Heterologous overproduction of new natural products in *Streptomyces*

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

von
Nikolas Peter Eckert

Saarbrücken
2022

Tag des Kolloquiums:	07.10.2022
Dekan:	Prof. Dr. Jörn Erik Walter
Berichterstatter:	Prof. Dr. Andriy Luzhetskyy
	Prof. Dr. Christoph Wittmann
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 Oktober 2017 bis Juni 2022.

“The only difference between screwing around and science is writing it down.”

Alexander Jason

Acknowledgements

First of all I would like to thank Andriy for giving me the opportunity to work on a PhD project and let me join the AMEG “family”. He was giving me good advices as well as some freedom to make my own experiences.

A BIG thanks to Liliya for being a great supervisor, being a good example on how to work efficiently, giving me often feedback, helping with problems as well as doing some activities outside of the laboratory.

Additionally, I would like to thank Stepan for helping me out quite a lot as well as doing some nice trips. Nazar for having some nice walks home. Constanze Paulus for spending time with me, doing sports, drinking a beer from time to time. Constanze Lasch for nice hiking trips, nature trips as well as “Dönerstag”. Marta for some “Feierabendbier” and cooking together. Marc for doing sports as well, always being fully motivated and for trusting me in being his first contact in cases of accidents. Maksym and Birgit for giving always some good advice. Anja for being always there for a shenanigans in the lab. So basically all work group members for making the time an enjoyable time. And last but not least Nils for some nice cooking evenings, poker, sport and making the time in the lab even more fun, with tip filling competitions and many more. There are many more things to mention but before it gets too long: Thank you for having me and being good colleagues as well as friends.

I would like to thank Prof. Dr. Christoph Wittmann for investing the time to review this thesis and accompany me through my thesis as well as excursions to BASF.

Zum Schluss möchte ich noch meiner Familie und meinen Freunden danken, für die Unterstützung und dass sie mich so hinnehmen wie ich bin.

Zusammenfassung

Sekundärmetabolite, die oft als Naturstoffe bezeichnet werden, werden oft von Actinobakterien (meist von Streptomyceten) produziert und sind eine wichtige Quelle für Antibiotika und andere Arzneimittel. Die Genomdaten von Actinobakterien haben gezeigt, dass ihr großes biosynthetisches Potenzial noch nicht ausgeschöpft ist. Der Hauptgrund dafür ist, dass die meisten für die Naturstoffproduktion verantwortlichen biosynthetischen Gencluster unter Laborbedingungen nicht aktiv sind oder ihr Expressionsniveau zu niedrig ist, um neue Moleküle nachweisen zu können.

Der erste Teil dieser Arbeit befasst sich mit der Entdeckung, Isolierung und Charakterisierung der neuen Naturstoffe Cyclofaulknamycin und Aorimycin. Die Entdeckung von Cyclofaulknamycin zeigt, dass selbst gut beschriebene Modellorganismen wie *Streptomyces albus* J1074, die in hunderten von Laboren weltweit untersucht werden, noch unbekannte bioaktive Moleküle verbergen können. Dies gilt umso mehr für neu isolierte Bakterien, was am Beispiel der neuartigen Aorimycine gezeigt wurde.

Der zweite Teil konzentriert sich auf bekannte Verbindungen mit unbekanntem pharmazeutischem Potenzial. Einer der Hauptgründe dafür ist die geringe Produktionsausbeute, die zu Engpässen bei der Profilierung von Verbindungen führt. Hier werden zufällig-rational konzipierte Promotoren eingesetzt, um die Naturstoff-Produktion zu steigern. Mit Hilfe dieser Methode wurden neue Derivate isoliert und weiter charakterisiert.

Abstract

Secondary metabolites, often referred to as natural products (NPs), produced by Actinobacteria (mostly by *Streptomyces*) are a major source for antibiotics and other drugs. The genome data of Actinobacteria revealed their great biosynthetic potential to be far from fully exploited. The major reason is that most corresponding biosynthetic gene clusters responsible for the NP production remain "silent" under laboratory conditions or their expression level is too low for the detection and isolation of new molecules.

The first part of this thesis is focussing on the discovery, isolation and characterisation of the new NPs cyclofaulknamycin and aorimycins. The discovery of cyclofaulknamycin shows that even well described model organisms like *Streptomyces albus* J1074, which is under investigations in hundreds of laboratories worldwide, have the potential to still conceal unknown bioactive molecules. This applies even more for newly isolated bacteria, which was shown by the example of the novel NPs aorimycins.

The second part focusses on known compounds with unknown pharmaceutical potential. One of the main reasons for that is a low production yield resulting in supply shortage for the profiling of compounds. Herein, a method is tested which utilised random rational designed promoters to increase the NP production level. By the use of this method new derivatives were isolated and further characterised.

Publications

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

Eckert N., Paulus C., Stierhof M., Horbal L., Luzhetskyy A., New PKS type II aorimycin with unique ring arrangements and heterologous expression thereof, *to be submitted*

Eckert N., Horbal L., Rebets Y., Zapp J., Herrmann J., Müller R., Luzhetskyy A., Discovery and isolation of novel highly active pamamycins through transcriptional gene cluster engineering, *to be submitted*

Conference contributions

Eckert N., Horbal L., Zapp J., Herrmann J., Müller R., Luzhetskyy A., Metabolic profile modification of the pamamycin production, through transcriptional gene cluster engineering, *HIPS symposium, Saarbrücken, Saarland*, **2019** (Poster)

Eckert N., Horbal L., Zapp J., Herrmann J., Müller R., Luzhetskyy A., Metabolic profile engineering through transcriptional gene cluster "refactoring"**,** *14th international symposium on the Genetics of industrial microorganisms (GIM), Pisa, Italy*, **2019** (Poster)

Eckert N., Horbal L., Zapp J., Herrmann J., Müller R., Luzhetskyy A., Metabolic profile modification of the pamamycin production, through transcriptional gene cluster engineering, *Doktorandentag, Saarbrücken, Saarland*, **2019** (Poster)

Table of Contents

1. Introduction

1.1 History and importance of natural products

In a broad sense, natural products (NPs) are substances which are produced by living organisms and can be classified in two groups: **Primary** and **secondary** metabolites. Primary metabolites are molecules that are crucial for the survival of organisms, for instance nucleotides, amino acids and carbohydrates. With the need for the production of more and even faster available meat products, companies started to add amino acids like lysine as food additives, which became crucial for current industry along the way. It belongs to the essential amino acids, meaning that it cannot be synthesised by the human body or insufficient amounts and therefore needs to be absorbed from external food supplies [1]. Amino acids are used as food additives not only for livestock but for humans as well [2]. The fermentative production of L-lysine for the year 2015 was 2.4 million tons and is expected to grow about 5-7 % each year [3], which is representing a market value of more than 3 billion US$. Contrary, secondary metabolites are substances which are not essential for the organism's survival but can give a certain benefit to the organism, e.g. siderophores, which help the organism to gather important ions or antimicrobial substances to contest with competing organisms. Synthesis of such substances has high energy costs for the producing organism, therefore their biosynthesis is strictly regulated [4]. An example for a famous secondary metabolite is penicillin. The advantage of this fungus derived secondary metabolite is the growth inhibition of gram positive bacteria, by inhibiting the bacterial cell wall biosynthesis [5]. Consequently, studies revealed the possibility to use it as an antibiotic for treatment of infection diseases [6]. Nonetheless, NPs are not only well known for their antibiotic activity, but were also found to interact extraordinarily good with proteins in general and thus living organisms [7-10]. Hence, they were found to possess antiparasitic [11], anticancer [12], antiviral [13] and immunosuppressive [14] properties, as was shown in numerous studies. Besides their application in the medical field NPs are also used in veterinary, agriculture [15] as well as cosmetic industry [16]. Both primary and secondary metabolites are of unmatched importance, which will only increase over time [17].

NPs have been used by humans knowingly or unknowingly since more than 3000 years. Skeletons found in Sudanese Nubia showed incorporation of tetracycline [18]. These bones were dated back to the years 350-550 CE. A second example for the already early use of NPs, is acetylsalicylic acid, the nowadays most commonly used pain killer [19]. The bark of

willow-trees, which is at the present day known to contain precursors of acetylsalicylic acid, was already employed 3500 years ago in Egypt, as stated in ancient texts [20]. Even though NPs have been used over a long period of time only the development of further technologies enabled the correct identification and characterisation of active agents and thus a more efficient use thereof. With the beginning of the 20th century it was observed that soil bacteria can exhibit inhibitory effects on the growth of other bacteria [21]. In combination with the growing technological possibilities, many more antimicrobial active compounds have been isolated. Especially in the field of actinomycetes a noteworthy person is Selman Waksman. He started the characterisation of bacteria inhabiting the soil and stated that about 20-50 % of those bacteria exhibit growth inhibiting effects on test strains [22]. These findings mark the start of the so called "golden age" of antibiotic discovery in the 1940s, which lasted for approximately 20 years [23]. In this time most of the major classes of known antibiotics have been discovered and characterised [24]. Nonetheless, the end of this "golden age" was marked by a high rediscovery rate of already known compounds. These findings led to a withdrawal of big pharma companies from the field of NPs research [15].

On the one hand, the discovery of antibiotics and the broad possibility to use them, led to a significant decrease of the mortality rate caused by bacterial infections [25]. But on the other hand, it also led to the faster "evolution" of antibiotic resistant pathogen [26]. If several resistances evolve in one organism or are transferred via horizontal gene transfer [27], a multi drug resistant organism can arise, like the methicillin resistant *Staphylococcus aureus* (MRSA) [28] or vancomycin-resistant enterococci (VRE) [29, 30]. Due to their broad resistance to commonly used antibiotics, an infection of those bacteria is especially difficult to treat. Since resistances arise also through unreasonable usage of most antibiotics, the world health organisation (WHO) recommends that certain antibiotics, like fosfomycin, daptomycin and carbapenems are held back as a kind of last resort antibiotics [31, 32]. Approximately 10 years (with a tendency to shorter time periods) can pass from the introduction of a new antibiotic to clinical usage, until the first detection of resistant bacteria [33]. Accordingly, it is only a matter of time, until resistant bacteria against new and last resort antibiotics will evolve and bacterial infections could become life threatening once again.

To overcome the high rediscovery rate of already known NPs and to satisfy the increasing need for new drugs, high-throughput screenings (HTS) of chemical compound libraries were performed [23]. After all the investments it was obvious that synthetic compounds are not

even closely comparable to NPs, as a source for biological active compounds [9], which is reflected by the fact that 70 % of all antibiotics are isolated or derived from NPs [34]. Compared to NPs, chemically synthesised compounds still look quite simple and when reconstructing complex NPs via organic synthesis, many labour intense reactions are involved. Therefore, the novel approach to screen for new compounds rendered not as promising as the previous approach developed by Waksman. Nevertheless the conventional drug discovery approach suffered from the high rate of rediscovery of known NPs, therefore an alternative was needed. Novel and cheap next generation sequencing methods revealed the huge biosynthetic potential of the already known and thoroughly characterised Actinobacteria strains. Well investigated strains like *Streptomyces coelicolor,* which were believed to only be able to produce a handful of different NPs, were revealed to have the biosynthetic potential to encode for more than 20 NPs, of which 15 should be transcribed according to gene expression analysis data and even fewer have been confirmed via structure analysis until now [35]. Further analysis of other *Streptomyces* revealed that an average *Streptomyces* strain can code for 20-50 biosynthesis gene clusters (BGCs), of which about 90 % are silent under laboratory conditions [36]. In combination with novel molecular biological tools (e.g. heterologous expression and plenty of other methods) those not expressed, commonly referred to as “silent” BGCs, are now being targeted for the search of new NPs. Those new genome mining approaches, in combination with the screening of new ecological niches and symbiotic strains add new possibilities for NP discovery [37].

1.2 Discovery of new NPs

All used methods for NP discovery can be clustered into two main approaches: Top-down and the bottom-up approach [38]. The top-down approach is characterised by a systematic screening of strains and NPs thereof, for certain activities. One of the first persons to establish a systematic screening was Waksman in the year 1940. His screening system involved the isolation of microbes from soil samples, cultivating them under different conditions and checking for inhibition zones around single colonies [39]. Following the active colonies could be tested against certain pathogenic bacteria. In combination with the exploration of new, underexplored ecological niches, this proofs to be still a powerful tool for the discovery of NPs [40, 41]. To cover the last part of the top-down approach the comparative metabolic profiling needs to be mentioned. In general, this method is quite broad and utilises different methods like HPLC-MS analysis to compare extracts with each other, to identify new

compounds. Those extracts can be from the same strain cultivated under different conditions, from different strains or from the same strain with different mutations.

Unlike the top-down approach, the bottom-up approach does not focus on the strain or different cultivation conditions to find new NPs. It starts with the identification of a potential interesting BGC, following its activation and screening for new NPs [38]. In bacteria, genes responsible for the synthesis of NPs are commonly clustered together and certain motives of key genes have been identified. There are several bioinformatic tools to identify BGCs based on their specific pattern. One of the most commonly used tools is antibiotics and secondary metabolites analysis shell (antiSMASH). It is an open source software developed for the identification of secondary metabolite clusters within genomes [42]. Other widely used similar tools and databases are MIBiG [43] and PRISM [44]. By the use of those tools the discovery of potential BGCs and following experiments can be considerably accelerated. The discovery that potentially more than 20 BGCs are encoded in only one *Streptomyces* led to the one strain many compounds approach (OSMAC) [45], in which a variety of different methods to activate those BGCs is used. OSMAC represents the bottom-up approach in its early stages. It can be simple variations of culture conditions like in the top-down approach or activation via molecular biological methods. With the help of those bioinformatic tools in combination with molecular biology it is now possible to focus on the clusters and screen for their encoded NPs. An additional advantage of this approach is the accessibility of strains which are not cultivable under laboratory conditions. It is estimated that about 99 % of all bacterial strains are not cultivatable and are only detectable by their DNA, which is referred to as environmental DNA (eDNA) [46, 47]. Using the above mentioned bioinformatic tools whole metagenomic libraries can be screened for interesting clusters. The downside of this method is obviously that only similar clusters can be detected, since the bioinformatic tools can only screen for similar genes of known clusters. This results in a low discovery rate of principally novel biosynthesises, which are more likely to be found by systematic or random expression approaches [48]. Should those detected gene clusters not be expressed under standard conditions or certain stress factors, there is an abundance of methods, as well as combinations thereof, to activate them.

Different methods to activate clusters exist and can be grouped in two categories: directed and undirected. Indifferently from the category, the methods to activate a cluster always focus on the transcription, translation or a combination of both. Undirected methods for random

mutagenesis can be applied for a strain which is not easily accessible by molecular biological techniques or when nothing is known about the respective cluster [49]. Commonly used methods for random mutagenesis of whole strains are UV radiation, chemical substances or molecular biologically by transposon mutagenesis. The expression of an already cloned cluster located on a vector can be optimised by random mutagenesis as well. This can lead to an activation of the production, as well to the identification of biosynthetic key enzymes. The drawback of this method are the unpredictability of what will happen as well as the accumulation of mutations which can be useful but also at a certain point hinder further improvements.

Due to ongoing advances in technology as well as knowledge about regulation networks, directed mutagenesis is nowadays a quite frequently used and viable tool, not only in prokaryotic but also eukaryotic cells [50]. It has been previously observed that the biosynthesis of NPs is tightly regulated within the organism [51]. The regulation in *Streptomyces* has been analysed for several strains and NPs. Within the regulators, positive regulators have been proven to activate silent gene clusters. The activator PapR2 from *S. pristinaespiralis* was able to activate the "silent" gene cluster of undecylprodigiosin in *S. lividans*, as well as the amicetin gene cluster in *S.* sp. SHP22-7 [52]. Deletion of negative regulator genes could also be a possible approach [53]. However, compared to the introduction of relatively conserved activators like the PapR2 activator, the deletion of more specific negative regulators is more labour intense and time consuming.

Another great tool, addressing the strict regulation of both transcription and translation, is the heterologous expression. With the help of several cloning methods from either a known strain or eDNA, BGCs can be transferred via the help of conjugation [54] or electroporation [55] into a more defined host strain. By doing so, the native strict regulation system can be circumvented and the NP can already be produced. If this is not the case, most host strains are accessible by molecular biological methods and the production can be optimised [56], by using well defined promoters [57] and ribosomal binding sites (RBS) [58]. Additionally, the host strain can be modified to optimise the production of certain kinds of NPs, targeting their corresponding precursor pool [59], or by deleting indigenous BGCs [60]. The model organism *S. albus* J1074 was chosen as such a "chassis" strain with a clean metabolic profile. 15 indigenous BGCs were deleted, enabling easier detection of NPs as well as less laborious downstream purification. To maximise the chances for a successful cluster expression clusters

can be artificially synthesised and synthetic promoters can be introduced or RBS and codon optimisation can be applied [61]. As far as this is a total rearrangement and adaptation of the native cluster to optimise the production in a native or heterologous host, this is referred to as "cluster refactoring" [62]. For certain modular NPs this method could be used in the future to artificially design a new NP.

With the availability of molecular biologically accessible host strains the optimisation of the BGCs itself got a lot easier. In combination with the above mentioned methods there is a vast range of possibilities to discover new NPs. However, in most cases the issue of low production yields remains.

1.3 NPs yield impacts pharmaceutical development

With the help of dereplication already known compounds can be identified by HPLC-MS analysis by their unique retention time, UV-Vis spectrum and mass. Therefore, re-isolation of already known compounds can be in some cases circumvented and new compounds identified. However, this method is only useful when good databases are available with detailed information on already known compounds. Plenty of compounds with a similar structure and retention time exist and possibly new compounds can be overseen. To be certain that the detected mass belongs to a desired compound more data are required. Therefore, the HPLC-MS analysis can be supplemented with methods like MS/MS, to analyse if the compound fragments in the predicted pattern or as an available control compound and has a similar retention time and UV-Vis spectrum. Constantly new methods are developed like the HPLC-SPE-NMR, to facilitate the faster analysis of crude extracts even further [63]. Furthermore, genomic data and bioinformatic tools can help in the prediction for possibly known or new compounds. Nonetheless purification of compounds and solving the structure via methods like nuclear magnetic resonance (NMR) is still a crucial and necessary step to confirm the structure of an identified compound. For NMR analysis approximately 0.5-10 mg of the compound is needed, depending on the molecular weight of the NP and the used system. Additionally, about 1-5 mg is needed for basic activity studies. Getting those amounts of new NPs can proof already quite laborious since most strains only produce in the mg/l scale, compared to optimised production strains with several g/l. Accordingly, approaches to increase the production yield are needed to get enough material to perform at least basic analysis. The discovery of new scaffolds is closely connected to low yields, since most NPs are barely produced and therefore not detectable. This close relation of discovery and

overproduction of new compounds is reflected in the usage of similar methods to attain them. In a broad sense the yield can be increased by two strategies: **Optimisation of medium and culture conditions** as well as **strain optimisation**. Both approaches have their benefits and disadvantages which will be presented in more detail below.

Different fermentation options for yield improvements

If a NP is discovered and the production yield proves to be very low, the first improvement to obtain the necessary quantities is commonly the upscaling of the culture volume. For example Kondo et al. 1988 obtained 55 mg of pamamycin-607 after cultivation of a wild-type strain in 200 l [64]. This represents a yield of 0.275 mg/l. The fermentation was performed in 500 ml shake flasks containing 125 ml of medium, which equals a total of 1600 shake flasks. Of course 55 mg is more than enough to perform structure analysis and several activity tests, but even for 5 mg, around 150 flasks (18.75 l) would have been needed. Another more recent example would be the newly characterised bonsecamin [65]. Unlike the previous mentioned strain, it was a silent gene cluster and was activated by heterologous expression, since it was "silent" in the parental strain. But similar to the above mentioned pamamycins the yield was with 0.07 mg/l rather low. To obtain enough substance for structure elucidation and activity testing's 10 l were cultivated in 100 flasks (500 ml) containing 100 ml medium each. Handling 100 or even more flask can be quite work-intensive, time consuming as well as prone to contaminations and expensive. Therefore, one possibility to reduce the work load and make the whole process more cost efficient is the use of fermentation systems. Different sizes are available from micro to bigger laboratory scale bioreactors, enabling a cultivation volume ranging from 0.2 ml [66] up to 50 l [67]. First of all big bioreactors allow a great reduction in labour, as far as only one system needs to be monitored and maintained. Another benefit is that unlike in shake flask fermentations, a lot of parameters can be measured and changed: pH, temperature, dissolved oxygen level, nutrition and cell density. Compared to a shake flask the growth of a culture can be easier monitored and changed as well, by keeping the pH at a constant level, feeding nutrition once it is depleted or changed dissolved oxygen levels to increase NP formation [68]. There are systems on the market which enable the measurement of conditions (pH, dissolved oxygen, cell density) during shake flask fermentations, but the possibility to change culture conditions in a bioreactor remains unmatched [69, 70]. The possibility to change and optimise the conditions during the fermentation is often reflected in a higher yield. Singh et al. 2017 managed to increase the yield of actinomycin D to 150 % in a fermenter compared to a flask, by optimisation of air flow and stirring rate [71]. Another

example for higher yields in fermenters is the production of the glycoprotein GP-1. Zhou et al. 2017 optimised the medium and compared the production in a flask to the production in a fermenter [72]. The flask had a yield of 0.601 mg/l whereas the fermenter had a yield of 2.54 mg/l, representing a 4.2 fold increase. However, the optimisation process of medium and basic cultivation conditions, especially for not characterised strains, is more time consuming and labour intense in big fermenters compared to smaller flasks. Shake flasks are already established since a long time and compared to bioreactor systems rather cheap and available for most laboratories. Accordingly, medium optimisation and characterisation is performed commonly in flasks, just after certain conditions are optimised the cultivation is transferred to a fermenter. However, the upscaling from a flask can be difficult since only few values from the flasks are known: temperature, rotation speed, medium composition, start and end pH value. Since for the structure elucidation and first characterisation of new NPs only small amounts are needed, the cultivation in shake flasks remains often the method of choice, as it is easily accessible for each lab and easier to screen for optimised media. To address the problem of small scale fermentations with fermenters new solutions have been found. Those systems are automated and enable the cultivation of 48 or even more cultures in parallel. Hortsch et al. 2011 compared the production of nikkomycin Z by the producer strain *S. tendae* with different cultivation systems [73]. They compared a small and a big scale of a shaken system and a stirred system with each other. The small scale fermenter contained 10 ml culture broth and was compared to a 1 l fermenter. The microtiter plate was filled with 1.5 ml medium compared to 50 ml in a flask. As a result it was shown that for the stirred fermentation the production yield was in both cases up to 300 mg/l. Unlike this, the production in flasks was with the same optimised medium lower and even in between the shake fermentations not comparable. The low volume fermentation only yielded 100 mg/l whereas the shake flask yielded 180 mg/l. *Streptomyces* tend to form agglomerated mycelia when grown in submerged cultures [74]. Therefore, the cells in the centre of the agglomerate are getting less oxygen and nutrition, which can result in less or not easily to predict production profiles. One possibility to counter this is the use of disruptive elements like glass beads during the fermentation, to keep the cells better dispersed resulting in a higher reproducibility [57, 75]. With the possibility for micro scale fermentations and higher comparability the use of bioreactors seems to be more promising and can also be used for medium optimisation in the future.

Medium optimisation for yield improvements

Apart from simple upscaling of the volume a powerful and essential tool for NP overproduction is the medium optimisation. Most commonly, medium optimisation is performed when a NP has been discovered and first analysis showed promising activities. A good example for the importance of medium optimisation is the study of Rateb et al. 2014 [76]. They used different carbon and nitrogen sources, as well as metal ions compositions, to screen for improved production of tirandamycin B. Their study shows the huge impact of optimised medium on the production. As a result they have increased production yield of tirandamycin B from 2.4 mg/l up to 22 mg/l by simply using different media. In addition they managed to obtain three new related compounds in enough quantities to characterise these. The method they used is referred to as one-variable-at-a-time as far as for one medium only one parameter is systematically changed to identify the effect. Nowadays bioinformatical and mathematical methods can help to optimise media faster [71]. One already commonly applied mathematical approach is the Response Surface Methodology (RSM) [77] and another one Design-of-Experiments (DoE) [78]. For example Gao et al. 2009 used this approach to increase the avermectin B1a production about 145 % [79]. With media optimisation it is not only possible to find the best medium composition but also to screen for external factors e.g.: trace elements, which could increase the production. Similar like in the HITES screening for cluster activation this can be performed to increase the yield of known compounds [80]. As it was shown by Kawai et al. 2007 supplementation of medium with low concentrations of scandium (10 – 100 µM) resulted in an up to 20 fold increased production of NPs actinorhodin and antimycins [81]. They postulated the activation of an alarmone cascade which is believed to activate the secondary metabolism. In combination with automated systems and small scale fermentation, this can be a crucial tool to improve yields, to broaden the analysis possibilities of those compounds, as well as make it more interesting for companies to conduct activity studies. Another part of medium optimisation would be the feeding of precursors. If the compound of interest belongs to a group of compounds with already known biosynthesis the production can be increased by choosing a substrate, which result in increased production of precursors or simply feeding the needed precursors. In combination with precursor feeding Jiang et al. 2013 managed to increase the yield of gougerotin to 250 % [82].

Genetic engineering for yield improvements

In most cases the production can be tuned up to several mg/l by the use of medium optimisation, but due to the strict regulation as well as possible toxicity of the produced NPs, this method has its limits [72, 76]. To overcome those limits and obtain a yield of several g/l the strain needs to be optimised as well [79]. The easiest method to use is a random mutagenesis. As far as no prior genetic knowledge of the strain or even the BGC is necessary to optimise the production with this method, it has been used prior to the availability of cheap sequencing. The study of Siddique et al. 2014 compares random mutagenesis of *Streptomyces avermitilis* 41445 by UV radiation and chemical agents and screened for overproduction of avermectin B1b [83]. After screening via HPLC they managed to obtain in both cases strains with a production above 200 mg/l, compared to a starting production of 17 mg/l. This represents a 12 fold increased production. The strain has been characterised before and medium optimisation was performed starting from 7 mg/l and yielding 500 mg/l in the optimised medium [84]. The starting production of avermectin B1b was 17 mg/l and the used medium was not optimised. The maximum yield of 250 mg/l was already half of the yield of the wildtype strain in optimised medium and a further medium optimisation could increase the possible yield into the g/l scale. Additionally, the strains can possibly be further increased by iterative use of UV radiation [85]. Random mutagenesis is a very powerful tool, especially if you consider that hundreds of genes can influence the expression level of the product. Therefore, it still remains quite commonly used in industry [86]. Despite that random mutagenesis has the disadvantage that during single or iterative mutation steps also detrimental mutations could be accumulated. To overcome those bottlenecks it is important to only introduce one mutation at a time and characterise the effect. For this purpose rational mutagenesis, studies of regulation within bacteria as well as other approaches have been used. However, identifying unknown regulation networks is quite tedious and time consuming, even with optimised methods like the transposon mutagenesis [87]. Once databases of certain mutations and their effect have been created possible targets for the production optimisation can be identified faster using bioinformatic tools, reducing the workload. Another example for the use of the transposon mutagenesis is the overexpression by introducing strong promoters in random positions in the genome. Mutants obtained in the experiment of Horbal et al. 2013 showed a 7 fold increased production of landomycin E compared to the wildtype [88]. With this approach also the localisation of the transposon insertion into the genome is possible. Subcloning of the transposon is also quite easy, allowing a fast forward determination of the

disrupted or upregulated genes. In this case it led to an upregulation of regulatory genes similar to TetR and LuxR family. With methods like this it is possible to perform random mutagenesis and to identify important regulator genes.

One more possibility to improve the transcription of the cluster and therefore the production of the NP is the introduction of synthetic promoters. Since the biosynthesis of NP is tightly regulated as well as the expression of genes match well to each other, the introduction of promoters can result in higher production [89]. However, it cannot be anticipated if the selected promoter of certain strength will lead to the desired overproduction and therefore the introduction of constructs with known strength is rather time consuming. To evade the labour intense work of cloning each promoter of a library and screening for the effect, another random rational approach has been developed. With this approach a library with random promoters of different strengths is incorporated into the cluster. Horbal et al. 2018 first proved the potential of this method by overproducing bottromycin up to 50 fold and identified 7 new derivatives [89].

Overproduction can also be achieved by focusing on the translation step, where the information encoded in RNA is converted into enzymes, responsible for the catalysis of reactions. One possibility to increase the translation is the optimisation of ribosomal binding sites (RBS). Luo et al. 2017 screened for optimised RBS and obtained up to 600 % increased production of the reporter protein [90]. A promising tool for in vivo screening of suitable RBS of *Streptomyces* also has been described [90, 91]. However, RBS are postulated to interact as well with promoters and therefore influence indirectly transcription as well and vice versa. This makes it difficult to trace back if a changed transcription or translation is responsible for the observed effect and some insulator elements have been investigated to reduce these effects [56]. Another promising method is the so called ribosome engineering [92]. Certain random mutations, enhancing production yields, were analysed and sequencing revealed their location in the genes coding for the ribosome. This method has been transferred to different strains und successfully used to overexpress different NPs [93]. By the use of the above mentioned methods the strict regulatory network can be bypassed and higher yields obtained. Another simple method to bypass the native strict regulation network is the heterologous expression. This tool is widely used for activation of “silent” BGCs but it can also be a powerful tool for the overexpression of NPs. Komatsu et al. 2010 managed to obtain about 300 % increased streptomycin production compared to the wildtype (WT) strain [94]. Many different methods

for improved production have been developed using a versatile spectrum of molecular biological techniques and mechanisms. An interesting approach is the introduction of multiple copies of the desired BGC. Murakami et al. 2011 managed to obtain a 20 fold increased production of actinorhodin compared to the wildtype strain with their developed method [95]. However, the generated modifications are quite often not that stable in the strains, as far as strains desire to be the "fittest" in a biological sense. Quite commonly they abolish or decrease the high metabolically burdensome production of NPs. Therefore, selection markers are often used in industry increasing the costs for the production, making marker free and stable expression systems to be sought after [96]. Many other possible targets (e.g.: Protein or mRNA stability, resistance genes,…) which influence the overproduction and methods to use them have been described or are currently developed but still need to find their application in industry [91].

