

## 1. Introduction

Cyclic peptides possess a wide variety of biological activities with high application potentials as antibiotics (e.g., vancomycin and teicoplanin) and phytopathogenic agents (e.g., cyclothiazomycin, neopeptin, and kutznerides) [1–3]. Their biosynthesis is accomplished via ribosomal [4] or nonribosomal biosynthetic machinery [5]. Peptides of both origins show extremely high structural diversity resulting from the presence of a large variety of nonproteinogenic amino acids that are either used for the biosynthesis of the peptides or generated as a result of post-translational modifications [4,6]. A growing group of peptide compounds with important biological activities contains a rare structural class of amino acids, derivatives of arginine with unique five- or six-membered cyclic guanidine moieties (enduracididine and capreomycin, respectively) [7,8]. For instance, mannopeptimycin and its semisynthetic analogs represent a new class of lipoglycopeptides with exceptional activity against MRSA, VRE, and penicillin-resistant *Streptococcus pneumoniae* [9,10]. Enduracidins A and B are depsipeptides containing two enduracididine moieties that are active against Gram-positive bacteria, including resistant strains and *Mycobacterium* species [11]. Teixobactin is a peptide-like compound with potent activity against different multidrug-resistant bacteria, including *M. tuberculosis* and *Clostridium difficile* [12], and does not induce the development of resistance [13]. Viomycin, which is an essential component in the drug cocktail currently used to treat *M. tuberculosis* infections, also contains the capreomycin amino acid moiety [14]. A broad spectrum

of evidence supports the notion that the presence of arginine residues in linear or cyclic antimicrobial peptides enhances their activity [15]. For example, the relevance of arginine to the activity of small peptides was demonstrated by less potent lysine-containing analogs of an 11-residue lactoferricin segment [16]. In addition to peptide compounds, other classes of natural products (NPs) containing arginine residues, or derivatives thereof, possess promising activities [8,17]. Recently, it was shown that the incorporation of arginine residue into the chemical structure of vancomycin confers it with cell-killing activity against carbapenem-resistant *E. coli*, which are Gram-negative bacteria [18]. Thus, NPs containing rare nonproteinogenic amino acid derivatives, including arginine, and their coding gene clusters, are of particular interest because they can serve as chemical scaffolds for biochemical and semisynthetic modifications to expand their biochemical diversity and as a source of genes for combinatorial biosynthesis. These features make them an important source of pharmacologically active and industrially relevant secondary metabolites.

In this article, we describe the utilization of two genes encoding enzymes responsible for the biosynthesis of the capreomycin moiety, alpha-ketoglutarate-dependent arginine beta-hydroxylase, and pyridoxal-phosphate-dependent aminotransferase, for the targeted genome mining of new peptide compounds containing the abovementioned amino acid. The genome analysis of a well-characterized *S. albus* strain revealed the presence of an arginine beta-hydroxylase-encoding gene homolog in one of the NRPS gene clusters. To identify the product of this cluster, the metabolomes of two previously developed *S. albus* mutant strains [19] with and without the abovementioned cluster were compared. The analysis and comparison of the secondary metabolite profiles of the two strains allowed us to identify compounds with molecular masses of 731.4 and 749.4 Da, which correspond to the molecular weights of the predicted but never-identified cyclic faulknamicin compound and the recently described linear faulknamicin, respectively [20]. This is the first identified cyclic peptide compound with the rare D-capreomycin amino acid. Furthermore, a gene encoding pyridoxal-phosphate-dependent aminotransferase located outside of the identified cluster is shown to be involved in the biosynthesis of the D-capreomycin moiety. Herein, we report the structure of this compound and describe its biosynthetic gene cluster, biosynthesis and bioactivity profile.

## 2. Materials and Methods

### 2.1. Bacterial Strains and Culture Conditions

The bacterial strains used in this study are listed in Table 1. The *E. coli* strains were grown in Luria–Bertani (LB) broth medium. When required, antibiotics (Carl Roth, Karlsruhe, Germany; Sigma-Aldrich, St. Louis, MO, USA) were added to the cultures at the following concentrations: 75  $\mu\text{g mL}^{-1}$  ampicillin, 50  $\mu\text{g mL}^{-1}$  kanamycin, 50 or 120  $\mu\text{g mL}^{-1}$  hygromycin, and 50  $\mu\text{g mL}^{-1}$  apramycin (Carl Roth, Karlsruhe, Germany; Sigma-Aldrich, St. Louis, MO, USA). *E. coli* GB05-red [21] was employed in Red/ET recombineering experiments [22].

For conjugation, the *Streptomyces albus* J1074, del9 and del10 [19] strains were grown on oatmeal or mannitol soy (MS) agar [23] for sporulation.

### 2.2. Recombinant DNA Techniques

Chromosomal DNA from *Streptomyces* strains and plasmid DNA from *E. coli* were isolated using standard protocols [23,24]. Restriction enzymes and molecular biology reagents were used as per the manufacturer's protocol (NEB, Ipswich, UK; Thermo Fisher Scientific, Waltham, MA, USA).

**Table 1.** Strains and plasmids used in this study.

Bacterial Strains and Plasmids	Description	Source or Reference
<i>S. albus</i> J1074	Isoleucine and valine auxotrophic derivative of <i>S. albus</i> G (DSM 40313) lacking Sall-restriction activity	Salas J., Oviedo, Spain
<i>S. albus</i> del9	<i>S. albus</i> G (DSM 40313) lacking 9 secondary metabolite gene clusters	This work
<i>S. albus</i> del10	<i>S. albus</i> G (DSM 40313) lacking 10 secondary metabolite gene clusters	This work
<i>S. albus</i> del9 delXNR_1347	<i>S. albus</i> del9 strain carrying the deletion of the XNR_1347 gene	This work
<i>E. coli</i> ET12567 (pUB307)	conjugative transfer of DNA	Kieser et al., 2000
<i>E. coli</i> GB05-red	Derivative of GB2005 containing integration of PBAD-ETgA operon	Zhang et al., 2000
pSMART	Chloramphenicol-resistant, general cloning BAC vector	Thermo Fisher Scientific
2K9/2	pSMART derivative containing a piece of the <i>S. albus</i> J1074 chromosome	Myronovskyi et al., 2018
2K9/2delXNR_1347	2K9/2 derivative containing deleted XNR_1347 gene	This work

### 2.3. Construction of the delXNR\_1347 BAC Vector

In the Red/ET recombination experiment, a linear DNA fragment containing an apramycin resistance marker and an origin of transfer (*oriT*) flanked by suitable homology arms was generated by PCR using the XNR\_1347RedFor and XNR\_1347RedRev primer pair (Table 2). PCR was carried out with Phusion DNA polymerase (Thermo Fisher Scientific, Waltham, MA, USA), according to the manufacturer's protocol. The PCR product was concentrated by ethanol precipitation prior to further use. In general, 300 µL of an overnight culture of *E. coli* GB05-red (Table 1) cells harboring the parent to be modified, 2K9/2 BAC, was inoculated into 15 mL of LB, and the culture was incubated on a shaker at 37 °C and 200 rpm for 2 h. Thereafter, 400 µL of 10% L-rhamnose was added to the culture to induce the expression of the recombinases. Cultivation was then continued for 45 min. The cells were subsequently harvested by centrifugation, washed twice with ice-cold distilled H<sub>2</sub>O, and resuspended in 600 µL of 10% ice-cold glycerol. The PCR product was mixed with the electrocompetent cells, which were then transferred to an ice-cold electroporation cuvette (1 mm). The mixture was subsequently electroporated at 1800 V (Eppendorf electroporator), followed by the addition of 750 µL LB. The cells were cultivated at 37 °C for 90 min before the culture was transferred to LB agar plates with apramycin. The plates were incubated at 37 °C overnight. Correct transformants were verified by the restriction analysis and sequencing of the isolated BAC DNA using the XNR\_1347F and XNR\_1347R primers (Table 2).

**Table 2.** Primers used in this work.

Primers	Sequence 5'-3'	Purpose
XNR_1347RedFor	TGGAGCGGGAGCGGAGCAGGACGAGCGGCGCGA GCGGGCGTGAGCCGCCGGGTGCCCGCGCCCCG GTCATATCCATCCTTTTTCGCACGATATAC	XNR_1347 gene deletion
XNR_1347RedRev	CGACGCCTGGCTGCCGGGGCTCCCGCGCGGGTCCG TAGAATCGGCCCCACCATGGCCTACCTCGACCACC AGATTACGCGCAGAAAAAAGGATCTC	
XNR_1347F XNR_1347R	AAGAAGCAGCTGGAGCGGGAG ACCATGGCCTACCTCGACCAC	XNR_1347 gene deletion check

#### 2.4. Construction of the *S. albus* delXNR\_1347 Mutant

The 2K9/2delXNR\_1347 (Table 1) gene disruption BAC was transferred from *E. coli* ET12567/pUB307 cells into *S. albus* del9 cells by means of conjugation [23]. Transconjugants were selected for resistance to apramycin ( $50 \mu\text{g mL}^{-1}$ ) and glucuronidase activity. The abovementioned cosmid did not contain the origin of replication for *Streptomyces*, all the obtained exconjugants were the result of a first crossover event. For the generation of the *S. albus*  $\Delta$ XNR\_1347 mutant, single-crossover apramycin and blue mutants were screened for the loss of glucuronidase activity (blue pigmentation) as the result of a double-crossover event. The replacement of the XNR\_1347 gene was confirmed by PCR using the XNR\_1347F and XNR\_1347R primer pair (Table 2).

#### 2.5. Metabolite Extraction and Analysis

*S. albus* strains were cultivated in 25 mL TSB medium for 48 h at 28 °C. The main cultures containing 50 mL of SG were inoculated with 2 mL of preculture. After 5 days of cultivation at 28 °C, the secreted metabolites were extracted with ethyl acetate and butanol, followed by solvent evaporation. The dry extracts were dissolved in 0.5 mL methanol, and 1  $\mu\text{L}$  of the dissolved sample was separated in a Dionex Ultimate 3000 RSLC system using a BEH C18,  $100 \times 2.1 \text{ mm}$ ,  $1.7 \mu\text{m}$  dp column (Waters Corporation, Milford, MA, USA). The separation of a 1  $\mu\text{L}$  sample was achieved via a linear gradient with a mobile phase of water/acetonitrile, each containing 0.1% formic acid, with a gradient from 5–95% acetonitrile applied over 9 min at a flow rate of 0.6 mL/min and 45 °C. High-resolution mass spectrometry was performed on an Accela UPLC system (Thermo Fisher Scientific, Waltham, MA, USA) coupled to a LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) operating in positive or negative ionization modes. Data were collected and analyzed with Thermo Xcalibur software, version 3.0 (Thermo Scientific, Waltham, MA, USA). The monoisotopic mass was searched in a natural product database.

#### 2.6. Isolation and Purification of Cyclic and Linear Iso-Faulknamycin

For faulknamycin production, 2 mL of a 2-day-old preculture grown in 50 mL of TSB media (Sigma-Aldrich, St. Louis, MO, USA) was inoculated into 100 mL of SG media, and the culture was grown for 5 days at 28 °C with agitation at 200 rpm. The methanol extract obtained after 10 L butanol extraction was used for the purification of the compounds in an Isolera™ One flash purification system (Biotage, Uppsala, Sweden) equipped with a CHROMABOND Flash RS 330 C18 ec column (Macherey-Nagel, Dueren, Germany) using a gradient of 5–95% aqueous methanol for 3 CV at a flow rate of 100 mL/min, with UV detection at 210 and 280 nm. The fractions containing both faulknamycin compounds were collected, pooled together, dried and dissolved in methanol. Thereafter, size-exclusion chromatography was performed on an LH20 Sephadex column (Sigma-Aldrich, Louis, MO, USA). Methanol was used as a solvent for elution. Fractions were collected every 12 min at a flow rate of 0.6 mL/min. The fractions containing faulknamycins were pooled together, evaporated, and dissolved in methanol. Then, the prepurified methanol extract was further separated by preparative reversed-phase (RP) HPLC (Waters 2545 Binary Gradient module, Waters, Milford, MA, USA) using a Nucleodur C18 HTec column ( $5 \mu\text{m}$ ,  $21 \times 250 \text{ mm}$ , Macherey Nagel, Dueren Germany) with a linear gradient of 0.1% formic acid solution in acetonitrile against 0.1% formic acid solution in water, applied at 5% to 95% over 33 min at a flow rate of 5 mL/min. UV spectra were recorded with a PAD detector (Photodiode Array Detector, Waters, Milford, MA, USA). The fractions containing the linear and cyclic faulknamycin compounds were dried and dissolved in methanol. The final purification step was performed in an RP HPLC system (Agilent Infinity 1200 series HPLC system) equipped with a Synergi™ 4  $\mu\text{m}$  Fusion-RP 80 Å  $250 \times 10$  (Phenomenex, Torrance, CA, USA) column using a linear gradient of [A]  $\text{H}_2\text{O}$  + 0.1% formic acid/[B] acetonitrile + 0.1% formic acid, applied at 5% to 95% [B] over 25 min at a flow rate of 3 mL/min, with a column oven temperature of 45 °C and UV detection at 210 nm, followed by fraction control via

HPLC-MS. Fractions were pooled to obtain the two pure faulknamycin isolates (1.2 mg of linear faulknamycin and 1.4 mg of cyclic faulknamycin).

### 2.7. Nuclear Molecular Resonance Spectroscopy (NMR)

NMR spectra were acquired on a Bruker Avance III, Ascent 700 MHz spectrometer (298 K) equipped with a 5 mm TCI cryoprobe (Bruker, BioSpin GmbH, Rheinstetten, Germany). The chemical shifts ( $\delta$ ) were reported in parts per million (ppm) relative to TMS. As a solvent, deuterated DMSO- $d_6$  ( $\delta$ H 2.50 ppm.,  $\delta$ C 39.51 ppm.) from Deutero (Kastellaun, Germany) was used. Edited-HSQC, HMBC,  $^1$ H- $^1$ H COSY and ROESY spectra were recorded using the standard pulse programs from Bruker TOPSPIN v.3.6 software.

### 2.8. Marfey's Method

Iso-faulknamycin, chymostatin and capreomycin were hydrolyzed in 100  $\mu$ L 6 N HCl at 110 °C for 45 min. While cooling down, the samples were dried for 15 min under nitrogen and dissolved in 110 mL water, after which 50  $\mu$ L of each sample was transferred to a 1.5 mL Eppendorf tube. Next, 20  $\mu$ L of 1 N  $\text{NaHCO}_3$  and 20  $\mu$ L of 1% L-FDLA or D-FDLA in acetone were added to the hydrolysates. The amino acid standards were prepared in the same way using only L-FDLA. The reaction mixtures were incubated at 40 °C for 90 min at 750 rpm and subsequently quenched with 2 N HCl to stop the reaction. The samples were diluted with 300  $\mu$ L ACN, and 1  $\mu$ L of each sample was analyzed in a maXis high-resolution LC-QTOF system using aqueous ACN with 0.1 vol% FA and an adjusted gradient of 5–10 vol% for 2 min, 10–25 vol% for 13 min, 25–50 vol% for 7 min and 50–95 vol% for 2 min. Sample detection was carried out at 340 nm.

### 2.9. Genome Mining and Bioinformatic Analysis

Genome screening was performed using the BLAST online tool ([blast.ncbi.nlm.nih.gov/Blast.cgi](http://blast.ncbi.nlm.nih.gov/Blast.cgi) (accessed on 15 June 2021)). The identified genomes were downloaded from the NCBI genome database ([www.ncbi.nlm.nih.gov/genome/](http://www.ncbi.nlm.nih.gov/genome/) (accessed on 15 June 2021)). Gene cluster analysis was performed using the antiSMASH online tool ([antismash.secondarymetabolites.org/#!/start](http://antismash.secondarymetabolites.org/#!/start) (accessed on 15 June 2021)) [25]. Analysis of genetic data was performed using Geneious software, version 11.0.3 (Biomatters Ltd., Auckland, New Zealand).

### 2.10. Antimicrobial Susceptibility Test

Minimum inhibitory concentrations (MICs) were determined according to standard procedures. Single colonies of the bacterial strains were suspended in cation-adjusted Müller-Hinton broth to achieve a final inoculum of approximately  $10^4$  CFU  $\text{mL}^{-1}$ . Serial dilutions of cyclofaulknamycin (0.03 to 64  $\mu\text{g}\cdot\text{mL}^{-1}$ ) were prepared in sterile 96-well plates, and the bacterial suspension was added. Growth inhibition was assessed after overnight incubation (16–18 h) at 30–37 °C. A panel consisting of the following microbial strains was tested: *Acinetobacter baumannii* DSM-30008, *E. coli* JW0451-2 ( $\Delta$ acrB), *S. aureus* Newman, *E. coli* BW25113 (wt), *M. smegmatis* mc2155, *P. aeruginosa* PA14, *B. subtilis* DSM-10, *Citrobacter freundii* DSM-30039, *Mucor hiemalis* DSM-2656, *Candida albicans* DSM-1665, *Cryptococcus neoformans* DSM-11959, and *Pichia anomala* DSM-6766.

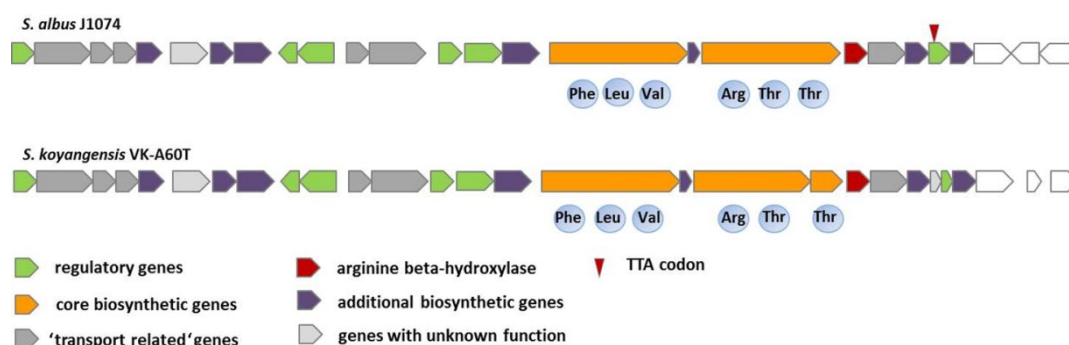
## 3. Results and Discussion

### 3.1. Mining of Actinobacterial Genomes for the Presence of Capreomycin Biosynthetic Genes

To identify the compounds containing the rare cyclic guanidino-amino acid capreomycin, two genes, *vioC* (encoding alpha-ketoglutarate-dependent arginine beta-hydroxylase) and *vioD* (encoding the PLP-dependent aminotransferase (PLP-aminotransferase) from the viomycin biosynthetic gene cluster [26], were used for the targeted genome analysis and cluster searches in the NCBI database. As a result of the in silico mining performed, two types of biosynthetic gene clusters (BGCs) were identified: one that contained both homologs, and other clusters that included only one of the two genes (the *vioC* homolog).



We focused our attention on the second group of clusters, which have not previously been characterized. Until recently, the products of both gene orthologs (VioC and VioD) were considered to be required for capreomycin biosynthesis [26]. During the preparation of this manuscript, an article describing the new hypothetical pathway for capreomycin biosynthesis was published [20]. From the identified clusters, two *flk* BGCs that had a very similar structure and NRPS domain organization attracted our attention; however, they were identified in different microorganisms, *S. albus* J1074 and *Streptomyces koyangensis* VK-A60T (Figure 1).



**Figure 1.** Structure of the *flk* clusters in the *S. albus* and *S. koyangensis* strains.

Furthermore, the recently published faulknamicin gene cluster (fau BGC) [20] appeared to be very similar to the *flk* BGCs. Based on adenylation domain analysis, the two clusters were predicted to encode hexapeptides composed of the same amino acids (Figure 1). The main difference between the two identified BGCs was the presence of two and three NRPS genes in the *S. albus* J1074 and *S. koyangensis* VK-A60T clusters, respectively (Figure 1). PLP-dependent aminotransferase genes were not present in these clusters or in the recently identified faulknamicin gene cluster. Tryon et al. suggested that the capreomycin moiety might be synthesized via the epimerization domain of the FauG NRPS. However, at least one homolog of the PLP-aminotransferase gene was identified within the *S. albus* J1074 (XNR\_1347) and *S. koyangensis* VK-A60T genomes after careful manual analysis. Taking into account this information, it was assumed that the XNR\_1347 product might be involved in the biosynthesis of the capreomycin moiety. The chemical products of the *flk* clusters were not known; however, they were assumed to be identical to faulknamicin based on the in silico prediction of the NRPS domain organization and the genes present in the clusters (Figure 1).

### 3.2. Analysis of an *flk* Gene Cluster in the *S. albus* Genome

The product of biosynthetic gene cluster 5 (*flk* cluster) from the well-studied *S. albus* strain has never been described before. Based on in silico analysis and comparison with a similar cluster from the *S. koyangensis* VK-A60T strain (Table S1, Figure 1), it was assumed that the *flk* cluster contained 23 genes encoding proteins that are sufficient for final peptide biosynthesis (Figure 1, Table S1). Among these genes, *flkP* and *flkS* encode two large multi-functional NRPSs (Figure 1, Table S1) that are composed of six biosynthetic modules and are therefore assumed to encode hexapeptides. Detailed analysis of the NRPSs, with the help of NRPS Predictor 2 ([abi-services.informatik.informati/tuebingen.de/nrps2/Controller?cmd=SubmitJob](http://abi-services.informatik.informati/tuebingen.de/nrps2/Controller?cmd=SubmitJob) (accessed on 15 June 2021)), allowed us to identify the adenylation domain (A-domain) specificity of the individual modules (Table 3). Based on the substrate specificity-conferring codes, it was assumed that domain A1 was responsible for the incorporation of phenylalanine, A3 for valine, and A5 and A6 for threonine. The remaining two domains, A2 and A4, are characterized by substrate promiscuity; however, they were predicted to most likely be responsible for the incorporation of leucine and arginine or its derivatives, respectively (Table 3). Sequence analysis of the FlkP NRPS revealed that it contained the following domains: three A domains, three peptidyl carrier protein (PCP) domains, one

epimerization domain and two condensation domains. These domains form three modules, including one loading module and two elongation modules, with the following domain organization: A<sub>Phe</sub>-PCP-C-A<sub>Leu</sub>-PCP-E-C-A<sub>Val</sub>-PCP. E represents an epimerization domain which is responsible for the conversion of L-amino acids into the D-form. The FlkP protein ends with a PCP domain; thus, its product should be offloaded to the next multidomain NRPS enzyme, which is FlkS. FlkS also contains three modules: C-A<sub>Arg</sub>-PCP-E-C-A<sub>Thr</sub>-PCP-C-A<sub>Thr</sub>-PCP-E. The presence of the two epimerization domains means that two D-amino acids are incorporated into the peptide during its biosynthesis. NRPS chain-terminating thioesterases are often found to exist as terminating domains in modular NRPSs [27]. However, no such domain was detected in the FlkP or FlkS proteins. Only the type II thioesterase homolog FlkX was identified in the *flk* cluster as a stand-alone gene, whose product is assumed to show proofreading activity. A homolog of the MppK alpha/beta hydrolase from the mannopeptimycin gene cluster, FlkO, was identified in the cluster and is assumed to be involved in the release of the hexapeptide from the biosynthetic machinery and its further cyclization. The *flkR* gene is located between two NRPS genes and encodes an MbtH homolog which is suggested to be involved in the promotion of NRPS production [28]. *FlkT* is a homolog of *vioC* [7] and encodes Fe(II)/alpha-ketoglutarate-dependent arginine beta-hydroxylase, which is involved in the biosynthesis of the capreomycin moiety [26]. No homolog of the PLP-dependent aminotransferase is present in the *flk* cluster or in the cluster from *S. koyangensis*. Using BLAST analysis, we identified an *XNR\_1347* gene encoding a PLP-dependent aminotransferase located outside of the *flk* gene cluster, which might be involved in capreomycin biosynthesis. Another possibility is that the biosynthesis of the capreomycin moiety is ongoing, as suggested by Tryon et al., 2020. *FlkG* and *flkH* encode methyl- and mannosyltransferases, respectively, and might be involved in post-translational modifications of the core peptide. Homologs of these genes are also present in the mannopeptimycin biosynthetic gene cluster [10].

**Table 3.** Analysis of the adenylation domain specificity.

Domain	NRPS Signature Position of the Residues										Predicted Substrate
	235	236	239	278	299	301	322	330	331	517	
A <sub>1</sub>	D	A	W	T	V	A	A	V	C	K	Phe
A <sub>2</sub>	D	G	M	L	V	G	A	V	V	K	Leu
A <sub>3</sub>	D	A	F	W	L	G	G	T	F	K	Val
A <sub>4</sub>	D	L	A	E	S	G	A	V	D	K	Arg, Orn, Lys
A <sub>5</sub>	D	F	W	S	V	G	M	V	H	K	Thr
A <sub>6</sub>	D	F	W	S	V	G	M	V	H	K	Thr

Two stand-alone putative FlkA and FlkW regulators, which belong to the DeoR/GlpR and ArsR families, respectively, are present in the cluster and might be involved in the regulation of biosynthesis. Furthermore, FlkW contains a TTA codon in the coding frame and is thus subject to *bldA* regulation [29]. Two pairs of histidine kinases and response regulators (FlkJ/FlkI and FlkN/FlkM) were detected in the cluster and might therefore also be involved in the regulation of gene expression.

Three genes, *flkB*, *flkV* and *flkD*, encode putative sugar transporters that might be involved in sugar uptake pathways. Transporters with similar putative functions (MppE and F) are also present in the mannopeptimycin gene cluster [10]. *FlkK* and *flkL* encode ABC transporters, and *flkU* encodes a major facilitator superfamily transporter and might thus be involved in antibiotic resistance.

### 3.3. Identification of the *flk* Cluster Product

Although *S. albus* has been subjected to a great deal of investigation as a model *Streptomyces* strain, the product of the *flk* cluster remains unknown. To identify this product, we took advantage of two constructed mutant strains, *S. albus* del9 and del10, published by Myronovsky et al., 2018. The *S. albus* del9 strain contains the *flk* gene cluster, whereas del10 is characterized by the deletion of this cluster [19]. Both strains were cultivated in SG production medium for 5 days, the culture broths were extracted with ethyl acetate and butanol separately, and the obtained extracts were analyzed using LC/MS (for details, see the Section 2). Comparison of the biosynthetic profiles of the two strains revealed the disappearance of the molecular ions  $[M + H]^+$  with masses of 750.4 and 732.4 Da from the metabolic profile of the del10 strain (Figure S21). Thus, the *flk* cluster was considered to be responsible for the biosynthesis of peptides with the above molecular masses. A search for compounds with such masses in the dictionary of natural products did not reveal any hits; therefore, the products were considered to be new. However, during the preparation of this manuscript, a linear compound with one of these masses (750.4 Da), referred to as faulknamycin, was published [20].

### 3.4. Purification and Structural Elucidation of Iso- and Cyclofaulknamycins

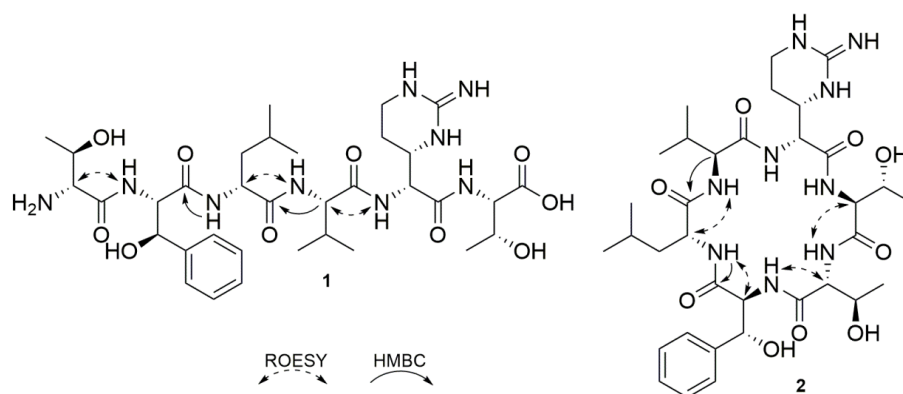
To gain insight into the structures of the produced compounds, the producer strain *S. albus* del9 was cultivated in 10 L of SG production medium, and the metabolites were extracted from the culture supernatant with butanol. The compounds with masses of 749.4 and 731.4 Da were successfully purified from the extract.

The structure of iso-faulknamycin (1) (Figure 2) was established by NMR spectroscopy and MS/MS fragmentation. The molecular formula was determined to be  $C_{34}H_{55}N_9O_{10}$  based on the identification of a mass of 749.40 Da and 12 degrees of unsaturation. In 2D NMR experiments (Figure S1, Table S2), including COSY (Figure S2), edited-HSQC (Figure S3) and HMBC (Figure S4) analyses, six amino acids assigned to leucine, valine, two threonines and two modified amino acids originating from phenylalanine and arginine were identified. The phenylalanine-derived unit showed HMBC correlations from its isochronic phenyl ring protons H-2 and H-6 (7.42, 2H) to a methine at  $\delta_C$  72.8 (Figure S4) in the side chain. Hence, it was assumed that a hydroxyl group in the  $\beta$  position resulted in the assignment of  $\beta$  phenylserine. Considering the four remaining unassigned nitrogens calculated from the molecular formula, it was concluded that one of the amino acids had to be related to arginine. Its  $\alpha$  CH proton ( $\delta_H$  4.70) revealed a COSY correlation to a neighboring methine at  $\delta_H$  3.73 (Figure S2). The remaining two degrees of unsaturation indicated an intramolecular ring closure of the guanidinium moiety and the  $\beta$ -position, leading to the assignment of capreomycinidine. The sample did not dissolve completely in DMSO- $d_6$ , which led to a lack of signal strength and prevented the full annotation of all carbon signals. However, by combining the observable HMBC and ROESY correlations, a major portion of the sequence could be assigned (Figure 2). The missing data were filled in by MS/MS fragmentation (Figures S17 and S18), revealing a final sequence of H<sub>2</sub>N Thr  $\beta$ -phenylserine-Leu-Val-capromycinidine-Thr-COOH.

Based on known hexapeptide structures, which are mostly cyclic, it was assumed the compound with a mass of 731.39 Da with a calculated formula of  $C_{34}H_{53}N_9O_9$  and 13 degrees of unsaturation represented the cyclic version of iso-faulknamycin 1. In comparison to 1, the observation of water loss and an additional degree of unsaturation indicated ring formation by intramolecular condensation. In 2D NMR experiments (Figures S6–S10, Table S3), it was possible to identify the same amino acids previously determined for iso-faulknamycin. Similar to linear hexapeptide 1, the cyclic version suffered from solubility issues, and the long-range HMBC and ROESY correlation had to be supported by MS/MS fragmentation (Figures S19 and S20). The cyclic formation of the new compound was confirmed by comparing the integral values of the NH signals and their ROESY correlations. The threonine NH<sub>2</sub> group of the linear peptide showed a shift of  $\delta_H$  7.88 and an integral value of two protons, while the same signal of the cyclic version showed a low-field shift to  $\delta_H$  8.52 with an integral of only one proton.



Furthermore, a ROESY correlation between this NH of the first threonine and the  $\alpha$ -CH proton of the second threonine was identified, both of which were still terminal in linear peptide 1 but directly adjacent in cyclic peptide 2 (Figure 2). In addition, the MS/MS fragmentation of the  $[M + 2H]^{2+}$  ion revealed intense a-ion and x-ion fragments, which could only be observed for the cyclic peptide (Table S6). This allowed us to confirm the assignment of cyclic version 2, which was named cyclofaulknamycin.



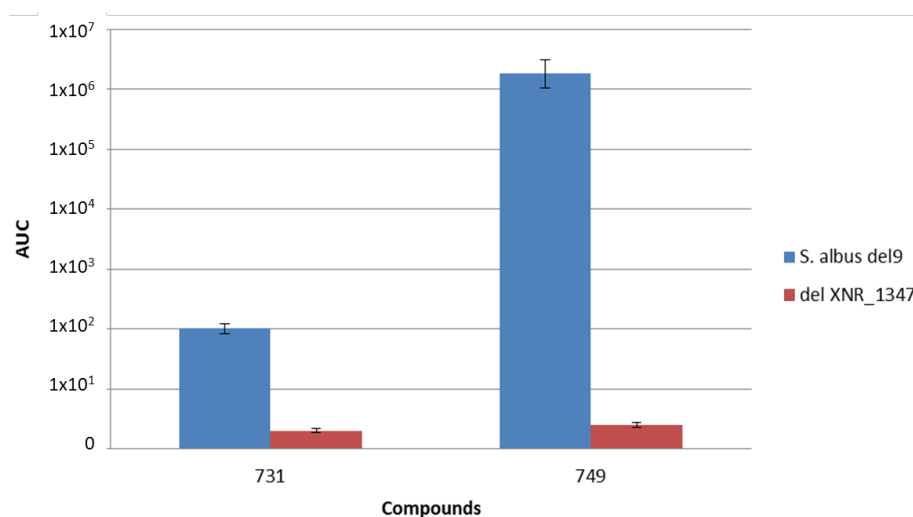
**Figure 2.** Structures of iso-faulknamicin (1) and cyclofaulknamycin (2) showing key HMBC and ROESY correlations.

Compounds 1 and 2 were named after a homologous structure that was recently published and designated faulknamycin [20]. The amino acid sequence and molecular formula of the abovementioned compound 1 appeared to be the same as those of faulknamycin. To determine whether 1 was identical to faulknamycin, the absolute stereochemistry of compound 1 was determined by Marfey's method (see Supplementary Materials) [30]. The amino acid mixture was compared with commercially acquired standards derivatized with L-FDLA (Figure S11). The configuration of capreomycin was determined as described previously by using standards derived from the hydrolysis of capreomycin and chymostatin and derivatization by L-FDLA and D-FDLA [20] (Figure S13). The amino acid  $\beta$ -phenylserine has been identified by Tryon et al. as the *D-threo*- isomer; therefore, *DL-threo*- $\beta$ -phenylserine was used as a standard to confirm the reported stereochemistry (Figures S14 and S15). This resulted in the identification of L-valine, D-leucine, L-threonine, *D-allo*-threonine, D-capreomycin and L-*threo*- $\beta$ -phenylserine as the constituent amino acids. *D-threo*- $\beta$ -phenylserine could not be identified as a result of the performed analysis; thus, compound 1 was concluded to be an isomer of faulknamycin and was named iso-faulknamicin (Figure S16). This finding is in line with the domain organization of the FlkP NRPS described above.

### 3.5. Deletion of the XNR1347 Gene in the *S. albus* del9 Genome

Only one PLP-dependent aminotransferase-encoding gene, *XNR\_1347*, has been found in the *S. albus* genome; therefore, it was assumed that its product might be involved in the biosynthesis of the capreomycin amino acid. To test this hypothesis, the deletion of this gene was performed in the del9 strain. For this purpose, the *S. albus* BAC library was used [19]. The coding sequence of the *XNR\_1347* gene in the 2K9/2 BAC was substituted with the apramycin resistance marker *aac(3)IV* and *oriT* via a Red/ET technology-based method (for details, see the Section 2), resulting in the construction of the 2K9/2del*XNR\_1347* deletion BAC (Table 2). The obtained construct was checked by restriction analysis as well as by partial sequencing. The cosmid was transferred into the *S. albus* del9 strain via conjugation. The obtained transconjugants were selected based on the resistance to apramycin and blue color, which denoted the presence of glucuronidase activity. Single-crossover mutants were screened for the loss of blue color resulting from a double-crossover event. As a result, the *S. albus*  $\Delta$ del9del*XNR\_1347* strain (Table 1) carrying the *XNR\_1347* gene deletion was created. The identity of the strain was confirmed by

PCR analysis (data not shown). LC-MS analysis of the extracts from the *S. albus*  $\Delta$ del9 and  $\Delta$ del9delXNR\_1347 strains (Figure 3) revealed 5- and 12-fold average decreases in the production of iso- and cyclofaulknamycins, respectively. Only traces of the compounds were detectable, which confirmed the hypothesis that the product of the XNR\_1347 gene is involved in the biosynthesis of the capreomycin moiety. Furthermore, this seems to be the main mechanism of capreomycin moiety formation in *S. albus*.



**Figure 3.** Faulknamycin production in the *S. albus* del9 and delXNR\_1347 strains: 731–cyclofaulknamycin; 749–iso-faulknamycin.

### 3.6. Proposed Biosynthesis of Cyclofaulknamycin

Tryon et al. (2020) never detected cyclofaulknamycin. In the current study, however, the cyclofaulknamycin product of the *flk* cluster was isolated, and its structure was elucidated with the aid of NMR and MS2 experiments (for details, see the section above). A detailed analysis of NRPS domain organization and cyclofaulknamycin chemical structure, as well as a comparison of the gene products in the cluster with previously described enzymes, allowed us to propose its biosynthesis process (Figure 4). The FlkP domain, in contrast to FauE and FauG from the *fau* cluster, contains a clear loading module which is responsible for the incorporation of the first amino acid, phenylalanine, or possibly its derivative, hydroxyphenylalanine, and two elongation modules that are responsible for leucine and valine incorporation. The epimerization domain is responsible for the conversion of L-leucine into its D-form. FlkP ends with the peptidyl carrier protein; therefore, its product has to be offloaded to the next NRPS multidomain enzyme, FlkS. The first module of FlkS is responsible for the incorporation of the capreomycin moiety, and the last two modules are responsible for the incorporation of two threonines. The biosynthesis of capreomycin starts with the hydroxylation of arginine with the help of the VioC homolog Fe(II)/alpha-ketoglutarate-dependent arginine beta-hydroxylase FlkT. Thereafter, the XNR\_1347 PLP-dependent aminotransferase catalyzes its conversion into capreomycin. The deletion of XNR\_1347 did not lead to a complete cessation of biosynthesis; therefore, it was assumed that some other aminotransferase homologs in the genome might be responsible for the biosynthesis of this amino acid or that it occurs spontaneously, as described by Tryon et al. The presence of epimerization domains in the first two modules is responsible for the conversion of capreomycin and threonine into their D-forms. No classical PKS/NRPS macrocyclizing thioesterase I is present in the cluster; thus, the serine hydrolase homolog FlkO is most likely responsible for the release and cyclization of faulknamycin.

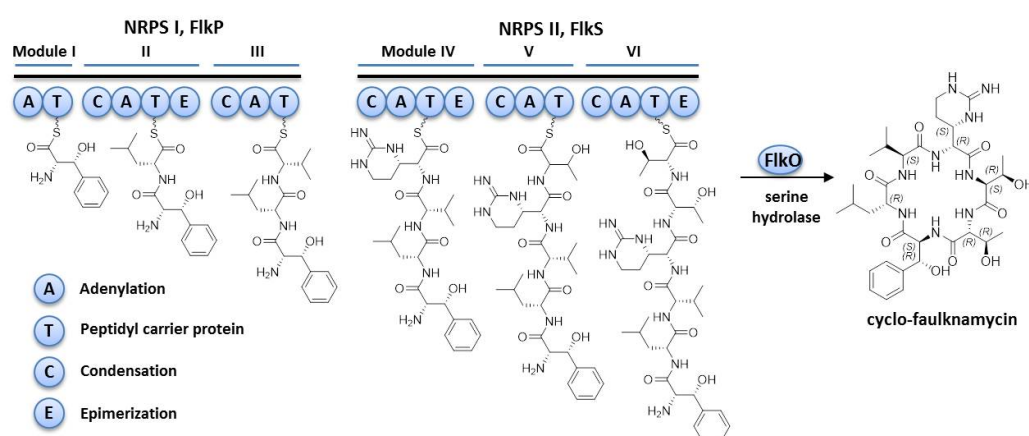


Figure 4. Proposed biosynthesis of cyclofaulknamycin.

### 3.7. Bioactivity Profile

Many known cyclic peptide NPs containing arginine derivatives, such as viomycin, mannopeptimycin, and capreomycin, are characterized by antibacterial activities [9,10]. Cyclofaulknamycin was identified in this study as a cyclic peptide and contains a capreomycin moiety; therefore, it was assumed to possess some interesting activities. Thus, it was tested against a panel of bacterial, yeast and mold strains, including *Acinetobacter baumannii* DSM-30008, *E. coli* JW0451-2 ( $\Delta$ acrB), *S. aureus* Newman, *E. coli* BW25113 (*wt*), *M. smegmatis* mc2155, *P. aeruginosa* PA14, *B. subtilis* DSM-10, *Citrobacter freundii* DSM-30039, *Mucor hiemalis* DSM-2656, *Candida albicans* DSM-1665, *Cryptococcus neoformans* DSM-11959, and *Pichia anomala* DSM-6766 (for details, see the Section 2). The tested concentrations were 0.03–64  $\mu\text{g}\cdot\text{mL}^{-1}$ . However, no activity was detected against the tested strains in the abovementioned range (data not shown).

## 4. Conclusions

Sequence databases are a potential source of new chemical scaffolds which can be accessed by different prioritization approaches [31,32]. Herein, we applied targeted genome mining based on the utilization of alpha-ketoglutarate-dependent arginine beta-hydroxylase and pyridoxal-phosphate-dependent aminotransferase genes responsible for the biosynthetic formation of the capreomycin moiety to identify new peptide NPs and their corresponding biosynthetic gene clusters. Using this approach, we identified two groups of clusters: one group contained both genes, and the other contained only the homolog encoding arginine beta-hydroxylase. The latter type of cluster was not known to exist until recently, because the products of both genes were considered to be required for the biosynthesis of L-capreomycin [26]. Detailed analysis of one of these clusters, the *flk* cluster, in the genome of the model organism *S. albus* enabled us to identify a cyclic peptide containing D-capreomycin (rather than L-capreomycin, as commonly found) for the first time. Only linear faulknamycin has previously been described to contain a D-capreomycin moiety [20]. The detailed analysis of the *flk* cluster allowed us to suggest the biosynthesis mechanisms of cyclofaulknamycin and its capreomycin moiety (Figure 4). For the first time, a PLP-dependent aminotransferase gene product generated from a site located outside of the cluster was shown to be involved in the biosynthesis of the capreomycin moiety of cyclofaulknamycin. However, cyclofaulknamycin biosynthesis was not completely blocked in the  $\Delta$ XNR\_1347 mutant, suggesting that other homologs of the aminotransferase gene in the genome might be involved in its biosynthesis or its synthesis proceeds partly via the mechanism described by Tryon et al. The analysis of the *S. lividans* genome, however, revealed a homolog of the XNR\_1347 gene showing 87% identity. Thus, it could be that the product of this gene might also be involved in the biosynthesis of capreomycin in *S. lividans*.

Comparisons of the cyclic hexapeptides encoding *flk* and mannopeptimycin gene cluster revealed a high degree of similarity. Both clusters exhibit some gene homologs that are involved in the biosynthesis and attachment of mannose, which suggests that both clusters originate from a common ancestor. Taking into account the similar core structures of the two peptide products, it might be interesting to use genes from both clusters, such as the isovaleryltransferase gene or the genes responsible for the biosynthesis and attachment of mannose, for combinatorial biosynthesis to generate unnatural analogs.

In summary, NPs with rare nonproteinogenic amino acid derivatives of arginine and their corresponding biosynthetic genes are of particular interest. They can serve as chemical scaffolds for semisynthetic modifications or as a natural library of genes for combinatorial biosynthesis to expand the structural diversity of nonribosomal peptides. In general, approaches based on searching for genes whose products are involved in the biosynthesis of interesting and unique moieties enable prioritization and targeted genome mining. This might represent a faster strategy for identifying new NPs with interesting structures and activities.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/microorganisms9081609/s1>, Table S1. Comparison of the cyclofaulknamicin gene cluster from *S. albus* with the cluster from *S. kozangensis*; Table S2. NMR data of iso-faulknamicin 1 (DMSO-d<sub>6</sub>, 1H: 700 MHz, 13C: 175 MHz); Table S3. NMR data of cyclofaulknamicin 2 (DMSO-d<sub>6</sub>, 1H: 700 MHz, 13C: 175 MHz); Table S4. Stereoisomers obtained after the hydrolysis of capreomycin and chymostatin and the equivalent enantiomers with equivalent retention time (RT); Table S5. Calculated and observed y-ion and b-ion fragments of iso-faulknamicin; Table S6. Calculated and observed a-ion and x-ion fragments within 5 ppm of cyclofaulknamicin; Figure S1. The 1H NMR spectrum of iso-faulknamicin 1 (DMSO-d<sub>6</sub>, 700 MHz); Figure S2. COSY spectrum of iso-faulknamicin 1 (DMSO-d<sub>6</sub>, 700 MHz); Figure S3. HSQC spectrum of iso-faulknamicin 1 (DMSO-d<sub>6</sub>, 700 MHz, 175 MHz); Figure S4. HMBC spectrum of iso-faulknamicin 1 (DMSO-d<sub>6</sub>, 700 MHz, 175 MHz); Figure S5. ROESY spectrum of iso-faulknamicin 1 (DMSO-d<sub>6</sub>, 700 MHz); Figure S6. The 1H spectrum of cyclofaulknamicin 2 (DMSO-d<sub>6</sub>, 700 MHz); Figure S7. COSY spectrum of cyclofaulknamicin 2 (DMSO-d<sub>6</sub>, 700 MHz); Figure S8. HSQC spectrum of cyclofaulknamicin 2 (DMSO-d<sub>6</sub>, 700 MHz, 175 MHz); Figure S9. HMBC spectrum of cyclofaulknamicin 2 (DMSO-d<sub>6</sub>, 700 MHz, 175 MHz); Figure S10. ROESY spectrum of cyclofaulknamicin 2 (DMSO-d<sub>6</sub>, 700 MHz); Figure S11. Marfey's chromatograms of the iso-faulknamicin hydrolysate (top) and the amino acid standards (bottom) derivatized with L-FDLA; Figure S12. Capreomycin from capreomycin (3) and chymostatin (4); Figure S13. Marfey's chromatograms showing the extracted mass of DLA-capreomycin of the hydrolyzed L-DLA-iso-faulknamicin and the references obtained from the hydrolysis of capreomycin and chymostatin derivatized with D-FDLA (DSR, DSS) and L-FDLA (LSS, LSR). DSR is equivalent to LRS; therefore, iso-faulknamicin contains the 2R 3S capreomycin; Figure S14. D-threo-β-phenylserine (5) and L-threo-β-phenylserine (6); Figure S15. Marfey's chromatograms showing the extracted mass of DLA-phenylserine of the hydrolyzed iso-faulknamicin and DL-threo-β-phenylserine, both derivatized with L-FDLA; Figure S16. Absolute configuration of iso-faulknamicin; Figure S17. Observed y-ion and b-ions from the MS/MS fragmentation (see Table S5) of iso-faulknamicin; Figure S18. MS/MS fragmentation spectrum of iso-faulknamicin ([M + H]<sup>+</sup> peak); Figure S19. Observed a-ion and x-ions from the MS/MS fragmentation (see Table S6) of cyclofaulknamicin; Figure S20. MS/MS fragmentation spectrum of cyclofaulknamicin ([M + 2H]<sup>2+</sup> peak).

**Author Contributions:** The experiments were designed and evaluated by L.H., M.S. and A.L. and the practical work was performed by L.H., N.E. and A.P. NMR experiments were set up, carried out and evaluated by M.S., and the data were reviewed by J.Z. The manuscript was drafted by L.H. and M.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Dias, D.A.; Urban, S.; Roessner, U. A historical overview of natural products in drug discovery. *Metabolites* **2012**, *2*, 303–336. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Marcone, G.L.; Binda, E.; Berini, F.; Marinelli, F. Old and new glycopeptide antibiotics: From product to gene and back in the post-genomic era. *Biotechnol. Adv.* **2018**, *36*, 534–554. [\[CrossRef\]](#)
3. Zhang, D.; Lu, Y.; Chen, H.; Wu, C.; Zhang, H.; Chen, L.; Chen, X. Antifungal peptides produced by actinomycetes and their biological activities against plant diseases. *J. Antibiot.* **2020**, *73*, 265–282. [\[CrossRef\]](#)
4. Hudson, G.A.; Mitchell, D.A. RiPP antibiotics: Biosynthesis and engineering potential. *Curr. Opin. Microbiol.* **2018**, *45*, 61–69. [\[CrossRef\]](#)
5. Strieker, M.; Tanović, A.; Marahiel, M.A. Nonribosomal peptide synthetases: Structures and dynamics. *Curr. Opin. Struct. Biol.* **2010**, *20*, 234–240. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Felnagle, E.A.; Jackson, E.E.; Chan, Y.A.; Podevels, A.M.; Berti, A.D.; McMahon, M.D.; Thomas, M.G. Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Mol. Pharm.* **2008**, *5*, 191–211. [\[CrossRef\]](#)
7. Thomas, M.G.; Chan, Y.A.; Ozanick, S.G. Deciphering tuberactinomycin biosynthesis: Isolation, sequencing, and annotation of the viomycin biosynthetic gene cluster. *Antimicrob. Agents Chemother.* **2003**, *47*, 2823–2830. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Atkinson, D.J.; Naysmith, B.J.; Furkert, D.P.; Brimble, M.A. Enduracidine, a rare amino acid component of peptide antibiotics: Natural products and synthesis. *Beilstein J. Org. Chem.* **2016**, *12*, 2325–2342. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Petersen, P.J.; Wang, T.Z.; Dushin, R.G.; Bradford, P.A. Comparative in vitro activities of AC98-6446, a novel semisynthetic glycopeptide derivative of the natural product mannopeptimycin, and other antimicrobial agents against gram-positive clinical isolates. *Antimicrob. Agents Chemother.* **2004**, *48*, 739–746. [\[CrossRef\]](#)
10. Magarvey, N.A.; Haltli, B.; He, M.; Greenstein, M.; Hucul, J.A. Biosynthetic pathway for mannopeptimycins, lipoglycopeptide antibiotics active against drug-resistant gram-positive pathogens. *Antimicrob. Agents Chemother.* **2006**, *50*, 2167–2177. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Yin, X.; Zabriskie, T.M. The enduracidin biosynthetic gene cluster from *Streptomyces fungicidicus*. *Microbiology* **2006**, *152*, 2969–2983. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Guo, C.; Mandalapu, D.; Ji, X.; Gao, J.; Zhang, Q. Chemistry and biology of teixobactin. *Chemistry* **2018**, *24*, 5406–5422. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Ling, L.L.; Schneider, T.; Peoples, A.J.; Spoering, A.L.; Engels, I.; Conlon, B.P.; Mueller, A.; Schäberle, T.F.; Hughes, D.E.; Epstein, E.; et al. A new antibiotic kills pathogens without detectable resistance. *Nature* **2015**, *517*, 455–459. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Goldstein, E.; Eagle, M.C.; LaCasse, M.L. In vitro chemotherapeutic combinations against isoniazid-resistant *Mycobacterium tuberculosis* and *Mycobacterium fortuitum*. *Appl. Microbiol.* **1971**, *22*, 329–333. [\[CrossRef\]](#)
15. Chan, D.I.; Prenner, E.J.; Vogel, H.J. Tryptophan- and arginine-rich antimicrobial peptides: Structures and mechanisms of action. *Biochim. Biophys. Acta* **2006**, *1758*, 1184–1202. [\[CrossRef\]](#)
16. Kang, J.H.; Lee, M.K.; Kim, K.L.; Hahm, K.S. Structure-biological activity relationships of 11-residue highly basic peptide segment of bovine lactoferrin. *Int. J. Pept. Protein Res.* **1996**, *48*, 357–363. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Sikorska, E.; Stachurski, O.; Neubauer, D.; Małuch, I.; Wyrzykowski, D.; Bauer, M.; Brzozowski, K.; Kamysz, W. Short arginine-rich lipopeptides: From self-assembly to antimicrobial activity. *Biochim. Biophys. Acta Biomembr.* **2018**, *1860*, 2242–2251. [\[CrossRef\]](#)
18. Antonoplis, A.; Zang, X.; Wegner, T.; Wender, P.A.; Cegelski, L. A vancomycin-arginine conjugate inhibits growth of carbapenem-resistant *E. coli* and targets cell-wall synthesis. *ACS Chem. Biol.* **2019**, *14*, 2065–2070. [\[CrossRef\]](#)
19. Myronovskiy, M.; Rosenkränzer, B.; Nadmid, S.; Pujic, P.; Normand, P.; Luzhetskyy, A. Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. *Metab. Eng.* **2018**, *49*, 316–324. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Tryon, J.H.; Rote, J.C.; Chen, L.; Robey, M.T.; Vega, M.M.; Phua, W.C.; Metcalf, W.M.; Ju, K.-S.; Kelleher, N.L.; Thomson, R.J. Genome mining and metabolomics uncover a rare d-capreomycin containing natural product and its biosynthetic gene cluster. *ACS Chem. Biol.* **2020**, *15*, 3013–3020. [\[CrossRef\]](#)
21. Fu, J.; Wenzel, S.C.; Perlova, O.; Wang, J.; Gross, F.; Tang, Z.; Yin, Y.; Stewart, A.F.; Müller, R.; Zhang, Y. Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. *Nucleic Acids Res.* **2008**, *36*, e113. [\[CrossRef\]](#)
22. Zhang, Y.; Muylers, J.P.P.; Testa, G.; Stewart, A.F. DNA cloning by homologous recombination in *Escherichia coli*. *Nat. Biotechnol.* **2000**, *18*, 1314–1317. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Kieser, T.; Bibb, M.J.; Buttner, M.J.; Chater, K.F.; Hopwood, D.A. *Practical Streptomyces Genetics*, 2nd ed.; John Innes Foundation: Norwich, UK, 2000.
24. Sambrook, J.; Russell, D. *Molecular Cloning: A Laboratory Manual*, 3rd ed.; Cold Spring Harbor Laboratory Press: Harbor, NY, USA, 2001.
25. Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-Powers, Y.; van Wezel, G.P. antiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **2021**, *49*, gkab335.
26. Yin, X.; McPhail, K.L.; Kim, K.J.; Zabriskie, T.M. Formation of the nonproteinogenic amino acid 2S,3R-capreomycin by VioD from the viomycin biosynthesis pathway. *ChemBioChem* **2004**, *5*, 1278–1281. [\[CrossRef\]](#)



- 
27. Kohli, R.M.; Walsh, C.T. Enzymology of acyl chain macrocyclization in natural product biosynthesis. *Chem. Commun.* **2003**, *3*, 297–307. [[CrossRef](#)]
  28. Baltz, R.H. Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery. *J. Ind. Microbiol. Biotechnol.* **2011**, *38*, 1747–1760. [[CrossRef](#)] [[PubMed](#)]
  29. Hackl, S.; Bechthold, A. The gene *bldA*, a regulator of morphological differentiation and antibiotic production in streptomyces. *Arch. Pharm.* **2015**, *348*, 455–462. [[CrossRef](#)]
  30. Harada, K.; Fujii, K.; Hayashi, K.; Suzuki, M.; Ikai, Y.; Oka, H. Application of D,L-FDLA derivatization to determination of absolute configuration of constituent amino acids in peptide by advanced Marfey's method. *Tetrahedron Lett.* **1996**, *37*, 3001–3004. [[CrossRef](#)]
  31. Rebets, Y.; Broetz, E.; Tokovenko, B.; Luzhetskyy, A. Actinomycetes biosynthetic potential: How to bridge in silico and in vivo? *J. Ind. Microbiol. Biotechnol.* **2014**, *41*, 387–402. [[CrossRef](#)]
  32. Belknap, K.C.; Park, C.J.; Barth, B.M.; Andam, C.P. Genome mining of biosynthetic and chemotherapeutic gene clusters in *Streptomyces* bacteria. *Sci. Rep.* **2020**, *10*, 2003. [[CrossRef](#)]

## Supplementary Materials

### Cyclofaulknamycin with the rare amino acid D-capreomycin isolated from a well-characterised *Streptomyces albus* strain

Horbal L.<sup>1,3§</sup>, Stierhof M.<sup>1§</sup>, Paluszczak A.<sup>1</sup>, Eckert N.<sup>1</sup>, Zapp. J., Luzhetskyy A.<sup>1,2\*</sup>

Department of Pharmaceutical Biotechnology, Saarland University, 66123 Saarbruecken, Germany, [Lihorbal@gmail.com](mailto:Lihorbal@gmail.com) (L.H.); [marc.stierhof@uni-saarland.de](mailto:marc.stierhof@uni-saarland.de) (M.S.); [anja.paluszczak@uni-saarland.de](mailto:anja.paluszczak@uni-saarland.de) (P.A.)

<sup>2</sup> Department of Pharmaceutical Biology, Saarland University, 66123 Saarbruecken, Germany; [j.zapp@mx.uni-saarland.de](mailto:j.zapp@mx.uni-saarland.de) (J.Z.)

<sup>3</sup> AMEG Department, Helmholtz Institute for Pharmaceutical Research Saarland, 66123 Saarbruecken Germany, [nikolas.eckert@uni-saarland.de](mailto:nikolas.eckert@uni-saarland.de) (N.E.)

\* Correspondence: [a.luzhetskyy@mx.uni-saarland.de](mailto:a.luzhetskyy@mx.uni-saarland.de); Tel.: +49-681-302-70200 (A.L.)

**Table S1.** Comparison of the cyclofaulknamycin gene cluster from *S. albus* with the cluster from *S. koyangensis*.

Gene in the <i>S. albus</i> genome	Gene name	Putative product	Locus in the genome of <i>S. koyangensis</i> VK-A60T
XNR_0965		urea ABC transporter permease subunit UrtC [Streptomyces albidoflavus]	
XNR_0966		urea ABC transporter ATP-binding protein UrtD [Streptomyces albidoflavus]	
XNR_0967		ATP-binding cassette domain-containing protein [Streptomyces albidoflavus]	
XNR_0968	<i>flkA</i>	DeoR/GlpR transcriptional regulator [Streptomyces sp. CS227]	D0C37_RS06225
XNR_0969	<i>flkB</i>	sugar ABC transporter substrate-binding protein [Streptomyces albidoflavus]	D0C37_RS06230
XNR_0970	<i>flkC</i>	integral membrane sugar transporter	D0C37_RS06235
XNR_0971	<i>flkD</i>	carbohydrate ABC transporter permease	D0C37_RS06240
XNR_0972	<i>flkE</i>	zinc-dependent alcohol dehydrogenase family protein	D0C37_RS06245
XNR_0973	<i>flkF</i>	DUF2029 domain-containing protein	D0C37_RS06250
XNR_0974	<i>flkG</i>	FkbM family methyltransferase	D0C37_RS06255
XNR_0975	<i>flkH</i>	glycosyltransferase family 4 protein alpha-(1-2)-phosphatidylinositol mannosyltransferase	D0C37_RS06260
XNR_0976	<i>flkI</i>	response regulator transcription factor	D0C37_RS06265
XNR_0977	<i>flkJ</i>	histidine kinase	D0C37_RS06270
XNR_0978	<i>flkK</i>	ABC transporter ATP-binding protein	D0C37_RS06275
XNR_0979	<i>flkL</i>	ABC type antimicrobial transporter permease, FtsX-like permease family protein	D0C37_RS06280
XNR_0980	<i>flkM</i>	LuxR family response regulator	D0C37_RS06285
XNR_0981	<i>flkN</i>	Two component sensor histidine kinase	D0C37_RS06290
XNR_0982	<i>flkO</i>	Alpha/beta hydrolase MppK beta-lactamase, serine hydrolase	D0C37_RS06295
XNR_0983	<i>flkP</i>	NRPS	D0C37_RS06300
XNR_0984	<i>flkR</i>	MbtH family protein	D0C37_RS06305
XNR_0985	<i>flkS</i>	NRPS	D0C37_RS06310
-	-	NRPS	D0C37_RS06315
XNR_0986	<i>flkT</i>	Fe(II)/alpha-ketoglutarate-dependent arginine beta-hydroxylase	D0C37_RS06320
XNR_0987	<i>flkU</i>	Major Facilitator Superfamily transporter	D0C37_RS06325
XNR_0988	<i>flkV</i>	Cytochrome 450	D0C37_RS06330
XNR_0989	<i>flkW</i>	ArsR family transcriptional regulator	D0C37_RS06335
XNR_0990	<i>flkX</i>	Thioesterase, alpha beta hydrolase	D0C37_RS06340
XNR_0991		Erythromycin esterase protein	
XNR_0992		class I SAM-dependent RNA methyltransferase	
XNR_0993		phenylalanine-specific permease	
XNR_0994		TrkA family potassium uptake protein	
XNR_0995		TrkA family potassium uptake protein	
XNR_0996		DUF3159 domain-containing protein	
XNR_0997		OB-fold nucleic acid binding domain-containing protein	
XNR_0998		Osmosensitive K <sup>+</sup> channel histidine kinase KdpD, response regulator	
XNR_0999		sensor histidine kinase KdpD	
XNR_1000		ABC transporter	
XNR_1001		Short chain dehydrogenase	
XNR_1002		TetR/AcrR family transcriptional regulator	
XNR_1003		DUF3710 domain-containing protein	
<b>XNR_1347</b>		<b>PLP-dependent aminotransferase (pyridoxal dependent)</b>	

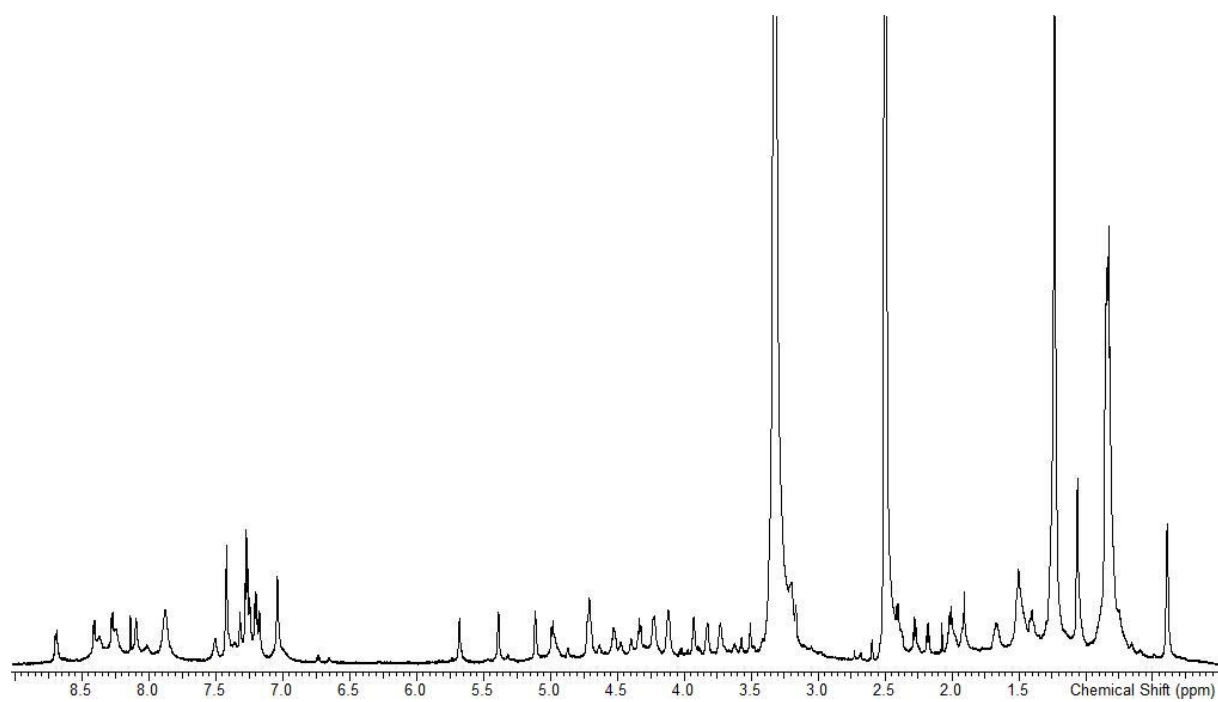


# 1. NMR Spectroscopy

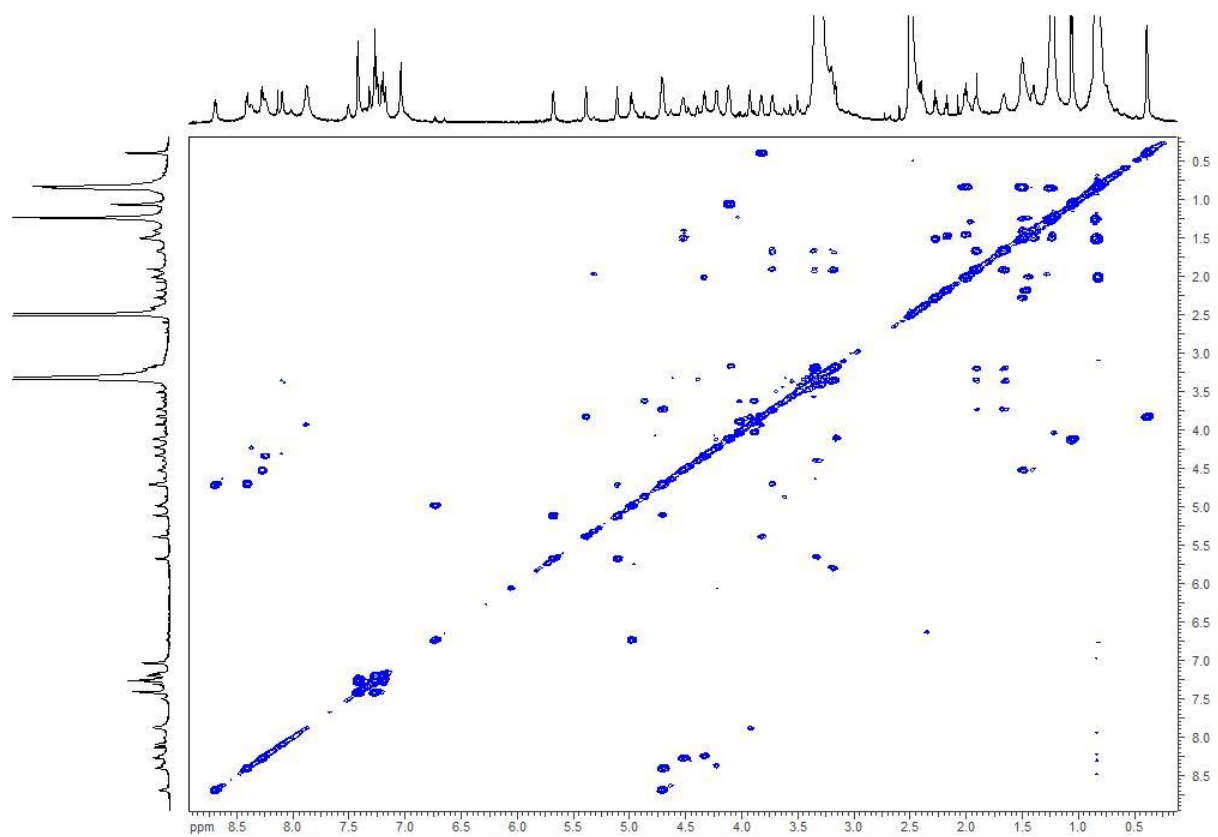
**Table S2.** NMR data of iso-faulknamycin **1** (DMSO-d<sub>6</sub>, <sup>1</sup>H: 700 MHz, <sup>13</sup>C: 175 MHz)

unit	δ <sub>c</sub>	δ <sub>H</sub> , multiplicity, (J in Hz)	COSY	ROESY	HMBC (H-)
1 – COOH	n.a. <sup>1</sup>				
2 – CH	58.1	4.22, m	3, 5		
3 – CH	66.4	4.12, m	2, 4		
4 – CH <sub>3</sub>	20.3	1.06, d (6.1)	3		2, 3
5 – NH		8.37	2		
6 – CO	n.a. <sup>1</sup>				
7 – CH	54.3	4.70, ovl. <sup>2</sup>	8, 14		
8 – CH	50.2	3.73, m	7, 9		8
9 – CH <sub>2</sub> a CH <sub>2</sub> b	22.0	1.93, ovl. <sup>2</sup> 1.67, m	8, 10		
10 – CH <sub>2</sub> a CH <sub>2</sub> b	36.1	3.37, ovl. <sup>2</sup> 3.19, m	9, 11		
11 – NH		8.09, bs	10		
12 – C		n.a. <sup>1</sup>			
12 – NH		n.a. <sup>1</sup>			
13 – NH		n.a. <sup>1</sup>			
14 – NH		8.41, d (8.2)	7	16	
15 – CO	171.3				
16 – CH	57.8	4.33, t (8.1)	17, 20	14	15, 21
17 – CH	30.9	2.01, m	16, 18, 19		
18 – CH <sub>3</sub>	19.3	0.84, ovl. <sup>2</sup>	17		16, 17, 19
19 – CH <sub>3</sub>	18.1	0.82, ovl. <sup>2</sup>	17		16, 17, 18
20 – NH		8.24, bs	16	22	
21 – CO	172.1				
22 – CH	51.2	4.52, m	23, 27	20	
23 – CH <sub>2</sub> a CH <sub>2</sub> b	41.8	1.49, ovl. <sup>2</sup> 1.40, m	22, 24		
24 – CH	24.2	1.50, ovl. <sup>2</sup>	23, 25, 26		
25 – CH <sub>3</sub>	23.2	0.86, ovl. <sup>2</sup>	24		23, 24, 26
26 – CH <sub>3</sub>	21.6	0.81, ovl. <sup>2</sup>	24		23, 24, 25
27 – NH		8.27, d (8.4)	22		28
28 – CO	169.0				
29 – CH	58.7	4.72, ovl. <sup>2</sup>	30, 37		28
30 – CH	72.8	5.11, m	29, 30-OH		31, 32/36
30 – OH		5.68, d (4.6)	30		29, 30
31 – C	142.1				
32/36 – CH	126.1	7.42, d (7.3)	33/35		30, 32, 34, 36
33/35 – CH	127.6	7.27, t (7.5)	32/36, 34		31, 33, 35
34 – CH	127.0	7.20, t (7.0)	33/35		32, 36
37 – NH		8.69, d (8.7)	29	39	
38 – CO	n.a. <sup>1</sup>				
39 – CH	57.4	3.93, m	40, 42	37	
40 – CH	64.6	3.83, m	39, 41, 40 OH		
40 – OH		5.39, bs	40		
41 – CH <sub>3</sub>	16.5	0.39, d (6.0)	40		39, 40
42 – NH <sub>2</sub>		7.88, bs	39		

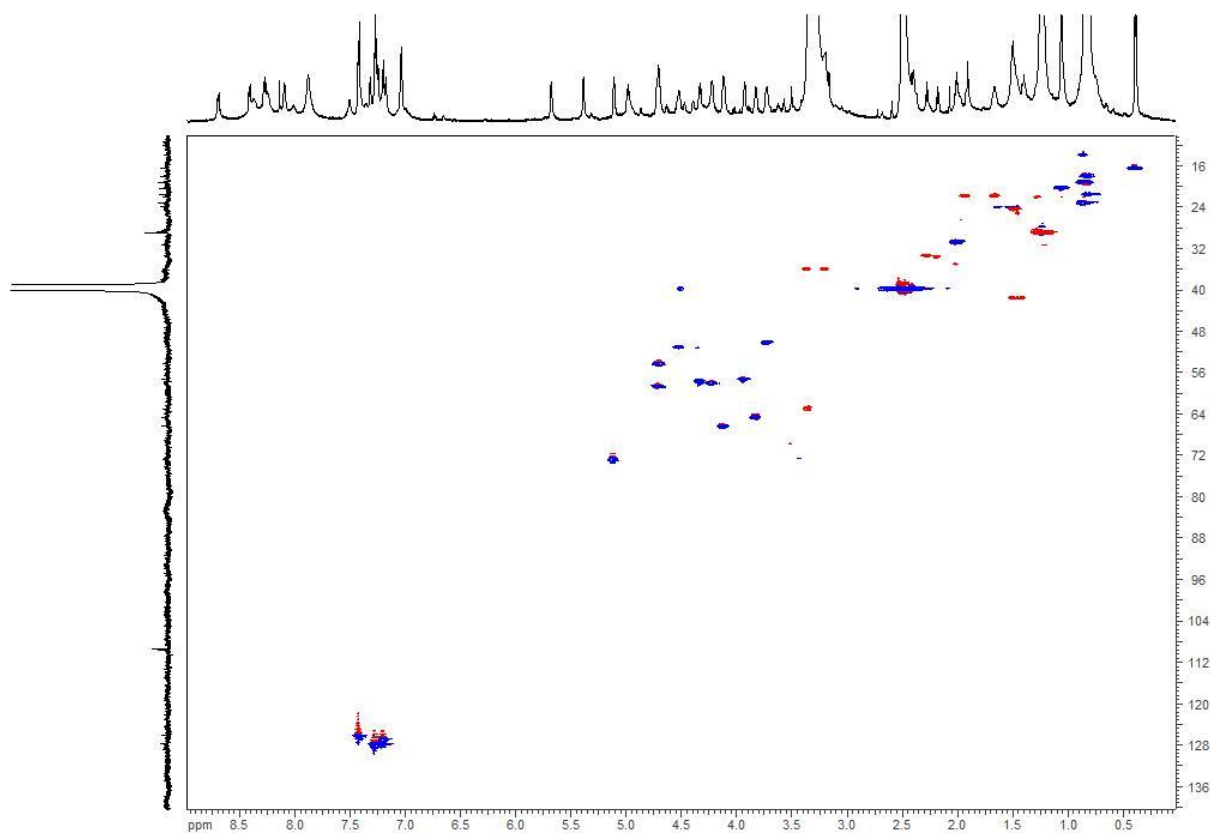
<sup>1</sup>n.a. = not available; <sup>2</sup>ovl. = signal overlap



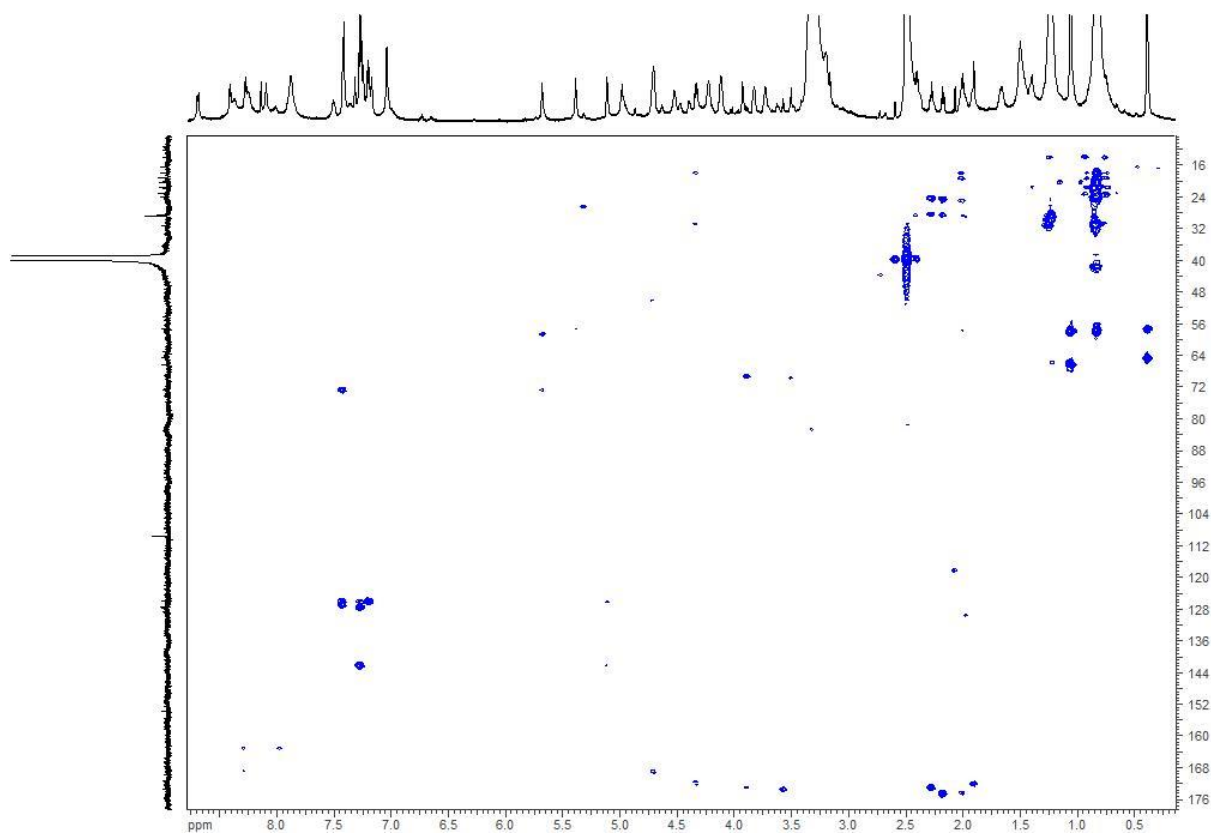
**Figure S1:**  $^1\text{H}$  NMR spectrum of iso-faulknamycin **1** (DMSO- $d_6$ , 700 MHz)



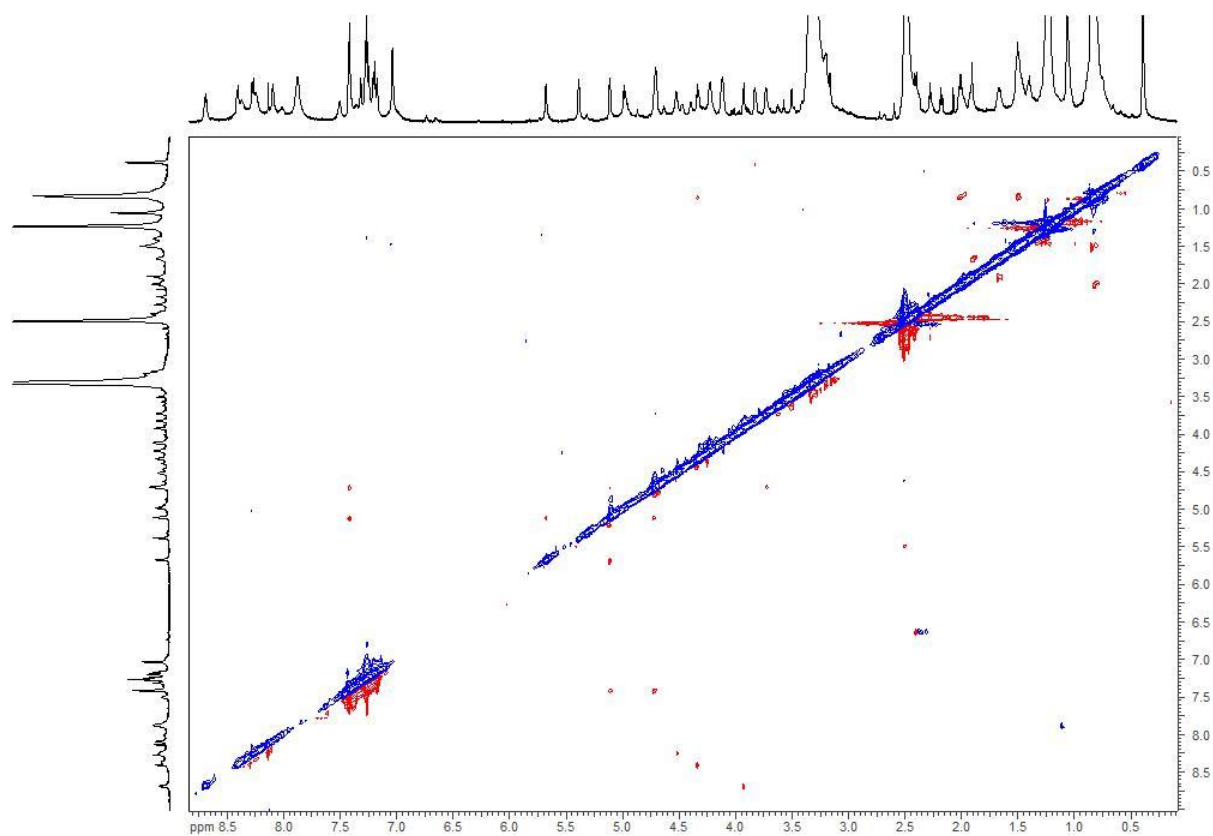
**Figure S2:** COSY spectrum of iso-faulknamycin **1** (DMSO- $d_6$ , 700 MHz)



**Figure S3:** HSQC spectrum of iso-faulknamycin **1** (DMSO-*d*<sub>6</sub>, 700 MHz, 175 MHz)



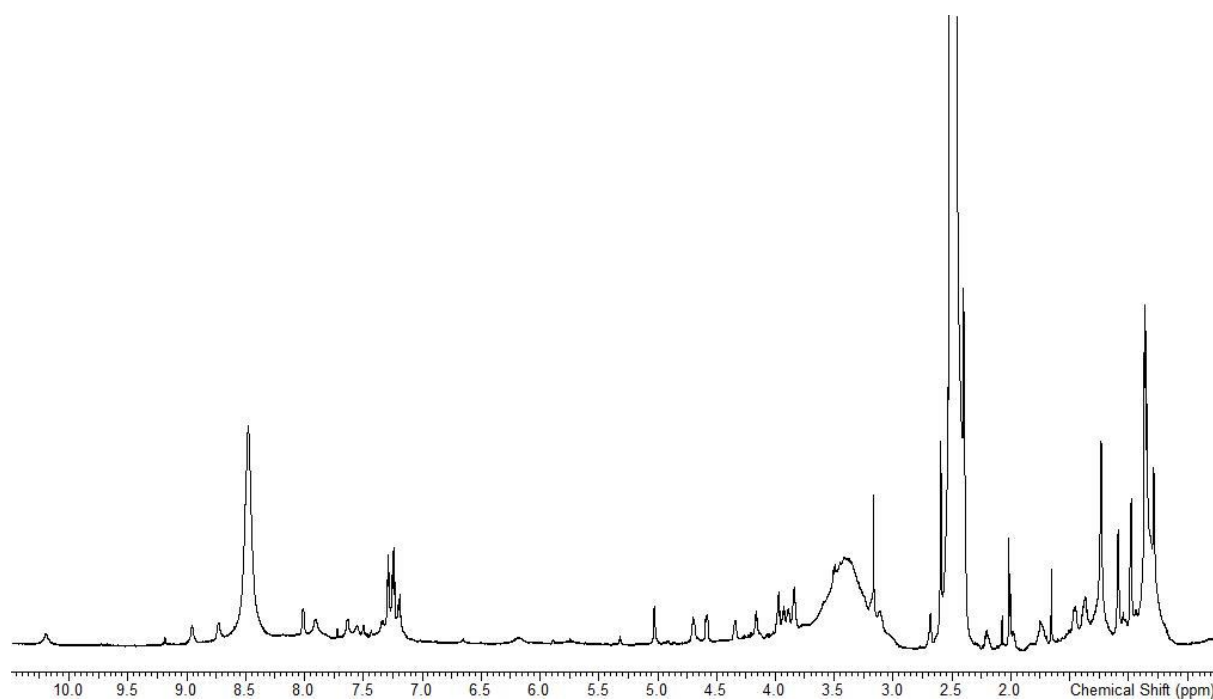
**Figure S4:** HMBC spectrum of iso-faulknamycin **1** (DMSO-*d*<sub>6</sub>, 700 MHz, 175 MHz)



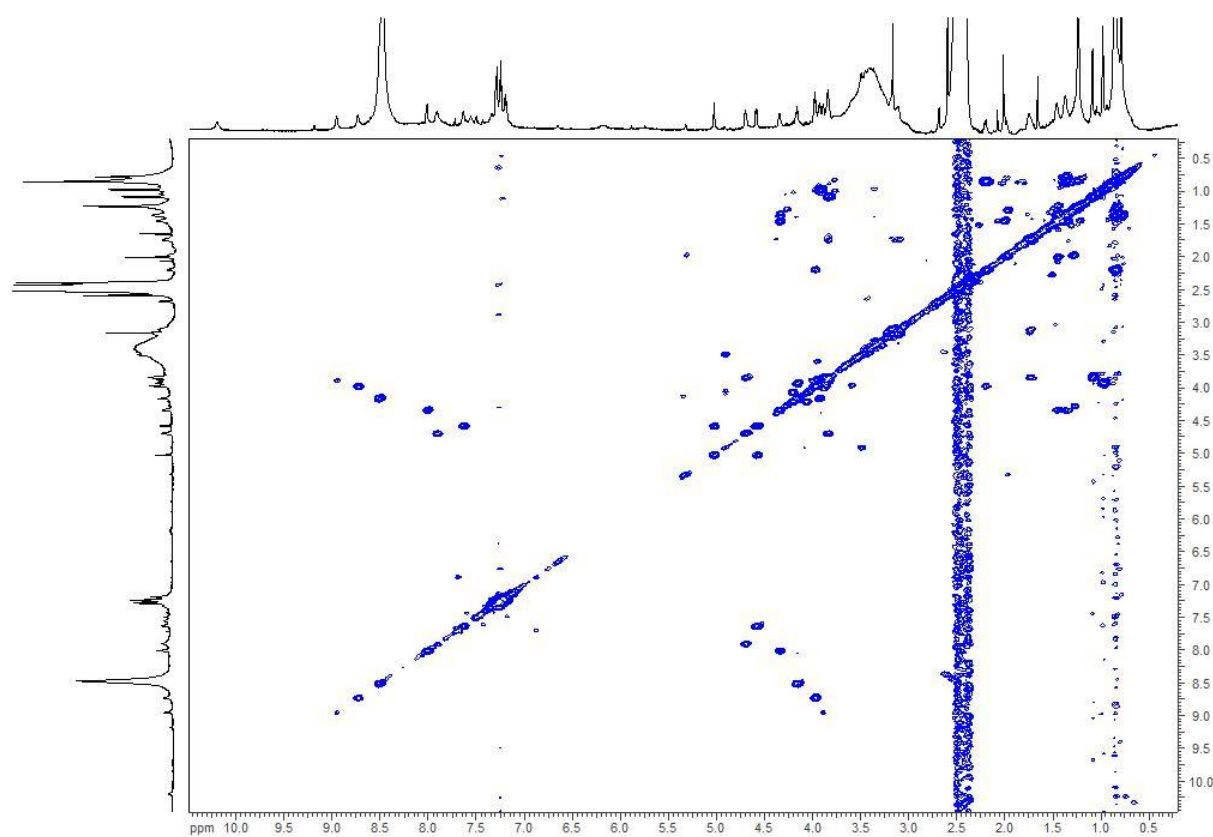
**Figure S5:** ROESY spectrum of iso-faulknamycin **1** (DMSO-d<sub>6</sub>, 700 MHz)

**Table S3:** NMR data of cyclofaulknamycin **2** (DMSO-d<sub>6</sub>, <sup>1</sup>H: 700 MHz, <sup>13</sup>C: 175 MHz)

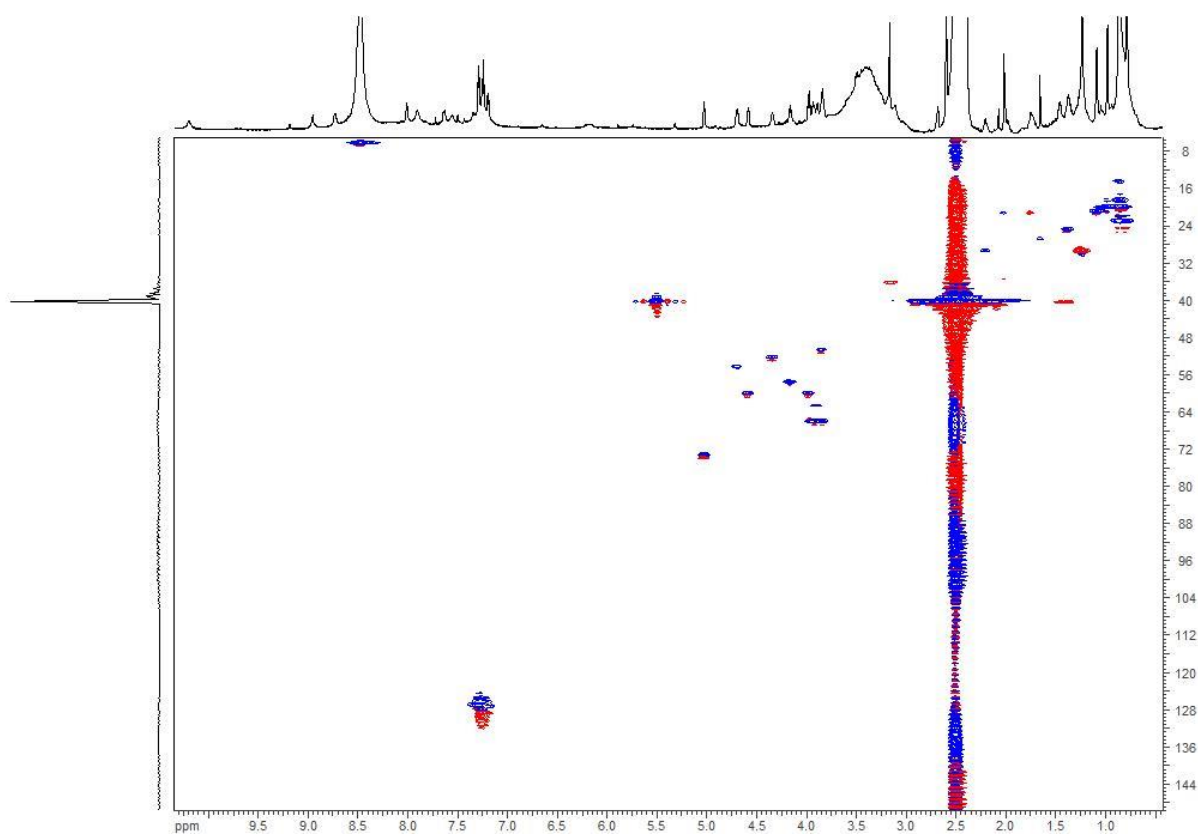
unit	δ <sub>c</sub>	δ <sub>H</sub> , multiplicity, (J in Hz)	COSY	ROESY	HMBC (H-)
1 – CO	171.6				
2 – CH	62.7	3.89, m	3, 5	42	
3 – CH	65.0	3.85, ovl	2, 4		
4 – CH <sub>3</sub>	20.9	1.09, d (6.0)	3		2, 3
5 – NH		8.94, bs	2		
6 – CO	n.a				
7 – CH	54.5	4.69, t (6.6)	8, 14		
8 – CH	50.7	3.84, ovl	7, 9		
9 – CH <sub>2</sub>	21.3	1.76, m	8, 10		
10 – CH <sub>2</sub>	36.3	3.11	9		
11 – NH		n.a			
12 – C	n.a				
12 – NH		n.a.			
13 – NH		n.a			
14 – NH		7.91, bs	7		
15 – CO	n.a				
16 – CH	60.1	3.97, dd (6.4, 11.6)	17, 20		21
17 – CH	29.4	2.2, m	16, 18, 19		
18 – CH <sub>3</sub>	20.0	0.86, ovl	17		16, 17, 19
19 – CH <sub>3</sub>	18.5	0.88, ovl	17		16, 17, 18
20 – NH		8.73	16	22	
21 – CO	172.4				
22 – CH	52.3	4.34, m	23, 27	20	
23 – CH <sub>2</sub> a CH <sub>2</sub> b	40.4	1.37, ovl 1.46, m	22, 24		21
24 – CH	24.9	1.38, ovl	23, 25, 26		21
25 – CH <sub>3</sub>	23.1	0.79, d (5.9)	24		23, 24, 26
26 – CH <sub>3</sub>	23.1	0.87, ovl	24		23, 24, 25
27 – NH		8.01, ovl	22	29	28
28 – CO	169.7				
29 – CH	60.0	4.58, dd (8.2, 3.9)	30, 37	27	28
30 – CH	73.3	5.03, d (3.0)	29		32, 36
31 – C	142.3				
32/36 – CH	127.2	7.29, d (7.1)	33/35		30, 32/36, 34
33/35 – CH	128.1	7.25, t (7.2)	32/36, 34		31, 33/35
34 – CH	127.5	7.20, t (7.1)	33/35		32, 36
37 – NH		7.64, bs	29	39	
38 – CO	n.a				
39 – CH	57.7	4.16, t (7.3)	40, 42	37	1
40 – CH	66.0	3.93, m	39, 41		
41 – CH <sub>3</sub>	20	0.98, d (6.2)	40		39, 40
42 – NH		8.52, ovl	39	2	



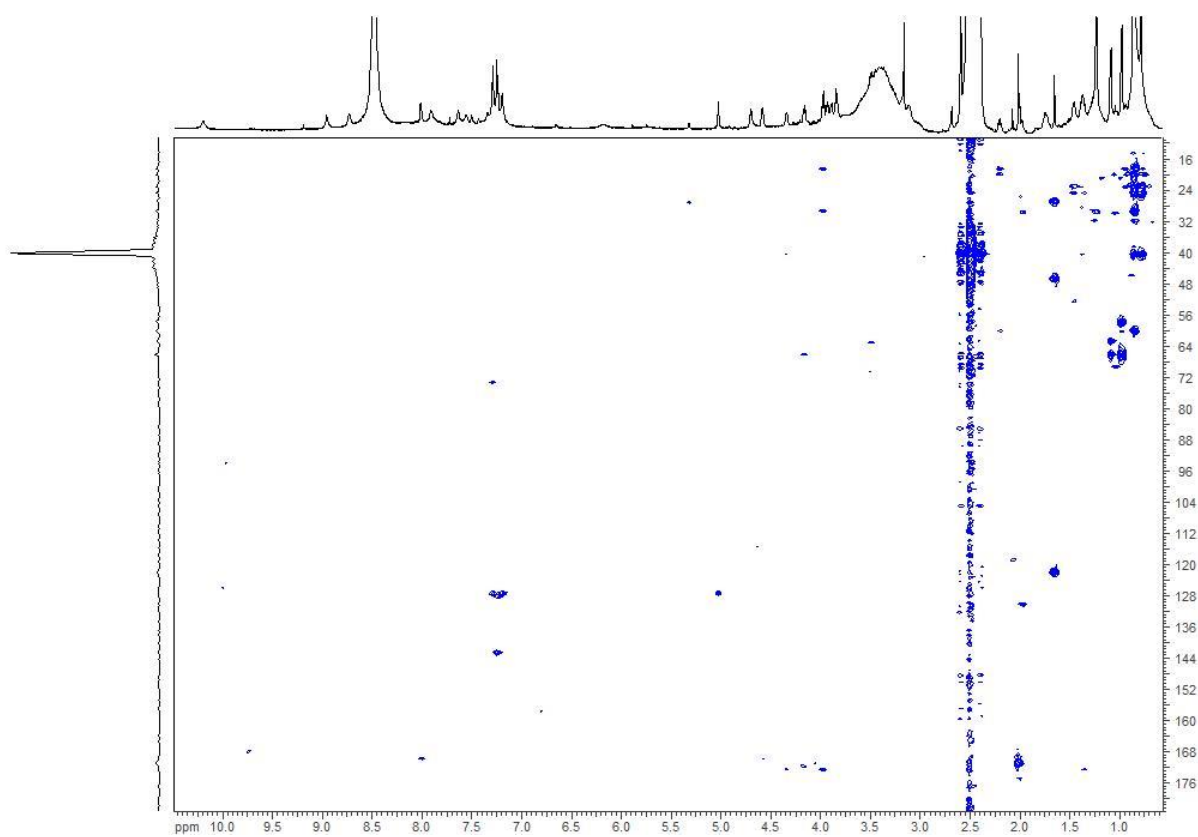
**Figure S6:**  $^1\text{H}$  spectrum of cyclofaulknamycin **2** (DMSO- $d_6$ , 700 MHz)



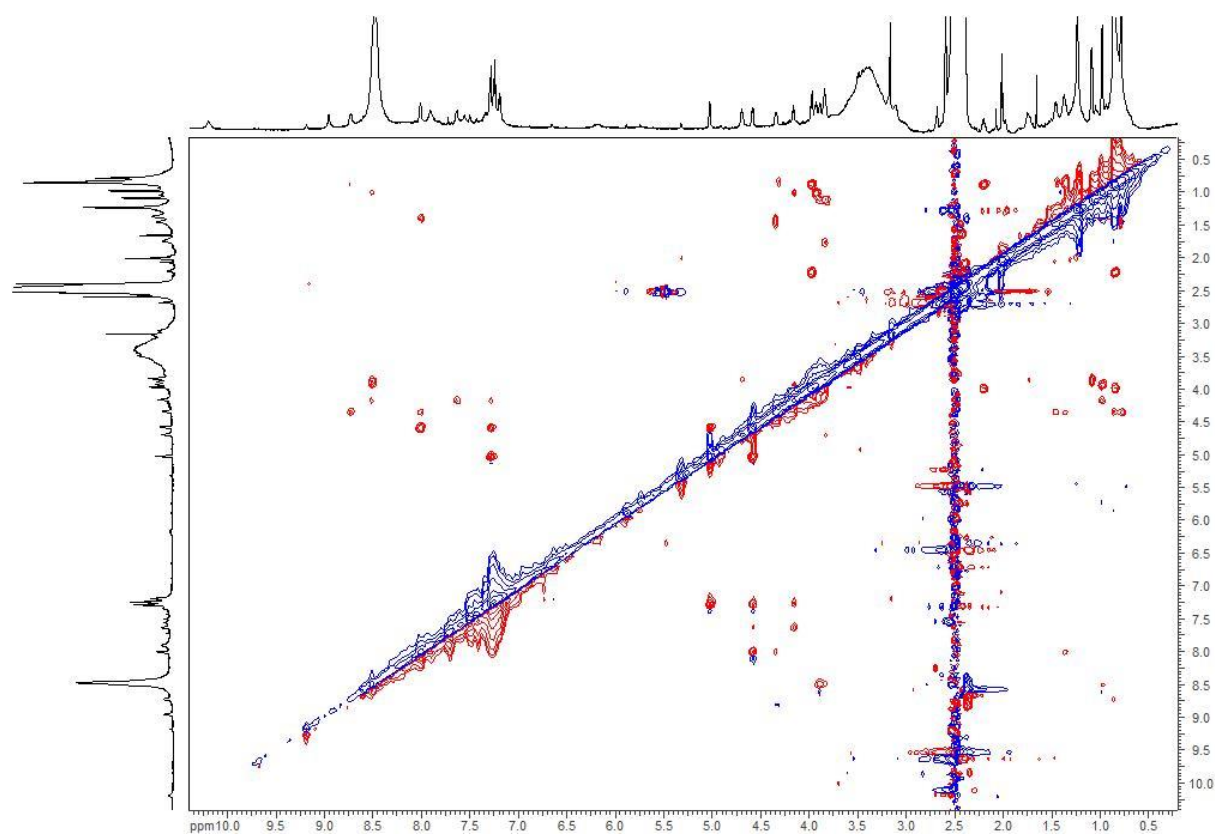
**Figure S7:** COSY spectrum of cyclofaulknamycin **2** (DMSO- $d_6$ , 700 MHz)



**Figure S8:** HSQC spectrum of cyclofaulknamycin **2** (DMSO- $d_6$ , 700 MHz, 175 MHz)



**Figure S9:** HMBC spectrum of cyclofaulknamycin **2** (DMSO- $d_6$ , 700 MHz, 175 MHz)

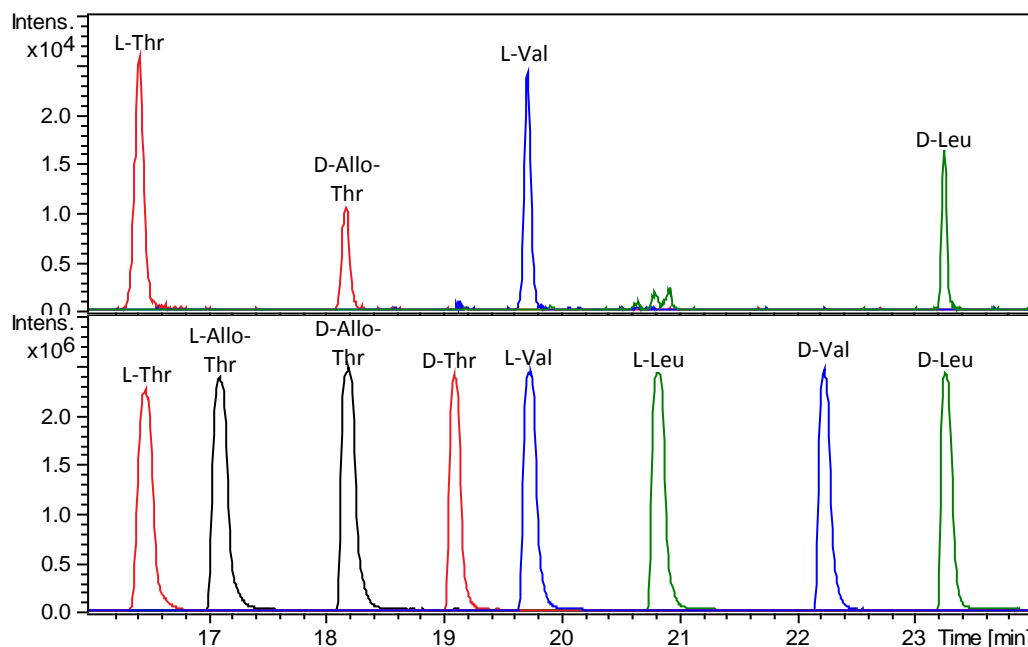


**Figure S1:** ROESY spectrum of cyclofaulknamycin **2** (DMSO-d<sub>6</sub>, 700 MHz)



## 2. Marfey's Analysis

Standards in D- and L-configuration were derivatized with L-FDLA and compared with the amino acids derived from hydrolysis of **1** (Fig. S11).

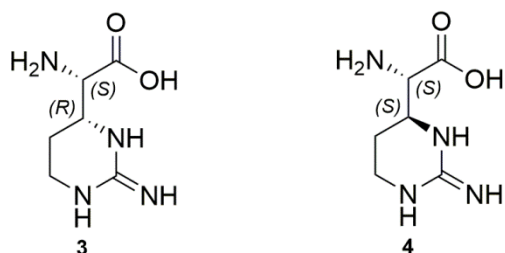


**Figure S2:** Marfey's chromatograms of the iso-faulknamycin hydrolysate (top) and the amino acids standards (bottom) derivatized with L-FDLA.

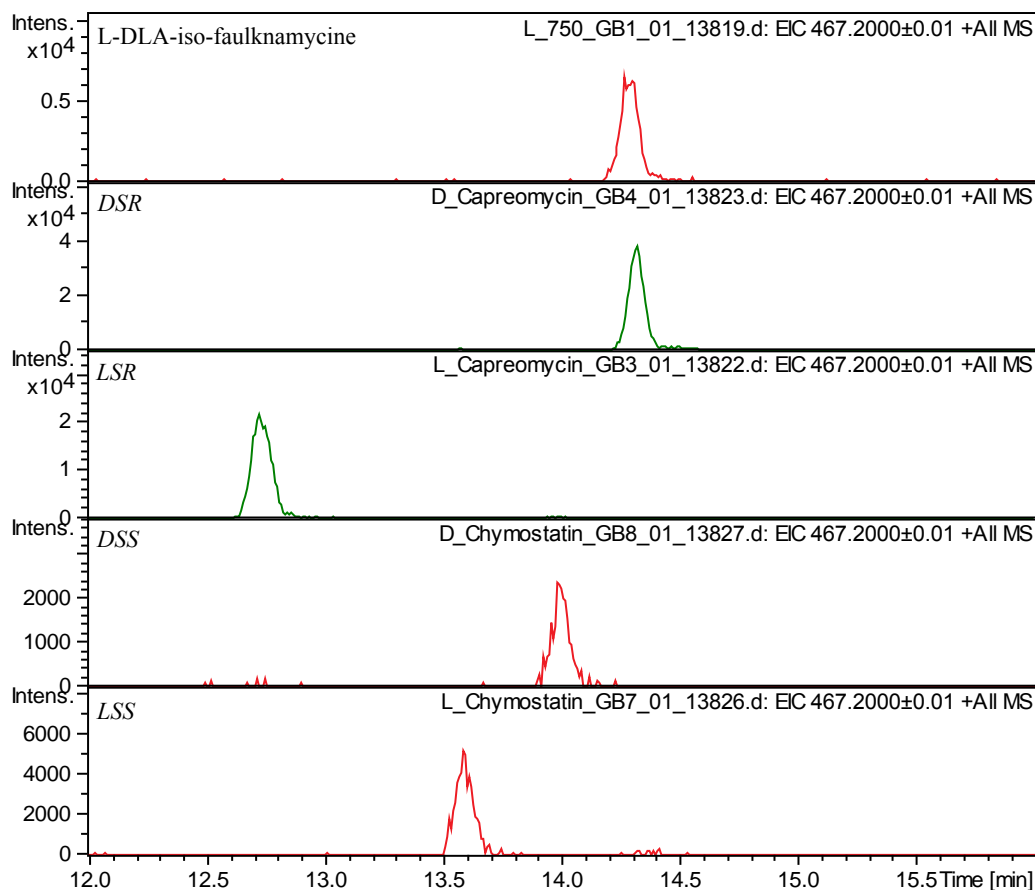
We obtained 2S-3R-capreomycinide **3** and 2S-3S-capreomycinide **4** by hydrolysis of capreomycin and chymostatin, respectively (fig. S12). The hydrolysates were derivatized with L-FDLA and D-FDLA, yielding all possible enantiomers with distinguishable retention time (table S4).

**Table S4:** Stereoisomers obtained after hydrolysis of capreomycin and chymostatin and the equivalent enantiomers with equivalent retention time (RT).

hyd. of capreomycin	L-FDLA	D-FDLA	hyd. of chymostatin	L-FDLA	D-FDLA
2S-3R-capreomycinide	<i>LSR</i>	<i>DSR</i>	2S-3S-capreomycinide	<i>LSS</i>	<i>DSS</i>
RT equivalent to	<i>DRS</i>	<i>LRS</i>	RT equivalent to	<i>DRR</i>	<i>LRR</i>

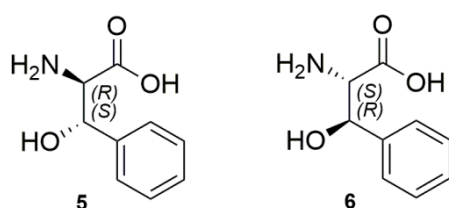


**Figure S12:** Capreomycinide from capreomycin (**3**) and chymostatine (**4**).

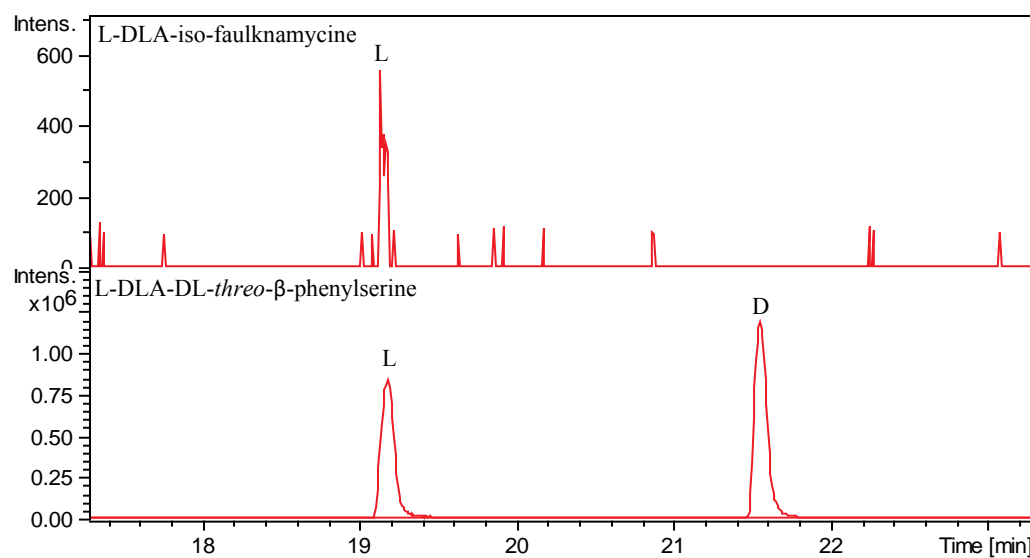


**Figure S13:** Marfey's chromatograms showing the extracted mass of DLA-capreomycidine of the hydrolysed L-DLA-iso-faulknamycin and the references obtained from hydrolysis of capreomycidine and chymostatin derivatized with D-FDLA ( DSR, DSS) and L-FDLA (LSS, LSR). DSR is equivalent to LRS, therefore iso-faulknamycin contains the 2R-3S-capreomycidine.

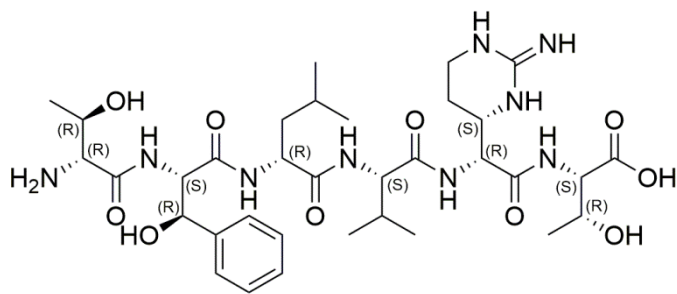
D-*threo*- $\beta$ -phenylserine **5** has been determined to be the constituent amino acid in faulknamycin (fig. S14). Therefore, we used DL-*threo*- $\beta$ -phenylserine as a reference to determine the stereochemistry of  $\beta$ -phenylserine of the hydrolysed iso-faulknamycin (Fig. S14, S16).



**Figure S14:** D-*threo*- $\beta$ -phenylserine (**5**) and L-*threo*- $\beta$ -phenylserine (**6**)

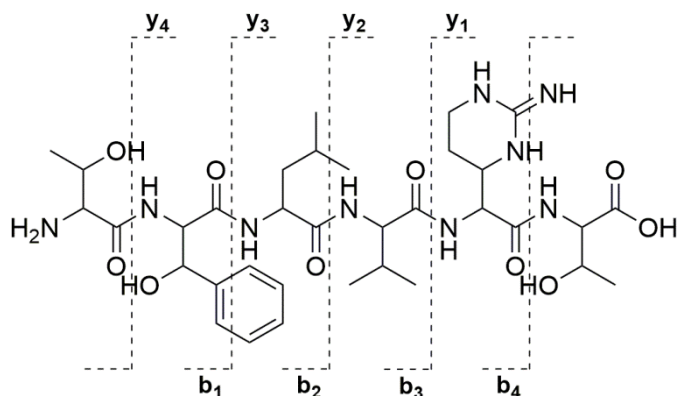


**Figure S15:** Marfey's chromatograms showing the extracted mass of DLA-phenylserine of the hydrolysed iso-faulknamycin and DL-*threo*- $\beta$ -phenylserine, both derivatized with L-FDLA.



**Figure S16:** Absolute configuration of iso-faulknamycin.

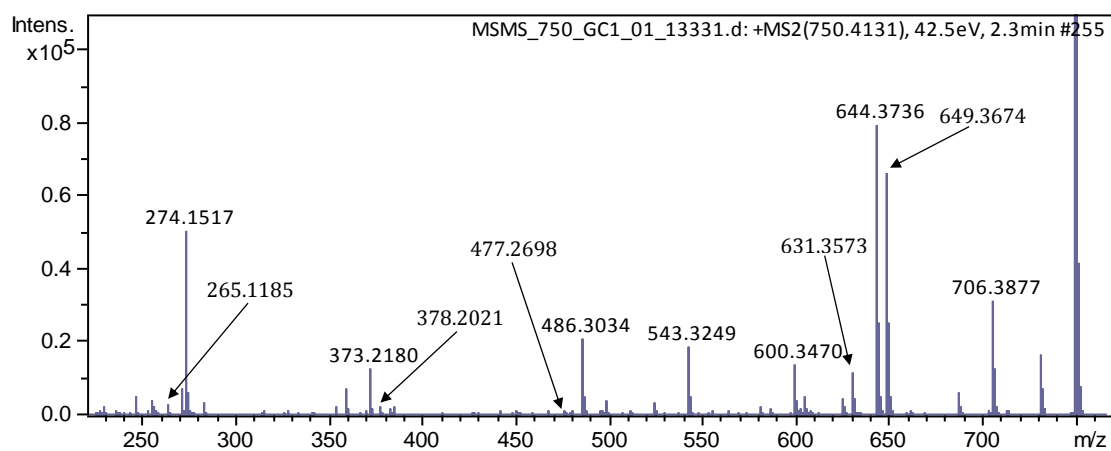
### 3. MS/MS Fragmentation



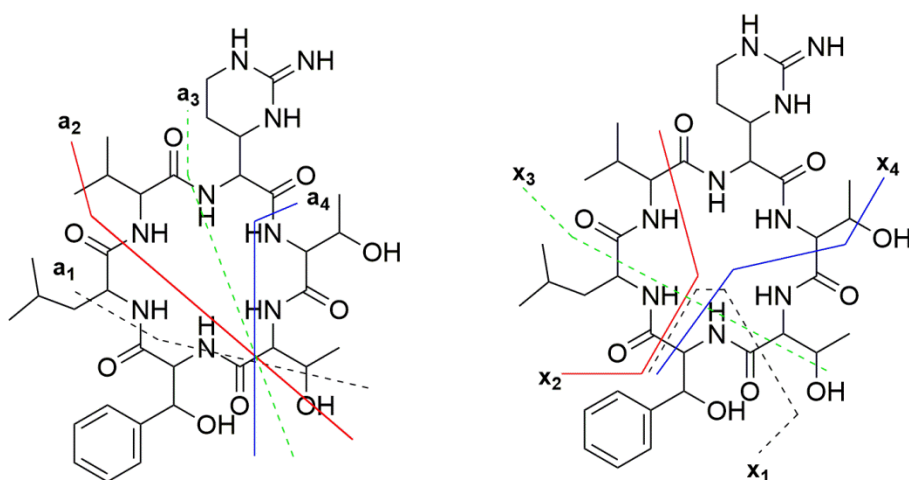
**Figure S3:** Observed y-ion and b-ions from the MS/MS fragmentation (see table 5) of iso-faulknamycin

**Table S5:** Calculated and observed y-ion and b-ion fragments of iso-faulknamycin.

ion type	y <sub>1</sub>	y <sub>2</sub>	y <sub>3</sub>	y <sub>4</sub>
[M] <sup>+</sup> calculated	274.1511	373.2190	486.3032	649.3675
[M] <sup>+</sup> observed	274.1517	373.2180	486.3034	649.3674
Δ [ppm]	-2.2	3.8	-0.4	0.2
ion type	b <sub>1</sub>	b <sub>2</sub>	b <sub>3</sub>	b <sub>4</sub>
[M] <sup>+</sup> calculated	265.1186	378.2019	477.2697	631.3556
[M] <sup>+</sup> observed	265.1185	378.2021	477.2698	631.3573
Δ [ppm]	0.4	-0.5	-0.2	-2.7



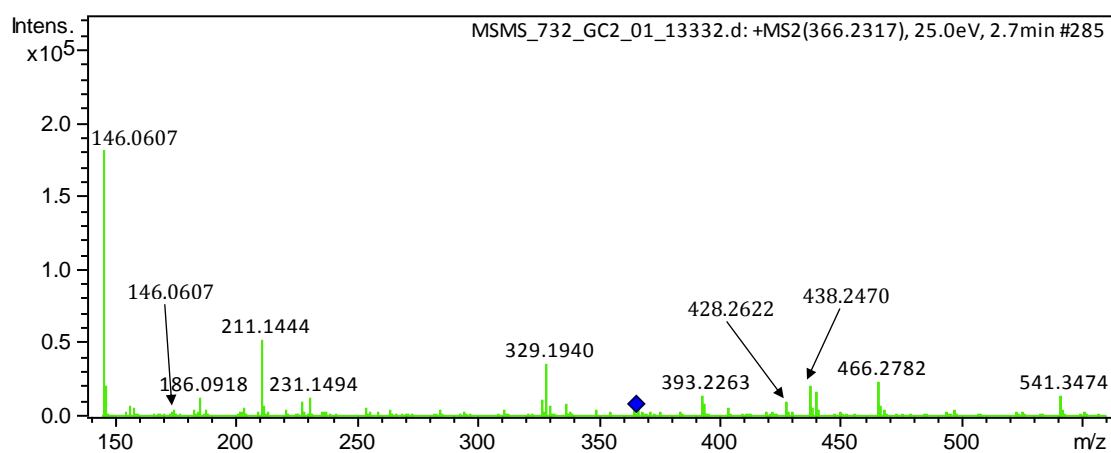
**Figure S18:** MS/MS fragmentation spectrum of iso-faulknamycin ([M+H]<sup>+</sup> peak).



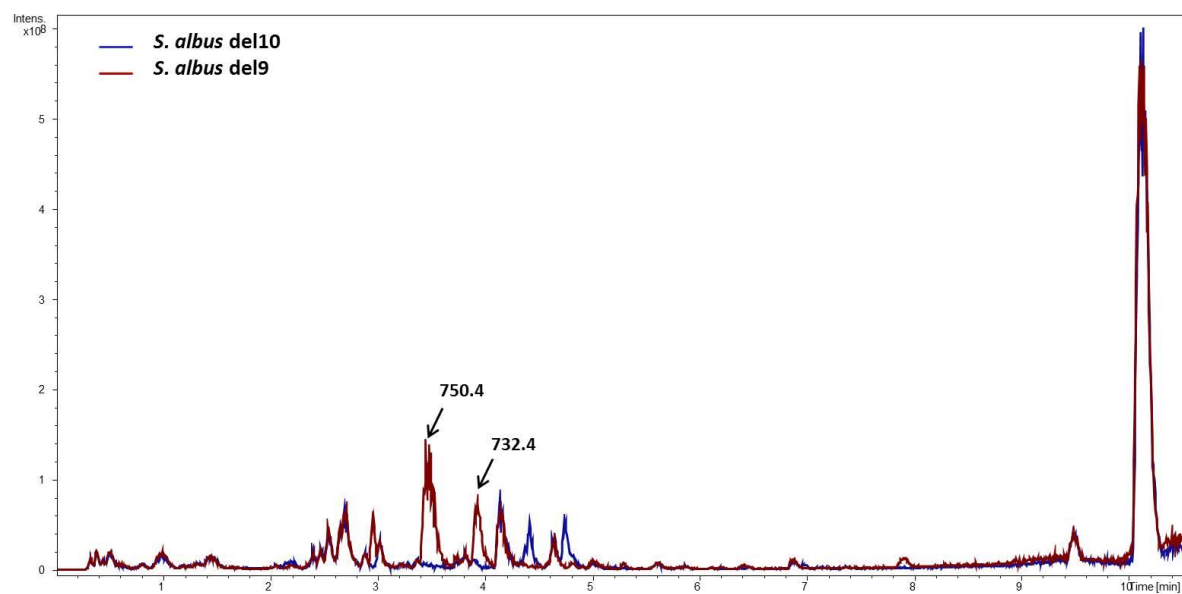
**Figure S19:** Observed a-ion and x-ions from the MS/MS fragmentation (see table 6) of cyclofaulknamycin

**Table S6:** Calculated and observed a-ion and x-ion fragments within 5 ppm of cyclofaulknamycin.

ion type	a1	a2	a3	a4
AA sequence	T <sub>1</sub> -T <sub>2</sub>	T <sub>1</sub> -T <sub>2</sub> -Cmp	T <sub>1</sub> -T <sub>2</sub> -Cmp-V	T <sub>1</sub> -T <sub>2</sub> -Cmp-V-L
[M] <sup>+</sup> calculated	175.1077	329.1932	428.2616	541.3457
[M] <sup>+</sup> observed	175.1082	329.1940	428.2622	541.3474
deviation [ppm]	2.9	2.4	1.4	3.1
ion type	x1	x2	x3	x4
AA sequence	βPhS(-H <sub>2</sub> O)	L-V	V-cmp-T <sub>2</sub> -T <sub>1</sub> (-H <sub>2</sub> O)	L-V-Cmp-T <sub>2</sub> (-H <sub>2</sub> O)
[M] <sup>+</sup> calculated	146.0600	211.1441	438.2459	466.2772
[M] <sup>+</sup> observed	146.0607	211.1444	438.2470	466.2782
deviation [ppm]	4.8	1.4	2.5	2.1



**Figure S20:** MS/MS fragmentation spectrum of cyclofaulknamycin ([M+2H]<sup>2+</sup> peak)



**Figure S21:** Secondary metabolite profile of the *S. albus* del9 and del 10 strains. Butanol extraction.

III

**Bonsecamin: A New Cyclic Pentapeptide Discovered through  
Heterologous Expression of a Cryptic Gene Cluster**

Constanze Lasch, Marc Stierhof, Marta Rodríguez Estévez, Maksym Myronovskyi, Josef Zapp  
and Andriy Luzhetskyy

*Microorganisms* **2021**, *9*(8), 1640

DOI: 10.3390/microorganisms9081640

Published online 31<sup>st</sup> July 2021

## Article

# Bonsecamin: A New Cyclic Pentapeptide Discovered through Heterologous Expression of a Cryptic Gene Cluster

Constanze Lasch <sup>1</sup>, Marc Stierhof <sup>1</sup>, Marta Rodríguez Estévez <sup>1</sup>, Maksym Myronovskiy <sup>1</sup>, Josef Zapp <sup>2</sup>  
and Andriy Luzhetskyy <sup>1,3,\*</sup>

<sup>1</sup> Department of Pharmaceutical Biotechnology, Saarland University, 66123 Saarbruecken, Germany; constanze.lasch@uni-saarland.de (C.L.); m.stierhof@t-online.de (M.S.);

marta.rodriguezestevez@uni-saarland.de (M.R.E.); maksym.myronovskiy@uni-saarland.de (M.M.)

<sup>2</sup> Department of Pharmaceutical Biology, Saarland University, 66123 Saarbruecken, Germany; j.zapp@mx.uni-saarland.de

<sup>3</sup> Helmholtz Institute for Pharmaceutical Research Saarland, 66123 Saarbruecken, Germany

\* Correspondence: a.luzhetskyy@mx.uni-saarland.de; Tel.: +49-681-302-70200

**Abstract:** The intriguing structural complexity of molecules produced by natural organisms is uncontested. Natural scaffolds serve as an important basis for the development of molecules with broad applications, e.g., therapeutics or agrochemicals. Research in recent decades has demonstrated that by means of classic metabolite extraction from microbes only a small portion of natural products can be accessed. The use of genome mining and heterologous expression approaches represents a promising way to discover new natural compounds. In this paper we report the discovery of a novel cyclic pentapeptide called bonsecamin through the heterologous expression of a cryptic NRPS gene cluster from *Streptomyces albus* ssp. *chlorinus* NRRL B-24108 in *Streptomyces albus* Del14. The new compound was successfully isolated and structurally characterized using NMR. The minimal set of genes required for bonsecamin production was determined through bioinformatic analysis and gene deletion experiments. A biosynthetic route leading to the production of bonsecamin is proposed in this paper.

**Keywords:** *Streptomyces*; NRPS; heterologous expression; cyclic peptide



**Citation:** Lasch, C.; Stierhof, M.; Estévez, M.R.; Myronovskiy, M.; Zapp, J.; Luzhetskyy, A. Bonsecamin: A New Cyclic Pentapeptide Discovered through Heterologous Expression of a Cryptic Gene Cluster. *Microorganisms* **2021**, *9*, 1640. <https://doi.org/10.3390/microorganisms9081640>

Academic Editors:  
Stéphane Cociancich and  
Valérie Leclère

Received: 1 July 2021  
Accepted: 28 July 2021  
Published: 31 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

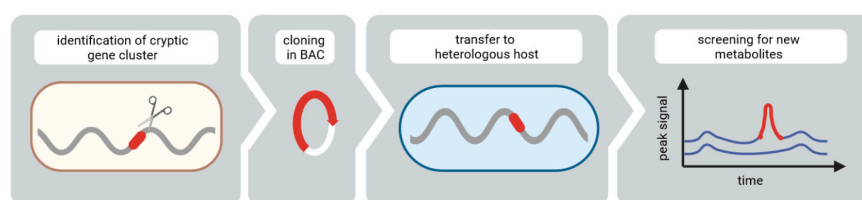
In the last two decades a considerable number of new small cyclic natural peptides produced by the bacterial genus of *Streptomyces* were discovered and published [1–8]. In nature, those molecules often have either toxic functions or serve the producing organism to coordinate the uptake of metal ions acting as a chelator [9,10]. Several naturally occurring cyclic peptides have proven their potential for use in pharmacotherapy. Remarkable antitumor and antibacterial activity in the recently discovered cyclic peptides chloptosin, hytramycins and mannopeptimycins, was reported by researchers [5,7,11]. Other compounds such as the antitumor agents actinomycin D and romidepsin, or immunosuppressant cyclosporine A, are already established as marketed drugs [12–15].

Cyclic peptides are often the product of enzymes belonging to the class of nonribosomal peptide synthetases (NRPS). NRPS are large modular enzymes or enzymatic complexes with each module responsible for the incorporation of a single amino acid residue into a nascent peptide chain. The individual modules can be split into separate domains. The typical elongation module contains a minimal set of condensation (C), adenylation (A) and peptidyl carrier (PCP) domains, where the A domain catalyzes the selective activation of an amino acid, the PCP domain holds the activated amino acid or the growing peptide chain and the C domain catalyzes the formation of the amide bond between two PCP-bound amino acid substrates. NRPS-derived molecules can display uncommon structural features, e.g., by incorporation of non-proteinogenic amino acids or when the amino acids



are structurally modified by tailoring domains. The chemical diversity of NRPS products is further expanded by the existence of hybrid NRPS clusters [16].

Recently we reported the successful expression of several cryptic secondary metabolite gene clusters of *Streptomyces albus* ssp. *chlorinus* in the engineered heterologous host strain *Streptomyces albus* Del14. This led to the identification of the biosynthetic genes of the characterized bioactive compounds, albucidin and nybomycin, as well as to the discovery of the new natural products, benzanthric acid, fredericamycin C<sub>2</sub> and dudomycins [17–21]. In this paper we report the discovery of a new cyclic pentapeptide, bonsecamin, through the heterologous expression of a cryptic NRPS gene cluster of *S. albus* ssp. *chlorinus* in *S. albus* Del14 (Figure 1). Bonsecamin was purified and its structure was elucidated by NMR. The bioinformatic analysis of the putative bonsecamin biosynthetic cluster as well as targeted gene inactivation experiments allowed the determination of the minimal set of genes responsible for the biosynthesis of the compound. A biosynthetic scheme leading to the production of bonsecamin is proposed in this paper.



**Figure 1.** General workflow for compound discovery using heterologous expression.

## 2. Materials and Methods

### 2.1. General Procedures

Table S1 provides an overview of all bacterial strains, BACs and plasmids used in this work. Lysogeny broth (LB) medium was used for cultivation of *Escherichia coli* strains [22]. Soy flour mannitol (MS) agar [23] and tryptic soy broth (TSB; Sigma-Aldrich, St. Louis, MO, USA) were used to grow *Streptomyces* strains. Metabolites were extracted from either liquid DNPM medium (40 g/L dextrin, 7.5 g/L soytone, 5 g/L baking yeast, and 21 g/L MOPS, pH 6.8 as aqueous solution) or defined medium DM (mannitol 5 g/L, amino acid 0.5 g/L, K<sub>2</sub>PO<sub>4</sub> 0.5 g/L, MgSO<sub>4</sub> × 7 H<sub>2</sub>O 0.2 g/L, FeSO<sub>4</sub> × 7 H<sub>2</sub>O 0.01 g/L). Before inoculation of DM, the cells of the preculture were washed three times with amino-acid-free DM. Amino acids L-lys, L-val, L-al, D-thr and D-ser were supplied to the defined medium as required by the experimental design. The antibiotics kanamycin, apramycin, ampicillin and nalidixic acid were added when needed.

### 2.2. DNA Isolation and Manipulation

Standard protocols were used to carry out DNA manipulation, transformation into *E. coli* and intergeneric conjugation between *E. coli* and *Streptomyces* [22–24]. The BACMAX™ DNA purification kit (Lucigen, Middleton, WI, USA) was used to isolate BAC DNA from a constructed genomic library of *Streptomyces albus* ssp. *chlorinus* NRRL B-24108. Cluster borders were determined by deletion of several genes downstream or upstream from the expected gene cluster with deletion of genes 15 to 28 leading to minimal BAC 2018\_del1 and deletion of genes 1 to 7 leading to BAC 2018\_del2. The genes were replaced by the resistance marker ampicillin through homologous recombination using the RedET system [25]. For amplification of the respective gene cassettes from the plasmid pUC19 PCR reactions were performed. The PCR primer pairs 20200815\_1\_fw / 20200815\_1\_rev and 20200815\_2\_fw / 20200815\_2\_rev were constructed with overhang regions allowing the site-specific introduction of the cassettes at both sides of the expected bonsecamin gene cluster. The success of the recombination was assessed by restriction mapping and DNA sequencing. The restriction enzymes purchased from ThermoFisher Scientific (Waltham, MA, USA) or New England BioLabs NEB (Ipswich, MA, USA) were used according to the manual. The deletion of three single genes was carried out as described above. To avoid

polar effects, additional *PmeI* recognition sites in primers 20201217\_1\_fw / 20201217\_1\_rev, 20201217\_2\_fw / 20201217\_2\_rev and 20201217\_4\_fw / 20201217\_4\_rev allowed the precise removal of the ampicillin resistance gene and religation to BACs 2O18\_delKR\_delbla, 2O18\_delPCP\_delbla, and 2O18\_delTE2\_delbla.

### 2.3. Metabolite Extraction

*Streptomyces* strains were grown in 15 mL of TSB medium in a 100 mL baffled flask for 1 to 2 days. 1 mL of this preculture was used to inoculate 100 mL of production medium in a 500 mL baffled flask. Cultures were incubated for 6 to 7 days at 28 °C and 180 rpm in an Infors multitron shaker (Infors AG, Basel, Switzerland) for metabolite production. The metabolites were extracted from the culture supernatant with an equal amount of *n*-butanol, dried at up to 50 °C and stored at 4 °C.

### 2.4. Mass Spectrometry (MS) Analysis of Metabolites

Prior to MS analysis the extracts were dissolved in methanol. MS experiments were performed on a Dionex Ultimate 3000 UPLC system (ThermoFisher Scientific, Waltham, MA, USA) with PDA detector (stationary phase 100 mm ACQUITY UPLC BEH C18 1.7 µm column (Waters Corporation, Milford, MA, USA), mobile phase: linear gradient of [A] ddH<sub>2</sub>O + 0.1% formic acid / [B] acetonitrile + 0.1% formic acid, 5% to 95% at a flow rate of 0.6 mL/min). For mass detection the system was further coupled to either an amaZon speed (Bruker, Billerica, MA, USA) or LTQ Orbitrap XL mass spectrometer (ThermoFisher Scientific, Waltham, MA, USA) applying standard settings of positive ionization and a mass range detection of *m/z* 200 to 2000. The data were analyzed by the softwares Compass Data Analysis 4.1 (Bruker) or Xcalibur 3.0 (ThermoFisher Scientific).

### 2.5. Extract Purification

The crude extracts from the 10 L cultures were dissolved in methanol. Four purification steps were carried out when using DNPM medium for cultivation: (1) Normal Phase (NP) Flash chromatography on an Isolera One system (Biotage, Uppsala, Sweden); stationary phase: SNAP Ultra 50 g (Biotage, Uppsala, Sweden), mobile phase: [A] *n*-hexane / [B] chloroform / [C] ethyl acetate / [D] methanol in linear gradients of [A]/[B] 10 column volumes (CV), [B]/[C] 15 CV, [C]/[D] 15 CV at a flow of 100 mL/min) was followed by (2) Size Exclusion Chromatography (SEC; stationary phase: Sephadex-LH20; mobile phase: isocratic elution using 100% methanol). Two Reversed Phase (RP) chromatography steps followed on a (3) Waters HPLC system (2545 Binary Gradient module, Waters, Milford, MA, USA); stationary phase: Nucleodur C18 HTec 250/21 5 µm (Macherey-Nagel, Düren, Germany); mobile phase: linear gradient of [A] H<sub>2</sub>O + 0.1% formic acid / [B] methanol + 0.1% formic acid, 5% to 95% [B] for 17 min at flow rate of 20 mL/min, mass detection *m/z* 430 using software MassLynx (Waters)) and an (4) Agilent Infinity 1200 series HPLC system (stationary phase: Synergi 4 µm Fusion-RP 80 Å 250 × 10 (phenomenex, Torrance, CA, USA); mobile phase: linear gradient of [A] H<sub>2</sub>O + 0.1% formic acid / [B] acetonitrile + 0.1% formic acid, 10% to 50% [B] for 15.5 min at flow rate of 4 mL/min, detection UV 201 nm followed by fraction control on HPLC-MS) yielding 0.7 mg of pure substance. In between all purification steps the fractions containing bonsecamin were pooled, dried (at up to 50 °C) and redissolved in methanol.

### 2.6. Structure Elucidation by NMR Spectroscopy, Optical Rotation and Marfey's Method

NMR spectra were acquired on a Bruker Avance III 700 MHz spectrometer at 298 K equipped with a 5 mm TCI cryoprobe. The chemical shifts (δ) were reported in parts per million (ppm) relative to TMS. As solvents, deuterated DMSO-*d*<sub>6</sub> (δH 2.50 ppm, δC 39.51 ppm) from Deutero (Deutero, Kastellaun, Germany) were used. Edited-HSQC, HMBC, <sup>1</sup>H-<sup>1</sup>H COSY, and N-HSQC spectra were recorded using the standard pulse programs from the TOPSPIN v.3.6 software. Optical rotations were measured using a Perkin Elmer Polarimeter Model 241 (Perkin Elmer, Ueberlingen, Germany).

For Marfey's method, bonsecamin was hydrolyzed in 100  $\mu$ L 6 N HCl at 110  $^{\circ}$ C for 1 h. While cooling down, the sample was dried for 15 min under nitrogen, dissolved in 110 mL water and 50  $\mu$ L each were transferred into 1.5 mL Eppendorf tubes. To the hydrolysate, 20  $\mu$ L of 1N NaHCO<sub>3</sub> and 20  $\mu$ L of 1% L-FDLA or D-FDLA in acetone were added, respectively. The amino acid standards were prepared the same way using L-FDLA only. The reaction mixtures were incubated at 40 $^{\circ}$ C for 90 min at 700 rpm and subsequently quenched with 2N HCl to stop the reaction. The samples were diluted with 300  $\mu$ L ACN and 1  $\mu$ L was analyzed by maXis high-resolution LC-QTOF system using aqueous ACN with 0.1 vol% formic acid and an adjusted gradient of 5–10 vol% for 2 min, 10–25 vol% for 13 min, 25–50 vol% for 7 min and 50–95 vol% for 2 min. Detection was carried out at 340 nm.

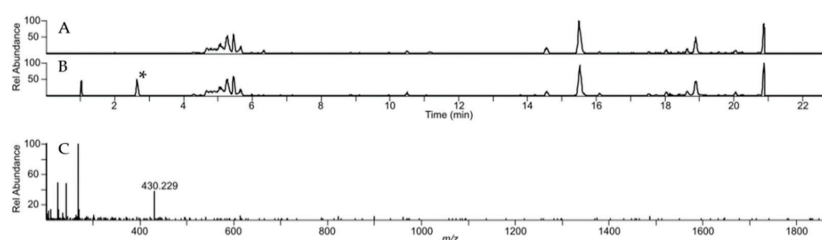
### 2.7. Genome Mining and Bioinformatic Analysis

The antiSMASH online tool was used to screen the genome of *S. albus* ssp. *chlorinus* (<https://antismash.secondarymetabolites.org/#!/start>, accessed on 10 July 2021) [26]. Analysis of the genetic data was performed by Geneious software, version 11.0.3 [27]. The genomic sequence of *Streptomyces albus* ssp. *chlorinus* can be accessed in GenBank under VJOK000000000. The Dictionary of Natural Products (DNP) 28.1 was used as reference database of so far characterized metabolites.

## 3. Results and Discussion

### 3.1. Identification and Expression of the NRPS Gene Cluster

Recently, the potential of the strain *Streptomyces albus* ssp. *chlorinus*, as a source of new and undiscovered biomolecules, was demonstrated [17–19]. Genome mining of this strain (GenBank accession number VJOK000000000) using antiSMASH software revealed a further uncharacterized biosynthetic gene cluster encoding a putative nonribosomal peptide synthetase (NRPS) [26]. A search in the previously constructed genomic BAC library of *S. albus* ssp. *chlorinus* uncovered a BAC clone 2O18 with the cloned DNA fragment covering the entire NRPS gene cluster. The BAC 2O18 was transferred into the heterologous host *Streptomyces albus* Del14 by conjugation [28]. The obtained exconjugant strain *S. albus* 2O18 as well as the control strain *S. albus* Del14 were cultivated in the production medium DNPM and secondary metabolites were extracted with *n*-butanol. The extracts were analyzed using high resolution LC-MS. The analysis of the peak profile of *S. albus* 2O18 revealed the presence of a new mass peak with an  $m/z$  of 430.229 eluting at hydrophilic conditions at the very front of the chromatogram (Figure 2). The identified peak was not detected in the extract of the reference strain without the BAC. The calculated molecular mass of 429.221 [ $\pm$ 5 ppm] of the identified compound was used for a search in the DNP database of natural products. This survey did not lead to any match with already registered metabolites from bacteria, implying that the structure of the molecule might be new. In order to gain insights into the structure of the identified compound, its purification for NMR analysis was carried out.



**Figure 2.** HPLC-MS analysis of bonsecamin production. A- and B-extracted ion chromatograms ( $430.5 \pm 0.5$  Da) of crude extracts of *S. albus* 2O18 and *S. albus* Del14, respectively. The new peak observed in the extract of *S. albus* 2O18 is marked with an asterisk (\*). C–Mass spectrum of the new peak observed in the extract of *S. albus* 2O18.

### 3.2. Purification and Structure Elucidation

To obtain the identified compound in an amount sufficient for structure elucidation, *S. albus* 2O18 was cultivated in 10L of DNPM medium. The metabolites were extracted with *n*-butanol. The compound was purified from the crude extract using normal-phase, size-exclusion and reversed-phase chromatography. Only a low submilligram amount of pure compound was obtained as a white powder and physically characterized. The optical rotation was determined as  $[\alpha]_D^{20} -13.5$  (c 0.12, MeOH) and the measured  $\lambda_{\max}$  (log  $\epsilon$ ) of the compound was 198 nm (2.11) in 11% ACN/H<sub>2</sub>O + 0.1% formic acid. The dried isolate proved stable for > 1 year when stored at  $-20^\circ\text{C}$  and did not show signs of degradation when dissolved in organic solvents (methanol/DMSO). It was further used for structure elucidation by NMR experiments, MS/MS fragmentation and FDLA derivatization (Table 1, Figure S1–S8).

**Table 1.** NMR data of bonsecamin in DMSO-*d*<sub>6</sub>.

Pos.		$\delta\text{C}/\delta\text{N}$ Type	$\delta\text{H}$ , mult. (J in Hz)	COSY	Key HMBC (H-)
2,3-DABA	1	170.5 *, C			
	2	58.53, CH	3.96, dd (7.5, 2.5)	3, 5	1
	3	56.23, CH	3.23, m	2, 4	23
	4	13.41, CH <sub>3</sub>	0.52, d (6.3)	3	2
	5	122.4, NH	7.06, d (7.5)	2	1, 6
Val	6	169.6 *, C			
	7	58.40, CH	4.01, t (9.1)	8, 11	6, 12
	8	30.24, CH	1.96, m	7, 9, 10	
	9	18.39, CH <sub>3</sub>	0.83, d (7.0)	8	7
	10	19.44, CH <sub>3</sub>	0.84, d (7.0)	8	7
	11	114.5, NH	7.84, d (9.1)	7	12
Ala	12	172.2 *, C			
	13	50.67, CH	4.21, dq (8.6, 7.5)	14, 15	12, 16
	14	17.46, CH <sub>3</sub>	1.32, d (7.5)	13	12
	15	120.5, NH	8.14, d (8.6)	13	16
Ala	16	169.8 *, C			
	17	50.56, CH	4.08, quin (7.2)	18, 19	16, 20
	18	17.23, CH <sub>3</sub>	1.24, d (7.2)	17	16
	19	121.7, NH	7.87, d (7.2)	17	16
Ser	20	172.9 *, C			
	21	62.38, CH	2.80, m	22, 24	3, 20
	22	61.04, CH <sub>2</sub>	3.44, m	21, 23	
			3.57, m		
	23	—, OH	5.01, m	22	
	24	48.3, NH	2.31, d (5.5)	21	2, 4, 20, 21, 22

\* Not visible in the <sup>13</sup>C NMR spectrum. The value was taken from the HMBC.

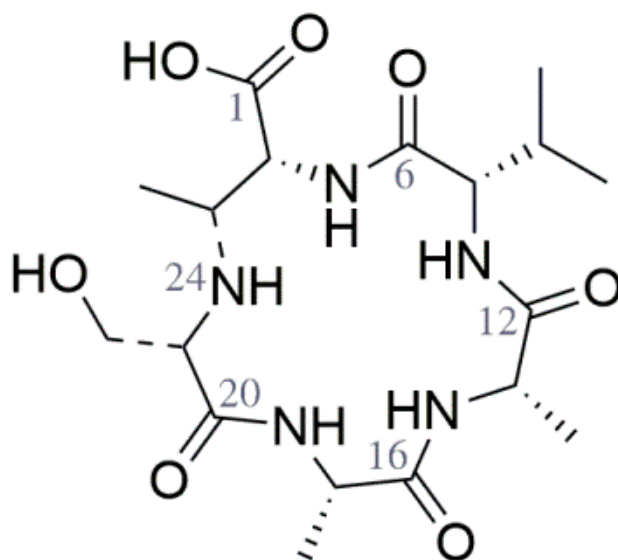
The molecular formula was calculated as C<sub>18</sub>H<sub>31</sub>N<sub>5</sub>O<sub>7</sub> with 6 degrees of unsaturation corresponding to the monoisotopic mass of 429.222 Da. The analysis of <sup>1</sup>H NMR in DMSO-*d*<sub>6</sub> revealed four doublet NH signals at  $\delta\text{H}$  7.06, 7.84, 7.87 and 8.14, indicating four peptide bonds. The measurement of <sup>15</sup>N-HSQC confirmed this assumption by showing correlations to  $\delta\text{N}$  114.5, 120.5, 121.7 and 122.4, and revealed an additional NH group at  $\delta\text{H}$  2.31 and  $\delta\text{N}$  48.3 suggesting a secondary amine (Figure S6). Analysis of <sup>1</sup>H and <sup>13</sup>C NMR, <sup>1</sup>H-<sup>1</sup>H COSY and edited-HSQC revealed five amino acids corresponding to valine, two alanines, serine and threonine. The peptide was assigned by long-range HMBC correlations leading to the sequence Ser-Ala-Ala-Val-Thr. The remaining degree of unsaturation indicated a cyclic structure with a ring closure between serine and threonine via the aforementioned amine. This could be concluded from the altered chemical shifts of serine CH- $\alpha$  ( $\delta\text{C}$  62.4,  $\delta\text{H}$  2.80) and threonine CH- $\beta$  ( $\delta\text{C}$  56.2,  $\delta\text{H}$  3.23). A final proof was provided by key correlations in the <sup>1</sup>H-<sup>1</sup>H COSY and HMBC spectra. The amine proton at  $\delta\text{H}$  2.31 was determined by a COSY correlation adjacent to CH- $\alpha$  of serine ( $\delta\text{H}$  2.80) and showed strong HMBC correlations to CH- $\alpha$  and CH<sub>3</sub> of threonine, but not to the carboxyl group. In addition, a

COSY correlation of the hydroxyl group of serine ( $\delta H$  5.01) to  $CH_2 \beta$  ( $\delta H$  3.44 and 3.57) was observed, ruling out ring closure by an ester or ether.

In summary, ring closure between serine and  $CH-\alpha$  of threonine was established. This will henceforth be referred to as 2, 3-diaminobutanoic acid (DABA). The missing COSY correlation between  $CH-\beta$  of DABA and the amine could be due to a dihedral angle of  $90^\circ$  between  $\delta H$  2.31 (NH, Ser) and  $\delta H$  3.23 ( $CH \beta$ , DABA) resulting in a very small  $3J$  coupling constant. The structure was confirmed by MS/MS fragmentation by showing a- and x-ions patterns commonly observed for cyclic structures (Figure S8). The spectral data of bonsecamin are shown in Table 1.

The absolute configuration was determined by Marfey's method. The peptide was treated with 6N hydrochloric acid at  $110^\circ C$  for 1 h. The hydrolysate and the amino acid standards were derivatized with D- and L-FDLA and analyzed by LC-MS (Figure S7). Alanine and valine were determined in an L-configuration and showed the expected ratio of 2:1.

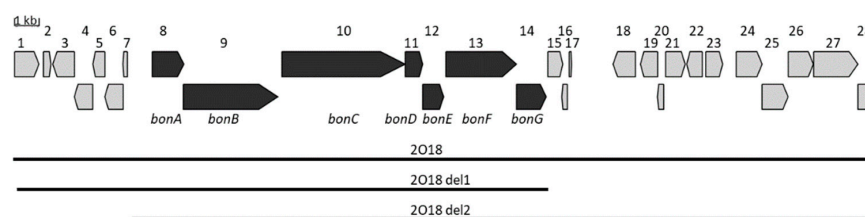
Serine was probably converted to 2,3-hydroxybutanoic acid and did not react with L-FDLA. Thus, its configuration could not be determined. The configuration of DABA was elucidated by the relative method using derivatization with D- and L-FDLA. When derivatized with D-FDLA, DABA showed a shorter retention time compared to L-FDLA, thus it had D-configuration. The absolute structure of bonsecamin is shown in Figure 3.



**Figure 3.** The structure of isolated bonsecamin.

### 3.3. Determination of the Minimal Gene Cluster

A sequence analysis of the 35 kb DNA fragment cloned in the BAC 2018 revealed 28 putative genes (Figure 4). AntiSMASH analysis did not predict any significant homology between the expressed cluster and any other already characterized gene cluster. However, this analysis revealed that a DNA sequence highly similar to genes 8 to 14 is present in a number of *Streptomyces* strains. This implies indirectly that genes 8 to 14 cloned in the BAC 2018 might correspond to the bonsecamin biosynthetic cluster. Genes 8 to 14 are further assigned as *bonA* to *bonG*. The genes *bonB*, *bonC* and *bonF* encode elements of an NRPS (Table S2). The gene *bonG* shows similarity to an amino acid ligase. The genes *bonD*, *bonE* and *bonA* encode two putative oxidoreductases and a transporter protein, respectively (Table S2).



**Figure 4.** Fragment of the *Streptomyces albus* ssp. *chlorinus* NRRL B-24108 chromosome cloned in BAC 2O18. The genes putatively involved in bonsecamin biosynthesis are highlighted in dark grey. The black bars indicate the chromosomal fragments cloned in BACs 2O18, 2O18\_del1 and 2O18\_del2 [29].

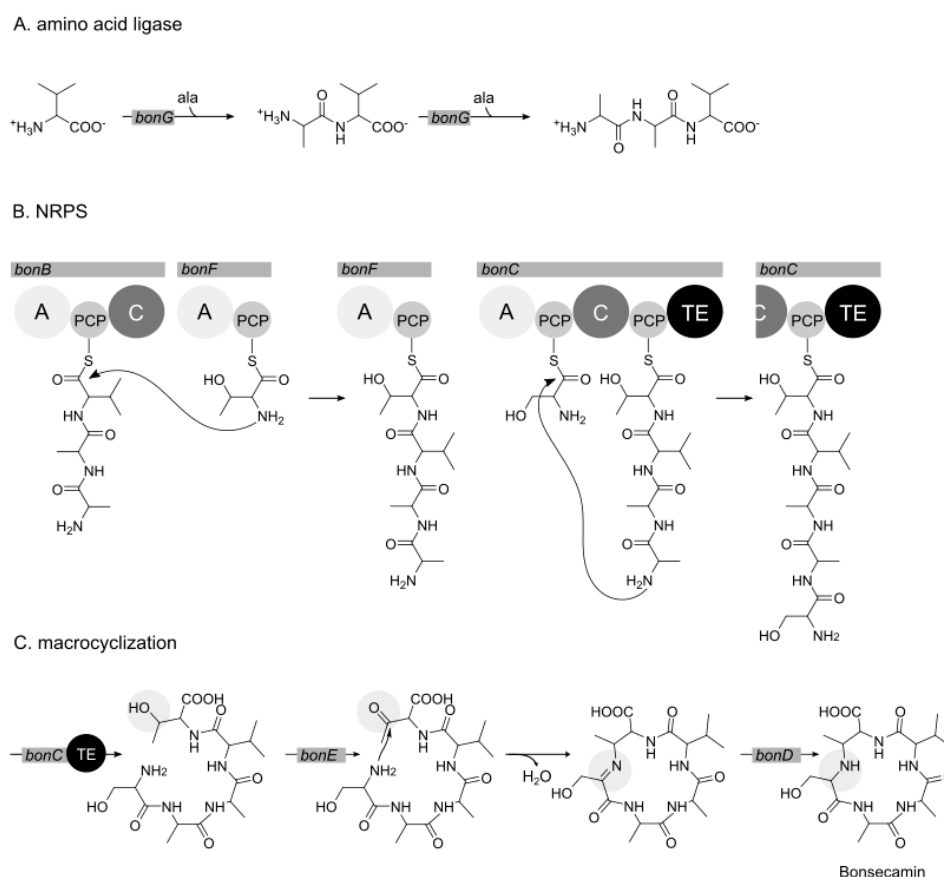
To prove the minimal set of genes responsible for bonsecamin synthesis, a set of gene deletion experiments were performed. The genes 15 to 28 which were predicted not to be involved in bonsecamin biosynthesis were deleted in the BAC 2O18 using RedET. The constructed recombinant BAC 2O18\_del1 was transferred into the host strain *S. albus* Del14 and the obtained strain *S. albus* 2O18\_del1 was tested for bonsecamin production. The results of the HPLC-MS analysis demonstrated that the production of bonsecamin was not affected in the *S. albus* 2O18\_del1 strain (Figure S9). This indicates that the genes 15–28 downstream of the *bonG* gene were not involved in bonsecamin production. Since *bonG* encoded a putative amino acid ligase and built an operon with the NRPS gene *bonF*, *bonG* was regarded as the last gene of the biosynthetic cluster (Figure 4). In order to determine the left border of the cluster, genes 1–7 upstream of the *bonA* gene were deleted in BAC 2O18. The constructed BAC 2O18\_del2 was transferred into the heterologous host *S. albus* Del14 and the obtained exconjugant strain was analyzed for bonsecamin production. No difference in bonsecamin production between the strains *S. albus* 2O18 and *S. albus* 2O18\_del2 was detected (Figure S9). This indicated that the deleted genes 1 to 7 were not involved in the production of bonsecamin. The gene *bonA* encoding a putative transporter was regarded as the first gene involved in bonsecamin production and *bonA* belongs to the same operon as the NRPS-encoding *bonB*, which further supports the assumption that the *bonA* gene constitutes the left border of the cluster. In general, our data indicated that the bonsecamin biosynthetic cluster encompassed the genes *bonA* to *bonG*.

### 3.4. Biosynthesis of Bonsecamin

Bonsecamin is a novel cyclic pentapeptide. The structure of bonsecamin implies that the compound might be synthesized through a linear peptide precursor–Ser-Ala-Ala-Val-Thr. The amino acid residues within the precursor are linked via conventional amide bonds. To generate the mature bonsecamin the linear precursor peptide likely undergoes intramolecular dehydrative cyclization. During this modification step the side chain of the threonine was joined with the free amino group of the serine residue.

The predicted minimal gene cluster for bonsecamin production consists of seven open reading frames from *bonA* to *bonG*. The *bonA* gene encodes a putative transporter. The genes *bonD* and *bonE* encode putative dehydrogenases and are presumably involved in the tailoring steps of the bonsecamin biosynthesis. The remaining four genes *bonB*, *bonC*, *bonF* and *bonG*, which encode putative elements of NRPS and an alanine ligase, might be involved in the biosynthesis of the linear bonsecamin precursor. The analysis of the NRPS genes revealed that *bonF* encodes A and PCP domains; *bonB* encodes A, PCP and C domains; and *bonC* encodes A, PCP, C and TE domains (Figure 5). This domain organization indicates that only three of five amino acid residues of bonsecamin were incorporated by the encoded NRPS. The A domains encoded by *bonC*, *bonB* and *bonF* were predicted to have substrate specificity towards the amino acids serine, valine and threonine, respectively. This prediction is in accordance with the amino acid composition of bonsecamin: Ser-Ala-Ala-Val-Thr. Two alanine residues are therefore expected to be incorporated into the bonsecamin precursor in an NRPS-independent manner. The putative alanine ligase encoded by the *bonG* gene might be responsible for the alanine incorporation.





**Figure 5.** Proposed biosynthetic scheme for bonsecamin production. A. Formation of the Ala-Ala-Val tripeptide precursor catalyzed by the putative alanine ligase encoded by *bonG*. B. Conversion of the tripeptide intermediate into the linear pentapeptide precursor catalyzed by the NRPS encoded by *bonB*, *bonF* and *bonC*. C. Cyclization of the linear bonsecamin precursor catalyzed by the products of *bonE* and *bonD*.

Taking into consideration the structure of the compound, domain organization and the predicted substrate specificity of the NRPS, the following bonsecamin biosynthetic scheme is proposed. The first step in the biosynthesis is the assembly of the Ala-Ala-Val tripeptide catalyzed by a putative alanine ligase encoded by *bonG*. The enzyme uses valine as a starter substrate and carries out two rounds of alanine ligation. The participation of amino acid ligases in the biosynthesis of natural products was reported before [30]. Amino acid ligases usually attach small, non-polar amino acids to the amino group of the substrate by forming an amide bond. The enzymes are characterized as being quite specific for their extension unit (here alanine), but exhibit only little substrate specificity for the starter substrate, which is extended [31]. The relaxed substrate specificity might explain the attachment of the alanine residue to the valine in the first elongation step and then to the Ala-Val dipeptide in the second elongation step.

The conversion of the synthesized Ala-Ala-Val tripeptide into the linear bonsecamin precursor is proposed to be catalyzed by the NRPS genes encoded within the cluster. The attachment of the threonine residue to the tripeptide is supposedly catalyzed by the products of *bonB* and *bonF*, which together encode the starter module and the first elongation module of the bonsecamin NRPS. The tripeptide is probably activated by the A domain of *bonB*, presumably specific for valine, and loaded on the corresponding PCP domain (Figure 5). The activated tripeptide is then elongated with threonine which is activated by the A domain of *bonF* and bound to its PCP domain, leading to the Ala-Ala-Val-Thr tetrapeptide (Figure 5).

To finalize the assembly of the linear pentapeptide precursor, an amide bond between the amino group of the serine and the carboxyl group of the alanine within the tetrapeptide product needs to be formed. We propose that the product of *bonC* catalyzes this reaction. *BonC* contains a canonical starter module with A and PCP domains and a shortened termination module with C, PCP and TE domains. The termination module of *bonC* lacks an A domain. The A domain of the starter module is predicted to activate serine and to load it on the corresponding PCP domain. To elongate the serine residue with the tetrapeptide the latter needs to be loaded on the PCP domain of the termination module of *bonC*. Since the termination module of *bonC* does not contain any A domain, we propose that the tetrapeptide is transferred from the PCP domain of *bonF* on the PCP domain of the termination module of *bonC*. The mechanism of this assumed transesterification step remains elusive. The condensation of the serine residue with the tetrapeptide to yield the linear pentapeptide Ser-Ala-Ala-Val-Thr is catalyzed by the C domain of *bonC*. The linear pentapeptide precursor is proposed to be released through the action of a dedicated TE domain within the termination module of *bonC* (Figure 5).

The released linear pentapeptide precursor needs to be cyclized to yield mature bonsecamin. The cyclization of bonsecamin occurs not through conventional amide bonds but through the secondary amine, which is formed between the side chain of the threonine and the amino group of the serine. We suppose that the conversion of the linear precursor into bonsecamin is catalyzed by the standalone enzymes encoded by *bonD* and *bonE*. Both enzymes were annotated to possess redox function. We assume that the first step towards cyclization is the oxidation of the hydroxyl group at the  $\beta$  carbon of the threonine, leading to a keto group with reactivity superior to the carboxylic acid carbon. This step might be catalyzed by the product of *bonE* which shows sequence similarity to 3-hydroxybutyrate dehydrogenase and 3-ketoacyl-ACP reductase—enzymes which catalyze similar reactions. Then the amino group of the serine residue reacts with the generated keto group under release of a water molecule to give an imine intermediate. Peptide macrocyclizations leading to imines are unusual, but have been described before [32–34]. The process of cyclization seems to happen spontaneously [32]. The generated imine group within the cyclized precursor is further reduced to a secondary amine leading to mature bonsecamin. We suppose that the final reduction step is catalyzed by the product of *bonD* which shows homology to the product of *lgrE*, involved in the reduction of linear gramicidin precursors [35].

The proposed scheme for bonsecamin biosynthesis includes a number of unusual enzymatic steps. In order to validate the proposed biosynthetic pathway, the inactivation of the *bonE* gene as well as the deletion of the second PCP domain or TE domain within *bonC* were undertaken and led to complete cessation of bonsecamin production (Figure S9). No putative derivatives or precursors of the compound could be detected in the culture broth of the engineered strains.

#### 4. Conclusions

In this article we reported the identification of the new compound bonsecamin after successful heterologous expression of a cryptic NRPS cluster of *S. albus* ssp. *chlorinus* NRRL B-24108. Bonsecamin is a cyclic pentapeptide with a secondary amine moiety formed between the side chain of a threonine and the amino group of a serine. The identified minimal set of biosynthetic genes indicated that bonsecamin was a result of the interplay between an amino acid ligase and a nonlinear NRPS.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/microorganisms9081640/s1>: Table S1: Strains, BACs, plasmids and primers used in this work; Table S2: Putative products of the genes in the DNA fragment encoding bonsecamin production; Figure S1:  $^1\text{H}$  NMR spectrum (700 MHz, DMSO- $d_6$ ) of bonsecamin. Figure S2:  $^{13}\text{C}$  NMR spectrum (700 MHz, DMSO- $d_6$ ) of bonsecamin. Figure S3:  $^1\text{H}$ - $^1\text{H}$  COSY spectrum (700 MHz, DMSO- $d_6$ ) of bonsecamin. Figure S4: Edited-HSQC spectrum (700 MHz, DMSO- $d_6$ ) of bonsecamin. Figure S5: HMBC spectrum (700 MHz, DMSO- $d_6$ ) of bonsecamin. Figure S6:  $^{15}\text{N}$ -HSQC spectrum (700 MHz, DMSO- $d_6$ ) of bonsecamin. Figure S7: LC-MS chromatograms of hydrolyzed bonsecamin derivatized



with D- or L-FDLA and the amino acid (aa) references derivatized with L-FDLA. Figure S8: MS/MS fragmentation of bonsecamin. Figure S9: Production of bonsecamin in *S. albus* Del14 mutant after gene deletion experiments.

**Author Contributions:** Bonsecamin was first identified in a mass chromatogram by M.R.E. All experiments except from NMR, polarimetry and Marfey's method were set up and evaluated by C.L., M.M. and A.L. and the practical work was performed by C.L. NMR experiments, polarimetry and Marfey's method were designed, carried out and the data evaluated by M.S. The final data were reviewed by J.Z. The manuscript was drafted by C.L., M.S. and M.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research has received funding from BMBF under Grant "EXPLOMARE" 031B0868A.

**Data Availability Statement:** The sequenced genome of *S. albus* ssp. *chlorinus* is available at GenBank under accession number VJOK00000000.

**Acknowledgments:** The authors thank Helmholtz Institute for Pharmaceutical Research Saarland, Saarbruecken, Germany (HIPS) for their support of NMR measurements. The strain *S. albus* ssp. *chlorinus* NRRL B-24108 was obtained from BASF SE Ludwigshafen, Germany.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Mohammadipanah, F.; Matasyoh, J.; Hamed, J.; Klenk, H.-P.; Laatsch, H. Persipeptides A and B, two cyclic peptides from *Streptomyces* sp. UTM 1154. *Bioorg. Med. Chem.* **2012**, *20*, 335–339. [[CrossRef](#)] [[PubMed](#)]
2. Wang, X.; Shaaban, K.A.; Elshahawi, S.I.; Ponomareva, L.V.; Sunkara, M.; Copley, G.C.; Hower, J.C.; Morris, A.J.; Kharel, M.K.; Thorson, J.S. Mullinamides A and B, new cyclopeptides produced by the Ruth Mullins coal mine fire isolate *Streptomyces* sp. RM-27-46. *J. Antibiot. (Tokyo)* **2014**, *67*, 571–575. [[CrossRef](#)] [[PubMed](#)]
3. Xiang, W.-S.; Wang, J.-D.; Wang, X.-J.; Zhang, J. Bingchamides A and B, two novel cyclic pentapeptides from the *Streptomyces* bingchenggensis: Fermentation, isolation, structure elucidation and biological properties. *J. Antibiot. (Tokyo)* **2009**, *62*, 501–505. [[CrossRef](#)]
4. Ji, Z.; Wei, S.; Fan, L.; Wu, W. Three novel cyclic hexapeptides from *Streptomyces alboflavus* 313 and their antibacterial activity. *Eur. J. Med. Chem.* **2012**, *50*, 296–303. [[CrossRef](#)]
5. Umezawa, K.; Ikeda, Y.; Uchihata, Y.; Naganawa, H.; Kondo, S. Chloptosin, an apoptosis-inducing dimeric cyclohexapeptide produced by *Streptomyces*. *J. Org. Chem.* **2000**, *65*, 459–463. [[CrossRef](#)]
6. Song, Y.; Li, Q.; Liu, X.; Chen, Y.; Zhang, Y.; Sun, A.; Zhang, W.; Zhang, J.; Ju, J. Cyclic Hexapeptides from the Deep South China Sea-Derived *Streptomyces scopuliridis* SCSIO ZJ46 Active Against Pathogenic Gram-Positive Bacteria. *J. Nat. Prod.* **2014**, *77*, 1937–1941. [[CrossRef](#)]
7. Cai, G.; Napolitano, J.G.; McAlpine, J.B.; Wang, Y.; Jaki, B.U.; Suh, J.-W.; Yang, S.H.; Lee, I.-A.; Franzblau, S.G.; Pauli, G.F.; et al. Hytramycins V and I, Anti-*Mycobacterium tuberculosis* Hexapeptides from a *Streptomyces hygroscopicus* Strain. *J. Nat. Prod.* **2013**, *76*, 2009–2018. [[CrossRef](#)]
8. He, H.; Williamson, R.T.; Shen, B.; Graziani, E.I.; Yang, H.Y.; Sakya, S.M.; Petersen, P.J.; Carter, G.T. Mannopeptimycins, Novel Antibacterial Glycopeptides from *Streptomyces hygroscopicus*, LL-AC98. *J. Am. Chem. Soc.* **2002**, *124*, 9729–9736. [[CrossRef](#)]
9. Dawson, R. The toxicology of microcystins. *Toxicon* **1998**, *36*, 953–962. [[CrossRef](#)]
10. Raymond, K.N.; Dertz, E.A.; Kim, S.S. Enterobactin: An archetype for microbial iron transport. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3584–3588. [[CrossRef](#)] [[PubMed](#)]
11. Singh, M.P.; Petersen, P.J.; Weiss, W.J.; Janso, J.E.; Luckman, S.W.; Lenoy, E.B.; Bradford, P.A.; Testa, R.T.; Greenstein, M. Mannopeptimycins, new cyclic glycopeptide antibiotics produced by *Streptomyces hygroscopicus* LL-AC98: Antibacterial and mechanistic activities, *Antimicrob. Agents Chemother.* **2003**, *47*, 62–69. [[CrossRef](#)] [[PubMed](#)]
12. Turan, T.; Karacay, O.; Tulunay, G.; Boran, N.; Koc, S.; Bozok, S.; Kose, M.F. Results with EMA/CO (etoposide, methotrexate, actinomycin D, cyclophosphamide, vincristine) chemotherapy in gestational trophoblastic neoplasia. *Int. J. Gynecol. Cancer* **2006**, *16*, 1432–1438. [[CrossRef](#)] [[PubMed](#)]
13. Ueda, H.; Manda, T.; Matsumoto, S.; Mukumoto, S.; Nishigaki, F.; Kawamura, I.; Shimomura, K. FR901228, a novel antitumor bicyclic depsipeptide produced by *chromobacterium violaceum* no. 968. III. Antitumor activities on experimental tumors in mice. *J. Antibiot. (Tokyo)* **1994**, *47*, 315–323. [[CrossRef](#)]
14. Borel, J.F.; Feurer, C.; Gubler, H.U.; Stähelin, H. Biological effects of cyclosporin A: A new antilymphocytic agent. *Agents Actions* **1976**, *6*, 468–475. [[CrossRef](#)]
15. Kolata, G. FDA Speeds Approval of Cyclosporin. *Science* **1983**, *221*, 1273. [[CrossRef](#)]
16. Sieber, S.A.; Marahiel, M.A. Learning from Nature's Drug Factories: Nonribosomal Synthesis of Macrocyclic Peptides. *J. Bacteriol.* **2003**, *185*, 7036–7043. [[CrossRef](#)]

17. Rodríguez Estévez, M.; Gummerlich, N.; Myronovskyi, M.; Zapp, J.; Luzhetskyy, A. Benzanthric Acid, a Novel Metabolite From *Streptomyces albus* Del14 Expressing the Nybomycin Gene Cluster. *Front. Chem.* **2020**, *7*, 896. [[CrossRef](#)] [[PubMed](#)]
18. Estévez, M.R.; Myronovskyi, M.; Rosenkränzer, B.; Paululat, T.; Petzke, L.; Ristau, J.; Luzhetskyy, A. Novel fredericamycin variant overproduced by a streptomycin-resistant *streptomyces albus* subsp. *Chlorinus Strain*. *Mar. Drugs* **2020**, *18*, 284. [[CrossRef](#)]
19. Lasch, C.; Stierhof, M.; Estévez, M.R.; Myronovskyi, M.; Zapp, J.; Luzhetskyy, A. Dudomycins: New Secondary Metabolites Produced after Heterologous Expression of an Nrps Cluster from *Streptomyces albus* ssp. *Chlorinus* Nr1 B-24108. *Microorganisms* **2020**, *8*, 1800. [[CrossRef](#)]
20. Myronovskyi, M.; Rosenkränzer, B.; Stierhof, M.; Petzke, L.; Seiser, T.; Luzhetskyy, A. Identification and Heterologous Expression of the Albucidin Gene Cluster from the Marine Strain *Streptomyces Albus* Subsp. *Chlorinus* NRRL B-24108. *Microorganisms* **2020**, *8*, 237. [[CrossRef](#)] [[PubMed](#)]
21. Rodríguez Estévez, M.; Myronovskyi, M.; Gummerlich, N.; Nadmid, S.; Luzhetskyy, A. Heterologous Expression of the Nybomycin Gene Cluster from the Marine Strain *Streptomyces albus* subsp. *chlorinus* NRRL B-24108. *Mar. Drugs* **2018**, *16*, 435. [[CrossRef](#)]
22. Green, M.R.; Sambrook, J. *Molecular Cloning: A Laboratory Manual*, 4th ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2012.
23. Kieser, T.; Bibb, M.J.; Buttner, M.J.; Chater, K.F.; Hopwood, D.A. *Practical Streptomyces Genetics. A Laboratory Manual.*; John Innes Foundation: Norwich, UK, 2000.
24. Rebets, Y.; Normanec, J.; Lutzetskyy, A.; Bernaerts, K.; Anné, J. Cloning and expression of metagenomic dna in *streptomyces lividans* and subsequent fermentation for optimized production. *Methods Mol. Biol.* **2017**, *1539*, 99–144.
25. Court, D.L.; Sawitzke, J.A.; Thomason, L.C. Genetic Engineering Using Homologous Recombination. *Annu. Rev. Genet.* **2002**, *36*, 361–388. [[CrossRef](#)]
26. Medema, M.H.; Blin, K.; Cimermancic, P.; De Jager, V.; Zakrzewski, P.; Fischbach, M.A.; Weber, T.; Takano, E.; Breitling, R. antiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **2011**, *39*, W339–W346. [[CrossRef](#)] [[PubMed](#)]
27. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [[CrossRef](#)] [[PubMed](#)]
28. Myronovskyi, M.; Rosenkränzer, B.; Nadmid, S.; Pujic, P.; Normand, P.; Luzhetskyy, A. Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. *Metab. Eng.* **2018**, *49*, 316–324. [[CrossRef](#)]
29. Harrison, K.J.; Crécy-Lagard, V. de; Zallot, R. Gene Graphics: A genomic neighborhood data visualization web application. *Bioinformatics* **2018**, *34*, 1406–1408. [[CrossRef](#)] [[PubMed](#)]
30. Parker, J.B.; Walsh, C.T. Action and Timing of BacC and BacD in the Late Stages of Biosynthesis of the Dipeptide Antibiotic Bacilysin. *Biochemistry* **2013**, *52*, 889–901. [[CrossRef](#)] [[PubMed](#)]
31. Antle, V.D.; Liu, D.; McKellar, B.R.; Caperelli, C.A.; Hua, M.; Vince, R. Substrate Specificity of Glycinamide Ribonucleotide Synthetase from Chicken Liver. *J. Biol. Chem.* **1996**, *271*, 8192–8195. [[CrossRef](#)] [[PubMed](#)]
32. Kopp, F.; Mahlert, C.; Grünwald, J.; Marahiel, M.A. Peptide Macrocyclization: The Reductase of the Nostocyclopeptide Synthetase Triggers the Self-Assembly of a Macrocyclic Imine. *J. Am. Chem. Soc.* **2006**, *128*, 16478–16479. [[CrossRef](#)]
33. Evans, B.S.; Ntai, I.; Chen, Y.; Robinson, S.J.; Kelleher, N.L. Proteomics-Based Discovery of Koranimine, a Cyclic Imine Natural Product. *J. Am. Chem. Soc.* **2011**, *133*, 7316–7319. [[CrossRef](#)] [[PubMed](#)]
34. Zipperer, A.; Konnerth, M.C.; Laux, C.; Berscheid, A.; Janek, D.; Weidenmaier, C.; Burian, M.; Schilling, N.A.; Slavetinsky, C.; Marschal, M.; et al. Human commensals producing a novel antibiotic impair pathogen colonization. *Nature* **2016**, *535*, 511–516. [[CrossRef](#)] [[PubMed](#)]
35. Schracke, N.; Linne, U.; Mahlert, C.; Marahiel, M.A. Synthesis of Linear Gramicidin Requires the Cooperation of Two Independent Reductases†. *Biochemistry* **2005**, *44*, 8507–8513. [[CrossRef](#)] [[PubMed](#)]



## Supplementary

# Bonsecamin: A New Cyclic Pentapeptide Discovered through Heterologous Expression of a Cryptic Gene Cluster

Constanze Lasch <sup>1</sup>, Marc Stierhof <sup>1</sup>, Marta Rodríguez Estévez <sup>1</sup>, Maksym Myronovskiy <sup>1</sup>, Josef Zapp <sup>2</sup> and Andriy Luzhetskyy <sup>1,3,\*</sup>

<sup>1</sup> Pharmaceutical Biotechnology, Saarland University, Saarbruecken, Germany; constanze.lasch@uni-saarland.de (C.L.), m.stierhof@t-online.de (M.S.), marta.rodriquezestevez@uni-saarland.de (M.R.), maksym.myronovskiy@uni-saarland.de (M.M.), a.luzhetskyy@mx.uni-saarland.de (A.L.)