Taking all those findings and developments into account, it can be said that the overproduction of especially promising compounds remains a bottleneck for further characterisation and clinical studies. As far as *Streptomyces* are versatile and their regulation is quite complex there will most likely be not only one solution for everything and a broad spectrum of methods to obtain overproduction is required and available. With growing knowledge of regulatory networks, biosynthesis and availability of automated systems, the overproduction of NPs will shift from random to targeted approaches, making use of the described methods. Thereby, making the characterisation of neglected compounds more cost-efficient and will reveal their potential applications.

1.4 Outline of this work

As it has been described above there are two main problems within the field of NPs. Firstly, new NP scaffolds need to be discovered and secondly, enough quantity of compounds need to be provided for further characterisations.

Chapter 2.1 and 2.2 focus on the problem of the identification of new target scaffolds. Primarily the work on the cyclofaulknamycin shows the hidden potential of even one of the most explored and well understood model organisms in this field (*S. albus* J1074). With the help of targeted genome mining the cyclic form of this compound could be proven as well as the cluster further characterised. In chapter 2.2 a unique type II polyketide from *Streptomyces* sp. IB2014/011-12 is described. This chapter proves that the strain isolation is still a good possibility to find interesting new scaffolds. Within this project the polyketide

aorimycin with a unique 9-membered macrocycle as well as D-forosamine moiety gets firstly isolated and described. The substance is not only isolated from the wildtype strain but heterologously as well. With the help of the simplified genetic accessibility of the heterologous host the biosynthesis was further analysed and further intermediates characterised.

The third part of this work is focussing on the yield improvement of the NPs pamamycins. Pamamycins have been isolated and characterised before. Thereby, they have proven to show very interesting activities against several human pathogenic bacteria, like *Mycobacterium smegmatis*, *Streptococcus pneumoniae* and *Staphylococcus aureus*. Further studies have been abolished most likely due to too low yields. In this work (chapter 2.3), the pamamycin production was shifted towards not characterised derivatives and further activity studies were performed. The derivatives were obtained by molecular biological introduction of random rational designed promoters. This example represents the importance on working with "neglected" compounds, to characterise them further and find possible applications for them.

1.5 References

1. Church, D.D., et al., *Essential Amino Acids and Protein Synthesis: Insights into Maximizing the Muscle and Whole-Body Response to Feeding.* Nutrients, 2020. **12**(12): p. 3717.
2. Karau, A. and I. Grayson, *Amino acids in human and animal nutrition.* Biotechnology of food and feed additives, 2014: p. 189-228.
3. Lee, J.-H. and V.F. Wendisch, *Production of amino acids–genetic and metabolic engineering approaches.* Bioresource technology, 2017. **245**: p. 1575-1587.
4. O'Connor, S.E., *Engineering of secondary metabolism.* Annual review of genetics, 2015. **49**: p. 71-94.
5. Yocum, R.R., J.R. Rasmussen, and J.L. Strominger, *The mechanism of action of penicillin. Penicillin acylates the active site of Bacillus stearothermophilus D-alanine carboxypeptidase.* Journal of Biological Chemistry, 1980. **255**(9): p. 3977-3986.
6. Ligon, B.L. *Penicillin: its discovery and early development*. in *Seminars in pediatric infectious diseases*. 2004. Elsevier.
7. Firn, R.D. and C.G. Jones, *Natural products–a simple model to explain chemical diversity.* Natural product reports, 2003. **20**(4): p. 382-391.
8. Ganesan, A., *The impact of natural products upon modern drug discovery.* Current opinion in chemical biology, 2008. **12**(3): p. 306-317.
9. Lahlou, M., *The success of natural products in drug discovery.* 2013.
10. Müller-Kuhrt, L., *Putting nature back into drug discovery.* Nature biotechnology, 2003. **21**(6): p. 602-602.
11. Sun, Y., et al., *'Streptomyces nanchangensis', a producer of the insecticidal polyether antibiotic nanchangmycin and the antiparasitic macrolide meilingmycin, contains multiple polyketide gene clusters.* Microbiology, 2002. **148**(2): p. 361-371.

12. Law, J.W.-F., et al., *Anticancer drug discovery from microbial sources: the unique mangrove streptomycetes*. Molecules, 2020. **25**(22): p. 5365.
13. Li, F., et al., *Anti-influenza A viral butenolide from Streptomyces sp. Smu03 inhabiting the intestine of Elephas maximus.* Viruses, 2018. **10**(7): p. 356.
14. Takahashi, K., et al., *Ushikulide C, a New Immunosuppressant from Streptomyces sp. IUK-102.* Natural Product Communications, 2013. **8**(6): p. 1934578X1300800634.
15. Katz, L. and R.H. Baltz, *Natural product discovery: past, present, and future.* Journal of Industrial Microbiology and Biotechnology, 2016. **43**(2-3): p. 155-176.
16. Dahal, R.H., et al., *Development of Multifunctional Cosmetic Cream Using Bioactive Materials from Streptomyces sp. T65 with Synthesized Mesoporous Silica Particles SBA-15.* Antioxidants, 2020. **9**(4): p. 278.
17. Drasar, P.B. and V.A. Khripach, *Growing importance of natural products research*. 2019, MDPI. p. 6.
18. Bassett, E.J., et al., *Tetracycline-labeled human bone from ancient Sudanese Nubia (AD 350).* Science, 1980. **209**(4464): p. 1532-1534.
19. Desborough, M.J. and D.M. Keeling, *The aspirin story–from willow to wonder drug.* British journal of haematology, 2017. **177**(5): p. 674-683.
20. Vlachojannis, J., F. Magora, and S. Chrubasik, *Willow species and aspirin: different mechanism of actions.* Phytotherapy Research, 2011. **25**(7): p. 1102-1104.
21. Frost, W.D., *The antagonism exhibited by certain saprophytic bacteria against the Bacillus typhosus Gaffky*. 1904: University of Chicago Press.
22. Waksman, S.A., A. Schatz, and D.M. Reynolds, *Production of antibiotic substances by actinomycetes.* Annals of the New York Academy of Sciences, 2010. **1213**(1): p. 112-124.
23. Hutchings, M.I., A.W. Truman, and B. Wilkinson, *Antibiotics: past, present and future.* Current opinion in microbiology, 2019. **51**: p. 72-80.
24. Aminov, R.I., *A brief history of the antibiotic era: lessons learned and challenges for the future.* Frontiers in microbiology, 2010. **1**: p. 134.
25. Adedeji, W., *The treasure called antibiotics.* Annals of Ibadan postgraduate medicine, 2016. **14**(2): p. 56.
26. Davies, J. and D. Davies, *Origins and evolution of antibiotic resistance.* Microbiology and molecular biology reviews, 2010. **74**(3): p. 417-433.
27. Heuer, H. and K. Smalla, *Horizontal gene transfer between bacteria.* Environmental biosafety research, 2007. **6**(1-2): p. 3-13.
28. Lade, H. and J.-S. Kim, *Bacterial Targets of Antibiotics in Methicillin-Resistant Staphylococcus aureus.* Antibiotics, 2021. **10**(4): p. 398.
29. Santajit, S. and N. Indrawattana, *Mechanisms of antimicrobial resistance in ESKAPE pathogens.* BioMed research international, 2016. **2016**.
30. Wang, C.-H., et al., *Defeating antibiotic-resistant bacteria: exploring alternative therapies for a post-antibiotic era.* International journal of molecular sciences, 2020. **21**(3): p. 1061.
31. Gupta, K., A. Gupta, and D. Shrivastava, *The last resort antibiotics: carbapenems.* Int. J. Adv. Res, 2017. **5**(4): p. 1410-1413.
32. Sharland, M., et al., *21st WHO Expert Committee on Selection and Use of Essential Medicines.* Classifying antibiotics in the WHO Essential Medicines List for optimal use-be AWaRe. Lancet Infect Dis, 2018. **18**(1): p. 18-20.
33. Ventola, C.L., *The antibiotic resistance crisis: part 1: causes and threats.* Pharmacy and therapeutics, 2015. **40**(4): p. 277.

34. Schneider, Y.K., *Bacterial Natural Product Drug Discovery for New Antibiotics: Strategies for Tackling the Problem of Antibiotic Resistance by Efficient Bioprospecting.* Antibiotics, 2021. **10**(7): p. 842.
35. Čihák, M., et al., *Secondary metabolites produced during the germination of Streptomyces coelicolor.* Frontiers in Microbiology, 2017. **8**: p. 2495.
36. Liu, Z., et al., *Recent advances in silent gene cluster activation in Streptomyces.* Frontiers in Bioengineering and Biotechnology, 2021. **9**.
37. Atanasov, A.G., et al., *Natural products in drug discovery: Advances and opportunities.* Nature Reviews Drug Discovery, 2021. **20**(3): p. 200-216.
38. Nguyen, C.T., et al., *Recent advances in strategies for activation and discovery/characterization of cryptic biosynthetic gene clusters in Streptomyces.* Microorganisms, 2020. **8**(4): p. 616.
39. Kresge, N., R.D. Simoni, and R.L. Hill, *Selman Waksman: the father of antibiotics.* Journal of Biological Chemistry, 2004. **279**(48): p. e7-e7.
40. Singh, V., et al., *Isolation, screening, and identification of novel isolates of actinomycetes from India for antimicrobial applications.* Frontiers in microbiology, 2016. **7**: p. 1921.
41. Takahashi, Y. and S. Omura, *Isolation of new actinomycete strains for the screening of new bioactive compounds.* The Journal of general and applied microbiology, 2003. **49**(3): p. 141-154.
42. Blin, K., et al., *antiSMASH 6.0: improving cluster detection and comparison capabilities.* Nucleic acids research, 2021: p. 1.
43. Kautsar, S.A., et al., *MIBiG 2.0: a repository for biosynthetic gene clusters of known function.* Nucleic acids research, 2020. **48**(D1): p. D454-D458.
44. Skinnider, M.A., et al., *Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences.* Nature communications, 2020. **11**(1): p. 1-9.
45. Bode, H.B., et al., *Big effects from small changes: possible ways to explore nature's chemical diversity.* ChemBioChem, 2002. **3**(7): p. 619-627.
46. Bodor, A., et al., *Challenges of unculturable bacteria: environmental perspectives.* Reviews in Environmental Science and Bio/Technology, 2020. **19**(1): p. 1-22.
47. Pedrós-Alió, C. and S. Manrubia, *The vast unknown microbial biosphere.* Proceedings of the National Academy of Sciences, 2016. **113**(24): p. 6585-6587.
48. Rodríguez Estévez, M., et al., *Heterologous expression of the nybomycin gene cluster from the marine strain Streptomyces albus subsp. chlorinus NRRL B-24108.* Marine drugs, 2018. **16**(11): p. 435.
49. Rutledge, P.J. and G.L. Challis, *Discovery of microbial natural products by activation of silent biosynthetic gene clusters.* Nature reviews microbiology, 2015. **13**(8): p. 509-523.
50. Bezie, Y., et al., *The potential applications of site-directed mutagenesis for crop improvement: A review.* Journal of Crop Science and Biotechnology, 2021. **24**(3): p. 229-244.
51. Liu, G., et al., *Molecular regulation of antibiotic biosynthesis in Streptomyces.* Microbiology and molecular biology reviews, 2013. **77**(1): p. 112-143.
52. Krause, J., et al., *Disclosing the potential of the SARP-type regulator PapR2 for the activation of antibiotic gene clusters in streptomycetes.* Frontiers in microbiology, 2020. **11**: p. 225.
53. Wei, J., L. He, and G. Niu, *Regulation of antibiotic biosynthesis in actinomycetes: perspectives and challenges.* Synthetic and systems biotechnology, 2018. **3**(4): p. 229-235.

54. Flett, F., V. Mersinias, and C.P. Smith, *High efficiency intergeneric conjugal transfer of plasmid DNA from Escherichia coli to methyl DNA-restricting streptomycetes.* FEMS microbiology letters, 1997. **155**(2): p. 223-229.
55. Pigac, J. and H. Schrempf, *A simple and rapid method of transformation of Streptomyces rimosus R6 and other streptomycetes by electroporation.* Applied and Environmental Microbiology, 1995. **61**(1): p. 352-356.
56. Bai, C., et al., *Exploiting a precise design of universal synthetic modular regulatory elements to unlock the microbial natural products in Streptomyces.* Proceedings of the National Academy of Sciences, 2015. **112**(39): p. 12181-12186.
57. Sohoni, S.V., et al., *Synthetic promoter library for modulation of actinorhodin production in Streptomyces coelicolor A3 (2).* PloS one, 2014. **9**(6): p. e99701.
58. Ji, C.-H., J.-P. Kim, and H.-S. Kang, *Library of synthetic Streptomyces regulatory sequences for use in promoter engineering of natural product biosynthetic gene clusters.* ACS synthetic biology, 2018. **7**(8): p. 1946-1955.
59. Ahmed, Y., et al., *Engineering of Streptomyces lividans for heterologous expression of secondary metabolite gene clusters.* Microbial cell factories, 2020. **19**(1): p. 1-16.
60. Myronovskyi, M., et al., *Generation of a cluster-free Streptomyces albus chassis strains for improved heterologous expression of secondary metabolite clusters.* Metabolic engineering, 2018. **49**: p. 316-324.
61. Groß, S., et al., *In vivo and in vitro reconstitution of unique key steps in cystobactamid antibiotic biosynthesis.* Nature communications, 2021. **12**(1): p. 1-15.
62. Tan, G.-Y. and T. Liu, *Rational synthetic pathway refactoring of natural products biosynthesis in actinobacteria.* Metabolic engineering, 2017. **39**: p. 228-236.
63. Clarkson, C., et al., *Discovering New Natural Products Directly from Crude Extracts by HPLC-SPE-NMR: Chinane Diterpenes in Harpagophytum p rocumbens.* Journal of Natural Products, 2006. **69**(4): p. 527-530.
64. Kondo, S., et al., *Isolation, physico-chemical properties and biological activity of pamamycin-607, an aerial mycelium-inducing substance from Streptomyces alboniger.* The Journal of antibiotics, 1988. **41**(9): p. 1196-1204.
65. Lasch, C., et al., *Bonsecamin: A New Cyclic Pentapeptide Discovered through Heterologous Expression of a Cryptic Gene Cluster.* Microorganisms, 2021. **9**(8): p. 1640.
66. Kensy, F., C. Engelbrecht, and J. Büchs, *Scale-up from microtiter plate to laboratory fermenter: evaluation by online monitoring techniques of growth and protein expression in Escherichia coli and Hansenula polymorpha fermentations.* Microbial Cell Factories, 2009. **8**(1): p. 1-15.
67. Olmos, E., et al., *Effects of bioreactor hydrodynamics on the physiology of Streptomyces.* Bioprocess and Biosystems Engineering, 2013. **36**(3): p. 259-272.
68. Yegneswaran, P., M. Gray, and B. Thompson, *Effect of dissolved oxygen control on growth and antibiotic production in Streptomyces clavuligerus fermentations.* Biotechnology Progress, 1991. **7**(3): p. 246-250.
69. Flitsch, D., et al., *Easy to use and reliable technique for online dissolved oxygen tension measurement in shake flasks using infrared fluorescent oxygen-sensitive nanoparticles.* Microbial cell factories, 2016. **15**(1): p. 1-11.
70. Schneider, K., et al., *Optical device for parallel online measurement of dissolved oxygen and pH in shake flask cultures.* Bioprocess and biosystems engineering, 2010. **33**(5): p. 541-547.
71. Singh, V., et al., *Strategies for fermentation medium optimization: an in-depth review.* Frontiers in microbiology, 2017. **7**: p. 2087.

72. Zhou, Y., et al., *Optimization of medium compositions to improve a novel glycoprotein production by Streptomyces kanasenisi ZX01.* AMB Express, 2017. **7**(1): p. 1-9.
73. Hortsch, R., H. Krispin, and D. Weuster-Botz, *Process performance of parallel bioreactors for batch cultivation of Streptomyces tendae.* Bioprocess and biosystems engineering, 2011. **34**(3): p. 297-304.
74. van Dissel, D., D. Claessen, and G.P. van Wezel, *Morphogenesis of Streptomyces in submerged cultures.* Advances in applied microbiology, 2014. **89**: p. 1-45.
75. Yepes-García, J., et al., *Morphological differentiation of streptomyces clavuligerus exposed to diverse environmental conditions and its relationship with clavulanic acid biosynthesis.* Processes, 2020. **8**(9): p. 1038.
76. Rateb, M.E., et al., *Medium optimization of Streptomyces sp. 17944 for tirandamycin B production and isolation and structural elucidation of tirandamycins H, I and J.* The Journal of antibiotics, 2014. **67**(1): p. 127-132.
77. Ibrahim, H.M. and E.E. Elkhidir, *Response surface method as an efficient tool for medium optimisation.* 2011.
78. Mandenius, C.-F., *Design-of-experiments for development and optimization of bioreactor media.* Edited by, 2016.
79. Gao, H., et al., *Medium optimization for the production of avermectin B1a by Streptomyces avermitilis 14-12A using response surface methodology.* Bioresource Technology, 2009. **100**(17): p. 4012-4016.
80. Tanaka, Y., T. Hosaka, and K. Ochi, *Rare earth elements activate the secondary metabolite–biosynthetic gene clusters in Streptomyces coelicolor A3 (2).* The Journal of antibiotics, 2010. **63**(8): p. 477-481.
81. Kawai, K., et al., *The rare earth, scandium, causes antibiotic overproduction in Streptomyces spp.* FEMS microbiology letters, 2007. **274**(2): p. 311-315.
82. Jiang, L., et al., *Combined gene cluster engineering and precursor feeding to improve gougerotin production in Streptomyces graminearus.* Applied microbiology and biotechnology, 2013. **97**(24): p. 10469-10477.
83. Siddique, S., et al., *Production and screening of high yield avermectin B1b mutant of Streptomyces avermitilis 41445 through mutagenesis.* Jundishapur journal of microbiology, 2014. **7**(2).
84. Burg, R.W., et al., *Avermectins, new family of potent anthelmintic agents: producing organism and fermentation.* Antimicrobial agents and Chemotherapy, 1979. **15**(3): p. 361-367.
85. Wang, Y., et al., *Iteratively improving natamycin production in Streptomyces gilvosporeus by a large operon-reporter based strategy.* Metabolic engineering, 2016. **38**: p. 418-426.
86. Parekh, S., V. Vinci, and R. Strobel, *Improvement of microbial strains and fermentation processes.* Applied microbiology and biotechnology, 2000. **54**(3): p. 287-301.
87. Xu, Z., et al., *Large-scale transposition mutagenesis of Streptomyces coelicolor identifies hundreds of genes influencing antibiotic biosynthesis.* Applied and environmental microbiology, 2017. **83**(6): p. e02889-16.
88. Horbal, L., et al., *A transposon-based strategy to identify the regulatory gene network responsible for landomycin E biosynthesis.* FEMS microbiology letters, 2013. **342**(2): p. 138-146.
89. Horbal, L., et al., *Secondary metabolites overproduction through transcriptional gene cluster refactoring.* Metabolic engineering, 2018. **49**: p. 299-315.
90. Luo, H.-D., et al., *Design of ribosome binding sites in Streptomyces coelicolor.* Current Proteomics, 2017. **14**(4): p. 287-292.

91. Horbal, L., T. Siegl, and A. Luzhetskyy, *A set of synthetic versatile genetic control elements for the efficient expression of genes in Actinobacteria.* Scientific reports, 2018. **8**(1): p. 1-13.
92. Tanaka, Y., et al., *Antibiotic overproduction by rpsL and rsmG mutants of various actinomycetes.* Applied and environmental microbiology, 2009. **75**(14): p. 4919-4922.
93. Ochi, K., *From microbial differentiation to ribosome engineering.* Bioscience, biotechnology, and biochemistry, 2007. **71**(6): p. 1373-1386.
94. Komatsu, M., et al., *Genome-minimized Streptomyces host for the heterologous expression of secondary metabolism.* Proceedings of the National Academy of Sciences, 2010. **107**(6): p. 2646-2651.
95. Murakami, T., et al., *A system for the targeted amplification of bacterial gene clusters multiplies antibiotic yield in Streptomyces coelicolor.* Proceedings of the National Academy of Sciences, 2011. **108**(38): p. 16020-16025.
96. Sevillano, L., M. Díaz, and R.I. Santamaría, *Development of an antibiotic marker-free platform for heterologous protein production in Streptomyces.* Microbial Cell Factories, 2017. **16**(1): p. 1-13.

2. Publications and articles to be submitted

2.1 Cyclofaulknamycin with the rare amino acid D-capreomycidine isolated from a well-characterized *Streptomyces albus* strain

Horbal L., Stierhof M., Palusczak A., Eckert N., Zapp J., Luzhetskyy A.

Microorganisms 2021, 9(8), 1609

DOI: 10.3390/microorganisms9081609

Published online 28.07.2021

2.1.1 Abstract

Targeted genome mining is an efficient method of biosynthetic gene cluster prioritization within constantly growing genome databases. Using two capreomycidine 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-capreomycidine 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-capreomycidine; cyclofaulknamycin; cyclopeptide; *Streptomyces*

2.1.2 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 capreomycidine, 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 capreomycidine 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 [7, 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 capreomycidine 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 faulknamycin compound and the recently described linear faulknamycin, respectively [20]. This is the first identified cyclic peptide compound with the rare D-capreomycidine 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-capreomycidine moiety. Herein, we report

the structure of this compound and describe its biosynthetic gene cluster, biosynthesis and bioactivity profile.

2.1.3 Material and Methods

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 μg mL^{-1} ampicillin, 50 μg mL^{-1} kanamycin, 50 or 120 μg mL^{-1} hygromycin, and 50 μg 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.

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
S. albus J1074	Isoleucine and valine auxotrophic derivative of *S. albus* G (DSM 40313) lacking *Sal*I-restriction activity	Salas J., Oviedo, Spain
S. albus del9	*S. albus* G (DSM 40313) lacking 9 secondary metabolite gene clusters	This work
S. albus del10	*S. albus* G (DSM 40313) lacking 10 secondary metabolite gene clusters	This work
S. albus del9 delXNR_1347	*S. albus* del9 strain carrying the deletion of the *XNR_1347* gene	This work
E. coli ET12567 (pUB307)	conjugative transfer of DNA	Kieser et al., 2000
E. coli 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 *S. albus* J1074 chromosome	Myronovskyi et al., 2018
2K9/2delXNR_1347	2K9/2 derivative containing deleted XNR_1347 gene	This work

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_2O, 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).

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 µg mL^{-1}) and glucuronidase activity. The abovementioned cosmid did not contain the origin of replication for *Streptomycetes*, all the obtained exconjugants were the result of a first crossover event. For the generation of the *S. albus* ΔXNR_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).

Table 2. Primers used in this work.

Primers	Sequence 5'-3'	Purpose
XNR_1347RedFor	TGGAGCGGGAGCGCGAGCAGGACGAGCGGCGCGAGCGGG CGTGAGCCGCCGGGTGCCCGCGCCCGGTCATATCCATCCTT TTTCGCACGATATAC	XNR_1347 gene deletion
XNR_1347RedRev	CGACGCCTGGCTGCCGGGGCTCCCGCGCGGGTCCGTAGAA TCGGCCCCACCATGGCCTACCTCGACCACCAGATTACGCG CAGAAAAAAAGGATCTC	
XNR_1347F	AAGAAGCAGCTGGAGCGGGAG	XNR_1347 gene deletion check
XNR_1347R	ACCATGGCCTACCTCGACCAC	

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 µL of the dissolved sample was separated in a Dionex Ultimate 3000 RSLC system using a BEH C18, 100 × 2.1 mm, 1.7 µm dp column (Waters Corporation, Milford, MA, USA). The separation of a 1 µl 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.

Isolation and Purification of Cyclic and Linear Isofaulknamycin

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 IsoleraTM 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 µm, 21 × 250 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 SynergiTM 4 µm Fusion-RP 80 Å 250 × 10 (Phenomenex, Torrance, CA, USA) column using a linear gradient of [A] H_2O + 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).

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 (δ) were reported in parts per million (ppm) relative to TMS. As a solvent, deuterated DMSO-d6 (δH 2.50 ppm., δC 39.51 ppm.) from Deutero (Kastellaun, Germany) was used. Edited-HSQC, HMBC, 1H-1H COSY and ROESY spectra were recorded using the standard pulse programs from Bruker TOPSPIN v.3.6 software.

Marfey's Method

Iso-faulknamycin, chymostatin and capreomycin were hydrolyzed in 100 µL 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 µL of each sample was transferred to a 1.5 mL Eppendorf tube. Next, 20 µL of 1 N $NaHCO_3$ and 20 µL 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 µL ACN, and 1 µL 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.

Genome Mining and Bioinformatic Analysis

Genome screening was performed using the BLAST online tool (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/ (accessed on 15 June 2021)). Gene cluster analysis was performed using the antiSMASH online tool (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).

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 104 CFU mL^{-1}. Serial dilutions of cyclofaulknamycin (0.03 to 64 $\mu g \cdot mL^{-1}$) 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 (ΔacrB), *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.

2.1.4 Results and Discussion

Mining of Actinobacterial Genomes for the Presence of Capreomycidine Biosynthetic Genes

To identify the compounds containing the rare cyclic guanidino-amino acid capreomycidine, 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 capreomycidine biosynthesis [26]. During the preparation of this manuscript, an article describing the new hypothetical pathway for capreomycidine 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).

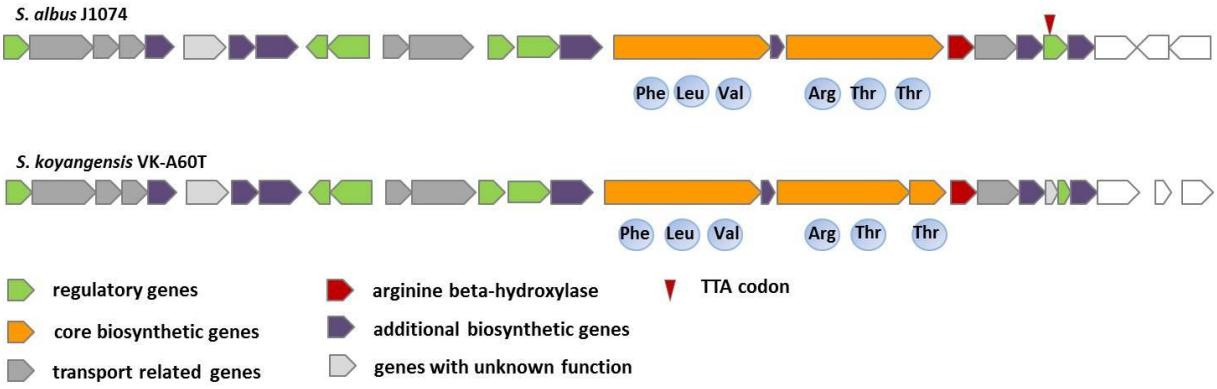

Figure 1. Structure of the *flk* cluster in the *S. albus* strain and *S koyangensis* strains.

Furthermore, the recently published faulknamycin 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 faulknamycin gene cluster. Tryon et al. suggested that the capreomycidine 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 capreomycidine moiety. The chemical products of the flk clusters were not known; however, they were assumed to be identical to faulknamycin based on the in silico prediction of the NRPS domain organization and the genes present in the clusters (Figure 1).

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

multifunctional 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 (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_{Phe}-PCP-C-A_{Leu}-PCP-E-C-A_{Val}-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_{Arg}-PCP-E-C-A_{Thr}-PCP-C-A_{Thr}-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* [8] and encodes Fe(II)/alpha-ketoglutarate-dependent arginine beta-hydroxylase, which is involved in the biosynthesis of the capreomycidine 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 capreomycidine biosynthesis. Another possibility is that the biosynthesis of the

capreomycidine 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 [9].

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_1	D	A	W	T	V	A	A	V	C	K	**Phe**
A_2	D	G	M	L	V	G	A	V	V	K	**Leu**
A_3	D	A	F	W	L	G	G	T	F	K	**Val**
A_4	D	L	A	E	S	G	A	V	D	K	**Arg, Orn, Lys**
A_5	D	F	W	S	V	G	M	V	H	K	**Thr**
A_6	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 [9]. *FlkK* and *flkL* encode ABC transporters, and *flkU* encodes a major facilitator superfamily transporter and might thus be involved in antibiotic resistance.

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 Myronovsyki 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].

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 δC 72.8 (Figure S4) in the side chain. Hence, it was assumed that a hydroxyl group in the β position resulted in the assignment of β 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 α CH proton (δH 4.70) revealed a COSY correlation to a neighboring methine at δH 3.73 (Figure S2). The remaining two degrees of unsaturation indicated an intramolecular ring closure of the guanidinium moiety and the β-position, leading to the assignment of capreomycidine. The sample did not dissolve completely in DMSO-d6, 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 H2N Thr β-phenylserine-Leu-Val-capromycidine-Thr-COOH.

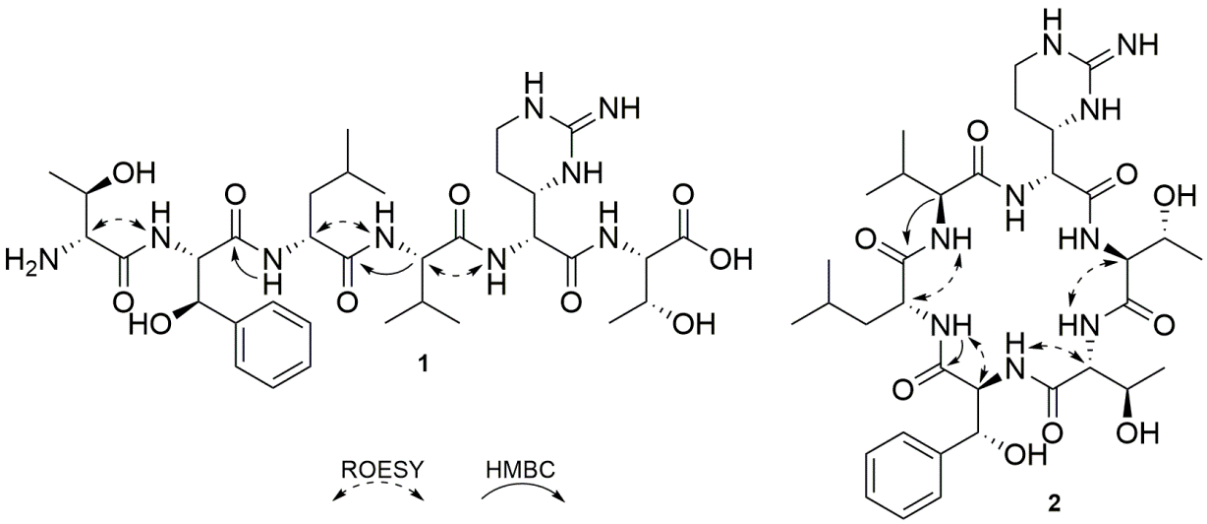

Figure 2. Structures of iso-faulknamycine (1) and cyclo-faulknamycine (2) showing key HMBC and ROESY correlations.