<sup>2</sup> Pharmaceutical Biology, Saarland University, Saarbruecken, Germany; j.zapp@mx.uni-saarland.de (J.Z.)

<sup>3</sup> Helmholtz Institute for Pharmaceutical Research Saarland, Saarbruecken, Germany

\* Correspondence: a.luzhetskyy@mx.uni-saarland.de; +49 681 302 70200 (A.L.)

## Supplementary.

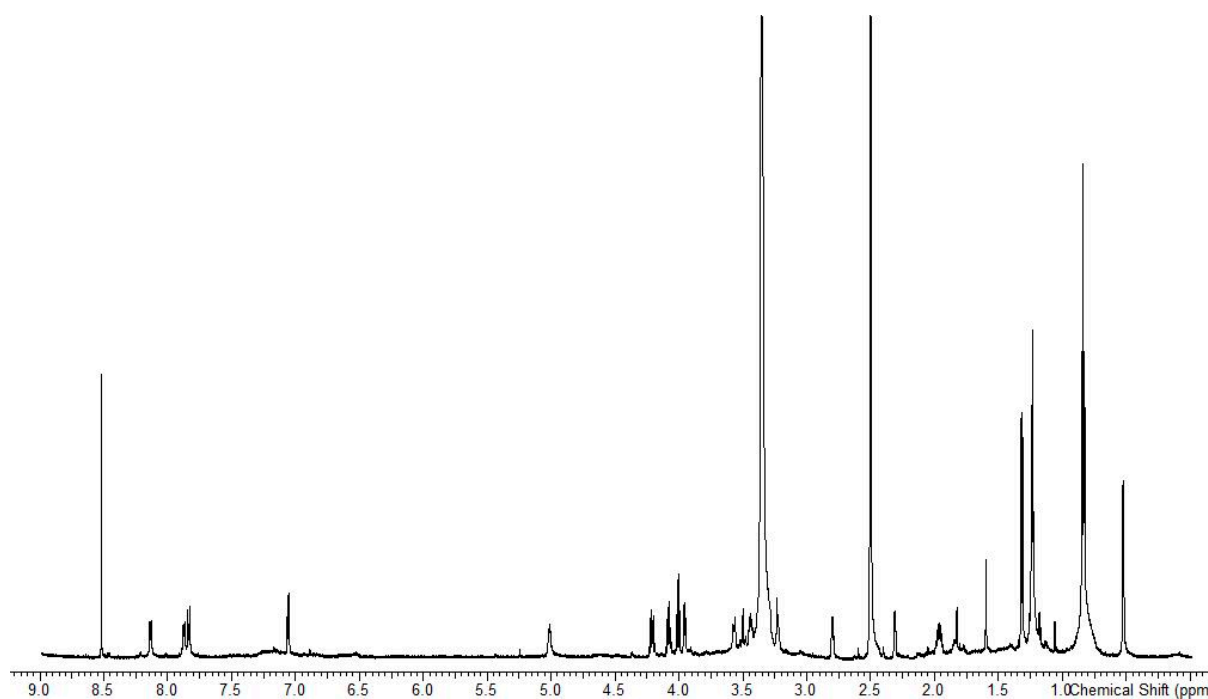
**Table S1.** Strains, BACs, plasmids and primers used in this work.

Material	Purpose
<b>A. Bacterial strains</b>	
<i>Streptomyces albus</i> Del14	heterologous host [1]
<i>Escherichia coli</i> GB05 RedCC	cloning host [Helmholtz-Institut für Pharmazeutische Forschung Saarland (HIPS)]
<i>Escherichia coli</i> ET12567 pUB307	alternate host intergeneric conjugation [2]
<b>B. BACs</b>	
2O18	heterologous expression of NRPS cluster
2O18_del1	determination downstream border of NRPS cluster
2O18_del2	determination upstream border of NRPS cluster
2O18_delKR_delbla	single gene inactivation
2O18_delPCP_delbla	single gene inactivation
2O18_delTE2_delbla	single gene inactivation
<b>C. Plasmids</b>	
pUC19	ampicillin resistance marker
<b>D. PCR primer Red/ET</b>	
20200815_1_fw [2O18_del1]	TAGTCCAGCGTCATCAGCGGGCGTCCGAGGCACTGCGGACCACGAGGCGCGTCAGGTGGCAC TTTTCG
20200815_1_rev [2O18_del1]	TCCGACGGCGGGCGGCCCGCACTAGGCTCGCCGCCATGACGGACGTCGACTTTTCTACGGGGT CTGAC
20210315_1_fw [2O18_del2]	CTATCGTCGCCACGCCTTGGTGACGGGAAATCCGGTGTGATGCCGGTGCCGTCAGGTGGCACT TTTTCG
20200815_2_rev [2O18_del2]	CACTGGATGCCCAGGCAGGGGGTACGCAGCATGACCGAGGAGGACGCGGCCTTTTCTACGGGG TCTGAC
20201217_1_fw [2O18_delKR_delbla]	GCTGGTGAACCCGCCGTCGACGGTGACCGTGAGCCGGTCACCTGGCGGGAGTTTAAACCGTC AGGTGGCACTTTTCG

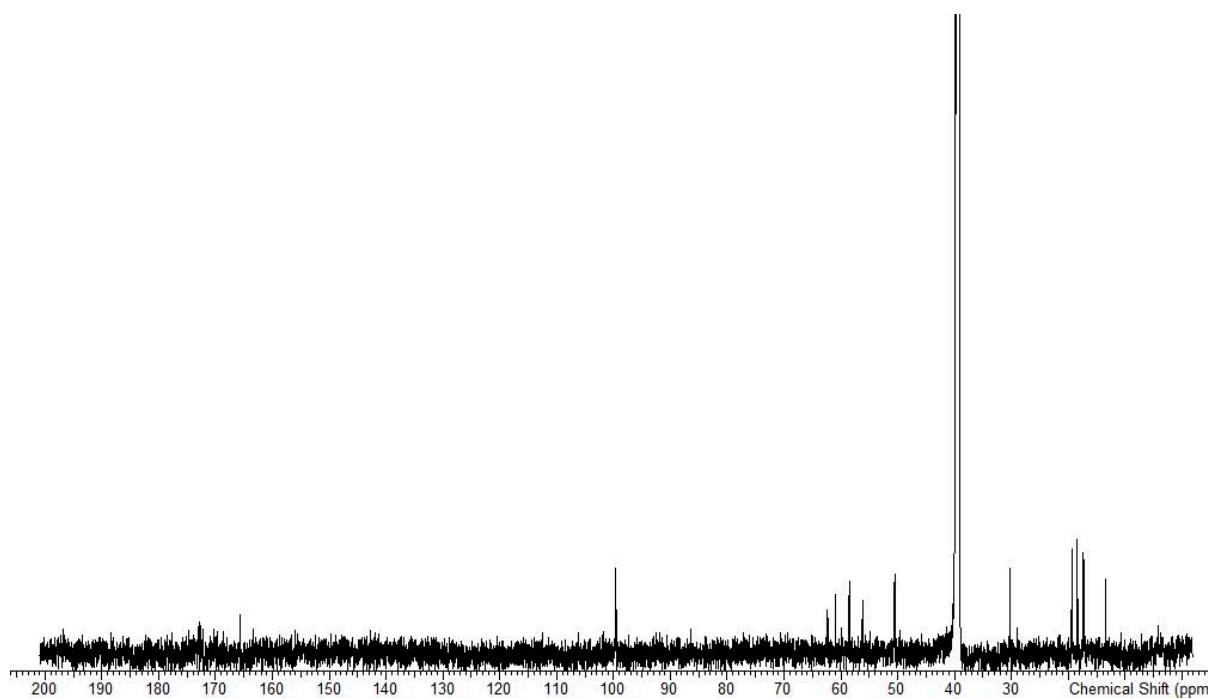
---

20201217_1_rev [2O18_delKR_delbla]	GGAGTGCTCACCGCGGGCGCCGCTCGCGGGCAAGGCCGCCGTCATCACGGGTTTAAACGACTTTT CTACGGGGTCTGAC
20201217_2_fw [2O18_delPCP_delbla]	GGCCAGGGCGGCCAGTTCGCCCAGCCGCGGGATGCGGGTGAGGTCGGTGAAGTTTAAACCGTC AGGTGGCACTTTTCG
20201217_2_rev [2O18_delPCP_delbla]	CGCGCGGTCTGGCAGCAGATCCTGGGGCTGACGGCGGAGGAGATCGGTGGTTTAAACGACTTT TCTACGGGGTCTGAC
20201217_4_fw [2O18_delTE2_delbla]	CAGTTCCGCGGTTCGCGGCCGGATTGCCGCGCACGAAGTAGTGGCCGCCCCGAGTTTAAACCGTC AGGTGGCACTTTTCG
20201217_4_rev [2O18_delTE2_delbla]	TGCGTGCCGTATCCGTGCGGGCACCCGGTCAACTTCAAACCGCTGGCCGGTTTAAACGACTTTT CTACGGGGTCTGAC

---



**Figure S1:**  $^1\text{H}$  NMR spectrum (700 MHz,  $\text{DMSO}-d_6$ ) of bonsecamin.



**Figure S2:**  $^{13}\text{C}$  NMR spectrum (700 MHz,  $\text{DMSO}-d_6$ ) of bonsecamin.

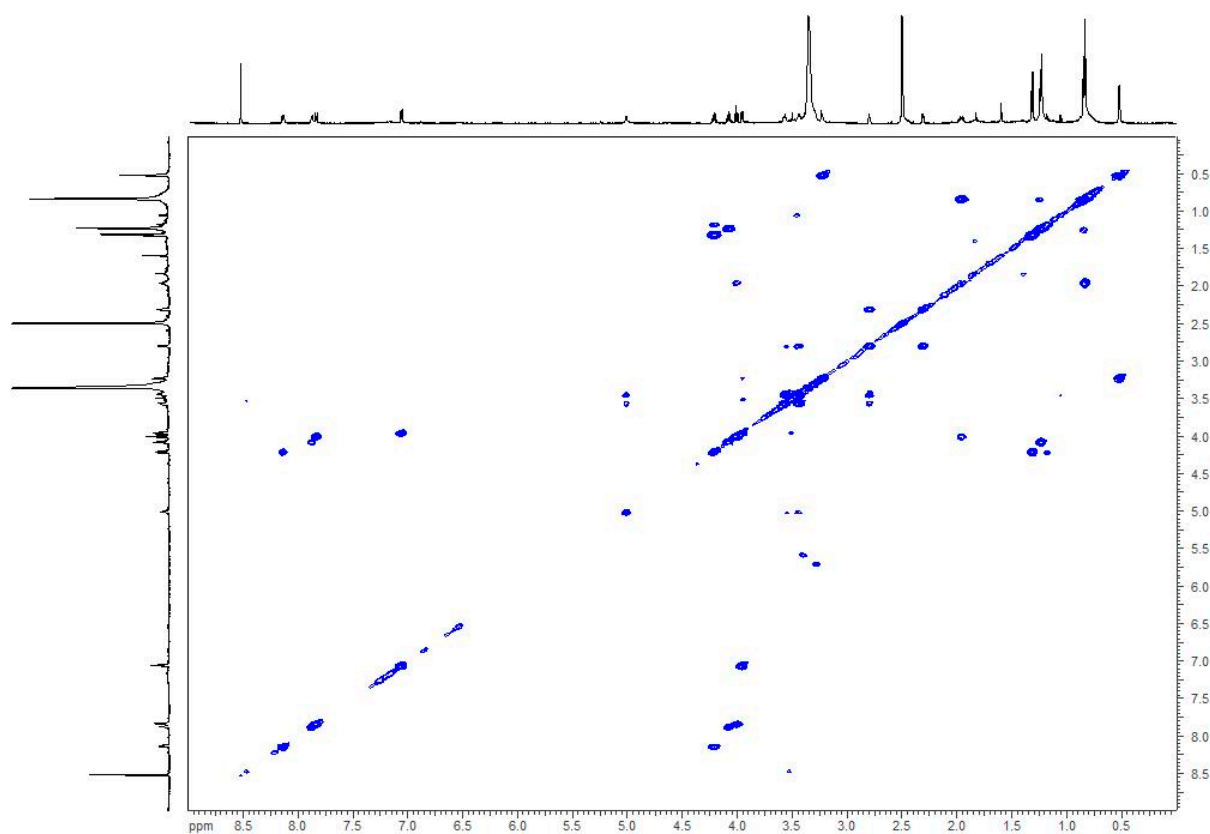


Figure S3:  $^1\text{H}$ - $^1\text{H}$  COSY spectrum (700 MHz,  $\text{DMSO-}d_6$ ) of bonsecamin.

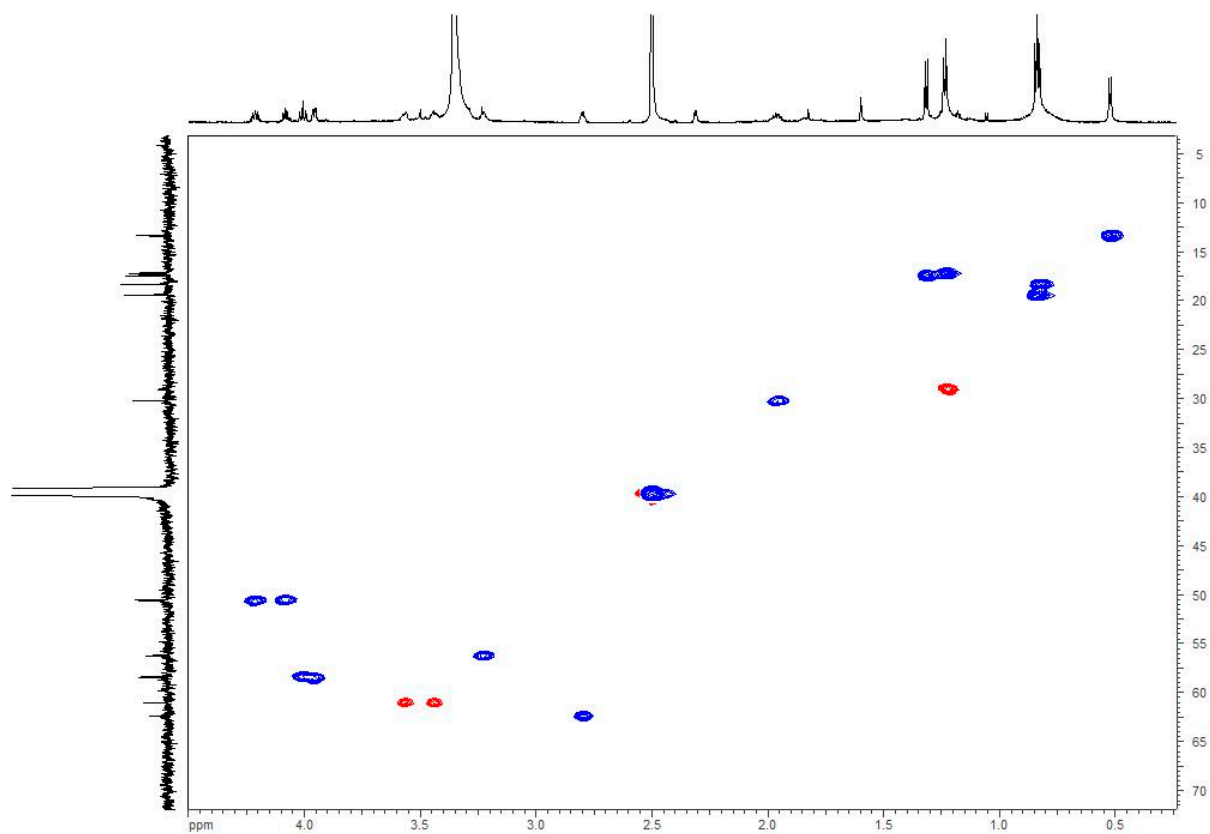


Figure S4: Edited-HSQC spectrum (700 MHz,  $\text{DMSO-}d_6$ ) of bonsecamin.

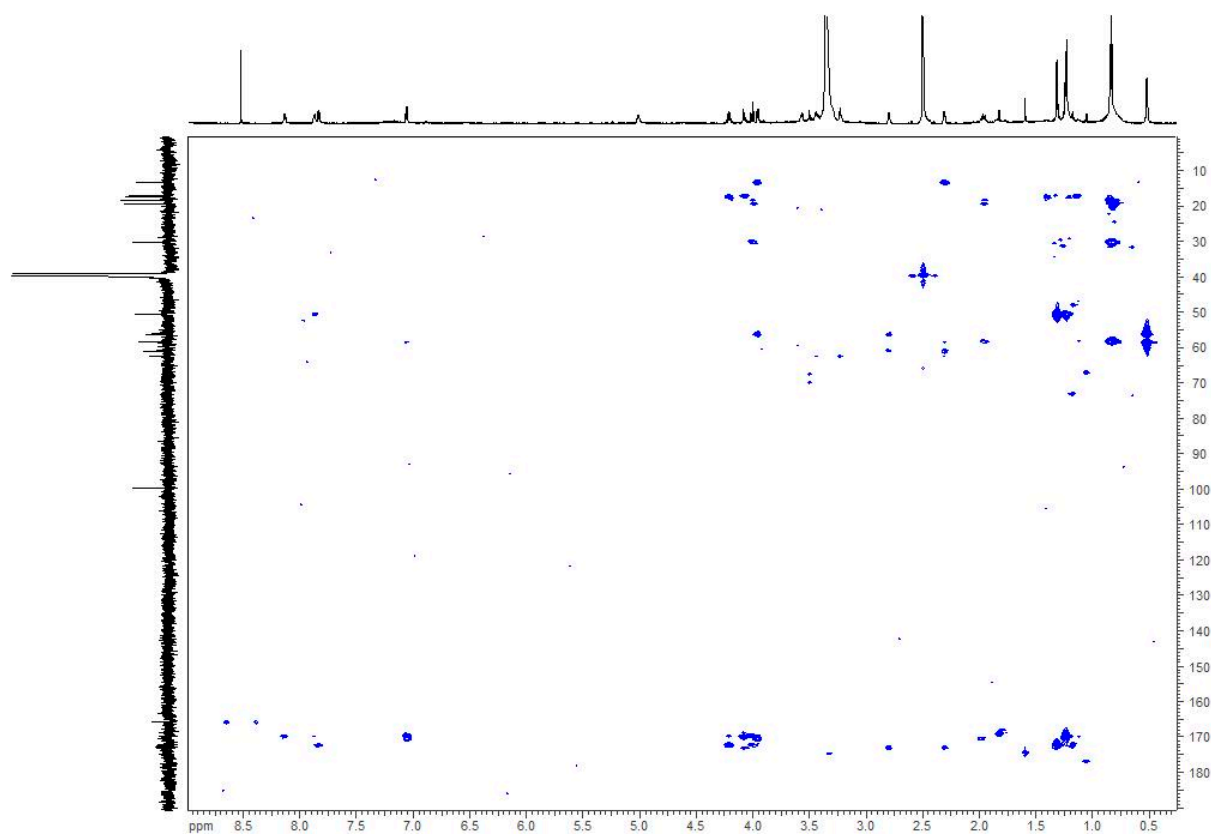


Figure S5: HMBC spectrum (700 MHz, DMSO-*d*<sub>6</sub>) of bonsecamin.

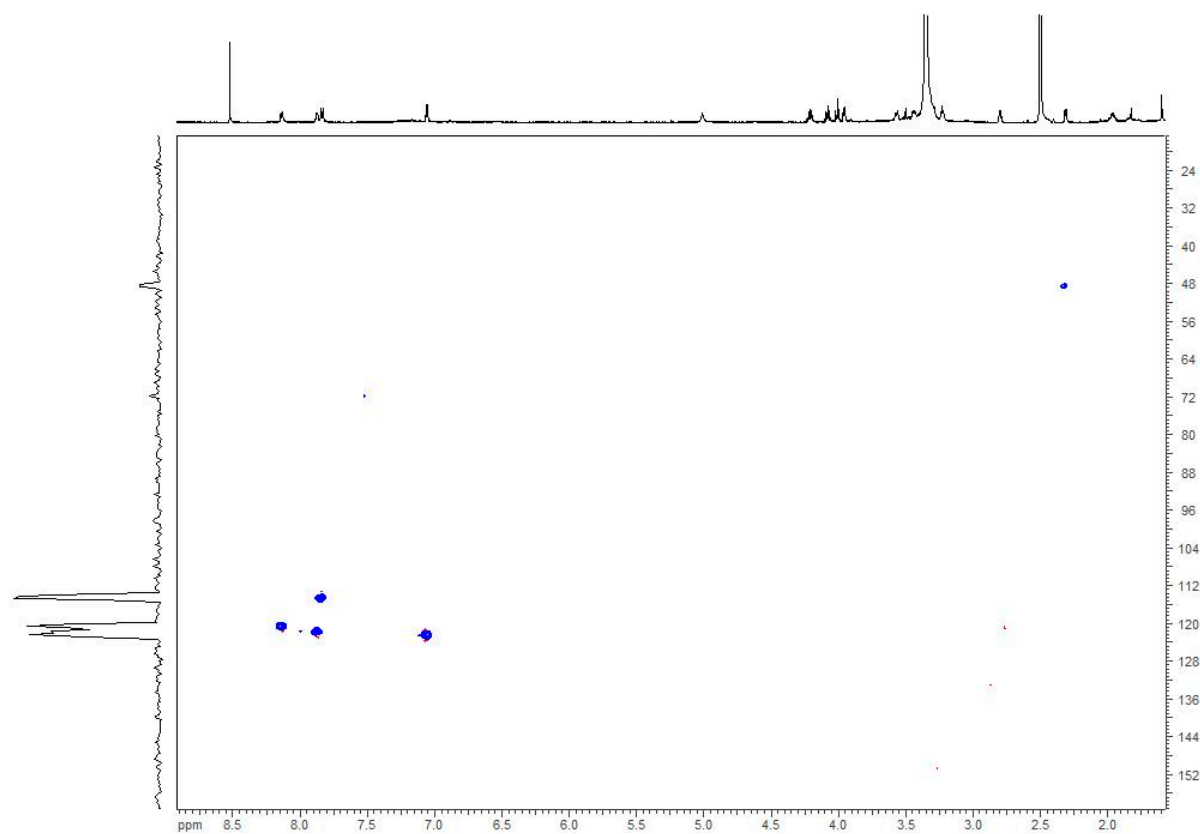
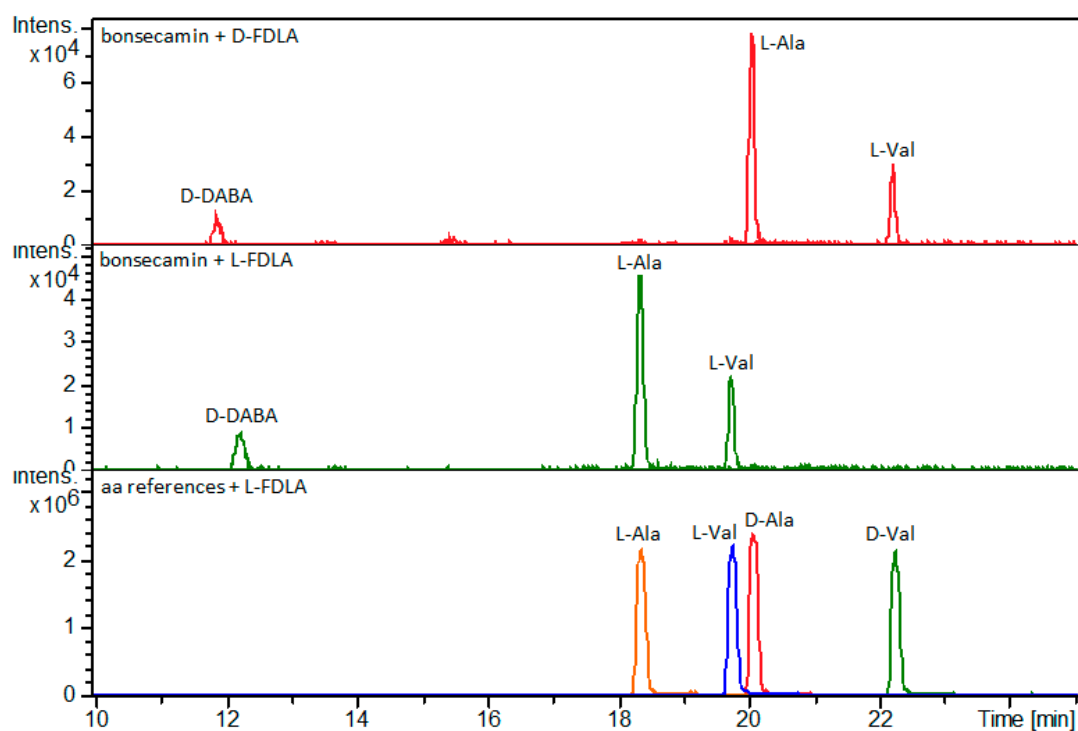
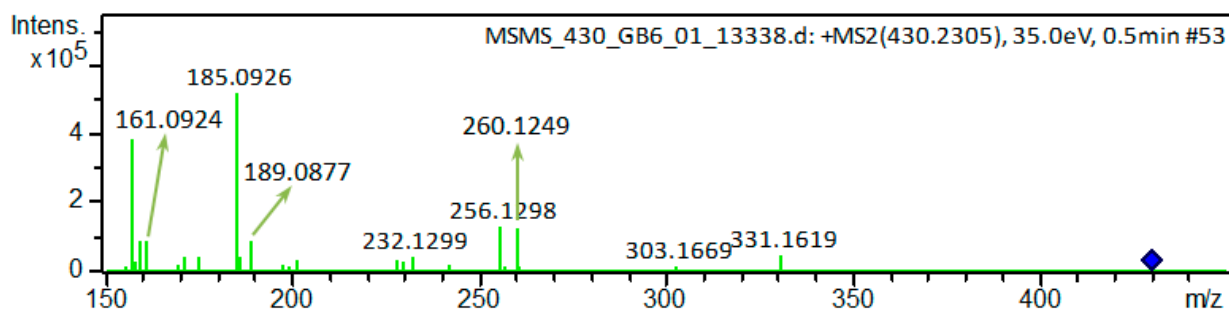
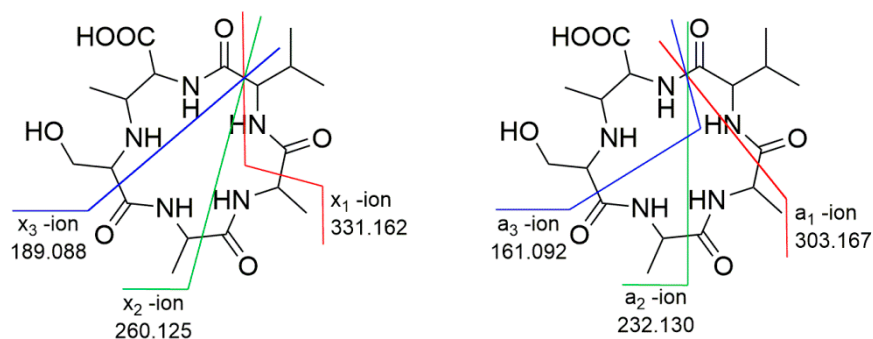


Figure S6: <sup>15</sup>N-HSQC spectrum (700 MHz, DMSO-*d*<sub>6</sub>) of bonsecamin.

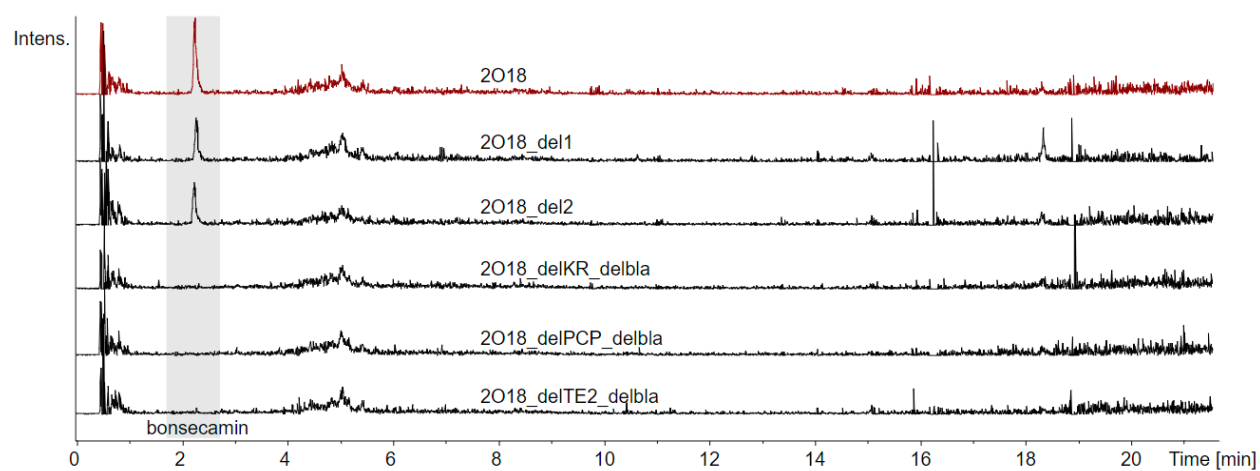


**Figure S7:** LC-MS chromatograms of hydrolyzed bonsecamin derivatized with D- or L-FDLA and the amino acid (aa) references derivatized with L-FDLA.



**Figure S8:** MS/MS fragmentation of bonsecamin.





**Figure S9.** Production of bonsecamin in *S. albus* Del14 mutant after gene deletion experiments. EIC extracted for masses [430-431].

**Table S2.** Putative products of the genes in the DNA fragment encoding bonsecamin production.

gene #	locus tag	putative product
1	SACHL_05130	Catalase
2	SACHL_05120	hypothetical protein
3	SACHL_05110	Cobalt import ATP-binding protein CbiO
4	SACHL_05100	Cobalt transport protein CbiQ
5	SACHL_05090	Cobalt transport protein CbiN
6	SACHL_05080	Cobalt transport protein CbiM
7	SACHL_05070	-
8 [bonA]	SACHL_05060	enterobactin exporter EntS
9 [bonB]	SACHL_05050	Tyrosidine synthase 3 - val
10 [bonC]	SACHL_05040	Tyrosidine synthase 3 - ser
11 [bonD]	SACHL_05030	Linear gramicidin dehydrogenase LgrE
12 [bonE]	SACHL_05020	(-)-trans-carveol dehydrogenase
13 [bonF]	SACHL_05010	Dimodular nonribosomal peptide synthase - thr
14 [bonG]	SACHL_05000	Alanine-anticapsin ligase BacD
15	SACHL_04990	hypothetical protein
16	SACHL_04980	hypothetical protein
17	SACHL_04970	-
18	SACHL_04960	P-aminobenzoate N-oxygenase AurF
19	SACHL_04950	hypothetical protein
20	SACHL_04940	hypothetical protein
21	SACHL_04930	hypothetical protein
22	SACHL_04920	CGNR zinc finger
23	SACHL_04910	(S)-2-haloacid dehalogenase
24	SACHL_04900	Putative phenylalanine aminotransferase
25	SACHL_04890	All-trans-nonaprenyl-diphosphate synthase (geranyl-diphosphate specific)
26	SACHL_04880	prenyltransferase
27	SACHL_04870	Squalene-hopene cyclase
28	SACHL_04860	2-octaprenyl-3-methyl-6-methoxy-1,4-benzoquinol hydroxylase

- [1] M. Myronovskyi, B. Rosenkränzer, S. Nadmid, P. Pujic, P. Normand, A. Luzhetskyy, Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters, *Metab. Eng.* 49 (2018) 316–324. <https://doi.org/10.1016/j.ymben.2018.09.004>.
- [2] F. Flett, V. Mersinias, C.P. Smith, High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes, *FEMS Microbiol. Lett.* 155 (2006) 223–229. <https://doi.org/10.1111/j.1574-6968.1997.tb13882.x>.

IV

**Dudomycins – New Secondary Metabolites Produced after  
Heterologous Expression of an Nrps Cluster from *Streptomyces albus*  
ssp. *Chlorinus* Nrri B-24108**

Constanze Lasch, Marc Stierhof, Marta Rodríguez Estévez, Maksym Myronovskyi,  
Josef Zapp and Andriy Luzhetskyy

*Microorganisms* **2020**, 8(11), 1800

DOI: 10.3390/microorganisms8111800

Published online 16<sup>th</sup> November 2020



## Article

# Dudomycins: New Secondary Metabolites Produced after Heterologous Expression of an Nrps Cluster from *Streptomyces albus* ssp. *Chlorinus* Nr1 B-24108

Constanze Lasch <sup>1</sup>, Marc Stierhof <sup>1</sup> , Marta Rodríguez Estévez <sup>1</sup>, Maksym Myronovskiy <sup>1</sup>, Josef Zapp <sup>2</sup> and Andriy Luzhetskyy <sup>1,3,\*</sup>

<sup>1</sup> Department of Pharmaceutical Biotechnology, Saarland University, 66123 Saarbrücken, Germany; constanze.lasch@uni-saarland.de (C.L.); m.stierhof@t-online.de (M.S.);

marta.rodriguezestevez@uni-saarland.de (M.R.E.); maksym.myronovskiy@uni-saarland.de (M.M.)

<sup>2</sup> Department of Pharmaceutical Biology, Saarland University, 66123 Saarbrücken, Germany; j.zapp@mx.uni-saarland.de

<sup>3</sup> AMEG Department, Helmholtz Institute for Pharmaceutical Research Saarland, 66123 Saarbrücken, Germany

\* Correspondence: a.luzhetskyy@mx.uni-saarland.de; Tel.: +49-681-302-70200

Received: 15 October 2020; Accepted: 10 November 2020; Published: 16 November 2020



**Abstract:** Since the 1950s, natural products of bacterial origin were systematically developed to be used as drugs with a wide range of medical applications. The available treatment options for many diseases are still not satisfying, wherefore, the discovery of new structures has not lost any of its importance. Beyond the great variety of already isolated and characterized metabolites, Streptomyces still harbor uninvestigated gene clusters whose products can be accessed using heterologous expression in host organisms. This work presents the discovery of a set of structurally novel secondary metabolites, dudomycins A to D, through the expression of a cryptic NRPS cluster from *Streptomyces albus* ssp. *Chlorinus* NRRL B-24108 in the heterologous host strain *Streptomyces albus* Del14. A minimal set of genes, required for the production of dudomycins, was defined through gene inactivation experiments. This paper also proposes a model for dudomycin biosynthesis.

**Keywords:** Streptomyces; NRPS; heterologous expression

## 1. Introduction

Biologically active natural products of microbial origin are the result of natural design and evolutionary optimization to target essential biological processes, and are, therefore, a valuable source of potential drug leads [1–3]. Since the 1950s, the bacterial genus of *Streptomyces* largely contributed to the pool of diverse structural scaffolds, some of which were developed as successful drugs, e.g., vancomycin (antibacterial) [4], avermectin (antiparasitic) [5,6] and actinomycin D (anticancer) [7,8], etc. As a result of extensive screening, the discovery of structurally novel compounds with new biological targets has become a challenging task nowadays. However, only a small part of the biosynthetic gene clusters encoded in microbial genomes is readily expressed in the laboratory, while the rest remain silent. It seems that under standard conditions, only a confined number of clusters leads to the production of natural products, which are then often rediscovered. Identification of unique biosynthetic pathways within genome sequence data and their targeted expression in optimized chassis strains is regarded as a most promising approach to access new biologically active scaffolds [9,10].

Recently we have reported the first studies on the genome mining of the strain *Streptomyces albus* ssp. *Chlorinus* NRRL B-24108. Heterologous expression of *S. albus* ssp. *Chlorinus* genes enabled identifying the biosynthetic gene clusters of the antibiotic nybomycin and the herbicide albucidin, as well as the

isolation and characterization of the novel compounds benzanthric acid and fredericamycin C2 [11–14]. This study reports on a biosynthetic gene cluster of *Streptomyces albus* ssp. *Chlorinus* encoding an uncharacterized nonribosomal peptide synthetase. The bioinformatic analysis did not reveal any characterized homologs of the studied cluster, implying that it might encode a novel natural product. Heterologous expression of the NRPS cluster in our optimized hosts *Streptomyces albus* Del14 [15] and *Streptomyces lividans* Del8 [16] led to identifying a set of four new compounds we named dudomycins. The compounds were purified, and their structures were elucidated in  $^1\text{H}$  and  $^{13}\text{C}$  nuclear magnetic resonance (NMR) experiments. The identified compounds are structurally related and consist of a core lysine and three branched-chain hydroxy fatty acid residues. A hypothesis on dudomycin biosynthesis is proposed based on the results of gene deletion experiments.

## 2. Materials and Methods

### 2.1. General Experimental Procedures

The strains, bacterial artificial chromosomes (BACs), and plasmids used in this work are listed in Table S3. *Escherichia coli* strains were cultured in lysogeny broth (LB) medium [17]. *Streptomyces* strains were grown on soy flour mannitol (MS) agar [18] and in liquid tryptic soy broth (TSB; Sigma-Aldrich, St. Louis, MO, USA). Liquid DNPM medium (40 g/L dextrin, 7.5 g/L soytone, 5 g/L baking yeast, and 21 g/L MOPS, pH 6.8 as aqueous solution) and defined medium DM (mannitol 5 g/L, amino acid 0.5 g/L,  $\text{K}_2\text{PO}_4$  0.5 g/L,  $\text{MgSO}_4 \times 7 \text{H}_2\text{O}$  0.2 g/L,  $\text{FeSO}_4 \times 7 \text{H}_2\text{O}$  0.01 g/L) were used for metabolite expression. When DM was used for production, the cells of the preculture were washed three times using amino acid-free defined medium prior to inoculation. Amino acids L-val, L-ile, L-lys, and D-lys were supplied to a defined medium as needed. The antibiotics kanamycin, apramycin, hygromycin, ampicillin, and nalidixic acid were added when required.

### 2.2. Isolation and Manipulation of DNA

DNA manipulation, transformation into *E. coli*, as well as intergeneric conjugation between *E. coli* and *Streptomyces*, were performed according to standard protocols [17–19]. BAC DNA from a constructed genomic library of *Streptomyces albus* ssp. *Chlorinus* NRRL B-24108 was isolated with the BACMAX™ DNA purification kit (Lucigen, Middleton, WI, USA). Deletion of several genes was performed on the BAC 4E8 itself using a two-step approach. Step one focused on deleting all genes 1 to 21, leading to BAC 4E8\_del1, step two addressed further deletion of genes 25 and 26 on BAC 4E8\_del1, resulting in BAC 4E8\_del2. The genes were replaced by resistance markers ampicillin and hygromycin through homologous recombination using the Red/ET system [20]. PCR was performed for amplification of the respective gene cassettes from plasmids pUC19 and pXCM hygformax. PCR primers 20190429\_1\_fw, 20190429\_1\_rev, 20190429\_2\_fw, and 20190429\_2\_rev were constructed with overhang regions for site-specific introduction of the cassettes left or right from the dudomycin cluster and simultaneous removal of the genes 1 to 21 or 25 and 26. Restriction mapping and sequencing were used to control the success of the recombination. Restriction enzymes from ThermoFisher Scientific (Waltham, MA, USA) or New England BioLabs NEB (Ipswich, MA, USA) were used according to the manufacturer's instruction.

### 2.3. Metabolite Extraction

For metabolite extraction, *Streptomyces* strains were grown in 15 mL of TSB in a 100 mL baffled flask for 1 to 2 days, and 1 mL of seed culture was used to inoculate 100 mL of production medium in a 500 mL baffled flask. Cultures were grown for 7 days at 28 °C and 180 rpm in an Infors multitron shaker (Infors AG, Basel, Switzerland). Metabolites were extracted from the culture supernatant with an equal amount of either *n*-butanol or ethyl acetate, evaporated at 40 °C and kept at storage condition 4 °C.

#### 2.4. Mass Spectrometry (MS) Metabolite Analysis

Dried extracts were dissolved in methanol prior to the mass spectrometry (MS) analysis. MS experiments were carried out on a Dionex Ultimate 3000 UPLC system (ThermoFisher Scientific, Waltham, MA, USA) coupled to PDA detector (stationary phase 30 or 100 mm ACQUITY UPLC BEH C18 1.7  $\mu$ m column (Waters Corporation, Milford, MA, USA), mobile phase: Linear gradient of [A] ddH<sub>2</sub>O + 0.1% formic acid/[B] acetonitrile + 0.1% formic acid, 5% to 95% at flow rate of 0.6 mL/min). Further mass detection was performed coupling either an amaZon speed (Bruker, Billerica, MA, USA) or LTQ Orbitrap XL mass spectrometer (ThermoFisher Scientific, Waltham, MA, USA) using positive ionization mode and mass range detection of m/z 200 to 2000. Data analysis was performed using software Compass Data Analysis v. 4.1 (Bruker) and Xcalibur v. 3.0 (ThermoFisher Scientific).

#### 2.5. Purification

The extract from the 10 L culture was dissolved in methanol. A first purification step was carried out using Size Exclusion Chromatography (SEC; stationary phase: Sephadex-LH20; mobile phase: isocratic elution using 100% methanol). Fractions containing dudomycins were pooled, dried, redissolved in methanol and undergone a second purification step: Reversed Phase (RP) HPLC (Agilent Infinity 1200 series HPLC system; stationary phase: Synergi<sup>TM</sup> 4  $\mu$ m Fusion-RP 80 Å 250  $\times$  10 (Phenomenex, Torrance, CA, USA); mobile phase: Linear gradient of [A] H<sub>2</sub>O + 0.1% formic acid/[B] acetonitrile + 0.1% formic acid, 30% to 95% [B] in 14.5 min at flow rate of 4 mL/min, column oven temperature 45 °C, detection UV 210 nm followed by fraction control on HPLC-MS). Fractions were pooled to obtain the four pure dudomycin isolates A to D.

#### 2.6. Nuclear Magnetic Resonance (NMR) Spectroscopy

The <sup>1</sup>H-NMR spectrum in CDCl<sub>3</sub> (Deutero, Kastellaun, Germany) was recorded on a Bruker Avance 500 spectrometer (Bruker, BioSpin GmbH, Rheinstetten, Germany) equipped with a 5 mm BBO probe at 298 K. The chemical shifts were reported in parts per million (ppm) relative to TMS. The spectra were recorded with the standard <sup>1</sup>H pulse program using 64 scans. All other NMR spectra were acquired on a Bruker Ascend 700 MHz NMR spectrometer at 298 K equipped with a 5 mm TXI cryoprobe. As a solvent, deuterated CD<sub>3</sub>OD was used. HSQC, HMBC, 1H-1H COSY spectra were recorded using the standard pulse programs from the TOPSPIN v. 3.6 software. Selective 1D TOCSY experiments were performed using mixing times of 120 ms.

#### 2.7. Genome Mining and Bioinformatic Analysis

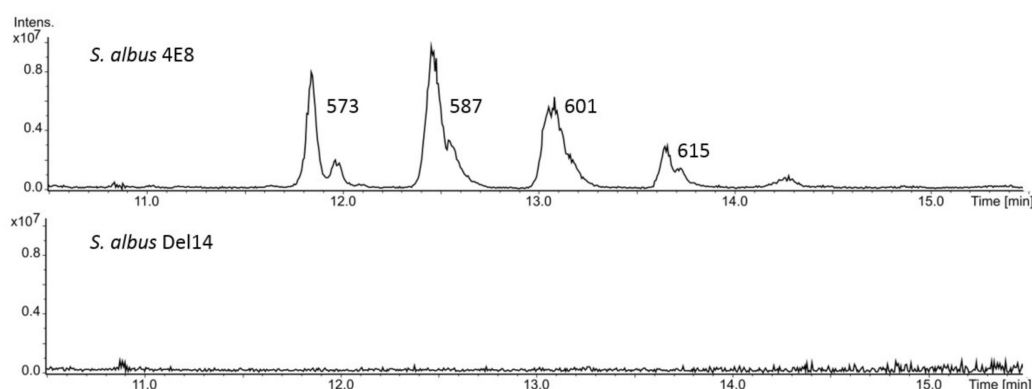
The genome of *S. albus* ssp. *Chlorinus* was screened for secondary metabolite biosynthetic gene clusters using the antiSMASH online tool (<https://antismash.secondarymetabolites.org/#!/start>) [21]. Analysis of genetic data was performed using Geneious software, v. 11.0.3 [22]. The genomic sequence of *Streptomyces albus* ssp. *Chlorinus* was deposited in GenBank under accession number VJOK000000000. For dereplication, the Dictionary of Natural Products (DNP) 28.1 was used as a database of known natural products.

### 3. Results and Discussion

#### 3.1. Identification and Expression of the NRPS Gene Cluster

Genome mining of *Streptomyces albus* ssp. *Chlorinus* NRRL B-24108 using AntiSMASH software revealed several cryptic biosynthetic gene clusters (BGCs) within its chromosome [22]. A BGC encoding a putative nonribosomal peptide synthetase (NRPS) caught our attention as the software did not detect any homology to already characterized BGCs [23]. A BAC 4E8 covering the entire NRPS cluster was identified in the previously constructed genomic library of *S. albus* ssp. *Chlorinus* (GenBank accession number VJOK000000000). The BAC 4E8 was transferred into the optimized heterologous host strains

*Streptomyces albus* Del14 and *Streptomyces lividans* Del8, leading to the respective exconjugant strains *Streptomyces albus* 4E8 and *Streptomyces lividans* 4E8. The obtained exconjugant strains were cultivated in the production medium DNPM, and the metabolites were extracted from the culture supernatant with ethyl acetate or *N*-butanol. High-resolution HPLC-MS analysis revealed the presence of four new peaks in the ethyl acetate and *N*-butanol extracts of both *S. albus* and *S. lividans* harboring the BAC 4E8 (Figure 1; Figures S1 and S2). The identified peaks could not be observed in the extracts of the control strains without the BAC. Analysis of the mass spectra of the identified peaks revealed the molecular ions  $[M + H^+]$  with the masses 573.411, 587.426, 601.442, and 615.457 Da. A search in the DNP database of natural products for the identified high-resolution masses did not generate any matches, implying that the identified compounds might be new. The mass differences of 14 Da between the individual compounds imply that they might differ by the presence of additional  $CH_2$  groups. The identified compounds were named dudomycins A, B, C, and D.



**Figure 1.** HPLC-MS analysis of dudomycin production by *S. albus* Del14 strain harboring the BAC (bacterial artificial chromosomes) 4E8. The strain *S. albus* Del14 without the BAC was used as a control. The extracted base peak chromatograms 573–574 Da, 587–588 Da, 601–602 Da, 615–616 Da are shown.

### 3.2. Purification and Structure Elucidation

To get insights into the structures of the produced compounds, the producer strain *S. albus* 4E8 was cultivated in 10 L of the production medium DNPM, and the metabolites were extracted from the culture supernatant with ethyl acetate. The strain *S. albus* 4E8 was preferred to *S. lividans* 4E8, due to its higher production level. The dudomycins A to D corresponding to the identified molecular ions  $[M + H^+]$  with the masses 573.411, 587.426, 601.442, and 615.457 Da were successfully purified from the extract: 1.2 mg of dudomycin A, 1.1 mg of dudomycin B, 0.7 mg of dudomycin C and 0.5 mg of dudomycin D were obtained. The structure elucidation of the isolated dudomycins was performed using 1D and 2D NMR with a special focus on 1D TOCSY experiments.

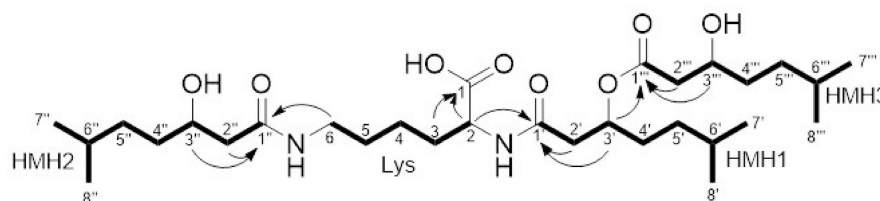
Due to its small molecular mass, dudomycin A was the first compound used for the structure elucidation. The molecular formula was calculated as  $C_{30}H_{56}O_8N_2$  based on the calculated high-resolution mass of 572.403 Da, indicating four degrees of unsaturation. Analysis of  $^1H$ ,  $^{13}C$  NMR, and edited HSQC led to 6 methyls, 13 methylenes, 7 methines, and 4 quaternary carbons. Two of the remaining five protons belonged to NH-groups, as the  $^1H$  NMR measurement in  $CDCl_3$  revealed the presence of two signals at  $\delta H$  6.81 and  $\delta H$  7.12 ppm (Figure S3), which disappeared in protic solvents. Due to broad signals and the compound's poor solubility in  $CDCl_3$ , all other spectra were recorded in  $CD_3OD$ .

Interpretation of HHCOSY revealed four discrete spin systems in the molecule. The first sequence starting from the methine signal at  $\delta H$  4.26 followed by four methylenes at  $\delta H$  1.83, 1.38, 1.51, and 3.17 was assigned to lysine, which was supported by HMBC data. The remaining three spin systems were mostly close to each other, leading to many overlaps of methylene and methyl resonances in the area of  $\delta H$  0.8–1.7 (Figure S4). However, well-separated methine proton signals at  $\delta H$  5.20,



3.95, 3.92 enabled selective 1D TOCSY experiments using these resonances as irradiation points. Their careful interpretation enabled the full assignment of all spin systems leading to three different 3-hydroxy-6-methyl heptanoic acid moieties (HMH1 to HMH3).

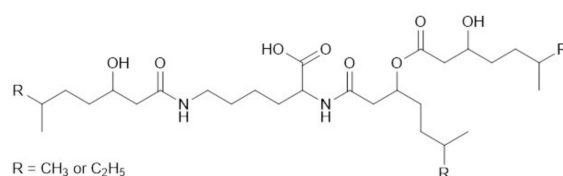
Long-range HMBC correlations of Lys-H-2 ( $\delta$ H 4.26,  $\delta$ C 56.2) to HMH1-C-1' ( $\delta$ C 171.9) and Lys-H-6 ( $\delta$ H 3.17,  $\delta$ C 40.4) to C-1' ( $\delta$ C 174.3) proved connections of HMH1 and HMH2 via amide bonds. The carbonyl group of lysine ( $\delta$ C 178.9) did not show any external connections, hence it was assigned as a free carboxyl group. The remaining long-range HMBC correlations between H-3' ( $\delta$ H 5.20,  $\delta$ C 73.3) and C-1''' ( $\delta$ C 173.2) indicated an ester bond between HMH1 and HMH3 leading to the final structure of dudomycin A (Figure 2; Figures S3–S10; Table S1).



**Figure 2.** Observed HMBC (arrows) and selective 1D TOCSY (bold lines) key correlations of dudomycin A.

The results of NMR analysis indicate that the dudomycins B ( $C_{31}H_{58}O_8N_2$ ), C ( $C_{32}H_{60}O_8N_2$ ), and D ( $C_{33}H_{62}O_8N_2$ ) are structurally related to dudomycin A. From these compounds, only dudomycin D (Table S2) was a pure substance, while dudomycins B and C were the mixtures of several isomers. Similar to dudomycin A, the structure of dudomycin D contains a lysine core, with three hydroxy fatty acids attached. However, in contrast to dudomycin A, three residues of 3-hydroxy-6-methyl octanoic acid (HMO1 to HMO3) can be found in the structure of dudomycin D instead of three residues of 3-hydroxy-6-methyl heptanoic acid (HMH1 to HMH3). The alkyl chain of the 3-hydroxy-6-methyl octanoic acid is extended by one additional  $CH_2$  group compared to 3-hydroxy-6-methyl heptanoic acid what explains the overall mass difference of 42 Da between dudomycins A and D.

NMR spectra of dudomycins B and C revealed that both samples consist of three isomers each. Similar to dudomycins A and D, dudomycins B and C also contain a lysine core in their structures. In contrast to dudomycins A and D, which contain either HMH or HMO residues bound to the core, dudomycins B and C contain a mixture of HMH and HMO residues in their structures. The results of  $^1H$  NMR and a multitude of 1D TOCSY measurements indicate that dudomycin B contains two HMH and one HMO residues, while dudomycin C contains one HMH and two HMO residues. These results are in accordance with the observed mass difference of 14 Da between Dudomycin A and B and of 28 Da between dudomycins A and C. Three isomeric forms are possible for both dudomycin B and C (Figure 3, Figures S11–S24).



**Figure 3.** The structures of isolated dudomycins. In the case of dudomycin A, all three R groups correspond to the  $CH_3$  group. Dudomycin B is a mixture of 3 constitutional isomers with two R groups corresponding to  $CH_3$  and one R group—to  $C_2H_5$ . Dudomycin C is a mixture of 3 constitutional isomers with one R group corresponding to  $CH_3$  and two R groups—to  $C_2H_5$ . In dudomycin D, all three R groups correspond to  $C_2H_5$ .

Experiments to determine the absolute configuration of dudomycins have not been performed, due to low amounts of the isolated compounds. The results of a feeding experiment imply that L-lysine



is used as a precursor for dudomycin biosynthesis (Figure S25). A higher dudomycin production level was observed upon the cultivation of *S. albus* 4E8 in a defined medium with L-lysine as a nitrogen source than in the medium with D-lysine.

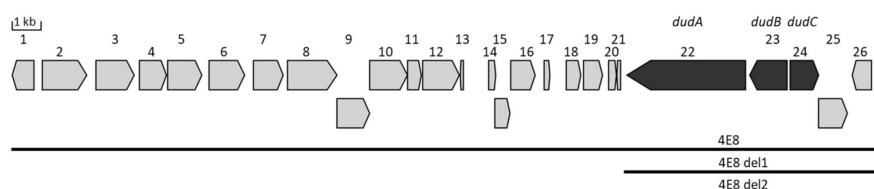
The occurrence of HMM and HMO in natural products is very rare, and to the best of our knowledge, the structures of dudomycin A to D have not been reported before.

### 3.3. Determination of the Minimal Dudomycin Gene Cluster

The 30 kb DNA fragment cloned in the BAC 4E8 contains 26 genes (Figure 4; Table 1). Gene 22 encodes a putative NRPS comprising condensation (C), adenylation (A), peptidyl carrier protein (PCP), and thioesterase (TE) domains. The A domain was predicted to have a weak preference for recognition of the amino acid ornithine. Due to the structural similarity of ornithine and lysine gene 22 encoding, an NRPS enzyme was regarded to be involved in dudomycin biosynthesis. A sequence analysis of the regions upstream and downstream of gene 22 was performed to identify the genes possibly involved in the production of dudomycins. This analysis did not reveal any gene whose product could be enzymatically involved in dudomycin biosynthesis. Sequence homology analysis revealed that the homologs of genes 22, 23, and 24 are clustered together in the genomes of seven different strains implying that those genes might be involved in the same pathway. Genes 23 and 24 encode a putative transport protein and a putative transcriptional regulator, respectively. To confirm that the genes upstream, gene 22 are not involved in dudomycin biosynthesis, the genes 1 to 21 were substituted in the BAC 4E8 with an ampicillin cassette using Red/ET. The constructed BAC 4E8\_del1 was transferred into *S. albus* Del14 strain yielding *S. albus* 4E8\_del1. HPLC-MS analysis of the metabolite production by the obtained strains did not detect any differences in dudomycin production compared to the *S. albus* 4E8 strain. This clearly indicated that the genes 1 to 21 encoded in the BAC 4E8 are not involved in the biosynthesis of dudomycins.

To find out if genes 25 and 26 are involved in dudomycin production, they were replaced in the BAC 4E8\_del1 with a hygromycin resistance cassette using Red/ET. The constructed BAC 4E8\_del2 contains genes 22, 23, and 24 only. The BAC was transferred into *S. albus* Del14, and the obtained strain *S. albus* 4E8\_del2 was checked for dudomycin production. HPLC-MS analysis revealed that deleting genes 25 and 26 did not affect the production of dudomycins (Figure S26), indicating that these genes do not take part in the biosynthesis of the compounds.

The results of gene deletion experiments demonstrate that genes 22, 23, and 24 (designated as *dudA*, *dudB*, and *dudC*), encoding the putative NRPS, transport protein, and transcriptional regulator (Table 1, Figure 4), suffice for the production of dudomycins. Since the products of *dudB* and *dudC* do not have an enzymatic function, we suppose that only the product of *dudA* is responsible for the biosynthesis of dudomycins. The products of *dudB* and *dudC* are likely involved in the transport of the biosynthetic products and in the regulation of dudomycin production, respectively. Since the inactivation of the genes *dudB* and *dudC* was not performed, the possibility that the genes are not essential for dudomycin production cannot be completely excluded.



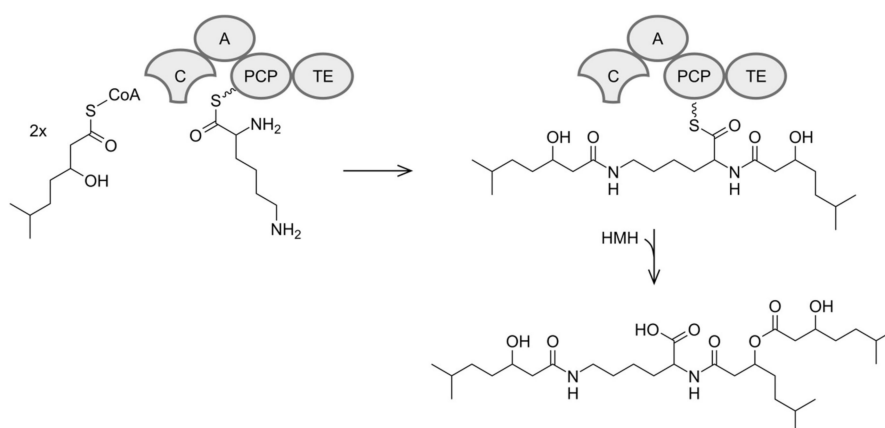
**Figure 4.** The chromosomal fragment of *Streptomyces albus* ssp. *Chlorinus* NRRL B-24108 containing the dudomycin gene cluster. The genes putatively involved in dudomycin biosynthesis are highlighted in dark grey. The chromosomal fragments cloned in BACs 4E8, 4E8\_del1 and 4E8\_del2 are shown with black bars.

**Table 1.** Proposed functions of the genes in the DNA fragment containing the dudomycin gene cluster.

Gene #	Locus Tag	Putative Function
1	SACHL_42600	glycosyltransferase
2	SACHL_42610	hypothetical protein
3	SACHL_42620	phosphatase
4	SACHL_42630	<i>N,N'</i> -diacetyllegionaminic acid synthase
5	SACHL_42640	hypothetical protein
6	SACHL_42650	hydrolase
7	SACHL_42660	membrane lipoprotein precursor
8	SACHL_42670	galactose/methyl galactoside import ATP-binding protein
9	SACHL_42680	ribose transport system permease protein
10	SACHL_42690	branched-chain amino acid transport system/permease component
11	SACHL_42700	cytidine deaminase
12	SACHL_42710	pyrimidine-nucleoside phosphorylase
13	SACHL_42720	hypothetical protein
14	SACHL_42730	hypothetical protein
15	SACHL_42740	hypothetical protein
16	SACHL_42750	hypothetical protein
17	SACHL_42760	hypothetical protein
18	SACHL_42770	hypothetical protein
19	SACHL_42780	ubiquinone biosynthesis <i>O</i> -methyltransferase
20	SACHL_42790	zinc metallo-peptidase
21	SACHL_42800	hypothetical protein
22 [ <i>dudA</i> ]	SACHL_42810	dimodular nonribosomal peptide synthase
23 [ <i>dudB</i> ]	SACHL_42820	inner membrane transport protein
24 [ <i>dudC</i> ]	SACHL_42830	transcriptional regulator
25	SACHL_42840	demethylrebeccamycin-D-glucose <i>O</i> -methyltransferase
26	SACHL_42850	hypothetical protein

### 3.4. Biosynthesis of Dudomycins

Structurally dudomycins consist of a lysine core and three hydroxy fatty acid residues, two of which are attached directly to the lysine moiety through amide bonds, while the third residue is esterified with one of the lysine bound hydroxy fatty acids (Figure 3). Two types of hydroxy fatty acids are used for the biosynthesis of dudomycins: The shorter HMH and the longer HMO. Three HMH residues are found in the smallest compound—dudomycin A, while three HMO residues are used to form the largest compound—dudomycin D. Dudomycins B and C contain both types of hydroxy fatty acids in their structures. The results of gene inactivation studies and sequence analysis demonstrated that only the product of the *dudA* gene is responsible for biosynthesis of dudomycins. The gene encodes an NRPS comprising four domains only: C, A, PCP, and TE. We propose that the A domain activates the L-lysine and loads it on the PCP module. The C domain then attaches two hydroxy fatty acid moieties to the two available amino groups of PCP-bound lysine. The product is released through the hydrolytic activity of the TE domain. We propose that the third hydroxy fatty acid is spontaneously esterified with HMH1 or HMO1, respectively (Figures 2 and 5). It remains unclear if the attachment of the third hydroxy fatty acid occurs before or after the release of the NRPS product. No peaks corresponding to the dudomycin derivatives without HMH3 or HMO3 residue could be identified by HPLC-MS analysis in the extracts of *Streptomyces* strains harboring the dudomycin biosynthetic cluster.

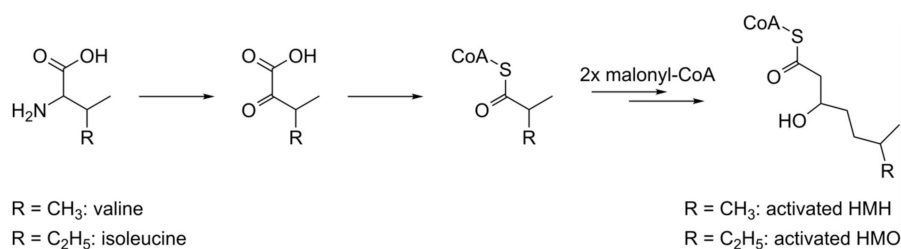


**Figure 5.** Proposed biosynthesis exemplarily for dudomycin A.

Attachment of two hydroxy fatty acid residues to a lysine, which is catalyzed by the single-module NRPS encoded by *dudA* is not typical. At least another two cases are known when a single NRPS module performs two rounds of condensation. During the biosynthesis of myxochelins a single-module NRPS encoded by the *mxcG* gene creates the amide linkage between two 2,3-dihydroxybenzoic acid residues and the two amino groups of lysine [24]. During the biosynthesis of vibriobactin the single-module NRPS encoded by the *vibF* gene also catalyzes two condensation events. However, in the case of vibriobactin, two condensation domains are present within *VibF* each of which is likely responsible for one condensation reaction [25].

Three genes, *dudA* to *dudC*, are required for the biosynthesis of dudomycins in *S. albus* Del14 and *S. lividans* Del8. The *dudA* gene encodes the above mentioned NRPS, and genes, *dudB* and *dudC*, encode the putative membrane transporter and transcriptional regulator. No genes involved in precursor supply could be identified in the DNA regions flanking genes, *dudA* to *dudC*, implying that all precursors required for the dudomycin biosynthesis, including the hydroxy fatty acids HMH and HMO are provided by the host metabolism.

*Streptomyces* are known for their extraordinary prevalence of branched-chain fatty acids, which are synthesized by type II fatty acid synthases [26]. Branched-chain amino acids often serve as precursors for biosynthesis of iso and anteiso carboxylic acids. Through oxidation of the amine group, the amino acids are converted into  $\alpha$ -keto acids, which are then used as starter units in bacterial fatty acid biosynthesis [27]. Valine and isoleucine with high probability serve as main biosynthetic precursors for HMH and HMO, respectively. These amino acids are converted to 3-methyl-2-oxobutanoic acid and 3-methyl-2-oxopentanoic acid through the action of valine dehydrogenase or homologous enzymes. The formed 2-keto carboxylic acids are decarboxylated and used as starter units by a type II fatty acid synthase which converts them to HMH and HMO after decarboxylative condensation of two malonyl units (Figure 6). The precursor role of L-valine and L-isoleucine in biosynthesis of HMH and HMO is indirectly confirmed by feeding studies [28]. During cultivation in a defined medium with L-valine as nitrogen source, the strain *S. albus* 4E8\_del2 produced mostly the smaller dudomycins (dudomycins A and B), which contain mainly HMH, which is derived from L-valine. With L-isoleucine, as a single nitrogen source, the strain produced mostly the bigger dudomycins C and D. These compounds contain mainly HMO, which is derived from L-isoleucine (Figure S25). Furthermore, heavily impaired production of dudomycins was observed when BAC 4E8 was expressed in an *S. albus* delVDH strain [29] with inactivated valine dehydrogenase gene *vdh* (Figure S27). This result suggests that the valine dehydrogenase is involved in the supply of HMH and HMO by oxidation of the branched-chain amino acids L-valine and L-isoleucine. Low dudomycin production by *S. albus* delVDH 4E8\_del2 also indicates that other homologs of the valine dehydrogenase with broad substrate specificity are encoded in the genome of *S. albus*, in accordance with previously published data [30].



**Figure 6.** Proposed biosynthesis of HMH and HMO.

#### 4. Conclusions

In this paper, we report the identification and successful heterologous expression of a new NRPS cluster leading to the production of a group of new compounds called dudomycins A to D. The single-module NRPS which is responsible for dudomycin production activates L-lysine and catalyzes its condensation with two hydroxy fatty acid CoA precursors. Such examples where a single condensation domain catalyzes two condensation steps are very rare. The branched hydroxy fatty acids used for the dudomycin biosynthesis are provided by the host's metabolism. The amino acids L-valine and L-isoleucine serve as main precursors for their biosynthesis.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2076-2607/8/11/1800/s1>. Figure S1: Production of dudomycins by different hosts using different organic solvents for extraction, Figure S2: Mass spectra of dudomycins (A–D), Figure S3: <sup>1</sup>HNMR (500 MHz, CDCl<sub>3</sub>) spectrum of dudomycin A, Figure S4: <sup>1</sup>HNMR (500 MHz, CD<sub>3</sub>OD) spectrum of dudomycin A, Figure S5: <sup>13</sup>C NMR (176 MHz, CD<sub>3</sub>OD) spectrum of dudomycin A, Figure S6: DEPT-135 (176 MHz, CD<sub>3</sub>OD) spectrum of dudomycin A, Figure S7: HSQC (700 MHz, CD<sub>3</sub>OD) spectrum of dudomycin A, Figure S8: COSY (700 MHz, CD<sub>3</sub>OD) spectrum of dudomycin A, Figure S9: HMBC (700 MHz, CD<sub>3</sub>OD) spectrum of dudomycin A, Figure S10: (A) <sup>1</sup>HNMR spectrum of dudomycin A showing the irradiation points B (Lys, H-2), C (HMH3, H-3''), D (HMH2, H-3'') and E (HMH1, H-3'). (B–E) Selective 1D TOCSY (700 MHz, CD<sub>3</sub>OD) spectra of the corresponding irradiation points, Figure S11: <sup>1</sup>HNMR (700 MHz, CD<sub>3</sub>OD) spectrum of dudomycin B, Figure S12: DEPT-135 (176 MHz, CD<sub>3</sub>OD) spectrum of dudomycin B showing a 2/1 ratio of the methyl signals of HMH and HMO, Figure S13: HSQC (700 MHz, CD<sub>3</sub>OD) spectrum of dudomycin B, Figure S14: COSY (700 MHz, CD<sub>3</sub>OD) spectrum of dudomycin B, Figure S15: HMBC (700 MHz, CD<sub>3</sub>OD) spectrum of dudomycin B, Figure S16: (A) <sup>1</sup>HNMR spectrum of dudomycin B showing the irradiation points C (H-3''/H-3''') and C (H-3'). (B–D) Selective 1D TOCSY (700 MHz, CD<sub>3</sub>OD) spectra of the corresponding irradiation points, Figure S17: <sup>1</sup>H NMR (700 MHz, CD<sub>3</sub>OD) spectrum of dudomycin C, Figure S18: DEPT-135 (176 MHz, CD<sub>3</sub>OD) spectrum of dudomycin C showing an approximate 1/2 ratio of the methyl signals of HMH and HMO, Figure S19: HMBC (700 MHz, CD<sub>3</sub>OD) spectrum of dudomycin C, Figure S20: (A) <sup>1</sup>HNMR spectrum of dudomycin C showing the irradiation points B (H-3''/H-3''') and C (H-3'). (B–D) Selective 1D TOCSY (700 MHz, CD<sub>3</sub>OD) spectra of the corresponding irradiation points, Figure S21: <sup>1</sup>HNMR (700 MHz, CD<sub>3</sub>OD) spectrum of dudomycin D, Figure S22: DEPT-135 (176 MHz, CD<sub>3</sub>OD) spectrum of dudomycin D, Figure S23: HSQC (700 MHz, CD<sub>3</sub>OD) spectrum of dudomycin D, Figure S24: (A) <sup>1</sup>HNMR spectrum of dudomycin D showing the irradiation points B (HMO2/3, H-3''/H-3'''), C (HMO1, H-3') and D (Lys, H-2). (B–D) Selective 1D TOCSY (700 MHz, CD<sub>3</sub>OD) spectra of the corresponding irradiation points; Figure S25: Production of dudomycin derivatives in defined medium (DM) with single sources of nitrogen, Figure S26: Production of dudomycins after gene deletion experiments, Figure S27: Decreased production of dudomycin derivatives by *S. albus* delVDH 4E8, Table S1: NMR data of dudomycin A in CD<sub>3</sub>OD, Table S2: NMR data of dudomycin D in CD<sub>3</sub>OD, Table S3: Strains, BACs, plasmids and primers used in this work, Tables S4: Proposed functions of the genes in the DNA fragment containing the dudomycin gene cluster.

**Author Contributions:** Dudomycin peaks were first identified by M.R.E. Except from NMR analysis, the experiments were designed and evaluated by C.L., M.M. and A.L. and the practical work performed by C.L. NMR experiments were set up, carried out and evaluated by M.S. and the data reviewed by J.Z. The manuscript was drafted by C.L., M.S. and M.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research has received funding from BMBF under Grant “EXPLOMARE” 031B0868A.

**Acknowledgments:** We thank Helmholtz Institute for Pharmaceutical Research Saarland, Saarbruecken, Germany (HIPS) for support of NMR measurements. The strain *S. albus* ssp. *chlorinus* NRRL B-24108 was provided by BASF SE Ludwigshafen, Germany.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Beutler, J.A. Natural Products as a Foundation for Drug Discovery. *Curr. Protoc. Pharmacol.* **2009**, *46*. [[CrossRef](#)] [[PubMed](#)]
2. Shen, B. A New Golden Age of Natural Products Drug Discovery. *Cell* **2015**, *163*, 1297–1300. [[CrossRef](#)] [[PubMed](#)]
3. Newman, D.J.; Cragg, G.M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83*, 770–803. [[CrossRef](#)] [[PubMed](#)]
4. Griffith, R.S. Vancomycin use—An historical review. *J. Antimicrob. Chemother.* **1984**, *14*, 1–5. [[CrossRef](#)]
5. Hotson, I.K. The avermectins: A new family of antiparasitic agents. *J. S. Afr. Vet. Assoc.* **1982**, *53*, 87–90.
6. Ottesen, E.A.; Campbell, W. Ivermectin in human medicine. *J. Antimicrob. Chemother.* **1994**, *34*, 195–203. [[CrossRef](#)]
7. D’Angio, G.J.; Evans, A.; Breslow, N.; Beckwith, B.; Bishop, H.; Farewell, V.; Goodwin, W.; Leape, L.; Palmer, N.; Sinks, L.; et al. The treatment of Wilms’ Tumor: Results of the second national Wilms’ Tumor study. *Cancer* **1981**, *47*, 2302–2311. [[CrossRef](#)]
8. Turan, T.; Karacay, O.; Tulunay, G.; Boran, N.; Koc, S.; Bozok, S.; Kose, M.F. Results with EMA/CO (etoposide, methotrexate, actinomycin D, cyclophosphamide, vincristine) chemotherapy in gestational trophoblastic neoplasia. *Int. J. Gynecol. Cancer.* **2006**, *16*, 1432–1438. [[CrossRef](#)]
9. Myronovskiy, M.; Luzhetskyy, A. Heterologous production of small molecules in the optimized *Streptomyces* hosts. *Nat. Prod. Rep.* **2019**, *36*, 1281–1294. [[CrossRef](#)]
10. Baltz, R. Renaissance in antibacterial discovery from actinomycetes. *Curr. Opin. Pharmacol.* **2008**, *8*, 557–563. [[CrossRef](#)]
11. Estévez, M.R.; Myronovskiy, M.; Gummerlich, N.; Nadmid, S.; Luzhetskyy, A. Heterologous Expression of the Nybomycin Gene Cluster from the Marine Strain *Streptomyces albus* subsp. *chlorinus* NRRL B-24108. *Mar. Drugs* **2018**, *16*, 435. [[CrossRef](#)] [[PubMed](#)]
12. Myronovskiy, M.; Rosenkränzer, B.; Stierhof, M.; Petzke, L.; Seiser, T.; Luzhetskyy, A. Identification and Heterologous Expression of the Albucidin Gene Cluster from the Marine Strain *Streptomyces albus* Subsp. *chlorinus* NRRL B-24108. *Microorganisms* **2020**, *8*, 237. [[CrossRef](#)] [[PubMed](#)]
13. Estévez, M.R.; Gummerlich, N.; Myronovskiy, M.; Zapp, J.; Luzhetskyy, A. Benzantronic Acid, a Novel Metabolite From *Streptomyces albus* Del14 Expressing the Nybomycin Gene Cluster. *Front. Chem.* **2020**, *7*. [[CrossRef](#)]
14. Estévez, M.R.; Myronovskiy, M.; Rosenkränzer, B.; Paululat, T.; Petzke, L.; Ristau, J.; Luzhetskyy, A. Novel fredericamycin variant overproduced by a streptomycin-resistant *Streptomyces albus* subsp. *chlorinus* Strain. *Mar. Drugs* **2020**, *18*, 284. [[CrossRef](#)] [[PubMed](#)]
15. Myronovskiy, M.; Rosenkränzer, B.; Nadmid, S.; Pujic, P.; Normand, P.; Luzhetskyy, A. Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. *Metab. Eng.* **2018**, *49*, 316–324. [[CrossRef](#)] [[PubMed](#)]
16. Ahmed, Y.; Rebets, Y.; Estévez, M.R.; Zapp, J.; Myronovskiy, M.; Luzhetskyy, A. Engineering of *Streptomyces lividans* for heterologous expression of secondary metabolite gene clusters. *Microb. Cell Fact.* **2020**, *19*, 5. [[CrossRef](#)]
17. Green, J.M.R.; Sambrook, J. *Molecular Cloning: A Laboratory Manual*, 4th ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2012.
18. Kieser, T.; Bibb, M.J.; Buttner, M.J.; Chater, K.F.; Hopwood, D.A. *Practical Streptomyces Genetics. A Laboratory Manual*; John Innes Foundation: Norwich, UK, 2000.
19. Rebets, Y.; Kormanec, J.; Luzhetskyy, A.; Bernaerts, K.; Anné, J. Cloning and expression of metagenomic dna in *Streptomyces lividans* and subsequent fermentation for optimized production. *Methods Mol. Biol.* **2017**, *1539*, 99–144. [[CrossRef](#)]
20. Court, D.L.; Sawitzke, J.A.; Thomason, L.C. Genetic Engineering Using Homologous Recombination. *Annu. Rev. Genet.* **2002**, *36*, 361–388. [[CrossRef](#)]
21. Medema, M.H.; Blin, K.; Cimermancic, P.; de Jager, V.; Zakrzewski, P.; Fischbach, M.A.; Weber, T.; Takano, E.; Breitling, R. antiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **2011**, *39*, W339–W346. [[CrossRef](#)]



22. Kears, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [[CrossRef](#)]
23. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
24. Gaitatzis, N.; Kunze, B.; Muller, R. In vitro reconstitution of the myxochelin biosynthetic machinery of *Stigmatella aurantiaca* Sg a15: Biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 11136–11141. [[CrossRef](#)] [[PubMed](#)]
25. Keating, T.A.; Marshall, C.G.; Walsh, C.T. Reconstitution and Characterization of the *Vibrio cholerae* Vibriobactin Synthetase from VibB, VibE, VibF, and VibH. *Biochemistry* **2000**, *39*, 15522–15530. [[CrossRef](#)] [[PubMed](#)]
26. Gago, G.; Diacovich, L.; Arabolaza, A.; Tsai, S.-C.; Gramajo, H. Fatty acid biosynthesis in actinomycetes. *FEMS Microbiol. Rev.* **2011**, *35*, 475–497. [[CrossRef](#)] [[PubMed](#)]
27. Kaneda, T. Iso- and anteiso-fatty acids in bacteria: Biosynthesis, function, and taxonomic significance. *Microbiol. Rev.* **1991**, *55*, 288–302. [[CrossRef](#)] [[PubMed](#)]
28. Tang, L.; Hutchinson, C.R. Regulation of expression of the valine (branched-chain amino acid) dehydrogenase-encoding gene from *Streptomyces coelicolor*. *Gene* **1995**, *162*, 69–74. [[CrossRef](#)]
29. Manderscheid, N. Strain Development for Heterologous Expression of Secondary Metabolite Clusters in Actinobacteria. Ph.D. Thesis, Universität des Saarlandes, Saarbrücken, Germany, 2015.
30. Vancurová, I.; Vancura, A.; Volc, J.; Neuzil, J.; Flieger, M.; Basarová, G.; Běhal, V. Isolation and characterization of valine dehydrogenase from *Streptomyces aureofaciens*. *J. Bacteriol.* **1988**, *170*, 5192–5196. [[CrossRef](#)]

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



Article

# Dudomycins: new secondary metabolites produced after heterologous expression of an NRPS cluster from *Streptomyces albus* ssp. *chlorinus* NRRL B-24108

Constanze Lasch <sup>1</sup>, Marc Stierhof <sup>1</sup>, Marta Rodríguez Estévez <sup>1</sup>, Maksym Myronovskiy <sup>1</sup>, Josef Zapp <sup>2</sup> and Andriy Luzhetskyy <sup>1,3,\*</sup>

<sup>1</sup> Pharmaceutical Biotechnology, Saarland University, Saarbruecken, Germany;

constanze.lasch@uni-saarland.de (C.L.), m.stierhof@t-online.de (M.S.),

marta.rodriguezestevez@uni-saarland.de (M.R.), maksym.myronovskiy@uni-saarland.de (M.M.),

a.luzhetskyy@mx.uni-saarland.de (A.L.)