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_2 group of the linear peptide showed a shift of δH 7.88 and an integral value of two protons, while the same signal of the cyclic version showed a low-field shift to δH 8.52 with an integral of only one proton. Furthermore, a ROESY correlation between this NH of the first threonine and the α-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.

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 capreomycidine was determined as described previously by using standards derived from the hydrolysis of capreomycidine and chymostatin and derivatization by L-FDLA and D-FDLA [20] (Figure S13). The amino acid β-phenylserine has been identified by Tryon et al. as the D-*threo*- isomer; therefore, DL-*threo*-β-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-capreomycidine and L-*threo*-β-phenylserine as the constituent amino acids. D-*threo*-β-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-faulknamycin (Figure S16). This finding is in line with the domain organization of the FlkP NRPS described above.

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 capreomycidine 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 2K9/2delXNR_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* Δdel9delXNR_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* Δdel9 and Δdel9delXNR_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 capreomycidine

moiety. Furthermore, this seems to be the main mechanism of capreomycidine moiety formation in *S. albus*.

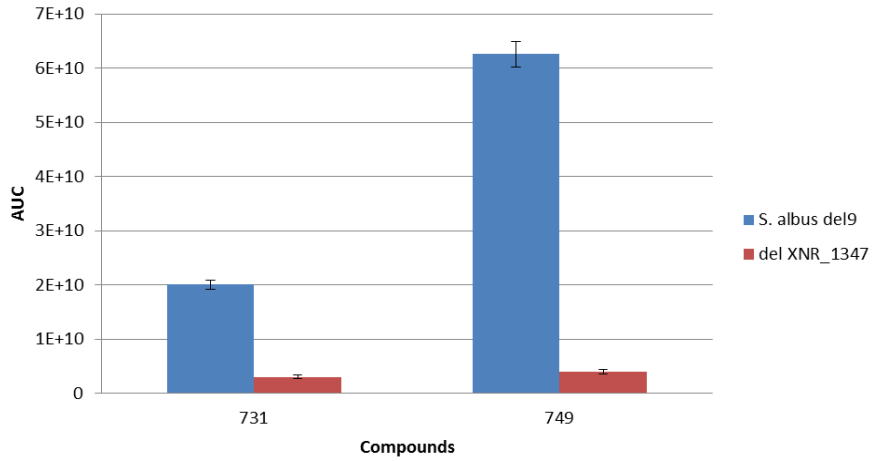

Figure 3. Faulknamycin production in the *S. albus* del9 and delXNR_1347 strains. 731 – cyclofaulknamycin; 749 – isofaulknamycin.

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 capreomycidine moiety, and the last two modules are responsible for the incorporation of two threonines. The biosynthesis of capreomycidine 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 capreomycidine. 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 capreomycidine 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.

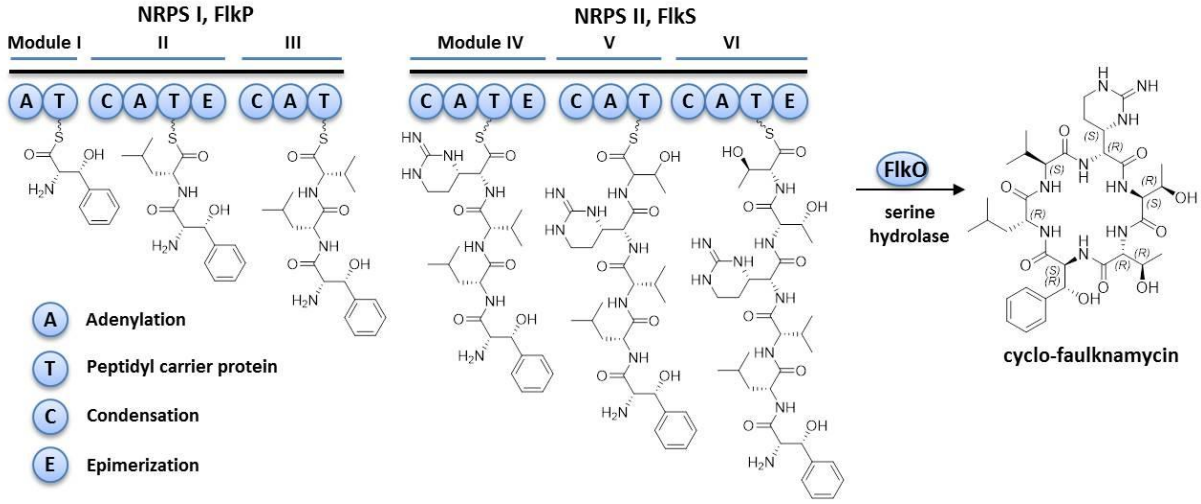

Figure 4. Proposed biosynthesis of cyclofaulknamycin.

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 capreomycidine 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 (ΔacrB), *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 μg·mL^{-1}. However, no activity was detected against the tested strains in the abovementioned range (data not shown).

2.1.5 Conclusion

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 capreomycidine 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-capreomycidine [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-capreomycidine (rather than L-capreomycidine, as commonly found) for the first time. Only linear faulknamycin has previously been described to contain a D-capreomycidine moiety [20]. The detailed analysis of the *flk* cluster allowed us to suggest the biosynthesis mechanisms of cyclofaulknamycin and its capreomycidine 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 capreomycidine moiety of cyclofaulknamycin. However, cyclofaulknamycin biosynthesis was not completely blocked in the delXNR_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 capreomycidine 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.

2.1.6 References

1. Dias, D.A., S. Urban, and U. Roessner, *A historical overview of natural products in drug discovery.* Metabolites, 2012. **2**(2): p. 303-336.
2. Marcone, G.L., et al., *Old and new glycopeptide antibiotics: from product to gene and back in the post-genomic era.* Biotechnology advances, 2018. **36**(2): p. 534-554.
3. Zhang, D., et al., *Antifungal peptides produced by actinomycetes and their biological activities against plant diseases.* The Journal of Antibiotics, 2020. **73**(5): p. 265-282.
4. Hudson, G.A. and D.A. Mitchell, *RiPP antibiotics: biosynthesis and engineering potential.* Current opinion in microbiology, 2018. **45**: p. 61-69.
5. Strieker, M., A. Tanović, and M.A. Marahiel, *Nonribosomal peptide synthetases: structures and dynamics.* Current opinion in structural biology, 2010. **20**(2): p. 234-240.
6. Felnagle, E.A., et al., *Nonribosomal peptide synthetases involved in the production of medically relevant natural products.* Molecular pharmaceutics, 2008. **5**(2): p. 191-211.
7. 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): p. 2325-2342.
8. Thomas, M.G., Y.A. Chan, and S.G. Ozanick, *Deciphering tuberactinomycin biosynthesis: isolation, sequencing, and annotation of the viomycin biosynthetic gene cluster.* Antimicrobial agents and chemotherapy, 2003. **47**(9): p. 2823-2830.
9. Magarvey, N.A., et al., *Biosynthetic pathway for mannopeptimycins, lipoglycopeptide antibiotics active against drug-resistant gram-positive pathogens.* Antimicrobial agents and chemotherapy, 2006. **50**(6): p. 2167-2177.
10. Petersen, P.J., et al., *Comparative in vitro activities of AC98-6446, a novel semisynthetic glycopeptide derivative of the natural product mannopeptimycin α, and other antimicrobial agents against gram-positive clinical isolates.* Antimicrobial agents and chemotherapy, 2004. **48**(3): p. 739-746.
11. Yin, X. and T.M. Zabriskie, *The enduracidin biosynthetic gene cluster from Streptomyces fungicidicus.* Microbiology, 2006. **152**(10): p. 2969-2983.
12. Guo, C., et al., *Chemistry and biology of teixobactin.* Chemistry–A European Journal, 2018. **24**(21): p. 5406-5422.
13. Ling, L.L., et al., *A new antibiotic kills pathogens without detectable resistance.* Nature, 2015. **517**(7535): p. 455-459.
14. Goldstein, E., M.C. Eagle, and M. LaCasse, *In vitro chemotherapeutic combinations against isoniazid-resistant Mycobacterium tuberculosis and Mycobacterium fortuitum.* Applied Microbiology, 1971. **22**(3): p. 329-333.

15. Chan, D.I., E.J. Prenner, and H.J. Vogel, *Tryptophan-and arginine-rich antimicrobial peptides: structures and mechanisms of action.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2006. **1758**(9): p. 1184-1202.
16. Kang, J.H., et al., *Structure–biological activity relationships of 11-residue highly basic peptide segment of bovine lactoferrin.* International journal of peptide and protein research, 1996. **48**(4): p. 357-363.
17. Sikorska, E., et al., *Short arginine-rich lipopeptides: From self-assembly to antimicrobial activity.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2018. **1860**(11): p. 2242-2251.
18. Antonoplis, A., et al., *Vancomycin–arginine conjugate inhibits growth of carbapenem-resistant E. coli and targets cell-wall synthesis.* ACS chemical biology, 2019. **14**(9): p. 2065-2070.
19. Myronovskyi, M., et al., *Generation of a cluster-free Streptomyces albus chassis strains for improved heterologous expression of secondary metabolite clusters.* Metabolic engineering, 2018. **49**: p. 316-324.
20. Tryon, J.H., et al., *Genome mining and metabolomics uncover a rare d-capreomycidine containing natural product and its biosynthetic gene cluster.* ACS chemical biology, 2020. **15**(11): p. 3013-3020.
21. Fu, J., et al., *Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition.* Nucleic acids research, 2008. **36**(17): p. e113-e113.
22. Zhang, Y., et al., *DNA cloning by homologous recombination in Escherichia coli.* Nature biotechnology, 2000. **18**(12): p. 1314-1317.
23. Kieser, T., et al., *Practical streptomyces genetics*. Vol. 291. 2000: John Innes Foundation Norwich.
24. Sambrook, J. and D. Russell, *Molecular cloning: a laboratory manual 3rd edition. 490 Coldspring*. 2001, Harbour Laboratory Press, UK.
25. Blin, K., et al., *antiSMASH 6.0: improving cluster detection and comparison capabilities.* Nucleic acids research, 2021. **49**(W1): p. W29-W35.
26. Yin, X., et al., *Formation of the nonproteinogenic amino acid 2S, 3R-capreomycidine by VioD from the viomycin biosynthesis pathway.* ChemBioChem, 2004. **5**(9): p. 1278-1281.
27. Kohli, R.M. and C.T. Walsh, *Enzymology of acyl chain macrocyclization in natural product biosynthesis.* Chemical Communications, 2003(3): p. 297-307.
28. Baltz, R.H., *Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery.* Journal of Industrial Microbiology and Biotechnology, 2011. **38**(11): p. 1747.
29. Hackl, S. and A. Bechthold, *The gene bldA, a regulator of morphological differentiation and antibiotic production in Streptomyces.* Archiv der Pharmazie, 2015. **348**(7): p. 455-462.
30. Harada, K.-i., et al., *Application of D, L-FDLA derivatization to determination of absolute configuration of constituent amino acids in peptide by advanced Marfey's method.* Tetrahedron letters, 1996. **37**(17): p. 3001-3004.
31. Belknap, K.C., et al., *Genome mining of biosynthetic and chemotherapeutic gene clusters in Streptomyces bacteria.* Scientific reports, 2020. **10**(1): p. 1-9.
32. Rebets, Y., et al., *Actinomycetes biosynthetic potential: how to bridge in silico and in vivo?* Journal of Industrial Microbiology and Biotechnology, 2014. **41**(2): p. 387-402.

2.1.7 Supplementary Material

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

Gene in the *S. albus* genome	Gene name	Putative product	Locus in the genome of *S. koyangensis* 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	*flk*A	DeoR/GlpR transcriptional regulator [*Streptomyces* sp. CS227]	D0C37_RS06225
XNR_0969	*flk*B	sugar ABC transporter substrate-binding protein [*Streptomyces albidoflavus*]	D0C37_RS06230
XNR_0970	*flk*C	integral membrane sugar transporter	D0C37_RS06235
XNR_0971	*flk*D	carbohydrate ABC transporter permease	D0C37_RS06240
XNR_0972	*flk*E	zinc-dependent alcohol dehydrogenase family protein	D0C37_RS06245
XNR_0973	*flk*F	DUF2029 domain-containing protein	D0C37_RS06250
XNR_0974	*flk*G	FkbM family methyltransferase	D0C37_RS06255
XNR_0975	*flk*H	glycosyltransferase family 4 protein alpha-(1-2)-phosphatidylinositol mannosyltransferase	D0C37_RS06260
XNR_0976	*flk*I	response regulator transcription factor	D0C37_RS06265
XNR_0977	*flk*J	histidine kinase	D0C37_RS06270
XNR_0978	*flk*K	ABC transporter ATP-binding protein	D0C37_RS06275
XNR_0979	*flkL*	ABC type antimicrobial transporter permease, FtsX-like permease family protein	D0C37_RS06280
XNR_0980	*flkM*	LuxR family response regulator	D0C37_RS06285
XNR_0981	*flkN*	Two component sensor histidine kinase	D0C37_RS06290
XNR_0982	*flkO*	Alpha/beta hydrolase MppK beta-lactamase, serine hydrolase	D0C37_RS06295
XNR_0983	*flkP*	NRPS	D0C37_RS06300
XNR_0984	*flkR*	MbtH family protein	D0C37_RS06305
XNR_0985	*flkS*	NRPS	D0C37_RS06310
-	-	NRPS	D0C37_RS06315
XNR_0986	*flkT*	Fe(II)/alpha-ketoglutarate-dependent arginine beta-hydroxylase	D0C37_RS06320
XNR_0987	*flkU*	Major Facilitator Superfamily transporter	D0C37_RS06325
XNR_0988	*flkV*	Cytochrome 450	D0C37_RS06330
XNR_0989	*flkW*	ArsR family transcriptional regulator	D0C37_RS06335
XNR_0990	*flkX*	Thioeterase, 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+ 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	
XNR_1347		PLP-dependent aminotransferase (pyrodoxal dependent)	

1. NMR Spectroscopy

Table S2. NMR data of iso-faulknamycin 1 (DMSO-d_6, ^{1}H: 700 MHz, ^{13}C: 175 MHz)

unit	δ_C	δ_H, multiplicity, (J in Hz)	COSY	ROESY	HMBC (H-)
1 – COOH	n.a.[1]				
2 – CH	58.1	4.22, m	3, 5		
3 – CH	66.4	4.12, m	2, 4		
4 – CH_3	20.3	1.06, d (6.1)	3		2, 3
5 – NH		8.37	2		
6 – CO	n.a.[1]				
7 – CH	54.3	4.70, ovl.[2]	8, 14		
8 – CH	50.2	3.73, m	7, 9		8
9 – CH_2a CH_2b	22.0	1.93, ovl.[2] 1.67, m	8, 10		
10 – CH_2a CH_2b	36.1	3.37, ovl.[2] 3.19, m	9, 11		
11 – NH		8.09, bs	10		
12 – C		n.a.[1]			
12 – NH		n.a.[1]			
13 – NH		n.a.[1]			
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_3	19.3	0.84, ovl.[2]	17		16 ,17 ,19
19 – CH_3	18.1	0.82, ovl.[2]	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_2a CH_2b	41.8	1.49, ovl.[2] 1.40, m	22, 24		
24 – CH	24.2	1.50, ovl.[2]	23, 25, 26		
25 – CH_3	23.2	0.86, ovl.[2]	24		23 ,24 ,26
26 – CH_3	21.6	0.81, ovl.[2]	24		23 ,24 ,25
27 – NH		8.27, d (8.4)	22		28
28 – CO	169.0				
29 – CH	58.7	4.72, ovl.[2]	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.[1]				
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 – CH3	16.5	0.39, d (6.0)	40		39, 40
42 – NH2		7.88, bs	39		

[1]n.a. = not available; [2]ovl. = signal overlap

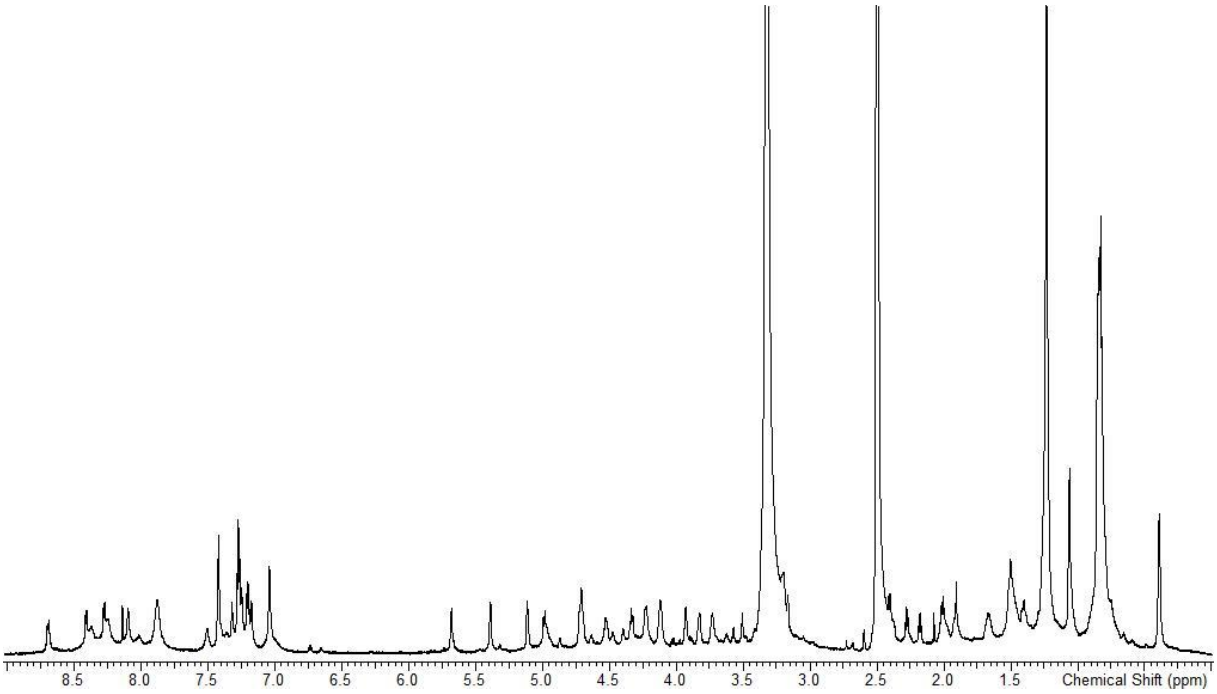

Figure S1: ^{1}H NMR spectrum of iso-faulknamycin 1 (DMSO-d_6, 700 MHz)

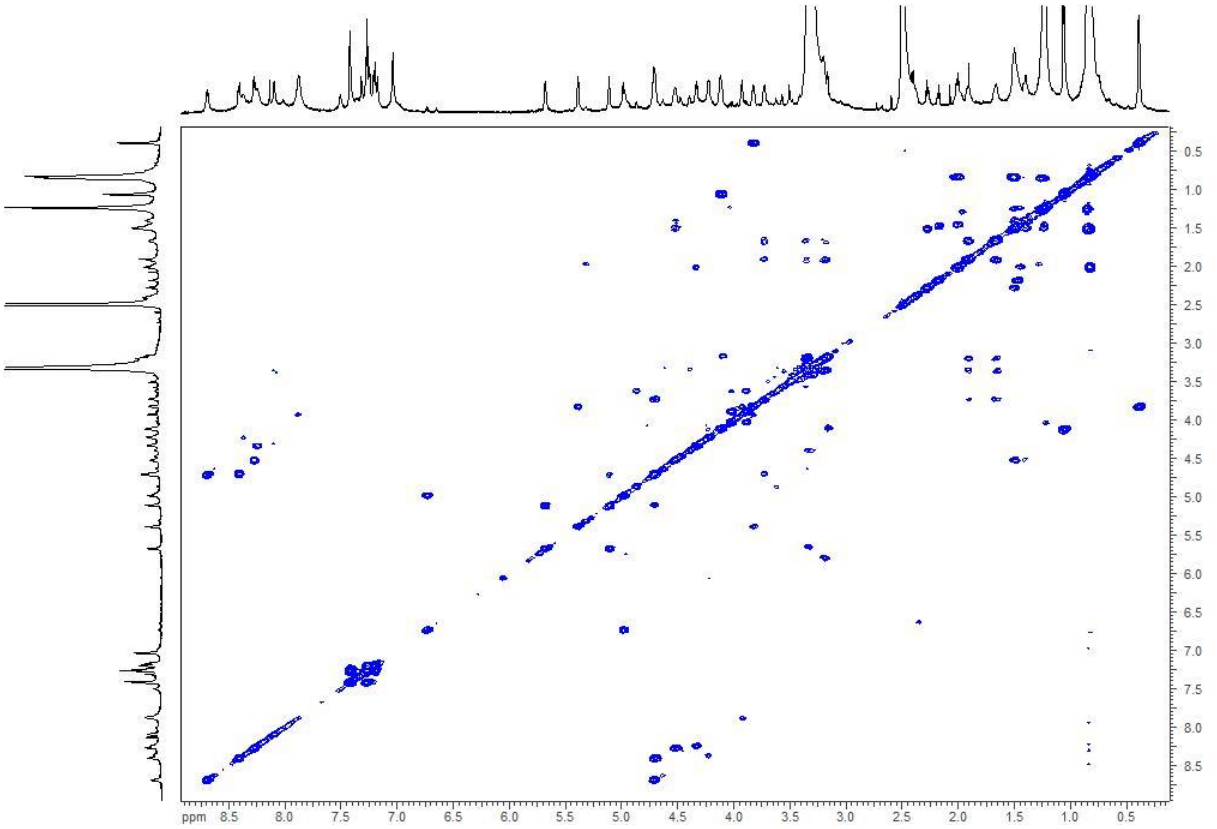

Figure S2: COSY spectrum of iso-faulknamycin 1 (DMSO-d_6, 700 MHz)

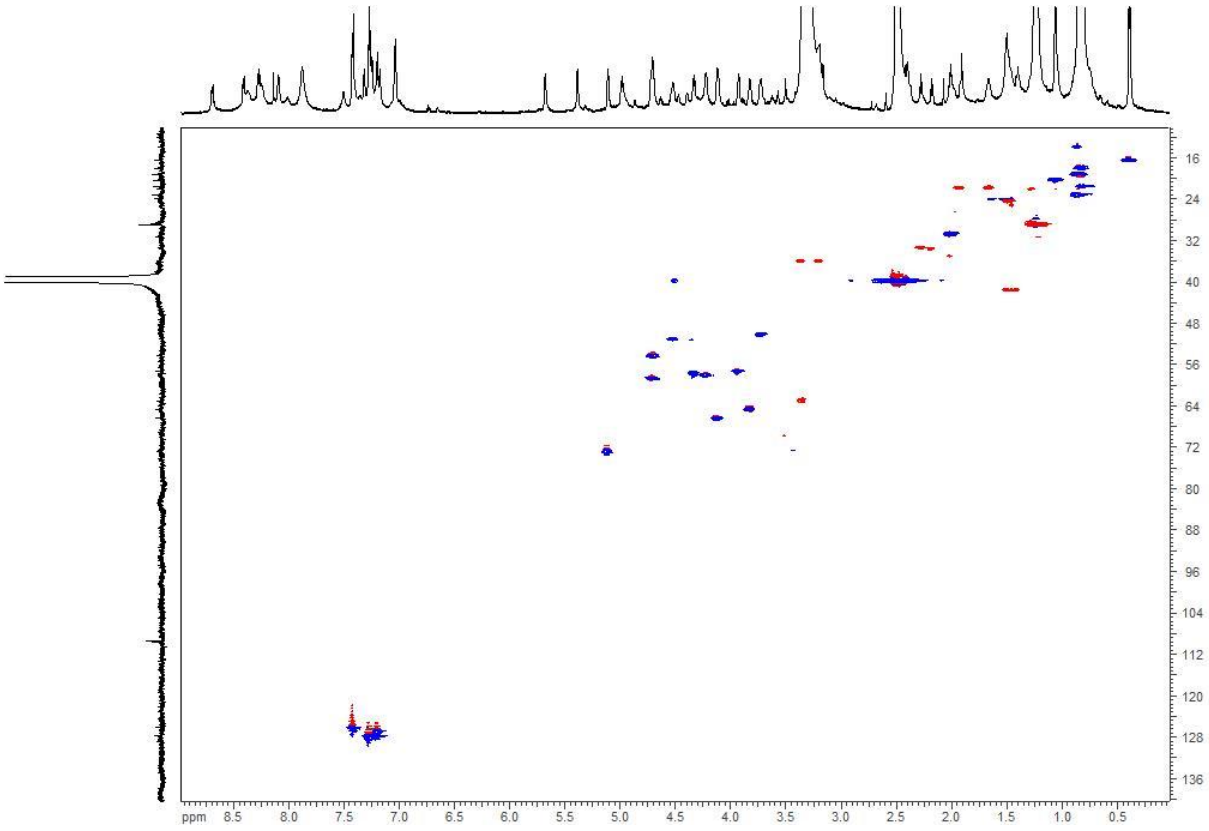

Figure S3: HSQC spectrum of iso-faulknamycin 1 (DMSO-d_6, 700 MHz, 175 MHz)

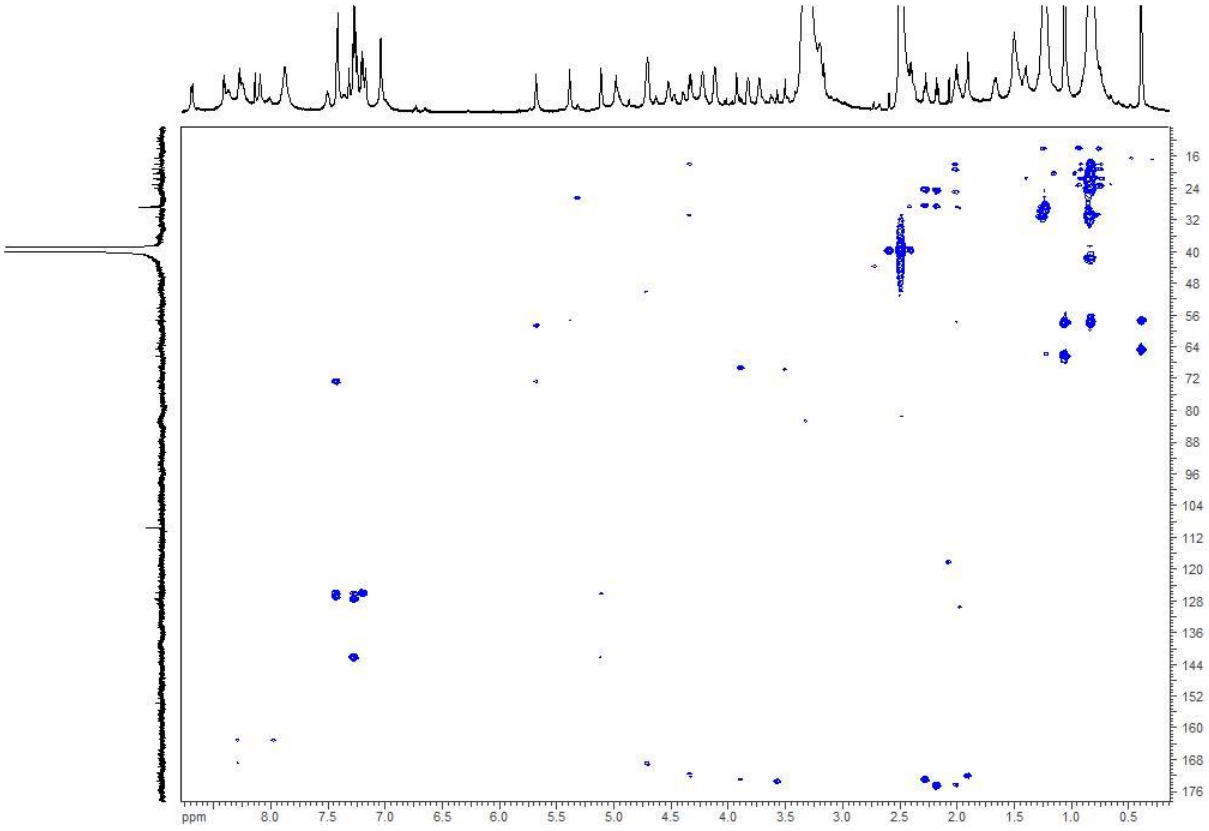

Figure S4: HMBC spectrum of iso-faulknamycin 1 (DMSO-d_6, 700 MHz, 175 MHz)