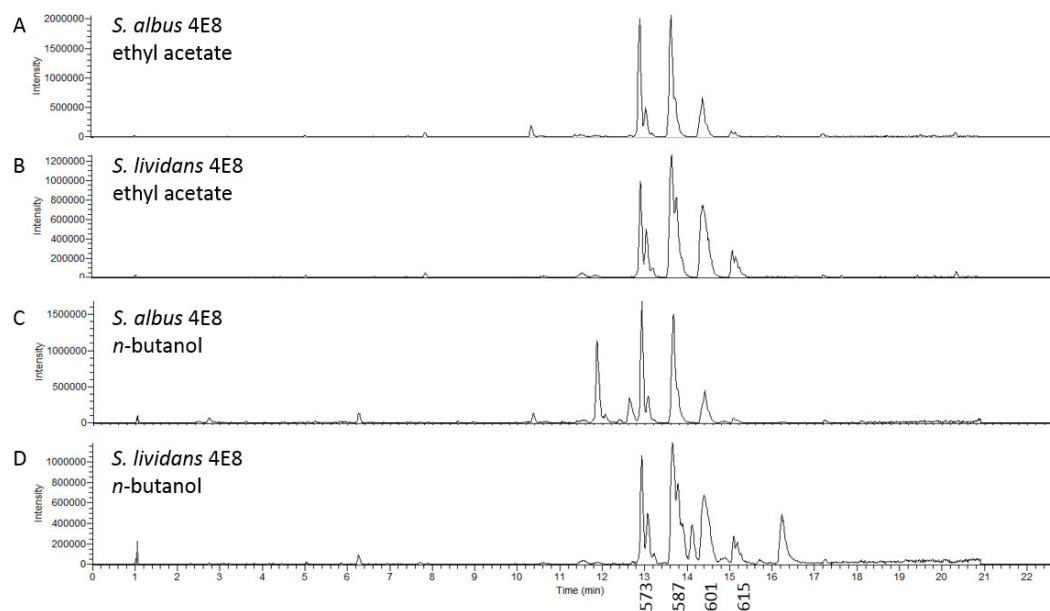
<sup>2</sup> Pharmaceutical Biology, Saarland University, Saarbruecken, Germany; j.zapp@mx.uni-saarland.de (J.Z.)

<sup>3</sup> Helmholtz Institute for Pharmaceutical Research Saarland, Saarbruecken, Germany

\* Correspondence: a.luzhetskyy@mx.uni-saarland.de; +49 681 302 70200 (A.L.)

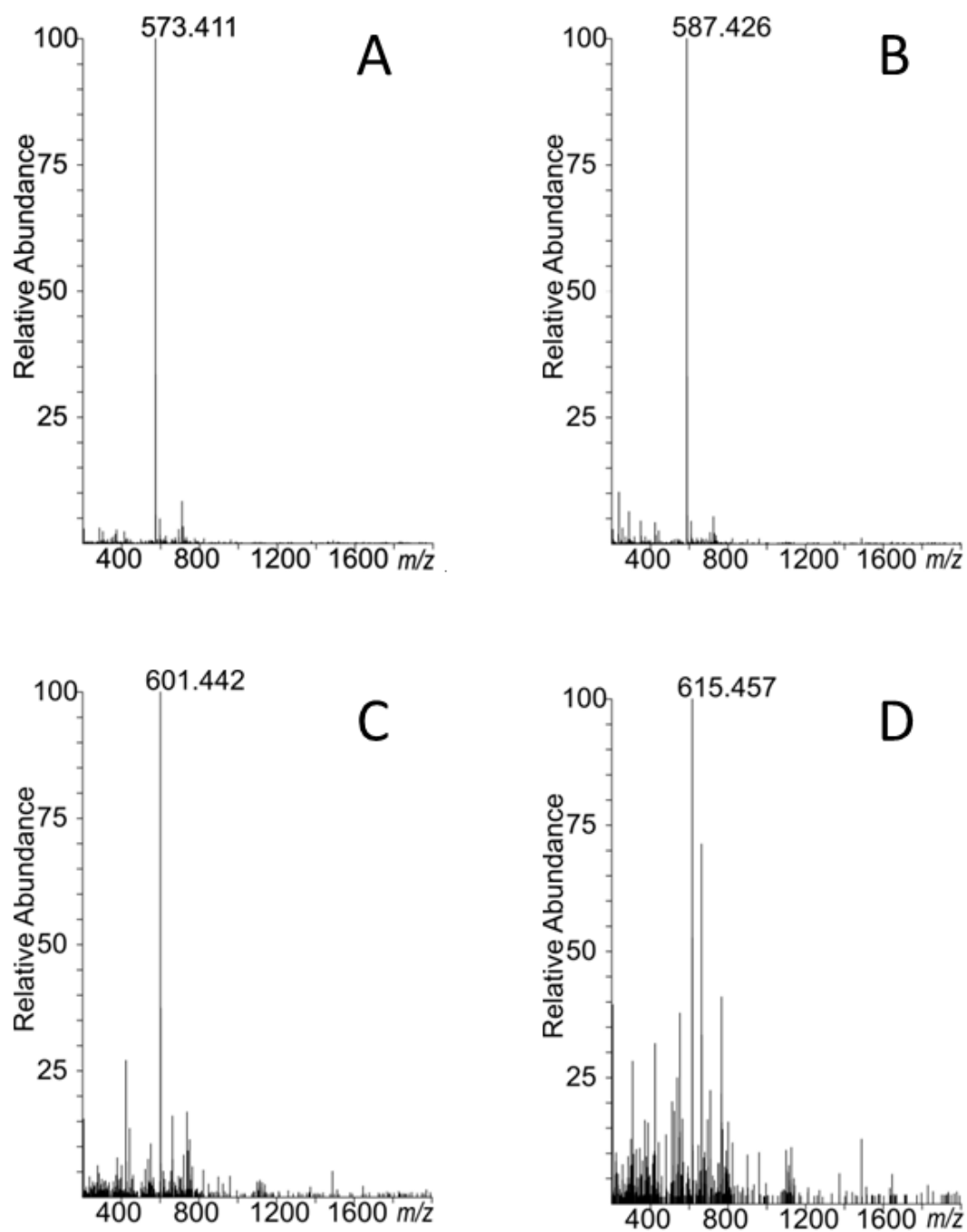
Received: date; Accepted: date; Published: date

**Supplementary.**

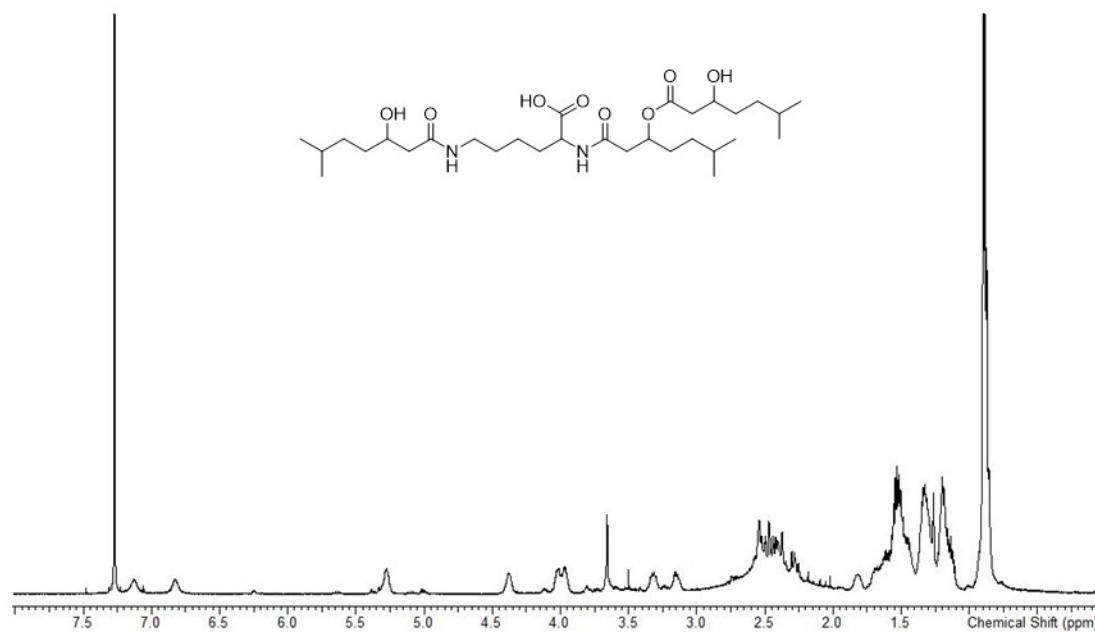


**Figure S1.** Production of dudomycins by different hosts using different organic solvents for extraction. Base Peak Chromatogram (BPC) extracted for masses  $M+H^+$  [573-574], [587-588], [601-602], [615-616]. Butanolic extractions led to more background peaks in crude extracts. *S. albus* 4E8 showed higher production levels compared to *S. lividans* 4E8.

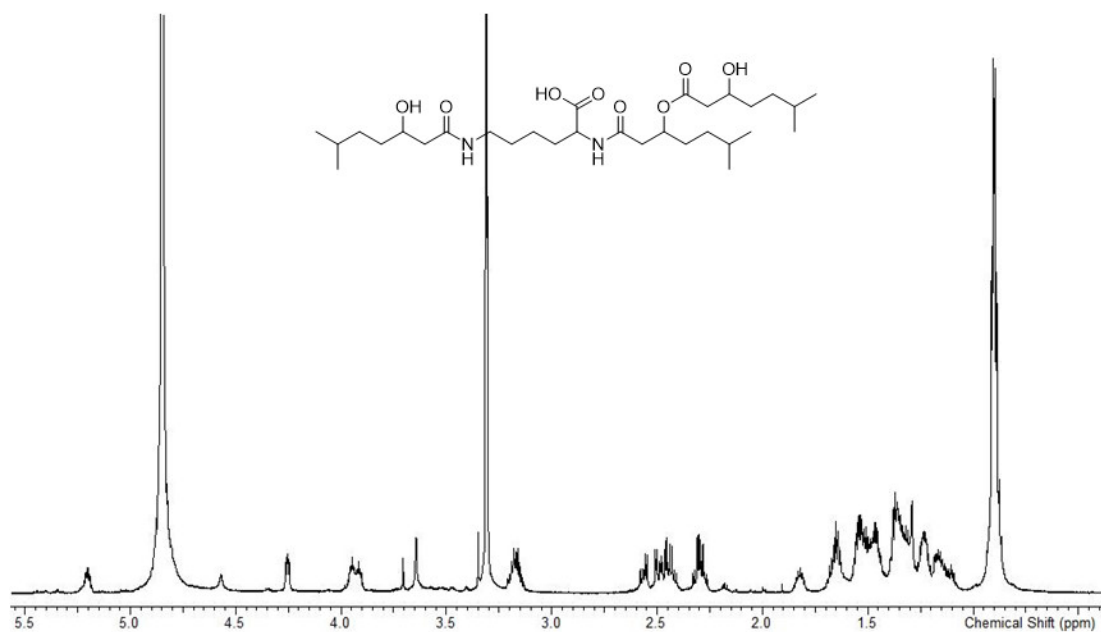




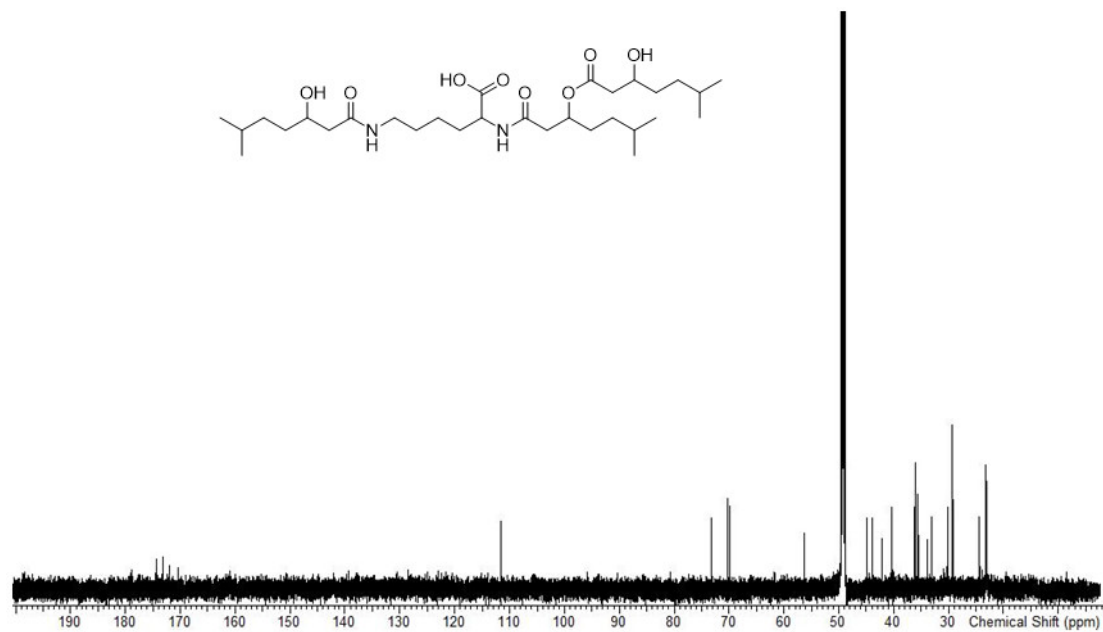
**Figure S2.** A to D: Mass spectra of dudomycins A to D.



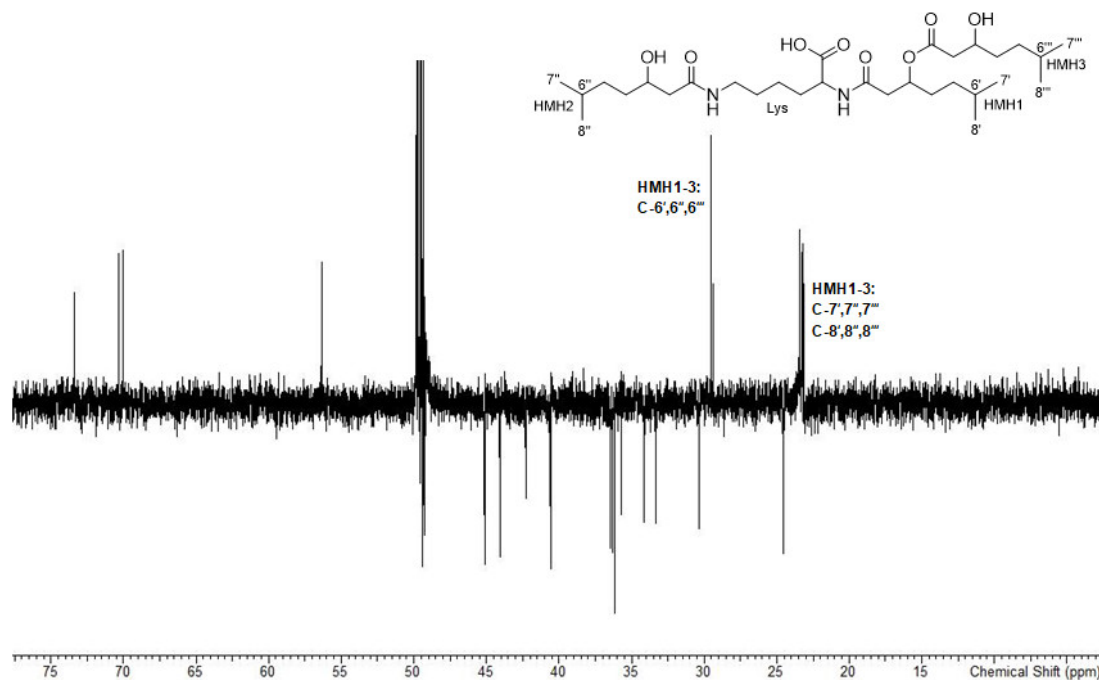
**Figure S3.** <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) spectrum of dudomycin A.



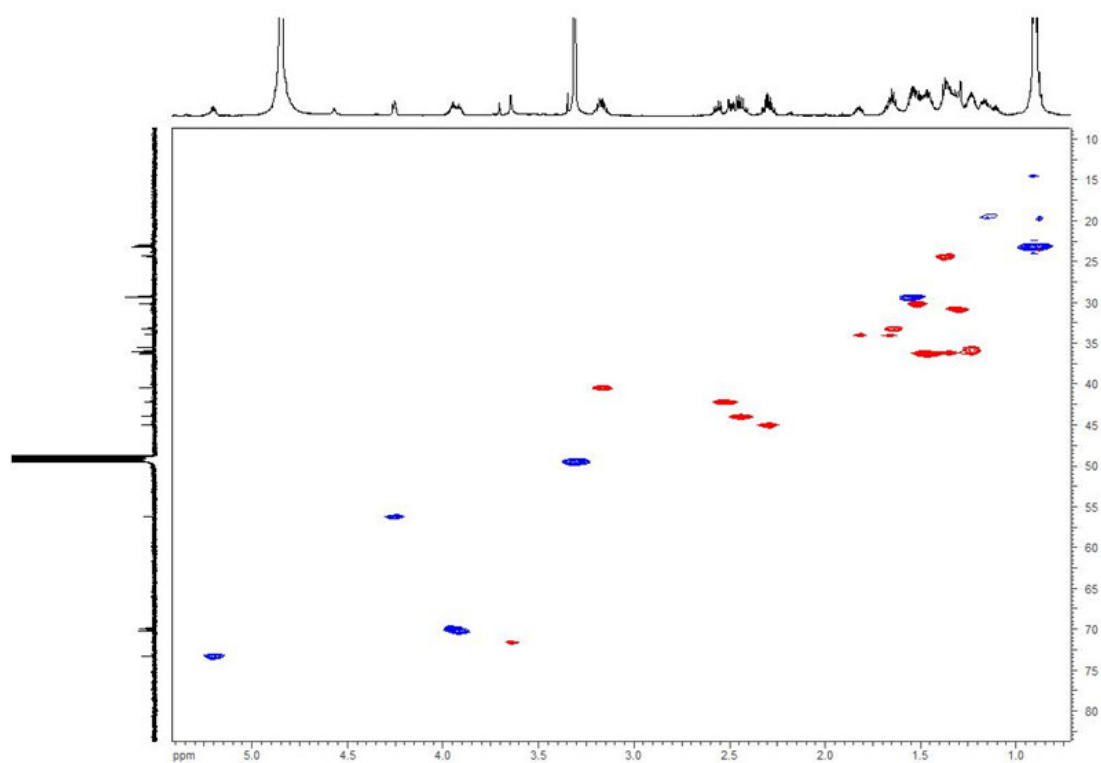
**Figure S4.** <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD) spectrum of dudomycin A.



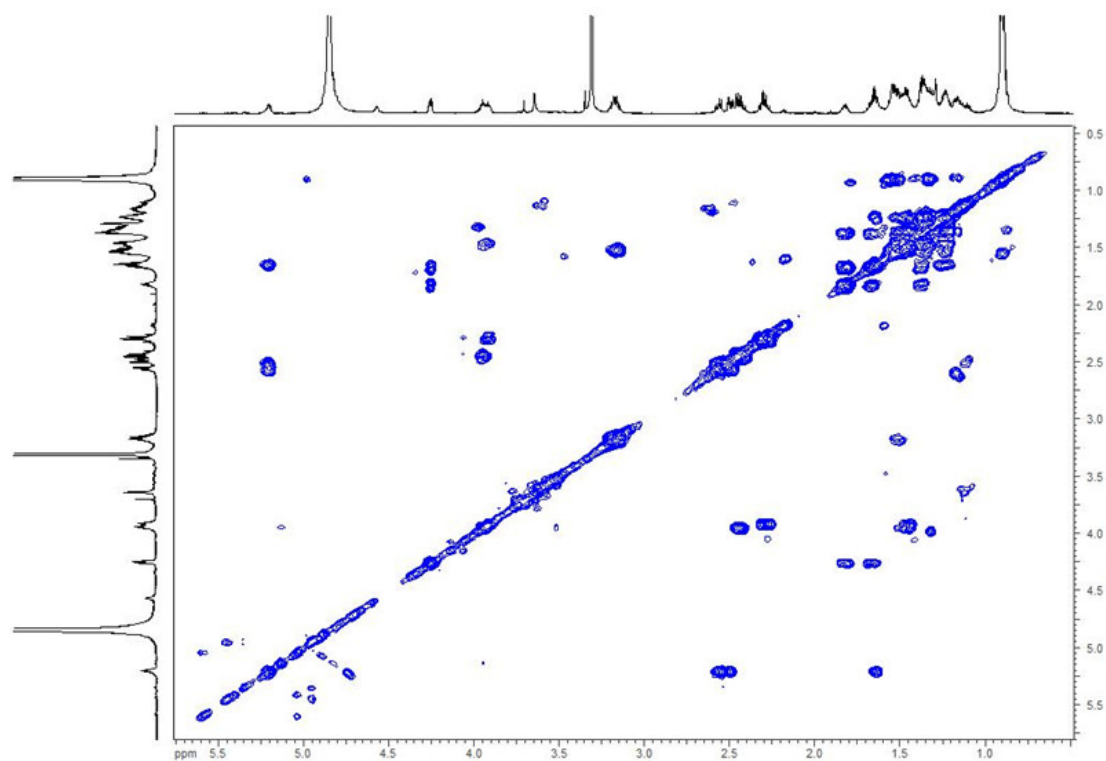
**Figure S5.** <sup>13</sup>C NMR (176 MHz, CD<sub>3</sub>OD) spectrum of dudomycin A.



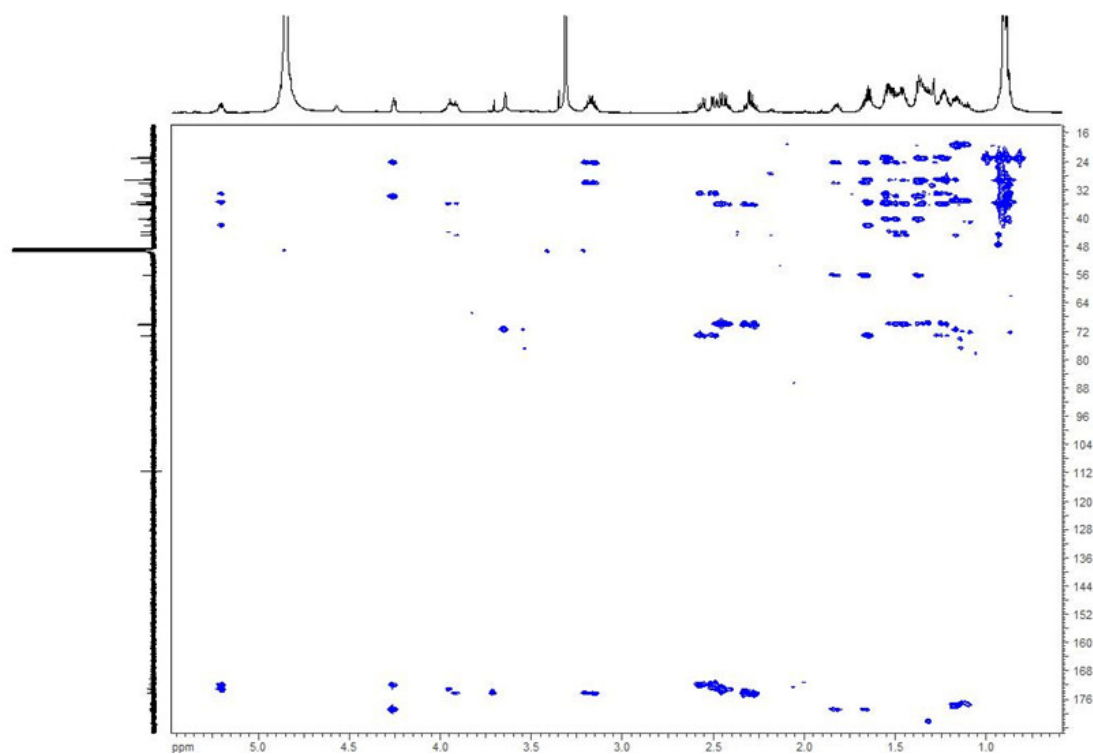
**Figure S6.** DEPT-135 (176 MHz, CD<sub>3</sub>OD) spectrum of dudomycin A.



**Figure S7.** HSQC (700 MHz, CD<sub>3</sub>OD) spectrum of dudomycin A.

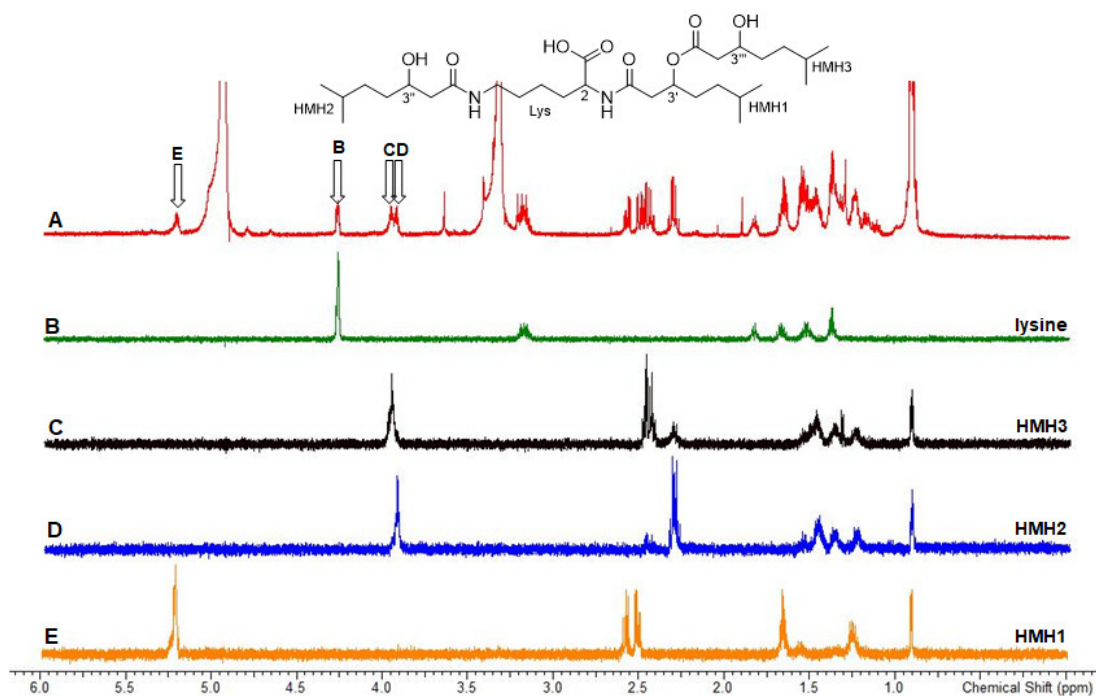


**Figure S8.** COSY (700 MHz, CD<sub>3</sub>OD) spectrum of dudomycin A.

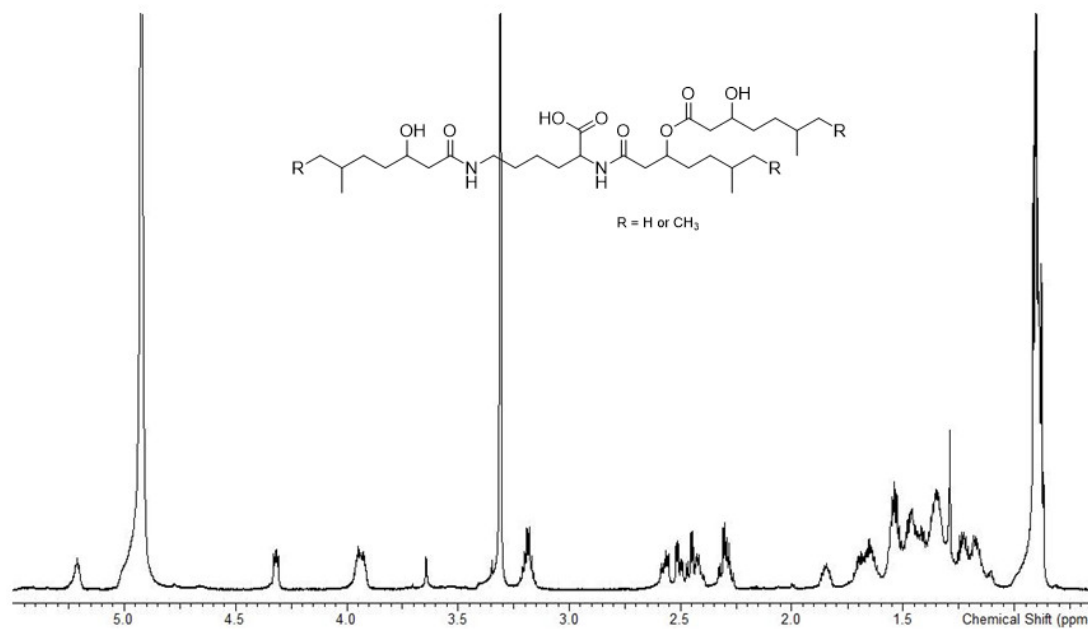


**Figure S9.** HMBC (700 MHz, CD<sub>3</sub>OD) spectrum of dudomycin A.

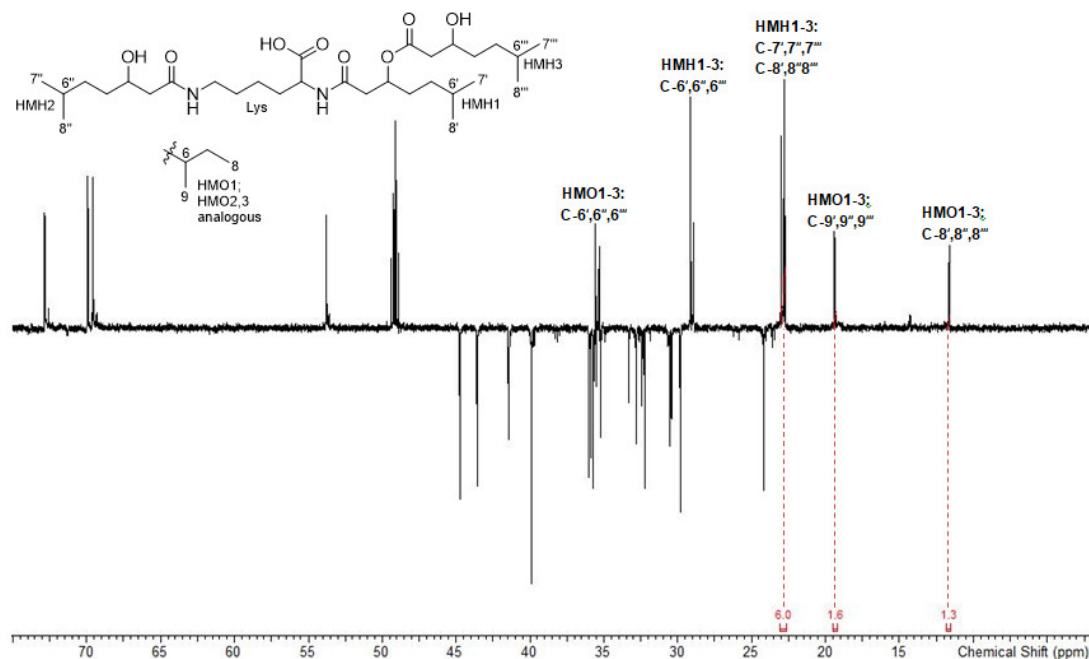




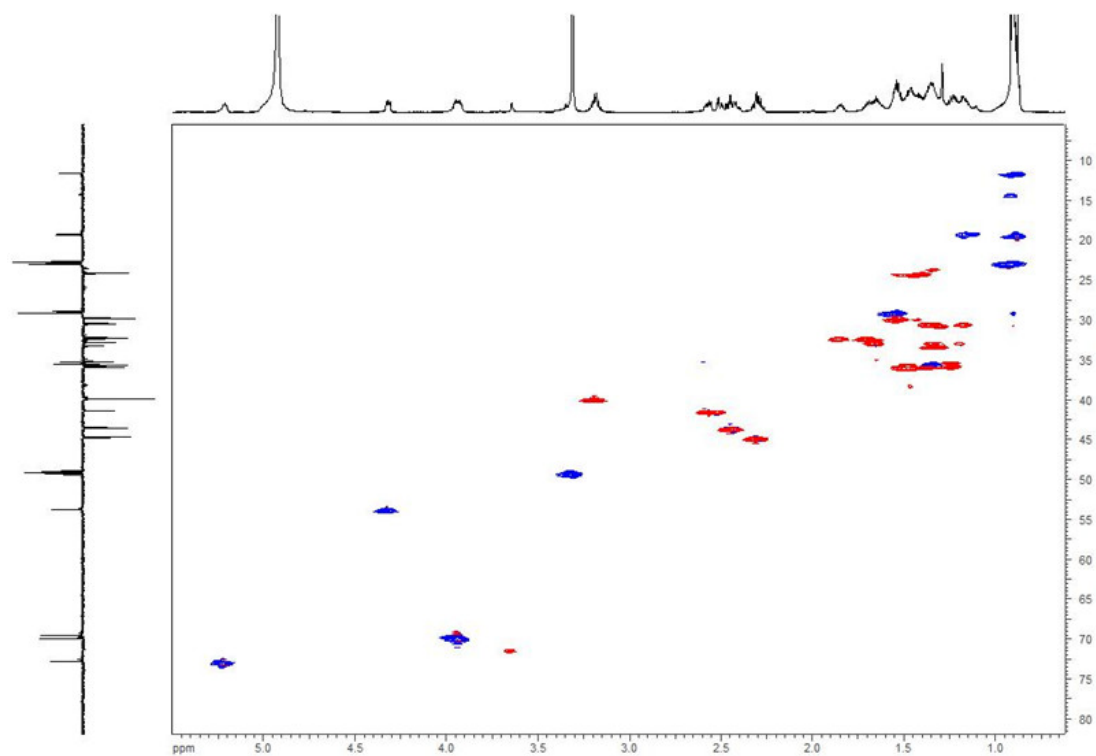
**Figure S10.** A. <sup>1</sup>H NMR spectrum of dudomycin A showing the irradiation points B (Lys, H-2), C (HMH3, H-3''), D (HMH2, H-3') and E (HMH1, H-3'). B-E. Selective 1D TOCSY (700 MHz, CD<sub>3</sub>OD) spectra of the corresponding irradiation points.



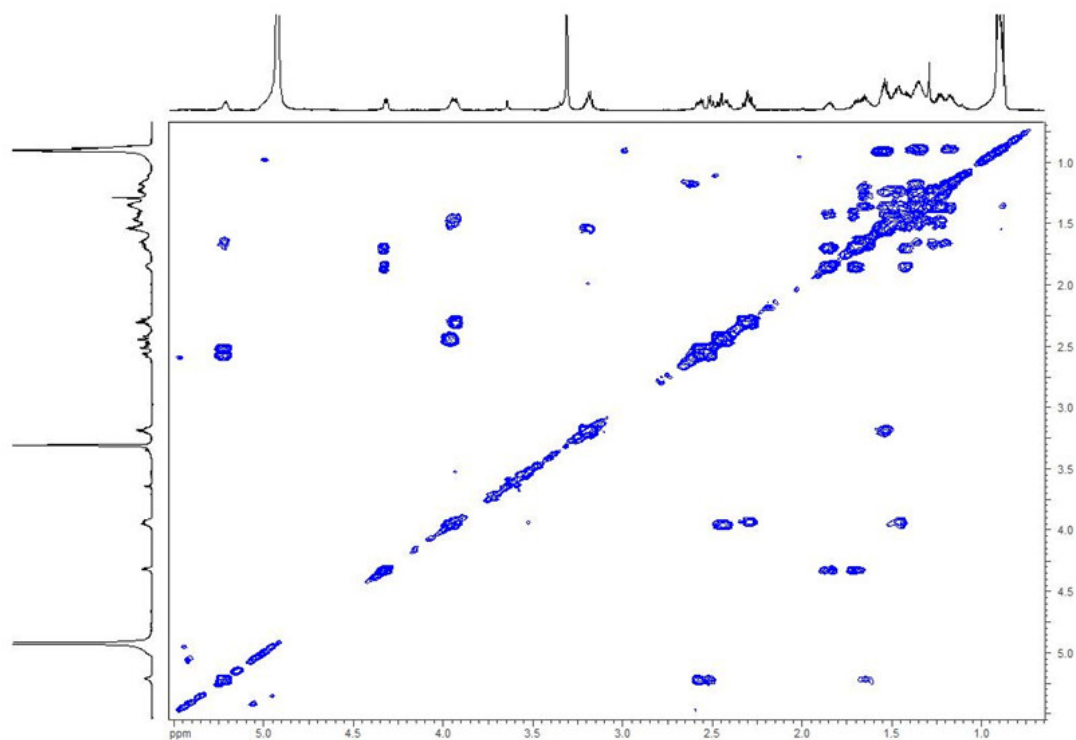
**Figure S11.** <sup>1</sup>H NMR (700 MHz, CD<sub>3</sub>OD) spectrum of dudomycin B.



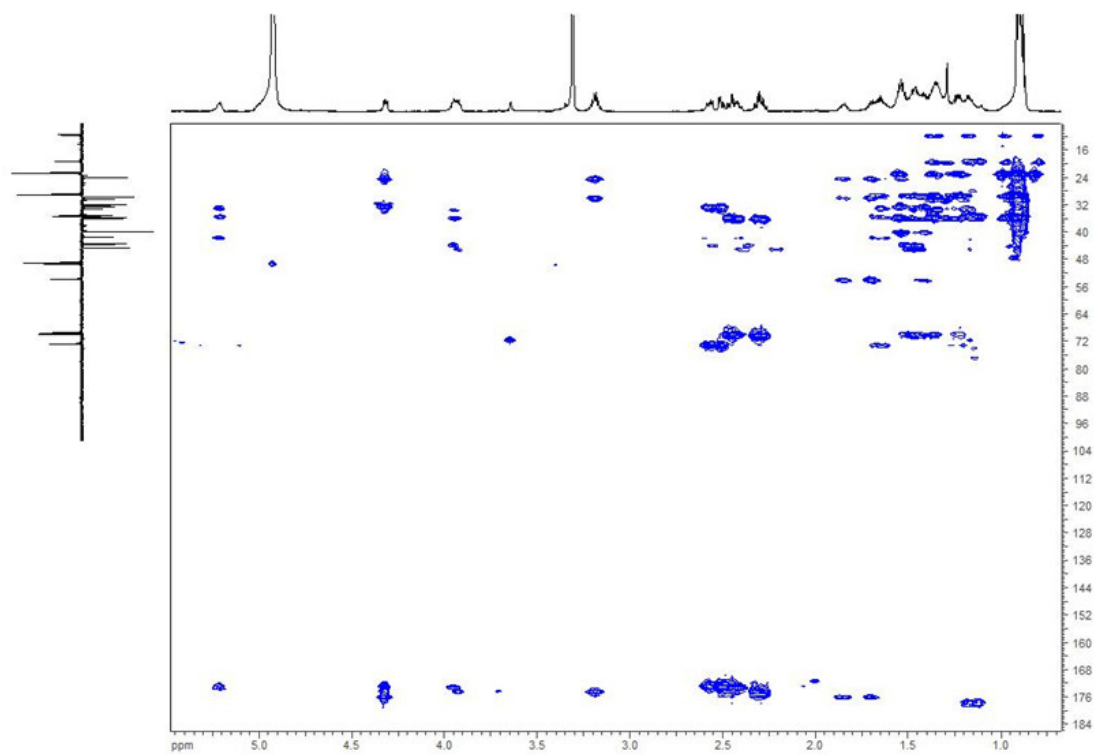
**Figure S12.** DEPT-135 (176 MHz, CD<sub>3</sub>OD) spectrum of dudomycin B showing a 2/1 ratio of the methyl signals of HMH and HMO.



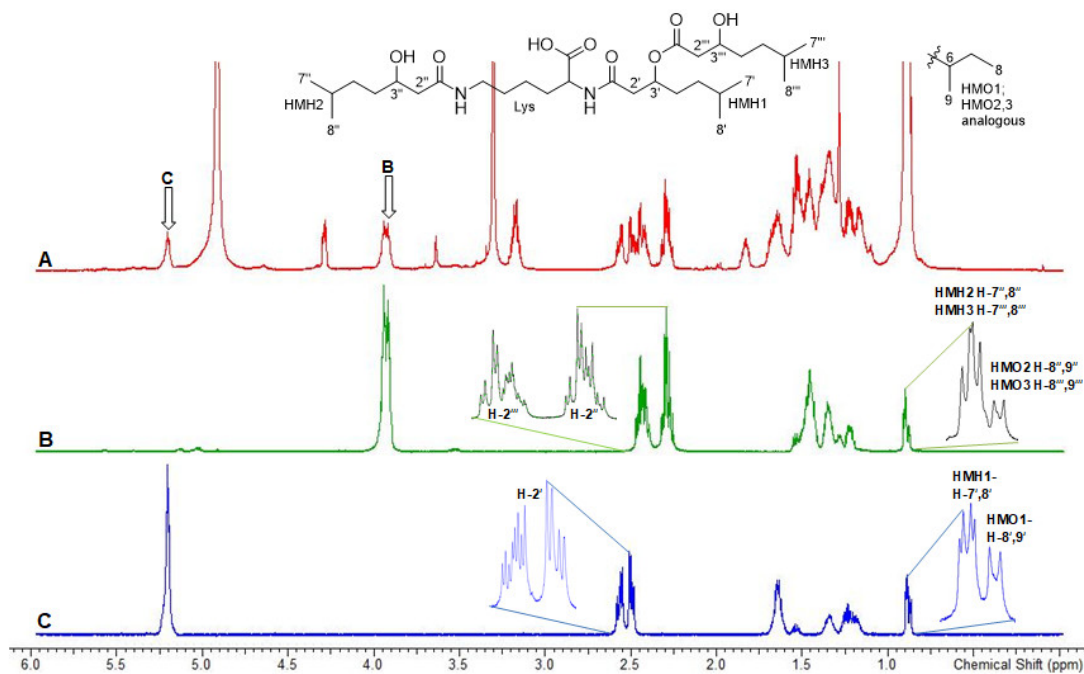
**Figure S13.** HSQC (700 MHz,  $\text{CD}_3\text{OD}$ ) spectrum of dudomycin B.



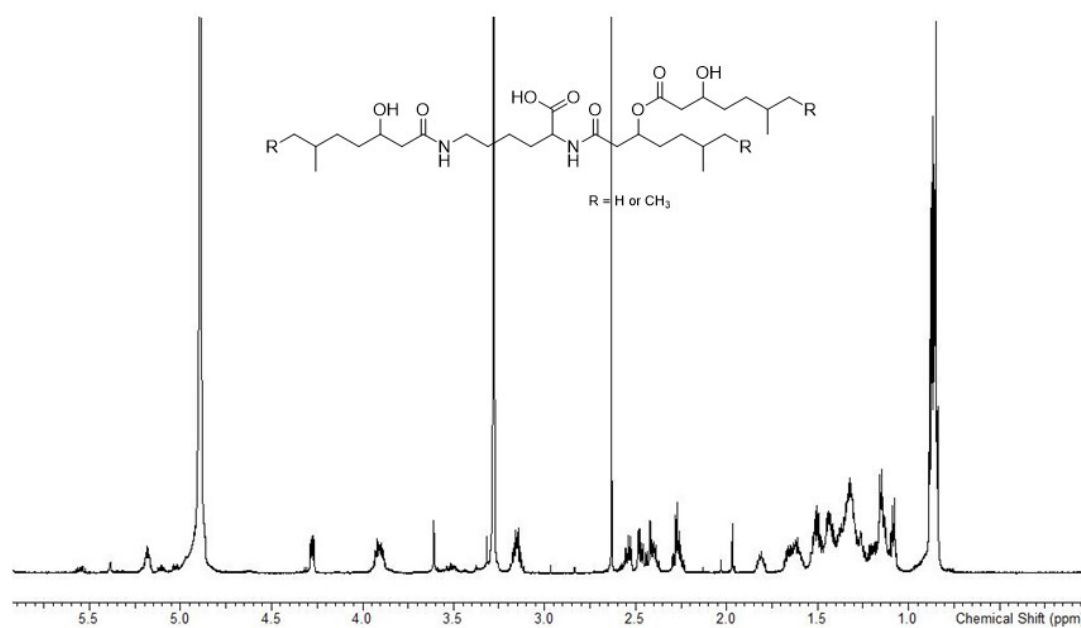
**Figure S14.** COSY (700 MHz, CD<sub>3</sub>OD) spectrum of dudomycin B.



**Figure S15.** HMBC (700 MHz,  $\text{CD}_3\text{OD}$ ) spectrum of dudomycin B.

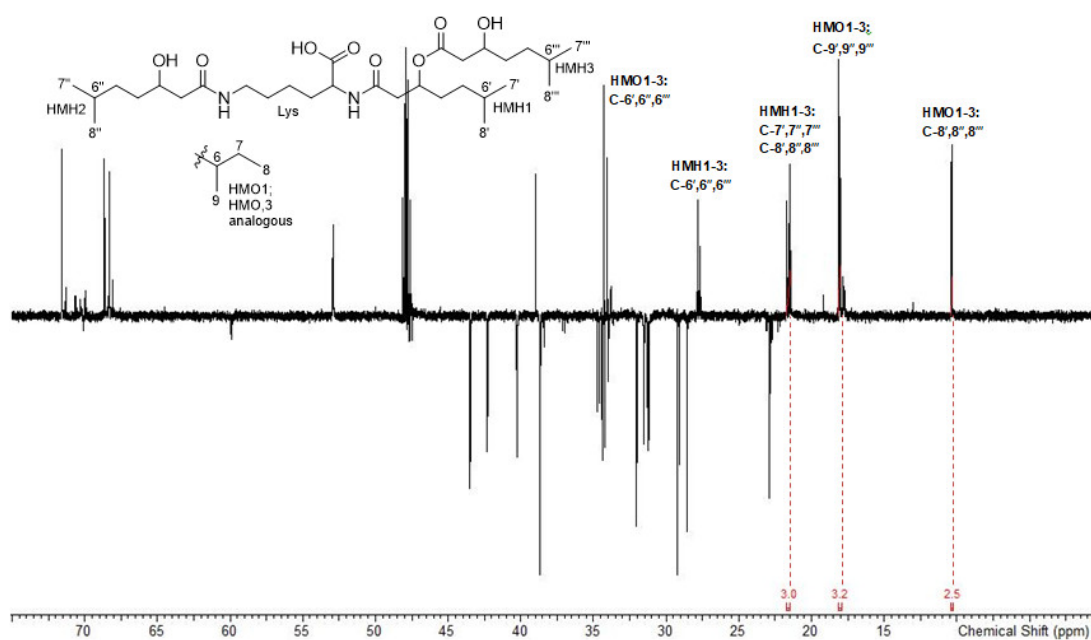


**Figure S16.** A.  $^1\text{H}$  NMR spectrum of dudomycin B showing the irradiation points B ( $\text{H-3''}/\text{H-3'''}$ ) and C ( $\text{H-3'}$ ). B-D. Selective 1D TOCSY (700 MHz,  $\text{CD}_3\text{OD}$ ) spectra of the corresponding irradiation points.

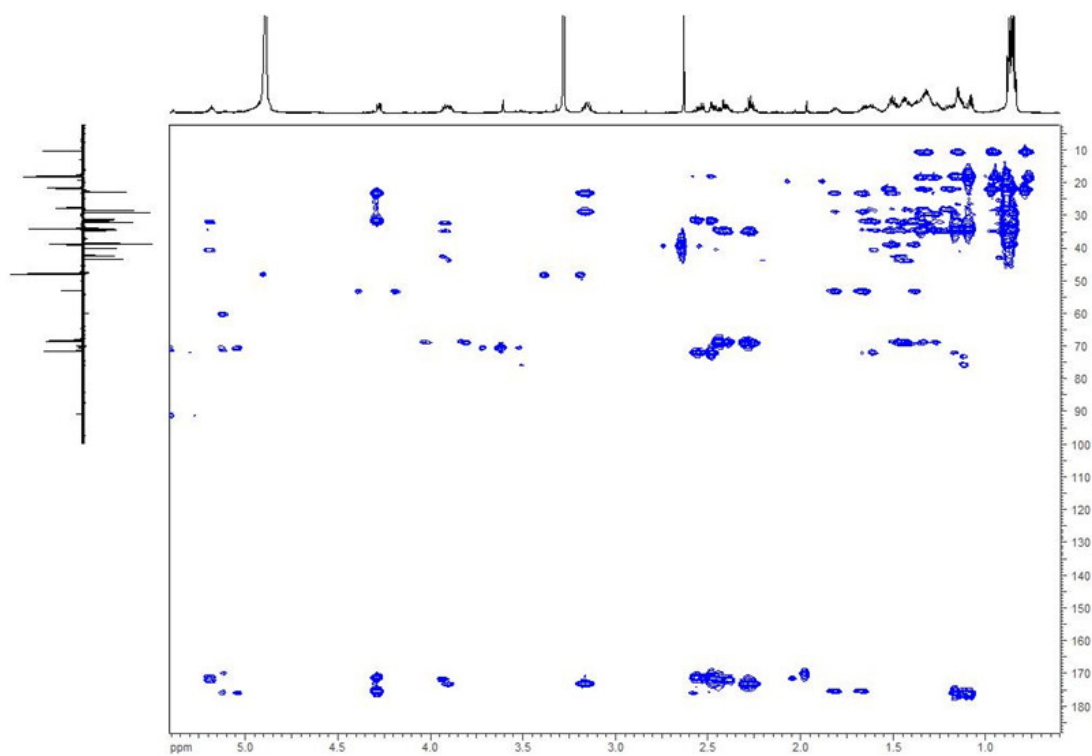


**Figure S17.**  $^1\text{H}$  NMR (700 MHz,  $\text{CD}_3\text{OD}$ ) spectrum of dudomycin C.

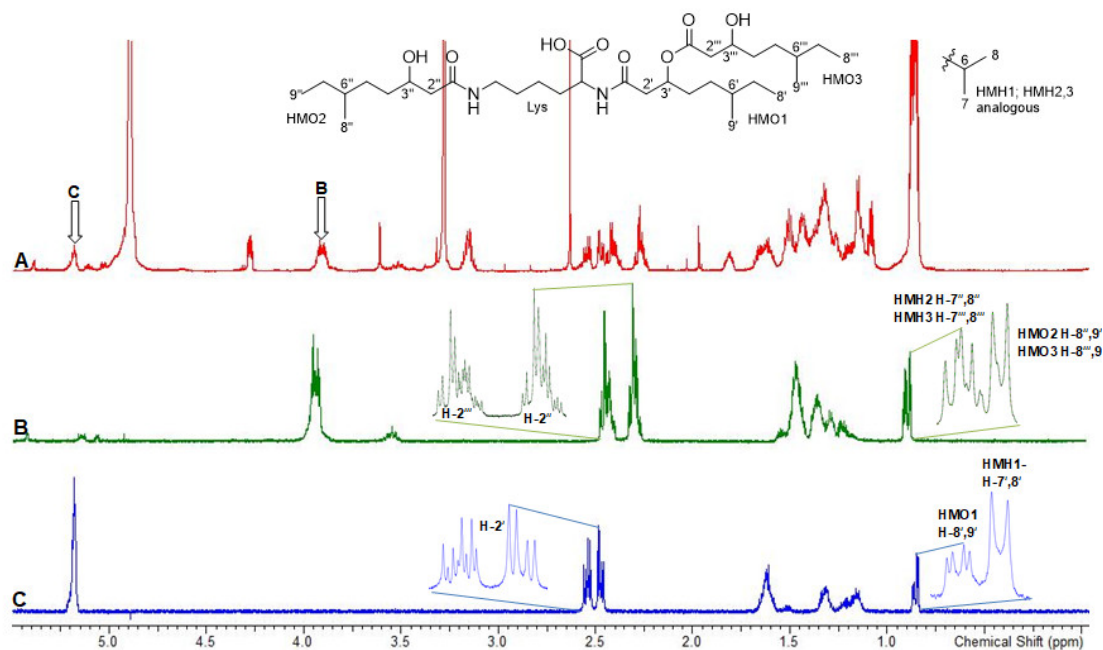




**Figure S18.** DEPT-135 (176 MHz, CD<sub>3</sub>OD) spectrum of dudomycin C showing an approximate 1/2 ratio of the methyl signals of HMH and HMO.

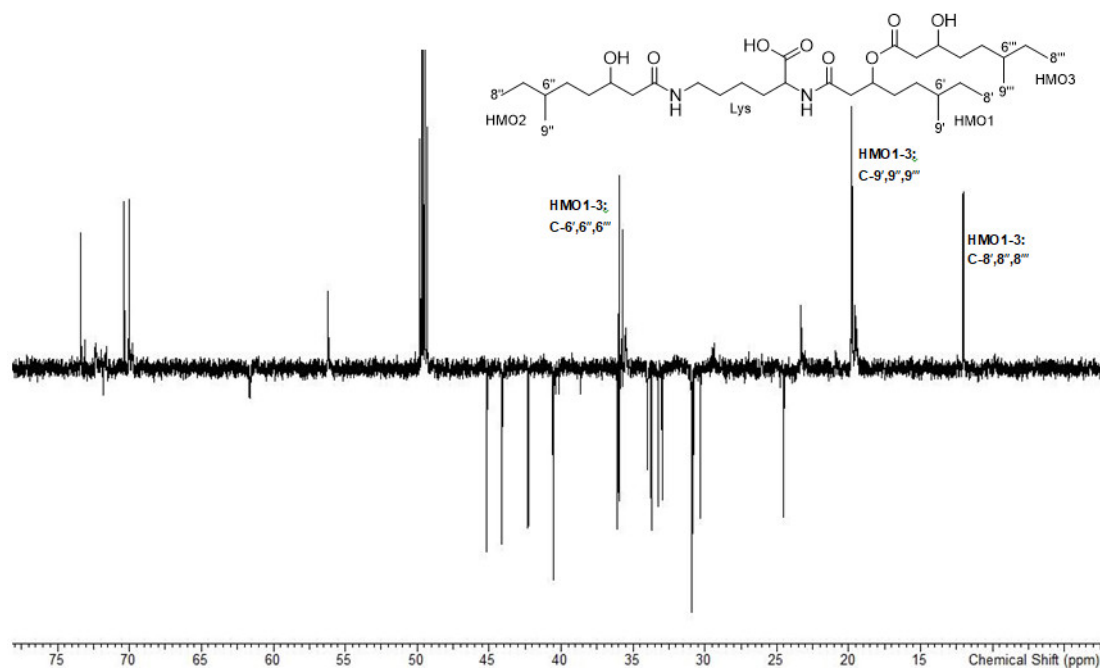


**Figure S19.** HMBC (700 MHz, CD<sub>3</sub>OD) spectrum of dudomycin C.

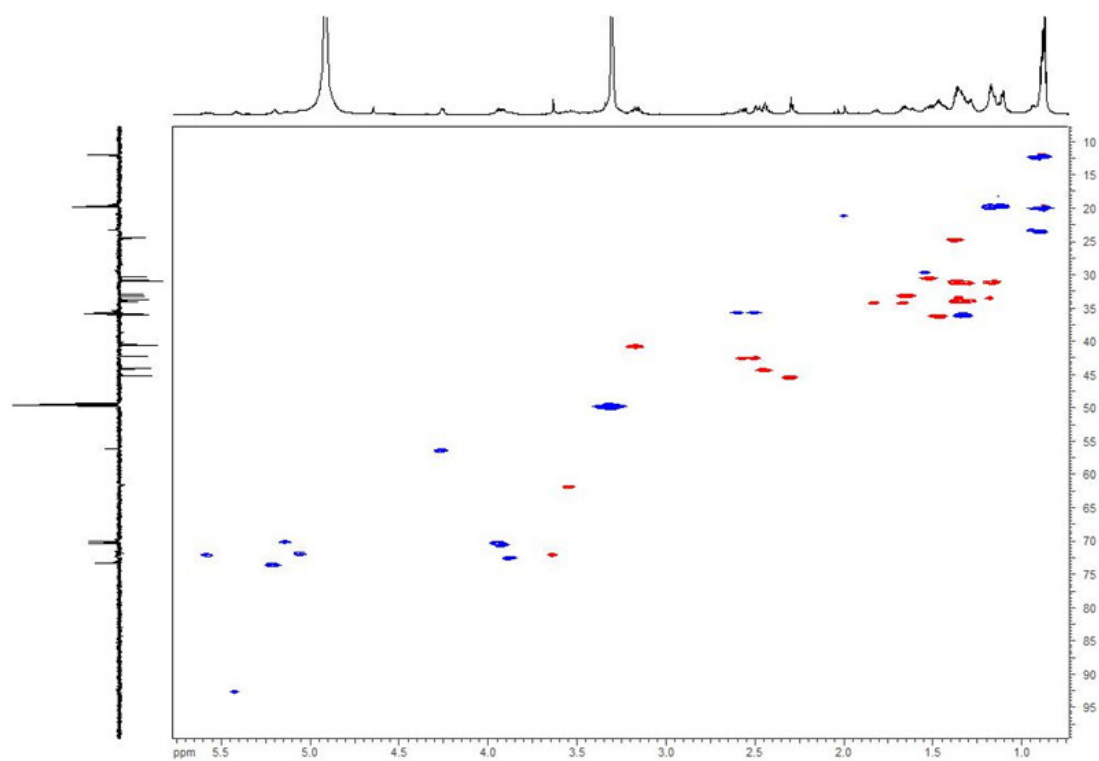


**Figure S20.** A.  $^1\text{H}$  NMR spectrum of dudomycin C showing the irradiation points B (H-3''/ H-3''') and C (H-3'). B-D. Selective 1D TOCSY (700 MHz,  $\text{CD}_3\text{OD}$ ) spectra of the corresponding irradiation points.

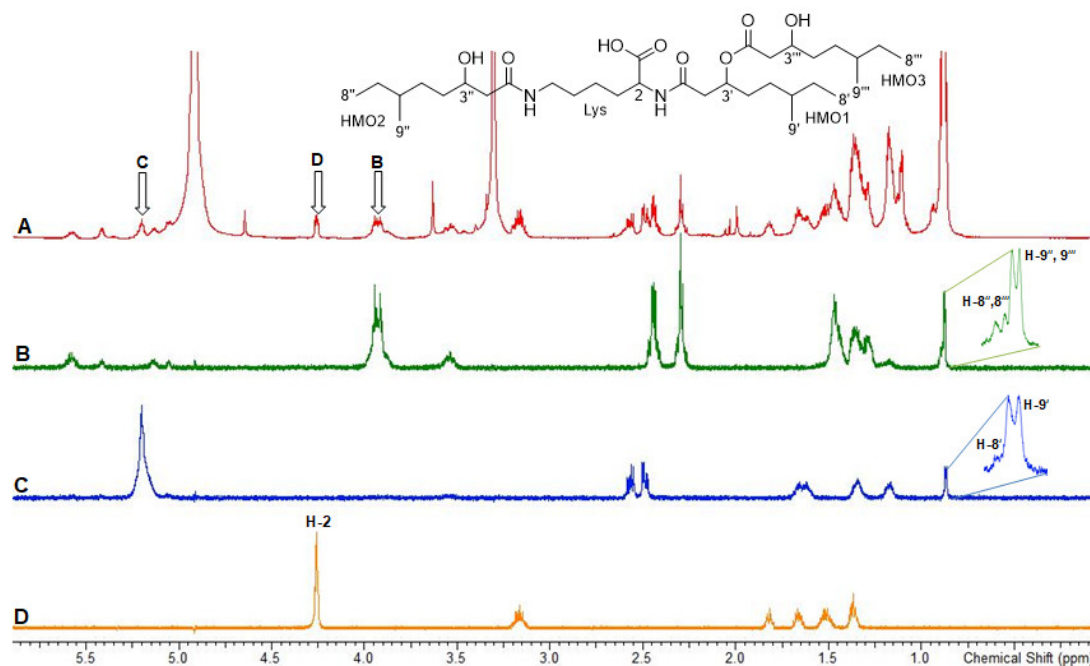




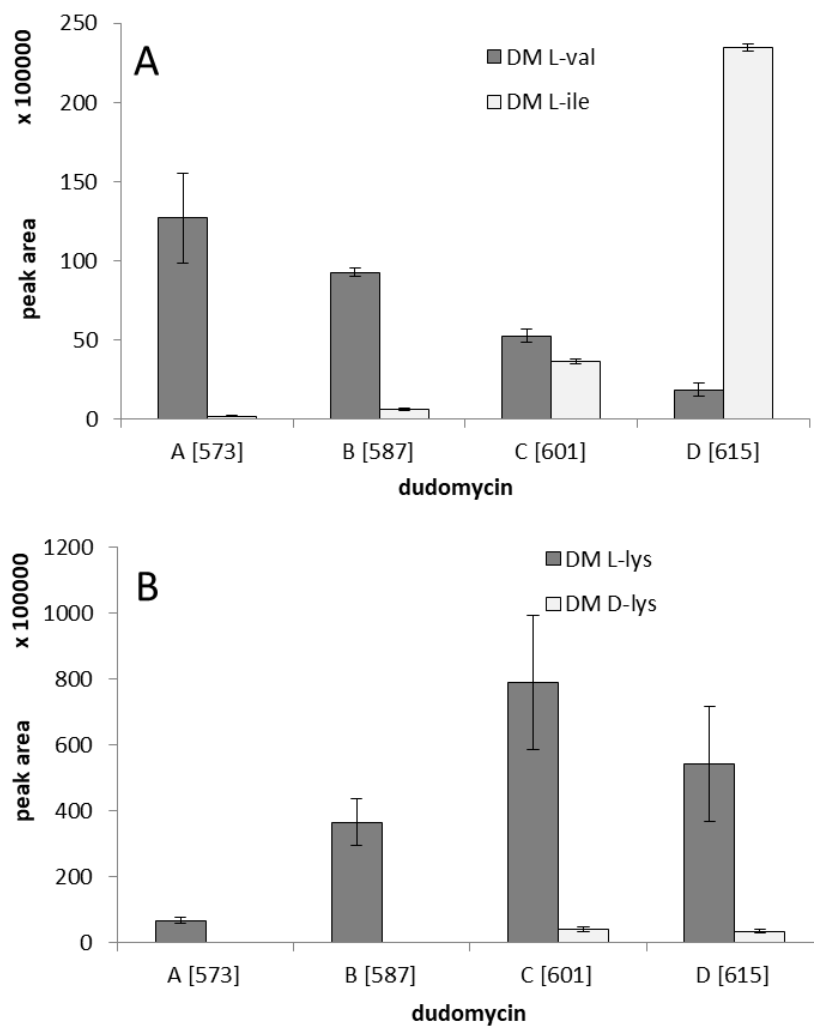
**Figure S22.** DEPT-135 (176 MHz, CD<sub>3</sub>OD) spectrum of dudomycin D.



**Figure S23.** HSQC (700 MHz,  $\text{CD}_3\text{OD}$ ) spectrum of dudomycin D.

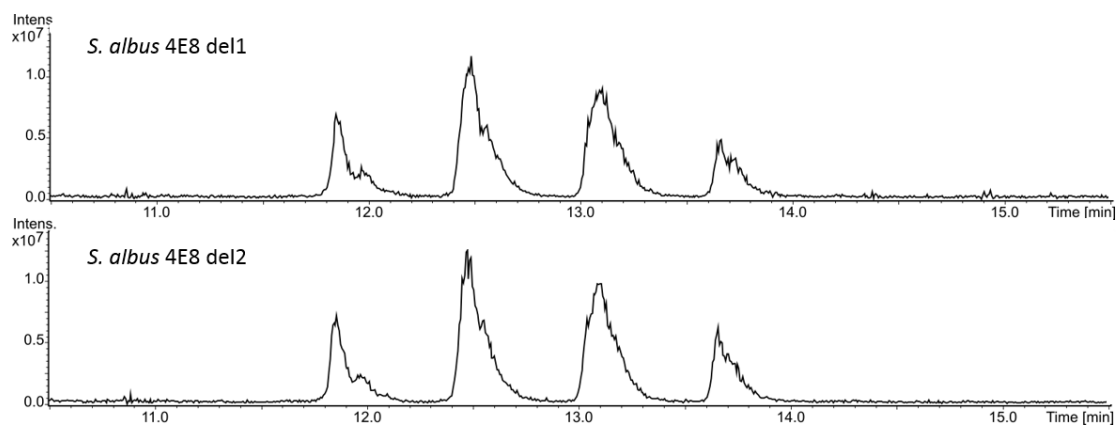


**Figure S24.** A.  $^1\text{H}$  NMR spectrum of dudomycin D showing the irradiation points B (HMO2/3, H-3''/ H-3'''), C (HMO1, H-3') and D (Lys, H-2). B D. Selective 1D TOCSY (700 MHz,  $\text{CD}_3\text{OD}$ ) spectra of the corresponding irradiation points.

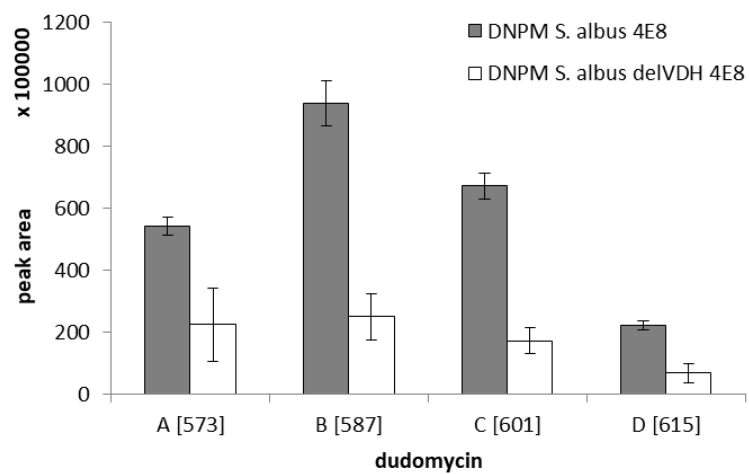


**Figure S25.** Production of dudomycin derivatives in defined medium (DM) with single sources of nitrogen.





**Figure S26.** Production of dudomycins after gene deletion experiments. BPC extracted for masses [573-574], [587-588], [601-602], [615-616].



**Figure S27.** Decreased production of dudomycin derivatives by *S. albus* delVDH 4E8.

**Table S1.** NMR data of dudomycin A in CD<sub>3</sub>OD.

unit	no.	$\delta_c$ , type	$\delta H$ , (J in Hz)
Lys	1	178.9, C	-
	2	56.2, CH	4.26, dd(4.8, 7.5)
	3	34.0, CH <sub>2</sub>	1.83, m; 1.677, m
	4	24.3, CH <sub>2</sub>	1.38, m
	5	30.2, CH <sub>2</sub>	1.51, m
	6	40.4, CH <sub>2</sub>	3.17, m
	NH-1	-	7.12, bs <sup>1</sup>
	NH-2	-	6.81, bs <sup>1</sup>
HMH1	1'	171.9, C	-
	2'	42.2, CH <sub>2</sub>	2.56, dd(7.7, 14.1); 2.50, dd(5.1, 14.4)
	3'	73.3, CH	5.20, dd(4.94, 7.5)
	4'	33.2, CH <sub>2</sub>	1.64, m
	5'	35.6, CH <sub>2</sub>	1.23, m
	6'	29.2, CH	1.54, m
	7'	23.1, CH <sub>3</sub>	0.88, d(6.5)
	8'	23.0, CH <sub>3</sub>	0.89, d(6.5)
HMH2	1''	174.3	-
	2''	44.9, CH <sub>2</sub>	2.32, dd(4.5, 13.9); 2.28, dd(8.0, 13.9)
	3''	70.2, CH	3.92, m
	4''	36.3, CH <sub>2</sub>	1.47, m
	5''	36.0, CH <sub>2</sub>	1.23, m
	6''	29.4, CH	1.56, m
	7''	23.1, CH <sub>3</sub>	0.90, d(4.4)
	8''	23.1, CH <sub>3</sub>	0.89, d(4.4)
HMH3	1'''	173.2	-
	2'''	43.9, CH <sub>2</sub>	2.47, dd(5.4, 14.8); 2.42, d(7.6, 14.9)
	3'''	69.9, CH	3.95, m
	4'''	36.2, CH <sub>2</sub>	1.49, m; 1.348, m
	5'''	36.0, CH <sub>2</sub>	1.23, m
	6'''	29.4, CH	1.56, m
	7'''	23.3, CH <sub>3</sub>	0.91, d(4.8)
	8'''	23.3, CH <sub>3</sub>	0.90, d(4.8)

<sup>1</sup>NH Signals measured in CDCl<sub>3</sub>

**Table S2.** NMR data of dudomycin D in CD<sub>3</sub>OD.

unit	no.	$\delta C^*$ , type	$\delta H$ , (J in Hz)
Lys	2	55.8, CH	4.26, dd (4.5, 6.8)
	3	33.6, CH <sub>2</sub>	1.82, m; 1.66, m
	4	24.1, CH <sub>2</sub>	1.38, m
	5	29.9, CH <sub>2</sub>	1.51, m
	6	40.1, CH <sub>2</sub>	3.17, m
HMO1	2'	41.9, CH <sub>2</sub>	2.57, dd (8.5, 14.4); 2.49, dd (4.9, 14.4)
	3'	72.9, CH	5.20, m
	4'	32.8, CH <sub>2</sub>	1.64, m
	5'	32.5, CH <sub>2</sub>	1.36, m; 1.18, m
	6'	35.3, CH	1.34, m
	7'	30.4, CH <sub>2</sub>	1.33, m; 1.17, m
	8'	11.6, CH <sub>3</sub>	0.88, ovl
	9'	19.3, CH <sub>3</sub>	0.87, d (6.0)
HMO2	2''	44.7, CH <sub>2</sub>	2.31, dd (4.9, 13.5); 2.28, dd (8.2, 13.6)
	3''	69.9, CH	3.92, m
	4''	35.6, CH <sub>2</sub>	1.47, m
	5''	33.3, CH <sub>2</sub>	1.30, m; 1.37, m
	6''	35.5, CH	1.34, m
	7''	30.5, CH <sub>2</sub>	1.35, m; 1.37, m
	8''	11.6, CH <sub>3</sub>	0.89, ovl
	9''	19.4, CH <sub>3</sub>	0.88, d (5.4)
HMO3	2'''	43.7, CH <sub>2</sub>	2.46, dd (5.5, 15.1); 2.424, d (8.2, 15.1)
	3'''	69.6, CH	3.94, m
	4'''	35.5, CH <sub>2</sub>	1.47, m
	5'''	33.3, CH <sub>2</sub>	1.30, m; 1.37, m
	6'''	35.5, CH	1.34, m
	7'''	30.5, CH <sub>2</sub>	1.35, m; 1.37, m
	8'''	11.6 CH <sub>3</sub>	0.89, ovl
	9'''	19.4, CH <sub>3</sub>	0.88, d (5.4)

\*from DEPT-135 measurement

**Table S3.** Strains, BACs, plasmids and primers used in this work.

Material	Purpose
<b>A. Bacterial strains</b>	
<i>Streptomyces albus</i> Del14	optimized heterologous host [1]
<i>Streptomyces lividans</i> Del8	optimized heterologous host [2]
<i>Streptomyces albus</i> J1074 delVDH	heterologous host [3]
<i>Escherichia coli</i> GB05 RedCC	cloning host [Helmholtz-Institut für Pharmazeutische Forschung Saarland (HIPS)]
<i>Escherichia coli</i> ET12567 pUB307	alternate host intergeneric conjugation [4]
<b>B. BACs</b>	
4E8	heterologous expression of NRPS cluster
4E8 del1	determination left border of NRPS cluster
4E8 del2	determination right border of NRPS cluster
<b>C. Plasmids</b>	
pUC19	ampicillin resistance marker
pXCM hygformax	hygromycin resistance marker
<b>D. PCR primer</b>	
20190429_1_fw [4E8 del1]	CAGGGAGGAGGCCACCCCGCGCCGGTAGAAGACGCCCAGCAGGCCGACCGCG TCAGGTGGCACTTTTCG
20190429_1_rev [4E8 del1]	GATCGCGTGCATGCCCAGATGGTCGCGCGCATACCGCCGTCGGAACGGGCGCG GATATCCTACTATGCCGAGGTATAATGTAGCCAGCGTGTTACCAATGCTTAATC AGTG
20190429_2_fw [4E8 del2]	GGCGACGGCTCCCCGGGCCCCGCGAGCCACGACTCCCGTCCCCACGGCCATCA GGCGCCGGGGGCGGTGT
20190429_2_rev [4E8 del2]	GCGATGAGTATCCGTACTCATGTCCGGGCCGCGGGTGCCGCCTAGCGTGAAAT ACTTGACATATCACTGT

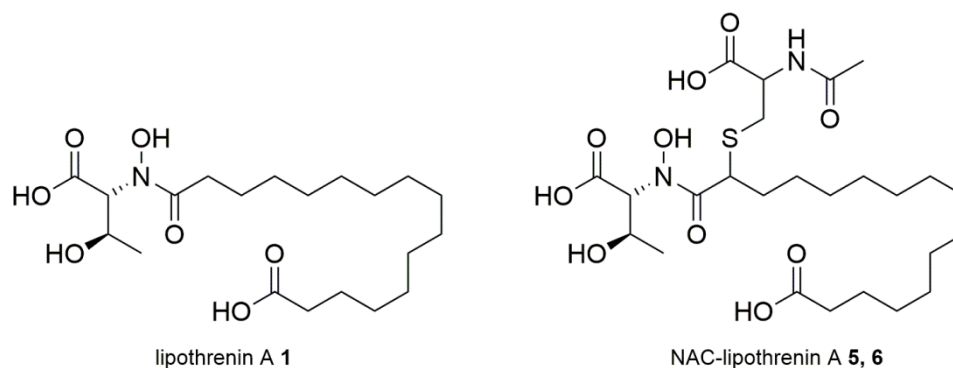
Table S4. Proposed functions of the genes in the DNA fragment containing the dudomycin gene cluster

gene #	locus tag	putative function
1	SACHL_42600	glycosyltransferase
2	SACHL_42610	hypothetical protein
3	SACHL_42620	phosphatase
4	SACHL_42630	N,N'-diacetyllegionaminic acid synthase
5	SACHL_42640	hypothetical protein
6	SACHL_42650	hydrolase
7	SACHL_42660	membrane lipoprotein precursor
8	SACHL_42670	galactose/methyl galactoside import ATP-binding protein
9	SACHL_42680	ribose transport system permease protein
10	SACHL_42690	branched-chain amino acid transport system / permease component
11	SACHL_42700	cytidine deaminase
12	SACHL_42710	pyrimidine-nucleoside phosphorylase
13	SACHL_42720	hypothetical protein
14	SACHL_42730	hypothetical protein
15	SACHL_42740	hypothetical protein
16	SACHL_42750	hypothetical protein
17	SACHL_42760	hypothetical protein
18	SACHL_42770	hypothetical protein
19	SACHL_42780	ubiquinone biosynthesis O-methyltransferase
20	SACHL_42790	zinc metallo-peptidase
21	SACHL_42800	hypothetical protein
22 [dudA]	SACHL_42810	dimodular nonribosomal peptide synthase
23 [dudB]	SACHL_42820	inner membrane transport protein
24 [dudC]	SACHL_42830	transcriptional regulator
25	SACHL_42840	demethylrebeccamycin-D-glucose O-methyltransferase
26	SACHL_42850	hypothetical protein

## REFERENCES

- [1] M. Myronovskyi, B. Rosenkränzer, S. Nadmid, P. Pujic, P. Normand, and A. Luzhetskyy, "Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters," *Metab. Eng.*, vol. 49, pp. 316–324, Sep. 2018, doi: 10.1016/j.ymben.2018.09.004.
- [2] Y. Ahmed, Y. Rebets, M. R. Estévez, J. Zapp, M. Myronovskyi, and A. Luzhetskyy, "Engineering of *Streptomyces lividans* for heterologous expression of secondary metabolite gene clusters," *Microb. Cell Fact.*, vol. 19, no. 1, p. 5, Dec. 2020, doi: 10.1186/s12934-020-1277-8.
- [3] N. Manderscheid, "Strain development for heterologous expression of secondary metabolite clusters in actinobacteria," Universität des Saarlandes, 2015.
- [4] F. Flett, V. Mersinias, and C. P. Smith, "High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes," *FEMS Microbiol. Lett.*, vol. 155, no. 2, pp. 223–229, Jan. 2006, doi: 10.1111/j.1574-6968.1997.tb13882.x.

### 3 Lipothrenins - N-hydroxylated threonine-hexadecanedioic-acids produced after heterologous expression of a fatty acid biosynthetic gene cluster



**Figure 1:** Two main compounds lipothrenin A and NAC-lipothrenin A after heterologous expression of a FA BGC.

#### 3.1 Introduction

The discovery of cryptic biosynthetic gene clusters (BGCs) in *Streptomyces coelicolor* two decades ago revealed a treasure chest of new natural products (NPs) hidden in the genome of streptomycetes.<sup>1</sup> Activation of cryptic BGCs by heterologous expression is one of the tools to access the encoded compounds.<sup>2, 3</sup> Selection of promising strains with biosynthetic gene clusters for heterologous expression can be guided by dereplication to get an insight to the range of known and unknown compounds of a strain. We screened a library of less studied *Streptomyces* strains by dereplication and genome mining to identify strains with a high potential to produce new NPs. From this library we selected *Streptomyces aureus* LU18118, since it produces a variety of compounds with unknown masses. A BAC library constructed for *S. aureus* LU18118 was expressed in the optimized host strains *S. lividans*  $\Delta 8$  and *S. albus*  $\Delta 14$  to determine BGCs which are responsible for the production of the new compounds. Heterologous expression of a fatty acid biosynthesis annotated gene cluster led to the production of new compounds with unidentified masses in both heterologous host strains *S. lividans*  $\Delta 8$  and *S. albus*  $\Delta 14$ . The encoded compounds from the BGC were identified as new lipothrenins (Figure 1). Lipothrenins consist of the core elements threonine and hexadecanedioic acid which are linked by an amide bond. Various modifications such as N-hydroxylation, dehydration and N-acetylcysteinylation (NAC) led to a total amount of 14 derivatives. The BGC borders and the biosynthesis were determined by deletion experiments. By this, a ketosynthase encoding gene, specific for the synthesis of linear FA compounds, was identified. Furthermore, stereospecific N-

oxygenation of the amid-nitrogen *via* an encoded diiron dioxygenase was demonstrated. Finally, a biosynthetic pathway towards lipothrenin and NAC-lipothrenins was proposed.

## **3.2 Materials and Methods**

### **3.2.1 General Experimental Procedures**

All strains and bacterial artificial chromosomes (BACs) used in this work are listed in (Table S1). *Escherichia coli* strains were cultured in LB medium.<sup>4</sup> *Streptomyces* strains were grown on soya flour mannitol agar (MS agar)<sup>5</sup> or in liquid tryptic soy broth (TSB; Sigma-Aldrich, St. Louis, MO, USA) for cultivation. Liquid DNPM medium (40 g/L dextrin, 7.5 g/L soytone, 5 g/L baking yeast, and 21 g/L MOPS, pH 6.8) was used for secondary metabolite expression. The antibiotics kanamycin, apramycin, ampicillin and nalidixic acid were supplemented when required.

### **3.2.2 Metabolite Extraction and Analysis by Mass Spectrometry (MS)**

*Streptomyces* strains were grown in TSB medium for 24 hours. Subsequently, the strains were inoculated in DNPM production medium and grown for 6 days at 28°C and 180 rpm. Metabolites were extracted from the supernatant with n-butanol. The extract dissolved in methanol was separated on a Dionex Ultimate 3000 UPLC system (Thermo Fisher Scientific, Waltham, MA, USA) equipped with an ACQUITY UPLC BEH C18 1.7  $\mu$ m column (30, 50 or 100 mm, Waters Corporation, Milford, MA, USA) using a linear gradient of 5-95 vol% aqueous acetonitrile (ACN) with 0.1 % formic acid (FA) at a flow rate of 0.6 mL/min and a column oven temperature of 45°C. Sample analysis was carried out on a coupled PDA detector followed by an amaZon speed (Bruker, Billerica, MA, USA), for production control. High-resolution masses were obtained from either an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) or a maXis high-resolution LC-QTOF system (Bruker, Billerica, MA, USA) using positive ionization mode and mass range detection of m/z 200 to 2000. Data analysis was performed using Compass Data Analysis v. 4.1 (Bruker) and Xcalibur v. 3.0 (Thermo Fisher Scientific).

### **3.2.3 Nuclear Magnetic Resonance (NMR) Spectroscopy and Optical Rotation (OD)**

NMR spectra were recorded on a Bruker Avance I 500 MHz (Bruker, BioSpin GmbH, Rheinstetten, Germany) equipped with a 5 mm BBO probe at 298 K. The chemical shifts ( $\delta$ ) were reported in parts per million (ppm) relative to TMS. As solvents, deuterated DMSO-*d*<sub>6</sub> ( $\delta_{\text{H}}$  2.50 ppm,  $\delta_{\text{C}}$  39.51 ppm) from Deutero (Kastellaun, Germany) was used. Edited-HSQC, HSQC-TOCSY, HMBC, <sup>1</sup>H-<sup>1</sup>H COSY, ROESY, NOESY and N-HSQC spectra were recorded using the standard pulse programs from TOPSPIN v.2.1 software. Optical rotations were measured on a Perkin Elmer Polarimeter Model 241 (Überlingen, Deutschland).



### 3.2.4 General isolation procedure of lipothrenins and NAC-lipothrenins

*Streptomyces* strains were grown in 5-10 L of DNPM production medium and extracted with *n*-butanol. The dry crude extract was dissolved in 100 mL methanol. The isolation process was guided by LC-MS to identify fractions. The extract was purified on a Isolera™ One flash purification system equipped with a Chromabond RS330 C<sub>18</sub> ec (Macherey-Nagel, Düren, Germany) using a gradient of 5-35 vol% aqueous methanol for 1 column volume (CV) followed by 35-80 vol% aqueous methanol for 7 CV at a flow rate of 100 mL/min and UV detection at 210 and 280 nm. Further impurities were removed by size exclusion chromatography (SEC; stationary phase: Sephadex-LH20) with isocratic elution using methanol.

Fractions with compounds were dissolved in methanol and reversed-phase HPLC was carried out on a Waters Autopurification System (Waters Corporation, Milford, MA, USA) with a SQD2-MS-Detector and equipped with a preparative VP 250/21 NUCLEODUR C<sub>18</sub> HTec 5 µm column (MACHERY NAGEL, Düren, Germany) using fraction specific gradients of aqueous methanol with 0.1% FA at a flow rate of 20 mL/min.

Final purification steps were carried out on an Agilent Infinity 1100 series reversed-phase HPLC system equipped with a Synergi™ 4 µm Fusion-RP C<sub>18</sub> 80 Å 250x10 mm column (Phenomenex, Torrance, CA, USA) using fraction specific gradients of aqueous ACN with 0.1% FA, a flow rate of 4 mL/min and an oven temperature of 20°C.

### 3.2.5 Marfey's Analysis

Lipothrenins were hydrolyzed in 100 µL 6 N HCl at 110°C for 1 hour. While cooling down, the sample was dried for 15 min under nitrogen, dissolved in 110 mL water and 50 µL each were transferred into 1.5 mL Eppendorf tubes. Next, 20 µL of 1N NaHCO<sub>3</sub> and 20 µL of 1% L-FDLA (N<sup>α</sup>-(5-Fluoro-2, 4-dinitrophenyl)-L-leucinamide) or D-FDLA in acetone were added to the hydrolysate, respectively. The amino acid standards were prepared the same way only using L-FDLA. The reaction mixtures were incubated at 40°C for 90 min at 700 rpm and subsequently quenched with 2N HCl to stop the reaction. The samples were diluted with 300 µL ACN and 1 µL was analyzed by MaXis high-resolution LC-QTOF system using aqueous ACN with 0.1 vol% FA and an adjusted gradient of 5-10 vol% in 2 min, 10-25 vol% in 13min, 25-50 vol% in 7 min and 50-95 vol% in 2 min. Sample detection was carried out at 340 nm.

### 3.2.6 Isolation and Manipulation of DNA

BAC extraction from a *S. aureus* LU18118 constructed genomic library (Intact Genomics, USA), DNA manipulation, *E. coli* transformation, and *E. coli*/*Streptomyces* intergeneric conjugation was performed according to standard protocols.<sup>4-6</sup> Plasmid DNA was purified with the BACMAX™ DNA

purification kit (Lucigen, Middleton, WI, USA). Restriction endonucleases were used according to manufacturer's recommendations (New England Biolabs, Ipswich, MA, USA).

BAC I6 derivatives with gene deletions were constructed using the RedET approach. For this, the ampicillin marker from pUC19 was amplified by PCR with primers harboring overhang regions complementary to the boundaries of the DNA to be deleted. Recombineering of the BAC was performed with the amplified fragments. The recombinant BACs were analyzed by restriction mapping and PCR. The primers used for recombineering purposes are listed in Table S2. Primers for sequencing the correct mutants are listed in Table S3.

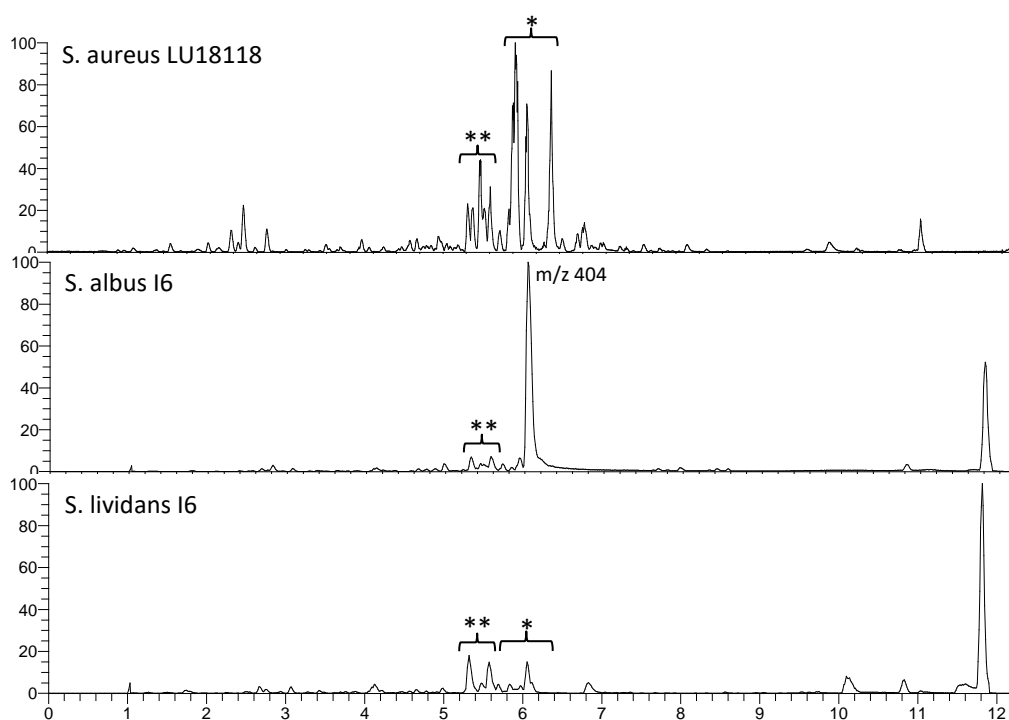
### 3.2.7 Genome Mining and Bioinformatics Analysis

The *S. aureus* LU18118 genome was screened for secondary metabolite biosynthetic gene clusters using the antiSMASH<sup>7</sup> online tool (<https://antismash.secondarymetabolites.org/#!/start>), basic local alignment search tool (BLAST)<sup>8</sup> and Geneious 11.0.3 software.<sup>9</sup>

## 3.3 Results and Discussion

### 3.3.1 Identification and expression of the lipothrenin gene cluster

From a strain library with less studied streptomyces, *S. aureus* LU18118 stood out due to the production of many compounds with unknown masses identified by dereplication (Figure 2). The genome of *S. aureus* LU18118 was analyzed *in silico* by antiSMASH to select BGCs encoding potentially new compounds. From a BAC library constructed for *S. aureus* LU18118, 13 BACs that cover 21 unknown BGCs were chosen for the heterologous expression in the optimized host strains *S. lividans* Δ8 and *S. albus* Δ14. The BACs were transferred into the heterologous hosts and the mutants were cultivated in DNPM production medium, extracted with n-butanol and the production of new metabolites was screened by LC-MS. Only BAC I6 was successfully expressed in both strains and led to the production of the unknown compounds which have also been produced by *S. aureus* LU18118. The new compounds can be divided into two groups due to the mass differences of 16 Da and 18 Da:  $m/z$  ( $[M+H]^+$ ) 404.266 (**1**), 388.271 (**2, 3**), 370.260 (**4**) Da and  $m/z$  565.278 (**5, 6**), 549.284 (**7-10**) and 531.272 (**11-14**) Da (Figure 2). Multiple peaks with same masses suggest the presence of isomers. All compounds showed UV/VIS bands at 204 nm and mass differences equivalent to oxygen or water. The new compounds were isolated from the wild type strain in parallel to the heterologous expression. All compounds were obtained in good amounts; thus, isolation from the heterologous hosts was not necessary.



**Figure 2:** LC-MS chromatograms of n-butanol extracts of *S. aureus* LU18118, *S. albus* I6 and *S. lividans* I6. All strains showed production of two groups of masses  $[M+H]^+$ : 370-404 Da (\*) and 531-565 Da (\*\*). In *S. albus* I6 m/z 404 Da was the main peak.

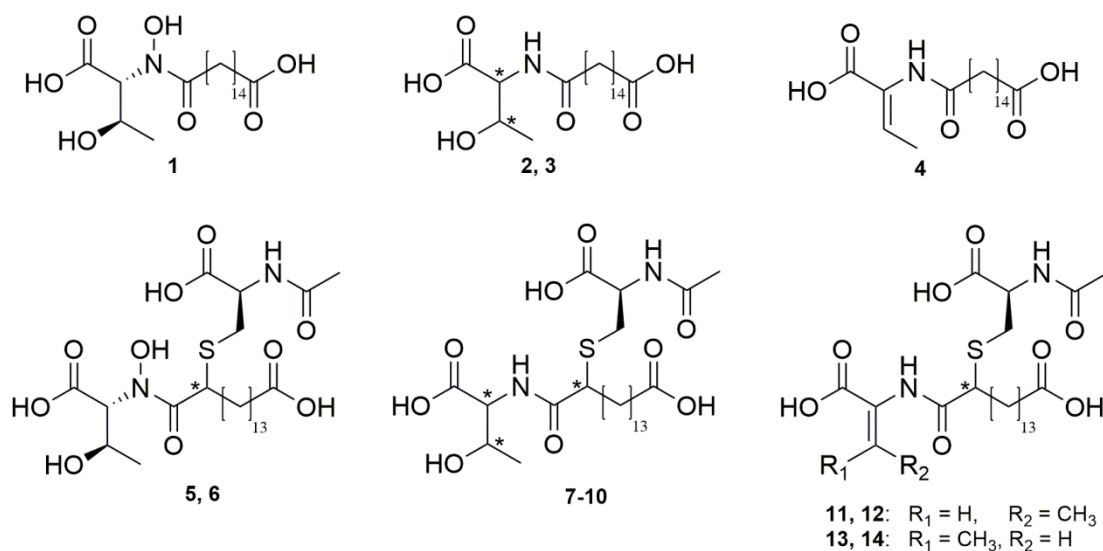
### 3.3.2 Isolation and structure elucidation of lipothrenins

*S. aureus* LU18118 was cultivated in 10 L DNPM medium and the metabolites were extracted with n-butanol. The crude extract was further purified by reversed-phase flash purification and size exclusion chromatography. The final HPLC purification step led to the successful isolation of compounds **1-14**. All structures (Figure 3) were determined by extensive analysis of 1D and 2D NMR (Table S4-S13, Figure S7-S55).

The molecular formula of **1** was calculated as  $C_{20}H_{37}NO_7$  with 3 degrees of unsaturation based on m/z 403.259 Da. Proton NMR data revealed a very large singlet at  $\delta_H$  1.25, which is common for the long methylene chain of fatty acids. The classic high field shifted doublet of the  $\omega$ -methyl group was absent, indicating another moiety at this position. HMBC correlations from the FA-chain associated methylenes  $CH_2$ -14 and  $CH_2$ -15 to  $\delta_C$  174.5 revealed a carboxyl group in  $\omega$ -position. Methylene  $CH_2$ -2 showed peak splitting, indicating a rigid structure attached to the adjacent carboxylate  $\delta_C$  173.5 (C-1). The remaining proton and carbon signals were assigned to one methyl, two methines and a carbon signal at  $\delta_C$  171.6. The spin system was assigned to threonine, which showed a connection to the FA-chain by an HMBC correlation of the  $\alpha$ -CH to  $\delta_C$  173.5. Based on the molecular formula and the remaining  $^{13}C$ -NMR signals, the FA-chain was determined to possess 14 methylenes and 2 carboxyl groups which led to hexadecanedioic acid. Despite using DMSO- $d_6$  as a solvent, no NH proton signals were observed. The unusually high shifted  $\alpha$ -CH at  $\delta_C$  63.3 and the remaining unassigned oxygen

from the calculated formula suggests that threonine carries an N-hydroxyl group. The previously described NMR data for N-hydroxyl threonine are in accordance to our data.<sup>10</sup> Threonine might be able to adopt a six-membered conformation stabilized through a hydrogen bridge between N-OH and the carboxyl group of threonine. The resulting rigid conformation would explain the previously observed peak splitting of methylene C-2 of FA. The new compound has no similarity to other lipo-amino acids and was named lipothrenin A (**1**).

Compounds **2** and **3** were purified as a mixture. The molecular formula of **2** and **3** was calculated as  $C_{20}H_{37}NO_6$  with 3 degrees of unsaturation based on  $m/z$  387.262 Da. Proton NMR revealed a mixture of two compounds highly similar to **1**, but showing additional NH signals at  $\delta_H$  7.86 and 7.66. Lipothrenins **2** and **3** showed differences in the shifts and coupling constants of the threonine moiety, indicating a mixture of two stereoisomers. This was analyzed by Marfey's method.<sup>11</sup> Compounds **1**, **2** and **3** were hydrolyzed with 6N hydrochloric acid and the hydrolysis product, together with the reference amino acids L-Thr, D-Thr, *L-allo*-Thr and *D-allo*-Thr, were derivatized with L-FDLA. Subsequent LC-MS measurements identified *D-allo*-Thr in lipothrenin A (**1**), while in the mixture of **2** and **3** *D-allo*-Thr and L-Thr were identified (Figure S5). Compared to **1**, the compounds lack the N-hydroxyl and were named D-lipothrenin B (**2**) and L-lipothrenin B (**3**).



	lipothrenin	NAC-lipothrenin*
D- <i>allo</i> -Thr, N-OH	<b>1</b>	<b>5, 6</b>
D- <i>allo</i> -Thr	<b>2</b>	<b>7, 9</b>
L-Thr	<b>3</b>	<b>8, 10</b>
Z-Dhb	<b>4</b>	<b>11, 12</b>
E-Dhb		<b>13, 14</b>

\*two isomers due to additional stereocenter at fatty acid chain

**Figure 3:** Structures of all lipothrenin and NAC-lipothrenin derivatives.

The molecular formula of **4** was calculated as  $C_{20}H_{35}NO_5$  with 4 degrees of unsaturation based on  $m/z$  369.252 Da. In the  $^{13}C$ -NMR spectra, two signals appeared at  $\delta_C$  129.0 and 130.6, while the  $\alpha$ - and  $\beta$ -CH groups in the  $^1H$  NMR spectrum vanished. The new spin system was established by HMBC and ROESY correlations of the NH (Table S6), revealing Z-dehydrobutyrine (Dhb). The compound was named Z-lipothrenin C (**4**).

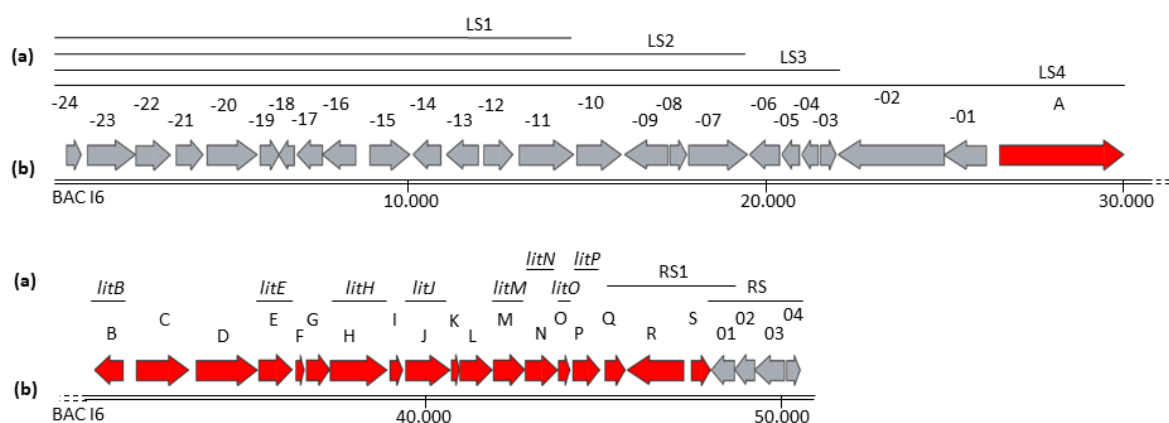
The molecular formulae of the heavier compounds were calculated as  $C_{25}H_{44}N_2O_{10}S$  (**5**, **6**),  $C_{25}H_{44}N_2O_9S$  (**7-10**) and  $C_{25}H_{42}N_2O_8S$  (**11-14**) based on  $m/z$  ( $[M+H]^+$ ) 564.271 (**5**, **6**), 548.277 (**7-10**) and 530.266 (**11-14**) Da. Molecular formulae of the heavier compounds **5-14** showed a difference of  $C_5H_7NO_3S$  compared to the lighter compounds **1-4**, indicating both groups are structurally related. Analysis of  $^1H$  NMR spectra of **5** and **6** revealed the previously described signals for lipothrenin A (**1**), and additional ones at  $\delta_H$  8.21 (NH), 4.31 (CH), 2.97 ( $CH_2a$ ), 2.74 ( $CH_2b$ ) and 1.84 ( $CH_3$ ). The signals were assigned to N-acetyl cysteine (NAC), which was determined at position CH-2 of the FA chain through analysis of COSY and HMBC correlations. Compounds **7-10** showed the same N-acetyl cysteine modification and were assigned as the NAC derivatives of **2** and **3**.

The NAC modifications introduce two additional stereocenters at  $\alpha$ -CH of cysteine or at CH-2 of the fatty acid that would lead 4 diastereomers. The appearance of two diastereomeric pairs of NAC-lipothrenins (see Figure 3) suggests only one of the additional stereocenters is present in both orientations. By using Marfey's method, we tried to identify the conformation of threonine and cysteine. NAC-lipothrenins **7-10** were digested by 6N HCl, derivatized with D- and L-FDLA and analyzed by LC-MS. Compounds **7** and **9** revealed D-*allo*-Thr as the constituent amino acid while **8** and **10** contained L-Thr. This agrees with the previous results found for **2** and **3**. Analysis of the coupling constant of CH- $\alpha$  and CH- $\beta$  showed a difference between D-*allo*-Thr ( $J = 3-3.5$ ) and L-Thr ( $J = 6-6.3$ ). The specific coupling constants of D-*allo*-Thr and L-Thr were used to retrospectively annotate the correct stereochemistry of **2** and **3**. The sulphur bond to C-2 was not cleaved by acid hydrolysis with HCl. This was expected since reductive conditions are required.<sup>12</sup> Therefore, we looked for the 2, 4-dinitrophenyl-L-leucinamide (DLA)-cys-FA conjugate (Figure S6) to investigate if there are any differences in the retention time of DLA-D-cys-FA and DLA-L-cys-FA. Derivatization of the hydrolysates **7-10** with L-FDLA did not lead to retention time differences, when observing the extracted mass of DLA-cys-FA. The same result was obtained after derivatization with D-FDLA, indicating that only one cysteine isomer is present in NAC-lipothrenin. However, there was a difference between L- and D-FDLA derivatized cys-FA, indicating Marfey's method also applies for the cys-FA conjugant (Figure S6). The retention time of L-DLA-cys-FA was shorter compared to the D-DLA-cys-FA conjugant. Therefore, cysteine is likely in L-configuration and the appearance of stereoisomers originates in the  $\alpha$ -CH of FA. Compounds **11-14** were assigned as the N-acetyl-cysteine derivatives of lipothrenin C. NAC-lipothrenins **11** and **12** eluted simultaneously and appeared as a mixture in NMR

measurements. The same result was obtained for **13** and **14**. Differences between **11/12** and **13/14** were determined in the chemical shift of CH- $\beta$  of Dba. Similar to **4**, compounds **11/12** showed a shift of  $\delta_c$  131.8 indicating Z-configuration. Compounds **13/14** showed a shift of  $\delta_c$  125.0, which could be due to isomerization to the E-conformer. This was confirmed by comparisons with predicted data from ACD labs ( $\delta_c(E) = 124.0$ ,  $\delta_c(Z) = 130.4$ ).

### 3.3.3 Identification of the lipothrenin biosynthetic genes

BAC I6 contains a 51 kb chromosomal fragment of *S. aureus* LU18118. The annotated FA cluster comprises 33 kb of the fragment, thus many genes might not take part in the biosynthesis. To experimentally determine the minimal set of lipothrenin biosynthetic genes, we performed sequential gene deletions of the left and right flanking regions of the chromosomal fragment. Single gene deletions within the minimal set were performed to selectively investigate biosynthetic functions (Figure 4). Modified BACs were transferred in *S. albus*  $\Delta$ 14 leading *S. albus* I6\_ $\Delta$  mutant strains, from hereon called I6\_ $\Delta$  for sake of simplicity (Table S1). Mutant strains were cultivated in DNPM and metabolite production was monitored by LC-MS.



**Figure 4:** Chromosomal fragment of *S. aureus* LU18118 with the lipothrenin biosynthetic genes. (a) Overview of the performed deletion within BAC I6; (b) All genes encoded within the chromosomal fragment; the minimal required genes for lipothrenin biosynthesis are marked red.

#### 3.3.3.1 Determination of the minimal set of lipothrenin biosynthesis genes

The FA cluster location was annotated by antiSMASH between 16 kb and 49 kb in the BAC insert. At the right flank, *litQ-litS* and *orf 01* encode genes with regulatory function which might be involved in the biosynthesis of lipothrenins.

**Table 1:** Genes on BAC I6 involved in lipothrenin biosynthesis (*litA-litS*).

Gene	length (aa)	putative function	coverage/identity	accession
<i>orf-06</i>	285	SDR reductase	99/99	WP_202415345
<i>orf-05</i>	171	Hypothetical protein	99/99	WP_161232279
<i>orf-04</i>	159	Rhodanese-like sulfurtransferase	99/97	WP_161108198
<i>orf-03</i>	164	Transcriptional regulator	98/98	GGT45137
<i>orf-02</i>	997	Nuclease SbcCD subunit C	99/97	GHE70600
<i>orf-01</i>	338	Nuclease SbcCD subunit D	99/99	WP_136237380
<i>litA</i>	1168	Bacterial transcriptional activator	95/99	WP_028959863
<i>litB</i>	269	Beta-ketoacyl ACP synthase III	99/98	WP_210637060
<i>litC</i>	496	MFS transporter permease	99/99	WP_210637059
<i>litD</i>	573	Fatty-acyl AMP ligase	99/89	WP_069171638
<i>litE</i>	326	Ribonucleotide-diphosphate reductase	99/99	WP_137209500
<i>litF</i>	88	Acyl carrier protein	98/91	WP_202278898
<i>litG</i>	228	SGNH/GDSL hydrolase	99/92	WP_202278897
<i>litH</i>	539	GMC family oxidoreductase	99/99	WP_210637056
<i>litI</i>	124	Hypothetical protein	99/82	WP_202278895
<i>litJ</i>	421	Acyl-CoA dehydrogenase	99/96	WP_202278894
<i>litK</i>	85	Hypothetical protein	-	-
<i>LitI</i>	311	Hypothetical protein	99/99	WP_210637055
<i>litM</i>	297	ThiF family adenyllyltransferase	99/99	WP_161108191
<i>litN</i>	311	Diiron oxygenase	99/97	WP_069171647
<i>litO</i>	110	DUF4873 domain protein	99/98	WP_210637054
<i>litP</i>	256	Dienelactone hydrolase	99/98	WP_028959855
<i>litQ</i>	205	Transcriptional regulator	99/94	WP_069171650
<i>litR</i>	526	Transcriptional regulator	98/98	WP_136237385
<i>litS</i>	181	Sigma-70 family	99/98	WP_164555762
<i>orf 01</i>	209	Translation regulator	99/99	WP_136237387
<i>orf 02</i>	180	Histidine phosphatase	99/91	WP_163088151
<i>orf 03</i>	281	Hypothetical protein	90/98	WP_210637257
<i>orf 04</i>	136	CoA-binding protein	99/98	WP_033276933

To investigate this, two deletion mutants were created covering *litQ-orf01* (I6\_ΔRS1) and *orf01-orf04* (I6\_ΔRS) (Figure S1). The production of **1-14** in I6\_ΔRS1 was abolished while production in I6\_ΔRS was decreased. Regulators *litQ-litS* seem to be essential for the expression of the *lit* cluster. However, the exact number of involved regulatory genes at the right flank needs to be determined in further deletion experiments. To determine the left border of the *lit* cluster, four mutants were created containing deletions of *orf-24 - orf-11* (I6\_ΔLS1), *orf-24 - orf-09* (I6\_ΔLS2), *orf-24 - orf-03* (I6\_ΔLS3) and *orf-24 - litA* (I6\_ΔLS4) (Figure S1). Lipothrenin production in the deletion mutants I6\_ΔLS1 and I6\_ΔLS2 was not altered. Similar to mutant I6\_ΔRS, lipothrenin production in mutant I6\_ΔLS3 was decreased. Mutant I6\_ΔLS4 finally led to a complete production abolishment of **1-14** indicating the left cluster border is located between *orf-02* and *litA*. Genes *orf-02* and *orf-01* encode for nucleases and do likely not express functions related to lipothrenin biosynthesis. Gene *litA* is encoding for an OmpR-like winged helix-turn-helix domain containing protein, a family of *Streptomyces* antibiotic regulatory proteins (SARP).<sup>13</sup> Both genes *litA* and *litR* are members of the SARP family and are likely

involved in the initiation of lipothrenin biosynthesis. In summary, the minimal set required for lipothrenin biosynthesis was roughly determined between *litA-litS* (Table 1).

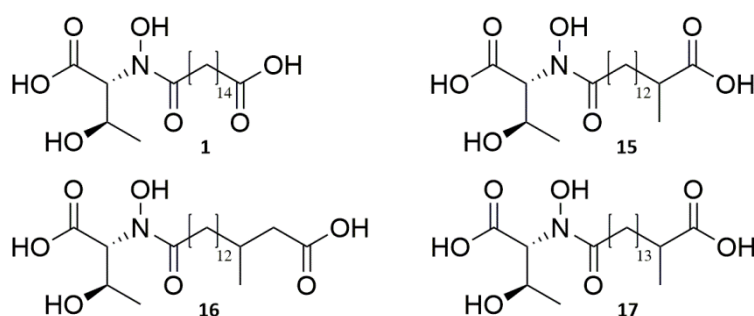
### 3.3.3.2 *LitB is involved in hexadecanedioic acid biosynthesis*

*De novo* FA synthesis in *Streptomyces* is catalyzed by components of type II FA synthase. In most of *Streptomyces* species, the core genes *fabD* (malonyl acyltransferase), *fabH* (keto synthase III), *acpP* (acyl carrier protein) and *fabF* (keto synthase II) are clustered, while enoylreductase, the ketoreductase and dehydratase genes are located elsewhere in the genome.<sup>14-16</sup> In the *lit* cluster, only three FA synthase related genes have been identified: ketosynthase *LitB*, acyl carrier protein *LitF* and dehydrogenase *LitJ*. *LitF* is likely involved in the activation of hexadecanedioic acid (see 1.3.3.4). Deletion of dehydrogenase gene encoding gene *litJ* led to a slight decrease in the production of lipothrenins, but a major impact on FA synthesis can be excluded. The protein size of *LitB* (269 aa) was substantially smaller to previously described *fabH* genes, suggesting an altered function. In *Streptomyces*, *FabH* uses the starter units acetyl-CoA, butyryl-CoA and isobutyryl-CoA to catalyze the biosynthesis of linear and branched chain fatty acids.<sup>17, 18</sup> The exclusive incorporation of the even numbered and unbranched hexadecanedioic acid in lipothrenin **1-4** indicates, that *litB* might be involved in the production of the linear dicarboxylic acid. To explore the influence of *litB* on the FA chain biosynthesis, we created the mutant strain I6\_Δ*litB*. As a result, in addition to the previously described lipothrenins **1-14**, I6\_Δ*litB* showed production of three new compounds with m/z 418.1 Da (**16**, **17**) and m/z 404.1 Da (**15**) (Figure S2). Isolation and structure elucidation by NMR spectroscopy led to the assignment of *iso*-lipothrenin A (**15**), an isomer of lipothrenin A (**1**). Compound **15** shares the same molecular formula (C<sub>20</sub>H<sub>38</sub>NO<sub>7</sub>) and mass (m/z 403.259 Da) compared to **1**. Analysis of the <sup>1</sup>H shifts and coupling constants revealed the same N-hydroxylated D-*allo*-threonine moiety as **1**. An additional methyl group at δ<sub>H</sub> 1.03 and a methine at δ<sub>H</sub> 2.28 indicated differences at the FA chain. HMBC and COSY correlations indicated that the ω-carboxyl group changed its position to C-14 of the FA chain. The altered FA of **15** was assigned as 2-methylpentadecanedioic acid (Table S14). The new compounds **16** and **17** showed calculated formulas of C<sub>21</sub>H<sub>39</sub>NO<sub>7</sub> based on the monoisotopic mass m/z 417.273. Analysis of <sup>1</sup>H NMR and HSQC showed the same N-hydroxylated D-*allo*-threonine moiety as **1**, but changes at the FA chain. Based on the COSY and HMBC data, the two new compounds were identified as 14-methyl lipothrenin A (**17**) and 15-methyl lipothrenin A (**16**) (Table S15, S16).

In summary, the mutant strain I6\_Δ*litB* showed production of the new lipothrenins **15**, **16** and **17** with branched chain fatty acid residues (Figure 5), while the linear lipothrenins **1-14** were significantly reduced. After deletion of *litB*, the production of **15**, **16** and **17** suggests that *LitB* was complemented by other KS genes encoded in the genome of *S. albus* Δ14. The results from the



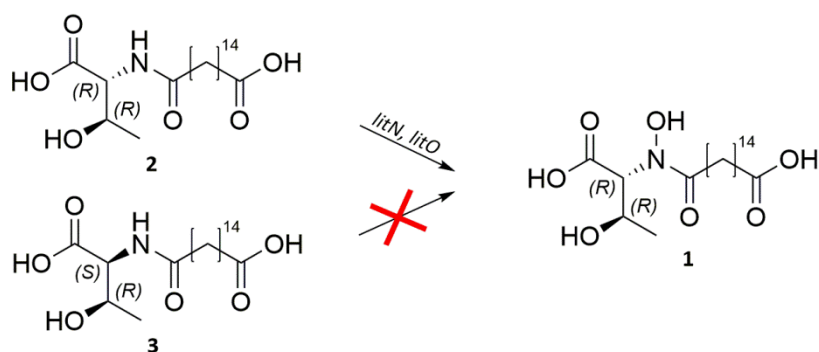
deletion experiment indicate that *LitB* is involved in the biosynthesis of the linear hexadecanedioic and might not be able to accept branched chain starter units.



**Figure 5:** Structures of lipothrenin A (**1**), *iso*-lipothrenin A (**15**), 14-Me lipothrenin A (**16**) and 15-Me lipothrenin A (**17**).

### 3.3.3.3 Stereoselective N-oxygenation

N-oxygenation is carried out by enzymes such as metal-dependent mono- and dioxygenases, flavin-dependent monooxygenases etc. Those enzymes are selectively activating and transferring oxygen to different kinds of nitrogen groups. By using antiSMASH and BLAST for sequence analysis, we identified the diiron monooxygenase encoding gene *litN* and a DUF4873-domain containing gene *litO* which is often associated with flavin-binding monooxygenases. The putative diiron oxygenase encoded by *litN* shows similarities to N-oxygenase AurF. This enzyme belongs to the family of non-heme dinuclear iron monooxygenases which catalyze conversion of aryl-amine substrates to aryl-nitro products.<sup>19, 20</sup> To investigate the involvement of *litN* and *litO* in the N-oxygenation, we created two mutant strains I6\_Δ*litN* and I6\_Δ*litO* containing deletions of *litN* and *litO*, respectively. Both deletion mutants showed high production of the non-hydroxylated isomers **2**, **3** and **7-10**. Simultaneously, the production of the N-hydroxylated compounds **1**, **5** and **6** was reduced in I6\_Δ*litO* and abolished in I6\_Δ*litN* (Figure S4). These results indicate that lipothrenins **2**, **3** and **7-10** might be precursors for compounds **1**, **5** and **6**. The results from the deletion indicate both enzymes *LitN* and *LitO* have an influence on the oxygenation of the amide nitrogen. The hypothetical DUF4873-domain containing protein *LitO* might not be directly involved in the N-oxygenation but likely has an effect on the enzymatic activity of the diiron oxygenase *LitN*, which presumably executes the N-oxygenation reaction. The assignment of the stereochemistry of **1**, **5** and **6** revealed the N-oxygenation occurs selectively at the amide nitrogen of lipothrenins with *D-allo*-Thr. Lipothrenins with N-hydroxylated *L*-Thr or Dbh were not observed. This indicates the N-oxygenation reaction occurs stereospecifically at compound **2** (Figure 6).



**Figure 6:** Diiron monooxygenase *litN* and DUF4873-domain *litO* catalyzes the stereoselective N-oxygenation of compound **2**.

#### 3.3.3.4 Biosynthetic route towards lipothrenins and NAC-lipothrenins

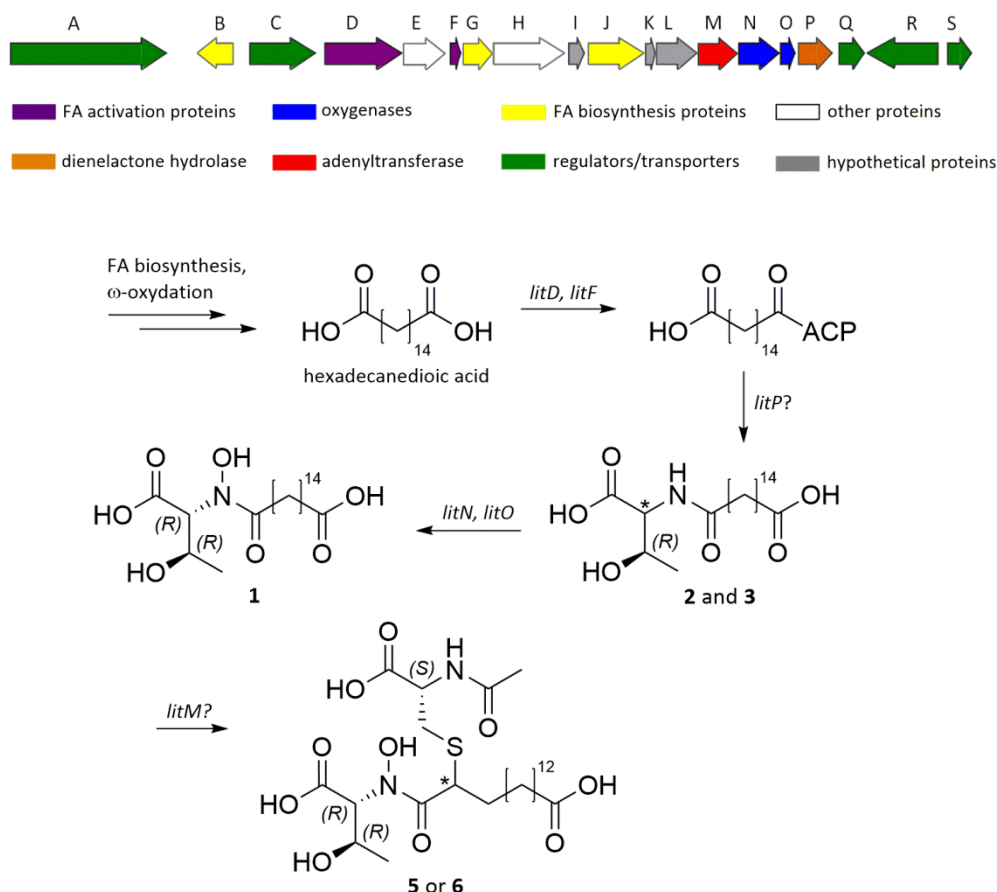
Lipothrenins are yet undescribed members of bacterial lipo-amino acids with an N-dicarboxylic fatty acid moiety. In the literature many amide-linked lipo-amino acids from bacteria have been described and some biosynthetic routes have been published.<sup>21-25</sup> Fatty acid chains are usually activated by fatty acyl-AMP ligase (FAAL) or fatty acyl-CoA ligase (FACL) followed by a transfer to ACP. The subsequent amide bond formation between the activated fatty acyl chain and amino acids is catalyzed by N-acyltransferases or N-acyl amino acid synthases.<sup>23, 26-28</sup> Isomerization of amino acids are catalyzed by racemases or epimerases before or after the amide bond formation.<sup>29-31</sup>

The lipothrenin gene cluster was analyzed to identify genes that match the above mentioned functions. The FAAL encoding gene *litD* and the ACP encoding gene *litF* likely catalyze the activation of the hexadecanedioic acid. A similar mechanism has been described for the lipidation reaction of daptomycin.<sup>32</sup> N-acyltransferase or N-acyl amino acid synthase genes responsible for the subsequent amide bond formation have not been identified within the lipothrenin BGC. Therefore, we performed a series of gene deletions within the *lit* cluster to determine candidates that are involved in the amide bond formation. The following genes were analyzed due to their unknown function in the lipothrenin biosynthesis: *LitE* (ribonucleotide-diphosphate reductase), *LitH* (GMC family oxidoreductase), *LitJ* (Acyl-CoA dehydrogenase), *LitM* (ThiF family adenylyltransferase) and *LitP* (dienelactone hydrolase). Lipothrenin production is expected to be abolished if one of the genes involved in the amide bond formation is deleted.

The constructed deletion mutant I6\_Δ*litE*, I6\_Δ*litH* and I6\_Δ*litJ* showed effects on the production level and metabolite composition of **1-14** (Figure S3). However, these genes are likely not involved in the amid bond formation since lipothrenins were still produced.

Production of **1-14** in the mutant strain I6\_Δ*litP* was completely abolished (Figure S3). BLAST analysis of *LitP* revealed high similarities to the dienelactone hydrolase from *Pseudomonas* sp. B13.<sup>33</sup> *LitP* belongs to the family of  $\alpha/\beta$  hydrolases and might in the amide bond formation. However, deletion

of *litP* could also have led to a polar effect on the adjacent transcriptional regulator *litQ*, thereby shutting off the *lit* cluster transcription. This effect has to be investigated in future experiments by *litP* complementation or removal of the resistance marker in I6\_Δ*litP*. If *litQ* is affected in I6\_Δ*litP*, production of **1-14** will remain absent.



**Figure 7:** Proposed biosynthetic route to lipothrenin A (**1**) and NAC-lipothrenin A/A<sub>1</sub> (**5, 6**).

Deletion mutant I6\_Δ*litM* showed high production of non-oxidized lipothrenins **2** and **3**, while other lipothrenin derivatives were not produced (Figure S3). The N-oxygenase genes *litN* and *litO* are placed within the same operon as *litM*; thus, the absence of the N-hydroxylated compound **1** can be explained by polar effects on *litN* and *litO*. The abolishment of the production of the NAC-modified lipothrenins **5-14** indicated influences of *litM* on N-acetyl-cysteinylolation. The *litM* gene encodes a ThiF family adenylyltransferase which catalyzes the adenylation by ATP of the carboxyl group of the C-terminal glycine of sulfur carrier protein ThiS.<sup>34</sup> However, the function of *LitM* in the NAC-modification is unclear. N-acetyl-cysteinylolation of natural products is a common process in streptomycetes that serves for detoxification and cell stress protection; *LitM* might be involved in this process.<sup>35, 36</sup> A function related to the amide bond formation can be excluded since lipothrenins B (**2, 3**) production was not abolished.

Genes responsible for the isomerization of L-threonine or the dehydration that leads to Z-dehydrobutyrine have not been identified through the performed deletion experiments. Therefore, many of the genes involved in lipothrenin biosynthesis might part of the primary metabolism since no equivalents have been identified in the lipothrenin gene cluster. The biosynthesis is accomplished as previously described by *litN* and *litO* completing the proposed biosynthesis route towards the final product lipothrenin A (**1**) (Figure 7).

### 3.4 Summary

Heterologous expression of BAC I6 from a genomic library of *S. aureus* LU18118 led to the discovery of new N-hydroxylated threonine-hexadecanedioic-acids named lipothrenins. All lipothrenin derivatives exist as an N-acetylcysteine (NAC) modified version which were named NAC-lipothrenins. The unbranched hexadecanedioic acid moiety is the same in all derivatives while the threonine moiety is varying (L-threonine, D-*allo*-threonine or Z/E – dehydrobutyrine). Only lipothrenins with D-*allo*-threonine carry the N-hydroxyl group (lipothrenin A). The lipothrenin gene cluster was analyzed by gene deletion experiments to propose a biosynthetic pathway towards lipothrenin A and NAC-lipothrenin A/A<sub>1</sub>. A minimal set of 19 genes (*litA-litS*) involved in the biosynthesis was roughly determined. The biosynthesis of hexadecanedioic acid is dependent on ketosynthase *LitB* which catalyzes the formation of the linear fatty acid product rather than the branched chain fatty acids which are usually observed in actinomycetes. The linear hexadecanedioic acid is activated by fatty acyl-AMP ligase *LitD* and acyl carrier protein *LitF*. Genes coding for enzymes that catalyzed the amide bond formation and isomerization of threonine were not identified by the performed deletion experiments or by *in silico* analysis. The stereoselective N-oxygenation is carried out by diiron oxygenase *litN* in combination with DUF4873 protein *LitO*. Deletion of *litM* led to the abolishment of the NAC modified lipothrenins. However, the function of *litM* in the N-acetyl cysteinylation of lipothrenins is unclear. Many of the genes involved in lipothrenin biosynthesis might part of the primary metabolism since no equivalents have been identified in the lipothrenin gene cluster by *in silico* analysis or deletion experiments. In addition, deleted genes might be complemented by genes outside of the cluster. Therefore, to draw a full picture of the lipothrenin biosynthesis, further experiments are required.

## References

1. Bentley, S.D.; *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2). *Nature* **2002**, *417* (6885), 141-147, doi.org/10.1038/417141a.
2. Myronovskyi, M.; *et al.* Heterologous production of small molecules in the optimized *Streptomyces* hosts. *Natural product reports* **2019**, *36* (9), 1281-1294, 10.1039/C9NP00023B.
3. Baltz, R.H. Renaissance in antibacterial discovery from actinomycetes. *Curr. Opin. Pharm.* **2008**, *8* (5), 557-563, 10.1016/j.coph.2008.04.008.
4. Green, J.M.R.; *et al.* *Molecular Cloning: A Laboratory Manual*, 4 ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2012.
5. Kieser, T.; *et al.* *Practical Streptomyces Genetics. A Laboratory Manual*; John Innes Foundation: Norwich, UK, 2000.
6. Rebets, Y.; *et al.* in *Metagenomics: Methods and Protocols*. (eds. W.R. Streit, R. Daniel) 99-144 Springer New York, New York, NY; 2017.
7. Medema, M.H.; *et al.* antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **2011**, *39* (suppl\_2), W339-W346, 10.1093/nar/gkr466.
8. Altschul, S.F.; *et al.* Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215* (3), 403-410, 10.1016/S0022-2836(05)80360-2.
9. Kears, M.; *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28* (12), 1647-1649, 10.1093/bioinformatics/bts199.
10. Whittle, M.; *et al.* Evaluation of Similarity Measures for Searching the Dictionary of Natural Products Database. *Journal of Chemical Information and Computer Sciences* **2003**, *43* (2), 449-457, 10.1021/ci025591m.
11. Fujii, K.; *et al.* Further application of advanced Marfey's method for determination of absolute configuration of primary amino compound. *Tetrahedron Lett.* **1998**, *39* (17), 2579-2582, 10.1016/S0040-4039(98)00273-1.
12. Mazingo, R.; *et al.* Hydrogenolysis of sulfur compounds by Raney nickel catalyst. *J. Am. Chem. Soc.* **1943**, *65* (6), 1013-1016, doi.org/10.1021/ja01246a005.
13. Wietzorrek, A.; *et al.* A novel family of proteins that regulates antibiotic production in streptomycetes appears to contain an OmpR-like DNA-binding fold. *Mol. Microbiol.* **1997**, *25* (6), 1181-1184, 10.1046/j.1365-2958.1997.5421903.x.
14. Revill, W.P.; *et al.* Purification of a malonyltransferase from *Streptomyces coelicolor* A3 (2) and analysis of its genetic determinant. *J. Bacteriol.* **1995**, *177* (14), 3946-3952, 10.1128/jb.177.14.3946-3952.1995.
15. Revill, W.P.; *et al.* Relationships between fatty acid and polyketide synthases from *Streptomyces coelicolor* A3 (2): characterization of the fatty acid synthase acyl carrier protein. *J. Bacteriol.* **1996**, *178* (19), 5660-5667, 10.1128/jb.178.19.5660-5667.1996.
16. Revill, W.P.; *et al.*  $\beta$ -Ketoacyl acyl carrier protein synthase III (FabH) is essential for fatty acid biosynthesis in *Streptomyces coelicolor* A3 (2). *J. Bacteriol.* **2001**, *183* (11), 3526-3530, 10.1128/JB.183.11.3526-3530.2001.

17. Kaneda, T. Iso-and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. *Microbiol. Mol. Biol. Rev.* **1991**, *55* (2), 288-302.
18. Han, L.; *et al.* Characterization of  $\beta$ -ketoacyl-acyl carrier protein synthase III from *Streptomyces glaucescens* and its role in initiation of fatty acid biosynthesis. *J. Bacteriol.* **1998**, *180* (17), 4481-4486, 10.1128/JB.180.17.4481-4486.1998.
19. Choi, Y.S.; *et al.* *In vitro* reconstitution and crystal structure of *p*-aminobenzoate *N*-oxygenase (AurF) involved in aureothin biosynthesis. *Proceedings of the National Academy of Sciences* **2008**, *105* (19), 6858-6863, 10.1073/pnas.0712073105.
20. Simurdiak, M.; *et al.* A New Class of Arylamine Oxygenases: Evidence that *p*-Aminobenzoate *N*-Oxygenase (AurF) is a Di-iron Enzyme and Further Mechanistic Studies. *ChemBioChem* **2006**, *7* (8), 1169-1172, 10.1002/cbic.200600136.
21. Asselineau, J. Bacterial lipids containing amino acids or peptides linked by amide bonds. *Fortschritte der Chemie organischer Naturstoffe/Progress in the Chemistry of Organic Natural Products* **1991**, 1-85.
22. Geiger, O.; *et al.* Amino acid-containing membrane lipids in bacteria. *Progress in lipid research* **2010**, *49* (1), 46-60, doi.org/10.1016/j.plipres.2009.08.002.
23. Brady, S.F.; *et al.* New natural product families from an environmental DNA (eDNA) gene cluster. *J. Am. Chem. Soc.* **2002**, *124* (34), 9968-9969, 10.1021/ja0268985.
24. Kudo, Y.; *et al.* Expansion of gamma-butyrolactone signaling molecule biosynthesis to phosphotriester natural products. *ACS Chemical Biology* **2020**, *15* (12), 3253-3261, 10.1021/acscchembio.0c00824.
25. Wang, M.; *et al.* Using the pimeloyl-CoA synthetase adenylation fold to synthesize fatty acid thioesters. *Nat. Chem. Biol.* **2017**, *13* (6), 660, 10.1038/nchembio.2361.
26. Macintyre, L.W.; *et al.* An *Ichip*-domesticated sponge bacterium produces an *N*-acyltyrosine bearing an  $\alpha$ -methyl substituent. *Organic letters* **2019**, *21* (19), 7768-7771, 10.1021/acs.orglett.9b02710.
27. Gould, T.A.; *et al.* Structure of the *Pseudomonas aeruginosa* acyl-homoserinelactone synthase LasI. *Mol. Microbiol.* **2004**, *53* (4), 1135-1146, 10.1111/j.1365-2958.2004.04211.x.
28. Weissenmayer, B.; *et al.* Identification of a gene required for the biosynthesis of ornithine-derived lipids. *Mol. Microbiol.* **2002**, *45* (3), 721-733, 10.1046/j.1365-2958.2002.03043.x.
29. Ogasawara, Y.; *et al.* Peptide epimerization machineries found in microorganisms. *Frontiers in microbiology* **2018**, *9*, 156, 10.3389/fmicb.2018.00156.
30. Lim, Y.-H.; *et al.* A new amino acid racemase with threonine  $\alpha$ -epimerase activity from *Pseudomonas putida*: purification and characterization. *J. Bacteriol.* **1993**, *175* (13), 4213-4217, 10.1128/jb.175.13.4213-4217.1993.
31. Walsh, C.T.; *et al.* Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angewandte Chemie International Edition* **2013**, *52* (28), 7098-7124.
32. Wittmann, M.; *et al.* Role of DptE and DptF in the lipidation reaction of daptomycin. *The FEBS journal* **2008**, *275* (21), 5343-5354, 10.1111/j.1742-4658.2008.06664.x.

33. Pathak, D.; *et al.* Refined structure of diene lactone hydrolase at 1.8 Å. *J. Mol. Biol.* **1990**, *214* (2), 497-525, 10.1016/0022-2836(90)90196-S.
34. Taylor, S.V.; *et al.* Thiamin biosynthesis in *Escherichia coli*: identification of ThiS thiocarboxylate as the immediate sulfur donor in the thiazole formation. *J. Biol. Chem.* **1998**, *273* (26), 16555-16560, 10.1074/jbc.273.26.16555.
35. Jothivasan, V.K.; *et al.* Mycothiol: synthesis, biosynthesis and biological functions of the major low molecular weight thiol in actinomycetes. *Natural product reports* **2008**, *25* (6), 1091-1117, 10.1039/B616489G.
36. Newton, G.L.; *et al.* Biosynthesis and functions of mycothiol, the unique protective thiol of Actinobacteria. *Microbiology and molecular biology reviews: MMBR* **2008**, *72* (3), 471, 10.1128/MMBR.00008-08.

## Supplementary

### Table of Contents

Lipothrenin - Supplementary .....	187
1 Heterologous expression and genetic manipulation .....	190
2 Physical information of isolated lipothrenins .....	195
3 NMR Spectroscopy .....	196
3.1 Lipothrenins isolated from <i>Streptomyces aureus</i> LU18118 .....	196
3.2 Lipothrenin derivatives produced by <i>Streptomyces aureus</i> I6_ΔKS .....	201
4 Stereochemistry Assignment by Marfey's Method .....	202
5 NMR Spectra .....	204
6 References.....	235

### Table of Tables

Table S1: Strains, Plasmids and BACS used in this study.....	190
Table S2: Used primers for RedET.....	191
Table S3: Used primers for screening correct mutants.....	192
Table S4: NMR data (500 MHz, DMSO- <i>d</i> 6) for lipothrenin A (1).....	196
Table S5: NMR data (500 MHz, DMSO- <i>d</i> 6) for D- and L-lipothrenin B mixture (2, 3).....	196
Table S6: NMR data (500 MHz, DMSO- <i>d</i> 6) for Z-lipothrenin C (4).....	197
Table S7: NMR data (500 MHz, DMSO- <i>d</i> 6) for 2-NAC-D-lipothrenin A (5).....	197
Table S8: NMR data (500 MHz, DMSO- <i>d</i> 6) for 2-NAC-D-lipothrenin A <sub>1</sub> (6) .....	198
Table S9: NMR data (500 MHz, DMSO- <i>d</i> 6) for 2-NAC-D-lipothrenin B (7).....	198
Table S10: NMR data (500 MHz, DMSO- <i>d</i> 6) for 2-NAC-L-lipothrenin B (8) .....	199
Table S11: NMR data (500 MHz, DMSO- <i>d</i> 6) for 2-NAC-D-/L-lipothrenin B <sub>1</sub> mixture (9, 10).....	199
Table S12: NMR data (500 MHz, DMSO- <i>d</i> 6) for 2-NAC-Z-lipothrenin C and C <sub>1</sub> mix (11, 12).....	200
Table S13: NMR data (500 MHz, DMSO- <i>d</i> 6) for 2-NAC-E-lipothrenin C and C <sub>1</sub> (13, 14) .....	200
Table S14: NMR data (500 MHz, DMSO- <i>d</i> 6) for <i>Iso</i> -lipothrenin A (15).....	201
Table S15: NMR data (500 MHz, DMSO- <i>d</i> 6) for 14-methyl-lipothrenin A (16).....	201
Table S16: NMR data (500 MHz, DMSO- <i>d</i> 6) for 15-methyl-lipothrenin A (17).....	201

### Table of Figures

Figure S1: LC-MS chromatograms of deletion mutants I6_ΔRS, I6_ΔRS1, I6_ΔLS1, I6_ΔLS2, I6_ΔLS3 and I6_ΔLS4 showing the extracted masses of the derivatives of lipothrenin A (1, 5, 6, red), lipothrenin B (2, 3, 7-10, green) and lipothrenin C (4, 11-14, blue).....	193
--	-----



Figure S2: LC-MS chromatograms of deletion mutants I6_ΔlitB showing the extracted masses of the derivatives of lipothrenin A (1, 5, 6, red), lipothrenin B (2, 3, 7-10, green), lipothrenin C (4, 11-14, blue) and 14/15-methyl-lipothrenin A (16, 17, black). .....	193
Figure S3: LC-MS chromatograms of deletion mutants I6_ΔlitJ, I6_ΔlitH, I6_ΔlitE, I6_ΔlitP and I6_ΔlitM showing the extracted masses of the derivatives of lipothrenin A (1, 5, 6, red), lipothrenin B (2, 3, 7-10, green), lipothrenin C (4, 11-14, blue) and new putative lipothrenin derivatives (orange). .....	194
Figure S4: LC-MS chromatograms of deletion mutants I6_ΔlitN and I6_ΔlitO showing the extracted masses of the derivatives of lipothrenin A (1, 5, 6, red), lipothrenin B (2, 3, 7-10, green) and lipothrenin C (4, 11-14, blue). .....	194
Figure S5: LC-MS chromatograms of hydrolysed lipothrenin A (a), B (b) 2-NAC-DL-lipothrenin B/B <sub>1</sub> (c-e) and the references L-, D-, L- <i>allo</i> - and D- <i>allo</i> -threonine (f) derivatized with L-FDLA.....	202
Figure S6: LC-MS chromatograms of the extracted masse of the DLA-cys-FA conjugate derived from hydrolysis of 2-NAC-DL-lipothrenin B/B <sub>1</sub> and derivatisation with L- and D-FDLA.....	203
Figure S7: <sup>1</sup> H NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of lipothrenin A (1). .....	204
Figure S8: <sup>13</sup> C NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of lipothrenin A (1). .....	204
Figure S9: <sup>1</sup> H- <sup>1</sup> H-COSY spectrum (500 MHz, DMSO-d <sub>6</sub> ) of lipothrenin A (1). .....	205
Figure S10: HSQC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of lipothrenin A (1). .....	205
Figure S11: HMBC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of lipothrenin A (1). .....	206
Figure S12: <sup>1</sup> H NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of D- and L-lipothrenin B (2, 3). .....	206
Figure S13: <sup>13</sup> C NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of D- and L-lipothrenin B (2, 3). .....	207
Figure S14: <sup>1</sup> H- <sup>1</sup> H-COSY spectrum (500 MHz, DMSO-d <sub>6</sub> ) of D- and L-lipothrenin B (2, 3). .....	207
Figure S15: HSQC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of D- and L-lipothrenin B (2, 3). .....	208
Figure S16: HMBC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of D- and L-lipothrenin B (2, 3). .....	208
Figure S17: <sup>1</sup> H NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of lipothrenin C (4). .....	209
Figure S18: <sup>13</sup> C NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of lipothrenin C (4). .....	209
Figure S19: <sup>1</sup> H- <sup>1</sup> H-COSY spectrum (500 MHz, DMSO-d <sub>6</sub> ) of lipothrenin C (4). .....	210
Figure S20: HSQC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of lipothrenin C (4). .....	210
Figure S21: HMBC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of lipothrenin C (4). .....	211
Figure S22: ROESY spectrum (500 MHz, DMSO-d <sub>6</sub> ) of lipothrenin C (4). .....	211
Figure S23: <sup>1</sup> H NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-lipothrenin A (5). .....	212
Figure S24: <sup>13</sup> C NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-lipothrenin A (5). .....	212
Figure S25: <sup>1</sup> H- <sup>1</sup> H-COSY spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-lipothrenin A (5). .....	213
Figure S26: HMBC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-lipothrenin A (5). .....	213
Figure S27: <sup>1</sup> H NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-lipothrenin A <sub>1</sub> (6). .....	214
Figure S28: <sup>13</sup> C NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-lipothrenin A <sub>1</sub> (6). .....	214
Figure S29: <sup>1</sup> H- <sup>1</sup> H-COSY spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-lipothrenin A <sub>1</sub> (6). .....	215
Figure S30: HSQC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-lipothrenin A <sub>1</sub> (6). .....	215
Figure S31: HMBC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-lipothrenin A <sub>1</sub> (6). .....	216
Figure S32: <sup>1</sup> H NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-D-lipothrenin B (7). .....	216
Figure S33: <sup>1</sup> H- <sup>1</sup> H-COSY spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-D-lipothrenin B (7). .....	217
Figure S34: HSQC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-D-lipothrenin B (7). .....	217
Figure S35: HMBC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-D-lipothrenin B (7). .....	218
Figure S36: ROESY spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-D-lipothrenin B (7). .....	218
Figure S37: <sup>1</sup> H NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-L-lipothrenin B (8). .....	219
Figure S38: <sup>1</sup> H- <sup>1</sup> H-COSY spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-L-lipothrenin B (8). .....	219

Figure S39: HSQC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-L-lipothrenin B (8). .....	220
Figure S40: HMBC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-L-lipothrenin B (8). .....	220
Figure S41: ROESY spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-L-lipothrenin B (8). .....	221
Figure S42: <sup>1</sup> H NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-D-/L-lipothrenin B <sub>1</sub> mixture (9, 10)... ..	221
Figure S43: <sup>13</sup> C NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-D-/L-lipothrenin B <sub>1</sub> mixture (9, 10).. ..	222
Figure S44: <sup>1</sup> H- <sup>1</sup> H-COSY spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-D-/L-lipothrenin B <sub>1</sub> mixture (9, 10). .....	222
Figure S45: HSQC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-D-/L-lipothrenin B <sub>1</sub> mixture (9, 10). ....	223
Figure S46: HMBC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-D-/L-lipothrenin B <sub>1</sub> mixture (9, 10). ....	223
Figure S47: <sup>1</sup> H NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-Z-lipothrenin C and C <sub>1</sub> mixture (11, 12). .....	224
Figure S48: <sup>13</sup> C NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-Z-lipothrenin C and C <sub>1</sub> mixture (11, 12). .....	224
Figure S49: <sup>1</sup> H- <sup>1</sup> H-COSY spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-Z-lipothrenin C and C <sub>1</sub> mixture (11, 12). .....	225
Figure S50: HSQC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-Z-lipothrenin C and C <sub>1</sub> mixture (11, 12).225	
Figure S51: HMBC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-Z-lipothrenin C and C <sub>1</sub> mixture (11, 12). .....	226
Figure S52: <sup>1</sup> H NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-E-lipothrenin C and C <sub>1</sub> mixture (13, 14). .....	226
Figure S53: <sup>1</sup> H- <sup>1</sup> H-COSY spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-E-lipothrenin C and C <sub>1</sub> mixture (13, 14). .....	227
Figure S54: HSQC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-E-lipothrenin C and C <sub>1</sub> mixture (13, 14).227	
Figure S55: HMBC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 2-NAC-E-lipothrenin C and C <sub>1</sub> mixture (13, 14). .....	228
Figure S56: <sup>1</sup> H NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of <i>iso</i> -lipothrenin A (15). .....	228
Figure S57: <sup>1</sup> H- <sup>1</sup> H-COSY spectrum (500 MHz, DMSO-d <sub>6</sub> ) of <i>iso</i> -lipothrenin A (15). .....	229
Figure S58: HSQC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of <i>iso</i> -lipothrenin A (15). .....	229
Figure S59: HMBC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of <i>iso</i> -lipothrenin A (15). .....	230
Figure S60: <sup>1</sup> H NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 14-methyl-lipothrenin A (16). .....	230
Figure S61: <sup>13</sup> C NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 14-methyl-lipothrenin A (16). .....	231
Figure S62: <sup>1</sup> H- <sup>1</sup> H-COSY spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 14-methyl-lipothrenin A (16). .....	231
Figure S63: HSQC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 14-methyl-lipothrenin A (16). .....	232
Figure S64: HMBC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 14-methyl-lipothrenin A (16). .....	232
Figure S65: <sup>1</sup> H NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 15-methyl-lipothrenin A (17). .....	233
Figure S66: <sup>13</sup> C NMR spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 15-methyl-lipothrenin A (17). .....	233
Figure S67: <sup>1</sup> H- <sup>1</sup> H-COSY spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 15-methyl-lipothrenin A (17). .....	234
Figure S68: HSQC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 15-methyl-lipothrenin A (17). .....	234
Figure S69: HMBC spectrum (500 MHz, DMSO-d <sub>6</sub> ) of 15-methyl-lipothrenin A (17). .....	235

# 1 Heterologous expression and genetic manipulation

**Table S1:** Strains, Plasmids and BACS used in this study

Strain/Plasmids/BACS	Description	Source
<b>Strains</b>		
Streptomyces aureus LU18118	wild type strain with lipothrenin cluster	BASF
Streptomyces albus $\Delta$ 14	Optimized heterologous host strain <sup>1</sup>	
Streptomyces lividans $\Delta$ 8	Optimized heterologous host strain <sup>2</sup>	
Escherichia coli GB05 RedCC	General cloning host	cloning host [HIPS*]
Escherichia coli ET12567 pUB307	Donor strain for intergeneric conjugation <sup>3</sup>	
<b>Plasmids and BACS</b>		
pUC19	Plasmid; General cloning vector; Used as a source of ampicillin-resistance marker for RedET experiments	Thermo Scientific
I6	BAC containing lipothrenin cluster	This study
I6_ $\Delta$ LS1	Derivative of I6; contains substitution of the genes orf(-24) – orf(-12) with ampicillin resistance gene	This study
i6_ $\Delta$ LS2	Derivative of I6; contains substitution of the genes orf(-24) – orf(-08) with ampicillin resistance gene	This study
I6_ $\Delta$ LS3	Derivative of I6; contains substitution of the genes orf(-24) – orf(-04) with ampicillin resistance gene	This study
I6_ $\Delta$ LS4	Derivative of I6; contains substitution of the genes orf(-24) – <i>litA</i> with ampicillin resistance gene	This study
I6_ $\Delta$ litB	Derivative of I6; contains substitution of the gene <i>litB</i> with ampicillin resistance gene	This study
I6_ $\Delta$ litE	Derivative of I6; contains substitution of the gene <i>litE</i> with ampicillin resistance gene	This study
I6_ $\Delta$ litH	Derivative of I6; contains substitution of the gene <i>litH</i> with ampicillin resistance gene	This study
I6_ $\Delta$ litJ	Derivative of I6; contains substitution of the gene <i>litJ</i> with ampicillin resistance gene	This study
I6_ $\Delta$ litM	Derivative of I6; contains substitution of the gene <i>litM</i> with ampicillin resistance gene	This study
I6_ $\Delta$ litN	Derivative of I6; contains substitution of the gene <i>litN</i> with ampicillin resistance gene	This study
I6_ $\Delta$ litO	Derivative of I6; contains substitution of the gene <i>litO</i> with ampicillin resistance gene	This study
I6_ $\Delta$ litP	Derivative of I6; contains substitution of the gene <i>litP</i> with ampicillin resistance gene	This study
I6_ $\Delta$ RS1	Derivative of I6; contains substitution of the genes <i>litQ</i> orf(01) with ampicillin resistance gene	This study
I6_ $\Delta$ RS	Derivative of I6; contains substitution of the genes orf(01) – orf(04) with ampicillin resistance gene	This study

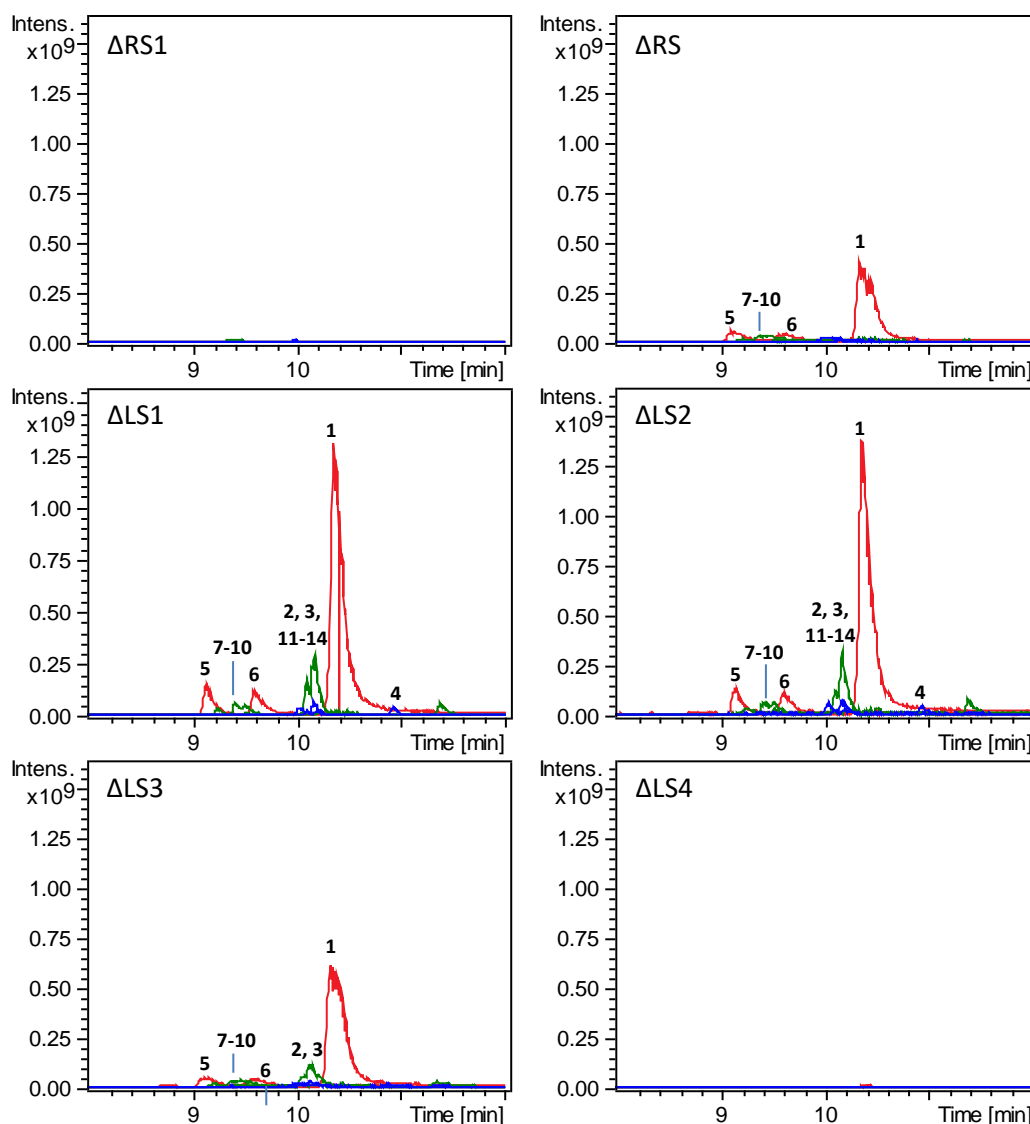
\*Helmholtz-Institut für Pharmazeutische Forschung Saarland

**Table S2:** Used primers for RedET.

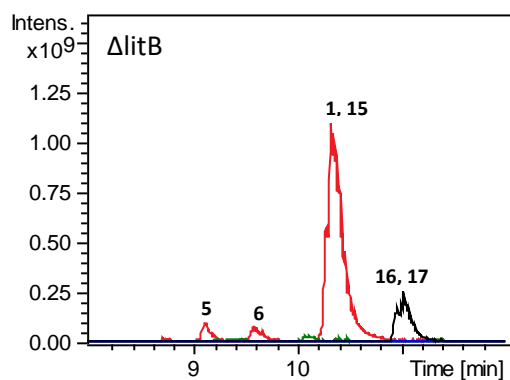
Gene #	Oligo Name	Sequence
<i>orf24-orf11</i>	$\Delta$ LS1_f	ATGAGTGAACCTTGCAGAGCGCAGTTCGGATCCGGGCGCCGGGCCCCG ACGGTTTAAACCGTCAGGTGGCACTTTTCG
<i>orf24-orf11</i>	$\Delta$ LS1_r	TTACTGACCGACGAGGCCCCGGCTGCAGCGTCTTCTGAACAGCACC GTGTTTAAACCTTACCAATGCTTAATCAGTG
<i>orf24-orf09</i>	$\Delta$ LS2_f	ATGAGTGAACCTTGCAGAGCGCAGTTCGGATCCGGGCGCCGGGCCCCG ACGGTTTAAACCGTCAGGTGGCACTTTTCG
<i>orf24-orf09</i>	$\Delta$ LS2_r	TCAGAGAGCCGCGAGTTCTGCCACAGGTCGTCCAGACCCAGCGAC CCGTTTAAACCTTACCAATGCTTAATCAGTG
<i>orf24-orf03</i>	$\Delta$ LS3_f	ATGAGTGAACCTTGCAGAGCGCAGTTCGGATCCGGGCGCCGGGCCCCG ACGGTTTAAACCGTCAGGTGGCACTTTTCG
<i>orf24-orf03</i>	$\Delta$ LS3_r	TCACCGCCCCACGGGACGGCGCGGCAGCGGAGGAGTACACCAC GCCGTTTAAACCTTACCAATGCTTAATCAGTG
<i>orf24-litA</i>	$\Delta$ LS4_f	ATGAGTGAACCTTGCAGAGCGCAGTTCGGATCCGGGCGCCGGGCCCCG ACGGTTTAAACCGTCAGGTGGCACTTTTCG
<i>orf24-litA</i>	$\Delta$ LS4_r	TTACGTGGTGAGGAGGGTTCGCAGGGTGGCTGCCTCCGGTACGCCG AGGTTTAAACCTTACCAATGCTTAATCAGTG
<i>litB</i>	$\Delta$ litB_f	TCAGGGGATCGGACCGTACCGTACGTCAGGCCCGCGCCG AAGGTTTAAACCGTCAGGTGGCACTTTTCG
<i>litB</i>	$\Delta$ litB_r	ATGGCCGCCACGGCCGCGGGGAAGGCGCTCGCCACGCGGGCGCG GCGGTTTAAACCTTACCAATGCTTAATCAGTG
<i>litE</i>	$\Delta$ litE_f	ATGAGCGTCACGACCCCCGGGCTCGTACTCGACGTACCGAGACGG CCCGTTTAAACCGTCAGGTGGCACTTTTCG
<i>litE</i>	$\Delta$ litE_r	TCAGGCCGCGCGGCCCGGATCGGGTGCGGCGCGTTCAGCGCTCG GCGTTTAAACCTTACCAATGCTTAATCAGTG
<i>litH</i>	$\Delta$ litH_f	ATGGGCGCCGGCTCACCAGCGCGGTACGTCTGGTCGGCGCGGGCT CGGGTTTAAACCGTCAGGTGGCACTTTTCG
<i>litH</i>	$\Delta$ litH_r	TCAGCGGCCGCGAGGGCCCCGCGCGGTGCTGCGCCGCTCGGC GAGGTTTAAACCTTACCAATGCTTAATCAGTG
<i>litJ</i>	$\Delta$ litJ_f	GTGATCGCGACCAACCACCGCCGTACCGAACGCCCGACCCCTCTC CGGTTTAAACCGTCAGGTGGCACTTTTCG
<i>litJ</i>	$\Delta$ litJ_r	TCACAGCCGCGCCCGTACAGCGCACCGGTAATGGCCAGCGCAAGC ACGTTTAAACCTTACCAATGCTTAATCAGTG
<i>litM</i>	$\Delta$ litM_f	ATGAGCACCAACCACCGCGCCAGGACGGTTCCGGCCGGTGTCCCCG AGGGTTTAAACCGTCAGGTGGCACTTTTCG
<i>litM</i>	$\Delta$ litM_r	TCACGACTCCTTGCTGTGTGCGCTGAGTGCTGTCCGAAAAGGCAT CGTTTAAACCTTACCAATGCTTAATCAGTG
<i>litN</i>	$\Delta$ litN_f	ATGACCACCGCCACCGACCGCCGGGCGCTCGCTCGGAGGACCCGC TCGGTTTAAACCGTCAGGTGGCACTTTTCG
<i>litN</i>	$\Delta$ litN_r	CTAGGCATCCTTGAGCCAGCCGCGGAGCGCCACACCTTCTCCGAGC GGTTTAAACCTTACCAATGCTTAATCAGTG
<i>litO</i>	$\Delta$ litO_f	GTGGACGATGACGGGTACGAAGGTCCGGCGGTGCTGGCGGTGGAA GGCGTTTAAACCGTCAGGTGGCACTTTTCG
<i>litO</i>	$\Delta$ litO_r	TCAGTTGCCGAGGTACCGAATTCGAACTTCTTCGATGGTGAACG GGTTTAAACCTTACCAATGCTTAATCAGTG
<i>litP</i>	$\Delta$ litP_f	ATGTGCTTCGACGAGGACGCCATTCCACCGGTACCGCAGGTTCCGG ATGGTTTAAACCGTCAGGTGGCACTTTTCG
<i>litP</i>	$\Delta$ litP_r	TCAGCCCGCCCCGACCCGGTCCACGAACCGCAGCACCCGGTCCCAG GCGTTTAAACCTTACCAATGCTTAATCAGTG
<i>orf01-orf04</i>	$\Delta$ RS1_f	TCAGGCGTCCCCGTACGCCTCCCCGCCAGCTCGAATCCGGCCGTCC CCGTTTAAACCGTCAGGTGGCACTTTTCG
<i>orf01-orf04</i>	$\Delta$ RS1_r	TCAGGCGAGGCGCGGGATCTCGATCGCGGGGACGCGGTCCATCACC ATGTTTAAACCTTACCAATGCTTAATCAGTG
<i>litP-orf01</i>	$\Delta$ RS_f	ATGCCACGACTCGTACGAGAACGCGTCCGACGACGCTGAGCGGT CCGGTTTAAACCGTCAGGTGGCACTTTTCG
<i>litP-orf01</i>	$\Delta$ RS_r	ATGCAGGACGAGTACCGCACCGTGGCCCGCGGGGTGTGCACGAG ACCGTTTAAACCTTACCAATGCTTAATCAGTG

**Table S3:** Used primers for screening correct mutants.

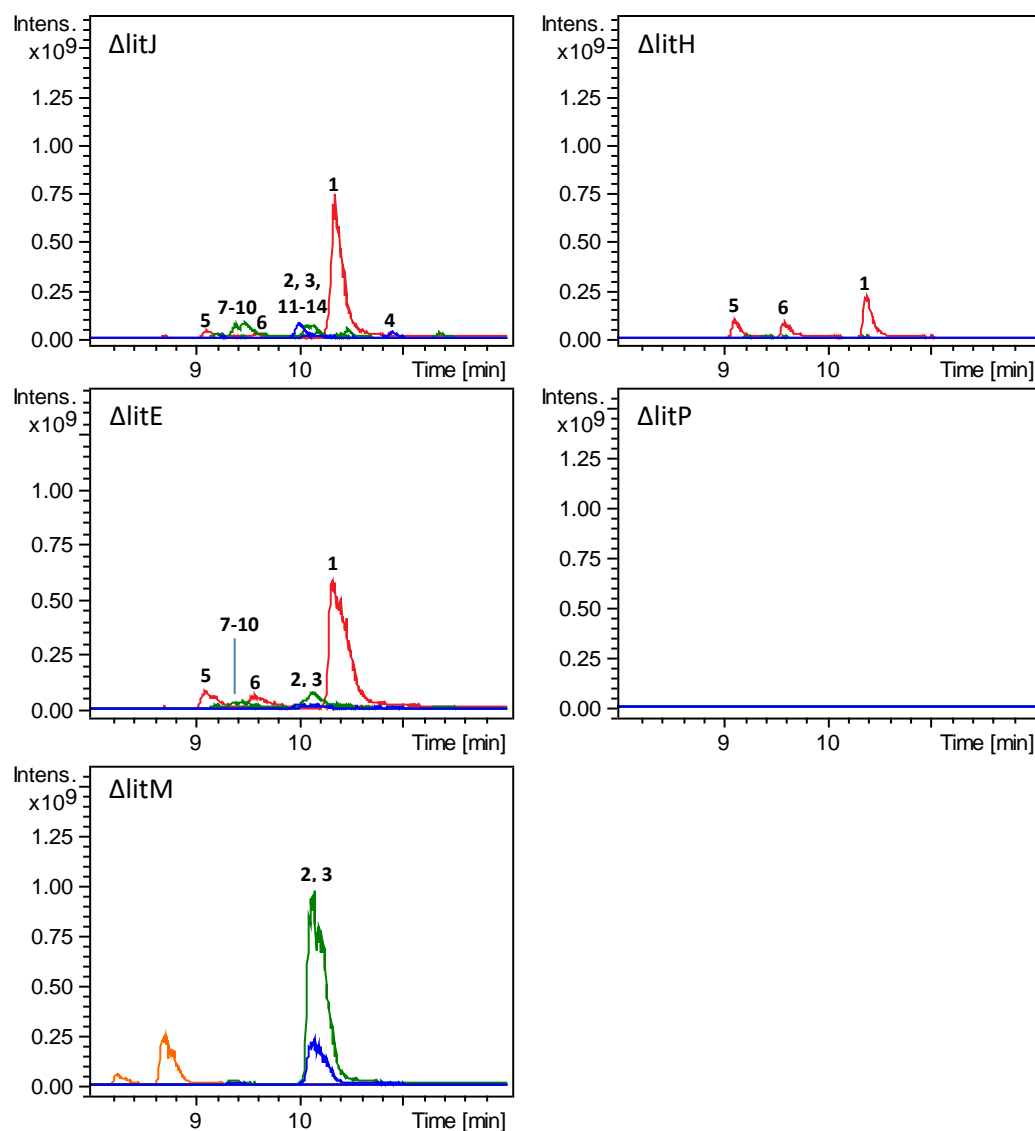
Gene #	Oligo Name	Sequence
<i>orf24-orf11</i>	$\Delta$ LS1_seqf	TTGCACAACATGGGGGATCA
<i>orf24-orf11</i>	$\Delta$ LS1_seqr	GAGCAGGTGGTTCAGGTAGC
<i>orf24-orf09</i>	$\Delta$ LS2_seqf	TCCTTGAGAGTTTTCGCCCC
<i>orf24-orf09</i>	$\Delta$ LS2_seqr	GTCTGTCTGCACCACCATGT
<i>orf24-orf03</i>	$\Delta$ LS3_seqf	GTGACACCACGATGCCTGTA
<i>orf24-orf03</i>	$\Delta$ LS3_seqr	CGAACGCGAACACGAACTG
<i>orf24-litA</i>	$\Delta$ LS4_seqf	GTGACACCACGATGCCTGTA
<i>orf24-litA</i>	$\Delta$ LS4_seqr	CATGTCCGACATCCTCGACC
<i>litB</i>	$\Delta$ litB_seqf	CGTCTGTACGTGCGTCAGG
<i>litB</i>	$\Delta$ litB_seqr	ACAGGAGTCCGTGTGAACAC
<i>litE</i>	$\Delta$ litE_seqf	GAACGATTCCGCGGACCTC
<i>litE</i>	$\Delta$ litE_seqr	GTCTCTGATGTCCTTGACCG
<i>litH</i>	$\Delta$ litH_seqf	TGACCAACGTCTTCGAGTCG
<i>litH</i>	$\Delta$ litH_seqr	GAACTGGAGGGAGAAGTCGC
<i>litJ</i>	$\Delta$ litJ_seqf	CCACATGACACGAAAGGCAG
<i>litJ</i>	$\Delta$ litJ_seqr	GTACGACACCGCGGAGAAG
<i>litM</i>	$\Delta$ litM_seqf	CGTCTGTCGCTGGAAGTAT
<i>litM</i>	$\Delta$ litM_seqr	CGTACAGGGAGATCCGCTTC
<i>litN</i>	$\Delta$ litN_seqf	GTACGTGGTCGAGATCCTGG
<i>litN</i>	$\Delta$ litN_seqr	CCATGCCTTCGATCCGGTAG
<i>litO</i>	$\Delta$ litO_seqf	ACAACCCGCTGATGTACGAG
<i>litO</i>	$\Delta$ litO_seqr	GTAGAACCCGTACAGGCCG
<i>litP</i>	$\Delta$ litP_seqf	CTACCGGATCGAAGGCATGG
<i>litP</i>	$\Delta$ litP_seqr	GCTCTGAAGAGGTGCTGCTT
<i>orf01-orf04</i>	$\Delta$ RS1_seqf	GTGAACACTGCGACAACCAC
<i>orf01-orf04</i>	$\Delta$ RS1_seqr	CAGTGCTGCAATGATACCGC
<i>litP-orf01</i>	$\Delta$ RS_seqf	CACAGCTTCTTCGACGCCAA
<i>litP-orf01</i>	$\Delta$ RS_seqr	CTGAAGGTGATCCGCGAGAC



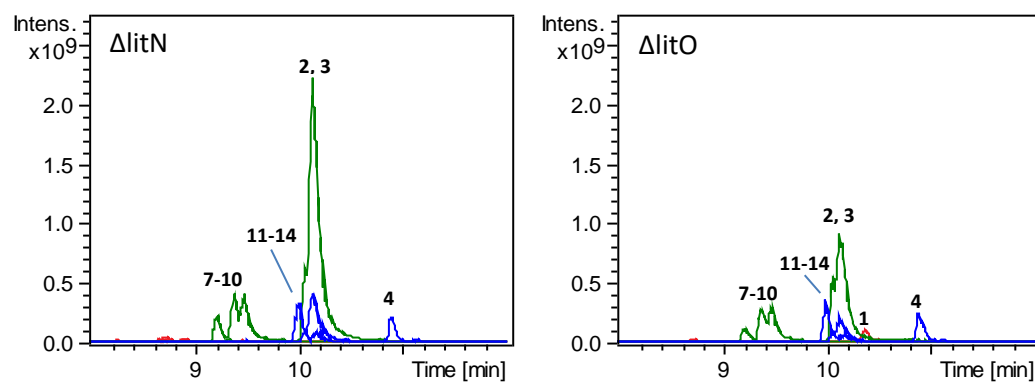
**Figure S1:** LC-MS chromatograms of deletion mutants I6\_ΔRS, I6\_ΔRS1, I6\_ΔLS1, I6\_ΔLS2, I6\_ΔLS3 and I6\_ΔLS4 showing the extracted masses of the derivatives of lipothrenin A (1, 5, 6, red), lipothrenin B (2, 3, 7-10, green) and lipothrenin C (4, 11-14, blue).



**Figure S2:** LC-MS chromatograms of deletion mutants I6\_ΔlitB showing the extracted masses of the derivatives of lipothrenin A (1, 5, 6, red), lipothrenin B (2, 3, 7-10, green), lipothrenin C (4, 11-14, blue) and 14/15-methyl-lipothrenin A (16, 17, black).