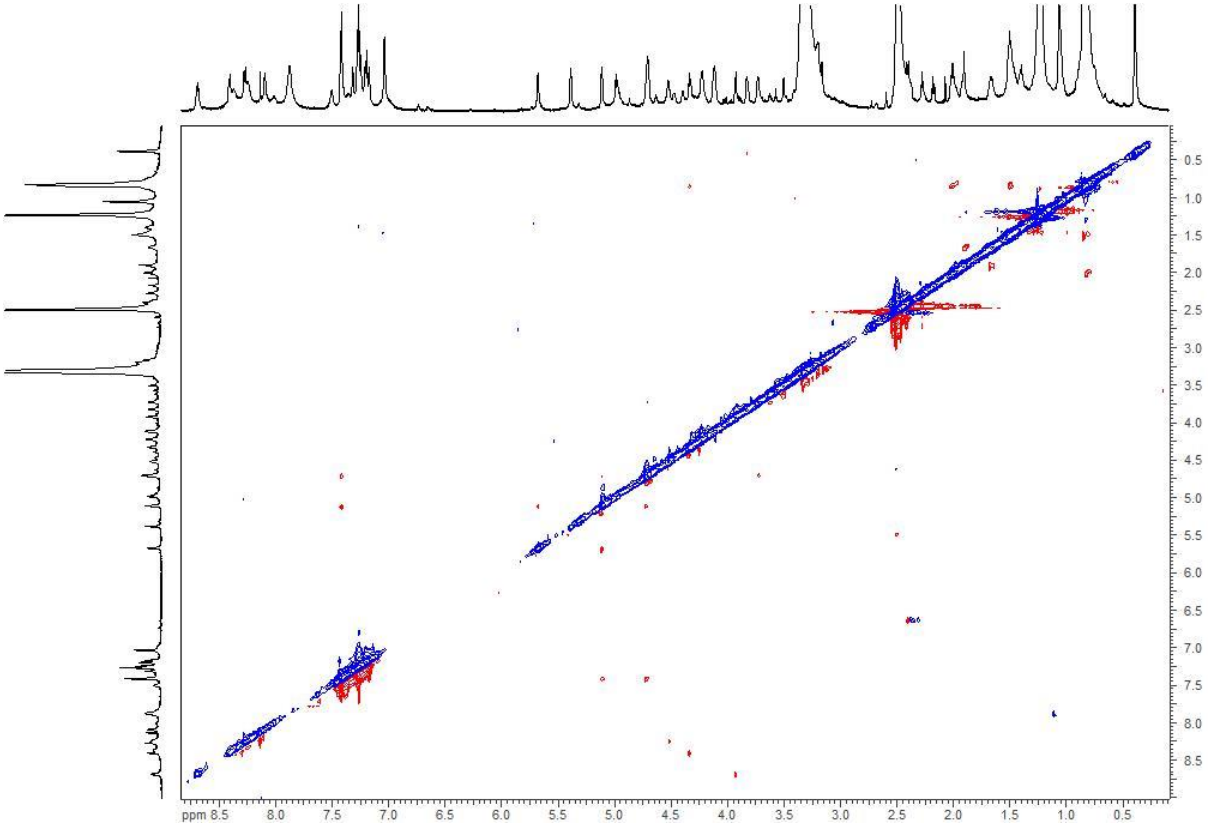

Figure S5: ROESY spectrum of iso-faulknamycin 1 (DMSO-d_6, 700 MHz)

Table S3: NMR data of cyclofaulknamycin 2 (DMSO-d_6, ^{1}H: 700 MHz, ^{13}C: 175 MHz)

unit	δ_C	δ_H, 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_3	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_2	21.3	1.76, m	8, 10		
10 – CH_2	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_3	20.0	0.86, ovl	17		16 ,17 ,19
19 – CH_3	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_2a CH_2b	40.4	1.37, ovl 1.46, m	22, 24		21
24 – CH	24.9	1.38, ovl	23 ,25 ,26		21
25 – CH_3	23.1	0.79, d (5.9)	24		23, 24 ,26
26 – CH_3	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_3	20	0.98, d (6.2)	40		39, 40
42 – NH		8.52, ovl	39	2	

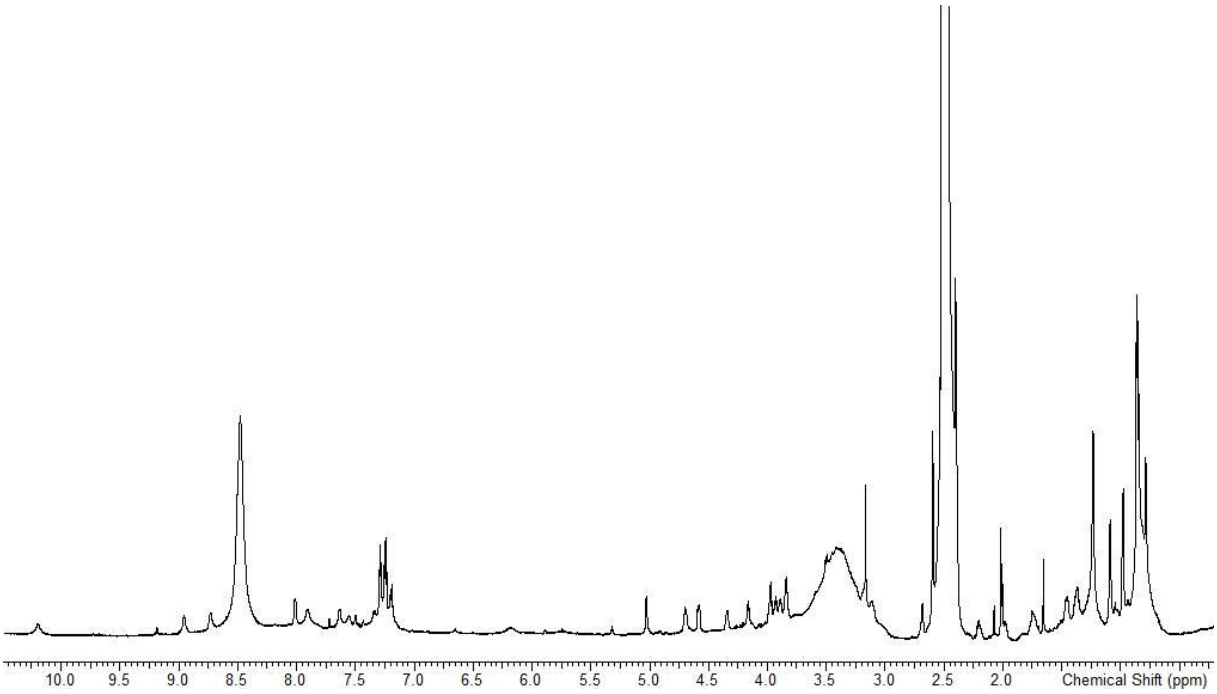

Figure S6: ^{1}H spectrum of cyclofaulknamycin 2 (DMSO-d_6, 700 MHz)

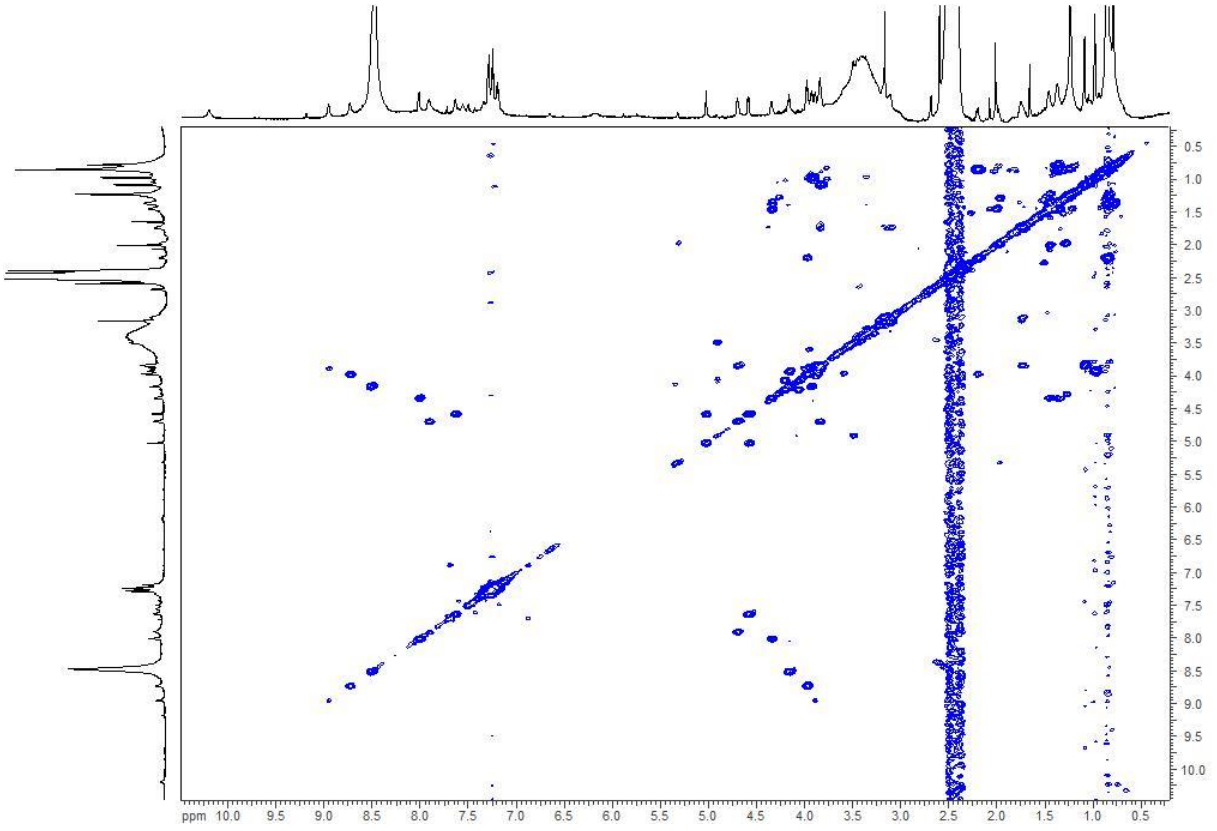

Figure S7: COSY spectrum of cyclofaulknamycin 2 (DMSO-d_6, 700 MHz)

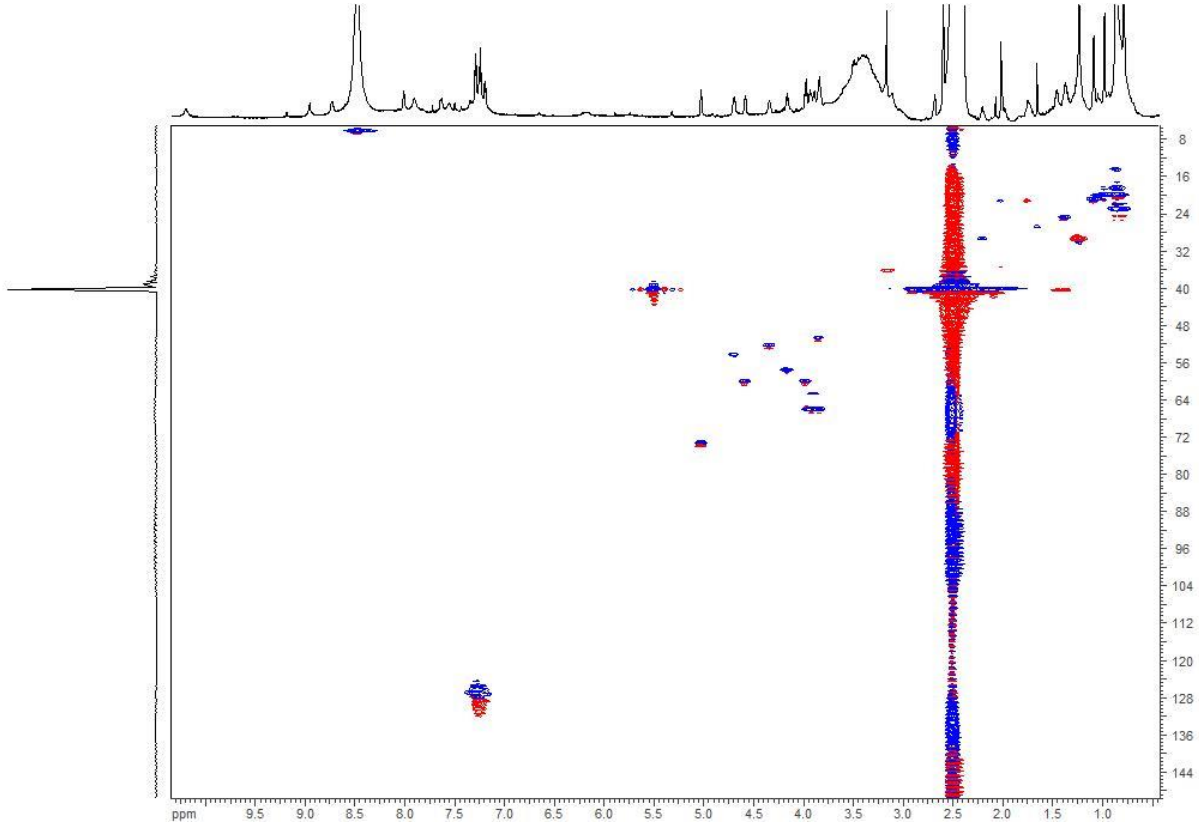

Figure S8: HSQC spectrum of cyclofaulknamycin 2 (DMSO-d_6, 700 MHz, 175 MHz)

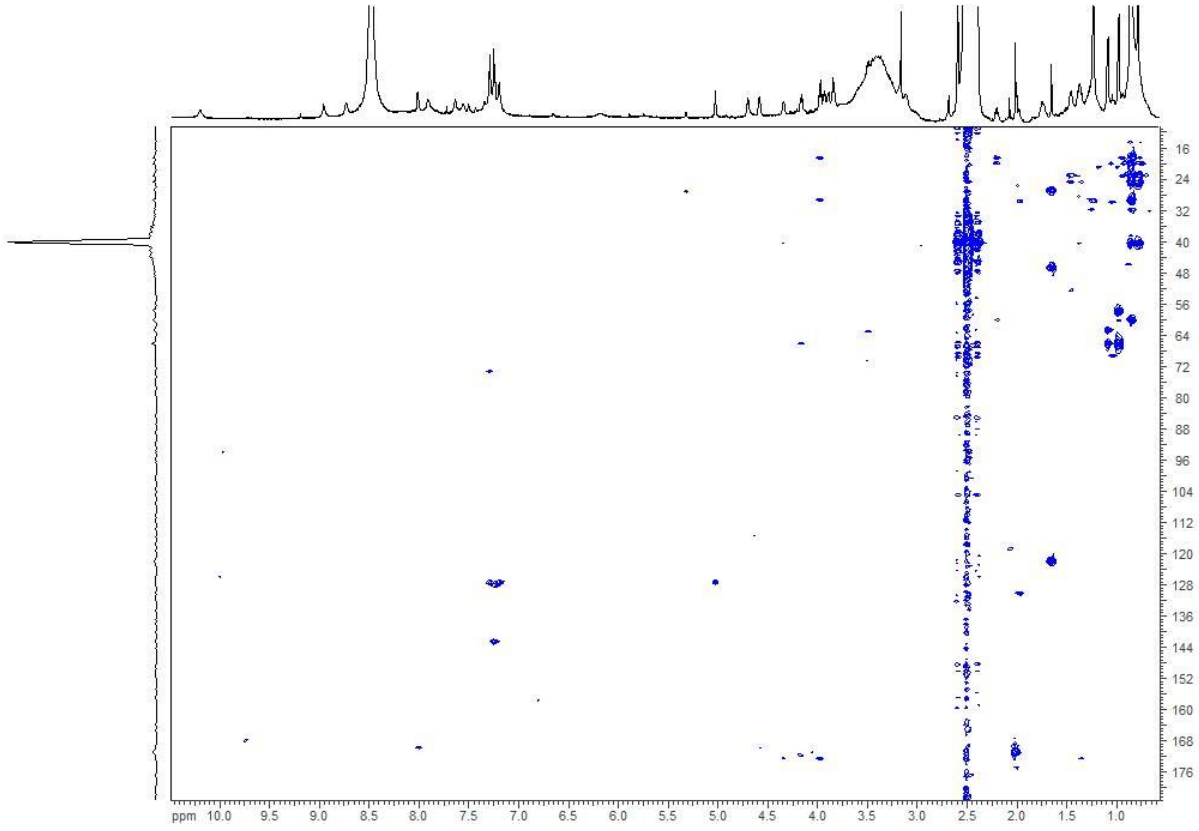

Figure S9: HMBC spectrum of cyclofaulknamycin 2 (DMSO-d_6, 700 MHz, 175 MHz)

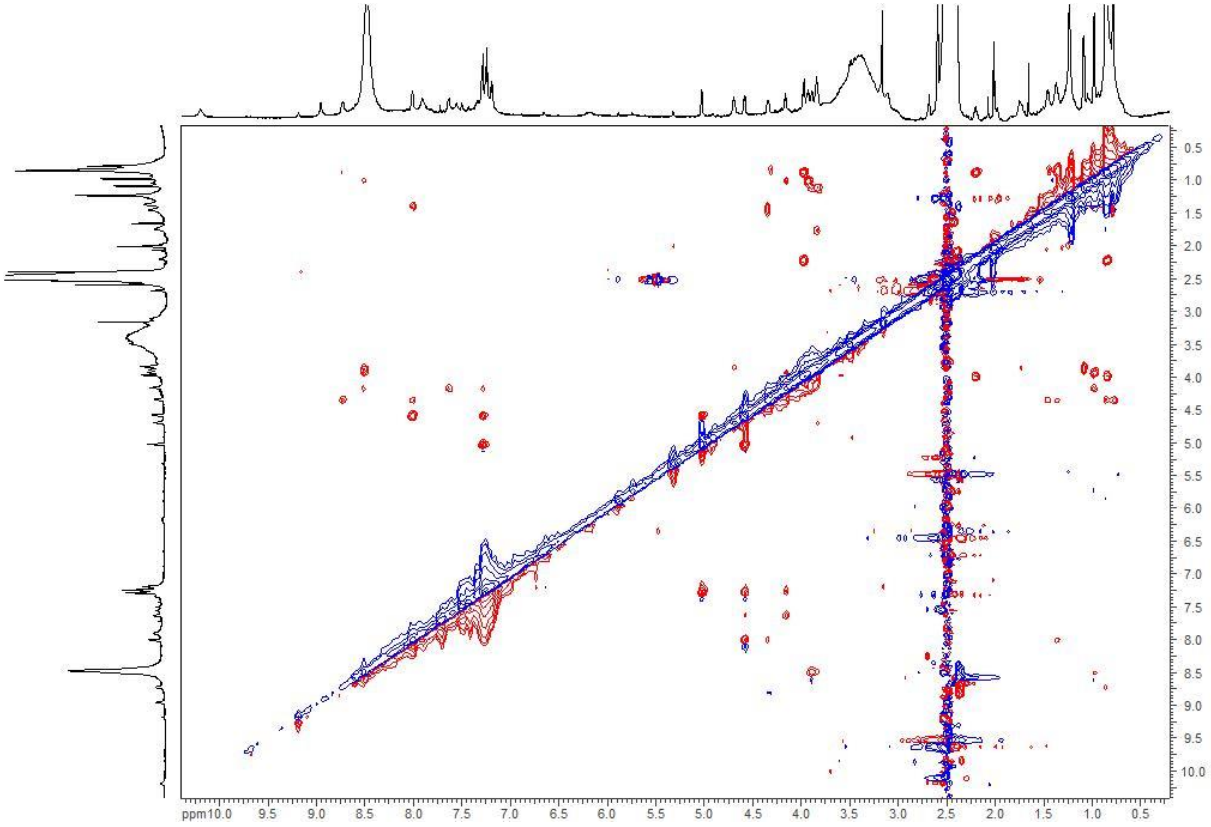

Figure S1: ROESY spectrum of cyclofaulknamycin 2 (DMSO-d_6, 700 MHz)

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).

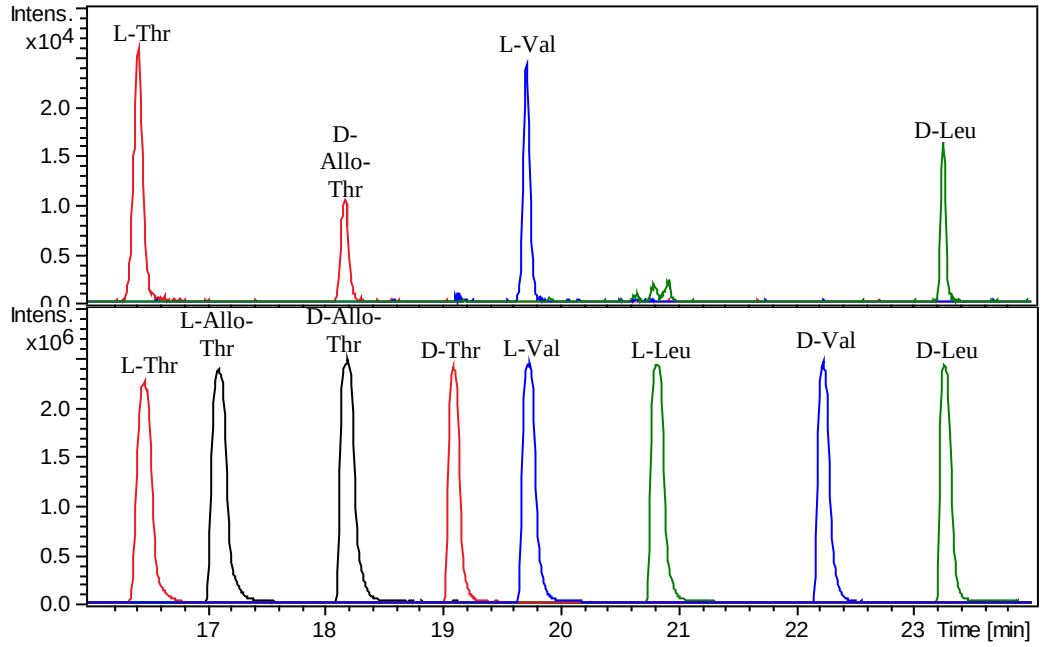

Figure S2: Marfey's chromatograms of the iso-faulknamycine hydrolysate (top) and the amino acids standards (bottom) derivatized with L-FDLA.

We obtained 2S-3R-capreomycidine **3** and 2S-3S-capreomycidine **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-capreomycidine	*LSR*	*DSR*	2S-3S-capreomycidine	*LSS*	*DSS*
RT equivalent to	*DRS*	*LRS*	RT equivalent to	*DRR*	*LRR*

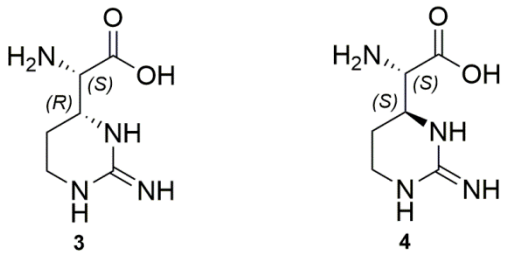

Figure S12: Capreomycidine from capreomycine (3) and chymostatine (4).

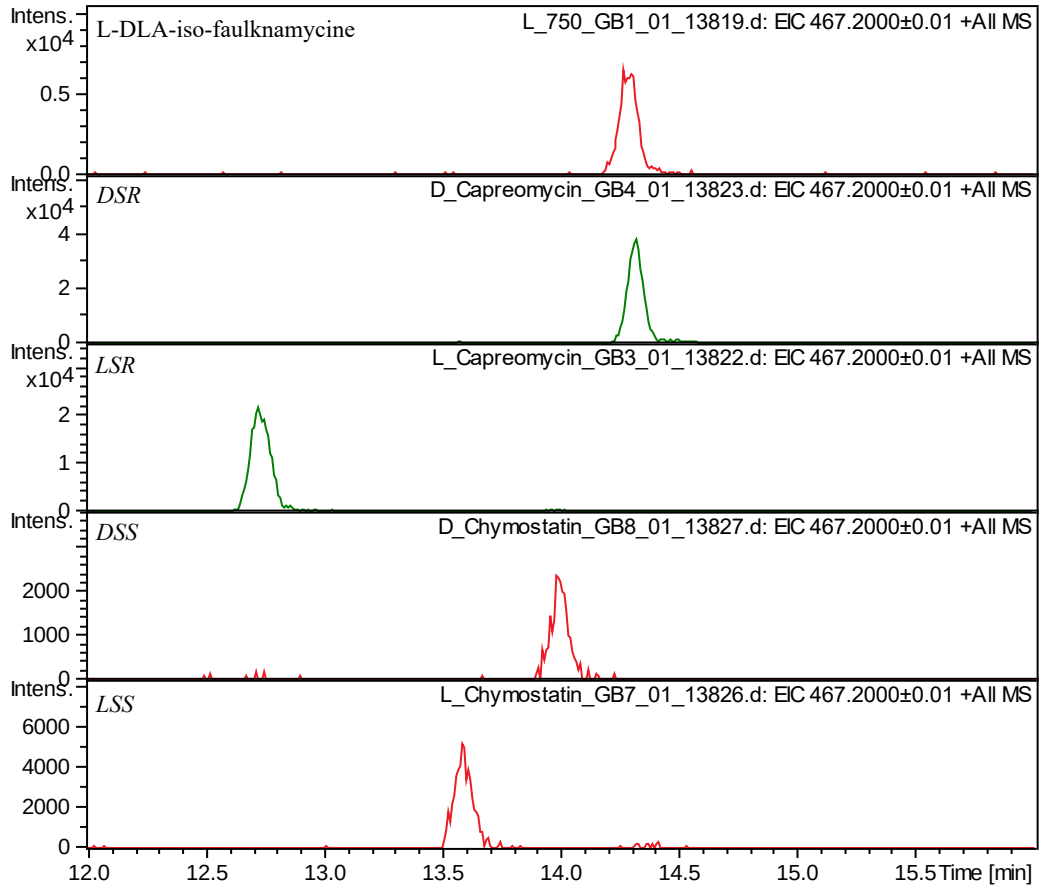

Figure S13: Marfey's chromatograms showing the extracted mass of DLA-capreomycidine of the hydrolysed L-DLA-iso-faulknamycine and the references obtained from hydrolysis of capreomycidine and chymostatin derivatized with D-FDLA (DSR, DSS) and L-FDLA (LSS, LSR). DSR is equivalent to LRS, therefore iso-faulknamycine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

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

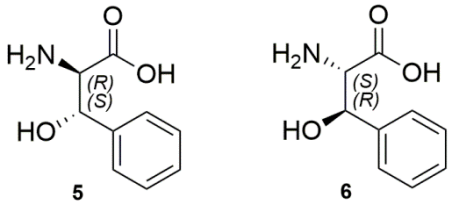

Figure S14: D-*threo*-β-phenylserine (5) and L-*threo*-β-phenylserine (6)

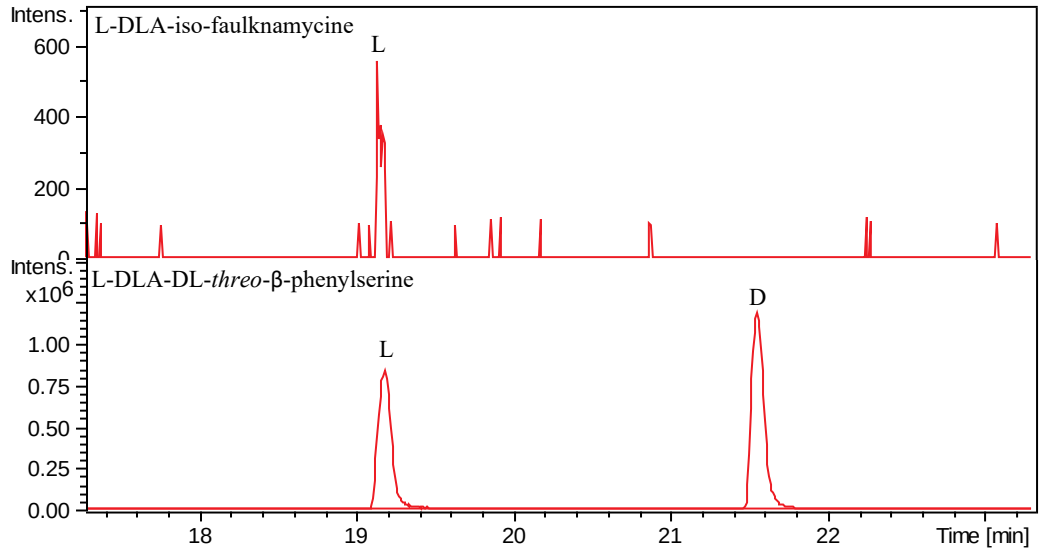

Figure S15: Marfey's chromatograms showing the extracted mass of DLA-phenylserine of the hydrolysed iso-faulknamycine and DL-*threo*-β-phenylserine, both derivatized with L-FDLA.

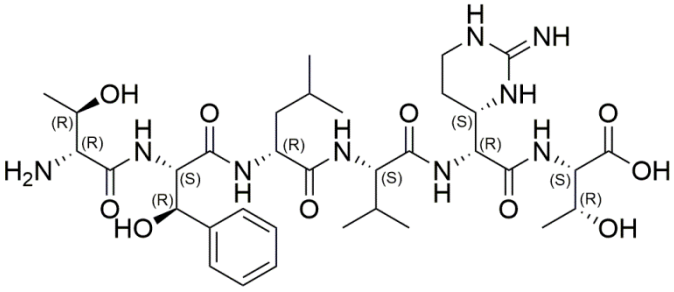

Figure S16: Absolute configuration of iso-faulknamycine.