**Figure S3:** LC-MS chromatograms of deletion mutants I6\_ΔlitJ, I6\_ΔlitH, I6\_ΔlitE, I6\_ΔlitP and I6\_ΔlitM showing the extracted masses of the derivatives of lipothrenin A (1, 5, 6, red), lipothrenin B (2, 3, 7-10, green), lipothrenin C (4, 11-14, blue) and shunt products (orange).



**Figure S4:** LC-MS chromatograms of deletion mutants I6\_ΔlitN and I6\_ΔlitO showing the extracted masses of the derivatives of lipothrenin A (1, 5, 6, red), lipothrenin B (2, 3, 7-10, green) and lipothrenin C (4, 11-14, blue).

## 2 Physical information of isolated lipothrenins

lipothrenin A (1): White powder; 3.0 mg;  $[\alpha]_D^{20} +7.5$  (c 0.24, MeOH); UV:  $\lambda = 200$  nm;  $^1\text{H}$  and  $^{13}\text{C}$  NMR data, see Table S4; ESI-TOF-MS m/z 404.266  $[\text{M}+\text{H}]^+$  (calc. for  $\text{C}_{20}\text{H}_{38}\text{NO}_7$  404.265).

lipothrenin B (2, 3): White powder; 11.7 mg; UV:  $\lambda = 200$  nm;  $^1\text{H}$  and  $^{13}\text{C}$  NMR data, see Table S5; ESI-TOF-MS m/z 388.271  $[\text{M}+\text{H}]^+$  (calc. for  $\text{C}_{20}\text{H}_{38}\text{NO}_6$  388.270).

Z-lipothrenin C (4): White powder; 2.0 mg; UV:  $\lambda = 200$  nm;  $^1\text{H}$  and  $^{13}\text{C}$  NMR data, see Table S6; ESI-TOF-MS m/z 370.260  $[\text{M}+\text{H}]^+$  (calc. for  $\text{C}_{20}\text{H}_{36}\text{NO}_5$  370.259).

2-NAC-D-lipothrenin A (5): White powder; 16.6 mg; UV:  $\lambda = 200$  nm;  $^1\text{H}$  and  $^{13}\text{C}$  NMR data, see Table S7; ESI-TOF-MS m/z 565.278  $[\text{M}+\text{H}]^+$  (calc. for  $\text{C}_{25}\text{H}_{44}\text{N}_2\text{O}_{10}\text{S}$  565.279).

2-NAC-D-lipothrenin A<sub>1</sub> (6): White powder; 13.9 mg; UV:  $\lambda = 200$  nm;  $^1\text{H}$  and  $^{13}\text{C}$  NMR data, see Table S8; ESI-TOF-MS m/z 565.278  $[\text{M}+\text{H}]^+$  (calc. for  $\text{C}_{25}\text{H}_{44}\text{N}_2\text{O}_{10}\text{S}$  565.279).

2-NAC-D-lipothrenin B (7): White powder; 5.3 mg; UV:  $\lambda = 200$  nm;  $^1\text{H}$  and  $^{13}\text{C}$  NMR data, see Table S9; ESI-TOF-MS m/z 549.284  $[\text{M}+\text{H}]^+$  (calc. for  $\text{C}_{25}\text{H}_{45}\text{N}_2\text{O}_9\text{S}$  549.285).

2-NAC-L-lipothrenin B (8): White powder; 5.8 mg; UV:  $\lambda = 200$  nm;  $^1\text{H}$  and  $^{13}\text{C}$  NMR data, see Table S9; ESI-TOF-MS m/z 549.284  $[\text{M}+\text{H}]^+$  (calc. for  $\text{C}_{25}\text{H}_{45}\text{N}_2\text{O}_9\text{S}$  549.285).

2-NAC-D/L-lipothrenin B mixture (9, 10): White powder; 13.0 mg; UV:  $\lambda = 200$  nm;  $^1\text{H}$  and  $^{13}\text{C}$  NMR data, see Table S11; ESI-TOF-MS m/z 549.284  $[\text{M}+\text{H}]^+$  (calc. for  $\text{C}_{25}\text{H}_{45}\text{N}_2\text{O}_9\text{S}$  549.285).

2-NAC-Z-lipothrenin C and C<sub>1</sub> mixture (11, 12): White powder; 6.5 mg; UV:  $\lambda = 200$  nm;  $^1\text{H}$  and  $^{13}\text{C}$  NMR data, see Table S12; ESI-TOF-MS m/z 531.272  $[\text{M}+\text{H}]^+$  (calc. for  $\text{C}_{25}\text{H}_{43}\text{N}_2\text{O}_8\text{S}$  531.274).

2-NAC-E-lipothrenin C and C<sub>1</sub> mixture (13, 14): White powder; 4.0 mg; UV:  $\lambda = 200$  nm;  $^1\text{H}$  and  $^{13}\text{C}$  NMR data, see Table S13; ESI-TOF-MS m/z 531.272  $[\text{M}+\text{H}]^+$  (calc.  $\text{C}_{25}\text{H}_{43}\text{N}_2\text{O}_8\text{S}$  531.274).

iso-lipothrenin A (15): White powder; 5.0 mg;  $[\alpha]_D^{20} +7.5$  (c 0.22, MeOH); UV:  $\lambda = 200$  nm;  $^1\text{H}$  and  $^{13}\text{C}$  NMR data, see Table S14; ESI-TOF-MS m/z 404.266  $[\text{M}+\text{H}]^+$  (calc. for  $\text{C}_{20}\text{H}_{38}\text{NO}_7$  404.265).

14-methyl-lipothrenin A (16): White powder; 19.1 mg;  $[\alpha]_D^{20} +7.3$  (c 1.71, MeOH); UV:  $\lambda = 200$  nm;  $^1\text{H}$  and  $^{13}\text{C}$  NMR data, see Table S15; ESI-TOF-MS m/z 418.280  $[\text{M}+\text{H}]^+$  (calc. for  $\text{C}_{21}\text{H}_{40}\text{NO}_7$  418.280).

15-methyl-lipothrenin A (17): White powder; 31.2 mg;  $[\alpha]_D^{20} +2.0$  (c 2.92, MeOH); UV:  $\lambda = 200$  nm;  $^1\text{H}$  and  $^{13}\text{C}$  NMR data, see Table S16; ESI-TOF-MS m/z 418.280  $[\text{M}+\text{H}]^+$  (calc. for  $\text{C}_{21}\text{H}_{40}\text{NO}_7$  418.280).

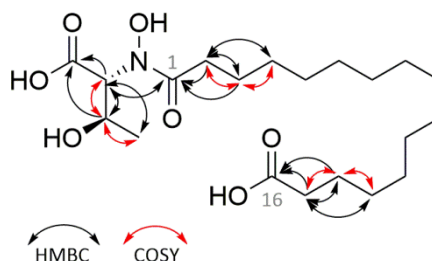


### 3 NMR Spectroscopy

#### 3.1 Lipothrenins isolated from *Streptomyces aureus* LU18118

**Table S4:** NMR data (500 MHz, DMSO-*d*<sub>6</sub>) for lipothrenin A (**1**)

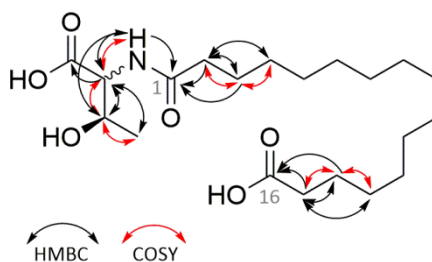
No	$\delta_C$ (type)	$\delta_H$ , mult. ( <i>J</i> in Hz)
Thr	171.6, C	
	63.3, CH	4.50, d (8.8)
	64.8, CH	4.03, dq (6.3, 8.8)
	20.4, CH <sub>3</sub>	1.01, d (6.3)
1	173.5, C	
2	31.6, CH <sub>2</sub>	2.34, m
3	24.2, CH <sub>2</sub>	1.47, m
4-13	28.6-29.1, CH <sub>2</sub>	1.25, bs (20H)
14	24.5, CH <sub>2</sub>	1.47, m
15	33.7, CH <sub>2</sub>	2.18, t (7.4)
16	174.5, C	



**Table S5:** NMR data (500 MHz, DMSO-*d*<sub>6</sub>) for D- and L-lipothrenin B mixture (**2**, **3**)

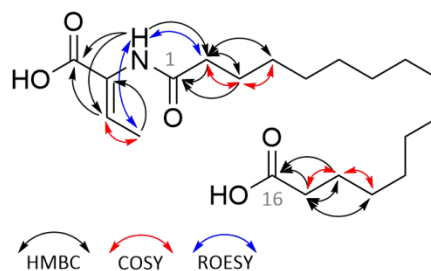
D <sup>1</sup> -lipothrenin B ( <b>2</b> )			L <sup>1</sup> -lipothrenin B ( <b>3</b> )	
No	$\delta_C$ (type)	$\delta_H$ , mult. ( <i>J</i> in Hz)	$\delta_C$ (type)	$\delta_H$ , mult. ( <i>J</i> in Hz)
Thr	172.4, C		172.4, C	
	58.3, CH	4.15, dd (8.4, 6.3)	57.4, CH	4.19, dd (3.4, 8.7)
	66.8, CH	3.84, q (6.3)	66.3, CH	4.09, dq (3.4, 6.3)
	19.8, CH <sub>3</sub>	1.06, d (6.4)	20.3 CH <sub>3</sub>	1.03, d (6.4)
	NH	7.86, d (8.2)	NH	7.66, d (8.7)
1	172.3, C		172.6, C	
2	35.1, CH <sub>2</sub>	2.12, m	35.1, CH <sub>2</sub>	2.12, m
		2.17, m		2.17, m
3	25.3, CH <sub>2</sub>	1.47, m	25.3, CH <sub>2</sub>	1.47, m
4-13	28.6-29.1, CH <sub>2</sub>	1.23, bs (20H)	28.6-29.1, CH <sub>2</sub>	1.23, bs (20H)
14	24.5, CH <sub>2</sub>	1.47, m	24.5, CH <sub>2</sub>	1.47, m
15	33.7, CH <sub>2</sub>	2.18, t (7.5)	33.7, CH <sub>2</sub>	2.18, t (7.5)
16	174.6, C	$\delta_H$ , mult. ( <i>J</i> in Hz)	174.6, C	

<sup>1</sup>Stereochemistry assignment based on 2-NAC-lipothrenine B



**Table S6:** NMR data (500 MHz, DMSO-*d*<sub>6</sub>) for Z-lipothrenin C (4)

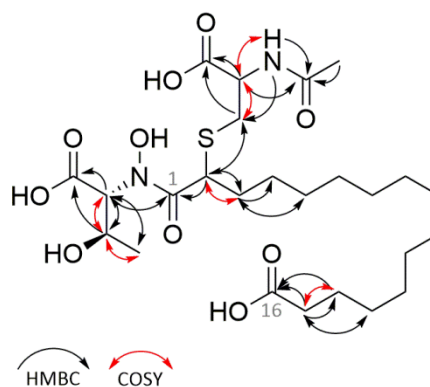
No	$\delta_c$ (type)	$\delta_H$ , mult. ( <i>J</i> in Hz)
Dhb	165.8, C	
	129.0, C	
	130.6, CH	6.46, q (7.1)
	13.5, CH <sub>3</sub>	1.62, d (7.1)
	NH	8.92, s
1	171.0, C	
2	35.1, CH <sub>2</sub>	2.19, t (7.3)
3	25.2, CH <sub>2</sub>	1.51, m
4-13	28.5-29.0, CH <sub>2</sub>	1.30, bs (20H)
14	24.5, CH <sub>2</sub>	1.47, m
15	33.7, CH <sub>2</sub>	2.18, t (7.3)
16	174.5, C	



## 1.2 2-NAC-Lipothrenins isolated from *Streptomyces aureus* LU18118.

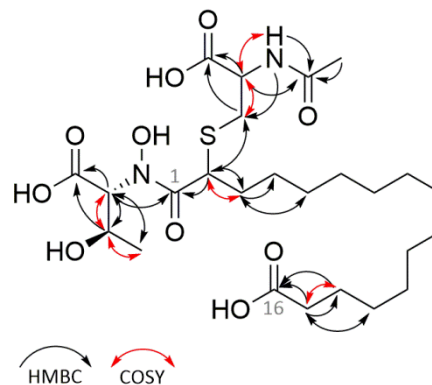
**Table S7:** NMR data (500 MHz, DMSO-*d*<sub>6</sub>) for 2-NAC-D-lipothrenin A (5)

No	$\delta_c$ (type)	$\delta_H$ , mult. ( <i>J</i> in Hz)
Thr	170.4, C	
	63.5, CH	4.61, d (8.5)
	66.4, CH	4.14, m (6.3, 8.5)
	20.1, CH <sub>3</sub>	1.09, d (6.3)
Cys	172.0, C	
	52.5, CH	4.31, dt (5.4, 8.2)
	31.1, CH <sub>2</sub>	2.97, dd (5.2, 13.4)
	NH	2.74, dd (8.6, 13.3)
Ac	169.4, C	
	22.4, CH <sub>3</sub>	1.84, s
1	171.9, C	
2	41.4, CH	3.92 dd (6.0, 9.0)
3	31.5, CH <sub>2</sub>	1.58, m
		1.76, m
4	26.7, CH <sub>2</sub>	1.31, m
5-13	28.6-29.1, CH <sub>2</sub>	1.23, bs
14	24.5, CH <sub>2</sub>	1.47, m
15	33.7, CH <sub>2</sub>	2.18, t (7.4)
16	174.6, C	

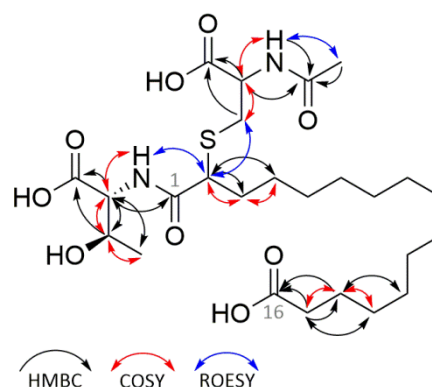


**Table S8:** NMR data (500 MHz, DMSO-*d*<sub>6</sub>) for 2-NAC-D-lipothrenin A<sub>1</sub> (**6**)

No	$\delta_c$ (type)	$\delta_H$ , mult. ( <i>J</i> in Hz)
Thr	170.5, C	
	64.1, CH	4.62, d (8.5)
	64.5, CH	4.18, m (6.6, 8.8)
	20.1, CH <sub>3</sub>	1.07, d (6.3)
Cys	172.3, C	
	52.1, CH	4.40, dt (4.7, 8.5)
	31.4, CH <sub>2</sub>	3.07, dd (4.9, 13.4)
	NH	2.78, dd (9.0, 13.4)
Ac	169.4, C	
	22.2, CH <sub>3</sub>	1.86, s
1	172.3, C	
2	41.6, CH	3.98 dd (6.3, 8.5)
3	31.7, CH <sub>2</sub>	1.56, m
		1.76, m
4	26.8, CH <sub>2</sub>	1.30, m
5 – 13	28.7 – 29.2, CH <sub>2</sub>	1.23, bs
14	24.6, CH <sub>2</sub>	1.47, m
15	33.7, CH <sub>2</sub>	2.18, t (7.4)
16	174.6, C	

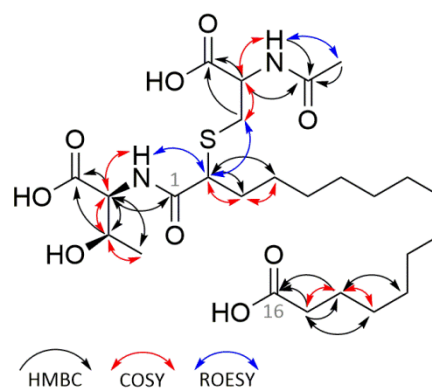
**Table S9:** NMR data (500 MHz, DMSO-*d*<sub>6</sub>) for 2-NAC-D-lipothrenin B (**7**)

No	$\delta_c$ (type)	$\delta_H$ , mult. ( <i>J</i> in Hz)
Thr	171.2, C	
	57.8, CH	4.16, dd (6.0, 8.3)
	66.2, CH	3.87, m (6.3)
	19.2, CH <sub>3</sub>	1.11, d (6.4)
	NH	8.09, d (8.1)
Cys	171.5, C	
	51.8, CH	4.32, dt (5.1, 8.2)
	31.2, CH <sub>2</sub>	2.94, dd (5.2, 13.4)
	NH	2.78, dd (8.6, 13.3)
Ac	168.6, C	
	21.8, CH <sub>3</sub>	1.84, s
1	170.7, C	
2	46.2, CH	3.43 dd (6.0, 9.0)
3	31.3, CH <sub>2</sub>	1.69, m
		1.50, m
4	26.0, CH <sub>2</sub>	1.25, m
5 – 13	27.8 – 28.9, CH <sub>2</sub>	1.23, bs
14	23.8, CH <sub>2</sub>	1.47, m
15	33.0, CH <sub>2</sub>	2.18, t (7.4)
16	173.8, C	



**Table S10:** NMR data (500 MHz, DMSO-*d*<sub>6</sub>) for 2-NAC-L-lipothrenin B (**8**)

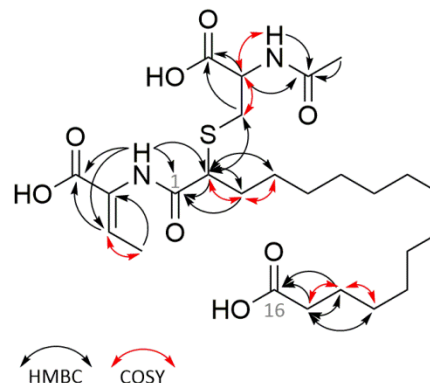
No	$\delta_c$ (type)	$\delta_H$ , mult. ( <i>J</i> in Hz)
Thr	171.4, C	
	56.9, CH	4.25, dd (3.0, 9.0)
	65.6, CH	4.16, dq (3.2, 6.6)
	19.6, CH <sub>3</sub>	1.03, d (6.4)
	NH	7.98, d (9.0)
Cys	171.5, C	
	52.1, CH	4.33, dt (5.2, 8.3)
	31.1, CH <sub>2</sub>	2.96, dd (5.1, 13.5)
		2.88, dd (8.7, 13.6)
	NH	8.19, d (7.9)
Ac	168.6, C	
	21.9, CH <sub>3</sub>	1.85, s
1	171.0, C	
2	46.4, CH	3.57, dd (5.8, 9.4)
3	31.2, CH <sub>2</sub>	1.69, m
		1.49, m
4	26.2, CH <sub>2</sub>	1.25, m
5 – 13	27.8 – 28.9, CH <sub>2</sub>	1.23, bs
14	23.8, CH <sub>2</sub>	1.47, m
15	33.0, CH <sub>2</sub>	2.18, (7.4)
16	173.8, C	

**Table S11:** NMR data (500 MHz, DMSO-*d*<sub>6</sub>) for 2-NAC-D-/L-lipothrenin B<sub>1</sub> mixture (**9**, **10**)

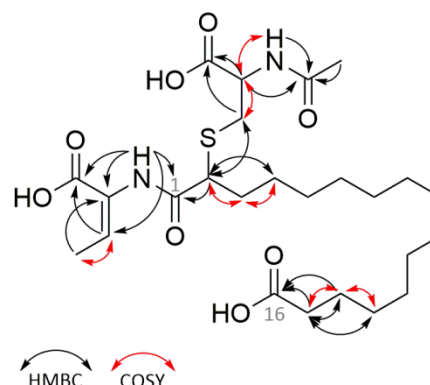
2-NAC-D-lipothrenin B <sub>1</sub>			2-NAC-L-lipothrenin B <sub>1</sub>	
No	$\delta_c$ (type)	$\delta_H$ , mult. ( <i>J</i> in Hz)	$\delta_c$ (type)	$\delta_H$ , mult. ( <i>J</i> in Hz)
Thr	171.9, C		172.1, C	
	58.4, CH	4.19, dd (6.3, 8.2)	57.6, CH	4.20, dd (3.6, 6.3)
	66.8, CH	3.87, m (6.3)	66.3, CH	4.16, dq (3.0, 6.6)
	19.7, CH <sub>3</sub>	1.08, d (6.0)	20.4, CH <sub>3</sub>	1.07, d (6.6)
	NH	8.13, d (8.8)	NH	7.95, d (8.8)
Cys	172.1, C		172.2, C	
	51.7, CH	4.37, m	51.8, CH	4.37, m
	31.7, CH <sub>2</sub>	2.99, dd (4.7, 13.2)	31.7, CH <sub>2</sub>	2.99, dd (4.7, 13.2)
		2.74, m		2.73, m
	NH	8.18, d (8.0)	NH	8.18, d (8.0)
Ac	169.3, C		169.3	
	22.3, CH <sub>3</sub>	1.84, s	22.3, CH <sub>3</sub>	1.84, s
1	171.3, C		171.9, C	
2	46.6, CH	3.42 dd (5.9, 8.9)	46.8, CH	3.54, dd (6.0, 8.8)
3	31.8, CH <sub>2</sub>	1.69, m	31.8, CH <sub>2</sub>	1.69, m
		1.49, m		1.49, m
4	26.7, CH <sub>2</sub>	1.22, m	26.7, CH <sub>2</sub>	1.22, m
5 – 13	28.6 – 29.1, CH <sub>2</sub>	1.23, bs	28.6 – 29.1, CH <sub>2</sub>	1.23, bs
14	24.5, CH <sub>2</sub>	1.47, m	24.5, CH <sub>2</sub>	1.47, m
15	33.7, CH <sub>2</sub>	2.18, t (7.4)	33.7, CH <sub>2</sub>	2.18, t (7.4)
16	174.5, C		174.5, C	

**Table S12:** NMR data (500 MHz, DMSO-*d*<sub>6</sub>) for 2-NAC-Z-lipothrenin C and C<sub>1</sub> mix (**11**, **12**)

No	$\delta_c$ (type)	$\delta_H$ , mult. ( <i>J</i> in Hz)
Dhb	165.5, C	
	128.5, C	
	131.8, CH	6.54, q (7.0)
	13.5, CH <sub>3</sub>	1.65, d (6.9)
	NH	9.19, s
Cys	172.2, C	
	51.8, CH	4.38, dt (4.7, 8.7)
	31.9, CH <sub>2</sub>	3.04, dd (4.7, 13.2)
	NH	2.78, dd (9.1, 13.2)
Ac	169.2, C	8.15, d (7.9)
	22.4, CH <sub>3</sub>	1.84, s
1	170.1, C	
2	46.6, CH	3.44, dd (5.0, 9.5)
3	31.6, CH <sub>2</sub>	1.73, m
		1.51, m
4	26.8, CH <sub>2</sub>	1.29, m
5 – 13	28.6– 29.0, CH <sub>2</sub>	1.23, bs
14	24.5, CH <sub>2</sub>	1.47, m
15	33.7, CH <sub>2</sub>	2.18, (7.4)
16	174.5, C	

**Table S13:** NMR data (500 MHz, DMSO-*d*<sub>6</sub>) for 2-NAC-E-lipothrenin C and C<sub>1</sub> (**13**, **14**)

No	$\delta_c$ (type)	$\delta_H$ , mult. ( <i>J</i> in Hz)
Dhb	165.2, C	
	128.2, C	
	125.0, CH	6.16, q (7.6)
	13.1, CH <sub>3</sub>	1.90, d (7.6)
	NH	9.37, s
Cys	171.7, C	
	52.3, CH	4.30, dt (5.3, 8.1)
	31.5, CH <sub>2</sub>	2.95, dd (5.6, 13.7)
	NH	2.81, dd (8.3, 13.2)
Ac	168.9, C	8.17, d (8.6)
	22.3, CH <sub>3</sub>	1.84, s
1	169.8, C	
2	47.0, CH	3.38, dd (6.0, 8.7)
3	31.4, CH <sub>2</sub>	1.70, m
		1.50, m
4	26.3, CH <sub>2</sub>	1.25, m
5 – 13	27.8– 28.8, CH <sub>2</sub>	1.23, bs
14	24.2, CH <sub>2</sub>	1.46, m
15	33.4, CH <sub>2</sub>	2.18, (7.5)
16	174.1, C	

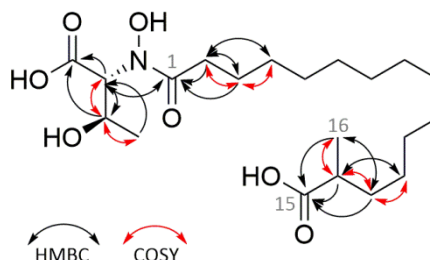


### 3.2 Lipothrenin derivatives produced by *Streptomyces aureus* I6\_ΔKS

**Table S14:** NMR data (500 MHz, DMSO-*d*<sub>6</sub>) for *Iso*-lipothrenin A (**15**)

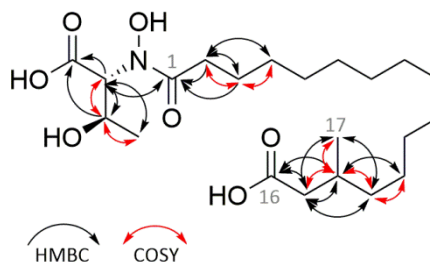
No	$\delta_c$ (type)	$\delta_H$ , mult. ( <i>J</i> in Hz)
Thr	170.7, C	
	63.2, CH	4.57, d (8.6)
	64.3, CH	4.09, dq (6.2, 8.8)
	20.1, CH <sub>3</sub>	1.03, d (ovl.)
1	173.5, C	
2	31.4, CH <sub>2</sub>	2.32, m
3	24.0, CH <sub>2</sub>	1.47, m
4-12	28.4-29.0, CH <sub>2</sub>	1.23, bs (18H)
13	33.2, CH <sub>2</sub>	1.51, m
14	38.5, CH	2.28, m
15	177.2, C	
16	16.8, CH <sub>3</sub>	1.03, d (ovl.)

ovl. = overlap



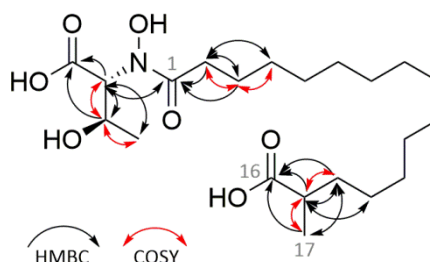
**Table S15:** NMR data (500 MHz, DMSO-*d*<sub>6</sub>) for 14-methyl-lipothrenin A (**16**)

No	$\delta_c$ (type)	$\delta_H$ , mult. ( <i>J</i> in Hz)
Thr	170.6, C	
	63.5, CH	4.61, d (8.2)
	64.3, CH	4.12, dq (6.4, 8.4)
	20.2, CH <sub>3</sub>	1.04, d (6.4)
1	173.9, C	
2	31.6, CH <sub>2</sub>	2.35, m
3	24.0, CH <sub>2</sub>	1.48, m
4-12	28.5-29.0, CH <sub>2</sub>	1.23, bs (18H)
13	36.1, CH <sub>2</sub>	1.11, m
		1.26, m
14	29.6, CH	1.79, m
15	41.4, CH <sub>2</sub>	1.98, dd (8.5, 15.0)
		2.17, dd (6.1, 15.0)
16	174.0, C	
17	19.5, CH <sub>3</sub>	0.86, d (6.7)

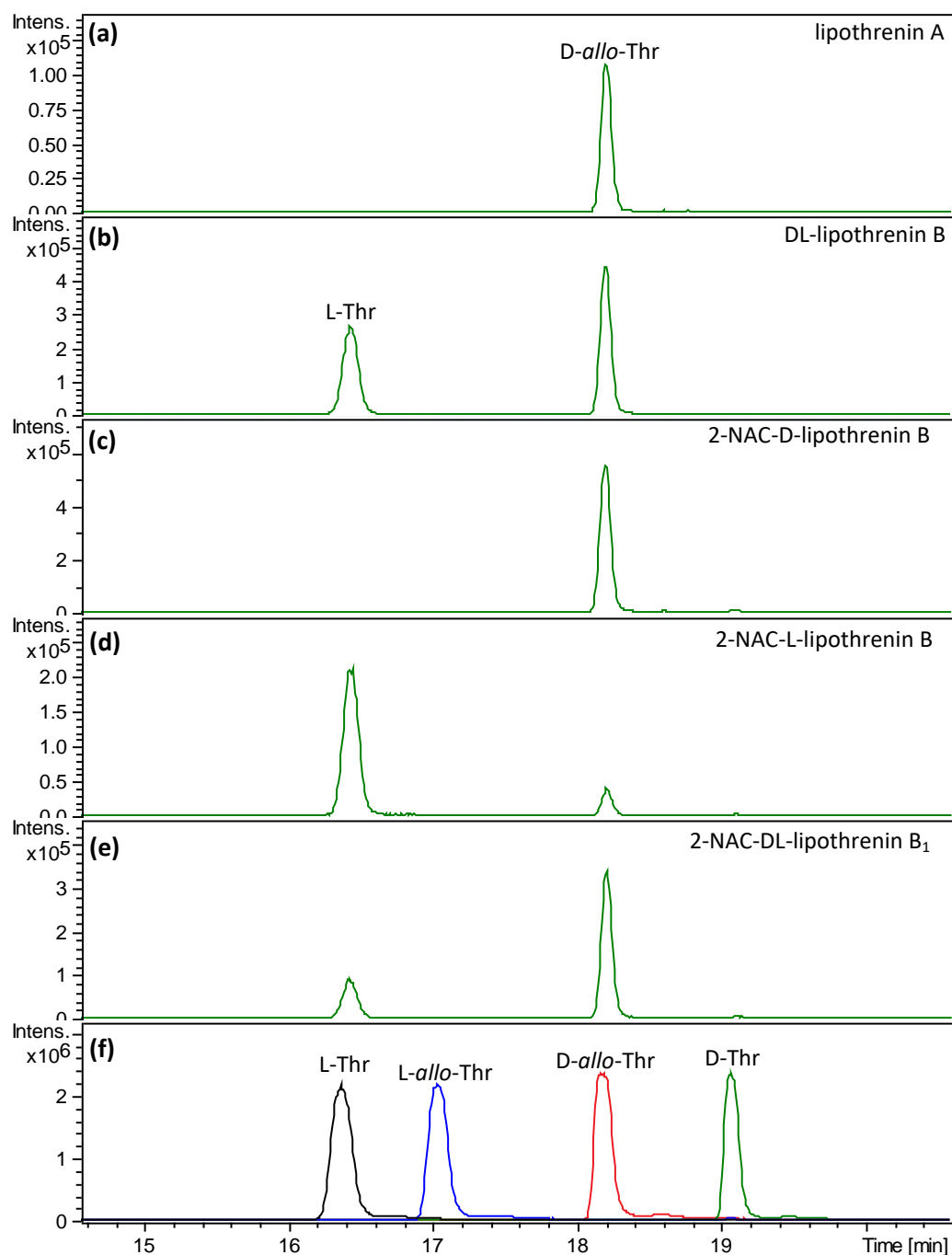


**Table S16:** NMR data (500 MHz, DMSO-*d*<sub>6</sub>) for 15-methyl-lipothrenin A (**17**)

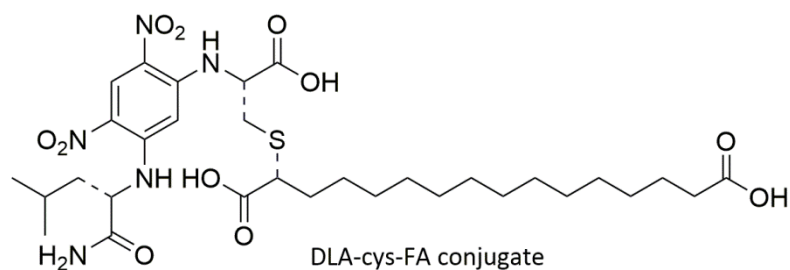
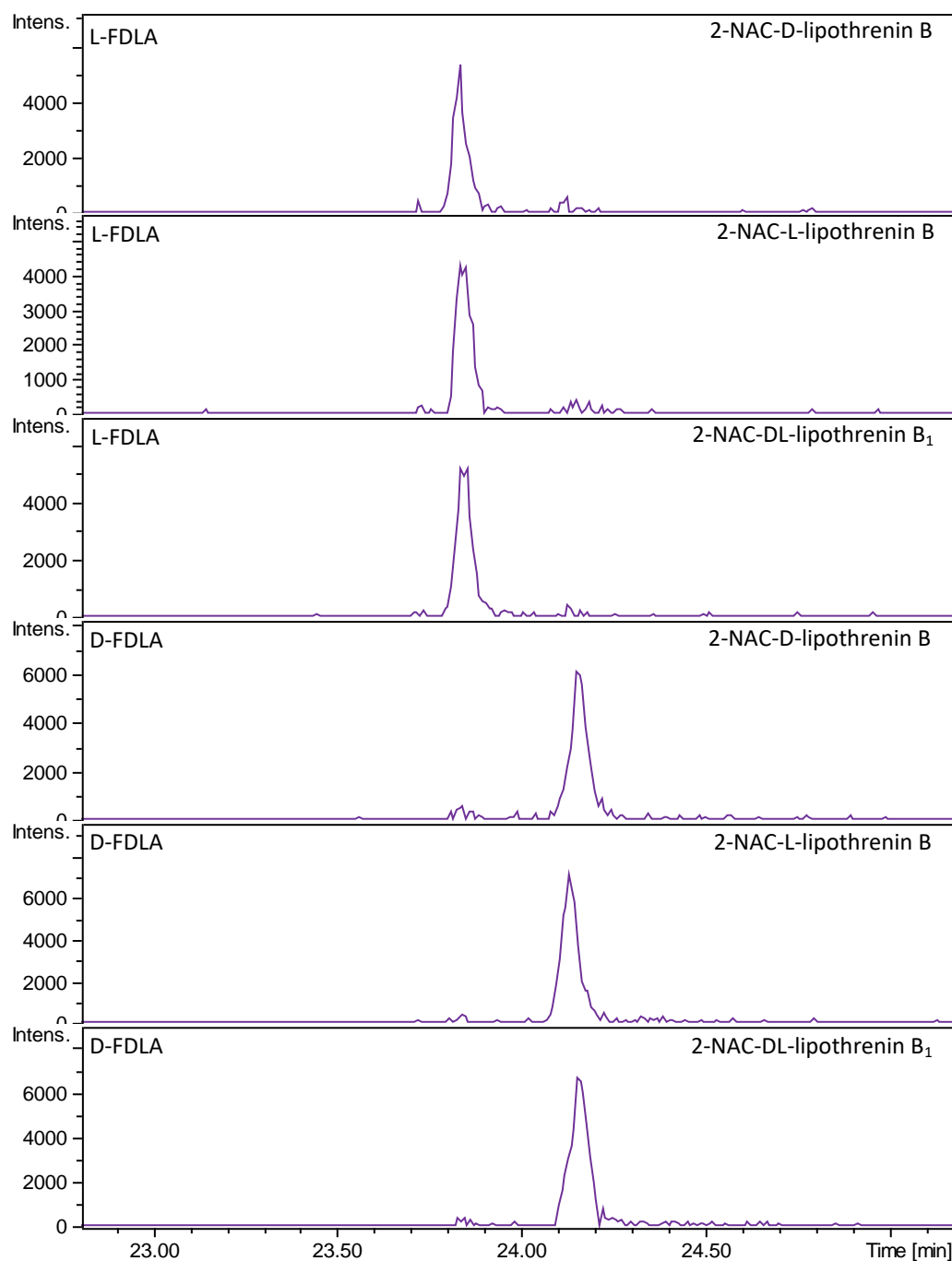
No	$\delta_c$ (type)	$\delta_H$ , mult. ( <i>J</i> in Hz)
Thr	170.6, C	
	63.5, CH	4.62, d (8.6)
	64.4, CH	4.13, dq (6.4, 8.6)
	20.2, CH <sub>3</sub>	1.04, d (6.4)
1	173.9, C	
2	33.7, CH <sub>2</sub>	2.17, t (7.3)
3	24.5, CH <sub>2</sub>	1.48, m
4-12	28.5-29.0, CH <sub>2</sub>	1.23, bs (18H)
13	26.7, CH <sub>2</sub>	1.23, m
14	33.3, CH <sub>2</sub>	1.52, m
		1.30, m
15	38.7, CH	2.27, m
16	177.5, C	
17	17.0, CH <sub>3</sub>	1.03, d (7.1)



## 4 Stereochemistry Assignment by Marfey's Method



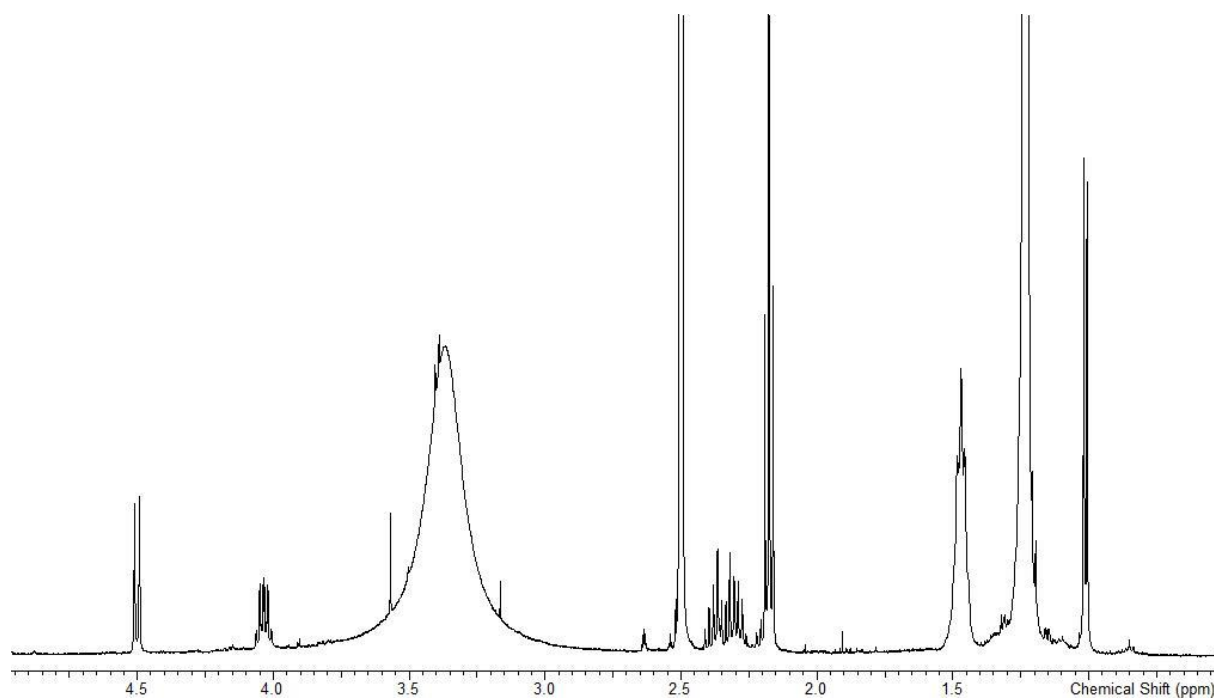
**Figure S5:** LC-MS chromatograms of hydrolysed lipothrenin A (a), B (b) 2-NAC-DL-lipothrenin B/B<sub>1</sub> (c-e) and the references L-, D-, L-*allo*- and D-*allo*-threonine (f) derivatized with L-FDLA.



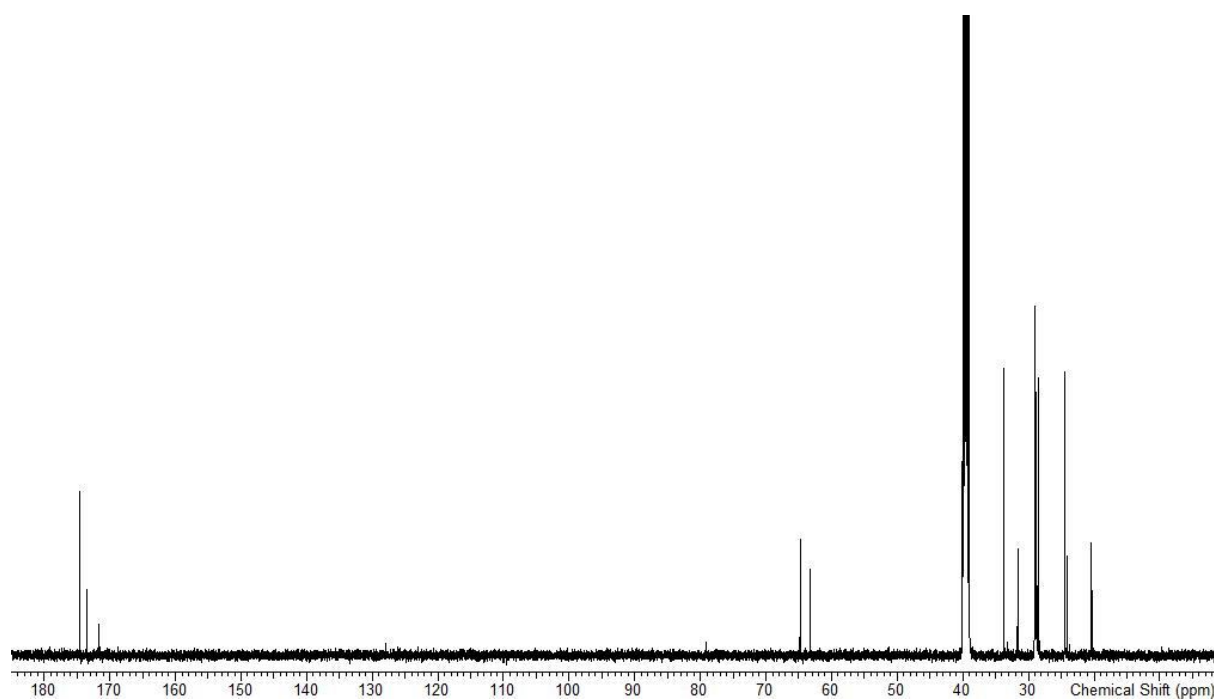
**Figure S6:** LC-MS chromatograms of the extracted masse of the DLA-cys-FA conjugate derived from hydrolysis of 2-NAC-DL-lipothrenin B/B<sub>1</sub> and derivatisation with L- and D-FDLA.



## 5 NMR Spectra



**Figure S7:**  $^1\text{H}$  NMR spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of lipothrenin A (1).



**Figure S8:**  $^{13}\text{C}$  NMR spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of lipothrenin A (1).

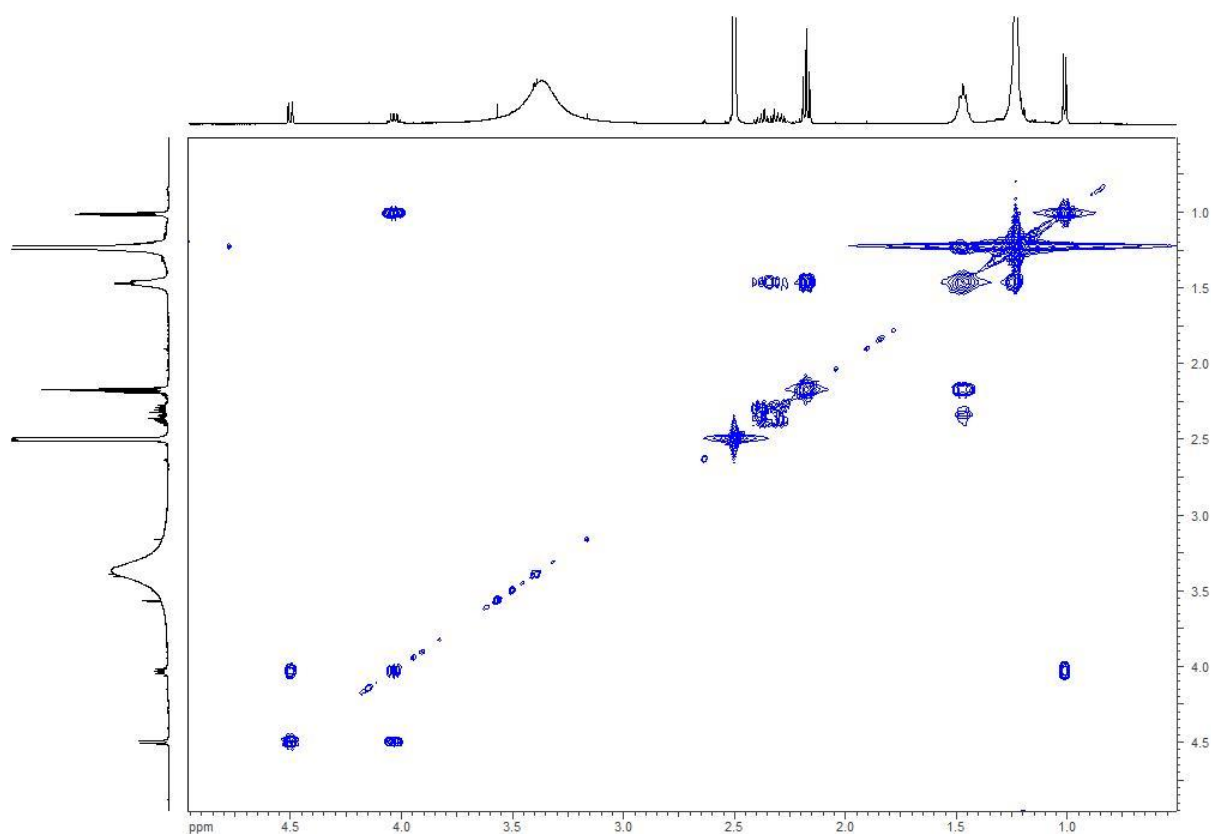


Figure S9:  $^1\text{H}$ - $^1\text{H}$ -COSY spectrum (500 MHz, DMSO- $d_6$ ) of lipothrenin A (1).

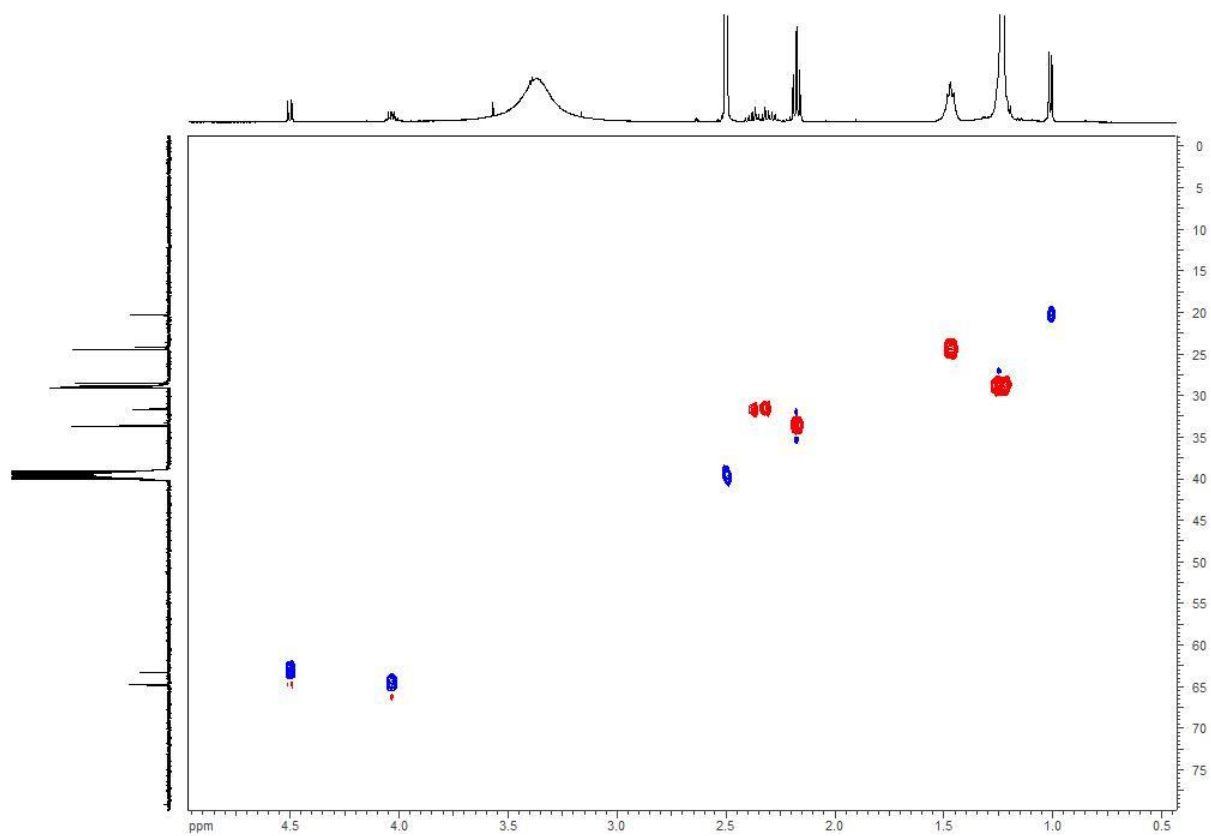
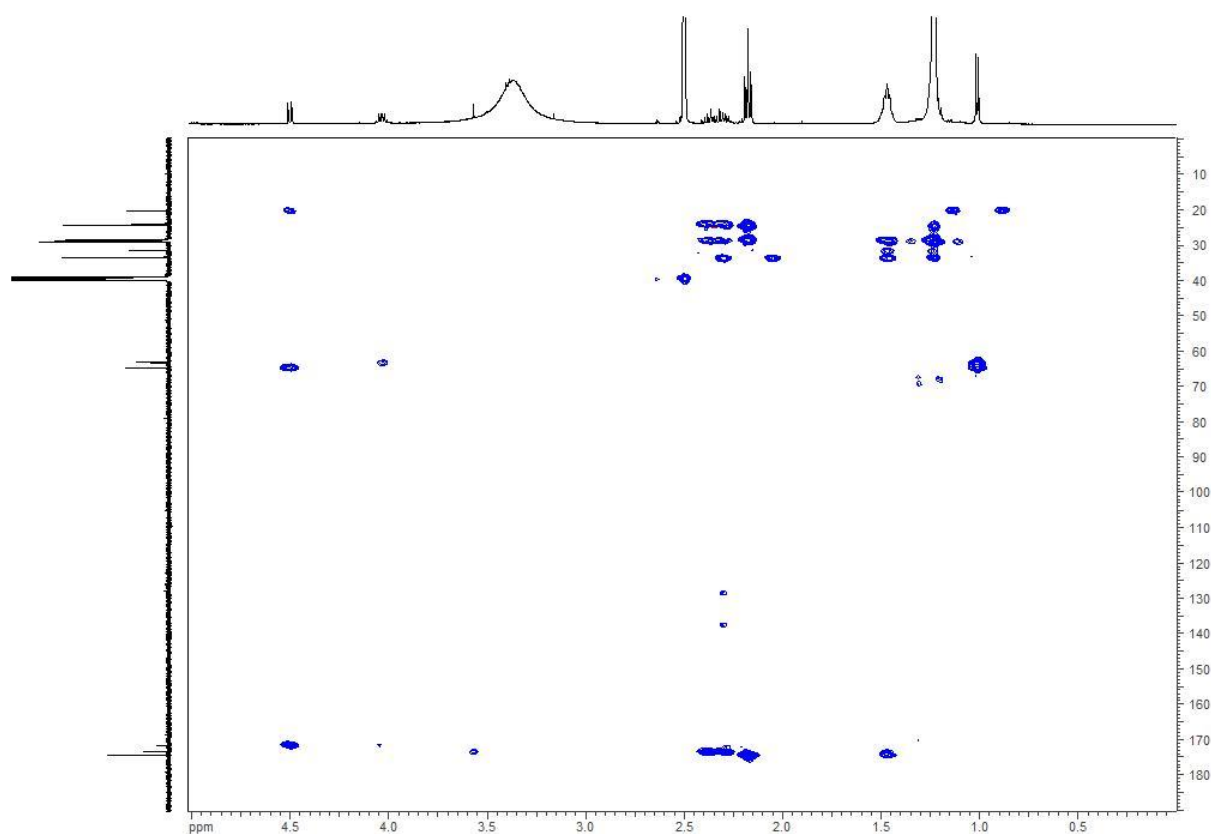
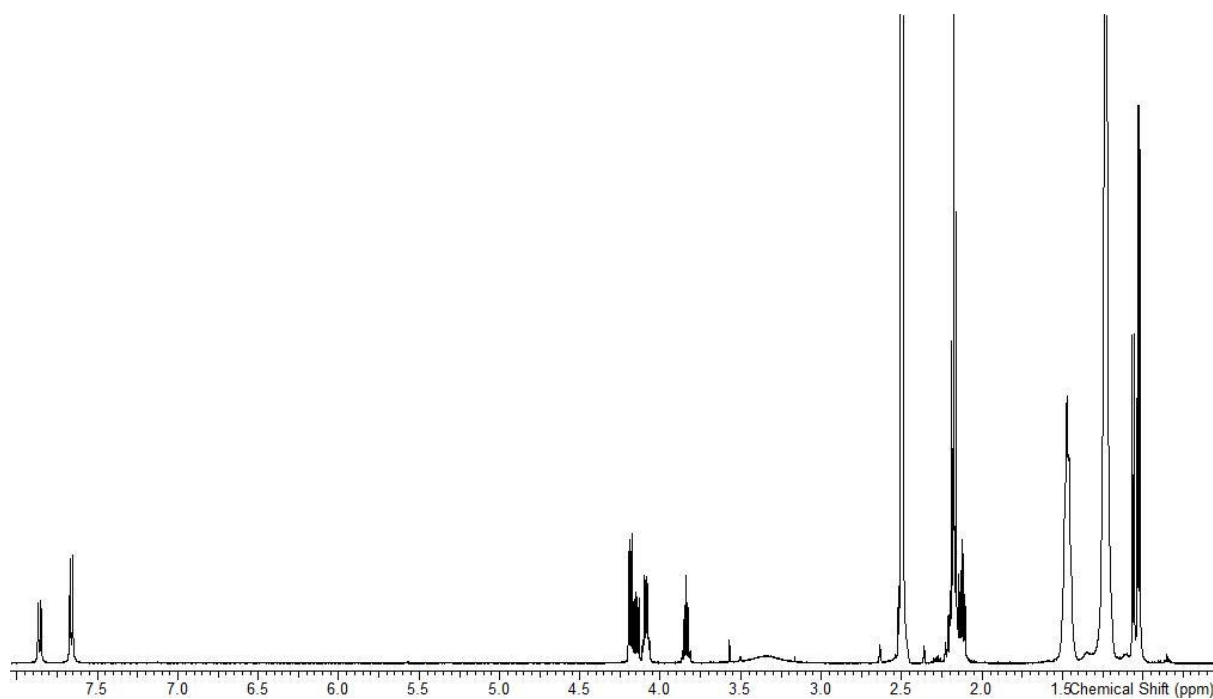


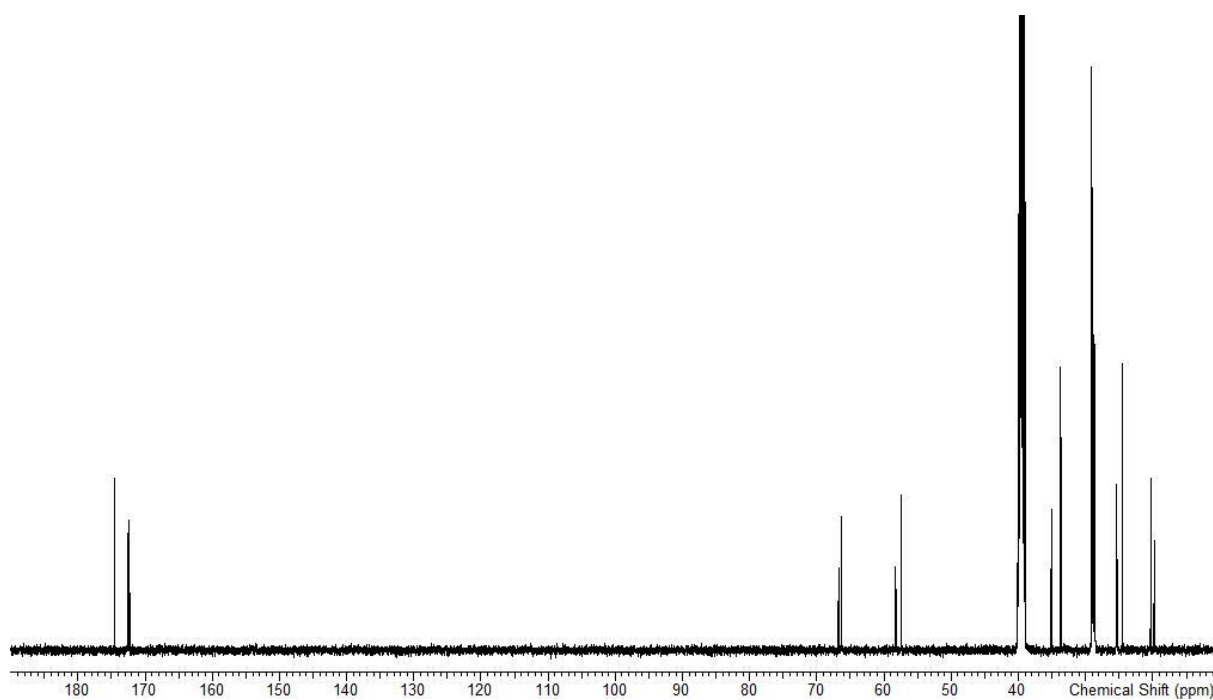
Figure S10: HSQC spectrum (500 MHz, DMSO- $d_6$ ) of lipothrenin A (1).



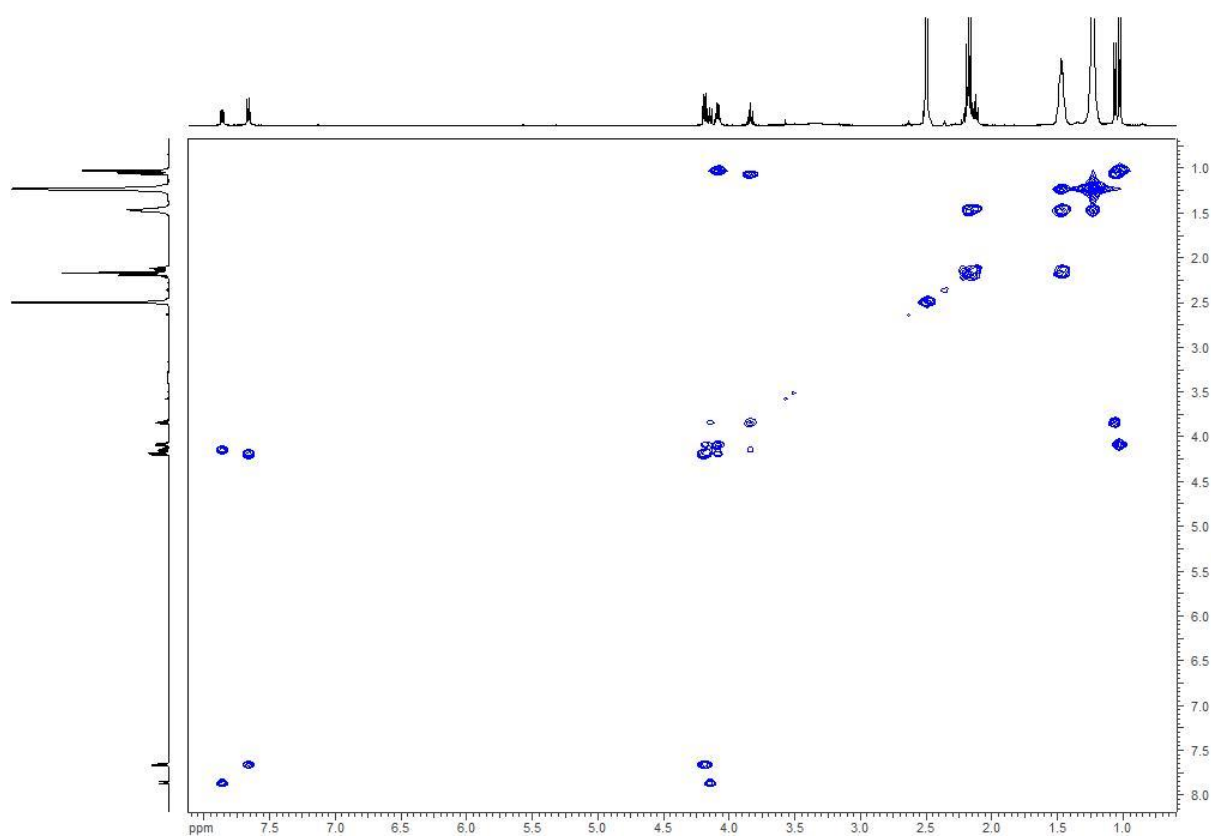
**Figure S11:** HMBC spectrum (500 MHz, DMSO- $d_6$ ) of lipothrenin A (**1**).



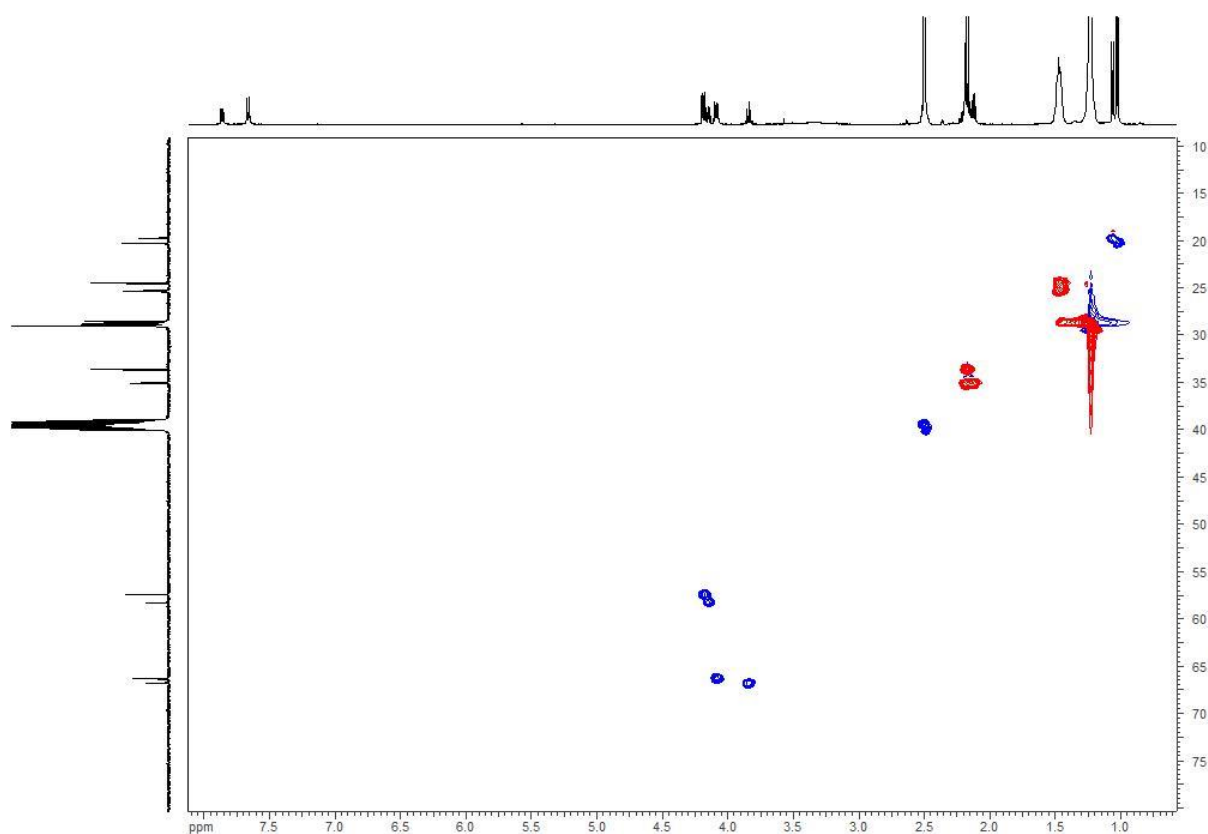
**Figure S12:**  $^1\text{H}$  NMR spectrum (500 MHz, DMSO- $d_6$ ) of D- and L-lipothrenin B (**2**, **3**).



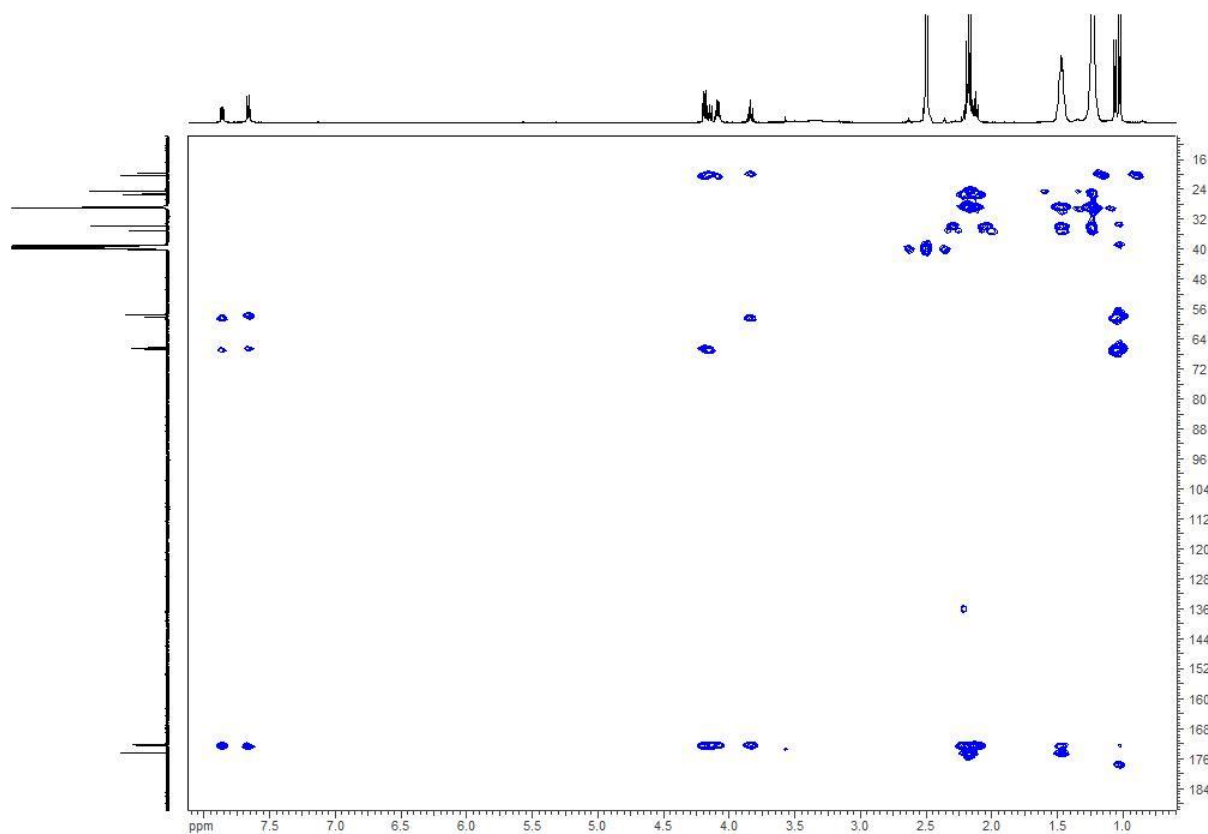
**Figure S13:**  $^{13}\text{C}$  NMR spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of D- and L-lipothrenin B (2, 3).



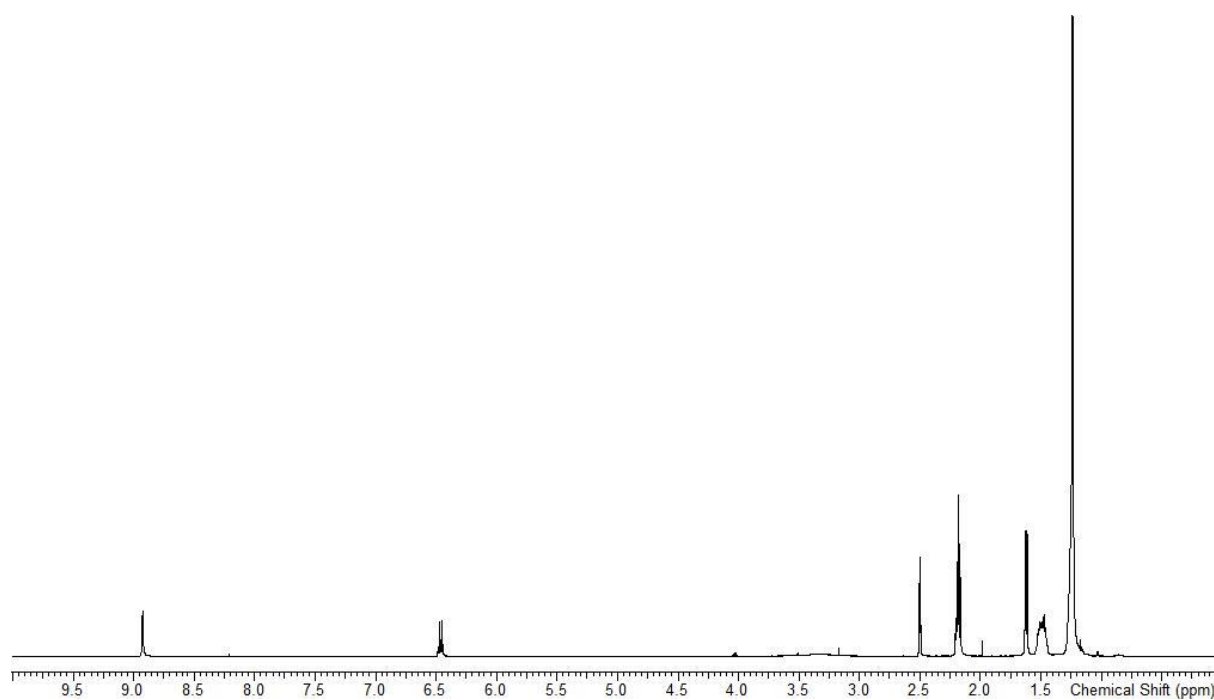
**Figure S14:**  $^1\text{H}$ - $^1\text{H}$  COSY spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of D- and L-lipothrenin B (2, 3).



**Figure S15:** HSQC spectrum (500 MHz, DMSO- $d_6$ ) of D- and L-lipothrenin B (**2**, **3**).



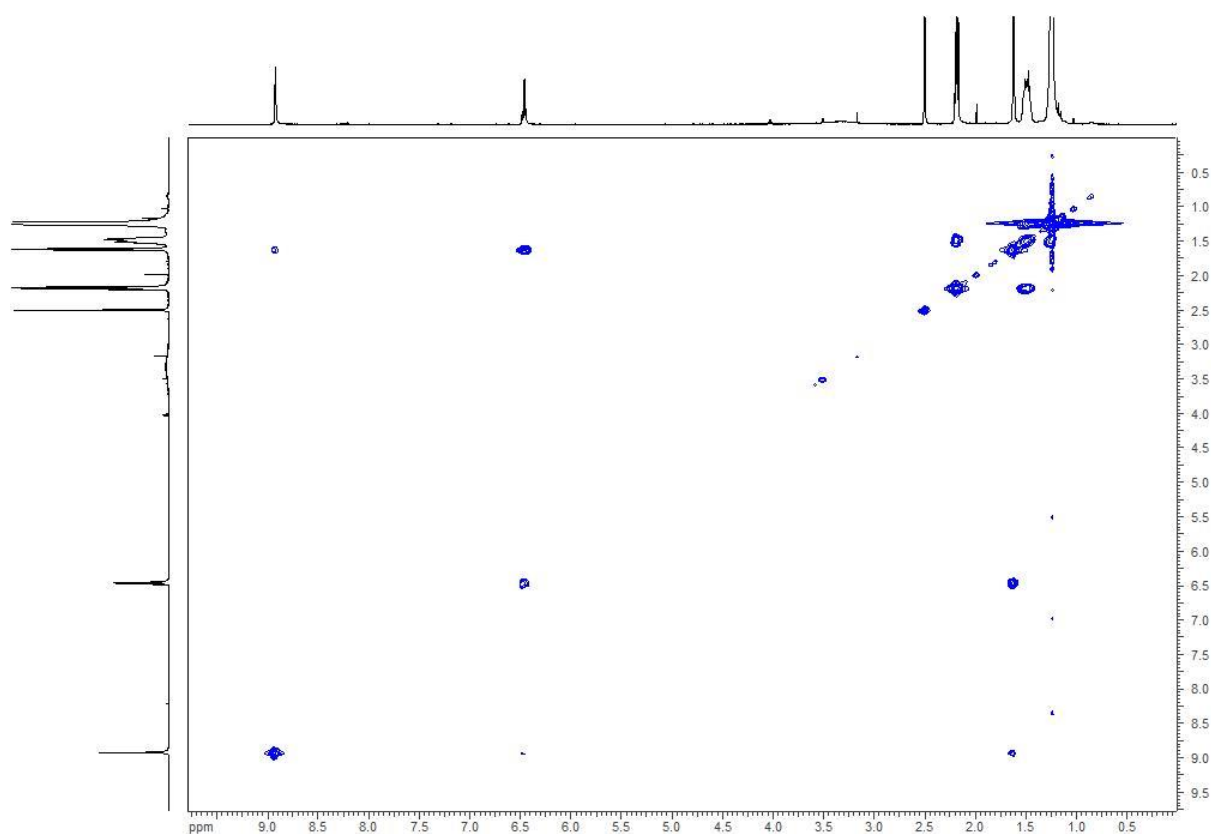
**Figure S16:** HMBC spectrum (500 MHz, DMSO- $d_6$ ) of D- and L-lipothrenin B (**2**, **3**).



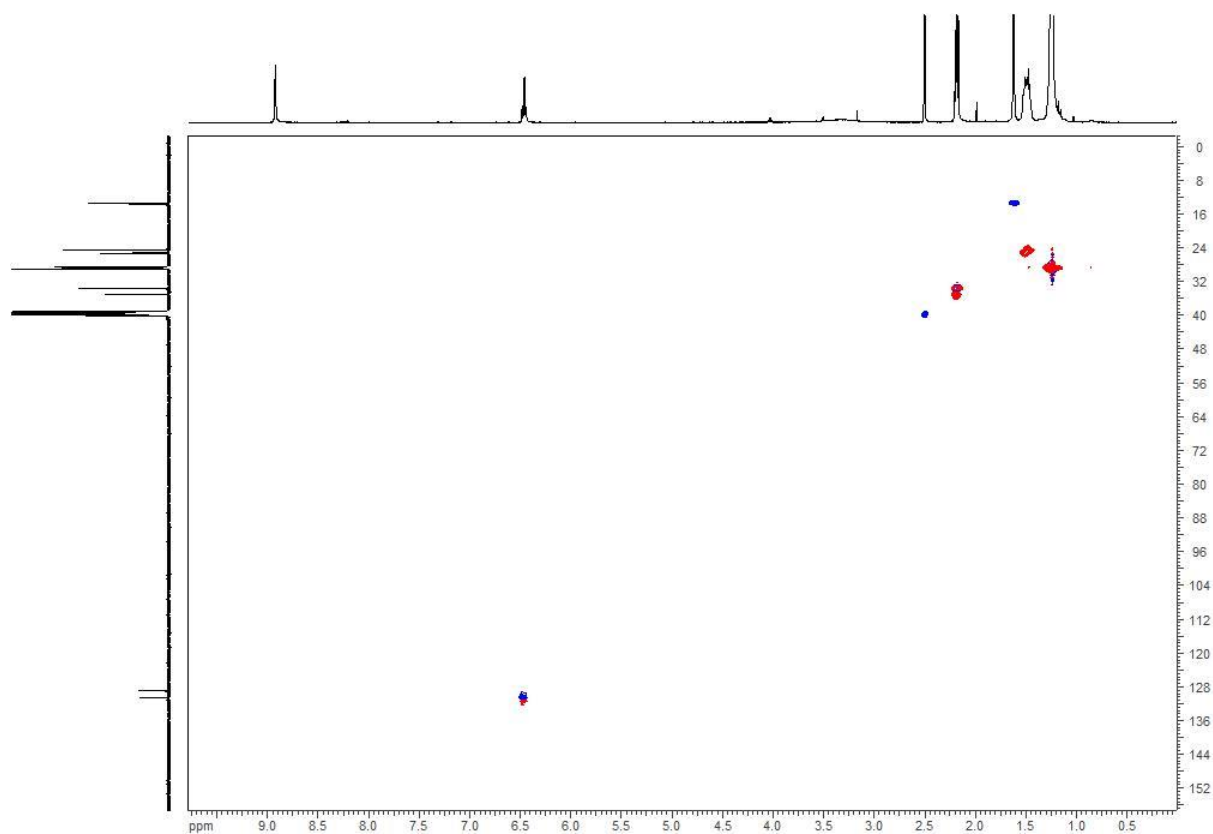
**Figure S17:**  $^1\text{H}$  NMR spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of lipothrenin C (**4**).



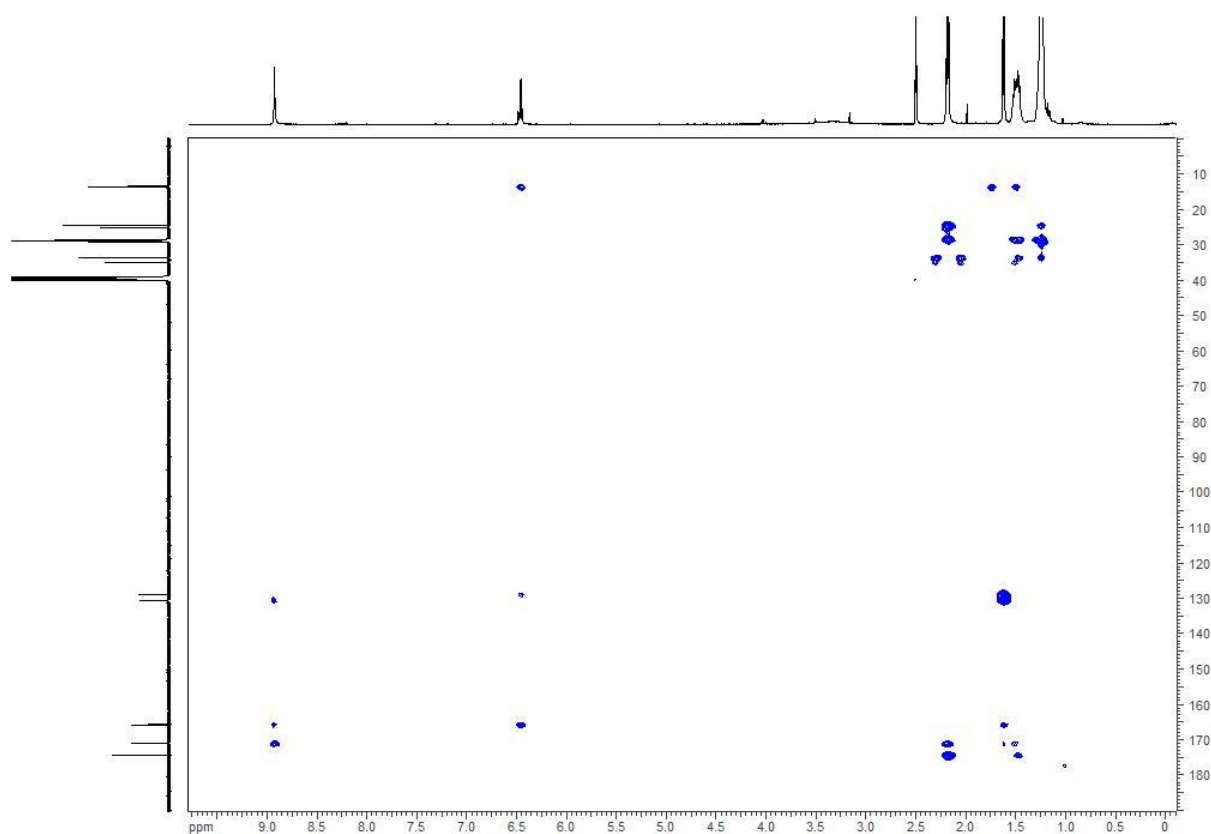
**Figure S18:**  $^{13}\text{C}$  NMR spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of lipothrenin C (**4**).



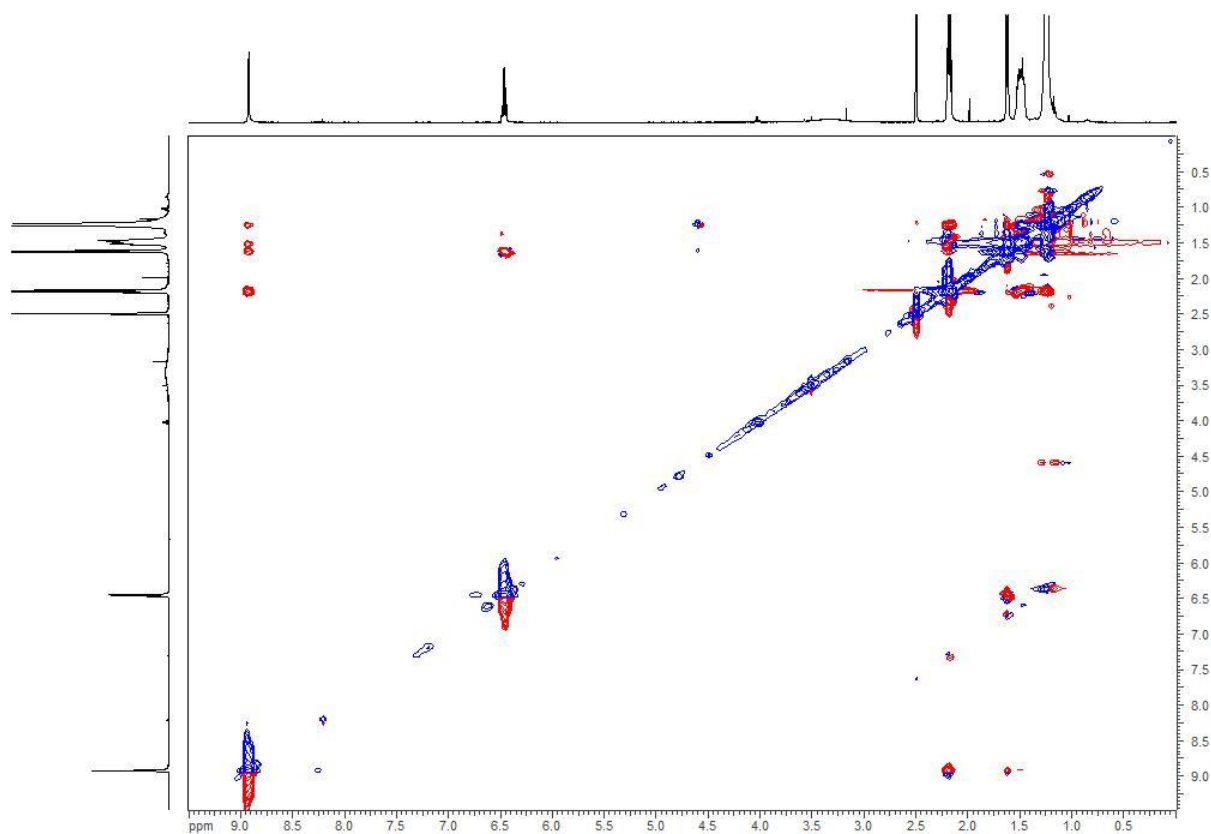
**Figure S19:**  $^1\text{H}$ - $^1\text{H}$ -COSY spectrum (500 MHz, DMSO- $d_6$ ) of lipothrenin C (4).