3. MS/MS Fragmentation

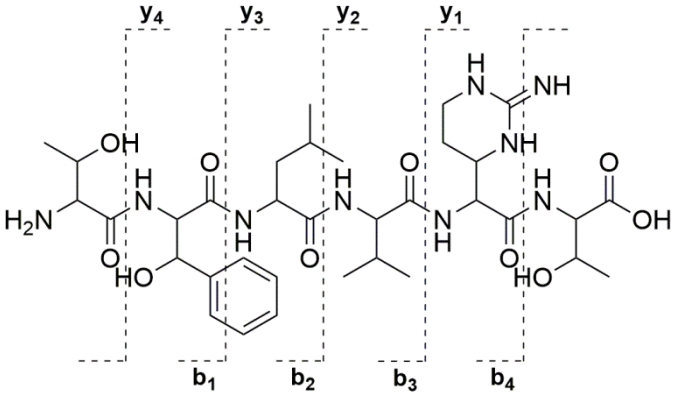

Figure S3: Observed y-ion and b-ions from the MS/MS fragmentation (see table 5) of iso-faulknamycine

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

ion type	y_1	y_2	y_3	y_4
[M]+ calculated	274.1511	373.2190	486.3032	649.3675
[M]+ observed	274.1517	373.2180	486.3034	649.3674
Δ [ppm]	-2.2	3.8	-0.4	0.2
ion type	**b_1**	**b_2**	**b_3**	**b_4**
[M]+ calculated	265.1186	378.2019	477.2697	631.3556
[M]+ observed	265.1185	378.2021	477.2698	631.3573
Δ [ppm]	0.4	-0.5	-0.2	-2.7

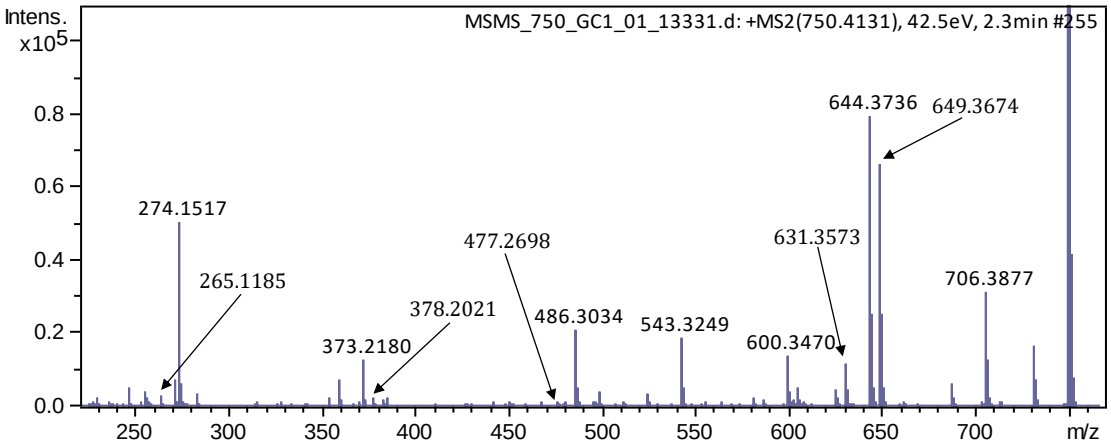

Figure S18: MS/MS fragmentation spectrum of iso-faulknamycine ($[M+H]^+$ peak).

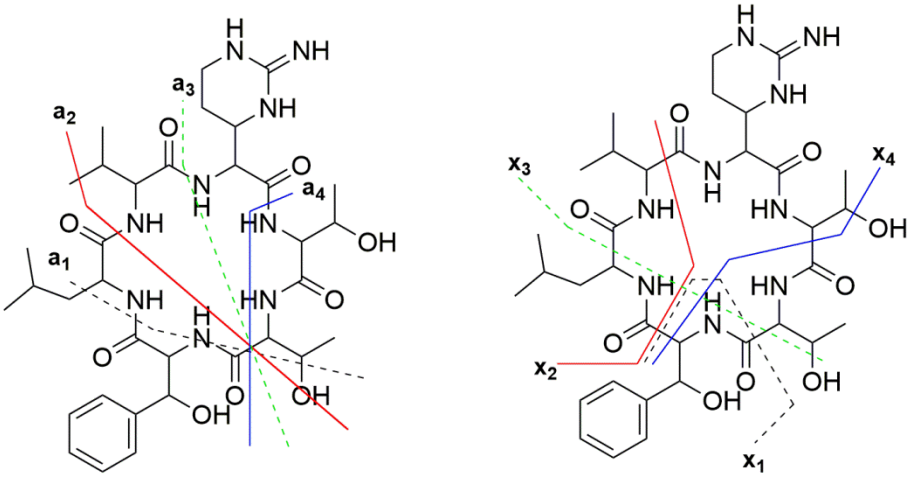

Figure S19: Observed a-ion and x-ions from the MS/MS fragmentation (see table 6) of cyclofaulknamycine

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

ion type	**a_1**	**a_2**	**a_3**	**a_4**
AA sequence	T_1-T_2	T_1-T_2-Cmp	T_1-T_2-Cmp-V	T_1-T_2-Cmp-V-L
[M]+ calculated	175.1077	329.1932	428.2616	541.3457
[M]+ observed	175.1082	329.1940	428.2622	541.3474
deviation [ppm]	2.9	2.4	1.4	3.1
ion type	**x_1**	**x_2**	**x_3**	**x_4**
AA sequence	βPhS(-H_2O)	L-V	V-cmp-T_2-T_1(-H_2O)	L-V-Cmp-T_2(-H_2O)
[M]+ calculated	146.0600	211.1441	438.2459	466.2772
[M]+ observed	146.0607	211.1444	438.2470	466.2782
deviation [ppm]	4.8	1.4	2.5	2.1

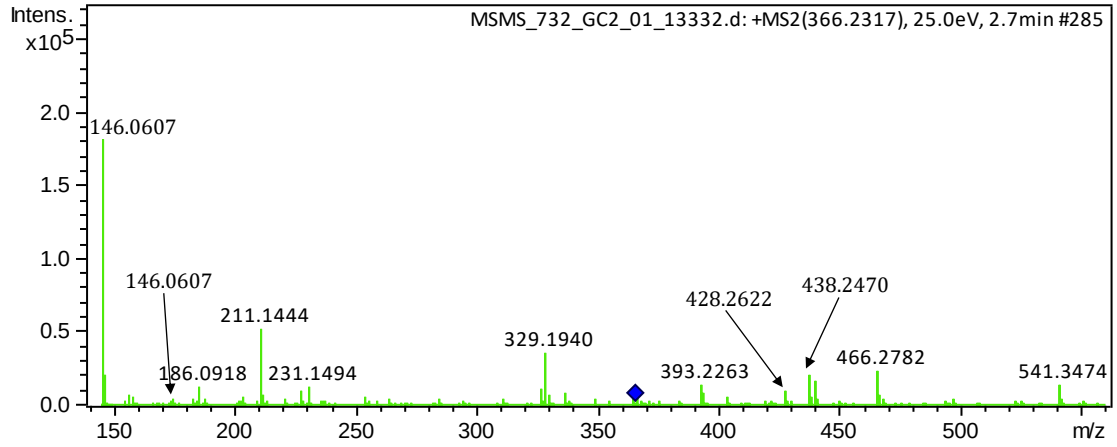

Figure S20: MS/MS fragmentation spectrum of cyclofaulknamycine ($[M+2H]^{2+}$ peak)

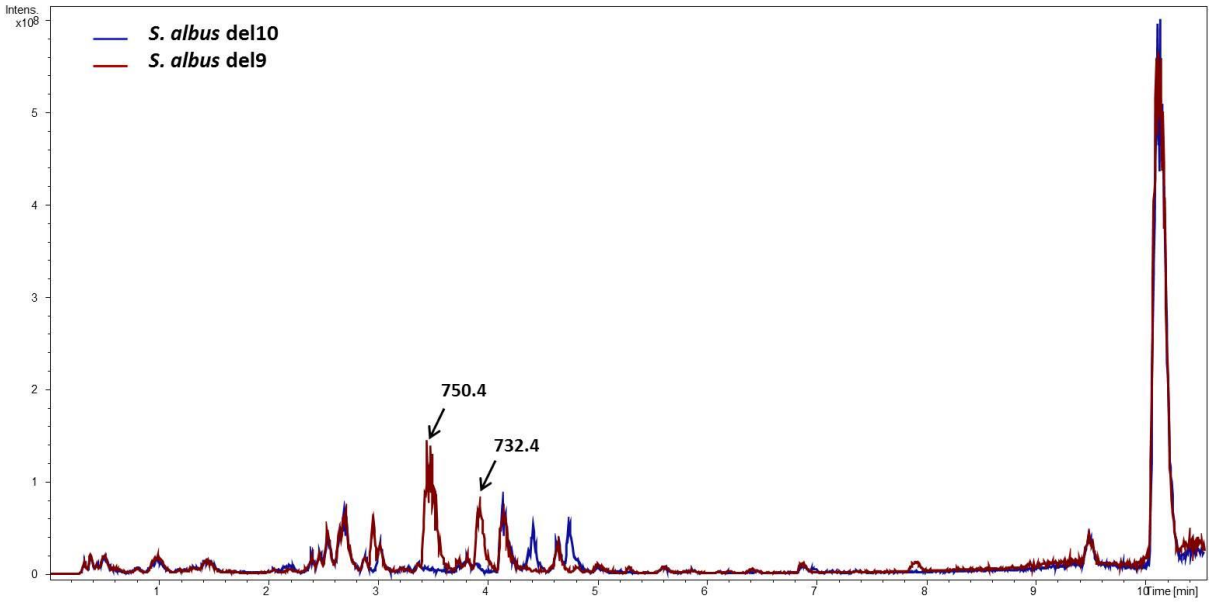

Figure 2. Secondary metabolite profile of the *S. albus* del9 and del10 strains. Butanol extraction.

2.2 Discovery and heterologous production of the new polyketide type II aorimycin with unique ring arrangements

Nikolas Eckert, Constanze Paulus, Marc Stierhof, Liliya Horbal, Andriy Luzhetskyy

To be submitted

2.2.1 Abstract

The strain *Streptomyces* sp. IB2014/011-12 has been previously described as a producer of new alpiniamide natural products (NPs). In this study we report the isolation of the unique polyketide type II aorimycin. It consists of a naphthalene core that bears a unique 9-membered ring with an epoxidised tetrahydrofuran ring and a D-forosamine residue. Based on the rare structure, we identified the respective biosynthesis gene cluster via genome mining and gene deletion experiments. The identity of the cluster was further proven by heterologous expression. Cyclisation steps during the biosynthesis of this compound were analysed in the heterologous host, by means of additional deletion experiments. Taking those findings into account the biosynthetic pathway was proposed. Aorimycin represents an example for diversification of core structures and the necessity to analyse the function of the respective enzymes, e.g. cyclases in more detail.

2.2.2 Introduction

A variety of bioactive NPs are produced by *Streptomyces*, such as polyketides, terpenes, NRPS and saccharides [1]. Based on the biosynthesis polyketides can be further divided in PKS type I, II and III [2]. A wide variety of natural compounds, such as e.g. landomycin [3], tetracycline [4] and daunorubicin [5], are synthesised by type II polyketide synthases (PKSs) and find therapeutic use as antitumor agents [6] or antibiotics [7]. Typical PK type II biosynthetic gene clusters (BGCs) contain a set of iteratively used core enzymes, ketosynthases (KSα, KSβ) and acyl carrier protein (ACP), responsible for the production of the core polyketide structure. The ACP is loaded with an acyl-CoA (typically acetyl-CoA), which is then condensed with several malonyl-coenzyme A (CoA) units catalysed by the KSs. Once the core structure is complete, it is modified by ketoreductases (KRs), aromatases/cyclases (Cycs) and oxygenases. At last an aromatic scaffold can be further decorated by oxygenases, glycosyl and methyl transferases as well as other enzymes leading to a final compound [8].

Considering that there is a huge diversity of natural products encoded by the type II PKS BGCs, even though the core genes are highly similar, makes the study of the cyclisation and tailoring reactions even more interesting. As described in Di Marco et al. 2007 the cyclisation of the first 3 aromatic rings can be quite versatile and aromatases, cyclases, KSs and KRs or combinations thereof could be catalysing these reactions [9]. Not only the cyclisation is an

interesting step also further modification of the compound by other enzymes like oxygenases have a huge impact on the diversity of this compound class. They can perform different reactions like for example epoxidations or Baeyer–Villiger reactions [10]. A glycosylation usually occurs at the later biosynthetic stage and very often the glycosyl moiety determines the activity of the compounds. Therefore, the glycosylation of compounds became an important tool to improve pharmaceutical properties of the drug.

In this study we report the isolation of the new polyketide aorimycin from *Streptomyces* sp. IB2014/011-12. It consists of a naphthalene core that bears a unique 9-membered ring with an epoxidised tetrahydrofuran ring and a rare D-forosamine residue. Based on the structure a genome mining was performed to identify the corresponding gene cluster. The corresponding biosynthesis gene cluster (BGC) was verified by gene knockouts, cloning and heterologous expression. Important cyclisation steps of the biosynthesis have been analysed and discussed based on the targeted inactivation of corresponding cyclases.

2.2.3 Material and methods

Bacterial strains and culture conditions

Used bacteria are listed in table S1. Cultivation of *Escherichia coli* strains was performed in liquid or solid lysogeny broth (LB) 2 % (w/w) agar. The strains were incubated at 37 °C and at 180 rpm for liquid cultures. Gene knockouts were performed in the *E. coli* GB05-redCC strain (Table S1) by targeted homologous recombination (Red/ET).

For sporulation *Streptomyces* strains were grown on mannitol soy agar (MS) (soy flour 20 g, agar 20 g, mannitol 20 g, tap water 1 l) for 7 days at 28 °C. The pre-culture for liquid cultivation was performed under the following conditions: 20 ml tryptic soy broth (TSB, Sigma-Aldrich, USA); 100 ml flask (4 baffles); glass beads (Carl Roth; 4 mm); 1 day; 29 °C; 180 rpm; antibiotics if necessary. Inoculation of the main culture was done with 1 ml pre-culture and incubated for 5 days under the same conditions in 50 ml DNPM medium (dextrin 40 g, soytone 7.5 g, fresh yeasts 5 g, MOPS 21 g and 1 l dH_2O; pH = 6.8) or defined medium (DM) (L-asparagin 0.5 g, K_2HPO_4 0.5 g, $MgSO_4$ * 7 H_2O 0.2 g, $FeSO_4$ * 7 H_2O 0.01 g, Glucose 10 g (added after autoclaving), Trace element solution 1 ml ($ZnCl_2$ 40 mg, $FeCl_3$ * 6 H_2O 200 mg and 10 mg of $CuCl_2$ * 2 H_2O, $MnCl_2$ * 4 H_2O, $Na_2B_4O_7$ * 10 H_2O and $(NH_4)_6Mo_7O_{24}$ * 4 H_2O) 1 l dH_2O), in a 300 ml flask with one baffle and glass beads.

Antibiotics (Roth, Germany) were added in the following concentrations, if needed: 50 µg/ml apramycin, 50 µg/ml kanamycin, 120 µg/ml hygromycin, 50 µg/ml chloramphenicol and 50 µg/ml nalidixic acid.

Recombinant DNA techniques

Chromosomal DNA from *Streptomyces* strains and plasmid DNA from *E. coli* were isolated using standard protocols [11]. Restriction enzymes and molecular biology reagents were used according to the manufacturer's protocol (Thermo Scientific, Germany).

Construction of a cosmid library and a BAC covering the whole cluster

A cosmid library of the strain *Streptomyces* sp. IB2014/011-12 was constructed using the vector cos15A-gus (Table S1) and the CopyControl fosmid library Kit (Lucigen, USA). After analysing the library 2 cosmids (12-A05; 06-C09) were selected and fused together to cover the whole potential cluster using transformation-associated recombination (TAR) [12]. For the fusing step, the BAC vector pSMART_cloni (Table S1) was used. The two cosmids and the vector were digested and afterwards mixed and transformed in *Saccharomyces cerevisiae* BY4742 [13]. Assembled BACs were checked by restriction mapping and PCR. The assembled BAC is referred to as BAC-11-12-07.

Gene knockouts using Red/ET method

For this purpose a hygromycin resistance cassette (Hyg^R) was amplified using oligonucleotides with 60 bp long homologous regions up and down-stream of the gene that should be deleted (Table S2). The obtained PCR fragment was used to perform recombination with the BAC-11-12-07 or the cosmid 12-A05. The integrase of the cosmid was replaced by a chloramphenicol resistance via Red/ET and used for wild type knockouts. After recombination, colonies were selected on LB agar plates with 120 µg/ml hygromycin or 50 µg/ml chloramphenicol. The obtained constructs were validated by PCR, sequencing and restriction digests.

For further screening the obtained constructs were transferred into *S. albus* J1074 Del14 via conjugation or *Streptomyces* sp. IB2014/011-12 for wildtype gene knockouts.

Extraction of aorimycin

Aorimycin was extracted from the supernatant of the culture broth with butanol and acetic acid. For small scale extractions 20 ml of a bacterial culture were centrifuged at 9000 g for 5 min at room temperature (RT). 10 ml supernatant was mixed with 10 ml butanol and 250 µl

acetic acid. Extracts were shaken for 1 h at 180 rpm and RT (Laboshake, Gerhardt GmBH, Dreieich), afterwards centrifuged under the same conditions. 5 ml of the butanol phase were evaporated, dissolved in 600 µl methanol, centrifuged (19000 g, 10 min, 4 °C) and subjected to HPLC-MS analysis.

For aorimycin isolation in larger amounts *Streptomyces* sp. IB2014/011-12 was cultivated in 5 l DNPM medium. Intermediates were purified of *S. albus* J1074 Del14 containing the BAC-11-12-07 with deletions, cultivated in 5 l defined medium. Subsequently aorimycin and intermediates were extracted as described above and the extracts further purified by normal phase, size exclusion and reverse phase chromatography.

HPLC-MS analysis of aorimycin production

Aorimycin extracts were analysed by HPLC-HRMS analysis (Dionex Ultimate 3000, Thermo Fisher Scientific; AmaZon Speed ETD, Bruker). Samples were eluted with solvent A (MilliQ, 0.1 % formic acid) and B (Acetonitrile, 0.1 % formic acid) and the gradient of solvent B from 5 to 95 % in 18 min was used. A Waters BEH C18 column (100 mm x 2.1 mm, 1.7 µm) was used, the column temperature was set to 45 °C and the flow was set to 0.6 ml/min. The data were analysed with the software Compass Data Analysis v. 4.1 (Bruker).

Samples from the recombinant strains were additionally analysed on a maXis 4G hr-ToF ultrahigh-resolution mass spectrometer using the Apollo II ESI source (Bruker Daltonics, Germany). Extracts were separated on UPLC system (Dionex Ultimate 3000, Thermo Fisher Scientific GmbH) with a Waters BEH C18 column (100 mm x 2.1 mm, 1.7 µm). The same solvents and conditions as above were used.

Purification of aorimycin

For normal phase and size exclusion chromatography the isolation process was guided by LC-MS to identify fractions containing the respective compound. Aorimycin and intermediates thereof were purified as described in Lasch et al. 2021 [14]. The crude extracts of the total culture volume was dissolved in methanol and 4 purification steps were performed.

Firstly, a normal phase chromatography using an Isolera One system (Biotage, Uppsala, Sweden) with an SNAP Ultra 50 g column (Biotage, Uppsala, Sweden) was performed. As mobile phase the following solvents were used: [A] n-hexane/[B] chloroform/[C] ethyl acetate/[D] methanol. The gradient was set as following with a flowrate of 100 ml/min: [A]/[B] 10 column volumes (CV), [B]/[C] 15 CV, [C]/[D] 15 CV.

Secondly, a size exclusion chromatography was performed. As the stationary phase Sephadex® LH 20 (Sigma Aldrich, Germany) was used and an isocratic elution was performed with methanol as mobile phase. The 63 cm long column was packed with 300 ml of bloated Sephadex® LH 20 material (methanol). The collected fractions were screened for the presence of the desired compound by HPLC-MS analysis and fractions containing the same compound were pooled for further purification steps.

Thirdly, a preparative HPLC-MS purification was performed using the Waters HPLC system (2545 Binary Gradient module, Waters, Milford, MA, USA). A Nucleodur C18 HTec 250/21 5 µm (Macherey-Nagel, Düren, Germany) column was used and the following solvents: [A] H_2O+0.1 % formic acid / [B] methanol+0.1 % formic acid. The gradient reached from 5 % to 95 % [B] over a time of 17 min at flow rate of 20 mL/min.

Lastly, the fractions containing the compound of interest were submitted to a last HPLC purification step with an Agilent Infinity 1200 series HPLC system. A Synergi 4 µm Fusion-RP 80 Å 250×10 (phenomenex, Torrance, CA, USA) column was used with the mobile phases: [A] H_2O+0.1 % formic acid/[B] acetonitrile+0.1 % formic acid (45 °C and a flowrate of 4 ml/min). Aberrantly to the method described in Lasch et al. 2021 the gradient reached from 5 % to 95 % and lasted over 18 min.

NMR analysis of aorimycins

Screening of the metabolome of *Streptomyces* sp. IB2014/011-12 from our strain library revealed two interesting compounds with masses of 559.20 Da and 573.22 Da. Analysis of their UV/VIS spectra indicated that both compounds might be polycyclic aromats due to the strong absorption at 266 and 418 nm. Dereplication of the compound with a mass of 559.20 Da revealed two compounds (Rhodomycin A and Oxaunomycin) with a similar mass, but different UV/VIS spectra while dereplication of 573.22 Da did not lead to any reasonable hits. To isolate the potentially new compounds, *Streptomyces* sp. IB2014/011-12 was cultivated in 5 l DNPM medium. Extraction and purification using the above mentioned methods led to the isolation of 4 mg of 559.20 Da and 3 mg of 573.22 Da. The structures were determined by 1D and 2D NMR experiments.

The molecular formular of 559.20 Da was calculated as $C_{28}H_{33}NO_{11}$ based on the monoisotopic mass m/z 559.20 Da. Analysis of edited HSQC, ^{1}H and ^{13}C NMR data revealed 4 methyls, 4 methylenes, 9 methines and 11 quaternary carbons. Two chemically equivalent

methyls at $\delta_{C,H}$ = 42.17, 2.77 indicate a dimethylamine group. By analysing HMBC and COSY correlations, the NMe_2 moiety was found to be attached to a spin system which resembles 2,3,4,6-tetradeoxy-4-(dimethylamino)-erythro-hexose (forosamine).

COSY correlations showed a spin system of three methines at δ_H = 7.08, 7.48, and 7.89 suggesting an aromatic ring structure. The ring system was established by analysing HMBC correlations. The aromaticity allowed the observation of J_4 couplings which were essential for the structure assignment since the ring system was lacking ^1H signals. The analysed spectral data indicated two annulated benzyl rings modified with hydroxyl groups at positions 6, 11 and 13 (Figure S1). The remaining two spin systems were assigned to a methyl 2-hydroxy-4-oxobutanoate and a 2-methyl-oxolan moiety. The final structure was assembled through long range HMBC correlations resulting in the final structure of 559.20 Da (Figure 1, A).

The molecular formular of 573.22 Da was calculated as $C_{29}H_{36}NO_{11}$ based on the monoisotopic mass m/z 573.22 Da and suggested an additional methyl-group. Analysis of 1D and 2D NMR data revealed a similar core structure as 559.20 Da. The difference of one methyl group resulted from an esterification of the free carboxyl group.

2.2.4 Results and discussion

Isolated compounds and elucidated structures

After cultivation of *Streptomyces* sp. IB2014/011-12 in DNPM medium, extraction and analysis of the HPLC-MS data, a peak with an accurate mass of 559.20 Da was identified. By comparing this mass to the Dictionary of Natural Products (Version 10.0, CRC Press, Boca Raton, FL, USA) database the best fitting compounds were rhodomycin A and oxaunomycin (559.20 Da). However, the expected UV/Vis spectra of these compounds did not matched the newly identified one. Additionally, another derivative with a similar UV spectrum and a difference in 14 Da (573.22 Da) was detected which did not lead to any reasonable hit in the database. Aiming to isolate these compounds, the strain was cultivated in 5 l DNPM medium, and the compound was purified over several chromatographic steps. In total 4 mg of aorimycin A and 3 mg of aorimycin B were purified and their unique structures determined by NMR experiments (Figure 1, A-B). Only two described compounds nargenicin and branimycins have a 9-membered ring, similar to the newly isolated aorimycins. Branimycins and nargenicin have been shown to be highly active against methicillin resistant Staphylococcus aureus (MRSA) [15], making further investigation of aorimycins interesting.

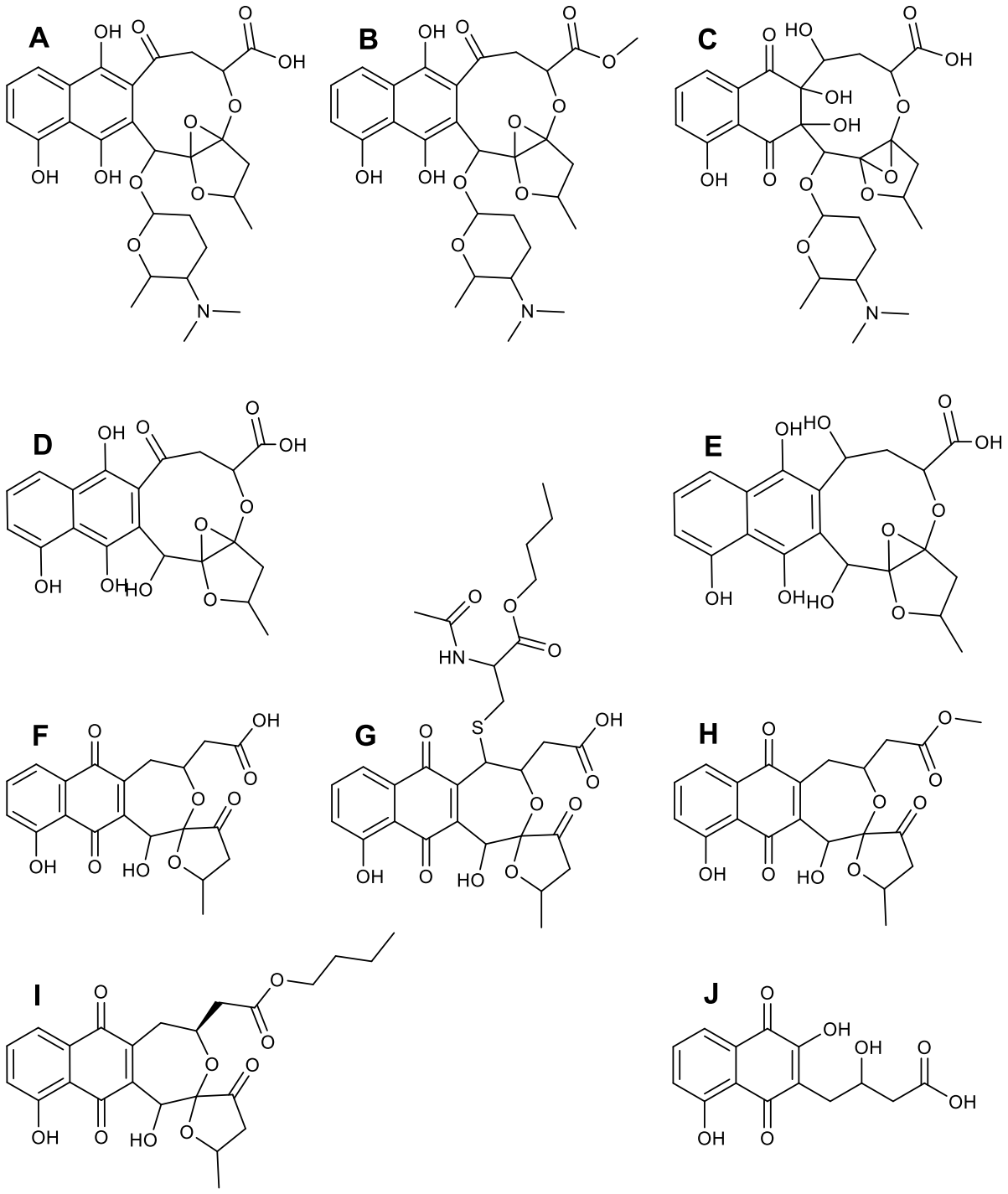

Figure 1: Elucidated structures of aorimycins and intermediates or derivatives thereof: A) Aorimycin A; B) Aorimycin B; C) Aorimycin C; D) Aorimycin-Int-6; E) Aorimycin-Int-5; F) Aorimycin-Int-1; G) Aorimycin-Int-4; H) Aorimycin-Int-3; I) Aorimycin-Int-2; J) Aorimycin-Shu

Identification and characterisation of the biosynthesis gene cluster

The unique structure and diversity of aorimycins made it interesting to investigate this biosynthesis further. Based on typical quinone structure, we hypothesised that aorimycin should originate from a type II PKS. Additionally the presence of a D-forosamin moiety

indicates that the aorimycin BGC should contain a corresponding glycosyltransferase and necessary deoxysugar biosynthetic genes [16]. The genome of the strain was already previously sequenced and is available in GenBank (QEIK00000000.1). The secondary metabolite biosynthesis gene cluster of the producing strain was examined using the bioinformatic online tool antiSMASH 6.0 [17]. One type II PKS BGC was found containing also an UDP-glycosyltransferase encoding gene. In order to perform knockouts and validate the identity of the BGC a cosmid library was constructed. The whole library was partially sequenced, mapped to the genome and two cosmids (12-A05 and 06-C09) were identified, covering the region of the putative cluster. To gain the information if the determined cluster is responsible for the aorimycin biosynthesis the gene *aor35*, which is coding for an UDP-glycosyltransferase, was knocked out via Red/ET. The finished construct for the knockout was transferred in *Streptomyces sp.* IB2014/011-12 by conjugation. The screening for double crossing over containing strains was simplified by the presence of the *gusA* gene and the hygromycin resistance cassette on the final construct. Strains with a double crossover were sensitive to hygromycin and showed no blue colour when cultivated on MS agar plates containing X-gluc. Positive colonies were additional screened by colony PCR. Those knockout strains were checked for validity, cultivated as mentioned above and screened for the presence of aorimycin (Figure 2).