**Figure S20:** HSQC spectrum (500 MHz, DMSO- $d_6$ ) of lipothrenin C (4).

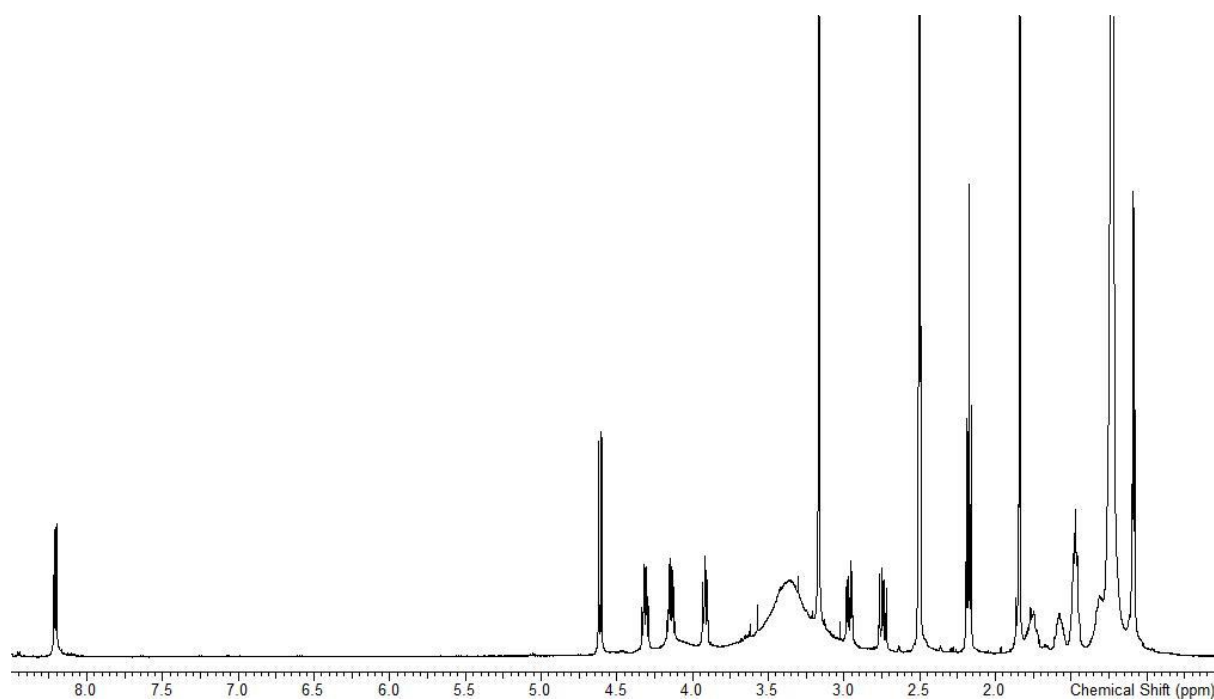


**Figure S21:** HMBC spectrum (500 MHz, DMSO- $d_6$ ) of lipothrenin C (**4**).

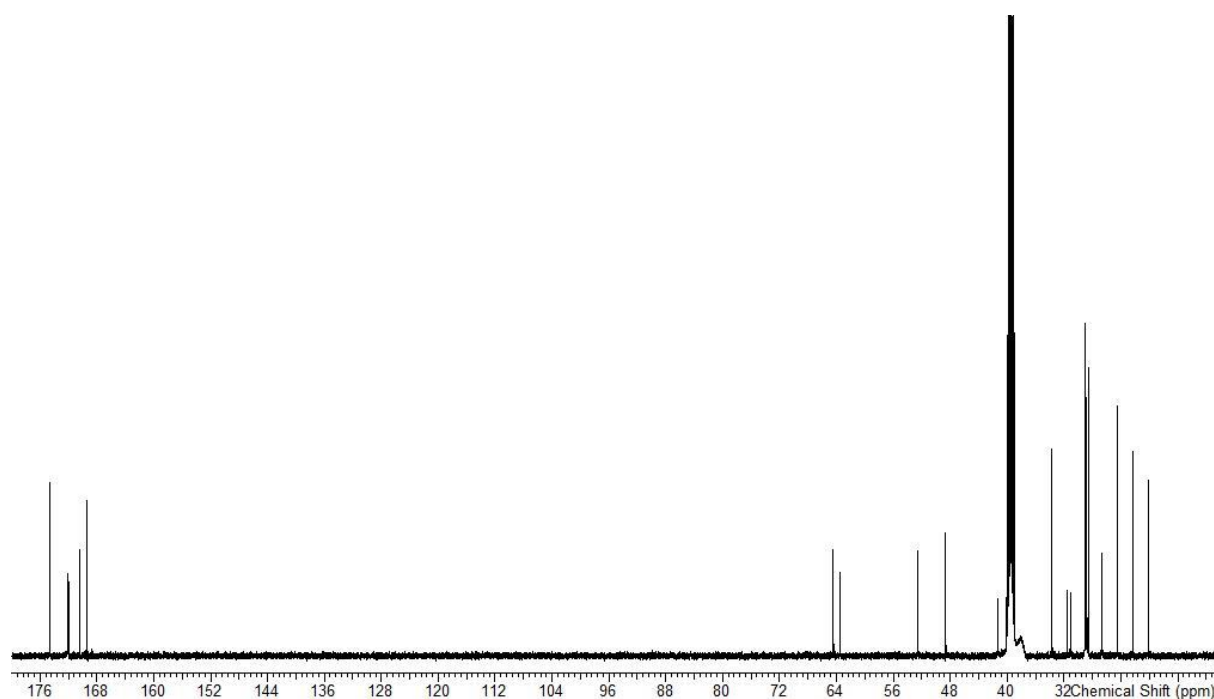


**Figure S22:** ROESY spectrum (500 MHz, DMSO- $d_6$ ) of lipothrenin C (**4**).

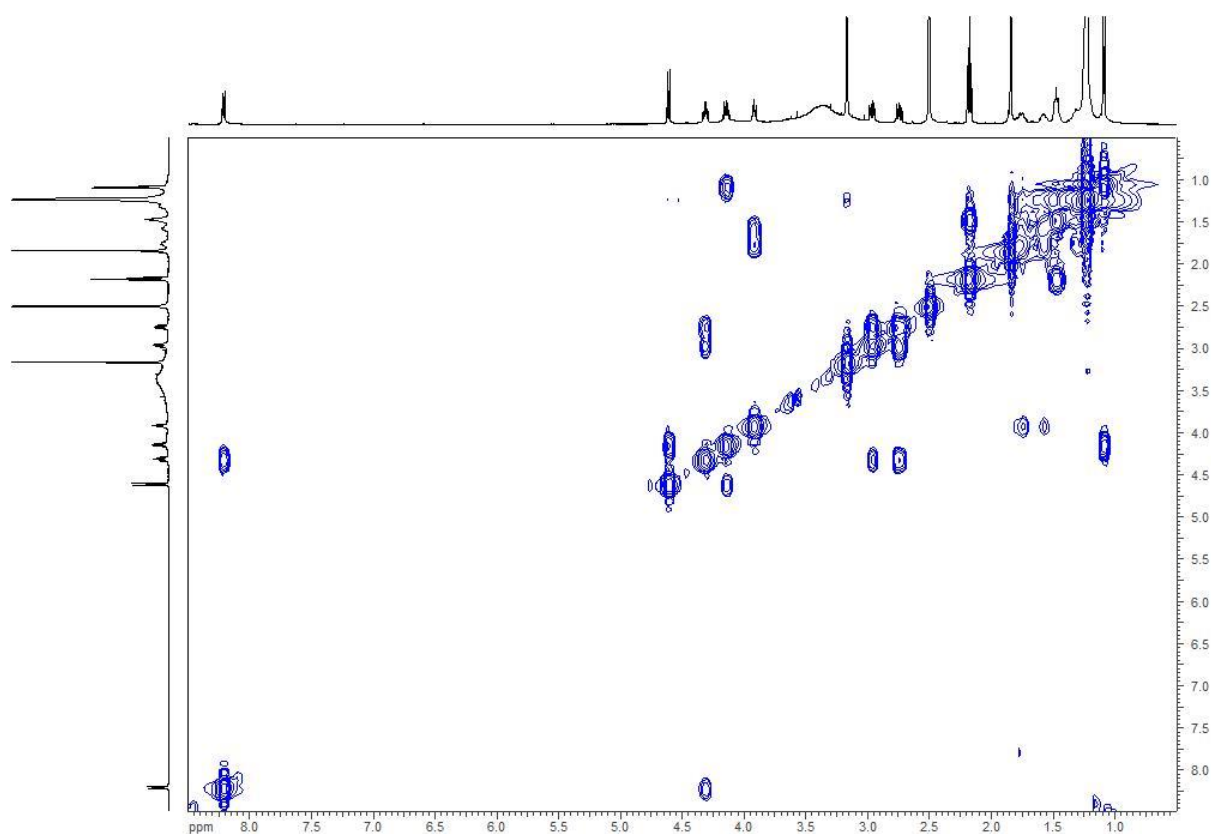




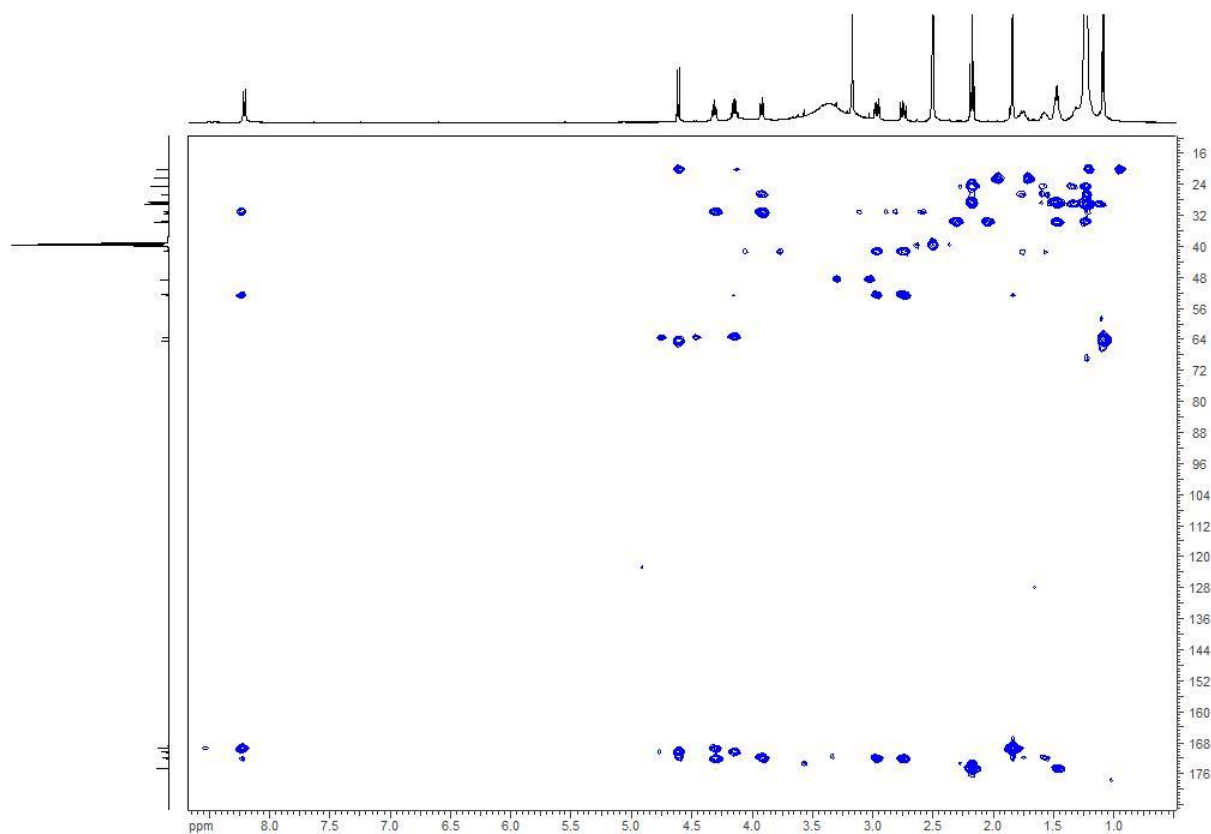
**Figure S23:**  $^1\text{H}$  NMR spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 2-NAC-lipothrenin A (**5**).



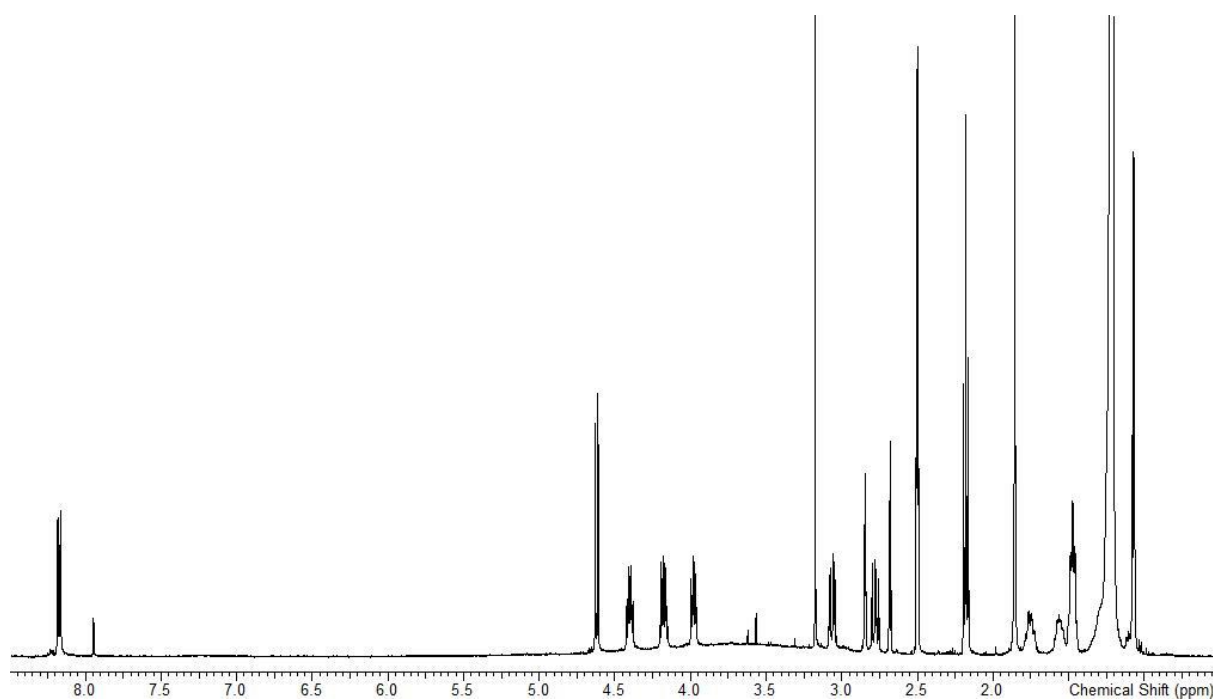
**Figure S24:**  $^{13}\text{C}$  NMR spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 2-NAC-lipothrenin A (**5**).



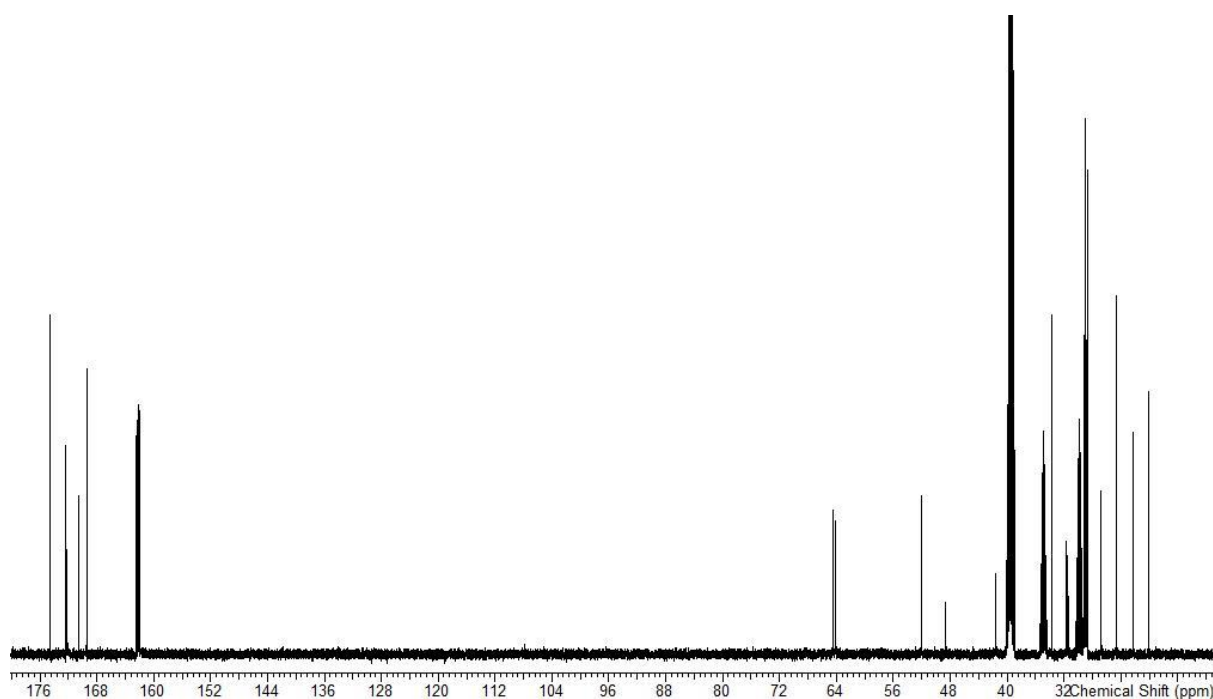
**Figure S25:**  $^1\text{H}$ - $^1\text{H}$ -COSY spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 2-NAC-lipothrenin A (5).



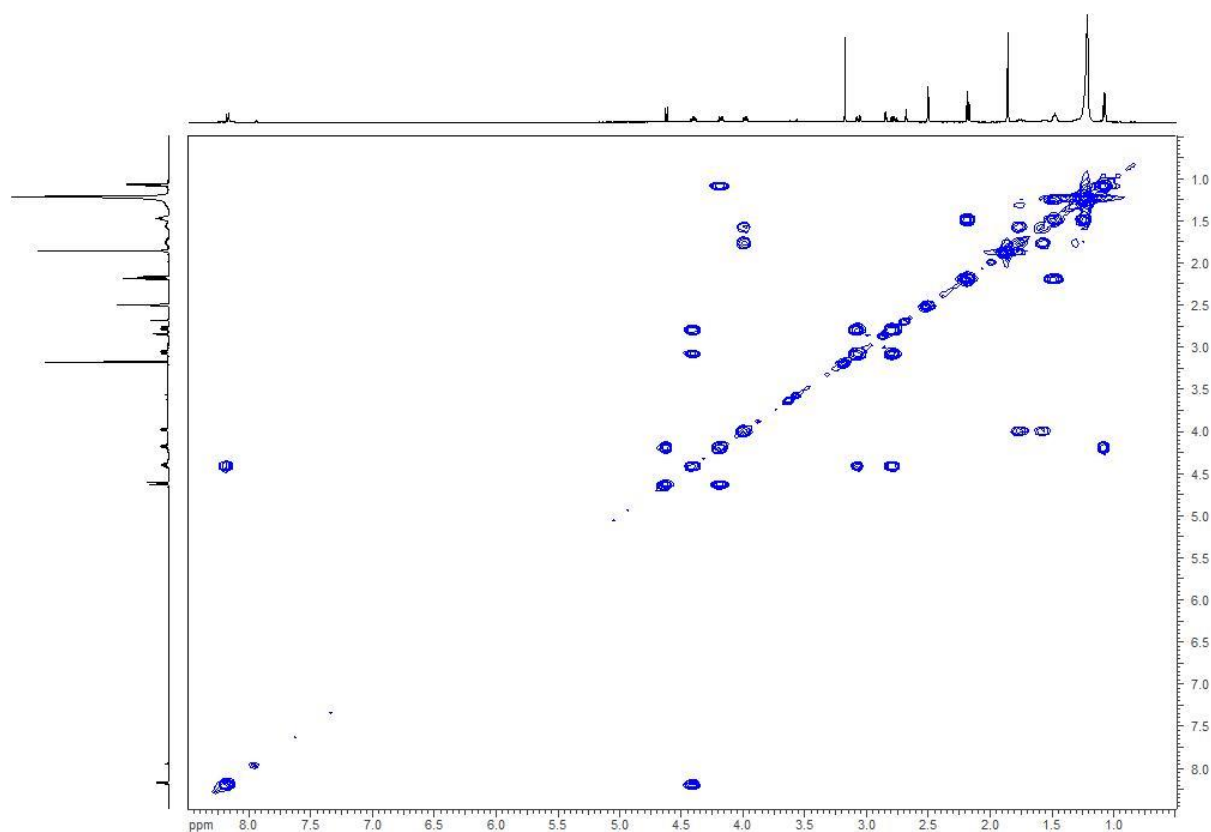
**Figure S26:** HMBC spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 2-NAC-lipothrenin A (5).



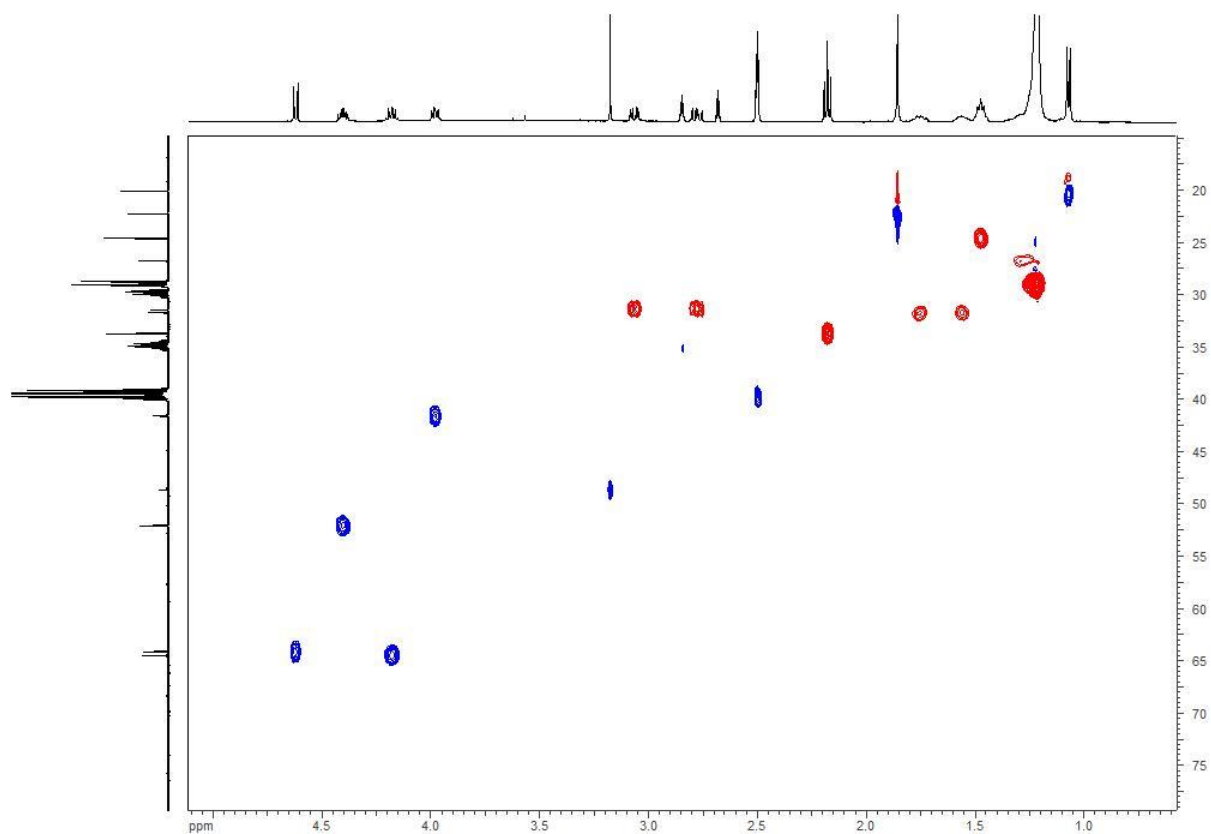
**Figure S27:**  $^1\text{H}$  NMR spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 2-NAC-lipothrenin  $\text{A}_1$  (**6**).



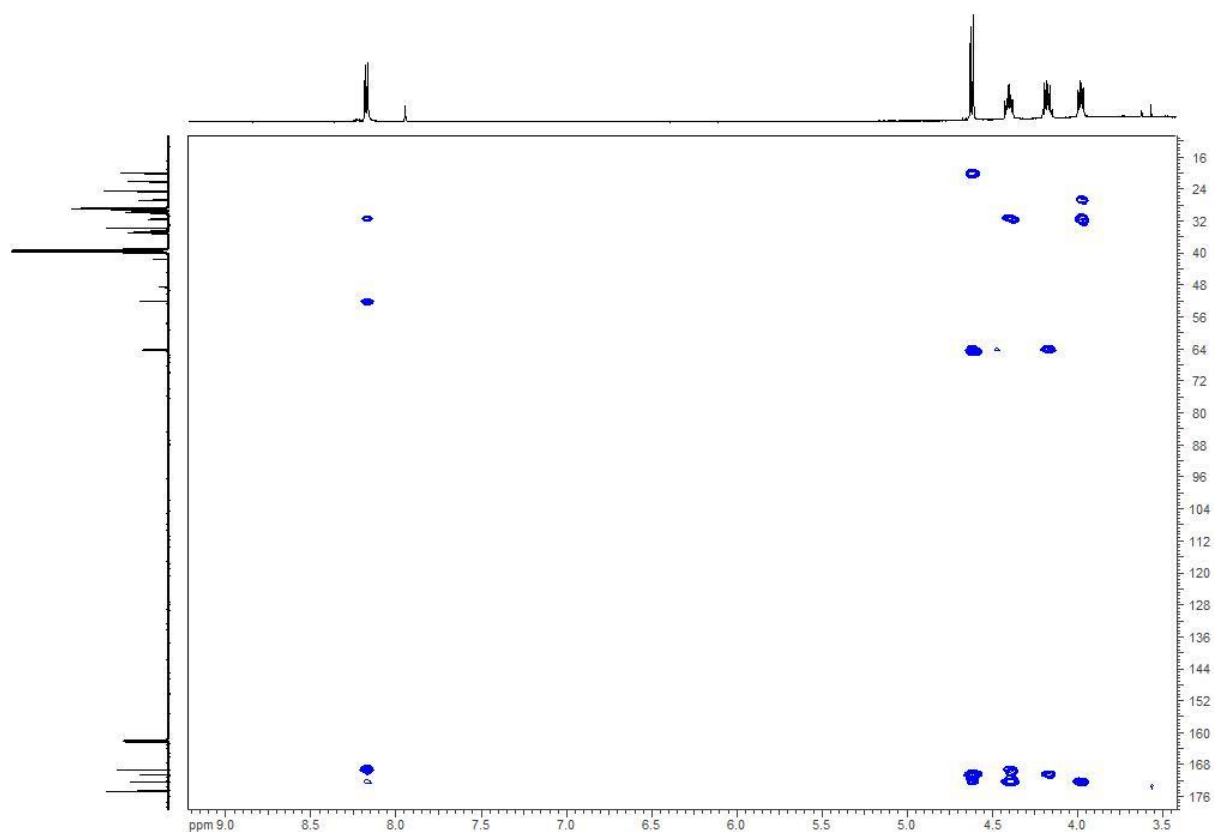
**Figure S28:**  $^{13}\text{C}$  NMR spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 2-NAC-lipothrenin  $\text{A}_1$  (**6**).



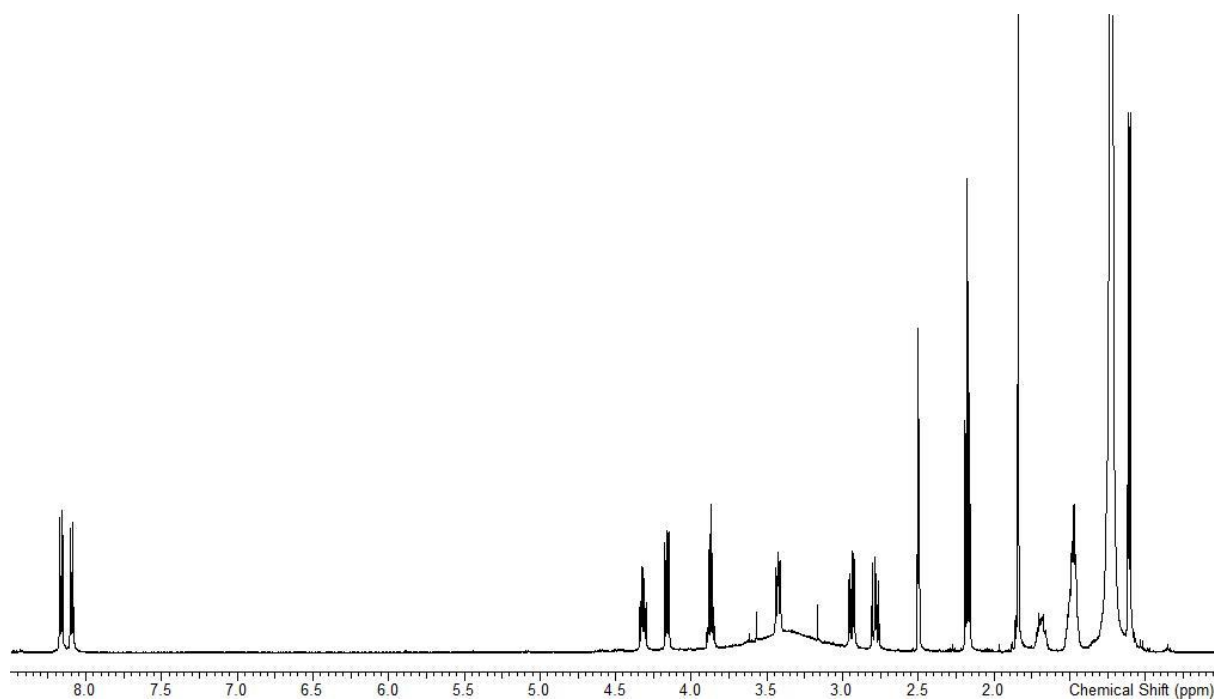
**Figure S29:**  $^1\text{H}$ - $^1\text{H}$ -COSY spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 2-NAC-lipothrenin  $\text{A}_1$  (**6**).



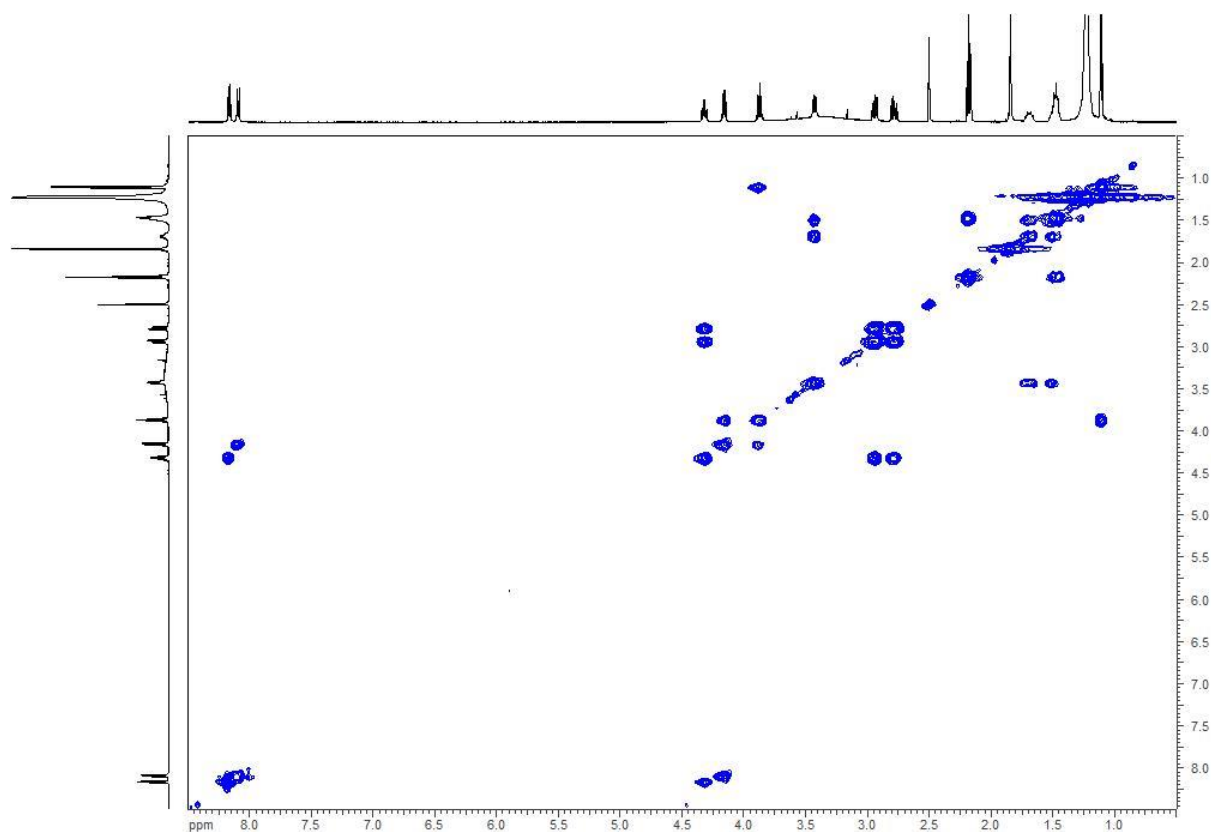
**Figure S30:** HSQC spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 2-NAC-lipothrenin  $\text{A}_1$  (**6**).



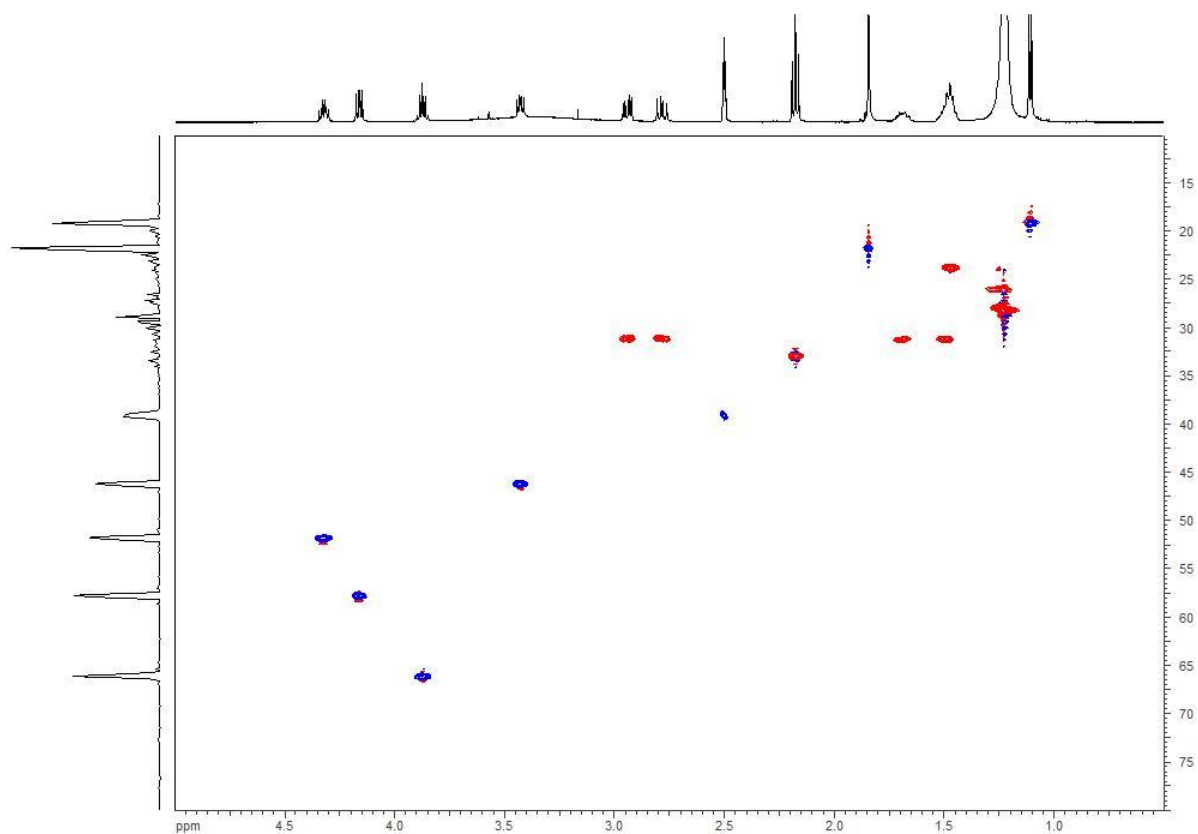
**Figure S31:** HMBC spectrum (500 MHz, DMSO- $d_6$ ) of 2-NAC-lipothrenin A<sub>1</sub> (**6**).



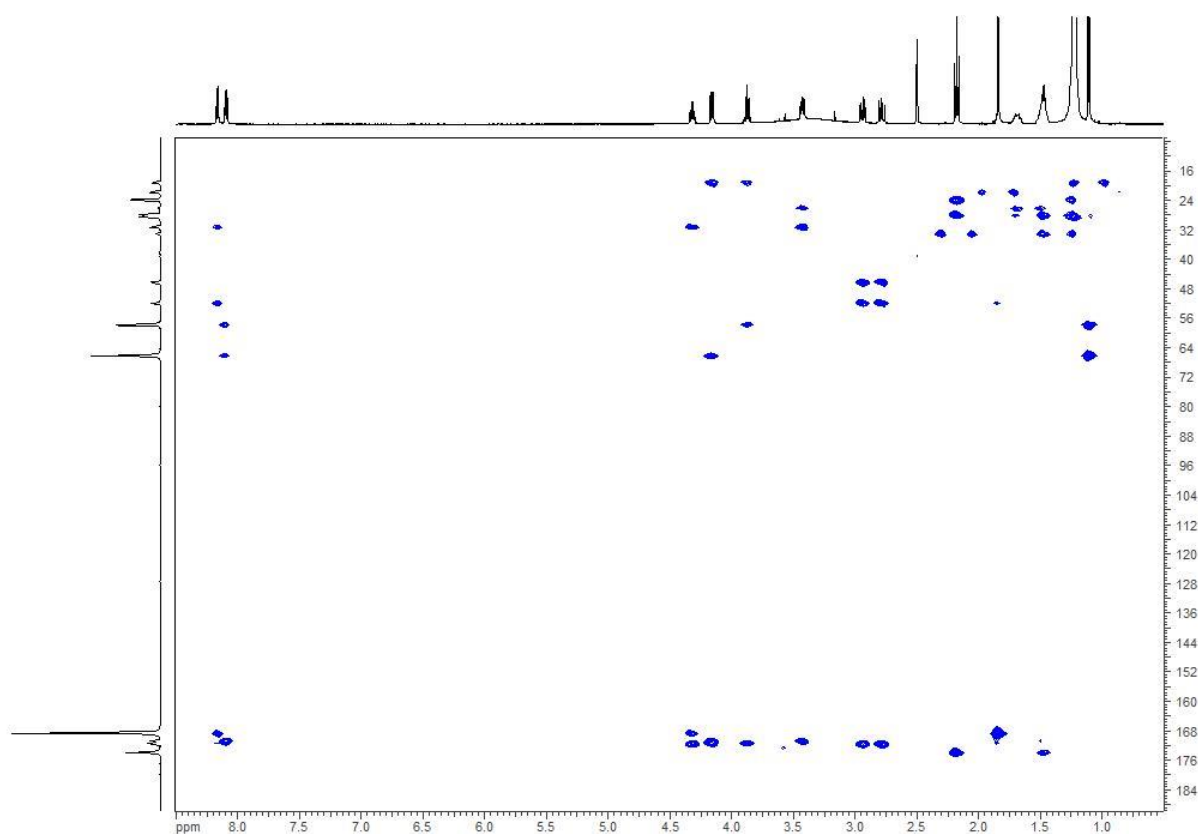
**Figure S32:**  $^1\text{H}$  NMR spectrum (500 MHz, DMSO- $d_6$ ) of 2-NAC-D-lipothrenin B (**7**).



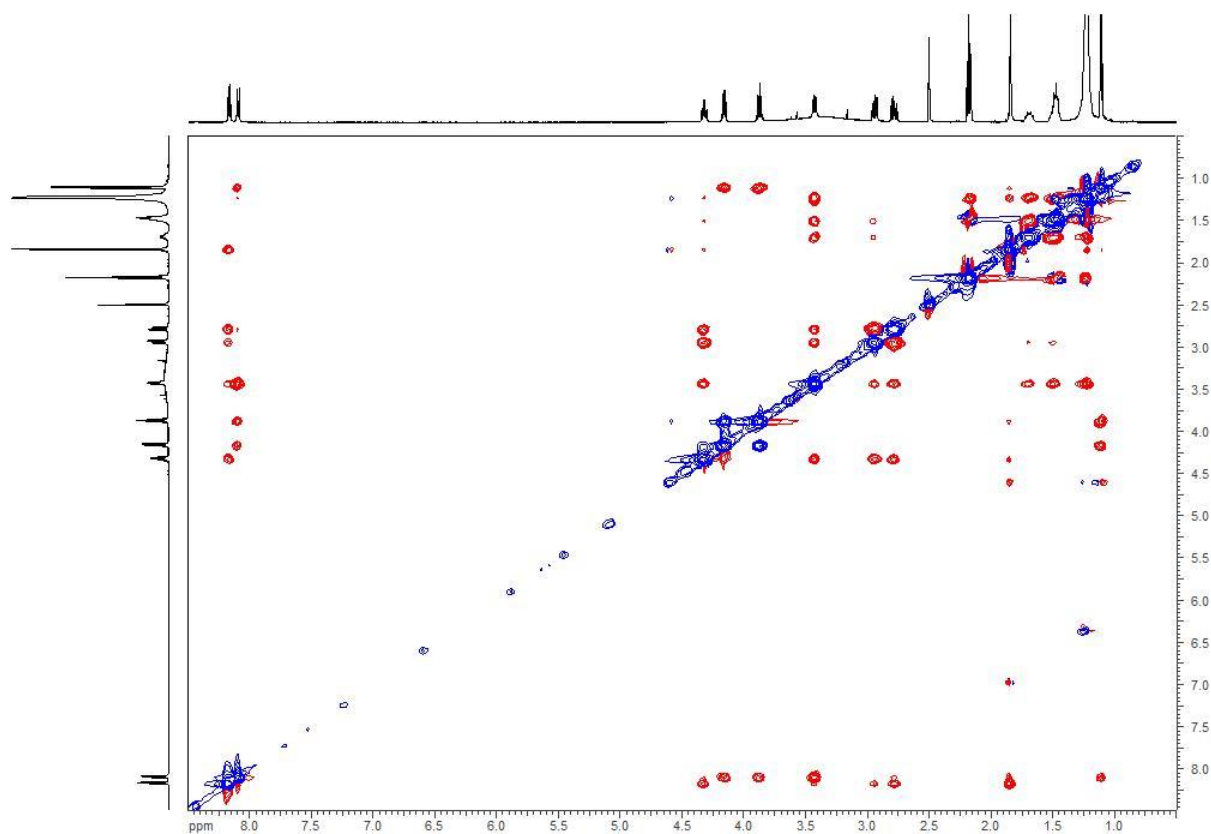
**Figure S33:**  $^1\text{H}$ - $^1\text{H}$ -COSY spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 2-NAC-D-lipothrenin B (**7**).



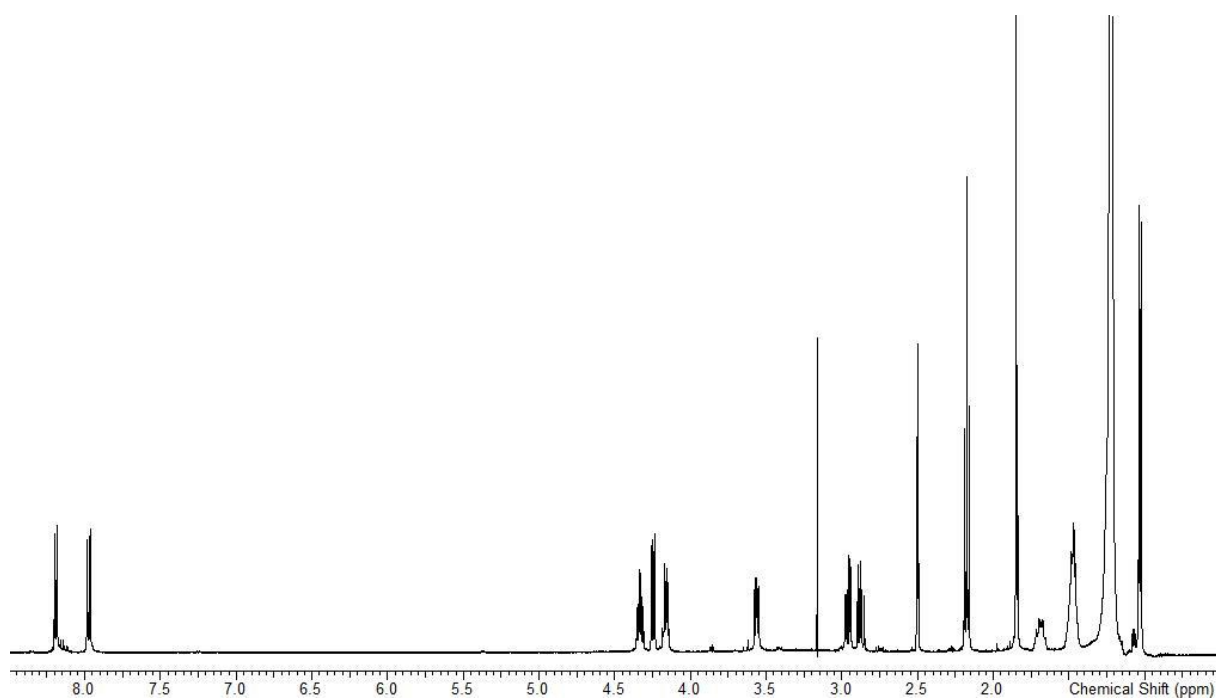
**Figure S34:** HSQC spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 2-NAC-D-lipothrenin B (**7**).



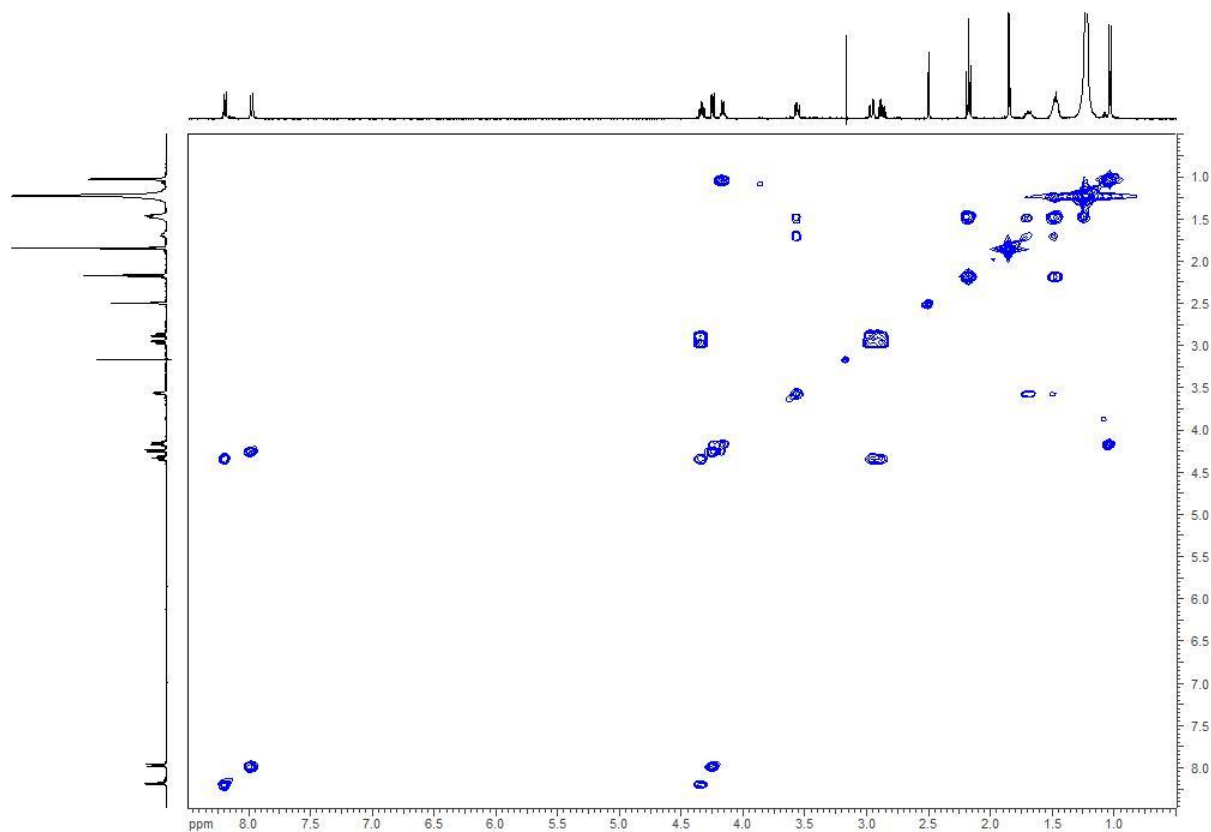
**Figure S35:** HMBC spectrum (500 MHz, DMSO- $d_6$ ) of 2-NAC-D-lipothrenin B (**7**).



**Figure S36:** ROESY spectrum (500 MHz, DMSO- $d_6$ ) of 2-NAC-D-lipothrenin B (**7**).

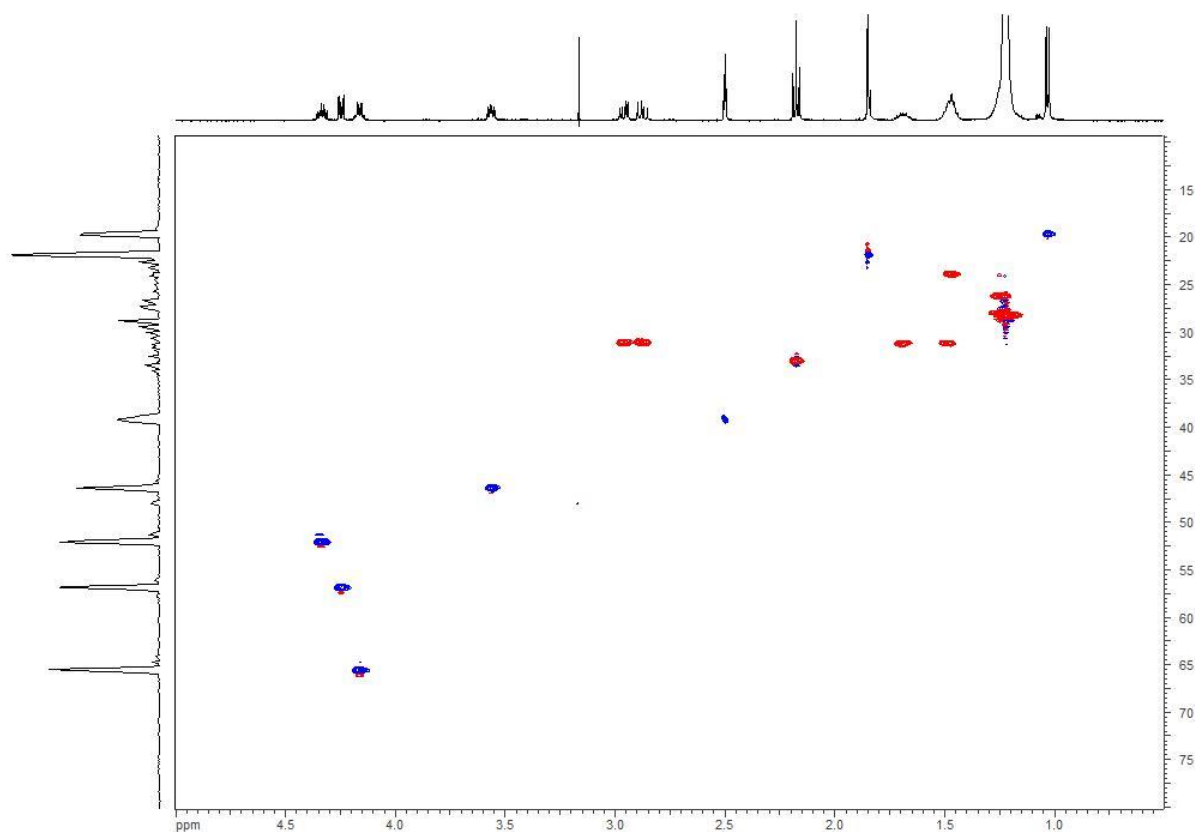


**Figure S37:**  $^1\text{H}$  NMR spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 2-NAC-L-lipothrenin B (**8**).

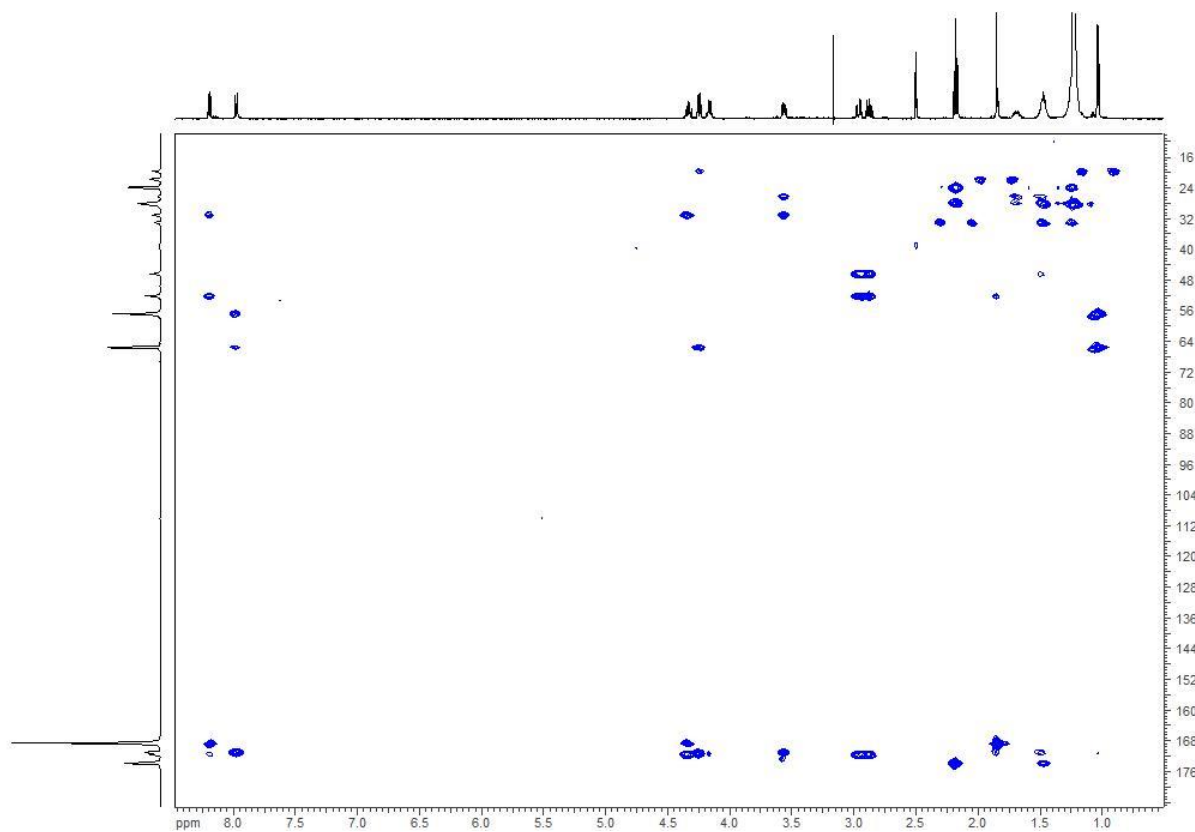


**Figure S38:**  $^1\text{H}$ - $^1\text{H}$ -COSY spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 2-NAC-L-lipothrenin B (**8**).

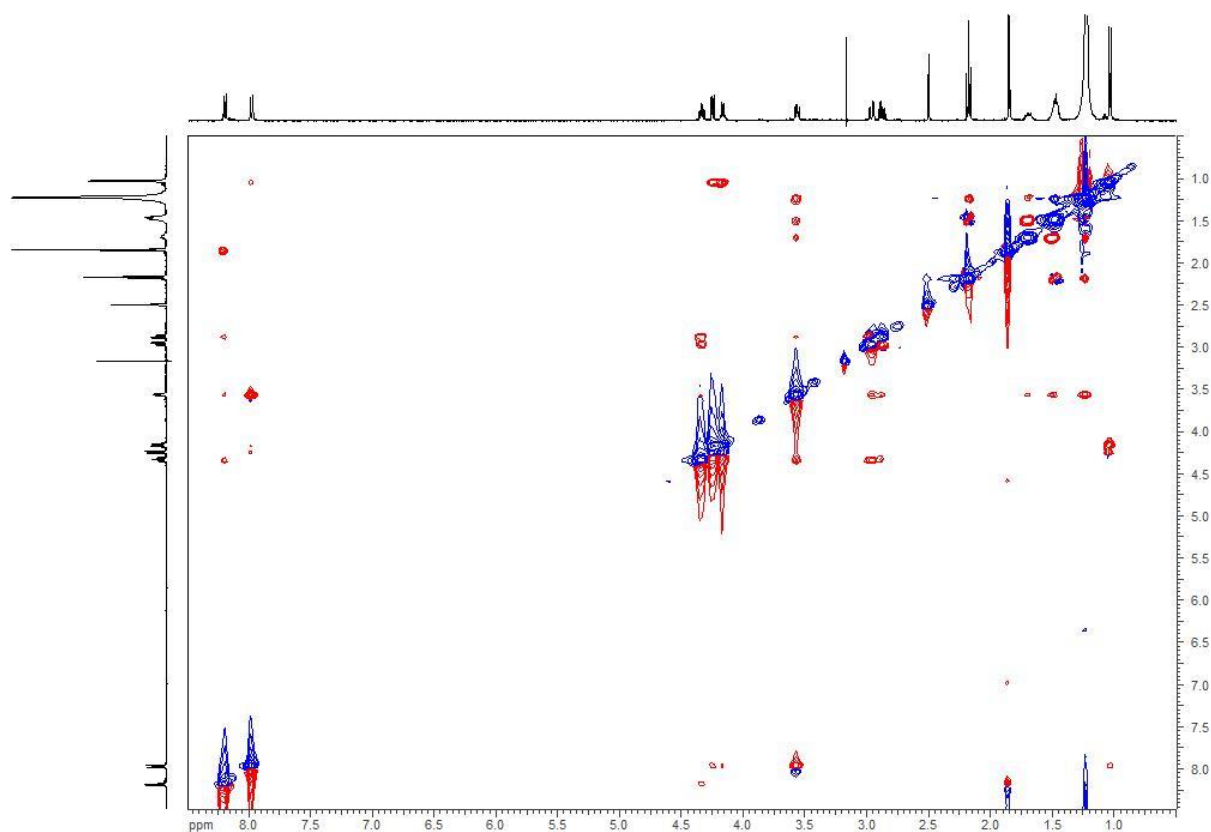




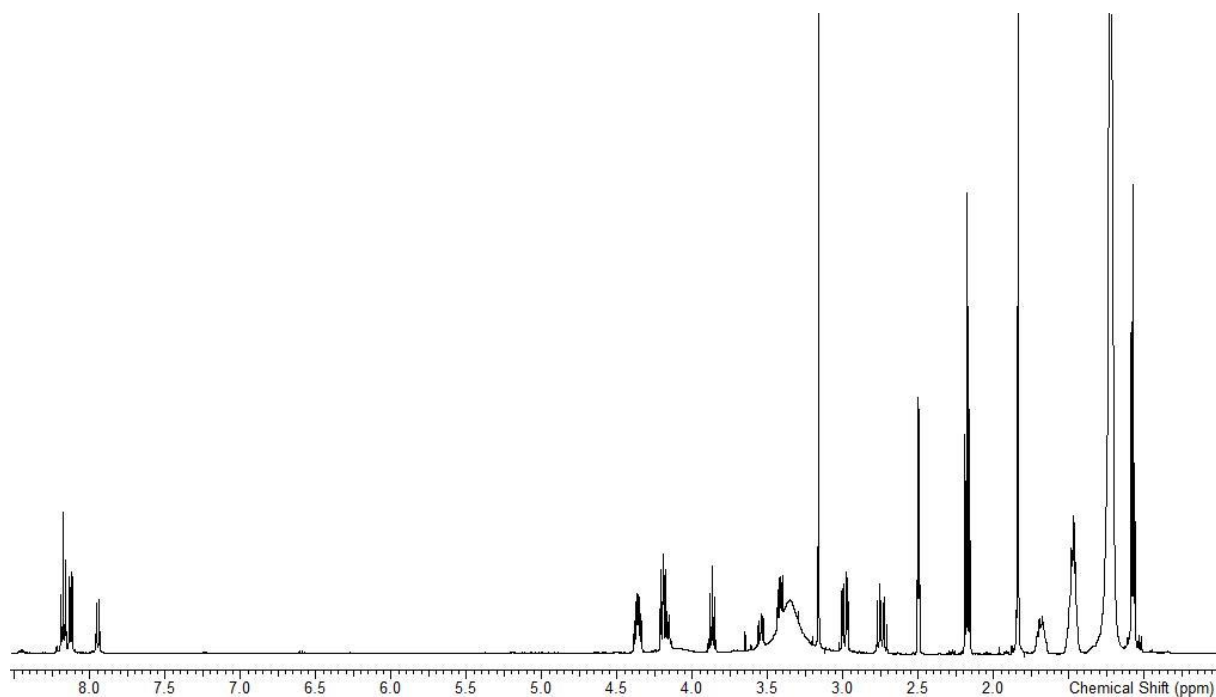
**Figure S39:** HSQC spectrum (500 MHz, DMSO- $d_6$ ) of 2-NAC-L-lipothrenin B (**8**).



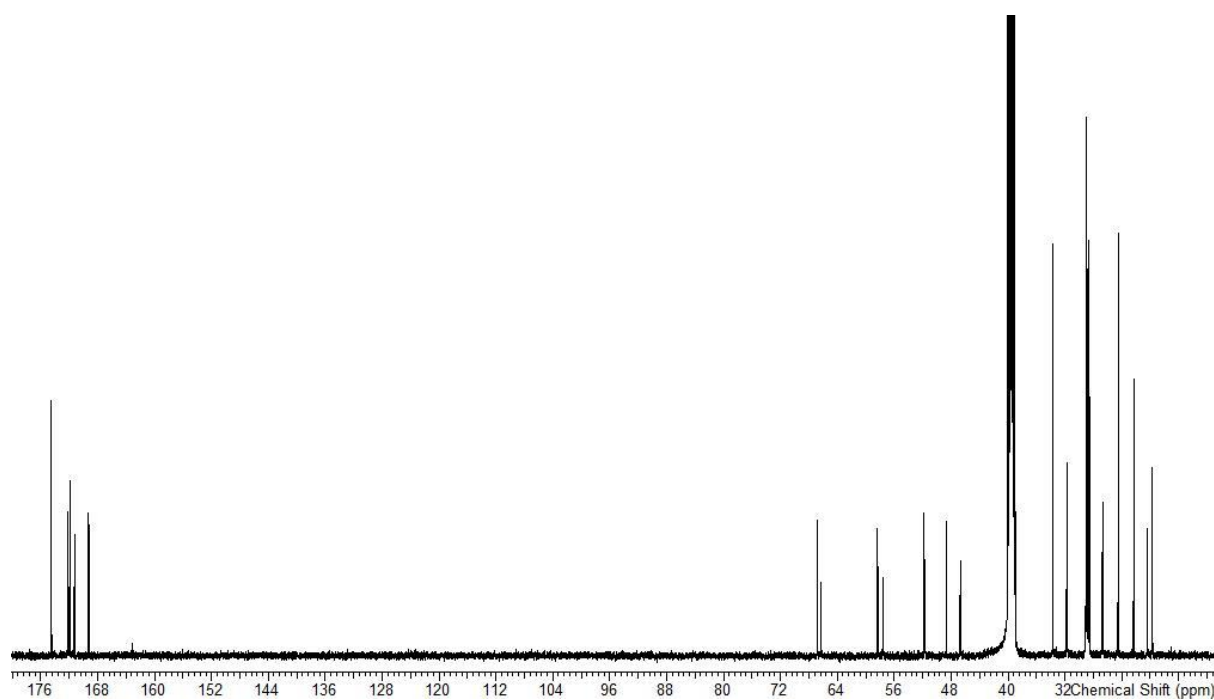
**Figure S40:** HMBC spectrum (500 MHz, DMSO- $d_6$ ) of 2-NAC-L-lipothrenin B (**8**).



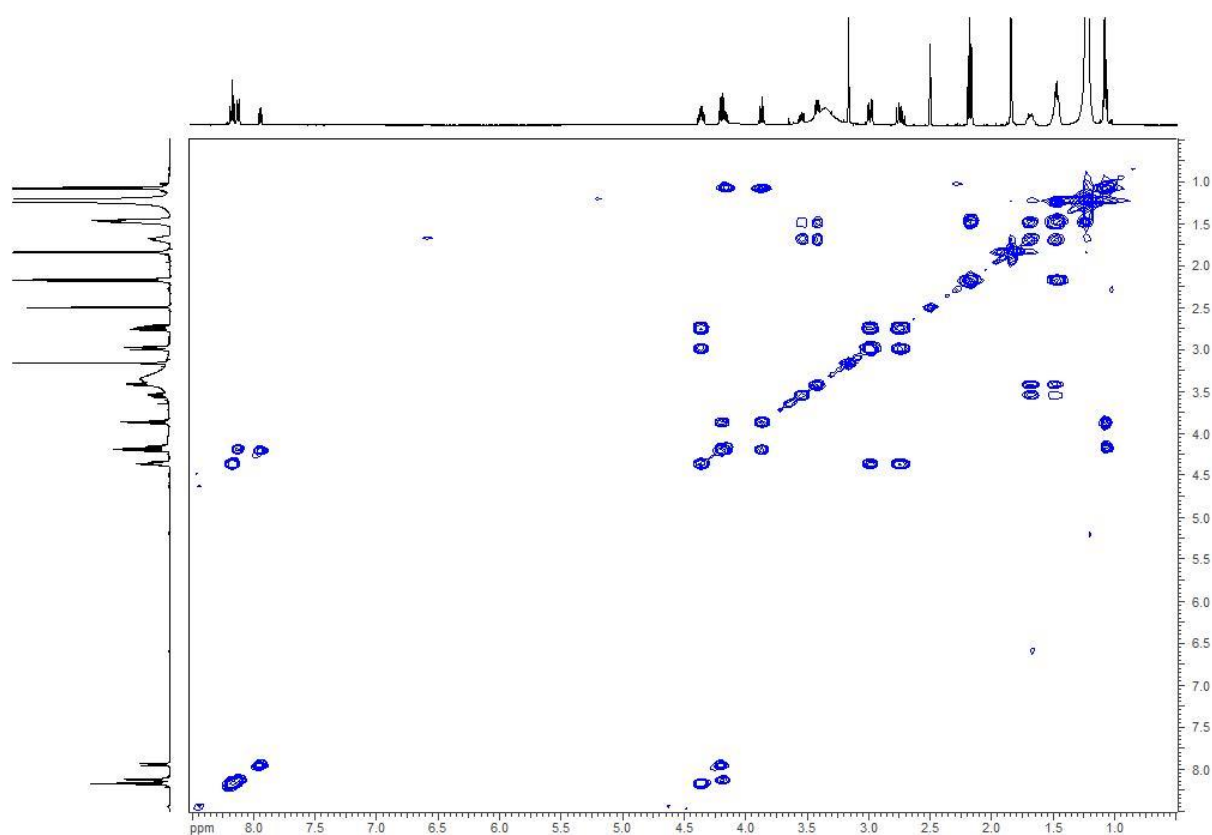
**Figure S41:** ROESY spectrum (500 MHz, DMSO- $d_6$ ) of 2-NAC-L-lipothrenin B (**8**).



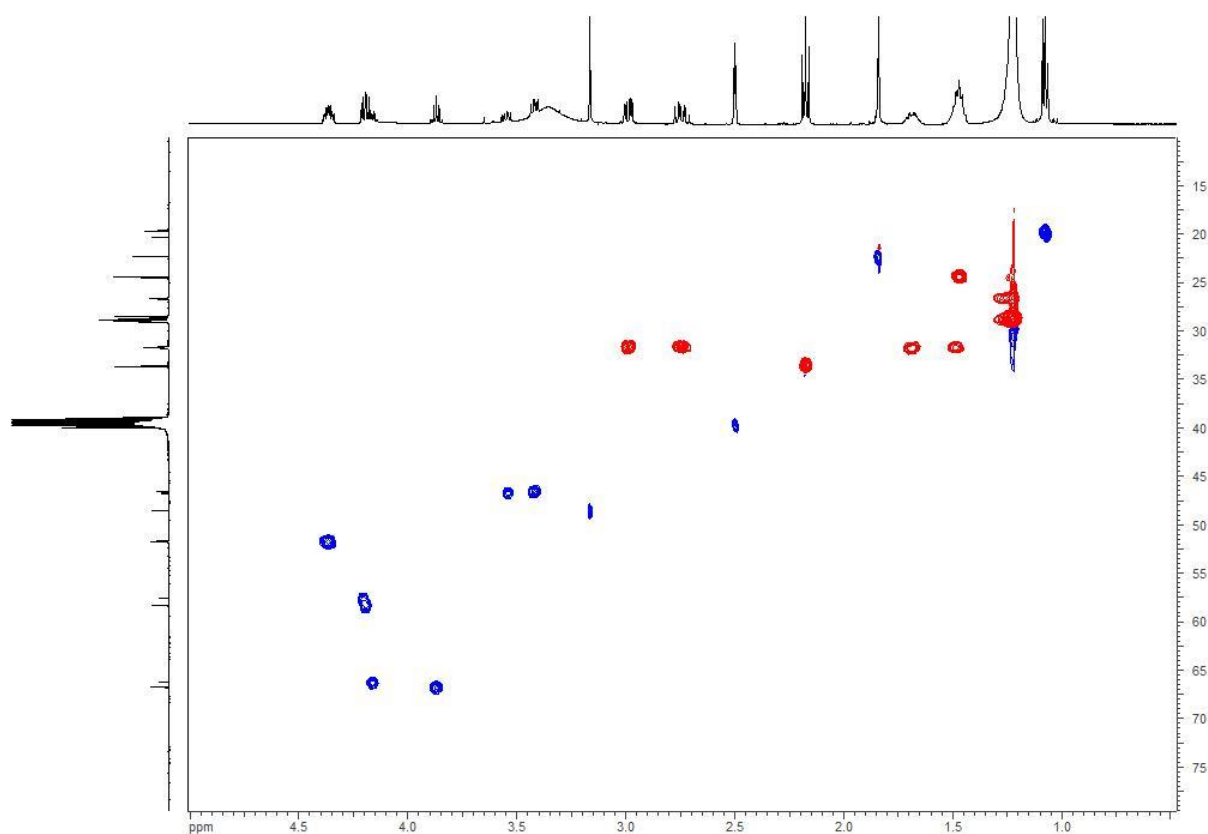
**Figure S42:**  $^1\text{H}$  NMR spectrum (500 MHz, DMSO- $d_6$ ) of 2-NAC-D-/L-lipothrenin B<sub>1</sub> mixture (**9**, **10**).



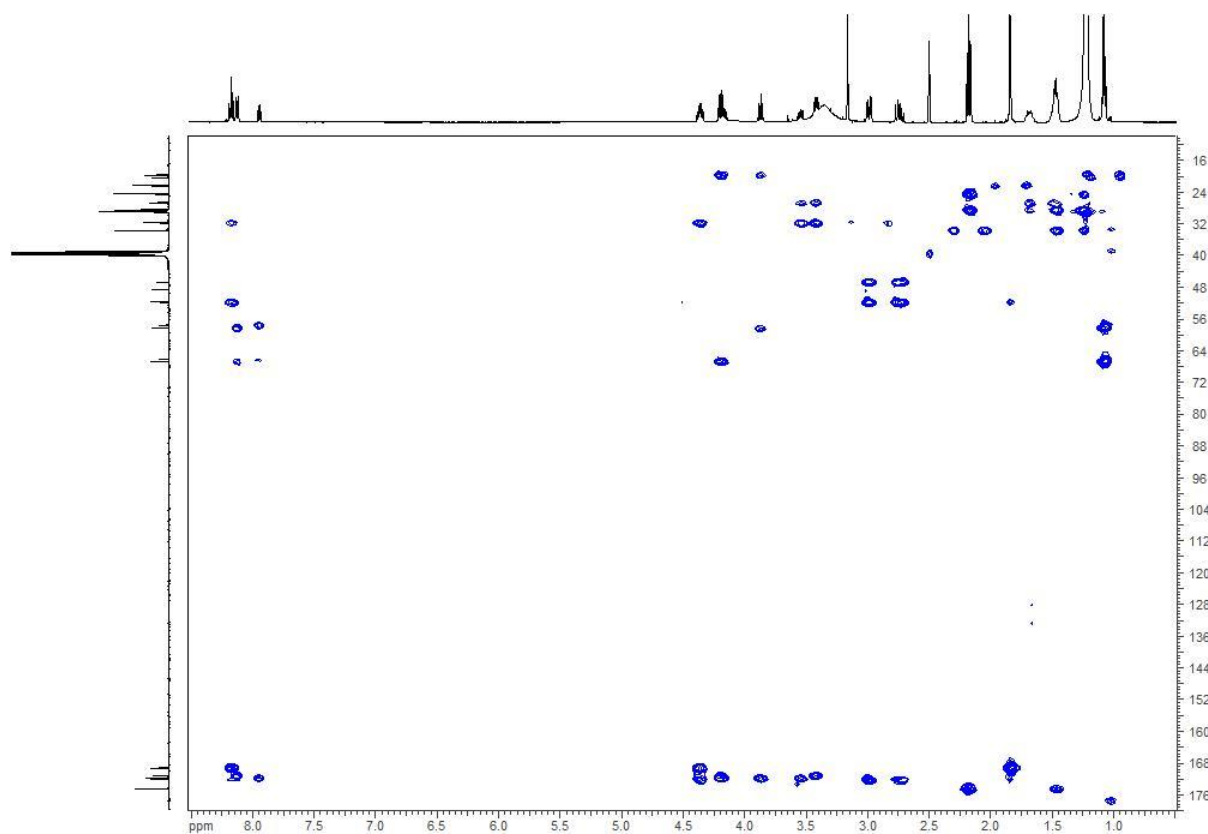
**Figure S43:**  $^{13}\text{C}$  NMR spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 2-NAC-D-/L-lipothrenin  $\text{B}_1$  mixture (**9**, **10**).



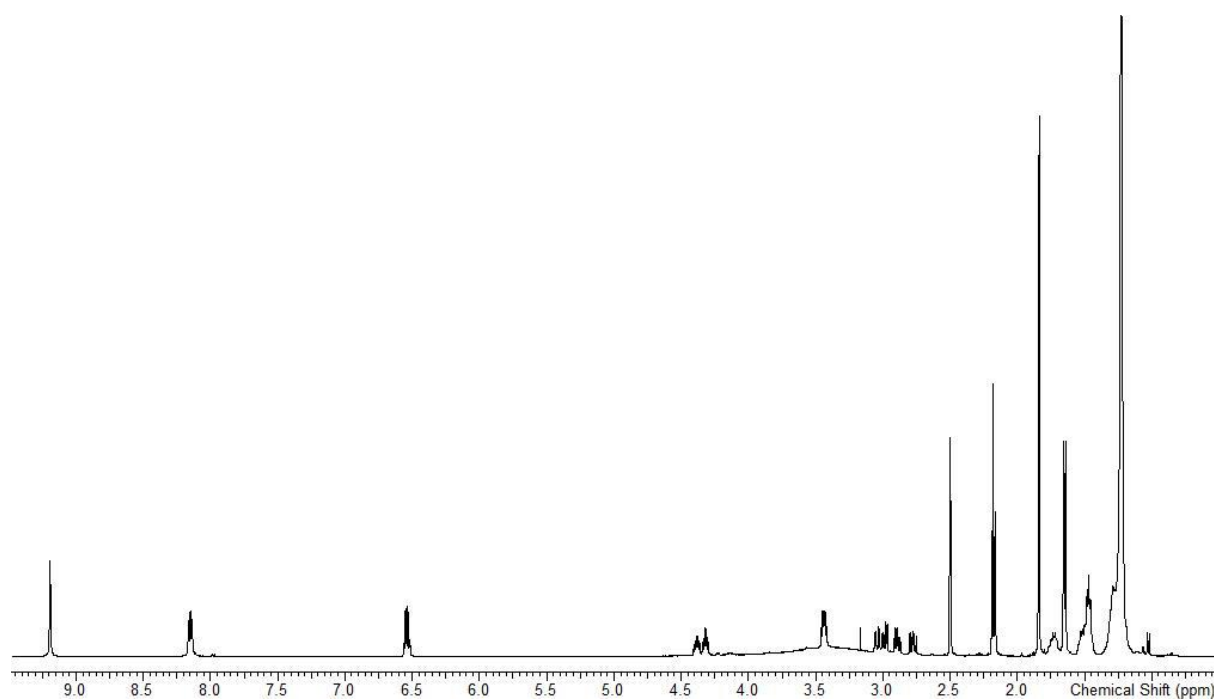
**Figure S44:**  $^1\text{H}$ - $^1\text{H}$  COSY spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 2-NAC-D-/L-lipothrenin  $\text{B}_1$  mixture (**9**, **10**).



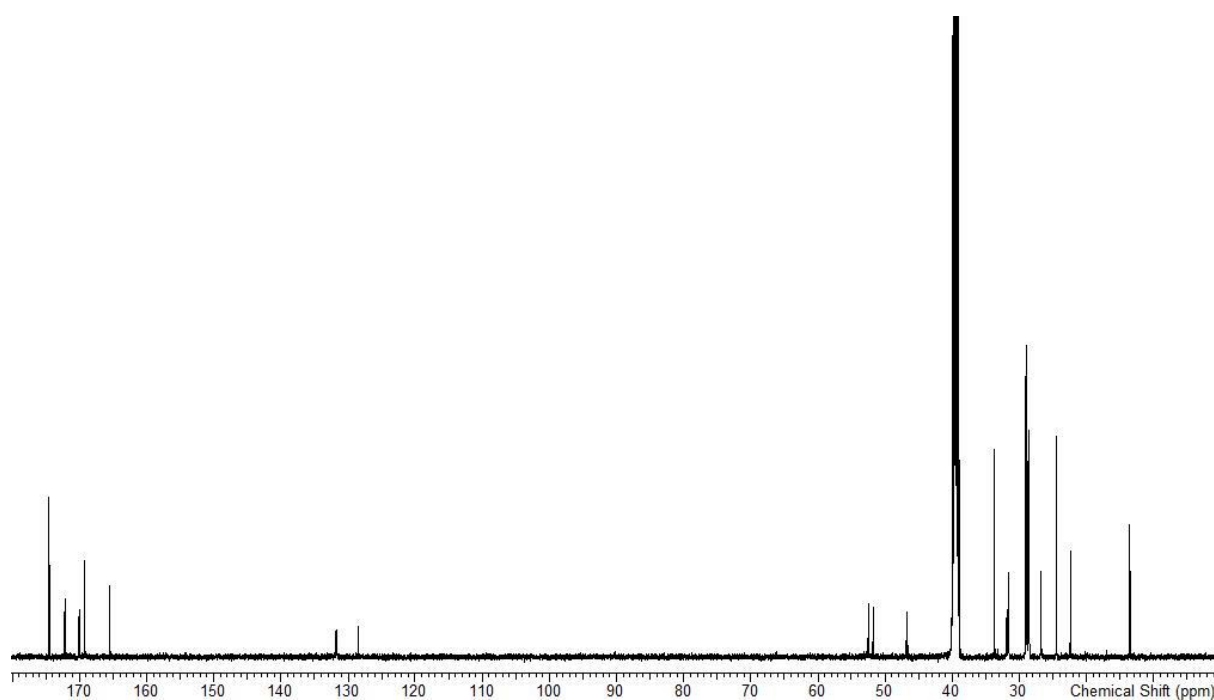
**Figure S45:** HSQC spectrum (500 MHz, DMSO- $d_6$ ) of 2-NAC-D-/L-lipothrenin B<sub>1</sub> mixture (**9**, **10**).



**Figure S46:** HMBC spectrum (500 MHz, DMSO- $d_6$ ) of 2-NAC-D-/L-lipothrenin B<sub>1</sub> mixture (**9**, **10**).



**Figure S47:**  $^1\text{H}$  NMR spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 2-NAC-Z-lipothrenin C and  $\text{C}_1$  mixture (**11**, **12**).



**Figure S48:**  $^{13}\text{C}$  NMR spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 2-NAC-Z-lipothrenin C and  $\text{C}_1$  mixture (**11**, **12**).