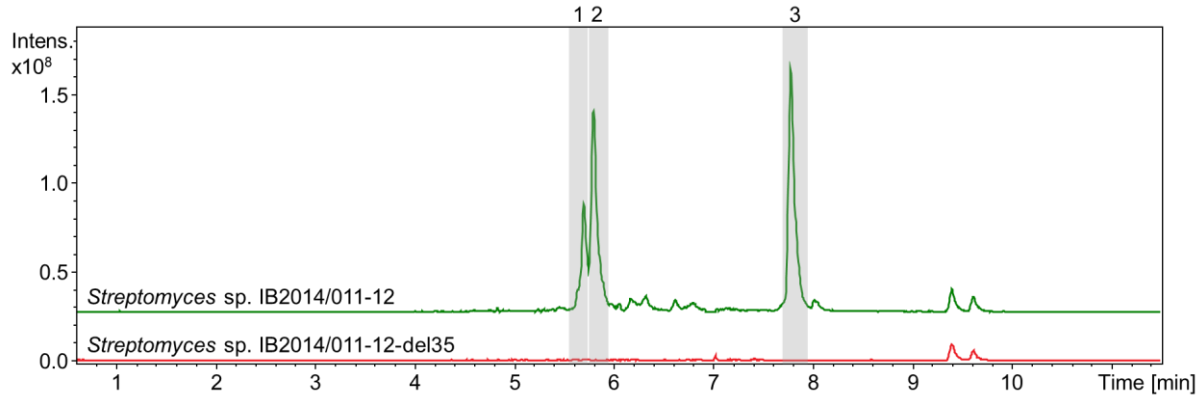

Figure 2: Extracted mass of aorimycin A and B from the HPLC-MS chromatograms of *Streptomyces* sp. IB2014/011-12 and deletion mutants cultivated in DNPM medium: Green *Streptomyces* sp. IB2014/011-12; Red *Streptomyces* sp. IB2014/011-12-del35. Visible masses [M+H]: 1: 574.22 Da; 2/3: 560.20 Da.

As shown in figure 2 there was no aorimycin production after deletion of gene *aor35*. This proves that the proposed cluster is responsible for the biosynthesis of aorimycins. Many genes, encoding for cyclases, ketoreductases, oxygenases as well as putative responsible KSs, were located upstream of the region covered by cosmid 12-A05. Therefore, it was fused by

the means of TAR in yeast with the cosmid 06-C09, covering the remaining region (*orf-3* until *aor31*). The genetic organisation of the aorimycin BGC as it is present on BAC-11-12-07 is shown in figure 3.

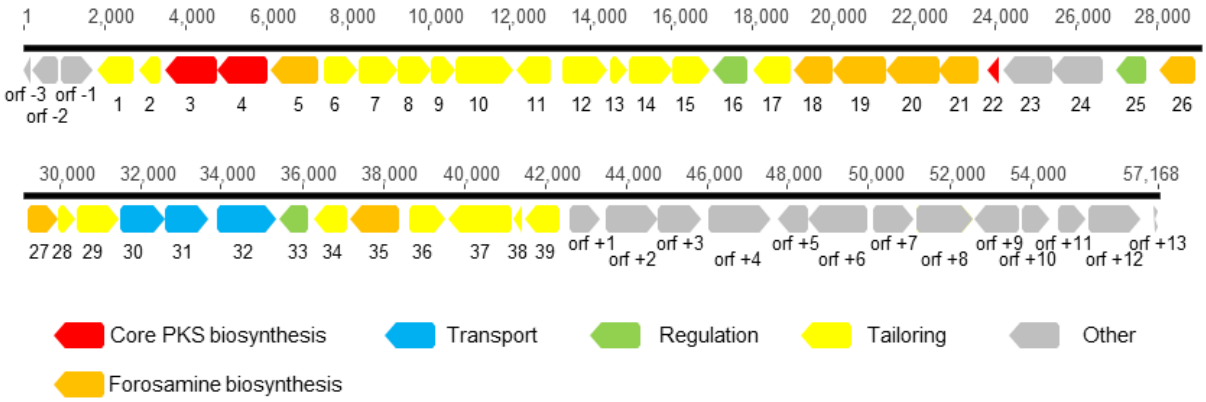

Figure 3: Organisation of the aorimycin biosynthesis gene cluster. Red: Core PKS; Orange: D-forosamine; Yellow: Tailoring enzymes; Green: Regulation; Grey: Not relevant for aorimycin biosynthesis.

The obtained BAC-11-12-07 was used for heterologous expression of the aorimycin BGC in the heterologous host *S. albus* J1074 Del14.

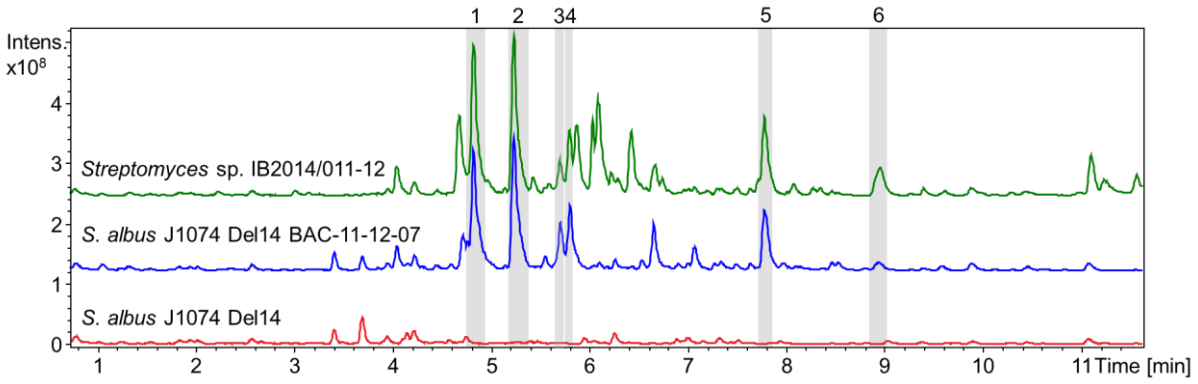

Figure 4: HPLC-MS chromatograms of different *Streptomyces* strains cultivated in DNPM medium: Green *Streptomyces* sp. IB2014/011-12; Blue *S. albus* J1074 Del14 BAC-11-12-07; Red *S. albus* J1074 Del14. Visible masses [M+H] are: 1: 576.20 Da; 2: 576.20 Da; 3: 574.22 Da; 4: 560.20 Da; 5: 560.20 Da; 6: 367.20 Da.

S. albus J1074 Del14 containing BAC-11-12-07 shows a similar metabolic profile to *Streptomyces* sp. IB2014/011-12 (Figure 4) and is thereby proven to produce aorimycins. Furthermore, other related new peaks with the masses [M+H] 576.20 and 367.20 Da were identified in *S. albus* J1074 Del14 BAC-11-12-07. The benefits of the heterologous expression were that unlike the wildtype strain, *S. albus* J1074 Del14 could be cultivated in defined medium. In combination with an easier accessibility for genetic manipulation and a cleaner metabolic background this made further studies of the biosynthesis and purification of

intermediates easier. Based on the heterologous production of aorimycins, it could be determined that BAC-11-12-07 contains all necessary genes for the biosynthesis. The proposed function of those genes is described in table 1 and additional open reading frames (ORFs) in table S3.

Table 1: Predicted function of the proteins encoded by the aorimycin BGC

Gene	Size [bp]	Function	Homologous Gene	Accession number	ID [%]	Coverage [%]
1	915	Cyclase	*HrsI1*	QBK46630.1	69	100
2	618	Flavin reductase	*HrsO*	QBK46639.1	48	61
3	1266	KS-Beta	*OxyB*	AAZ78326.1	67	100
4	1272	KS-Alpha	*PgaA*	AAK57525.1	75	98.6
5	1152	dTDP-D-glucose-4-aminotransferase	*HrsE*	QBK46623.1	71	98.4
6	786	Ketoreductase	*PgaD*	AAK57528.1	84	100
7	951	Cyclase/Dehydrase	*PgaL*	AAK57529.1	79	99.4
8	846	Ketoreductase	*Aur1G*	AAX57194.1	32	71.2
9	579	Hypothetical protein	Hypothetical protein	AAK61714.1	60.61	34
10	1455	Oxidase	*XanO9*	ADE22318.1	45	100
11	825	Oxidoreductase	*Orf23* (Aranciamycin)	ABL09970.1	35	90.9
12	1128	oxygenase-dehydratase	*LanZ5*	AAD13564.1	44	100
13	444	Cyclase	*Arm4*	AHA81974.1	41	97.3
14	1071	Monooxygenase	*RslO1*	AHL46716.1	60	98
15	945	3-hydroxybutyryl-CoA dehydrogenase	*HrsK4*	QBK46641.1	65	95.5
16	816	SARP/transcriptional regulator	*HrsH*	QBK46629.1	69	96.3
17	933	Thioesterase	*May4*	AVO00803.1	56	89.7
18	966	dTDP-D-glucose 3-ketoreductase	*LanT*	AAD13550.1	68	92.8
19	1323	dTDP-D-glucose-2,3-dehydratase	*Aur1S*	ACK77745.2	79	97
20	1305	dTDP-D-glucose-3,4-dehydratase	*LanQ*	AAD13547.1	84	99.8
21	1029	dTDP-D-glucose-4,6-dehydratase	*PgaH1*	AHW57787.1	80	94.2
22	270	ACP	*PgaC*	AAK57527.1	71	96.6
23	1218	KS-Beta	*PgaB*	AAK57526.1	64	100
24	1266	KS-Alpha	*PgaA*	AAK57525.1	70	100
25	765	Response regulator	*PgaR1*	AHW57766.1	67	86.2
26	897	dTDP-1-glucose_synthase	*AclY*	BAB72036.1	72	96
27	714	dTDP-desosamine-N-dimethyltransferase	*EryCVI*	CAM00059.1	57	99.6
28	417	PPOX class F420-dependent oxidoreductase	*AzicX*	ADB02839.1	52	80.2
29	1050	Monooxygenase	*MsnO8*	AHL46694.1	50	100
30	1074	HAF domain containing protein	exported protein of unknown function	CAD5931489.1	99.72	100
31	1080	PKD domain containing protein	PKD domain containing protein	CEL20606.1	57.45	90
32	1494	Transporter/Permease	*ElmE*	CAP12594.1	39	100
33	681	TetR Regulator	*Aur1R*	ADM72853.2	56	88.5
34	831	Monooxygenase	*kstA11*	AFJ52679.1	51	101.8
35	1311	UDP-glucosyl transferase	*apoGT1*	AEP40933.1	37	88.5
36	921	Oxidoreductase	*Lct43*	ABX71126.1	34	101.6
37	1566	acetyl-CoA carboxylase/Carboxyltranferase	*SprU*	BAV17019.1	93	94
38	198	Acyl-CoA carboxylase epsilon subunit	Acyl-CoA carboxylase epsilon subunit	MYW30096.1	95.38	100
39	897	Cyclase/Dehydratase	*FlsD*	ALJ99851.1	37	100

Deletion experiments and biosynthesis

To gain deeper insights into the complex biosynthesis and cyclisation steps of aorimycin, knockout experiments were performed. For each selected gene a hygromycin resistance cassette was amplified with the respective primer pairs (Table S2), designed for homologous recombination. The amplified cassette was used to replace the native gene in the BAC-11-12-07 via Red/ET. Following, 16 mutants based on BAC-11-12-07 with the following deleted genes or multiple genes, were constructed: *orf-1* until *orf-3*; *aor1* until *orf-3*; *aor4*; *aor5*; *aor7*; *aor9*; *aor12*; *aor13*; *aor15*; *aor18*; *aor24*; *aor31* until *orf+13*; *aor35*; *aor39* until *orf+13*; *orf+1* until *orf+13*; *orf-1* until *orf-3* and *orf+1* until *orf+13*. After validation of the deletions, the recombinant BACs were conjugated into *S. albus* J1074 Del14, the respective strains cultivated and the produced secondary metabolites were analysed with HPLC-MS. The results of the deletion experiments are summarised in table S4.

In order to determine the borders of the aorimycin BGC we have performed several deletions of the flanking regions. Deletion of the regions *orf-1* until *orf-3* and *orf+1* until *orf+13* resulted in the same shifted production spectrum of aorimycins with an increased production of two previously barely expressed intermediates. Combination of both deletions in one vector resulted in the same production spectrum. Further deletions of *aor1* or *aor39* led to new intermediates and clearly reduced aorimycin A and B production. For the deletion of *aor1* no aorimycin production was visible at all (Figure S12). Based on those deletions the minimal *aor* BGC could be determined ranging from *aor1* until *aor39*. Most likely some regulatory elements are located upstream as well as downstream of this region, resulting in a changed transcription and therefore a changed production profile once the flanking regions are deleted.

Exceptionally, within the minimal *aor* BGC, we observed two sets of KSα and KSβ encoding genes (*aor3*, *aor4* and *aor23*, *aor24*) supposed to control the biosynthesis of the core polyketide chain (Figure 3/Table 1). Typically for the biosynthesis of polyketide type II compounds, only one pair of KSs and one ACP are responsible for the synthesis of the core polyketide. Therefore, genes *aor4* and *aor24* were deleted and the production of aorimycins was analysed. It could be clearly seen that after deletion of *aor4* no aorimycin production was detectable. In contrast, a deletion of *aor24* did not influence the aorimycin production. The strain containing the modified BAC with this deletion showed a similar production profile as the control strain (Figure S13). Hence, indicating that either another NP is produced by these KSs or perhaps they were integrated there randomly via crossover or another type of

recombination. This finding is surprising since aor22, 23 and 24 represent classical arrangement of minimal PKSII gene set in contrast to aor3 and 4, where ACP encoding gene is missing or at least not in close proximity.

D-forosamine biosynthesis has been previously analysed and corresponding responsible genes have been identified [16]. *Aor35* encoding for a UDP-glycosyltransferase has been previously deleted to prove the identity of the gene cluster. By analysis of the BGC in total 8, to the described genes needed for D-forosamine biosynthesis, homologous genes have been found (*aor26*, *aor21*, *aor19*, *aor18*, *aor20*, *aor5*, *aor27* and *aor35*). However, *aor5* and *aor18* both showed similarity to a gene encoding for the dTDP-D-glucose-4-aminotransferase. To analyse this redundancy we have deleted them both independently from the *aor* cluster. When the deletion strains are analysed, it can be seen that the deletion of *aor18* only leads to a shifted production of aorimycins and no abolishment (Figure S14). Therefore, it can be inferred that this gene is not essential for the biosynthesis of aorimycins. Contrary to this the deletions of *aor5* and *aor35* looked very similar and led to no visible aorimycin production. This proves that *aor5* is catalysing a more specific reaction during D-forosamine biosynthesis. Further analysis of closer related clusters within the MIBIG database also supported those findings, and show similarities of *aor5* to dTDP-glucose-4-aminotransferase and of *aor18* to dTDP-D-glucose ketoreductase.

Taking the above mentioned findings into account, as well as the analysis of other gene knockouts (Table S4) and purified compounds figure 1, the biosynthesis was postulated (Figure 5). As for polyketide type II compounds typical, the KSα, KSβ and ACP work together to condense in this case 9 malonyl-CoA and 1 acetly-CoA to the core polyketide chain (Figure 5; A). Deletion of *aor7* led to the detection of new compounds with masses [M+H] of 405.20, 369.20, 371.20, 387.20 and 367.09 Da. The *aor7* mutant did not produce any final compounds (Figure S15). It was further observed that one peak of the respective mass [M+H] 405.20 Da ranged over a retention time of 40 seconds. Thereby, indicating the presence of either one compound with high hydrophobicity or several different compounds with a high similarity. In other studies the deletion of cyclases responsible for the first cyclisation led to formation of shunt products with random cyclisation [18]. Further comparison of the genes *aor6* and *aor7* with similar genes of known function from related BGCs showed similarities to enzymes responsible for the first ring cyclisation. Thus the first

ring cyclisation of the core polyketide is suggested to be catalysed by the C-9 ketoreductase and a cyclase/dehydrase encoded by *aor6* and *aor7* (Figure 5, B).

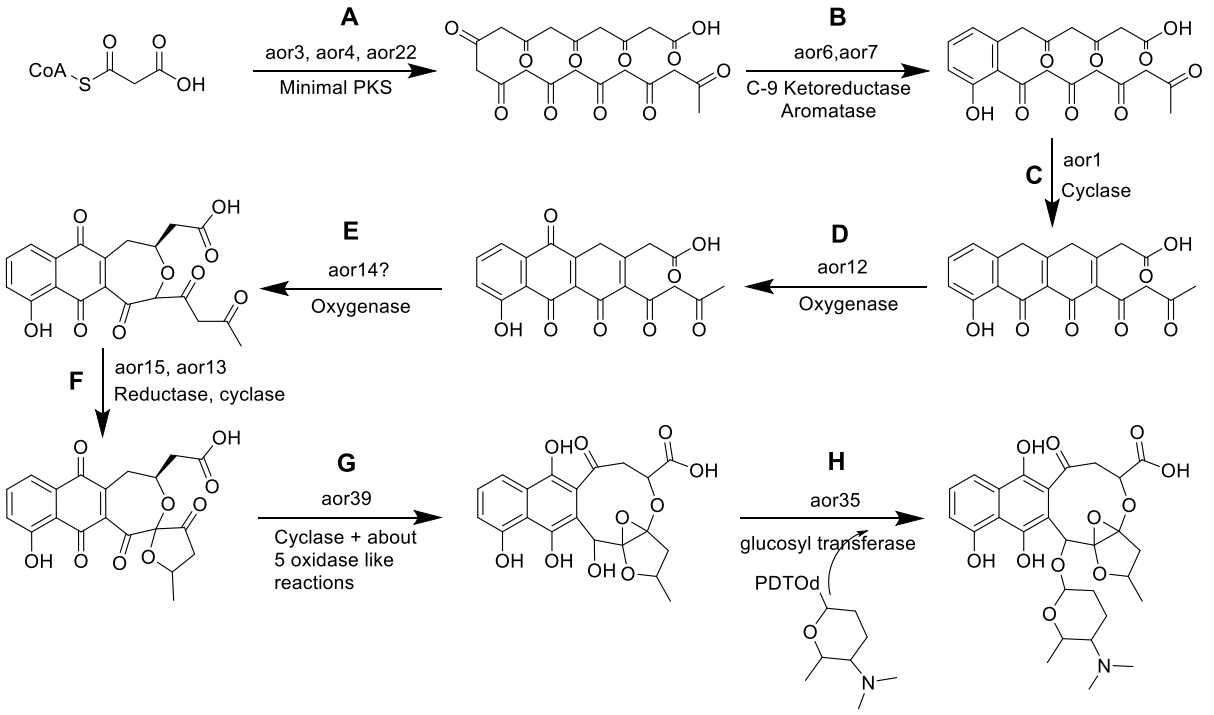

Figure 5: Proposed biosynthesis pathway of aorimycins. Based on deletion experiments and purified intermediates thereof, in combination with bioinformatic analysis (AntiSMASH; MIBIG; NCBI-Blast) of the genes the biosynthesis was proposed.

The third step (Figure 5, C) is a cyclisation of the second and third ring catalysed by the cyclase Aor1. Analysis of mutants with a deletion of *aor1* led to no observable production of aorimycin, instead intermediates with masses of [M+H] 371.20, 367.09 and 369.20 Da were accumulated. Contrary to the deletion of *aor7*, knockout of *aor1* (Figure S12) led to fewer intermediates, indicating that the possibility of random cyclisation has been already narrowed, by the closure of the first ring. By comparing *aor1* to other genes of similar BGCs, *aor1* showed similarity to genes encoding for 2,3-cyclases, responsible for second and third ring formation. Subsequently we propose that the second ring gets oxidised by the bifunctional oxygenase/dehydratase Aor12 (Figure 5, D). Deletion of *aor12* and cultivation of the mutant resulted in a similar production spectrum as the deletion of *aor1* (Figure S16). Blast analysis of this gene showed similarities to either a hydrolase or an oxygenase. Med-Orf7 is one of those similar enzymes and is proposed to introduce an oxygen-group at the second ring [19]. Furthermore, no intermediate was isolated without two oxygens present at the second ring, which is why we postulate the oxygenation of the second ring, right after the formation of ring two and three.

Aorimycin-Int-1-5 (Figure 1, F-I) were isolated, showing that the third ring was oxidised to a 7-membered ring. Through bioinformatic comparison of possible candidate genes for such a Baeyer-Villiger reaction (monooxygenases and oxygenases) *aor14* encoding for a monooxygenase was found to be the most suiting. Aor14 showed similarities to monooxygenases catalysing a similar reaction, like RslO9 [20] or UrdM [21] (Figure 5, E).

Deletion of *aor13* (cyclase) and *aor15* (oxidoreductase) showed similar metabolic profiles (Figures S17, S18). However, the mutant with a knockout of *aor15* showed no detectable production of final aorimycins A/B and the one with *aor13* deletion did show low quantities thereof. The isolated compound aorimycin-Int-1 (Figure 1, F) shows a 5-membered ring attached to the third (7-membered) ring. When comparing the ketoreductase encoding gene *aor15* with genes of known BGCs, similarities to genes *med-ORF12* [22] and *ActVI-Orf1* [23] are apparent. Those genes reduce a Keto-group to a hydroxyl-group, to make it accessible for a cyclisation reaction. If the activation of the keto-group cannot be performed, no cyclisation of the 5-membered ring can be performed, leading to no aorimycin production. However, once the reduction can be performed but the gene encoding for the cyclase (*aor13*) is deleted low quantities of aorimycins are detectable. This indicates that with less efficiency the cyclisation can happen spontaneously or is catalysed by a different enzyme, after the activation step by Aor15. Hence, we suggest the involvement of Aor13 and Aor15 in the 5-membered ring formation (Figure 5, F). These findings were supported by the presence of Aorimycin-Int-1 (Figure 1, F) with a 7- and 5-membered ring, indicating that these reactions should be catalysed prior to the final formation of the 9-membered ring.

Following the cyclisation to a 9-membered ring by involvement of the cyclase Aor39 is performed (Figure 5, G). *S. albus* J1074 Del14 containing the cosmid 06-C09 (deletion of *aor31* until *orf+13*) was cultivated and the production spectrum looked nearly identical to the one of the mutant containing BAC-11-12-07 with the deletion of *aor39*. However, closer analysis revealed the difference of the intermediates. By comparing the intermediates it could be proposed that the genes *aor34* (monooxygenase) and *aor36* (oxidoreductase) could be responsible for the modifications to obtain the 9-membered ring. After deletion of *aor39* still aorimycin-Int-5 and -6 (Figure 1, D-E) could be isolated and aorimycins A and B are still present in the extract. This shows that the cyclase Aor39 is important for the 9-ring formation, but can be replaced by other enzymes or spontaneous cyclisation reactions. Most likely the keto-group needs to be activated by an oxygenase, like the activation by Aor15 and the

following cyclisation by Aor13. Additionally, the epoxy-group needs to be attached and the ring system rearranged. We estimate the involvement of genes *aor28* and *aor29* in these reactions.

At last we propose the glycosylation by Aor35, to form the final aorimycin (Figure 5, H). Mutants with deletions of the genes *aor35* and *aor5* have nearly identical metabolic profiles. This shows that only the final activated D-forosamine is recognised by the glycosyltransferase as a sugar donor. The biosynthetic genes for the D-forosamine formation have been previously described. Unlike *aor35*, *aor18* and *aor5* they only have been validated by bioinformatic analysis. The biosynthesis of D-forosamine is shown in figure S20.

Additionally isolated compounds aorimycin-Int-2,-3 and -4 can be originating from a detoxification process of the heterologous host itself. This is especially the case of aorimycin-Int-4 which looks very much like the interference of the well described detoxification with mycothiol [24]. Aorimycin-Shu looks to be originating from a premature termination of the biosynthesis chain or a degradation product from some intermediates. The structure looks similar to those of shunt products after the deletion of enzymes responsible for the first cyclisation steps [18, 25]. This could be originating from an unbalanced expression of the in the biosynthesis involved genes or not properly sufficient precursor supply.

2.2.5 Conclusion

In this article we report the identification of a new compound class with for type II polyketides unique 9-membered ring. Additionally, we were able to identify and express the BGC heterologously. Based on deletion experiments of several genes the minimal BGC was determined and the biosynthesis was proposed.

2.2.6 References

1. Katz, L. and R.H. Baltz, *Natural product discovery: past, present, and future.* Journal of Industrial Microbiology and Biotechnology, 2016. **43**(2-3): p. 155-176.
2. Shen, B., *Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms.* Current opinion in chemical biology, 2003. **7**(2): p. 285-295.
3. Yushchuk, O., et al., *Landomycin biosynthesis and its regulation in Streptomyces.* Applied microbiology and biotechnology, 2019. **103**(4): p. 1659-1665.
4. Zhang, W., et al., *Engineered biosynthesis of a novel amidated polyketide, using the malonamyl-specific initiation module from the oxytetracycline polyketide synthase.* Applied and environmental microbiology, 2006. **72**(4): p. 2573-2580.

5. Vasanthakumar, A., K. Kattusamy, and R. Prasad, *Regulation of daunorubicin biosynthesis in Streptomyces peucetius–feed forward and feedback transcriptional control.* Journal of basic microbiology, 2013. **53**(8): p. 636-644.
6. Yang, X., B. Fu, and B. Yu, *Total synthesis of landomycin A, a potent antitumor angucycline antibiotic.* Journal of the American Chemical Society, 2011. **133**(32): p. 12433-12435.
7. Di Marco, A., G. Cassinelli, and F. Arcamone, *The discovery of daunorubicin.* Cancer treatment reports, 1981. **65**: p. 3-8.
8. Zhang, Z., H.-X. Pan, and G.-L. Tang, *New insights into bacterial type II polyketide biosynthesis.* F1000Research, 2017. **6**.
9. Hertweck, C., et al., *Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork.* Natural product reports, 2007. **24**(1): p. 162-190.
10. Katsuyama, Y., et al., *Involvement of the Baeyer–Villiger Monooxygenase IfnQ in the Biosynthesis of Isofuranonaphthoquinone Scaffold of JBIR-76 and-77.* ChemBioChem, 2016. **17**(11): p. 1021-1028.
11. Kieser, T., et al., *Practical streptomyces genetics*. Vol. 291. 2000: John Innes Foundation Norwich.
12. Kouprina, N. and V. Larionov, *Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology.* Chromosoma, 2016. **125**(4): p. 621-632.
13. Voitsekhovskaia, I., et al., *Baikalomycins AC, New Aquayamycin-Type Angucyclines Isolated from Lake Baikal Derived Streptomyces sp. IB201691-2A.* Microorganisms, 2020. **8**(5): p. 680.
14. Lasch, C., et al., *Bonsecamin: A New Cyclic Pentapeptide Discovered through Heterologous Expression of a Cryptic Gene Cluster.* Microorganisms, 2021. **9**(8): p. 1640.
15. Sohng, J.K., et al., *Production, isolation and biological activity of nargenicin from Nocardia sp. CS682.* Archives of pharmacal research, 2008. **31**(10): p. 1339-1345.
16. Hong, L., et al., *In vitro characterization of the enzymes involved in TDP-D-forosamine biosynthesis in the spinosyn pathway of Saccharopolyspora spinosa.* Journal of the American Chemical Society, 2008. **130**(14): p. 4954-4967.
17. Blin, K., et al., *antiSMASH 6.0: improving cluster detection and comparison capabilities.* Nucleic acids research, 2021: p. 1.
18. Zawada, R.J. and C. Khosla, *Domain analysis of the molecular recognition features of aromatic polyketide synthase subunits.* Journal of Biological Chemistry, 1997. **272**(26): p. 16184-16188.
19. Ichinose, K., et al., *Cloning, sequencing and heterologous expression of the medermycin biosynthetic gene cluster of Streptomyces sp. AM-7161: towards comparative analysis of the benzoisochromanequinone gene clusters.* Microbiology, 2003. **149**(7): p. 1633-1645.
20. Tsypik, O., et al., *Oxidative carbon backbone rearrangement in rishirilide biosynthesis.* Journal of the American Chemical Society, 2020. **142**(13): p. 5913-5917.
21. Tolmie, C., M.S. Smit, and D.J. Opperman, *Native roles of Baeyer–Villiger monooxygenases in the microbial metabolism of natural compounds.* Natural product reports, 2019. **36**(2): p. 326-353.
22. He, Q., et al., *Functional characterization of a ketoreductase-encoding gene med-ORF12 involved in the formation of a stereospecific pyran ring during the biosynthesis of an antitumor antibiotic medermycin.* PloS one, 2015. **10**(7): p. e0132431.

23. Ichinose, K., et al., *Functional complementation of pyran ring formation in actinorhodin biosynthesis in Streptomyces coelicolor A3 (2) by ketoreductase genes for granaticin biosynthesis.* Journal of bacteriology, 2001. **183**(10): p. 3247-3250.
24. Newton, G.L., Y. Av-Gay, and R.C. Fahey, *A novel mycothiol-dependent detoxification pathway in mycobacteria involving mycothiol S-conjugate amidase*. Biochemistry, 2000. **39**(35): p. 10739-10746.
25. Bräuer, A., et al., *Structural snapshots of the minimal PKS system responsible for octaketide biosynthesis.* Nature Chemistry, 2020. **12**(8): p. 755-763.

2.2.7 Supplementary

Table S1: Bacterial strains and plasmids

Bacterial strains	Description	Source or reference
E. coli ET12567 (pUB307)	Conjugative transfer of DNA	[1]
E. coli GB05-redCC	Derivative of GB2005 containing integration of PBAD-ETgA operon	[2]
S. albus J1074 Del14	*Streptomyces* strain used for heterologous expression of different vectors	[3]
Streptomyces sp. IB2014/011-12	WT strain isolated as a symbiont from *Trichoptera* sp. (larvae)	[4]
Saccharomyces cerevisiae BY4742	Auxotrophic strain used for homologous recombination in yeast	[5]
Vectors	**Description**	**Source or reference**
cos15A-gus	Synthetic vector used for cosmid library construction	To be submitted
12-A05	Derivative of cos15A-gus with inserted chromosomal DNA of *Streptomyces* sp. IB2014/011-12 covering the region *aor16* until *orf+13*	This study
06-C09	Derivative of cos15A-gus with inserted chromosomal DNA of *Streptomyces* sp. IB2014/011-12 covering the region *orf-3* until *aor31*	This study
12-A05-del35	Derivative of 12-A05 with deleted gene *aor35* and chloramphenicol resistance	This study
pSMART_cloni	Derivative of pSMART-BAC for amplification of larger insert DNA	To be submitted
pCC1Bac	Chloramphenicol resistance containing vector	Epicentre
BAC-11-12-07	pSMART_cloni with inserted chromosomal DNA of *Streptomyces* sp. IB2014/011-12 covering the region *orf-3* until *orf+13*	This study
BAC-11-12-07-delorf-3-orf-1	BAC-11-12-07 with deleted region *orf-3* until *orf-1*	This study
BAC-11-12-07-del1	BAC-11-12-07 with deleted region *orf-3* until *aor1*	This study
BAC-11-12-07-del4	BAC-11-12-07 with deletion of *aor4*	This study
BAC-11-12-07-del5	BAC-11-12-07 with deletion of *aor5*	This study
BAC-11-12-07-del7	BAC-11-12-07 with deletion of *aor7*	This study
BAC-11-12-07-del9	BAC-11-12-07 with deletion of *aor9*	This study
BAC-11-12-07-del12	BAC-11-12-07 with deletion of *aor12*	This study
BAC-11-12-07-del13	BAC-11-12-07 with deletion of *aor13*	This study
BAC-11-12-07-del15	BAC-11-12-07 with deletion of *aor15*	This study
BAC-11-12-07-del18	BAC-11-12-07 with deletion of *aor18*	This study
BAC-11-12-07-del24	BAC-11-12-07 with deletion of *aor24*	This study
BAC-11-12-07-del35	BAC-11-12-07 with deletion of *aor35*	This study
BAC-11-12-07-del39	BAC-11-12-07 with deleted region *aor39* until *orf+13*	This study
BAC-11-12-07-delorf+1-orf+13	BAC-11-12-07 with deleted region *orf+1* until *orf+13*	This study
Minimal BAC-11-12-07 (delorf-3-orf-1+orf+1-orf+13)	BAC-11-12-07 with deleted regions *orf-3* until *orf-1* and *orf+1* until *orf+13*	This study

Table S2: Oligonucleotides used in this study

Pair Nr.	Name	Sequence (5'→3')	Purpose
1	del-o3-1-F	GCGGTTTCACGTTACCCATTGTGTAACTATAACGGTCCTA AGGTAGCGAAAGATCTGTTTTGTTTAAACAAAATACTTGA CATATCACTG	Primer for replacement of the upstream region *orf-3* until *orf-1* with a Hyg^R for deletion experiments in BAC-11-12-07
	del-o3-1-R	GGCCGCGCTGGCTGGCGACCGGCGGGCCCGGGGAACGCC GGGCGGGGCAGCCCGCCCCTAGTTTAAACCGCCGGGATG TATCAGGCGCC	
2	del-o3-1-aor1-F	GCGGTTTCACGTTACCCATTGTGTAACTATAACGGTCCTA AGGTAGCGAAAGATCTGTTTGTTTAAACCGCCGGGATGTA TCAGGCGCC	Primer for replacement of the upstream region *orf-3* until *aor1* with a Hyg^R for deletion experiments in BAC-11-12-07
	del-o3-1-aor1-R	TCGCGGGTCCCCGGCTCCCACGCCGGCCGAGCACACGAG CAAGGGACCGTGACCGTCATGTGTTTAAACAAAATACTTG ACATATCACTG	
3	del-aor4-F	CGGCGGTCTTCGGGGCGGGAGCGGCGTCCCCAAAGGGGC GGGAGAGTGCCTCGGCGGTCACGCCGGGATGTATCAGGC GCC	Primer for replacement of *aor4* with a Hyg^R for deletion experiments in BAC-11-12-07
	del-aor4-R	GTTCCCTCCCGTACCGAACCGACCGCATCCGGAGAGCCTC ACCGTGAAGGACACCCCATGAAATACTTGACATATCACTG	
4	del-aor5-F	CCTTCACGGTGAGGCTCTCCGGATGCGGTCGGTTCGGTAC GGGAGGGAACGCCGGGCTCACGCCGGGATGTATCAGGCG CC	Primer for replacement of *aor5* with a Hyg^R for deletion experiments in BAC-11-12-07
	del-aor5-R	CGGTGTCGAGCGGCCCTTCTCAAGCTCGACTGCGACATCA GGACTTGAGAGGCGGCCATGAAATACTTGACATATCACT G	
5	del-aor7-F	ACTTCTGACCGCACCGCAGCCGGACGACGAACCGTTCCCG ACCAAGGAGTGCCCGCCATGGCTAGCACGTCAGGTGGCA CTTTTCGG	Primer for replacement of *aor7* with a Hyg^R for deletion experiments in BAC-11-12-07
	del-aor7-R	CGTGACCGAGCCCCCGGGAGGCTCCGGTGACGAACCAGC AGCGTGACATGGGCCGGCTCAGCTAGCTTACCAATGCTTA ATCAGTGA	
6	del-aor9-F	GGAGGCCTGGAACCGGCTGGCCCGCACCACGGACTCACC CCCGCCGGAGGCGTCGTGAGGGCTAGCACGTCAGGTGGC ACTTTTCGG	Primer for replacement of *aor9* with a Hyg^R for deletion experiments in BAC-11-12-07
	del-aor9-R	GGAGGGCGGCGGCGCCCGCGGCGGCGGTGATCTGGAGCG CGTGACGTCGGTTCAGCATCAGCTAGCTTACCAATGCTTA ATCAGTGA	
7	del-aor12-F	TCCCATGGCCTGCCCGAGGTGCCCGCGTCACCGCGACGAT GGAAGAGAGTGGTGCTGATGGCTAGCACGTCAGGTGGCA CTTTTCGG	Primer for replacement of *aor12* with a Hyg^R for deletion experiments in BAC-11-12-07
	del-aor12-R	ATGCGCACTGCCTTTCCTGAGTACGCCGGAGAACCGGCTG CGGTTGCTGGTCTGAGGTCAGCTAGCTTACCAATGCTTAA TCAGTGA	
8	del-aor13-F	GACCTCAGACCAGCAACCGCAGCCGGTTCTCCGGCGTACT CAGGAAAGGCAGTGCGCATGGCTAGCACGTCAGGTGGCA CTTTTCGG	Primer for replacement of *aor13* with a Hyg^R for deletion experiments in BAC-11-12-07
	del-aor13-R	CGGACCTACCGTGGGAAAGAAGACGATCCCGAATTTCAT GGCGGCTCCTCAGCCTCGTCAGCTAGCTTACCAATGCTTA ATCAGTGA	
9	del-aor15-F	GACCATGCGGCTGATGGCCGAACAGGTGCTGCCCAAGCT GAACGGAGGCCGCGGATGAGCGCTAGCACGTCAGGTGGC ACTTTTCGG	Primer for replacement of *aor15* with a Hyg^R for deletion experiments in BAC-11-12-07
	del-aor15-R	GGAACCGTGAAGGTGTCCGGCACACGGTCCGGGGGGTCA CTCCTCCGGCCCGCGGGGTCAGCTAGCTTACCAATGCTTA ATCAGTGA	
10	del-aor18-F	CGCTCCTGCCGGTGACGGTGGTCAGGGGACGGTACGTGT GGGGGCGGCGATCCCGGCTCATGTAACGCACTGAGAAGC CC	Primer for replacement of *aor18* with a Hyg^R for deletion experiments in BAC-11-12-07
	del-aor18-R	GAACGTCCAGGCCCGCACGCTCCTGTCCTGCCTGACCACG GGGGCGGTACGGCTCTGATGTCTTGTAGGCTGGAGCTGCT TC	
11	del-aor24-F	TCACCCCCGTGCCGCTGGGGGCCAGGACACCGATTCCGGT GATCACCGCTCCGGCGCTCATGTAACGCACTGAGAAGCCC	Primer for replacement of *aor24* with a Hyg^R for deletion experiments in BAC-11-12-07
	del-aor24-R	GATGCCCCGACGCGAAGTGAGCGCGCACCTCACCCGGAC GACATCGGGACGGTACGCATGTCTTGTAGGCTGGAGCTGC TTC	
12	del-aor35-F	AATGGATTCTTGAACCTCAATCACAATTGAGGTCAAGTCG AAGTGAAGTGATGTGCATCACGCCGGGATGTATCAGGCG	Primer for replacement of *aor35* with a Hyg^R for deletion experiments in BAC-11-

		CC	12-07
	del-aor35-R	TTCGAGTCCGCGTCGAGCGGAGGCCTCAAGGATCGTCGG GTCACCGGTGTGTACCGGATGAAATACTTGACATATCACT G	
13	del+o1-13-aor39F	GTCCCATCACGGTCTCCGCTGCCCCTCGACACCCGCTCGC ACCTTTCTGGAGCCGGCTCAGTTTAAACCGCCGGGATGTA TCAGGCGCC	Primer for replacement of the downstream region *aor39* until *orf+13* with a Hyg^R for deletion experiments in BAC-11-12-07
	del+o1-13-aor39R	TATCCTATAAATATAACGTTTTTGAACACACATGAACAAG GAAGTACATTAATTAAGTTTTGTTTAAACAAAATACTTGA CATATCACTG	
14	del+o1-13F	CGGCGGTTGACTGACGCCTGATCAACGCTGTCACGCTTCG TCCCGGGAGACACCTTGTTGTGTTTAAACAAAATACTTGA CATATCACTG	Primer for replacement of the downstream region *orf+1* until *orf+13* with a Hyg^R for deletion experiments in BAC-11-12-07
	del+o1-13R	TATCCTATAAATATAACGTTTTTGAACACACATGAACAAG GAAGTACATTAATTAAGTTTGTTTAAACCGCCGGGATGTA TCAGGCGCC	
15/16	del-up-c-F	ACCAACTCGTGTAGGTCC	Primer pairs for checking for deletion of upstream region *orf-3* until *orf-1* and *orf-3* until *aor1*
	del-o3-1-c-R	ATGATCACCTTCCACGGG	
	del-o-3-aor1-c-R	TCTCCCGAATCGCGGGTC	
17	del-aor4-c-F	ATGACCTGGTCCCGCAGG	Primer pairs for checking for deletion of *aor4*
	del-aor4-c-R	ATCCGGAGAGCCTCACCG	
18	del-aor5-c-F	TGTCGTAGGCGTGCGGTC	Primer pairs for checking for deletion of *aor5*
	del-aor5-c-R	TCGATCCGGTGTCGAGCG	
19	del-aor7-c-F	TATGTCGAGACGCCCATG	Primer pairs for checking for deletion of *aor7*
	del-aor7-c-R	TTGTTCAGGACGACGTCG	
20	del-aor9-c-F	AGTGCCGAGAGCATGGTG	Primer pairs for checking for deletion of *aor9*
	del-aor9-c-R	TGACGATCTCGACCTGTG	
21	del-aor12-c-F	TCATACGGGTCTCCAAGG	Primer pairs for checking for deletion of *aor12*
	del-aor12-c-R	TTCGAGCCGCTTCTTGCG	
22	del-aor13-c-F	TCAGAACATGATCGCGGG	Primer pairs for checking for deletion of *aor13*
	del-aor13-c-R	AGATGTTGTCGAGCATGG	
23	del-aor15-c-F	ATCCACTTCAGCTACCCG	Primer pairs for checking for deletion of *aor15*
	del-aor15-c-R	AATGGTTCCACTGCCAGC	
24	del-aor18-c-F	GCGGATCTGCGGAGACTC	Primer pairs for checking for deletion of *aor18*
	del-aor18-c-R	CTATCTGAACGTCCAGGCCC	
25	del-aor24-c-F	GGCCAGGACACCGATTCC	Primer pairs for checking for deletion of *aor24*
	del-aor24-c-R	CTCACCCGGACGACATCG	
26	del-aor35-c-F	CAACACATGCTCCTTGTCGG	Primer pairs for checking for deletion of *aor35*
	del-aor35-c-R	CTCAAGGATCGTCGGGTCAC	
27/28	del-aor39-o13-c-R	ATATCGCTCGCATCACTCAC	Primer pairs for checking for deletion of downstream region *aor39* until *orf+13* and *orf+1* until *orf+13*
	del-o1-o13-c-R	ATGACGTATCTCCCGATACC	
	del-down-c-R	GTTAACTGCGGTCAAGATAT	
29	Chlor-F	ATGGAGAAAAAAATCACTGG	Primer pair for amplification of chloramphenicol resistance cassette
	Chlor-R	TTACGCCCCGCCCTGCCAC	

Table S3: Open reading frames (orf) of BAC-11-12-07

Gene	Size [bp]	Function	Homolouge Gene	Accession number	ID [%]	Coverage [%]
orf-3	183	Hypothetical protein	Hypothetical protein	NYS18013.1	88.33	98
orf-2	675	AIM24 domain containing protein	AIM24 domain containing protein	AGK80888.1	93.31	99
orf-1	759	peptidyl-tRNA hydrolase	peptidyl-tRNA hydrolase	NDZ63354.1	98.41	98
orf+1	750	DUF4142 domain-containing protein	DUF4142 domain-containing protein	NDZ63356.1	99.6	100
orf+2	1296	DUF692 domain-containing protein	DUF692 domain-containing protein	NED0670.1	84.47	94
orf+3	1047	TIGR04222 domain-containing protein	TIGR04222 domain-containing protein	NUV90407.1	72.87	73
orf+4	1569	alpha/beta hydrolase fold	alpha/beta hydrolase fold	NDZ63359.1	99.43	99
orf+5	729	Chlorite dismutase family protein	Chlorite dismutase family protein	MBD2830076.1	96.28	99
orf+6	1488	protoporphyrinogen oxidase	protoporphyrinogen oxidase	RLV70358.1	98.18	99
orf+7	1020	DUF4349 domain-containing protein	DUF4349 domain-containing protein	MXG28800.1	81.53	72
orf+8	1362 - 1380	Flavoprotein oxidoreductase	Flavoprotein oxidoreductase	RLV70356.1	96.69	99
orf+9	1110	uroporphyrinogen decarboxylase	uroporphyrinogen decarboxylase	QTA3546704.1	98.10	99
orf+10	660	DUF3000 domain-containing protein	DUF3000 domain-containing protein	NED06752.1	93.18	99
orf+11	663	LuxR;response regulator transcription factor	*WhiI*	AIM19628.1	95.91	99
orf+12	1275	ribonuclease D	ribonuclease D	NDZ63367.1	99.53	99
orf+13	117-129	acetyl-CoA acetyltransferase	acetyl-CoA acetyltransferase	ARX87150.1	92.31	100

Table S4: Visible compounds and corresponding retention time produced by *S. albus* J1074 Del 14 containing the aorimycin BGC with various deletions. - = No production; + = Visible when extracted; ++ = High production

Compound [M+H] Da / Vector with or without deletion	576.20	574.22	560.20	389.20	371.11	400.30	574.22	373.20	560.20	564.08	371.20	530.11	403.08/ 401.07	560.20	288.92	367.09/ 530.10
WT (*Streptomyces* sp. IB2014/011-12)	+	+	+	-	-	-	+	-	+	-	-	-	-	+	-	+
BAC 11-12-07	++	+	++	-	-	-	+	-	+	-	-	-	-	++	-	+
Del *orf -1* until *orf -3*	++	+	+	-	-	-	++	-	++	+	-	-	-	+	-	+
Del *aor1* until *orf -3*	-	-	-	++	++	-	-	+	-	-	+	-	-	-	-	++
Del *aor4*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Del *aor5*	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-	++
Del *aor7*	-	-	-	-	-	-	-	+	-	-	++	-	-	-	-	-
Del *aor9*	++	+	++	-	-	-	++	-	++	+	-	-	-	+	-	+
Del *aor12*	-	-	-	-	-	-	-	++	-	-	-	++	-	-	++	+
Del *aor13*	+	-	-	-	-	-	+	+	+	-	-	-	-	+	-	++
Del *aor15*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	++
Del *aor18*	+	+	+	-	-	-	-	+	+	+	-	-	-	+	-	+
Del *aor24*	+	+	+	-	-	-	++	-	+	+	-	-	-	+	-	+
Del *aor31* until *orf +13*	-	-	-	-	-	+	-	+	-	++	-	-	++	-	-	+
Del *aor35*	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-	++
Del *aor39* until *orf +13*	++	+	+	-	-	+	+	++	++	++	-	-	++	++	-	+
Del *orf +1* until *orf +13*	++	+	+	-	-	-	+	-	++	+	-	-	-	+	-	+
Minimal - BAC 11-12-07	++	+	+	-	-	-	++	-	++	+	-	-	-	++	-	+
Retention time [min]	**5.2**	**5.6**	**5.7**	**5.7**	**5.9**	**5.9**	**6**	**6.2**	**6.3**	**6.6**	**6.9**	**7**	**7.2**	**7.8**	**8.4**	**8.9**

Table S5: NMR data of aorimycin A amd B (DMSO-d_6, ^{1}H: 700 MHz, ^{13}C: 175 MHz)

	Aorimycin A		Aorimycin B	
No	δ(^{13}C)[ppm]	δ(^{1}H) [ppm], mult (J)	δ(^{13}C) [ppm]	δ(^{1}H) [ppm], mult(J)
1	176.75, C	-	172.5, C	-
2	74.17, CH	5.07, t (4.4)	73.15, CH	5.39, dd (8.2, 3.8)
3	39.55, CH_2	3.31, dd (16.5, 4.8) 2.84, dd (16.5, 4.1)	36.70, CH_2	3.20, dd (16.7, 3.8) 2.82, dd (16.6, 8.0)
4	202.17, C	-	200.17, C	-
5	106.81, C	-	106.22, C	-
6	156.68,C	-	157.49, C	-
7	128.92, C	-	128.83, C	-
8	116.55, CH	7.89, dd (8.3, 0.8)	116.77, CH	7.90, dd (8.2, 1.0)
9	129.58, CH	7.48, t (8.1)	129.92, CH	7.50, t (8.0)
10	116.85, CH	7.07, dd (7.7, 0.8)	117.54, CH	7.10, dd (7.9, 1.0)
11	156.39, C	-	156.49, C	-
12	117.59, C	-	117.83, C	-
13	143.02, C	-	143.38, C	-
14	108.42, C	-	107.93, C	-
15	74.05, CH	4.02, s	74.50, CH	4.05, s
16	104.52, C	-	104.76, C	-
17	96.51, C	-	96.83, C	-
18	48.05, CH_2	2.31, dd (13.8, 3.0) 1.96, dd (13.6, 11.7)	48.05, CH_2	2.34, dd (13.9, 2.8) 1.99, dd (13.6, 11.7)
19	67.39, CH	3.64, m	67.69, CH	3.69, m
20	20.92, CH_3	1.19, d (6.1)	20.86, CH_3	1.21, d (6.3)
21	98.09, CH	5.32, d (2.8)	98.49, CH	5.37, ovl
22	30.17, CH_2	1.98, ovl. 1.70, ovl	30.00, CH_2	2.02, ovl 1.73, ovl
23	19.39, CH_2	1.78, ovl.	16.80, CH_2	1.73, ovl
24	68.51, CH	2.91, ovl.	67.53, CH	2.93, ovl
25	65.87, CH	4.52, m	65.79, CH	4.20, m
26	19.37, CH_3	1.34, d (6.2)	18.49, CH_3	1.23, d (6.3)
27/27'	42.17, 2xCH_3	2.77, s	40.40, 2xCH_3	2.66, s
28			52.77, CH_3	3.76, s

*signal overlap

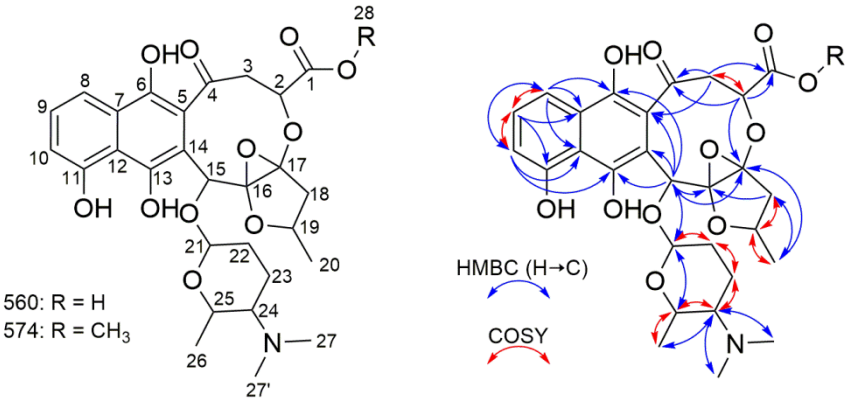

Figure S1: Structures of 560 and 574 with key COSY and HMBC correlations.

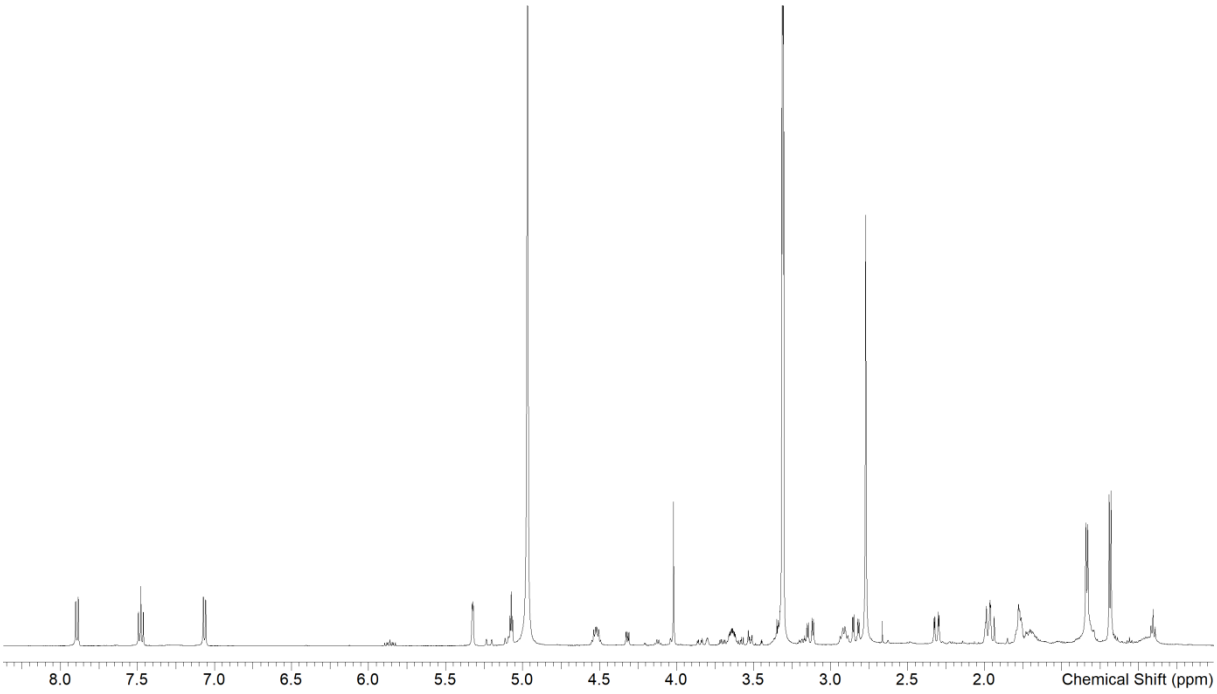

Figure S2: ^{1}H NMR of 560 in CD_3OD.

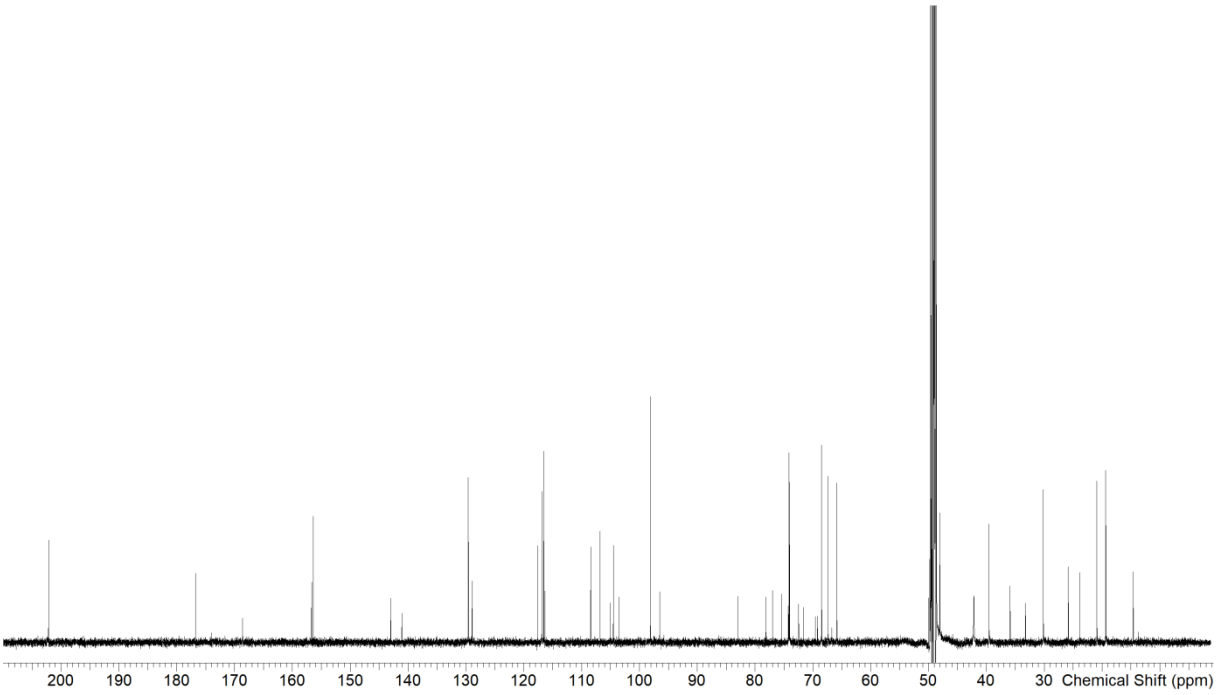

Figure S3: ^{13}C NMR of 560 in CD_3OD.

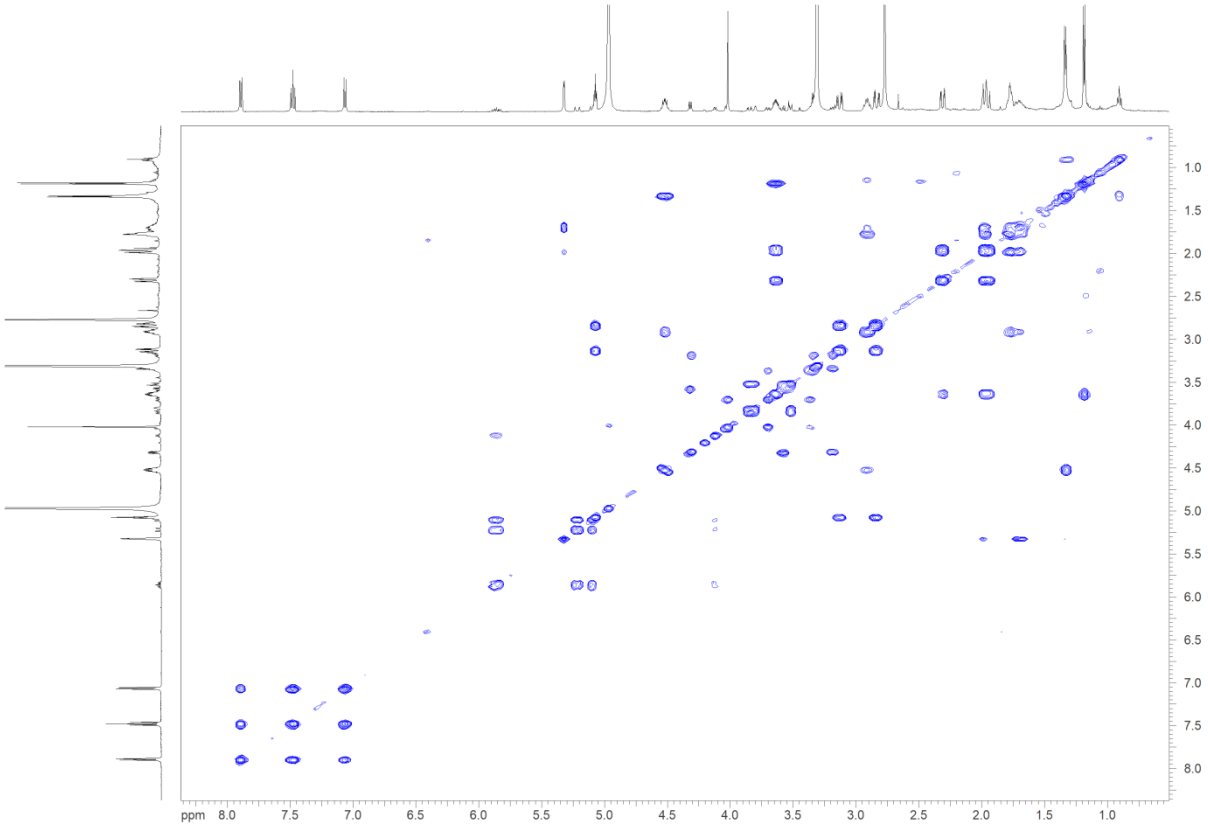

Figure S4: H-H COSY of 560 in CD_3OD.