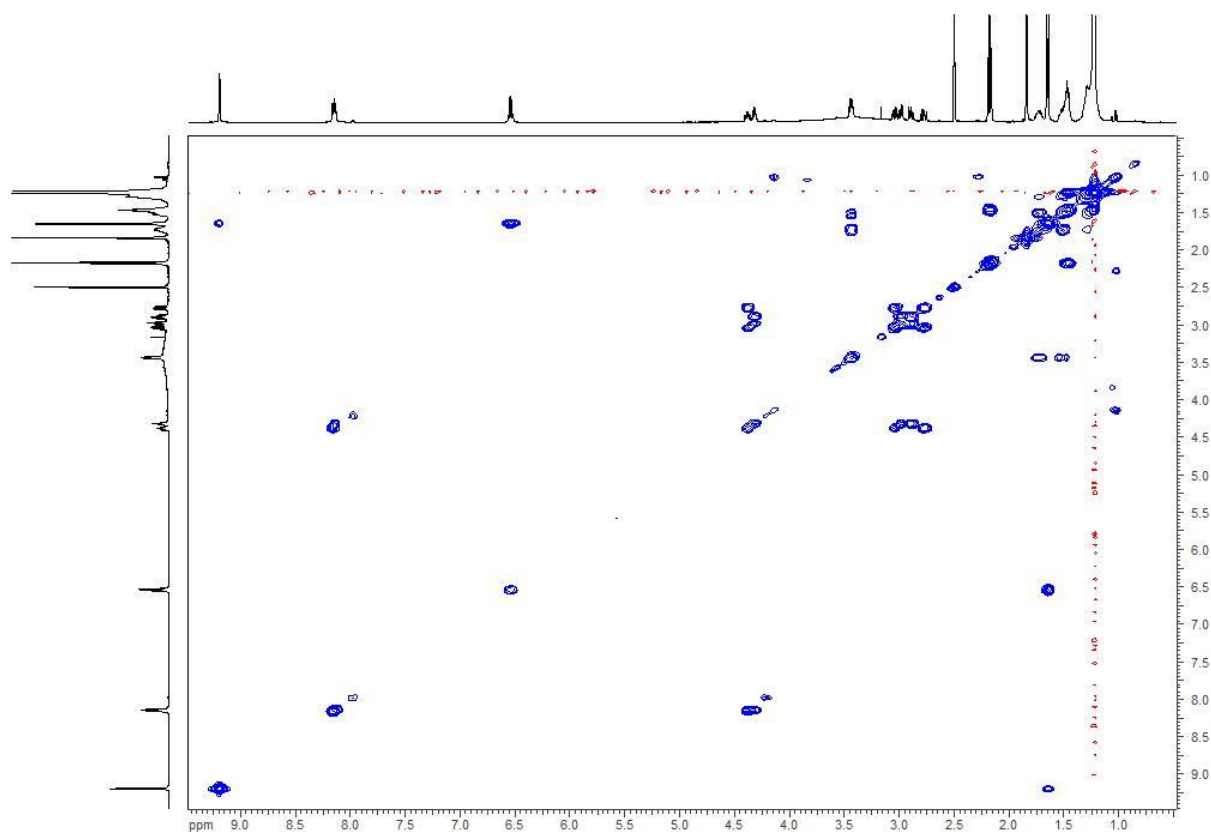


Figure S49:  $^1\text{H}$ - $^1\text{H}$ -COSY spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 2-NAC-Z-lipothrenin C and  $\text{C}_1$  mixture (**11**, **12**).

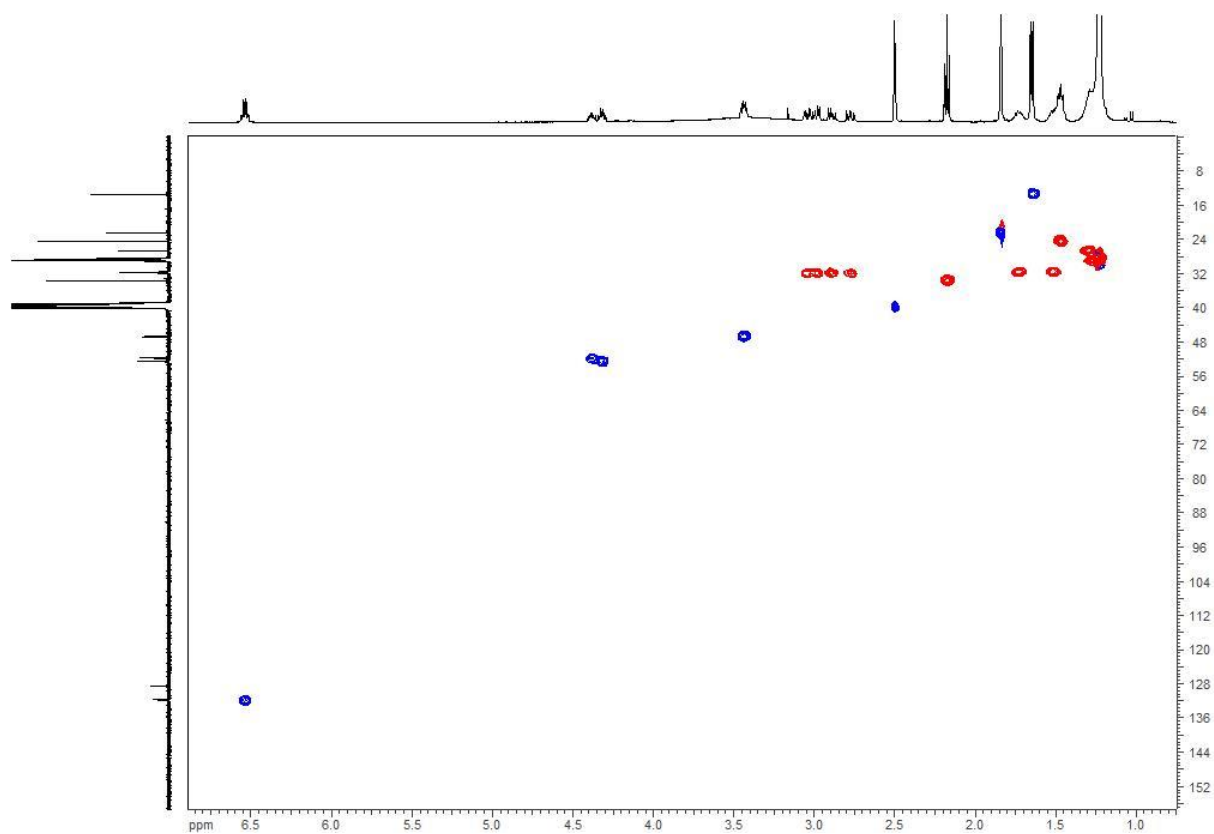
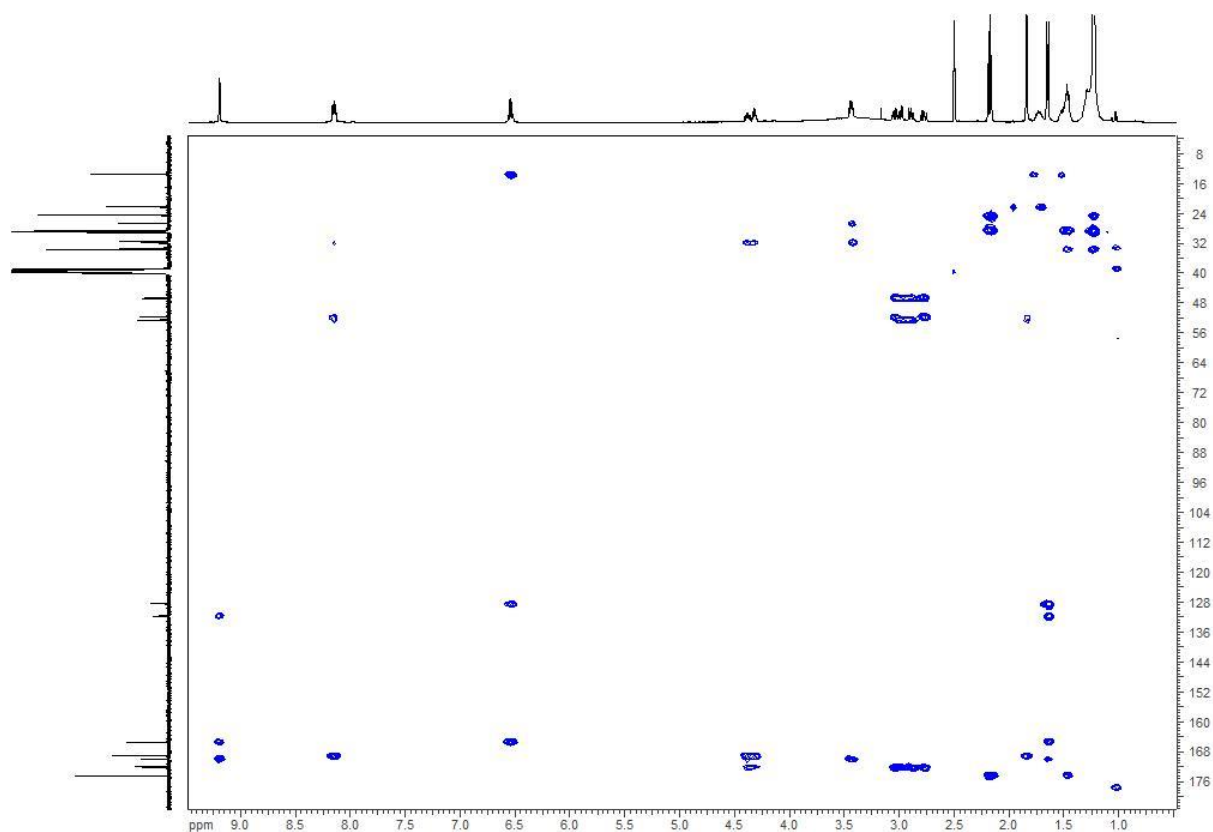
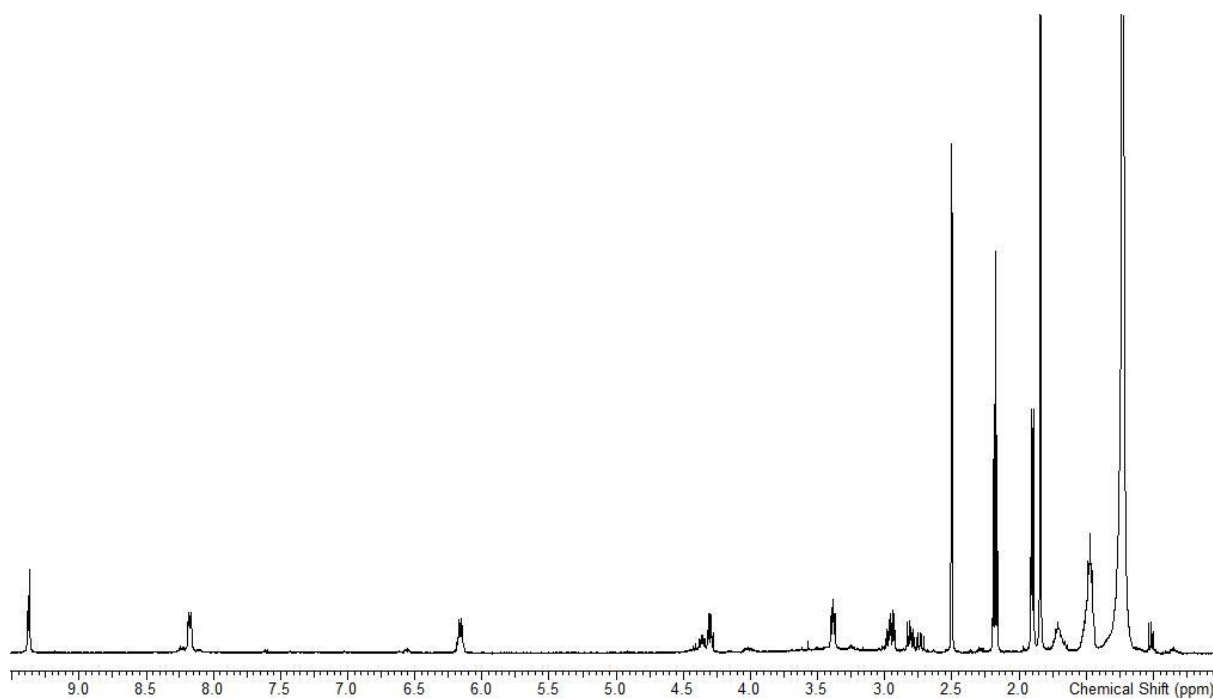


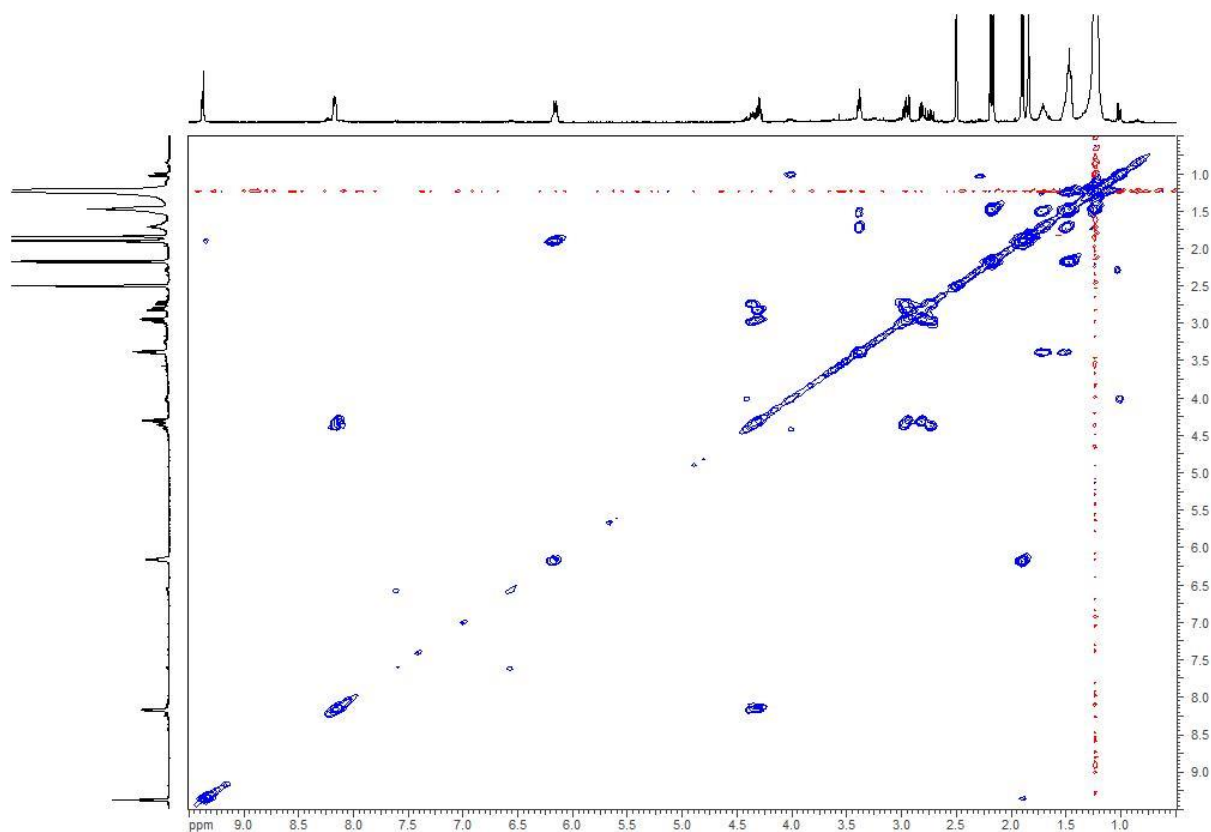
Figure S50: HSQC spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 2-NAC-Z-lipothrenin C and  $\text{C}_1$  mixture (**11**, **12**).



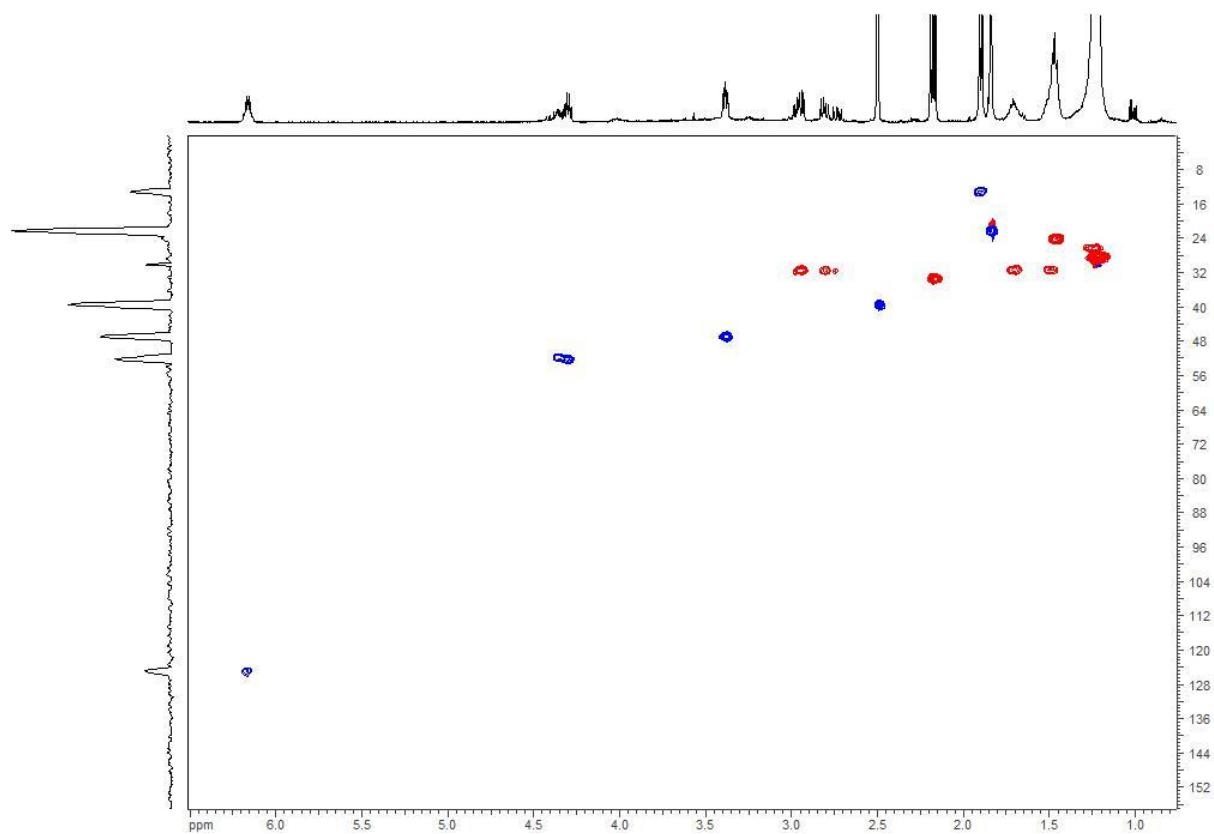
**Figure S51:** HMBC spectrum (500 MHz, DMSO- $\text{d}_6$ ) of 2-NAC-Z-lipothrenin C and  $\text{C}_1$  mixture (**11**, **12**).



**Figure S52:**  $^1\text{H}$  NMR spectrum (500 MHz, DMSO- $\text{d}_6$ ) of 2-NAC-E-lipothrenin C and  $\text{C}_1$  mixture (**13**, **14**).

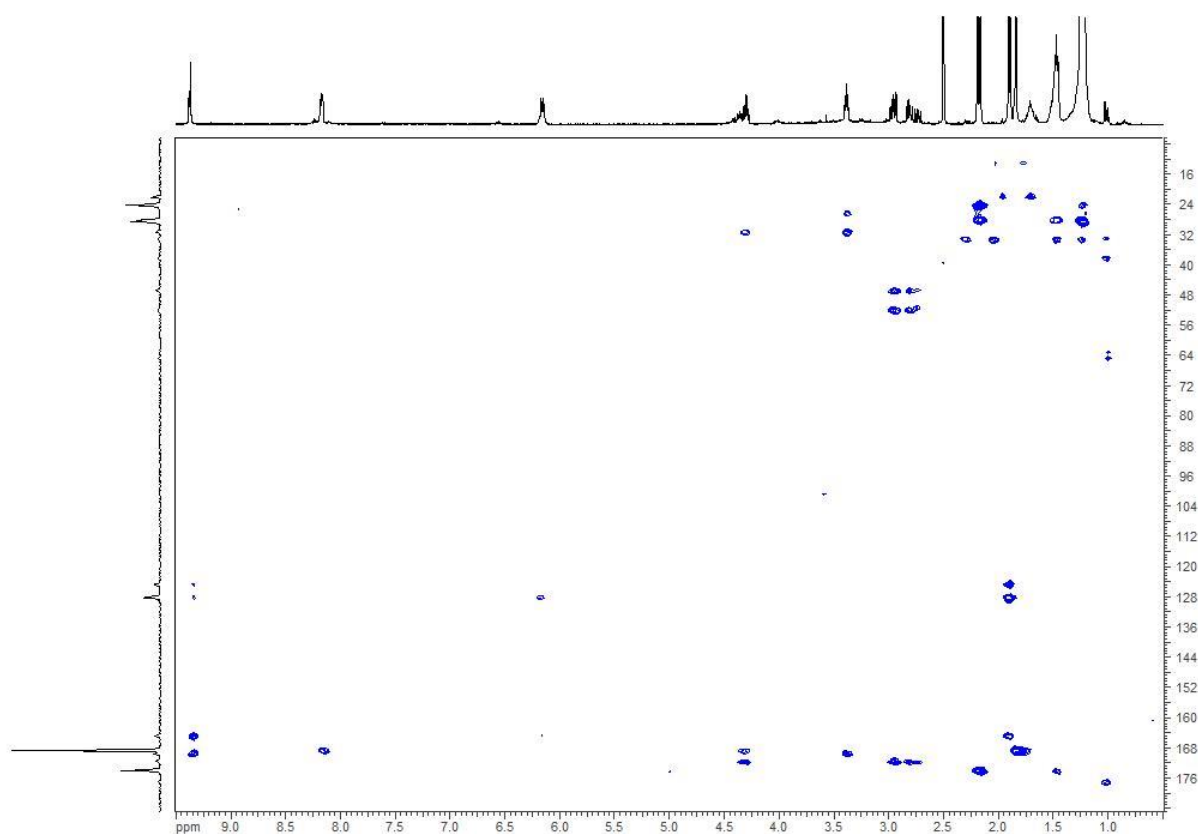


**Figure S53:**  $^1\text{H}$ - $^1\text{H}$ -COSY spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 2-NAC-*E*-lipothrenin C and  $\text{C}_1$  mixture (**13**, **14**).

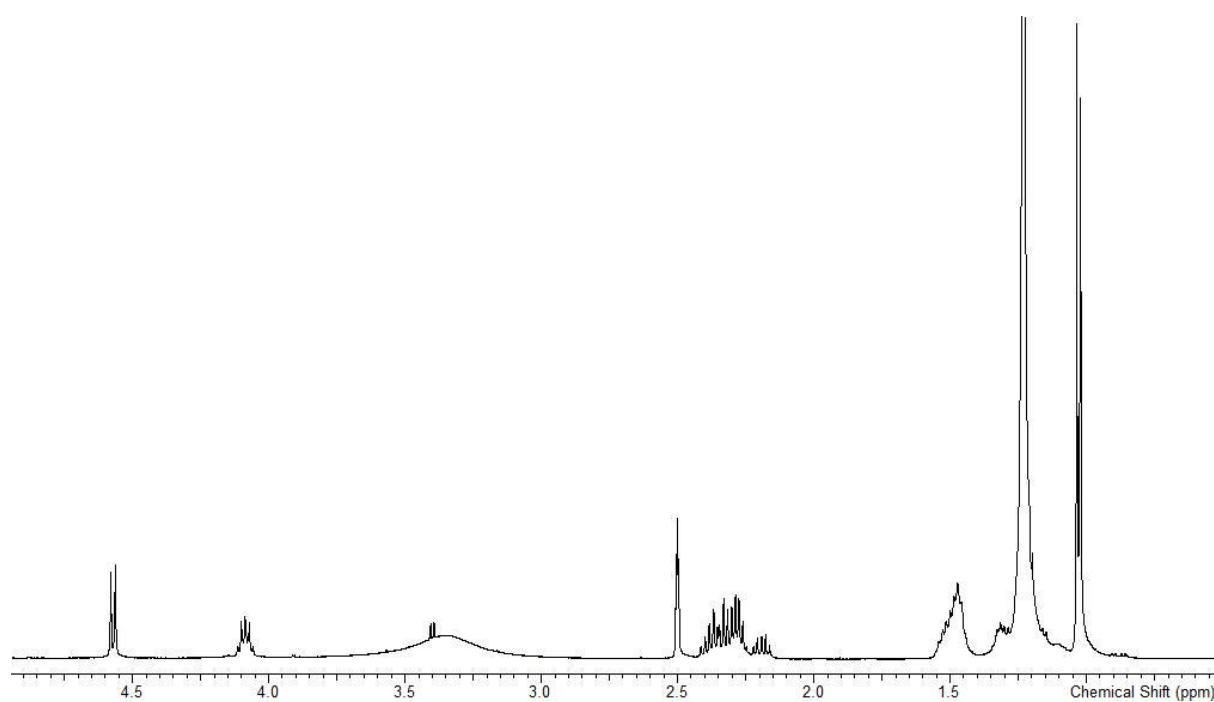


**Figure S54:** HSQC spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 2-NAC-*E*-lipothrenin C and  $\text{C}_1$  mixture (**13**, **14**).





**Figure S55:** HMBC spectrum (500 MHz, DMSO- $\text{d}_6$ ) of 2-NAC-*E*-lipothrenin C and  $\text{C}_1$  mixture (**13**, **14**).



**Figure S56:**  $^1\text{H}$  NMR spectrum (500 MHz, DMSO- $\text{d}_6$ ) of *iso*-lipothrenin A (**15**).

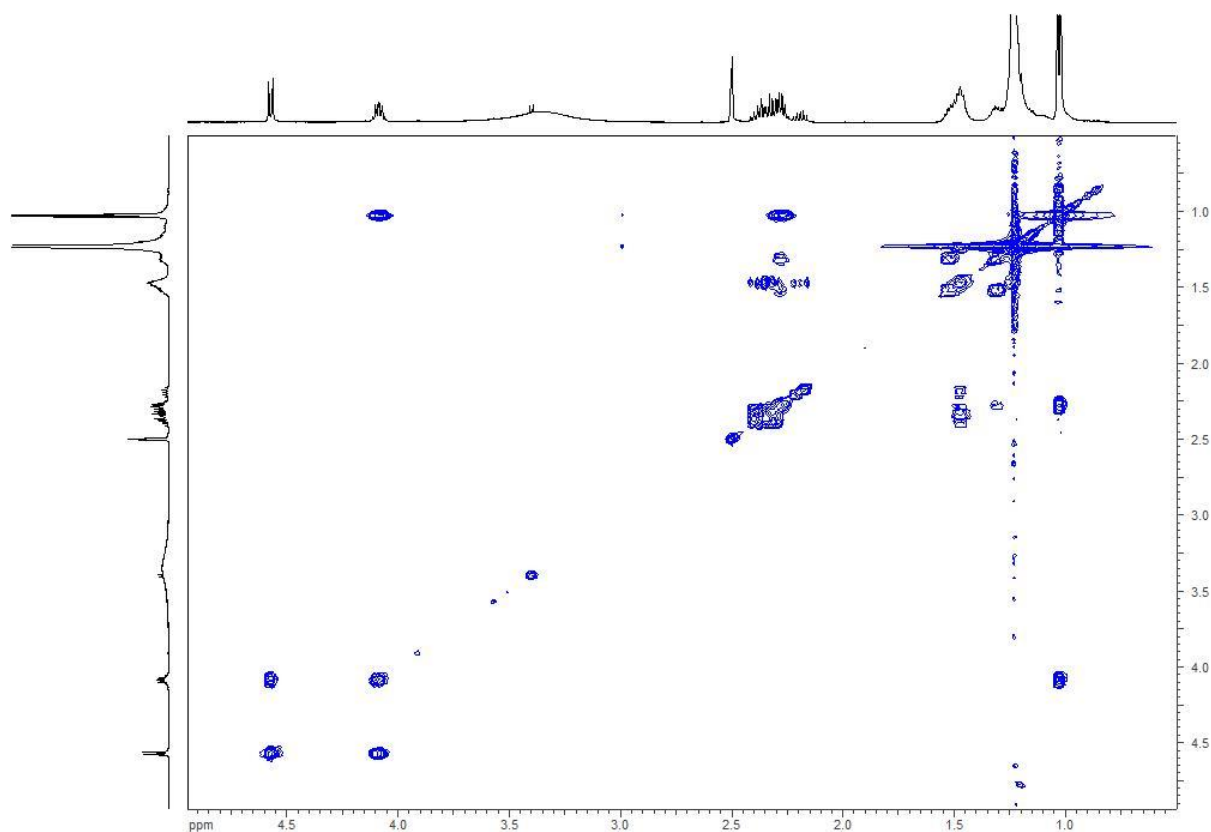


Figure S57:  $^1\text{H}$ - $^1\text{H}$ -COSY spectrum (500 MHz, DMSO- $d_6$ ) of *iso*-lipothrenin A (15).

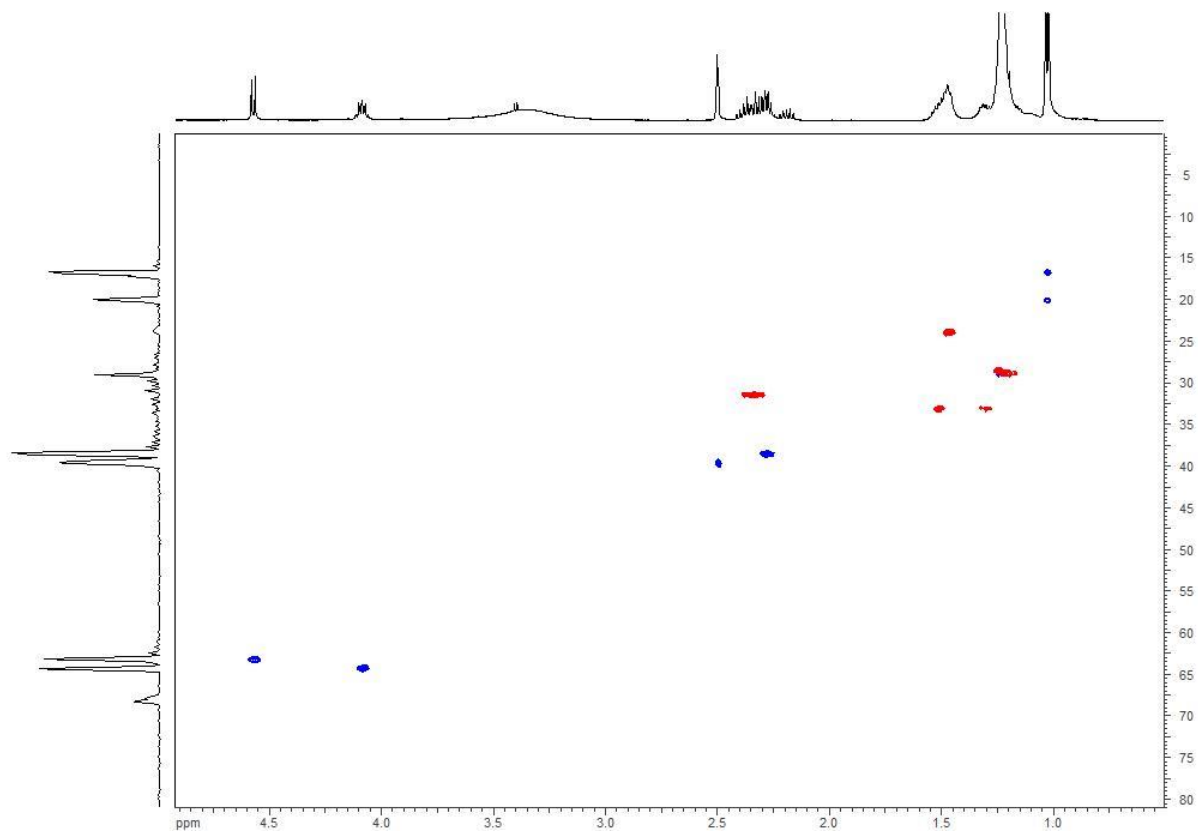
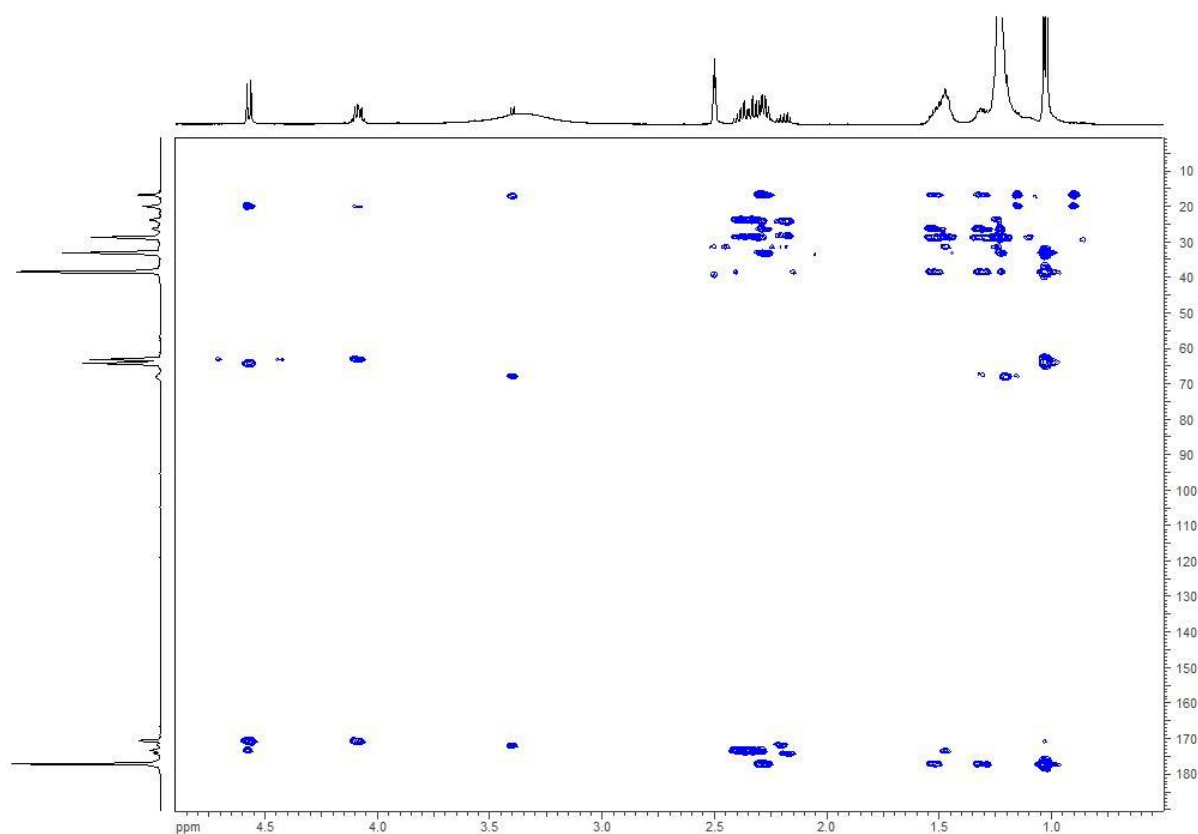
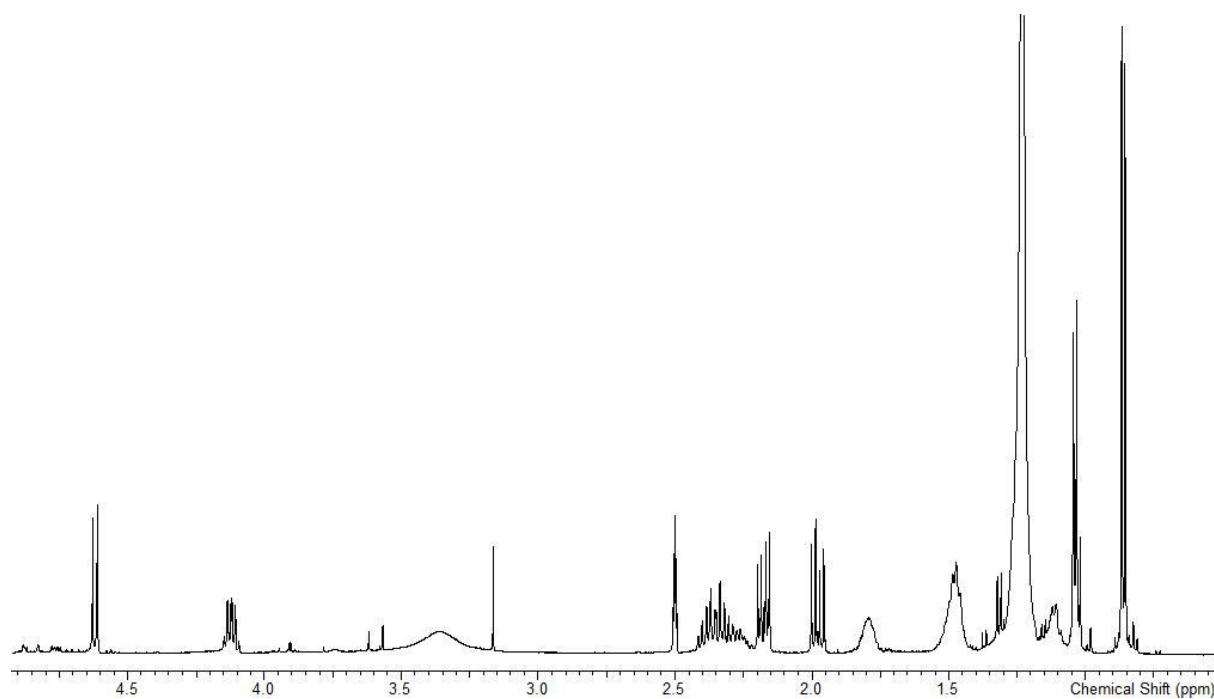


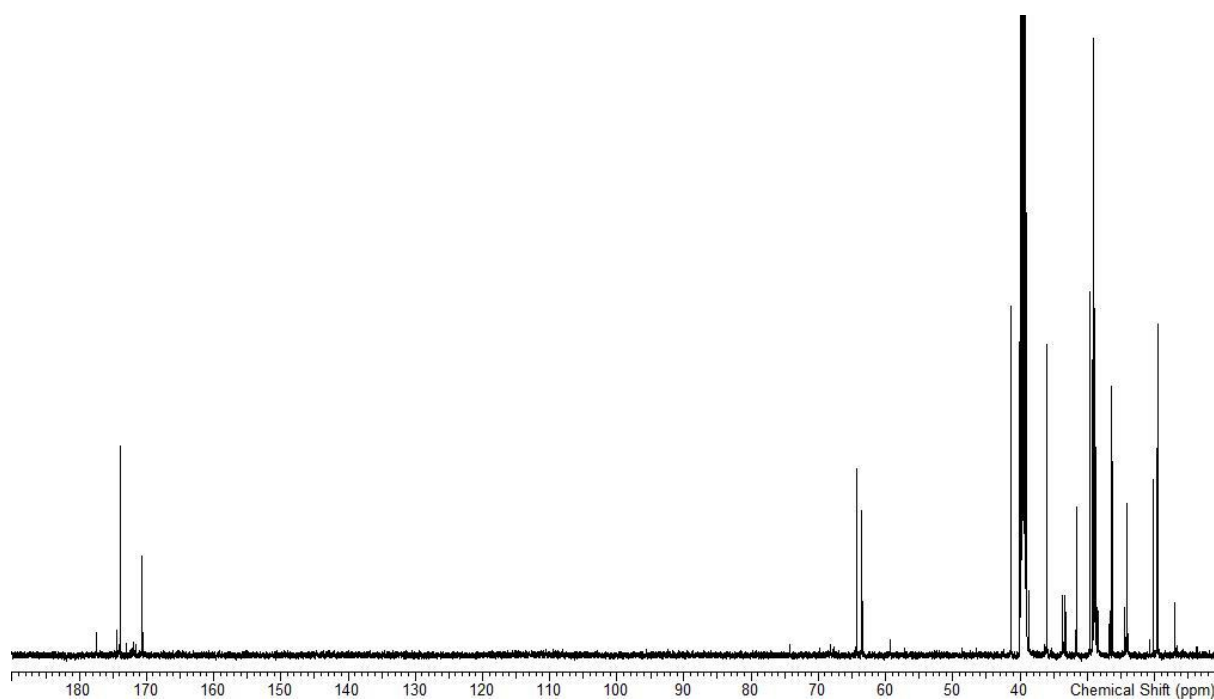
Figure S58: HSQC spectrum (500 MHz, DMSO- $d_6$ ) of *iso*-lipothrenin A (15).



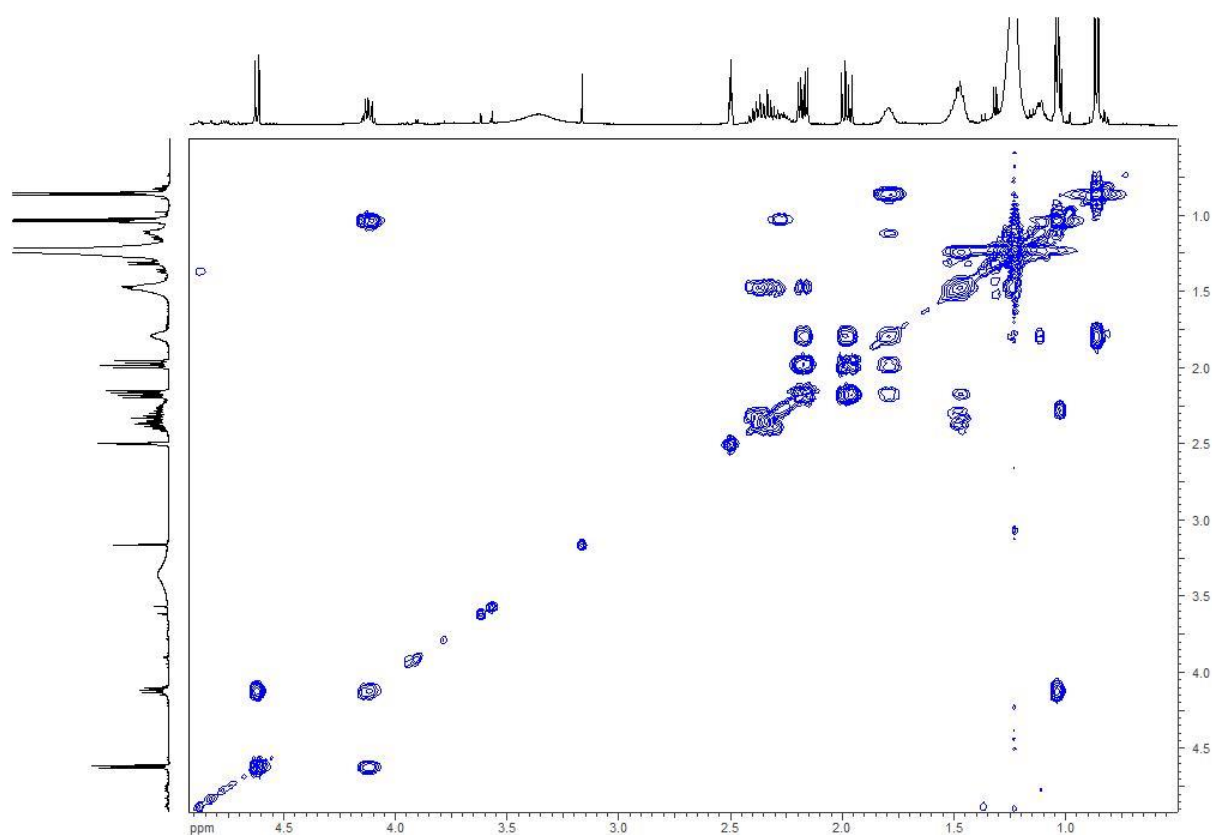
**Figure S59:** HMBC spectrum (500 MHz, DMSO- $\text{d}_6$ ) of *iso*-lipothrenin A (**15**).



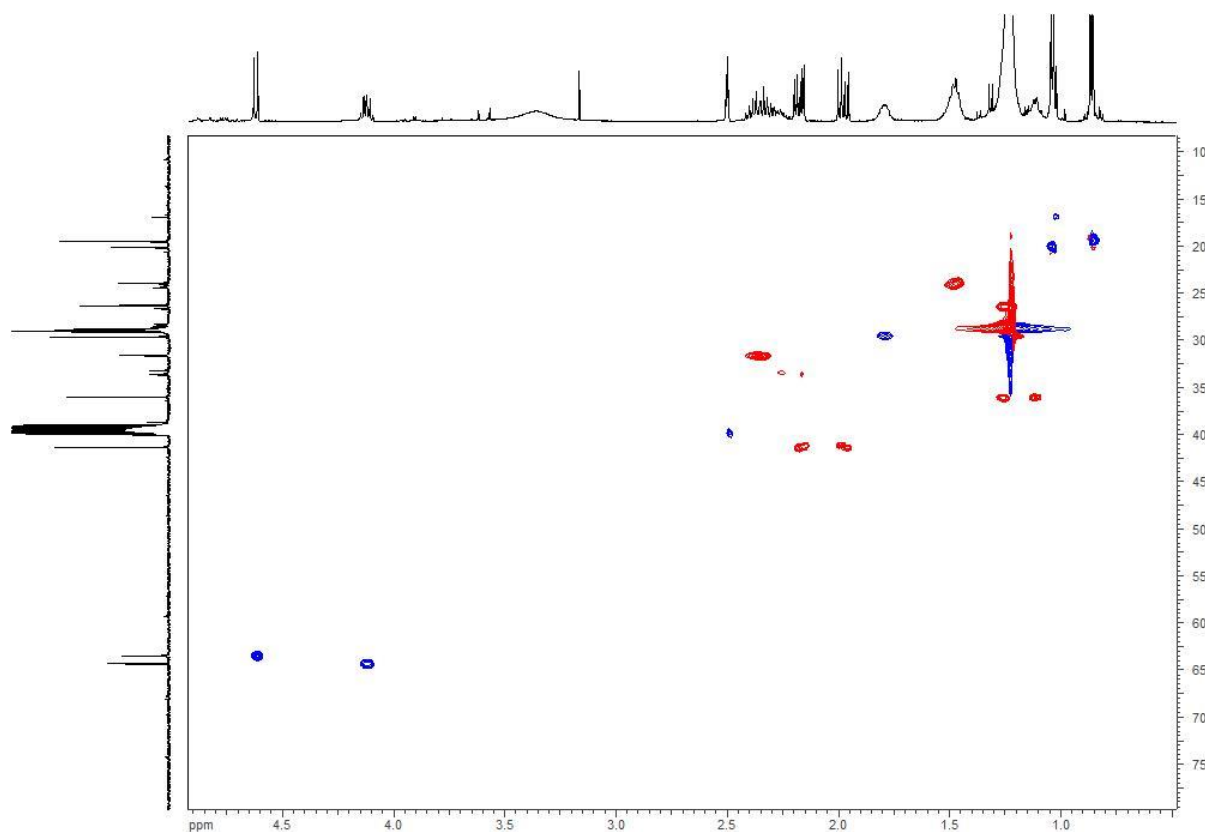
**Figure S60:**  $^1\text{H}$  NMR spectrum (500 MHz, DMSO- $\text{d}_6$ ) of 14-methyl-lipothrenin A (**16**).



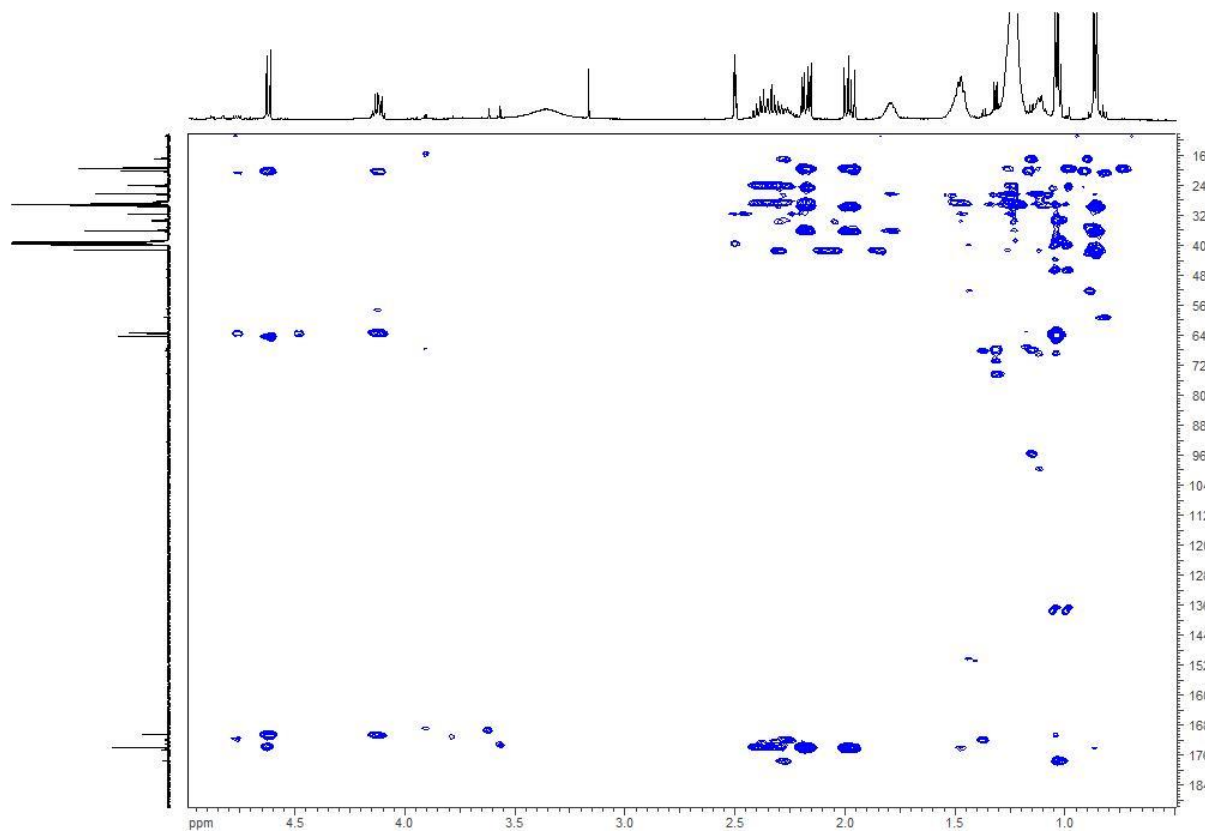
**Figure S61:**  $^{13}\text{C}$  NMR spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 14-methyl-lipothrenin A (**16**).



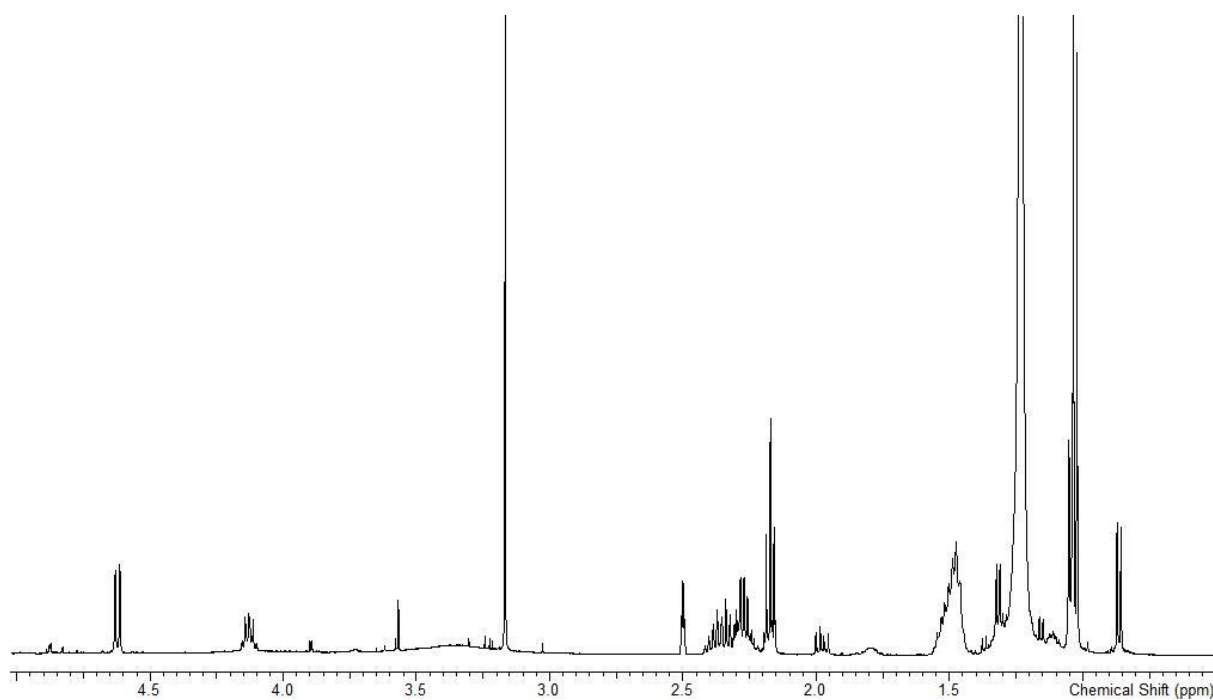
**Figure S62:**  $^1\text{H}$ - $^1\text{H}$ -COSY spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 14-methyl-lipothrenin A (**16**).



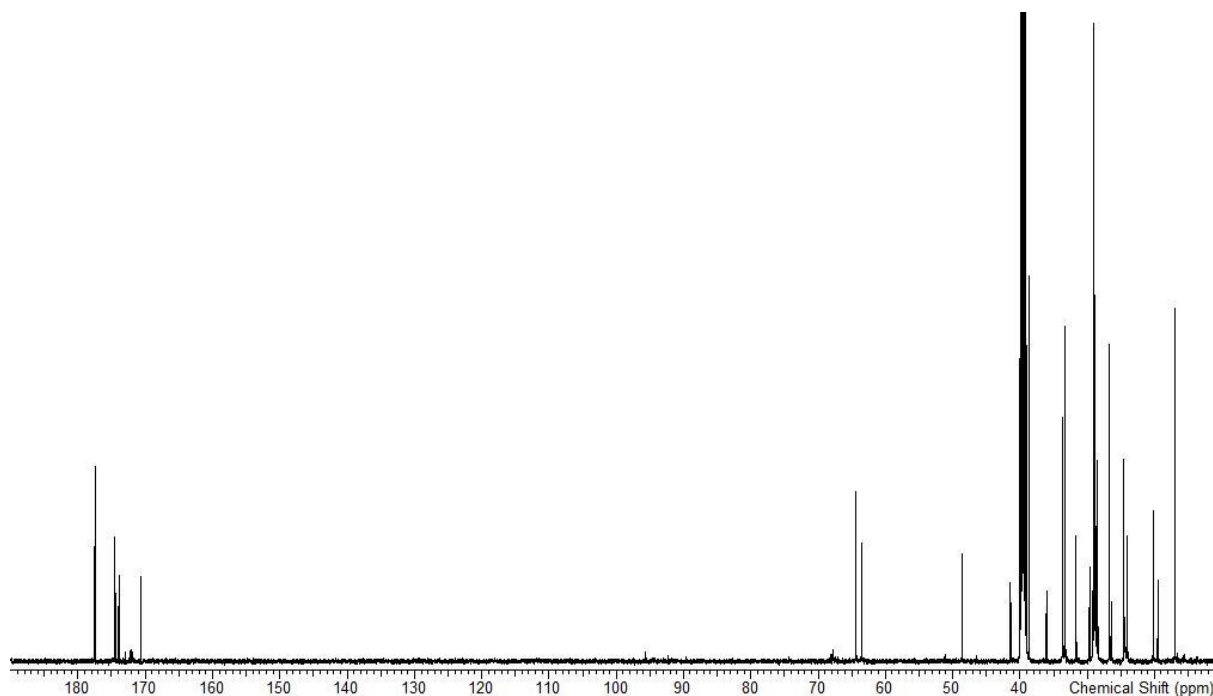
**Figure S63:** HSQC spectrum (500 MHz, DMSO- $d_6$ ) of 14-methyl-lipothrenin A (**16**).



**Figure S64:** HMBC spectrum (500 MHz, DMSO- $d_6$ ) of 14-methyl-lipothrenin A (**16**).



**Figure S65:**  $^1\text{H}$  NMR spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 15-methyl-lipothrenin A (**17**).



**Figure S66:**  $^{13}\text{C}$  NMR spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 15-methyl-lipothrenin A (**17**).

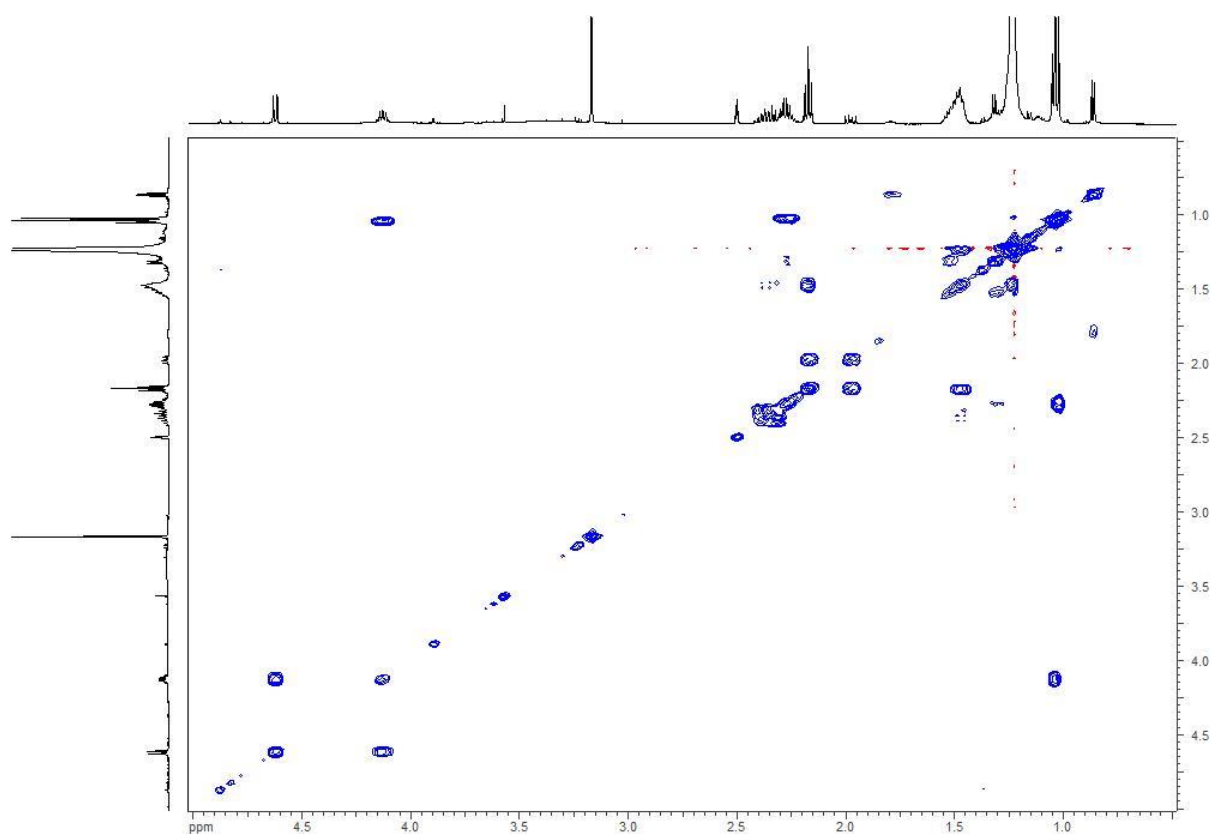


Figure S67:  $^1\text{H}$ - $^1\text{H}$ -COSY spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 15-methyl-lipothrenin A (**17**).

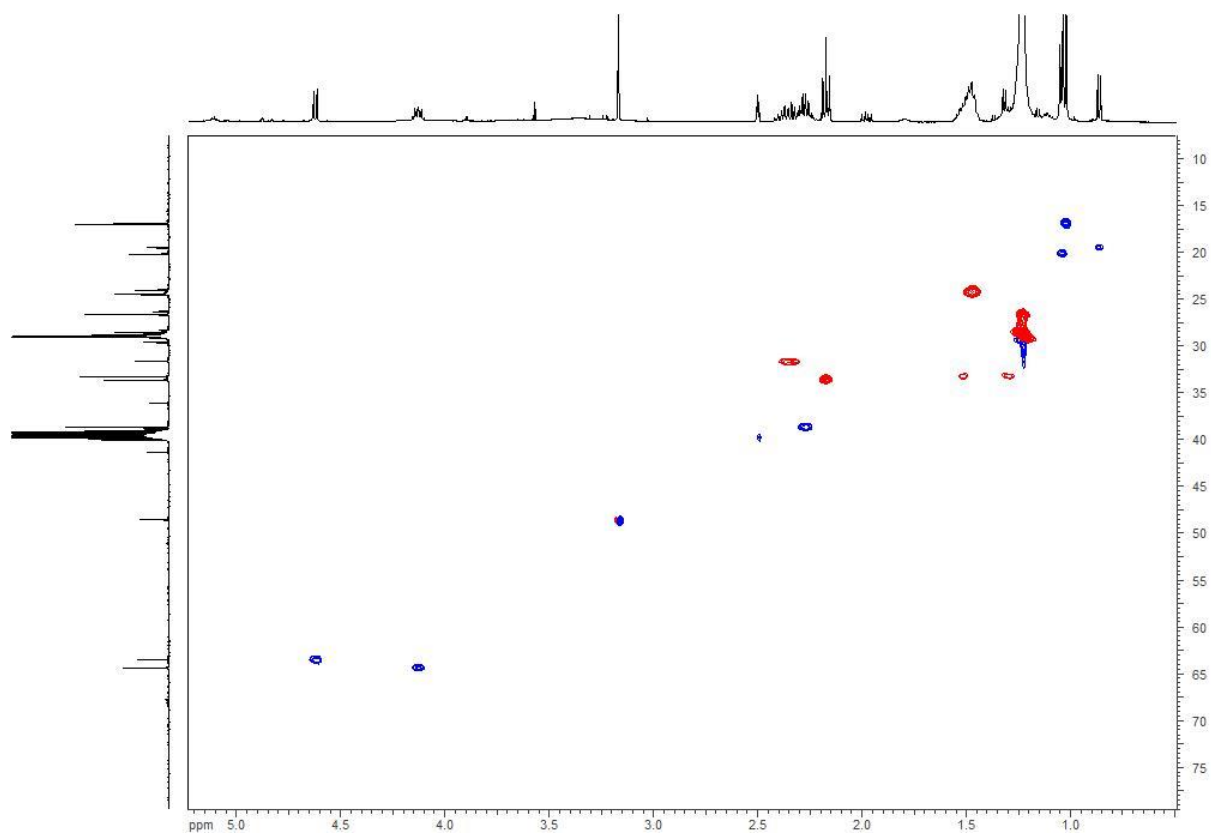


Figure S68: HSQC spectrum (500 MHz,  $\text{DMSO-d}_6$ ) of 15-methyl-lipothrenin A (**17**).

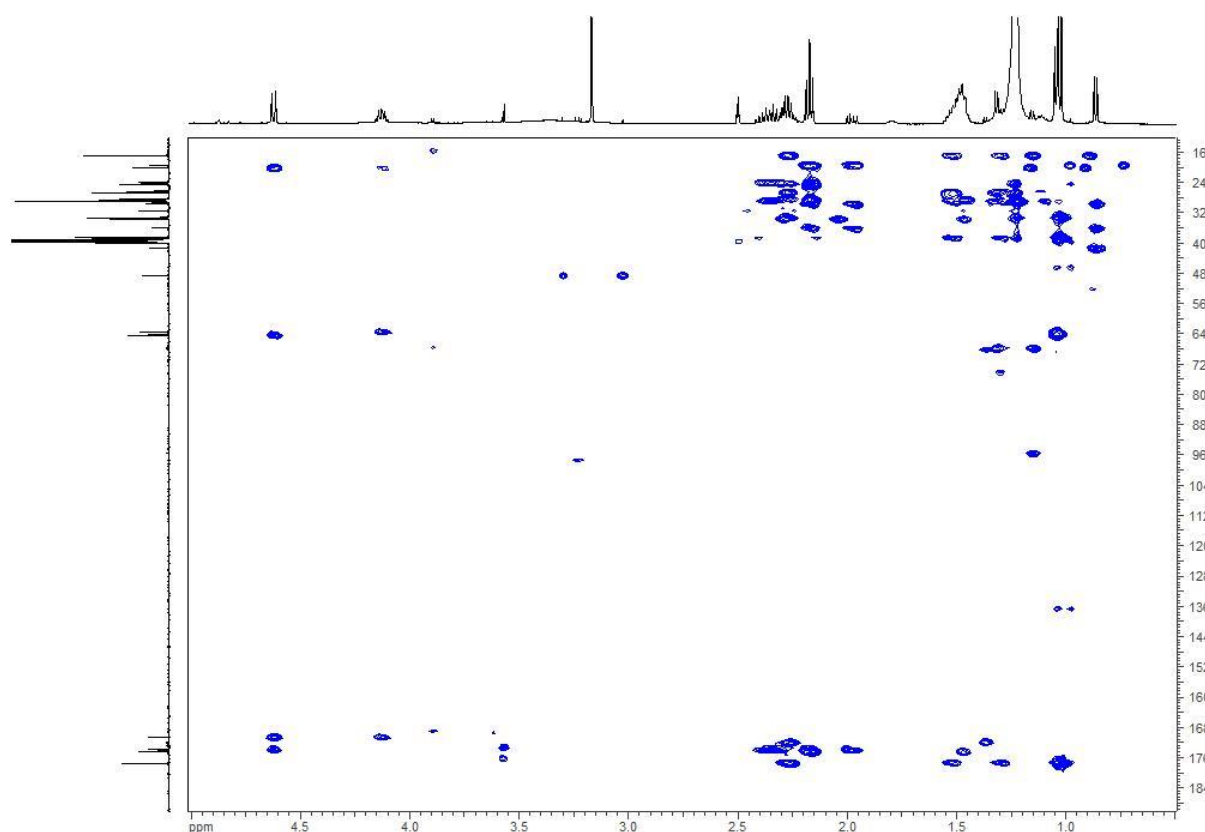


Figure S69: HMBC spectrum (500 MHz, DMSO- $d_6$ ) of 15-methyl-lipothrenin A (**17**).

## 6 References

1. Myronovskiy, M.; *et al.* Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. *Metab. Eng.* **2018**, *49*, 316-324, 10.1016/j.ymben.2018.09.004.
2. Ahmed, Y.; *et al.* Engineering of *Streptomyces lividans* for heterologous expression of secondary metabolite gene clusters. *Microbial cell factories* **2020**, *19* (1), 1-16, 10.1186/s12934-020-1277-8.
3. Flett, F.; *et al.* High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. *FEMS Microbiol. Lett.* **1997**, *155* (2), 223-229, 10.1111/j.1574-6968.1997.tb13882.x.



## 4 Summary and Conclusion

New chemical entities are urgently needed in the everlasting fight against microbial infections. *Streptomyces* are a valuable source of structurally diverse natural products (NPs) with a variety of biological activities. After years of research, the biosynthetic potential of this genus was assumed to be exhausted. However, advances in analytical chemistry, genetics and biotechnology revealed a treasure chest full of undiscovered NPs hidden in the genome of *Streptomyces*. Herein, we applied dereplication, genome mining and heterologous expression to find new natural products and their corresponding biosynthetic gene clusters (BGCs) from *Streptomyces*. Furthermore, the biosynthetic pathways of new compounds were analyzed to gain a deeper understanding about NP biosynthesis, to find new biosynthetic mechanisms and to discover new enzymes. The aim of the thesis was to enrich the field of natural product research by contributing new chemical structures, biosynthetic gene clusters and enzymes that can potentially be used for combinatorial biosynthesis or enzyme catalysis. For this purpose we screened a library of less studied *Streptomyces* strains by dereplication of their metabolome. Thereby, two strains were selected due to their production of unidentified compounds: *Streptomyces aurantiacus* LU19075 and *Streptomyces aureus* LU18118.

In culture broth of *S. aurantiacus* LU19075, three new compounds with similar masses were identified. Isolation and structure elucidation led to the identification of new depsibosamycins (**7**) which are cyclic variants of the previously discovered linear bosamycins.<sup>1</sup> Depsibosamycins are octapeptides containing the following amino acid sequence: L-Tyr, D-Tyr, L-Leu, L-erythro- $\beta$ -OH-Asp, L-Ser, 5-MeO-D-Tyr, L-Leu, and Gly. In contrast to the linear bosamycins, which were also produced by the strain, the cyclic depsibosamycins show a side-chain-to-tail lactonisation of serine and glycine (Figure 1). The discovery of depsibosamycin and the structural similarities to hyeptin is of interest since cyclic peptides more likely possess biological activities. However, the degradation of the compound prevented any biological testing. Genome mining revealed an NRPS cluster which showed high similarities (98%) to the previously reported bosamycin BGC.<sup>2</sup> This seemed strange since the previous study only reported the linear bosamycin as the product of the cluster. Heterologous expression of the depsibosamycin BGC from a genomic library constructed for *S. aurantiacus* LU19075 revealed high concentrations of the cyclic compounds in the culture broth of the heterologous host, while the linear versions were barely detectable. This suggests the cyclic depsibosamycins are the actual product of the cluster rather than the linear bosamycins. To verify this, we compared the time course of production of depsibosamycin and bosamycins. Thereby, we observed that the concentration of depsibosamycin decreases over time while bosamycins increased. This indicates that the linear bosamycins are a product of physicochemical or enzymatic degradation

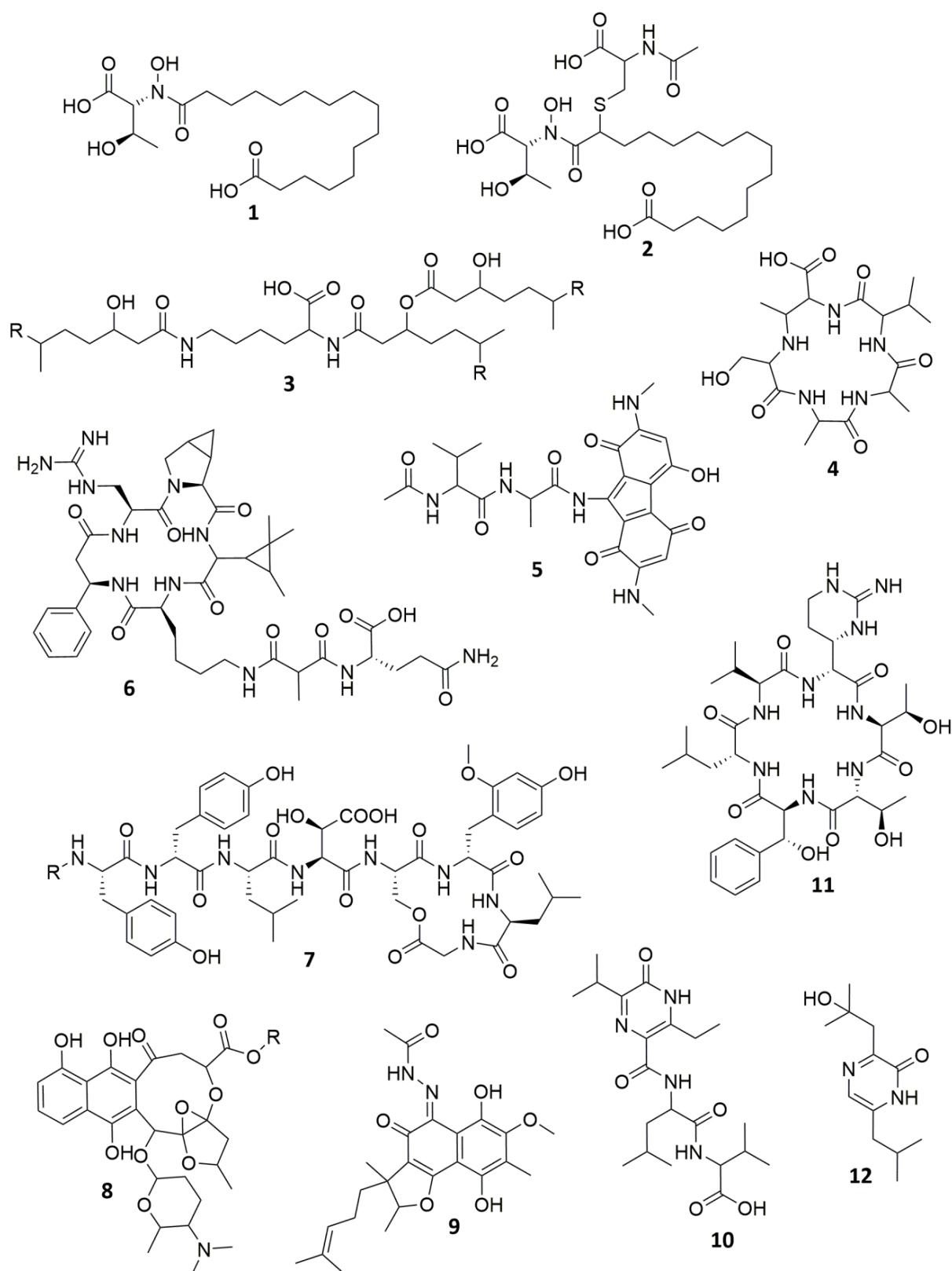
of the cyclic depsibosamycins. In summary, we found out that the cyclic depsibosamycins are the actual product of its corresponding BGC rather than the linear bosamycins. This might be also the case for the previously discovered bosamycin gene cluster.

The second strain *S. aureus* LU18118 was selected due to a plethora of unidentified compounds in its culture broth. We investigated a genomic library constructed for *S. aureus* LU18118 to identify biosynthetic gene clusters encoding some of the found compounds. Potential clusters were annotated by antiSMASH and a selection of clusters was expressed in the optimized heterologous hosts *S. albus*  $\Delta 14$  and *S. lividans*  $\Delta 8$ . As a result, a fatty acid cluster on BAC I6 was found producing new lipothrenins. Lipothrenins consist of hexadecanedioic acid which is amide linked to L-threonine, D-*allo*-threonine or dehydrobutyrine (Figure 1). The proposed main product lipothrenin A (**1**) carries an N-hydroxyl group which is exclusively found on D-*allo*-threonine. All lipothrenin derivatives exist as a heavier version carrying an N-acetyl cysteine (NAC) at the  $\alpha$ -position of the fatty acid chain (**2**). The lipothrenin biosynthetic gene cluster (*lit* BGC) was analyzed by sequence alignments tools and gene deletion studies to predict the minimal *lit* BGC and to assign the biosynthesis of lipothrenins. Gene deletions of the left and right shoulders revealed that the minimal cluster includes 19 genes (*liA-litS*). Ketosynthase *LitB* is involved in fatty acid biosynthesis where it catalyzes the formation of the linear hexadecanedioic acid rather than the branched chain fatty acids which are usually observed in actinomycetes. Hexadecanedioic acid is activated by fatty acyl-AMP ligase *LitD* and acyl carrier protein *LitF*. Enzymes involved in the amide bond linkage of the threonine isomers and the isomerization of L-threonine could not be identified within the *lit* BGC. The stereoselective N-oxygenation of D-*allo*-threonine is catalyzed by diiron oxygenase *LitN* in combination with DUF4873 protein *LitO*. The N-acetyl cysteinylolation is likely originating from a common process in streptomycetes that serves for detoxification and cell stress protection. In summary, new lipothrenins and a corresponding lipothrenin biosynthetic gene cluster have been discovered. Furthermore, a lipothrenin biosynthesis pathway has been described. Highlights in the biosynthesis are the ketosynthase *LitB* and the N-oxygenation enzymes *LitN/LitO*. *LitB* could be used to engineer fatty acid production to achieve the selective production of linear fatty acids. *LitN/LitO*, on the other hand, might be of interesting as enzyme catalysts for late stage modifications in natural product total synthesis.

*De novo* structure elucidation of new compounds from colleagues was a major part of this thesis. In one of the collaborations, new compounds were discovered by target genome mining of genes which are responsible for the biosynthesis of the cyclic arginine derivative capreomycin. Non-ribosomal peptides which contain the rare amino acid capreomycin are of interest since they express important biological activities.<sup>3, 4</sup> The biosynthesis of capreomycin is catalyzed by alpha-

ketoglutarate-dependent arginine beta-hydroxylase and pyridoxal-phosphate (PLP)-dependent aminotransferase. Targeted genome mining for these genes revealed an unknown NRPS gene cluster in *S. albus* J1074 carrying the arginine beta-hydroxylase-encoding gene homolog. The NRPS cluster products were identified by comparisons of the production profile of two previously developed *S. albus* mutant strains with and without deletion of the NRPS cluster. Two new compounds were isolated and their structures, including absolute configurations, were determined by NMR, MS/MS fragmentation and Marfey's method. This led to the discovery of new pentapeptides cyclofaulnamycin (**11**) and the linear iso-faulnamycin (Figure 1).<sup>5</sup> The compounds contain the following amino acids: D-capreomycidine, L-threonine, D-*allo*-threonine, L-*threo*- $\beta$ -phenylserine, D-leucine and L-valine. The peptide is similar to the previously discovered faulnamycine which comprises the same amino acids except it is containing D-*threo*- $\beta$ -phenylserine instead of the L-isomer.<sup>6</sup> Tryon *et al.* suspected the presence of a cyclic derivative because the structure of faulnamycin could not be correctly projected onto the corresponding biosynthetic gene cluster. However, in contrast to our studies, a cyclic faulnamycine could not be identified. Tyron *et al.* also proposed the cyclisation of capreomycidine is catalyzed by an epimerase domain within the faulnamycin BGC since the essential pyridoxal-phosphate (PLP)-dependent aminotransferase is missing in the cluster.<sup>6</sup> We identified this gene outside of the cluster in the genome of the cyclofaulnamycin producing *S. albus* strain. Deletion of the gene led to a 5-12 fold decrease of the production of iso- and cyclofaulnamycin. This confirms that the main mechanism towards D-capreomycidine involves the PLP-dependent aminotransferase. In summary, by target genome mining of capreomycidine biosynthesis genes, new iso- and cyclofaulnamycin have been discovered. The cyclic cyclofaulnamycin is the expected main product of the cluster while the linear version is likely formed by physicochemical or enzymatic degradation. Furthermore, we found out that the cyclization reaction which results in D-capreomycidine is mainly catalyzed by a PLP-dependent aminotransferase which was identified outside of the cluster.

In other collaboration with co-workers, structure elucidation by NMR of new compounds has revealed the impressive variety of chemical scaffolds which can be produced by streptomycetes (Figure 1). A highlight is the cyclic non-ribosomal peptide cyclohuinilsiopeptin (**6**), which contains four unnatural amino acids: 3-azabicyclo[3.1.0]hexane-2-carboxylic acid (AHCA),  $\alpha$ -amino-2,2,3-trimethyl cyclopropaneacetic acid (ATCA), 2-amino-3-guanidinopropanoic (AGA) acid and D- $\beta$ -phenylalanine. The unique amino acid residues are of high interest for chemists and natural product scientist since they are challenging to synthesize *in vivo* and *in vitro*. Another highlight was the hipposudoric acid derivative **5** which has been discovered by heterologous expression of a gene cluster in *S. albus*  $\Delta$ 14. The compound is a chromophore which appears dark green and shows sensitivity towards pH changes like an indicator.



**Figure 1:** All compounds which have been structurally characterized within the time period of this thesis: lipothrenin A (**1**) and NAC-lipothrenin A (**2**) (unpublished), dudomycins<sup>7</sup> (**3**), bonsecamin<sup>8</sup> (**4**), hipposudoric acid derivative (**5**) (unpublished), cyclohuinilsoeptin (**6**) (unpublished), depsibosamycins<sup>1</sup> (**7**), new PKS II compound (**8**) (unpublished), furaquinocin derivative (**9**) (unpublished), JBIR-56 derivative (**10**) (unpublished), cyclo-faulknamycin<sup>5</sup> (**11**) and flavacol derivative (**12**) (unpublished).

Further structure elucidations led to the identification of the NRPS compounds dudomycin (**3**) and bonsecamin (**4**) which have been found after heterologous expression of cryptic BGCs, a new PKS II compound **8** with a unique polycyclic ether moiety, a furaquinocin C derivative (**9**) with a rare acetylhydrazin moiety, a derivative of JBIR-56 (**10**) with an undescribed biosynthesis and a flavacol derivative (**12**) (Figure 1).

In the course of this work, *Streptomyces* have been shown to be still a valuable source of natural products even after years of intensive study. Dereplication and heterologous expression of silent gene clusters have been proven to be powerful methods to find new NPs. Both methods led to the discovery of new natural products, new biosynthetic gene clusters and even some interesting gene functions. Despite the discovery process has accelerated over the years due to constant development of instruments and methods, nowadays industrial drug discovery and development based on natural products is still barely economically feasible. Meanwhile, antibiotic resistance is on the rise and soon going to cast a shadow even on the corona pandemic. Natural product research can be one of the solutions to this problem, but there is still a lot of work to be done to accelerate natural product discovery especially through innovative downstream processing.

## References

1. Stierhof, M.; *et al.* Discovery and Heterologous Production of New Cyclic Depsibosamycins. *Microorganisms* **2021**, *9* (7), 1396, 10.3390/microorganisms9071396.
2. Xu, Z.F.; *et al.* Discovery and biosynthesis of bosamycins from *Streptomyces* sp. 120454. *Chemical Science* **2020**, *11* (34), 9237-9245, 10.1039/D0NP90037K.
3. Atkinson, D.J.; *et al.* Enduracididine, a rare amino acid component of peptide antibiotics: Natural products and synthesis. *Beilstein journal of organic chemistry* **2016**, *12* (1), 2325-2342, 10.3762/bjoc.12.226.
4. Thomas, M.G.; *et al.* Deciphering tuberactinomycin biosynthesis: isolation, sequencing, and annotation of the viomycin biosynthetic gene cluster. *Antimicrobial agents and chemotherapy* **2003**, *47* (9), 2823-2830, 10.1128/AAC.47.9.2823-2830.2003.
5. Horbal, L.; *et al.* Cyclofaulknamycin with the Rare Amino Acid D-capreomycinidine Isolated from a Well-Characterized *Streptomyces albus* Strain. *Microorganisms* **2021**, *9* (8), 1609, 10.3390/microorganisms9081609.
6. Tryon, J.H.; *et al.* Genome mining and metabolomics uncover a rare D-capreomycinidine containing natural product and its biosynthetic gene cluster. *ACS chemical biology* **2020**, *15* (11), 3013-3020, 10.1021/acscchembio.0c00663.
7. Lasch, C.; *et al.* Dudomycins: New Secondary Metabolites Produced after Heterologous Expression of an Nrps Cluster from *Streptomyces albus* ssp. *Chlorinus* Nr1 B-24108. *Microorganisms* **2020**, *8* (11), 1800, 10.3390/microorganisms8111800.
8. Lasch, C.; *et al.* Bonsecamin: A New Cyclic Pentapeptide Discovered through Heterologous Expression of a Cryptic Gene Cluster. *Microorganisms* **2021**, *9* (8), 1640, 10.3390/microorganisms9081640.