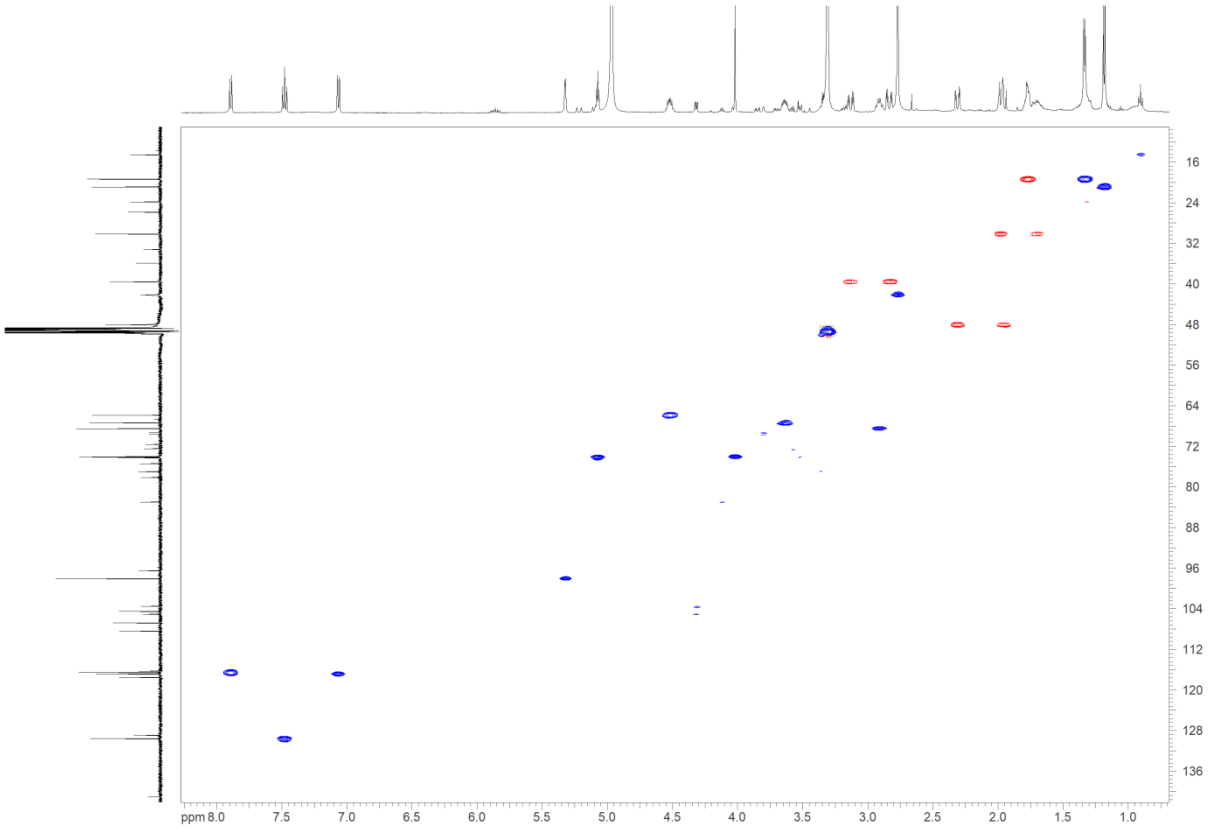

Figure S5: Edited HSQC of 560 in CD_3OD.

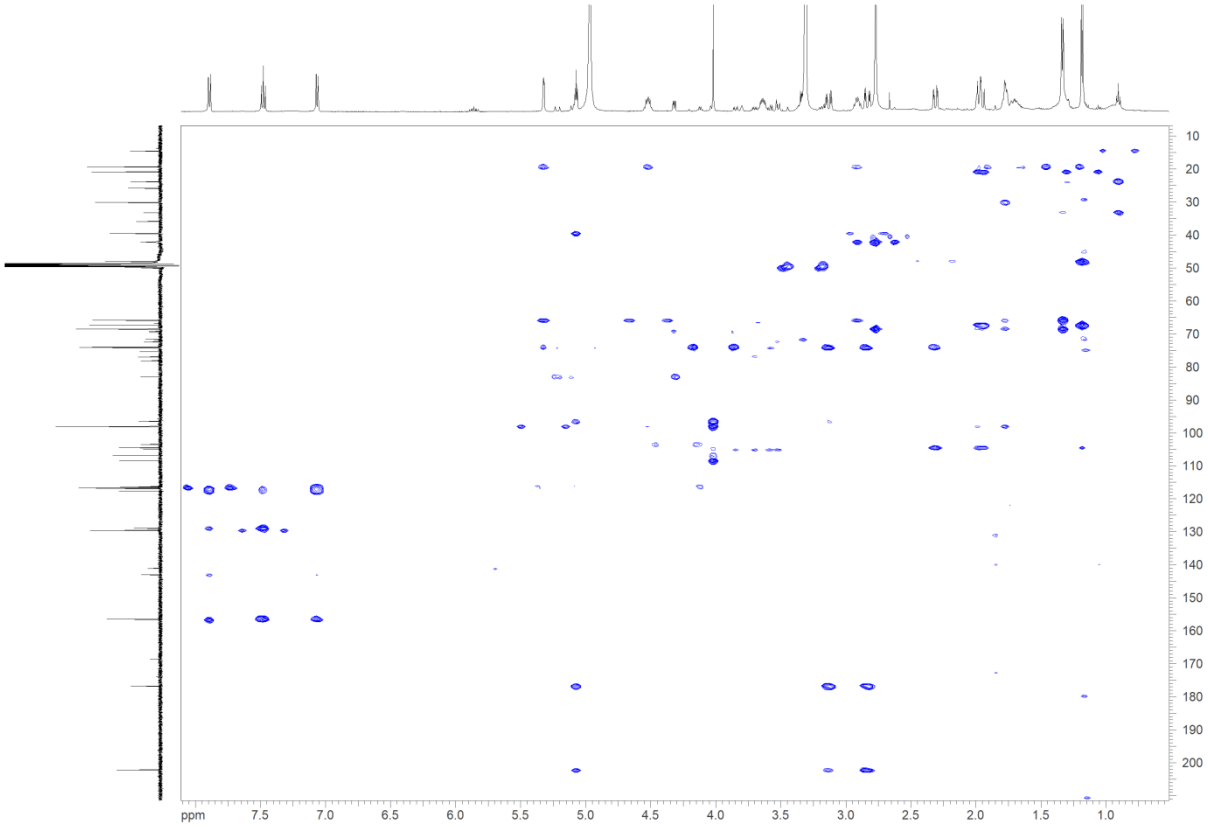

Figure S6: HMBC of 560 in CD_3OD.

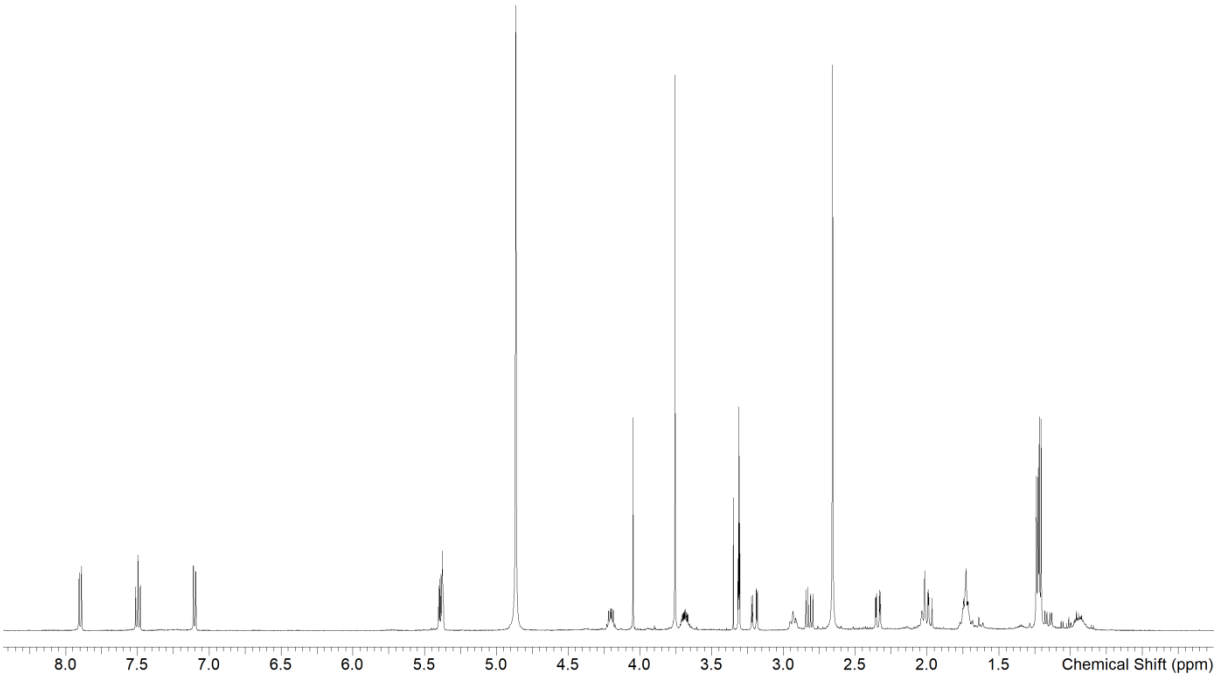

Figure S7: ^1H NMR of 574 in CD_3OD.

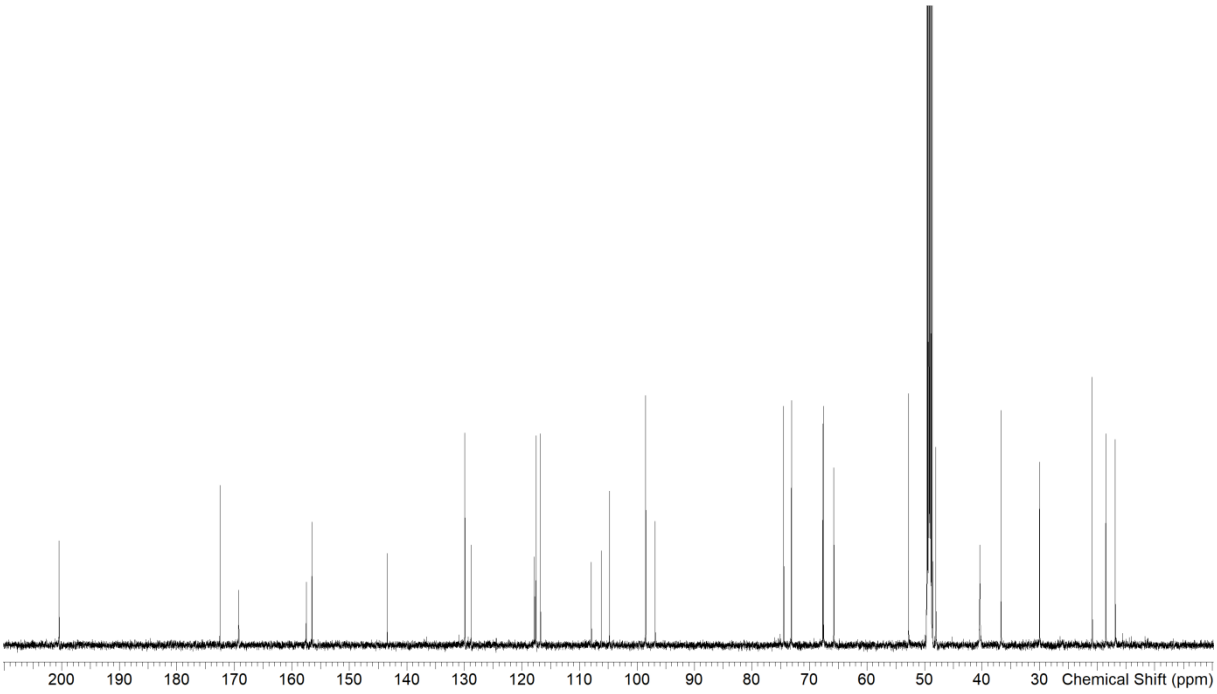

Figure S8: ^{13}C NMR of 574 in CD_3OD.

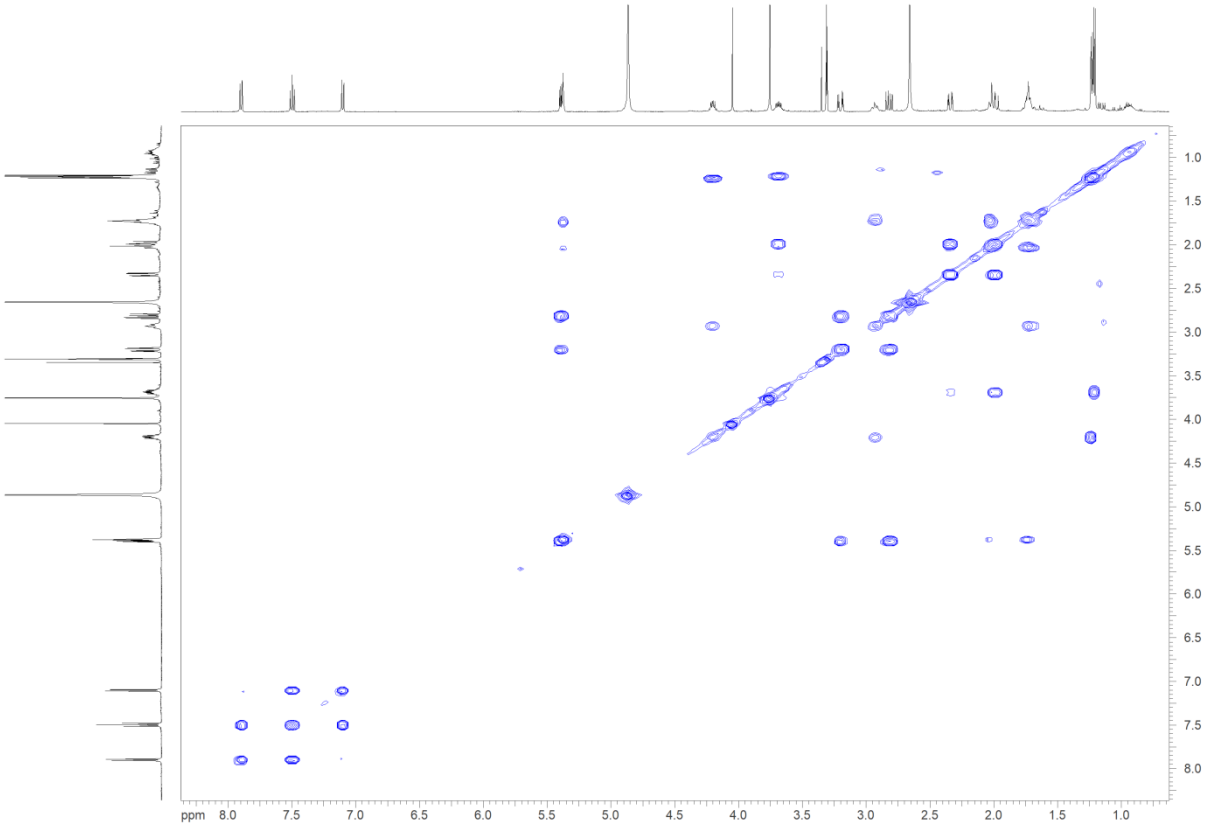

Figure S9: H-H COSY of 574 in CD_3OD.

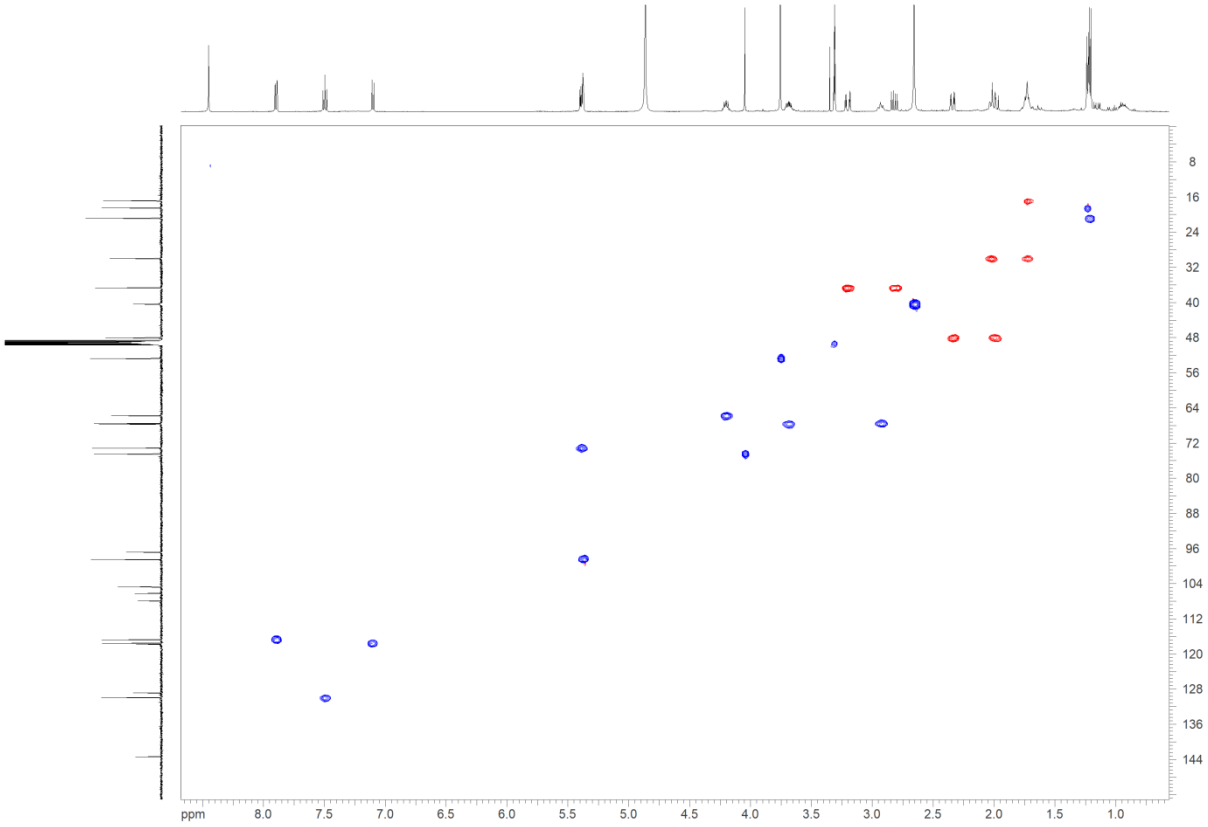

Figure S10: Edited HSQC of 574 in CD_3OD.

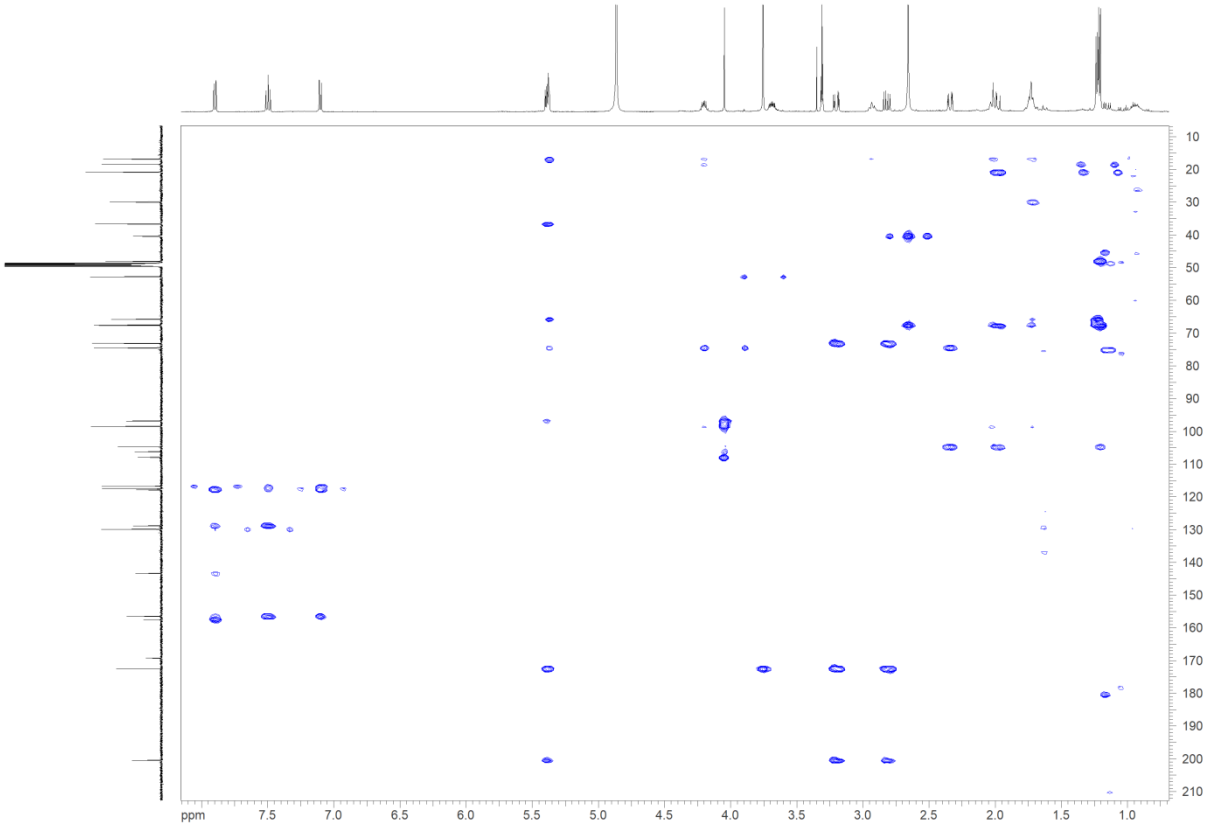

Figure S11: HMBC of 574 in CD_3OD.

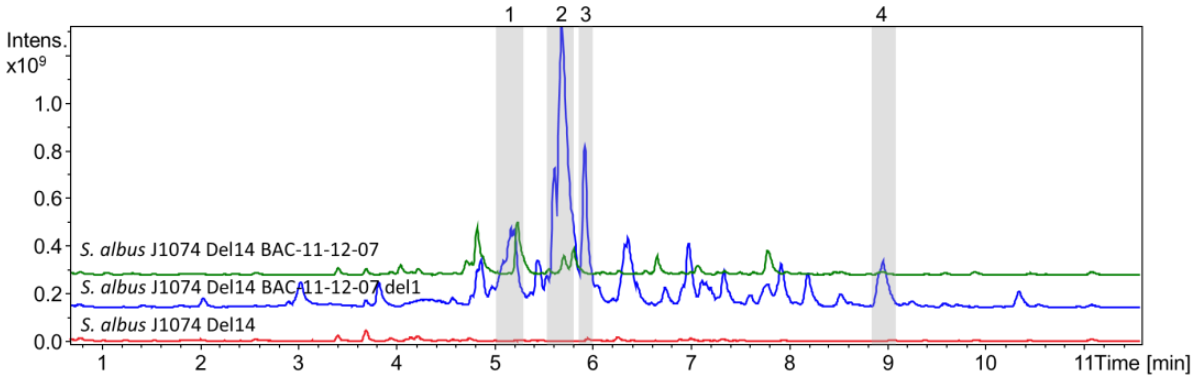

Figure S12: HPLC-MS chromatogram of different *S. albus* J1074 Del14 strains cultivated in DNPM medium: Red *S. albus* J1074 Del14; Blue *S. albus* J1074 Del14 BAC-11-12-07 del1; Green *S. albus* J1074 Del14 BAC-11-12-07. Highlighted masses [M+H] are: 1: 369/530 Da; 2: 371 Da; 3: 371 Da; 4: 367/530 Da.

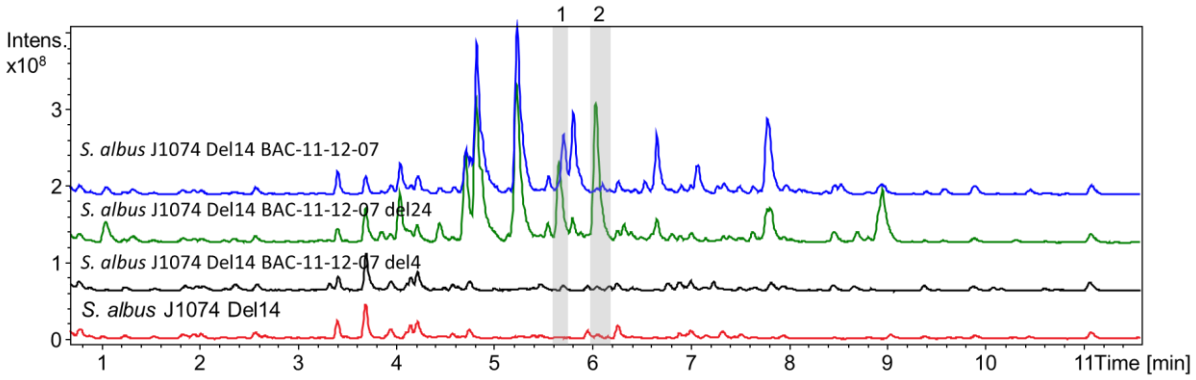

Figure S13: HPLC-MS chromatogram of different *S. albus* J1074 Del14 strains cultivated in DNPM medium: Red *S. albus* J1074 Del14; Green *S. albus* J1074 Del14 BAC-11-12-07 del24; Black *S. albus* J1074 Del14 BAC-11-12-07 del4; Blue *S. albus* J1074 Del14 BAC-11-12-07. Highlighted masses [M+H] are: 1: 574 Da; 2: 574 Da.

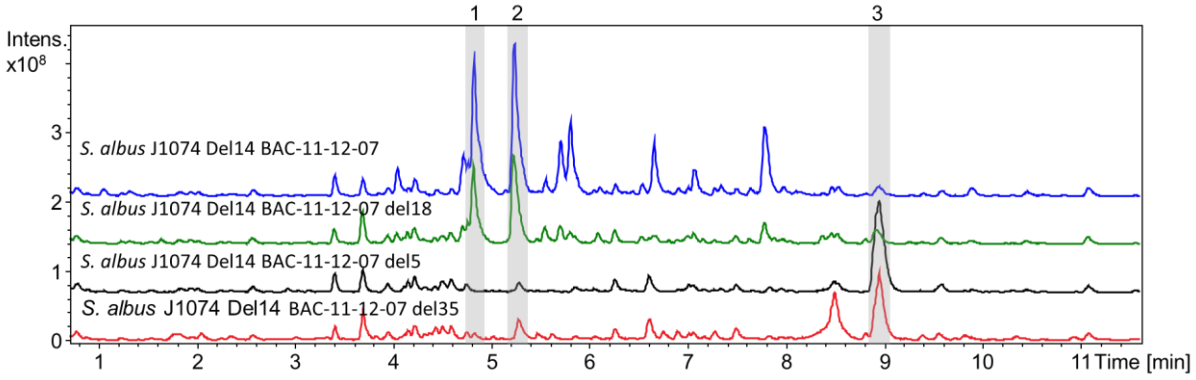

Figure S14: HPLC-MS chromatogram of different *S. albus* J1074 Del14 strains cultivated in DNPM medium: Red *S. albus* J1074 Del14 BAC-11-12-07 del35; Green *S. albus* J1074 Del14 BAC-11-12-07 del18; Black *S. albus* J1074 Del14 BAC-11-12-07 del5; Blue *S. albus* J1074 Del14 BAC-11-12-07. Highlighted masses [M+H] are: 1: 576 Da; 2: 576/293 Da; 3: 367 Da.

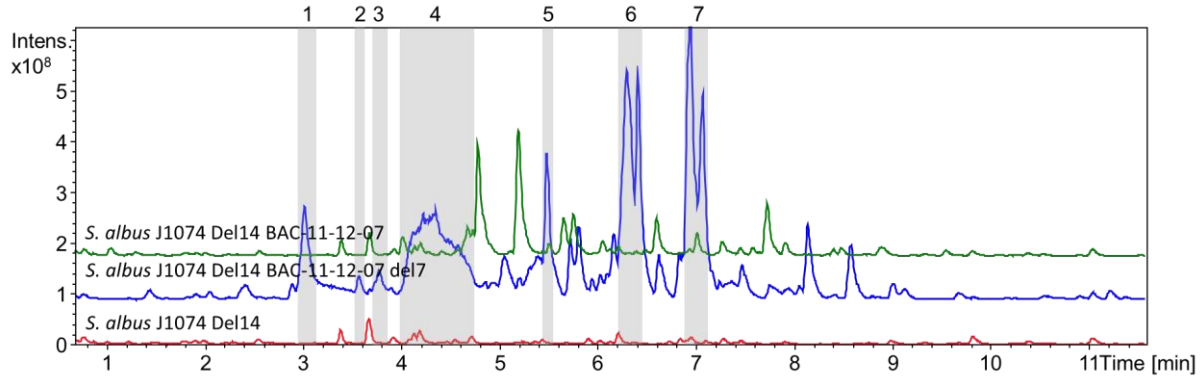

Figure S15: HPLC-MS chromatogram of different *S. albus* J1074 Del14 strains cultivated in DNPM medium: Red *S. albus* J1074 Del14; Blue *S. albus* J1074 Del14 BAC-11-12-07 del7; Green *S. albus* J1074 Del14 BAC-11-12-07. Highlighted masses [M+H] are: 1,2,3,4: 405 Da; 5: 369 Da; 6: 371 Da; 7: 387 Da.

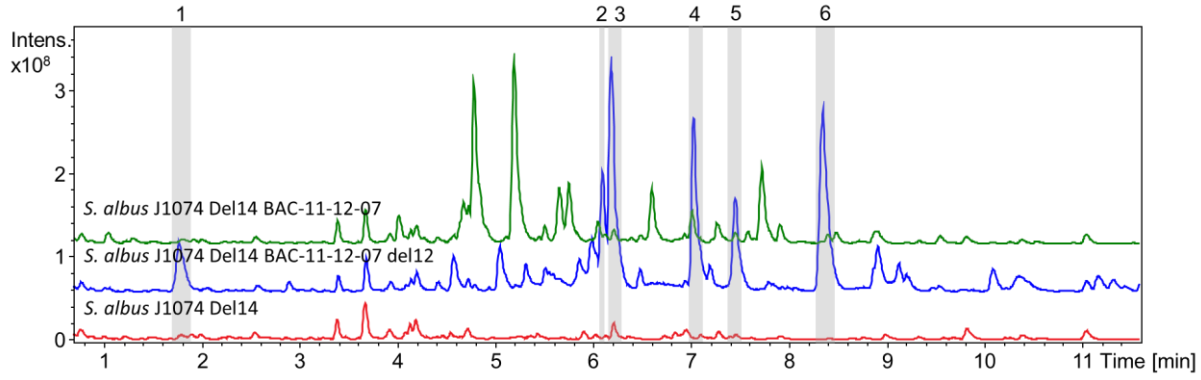

Figure S16: HPLC-MS chromatogram of different *S. albus* J1074 Del14 strains cultivated in DNPM medium: Red *S. albus* J1074 Del14; Blue *S. albus* J1074 Del14 BAC-11-12-07 del12; Green *S. albus* J1074 Del14 BAC-11-12-07. Highlighted masses [M+H] are: 1: 270 Da; 2: 387 Da; 3: 373 Da; 4: 530 Da; 5: 530 Da; 6: 289 Da.

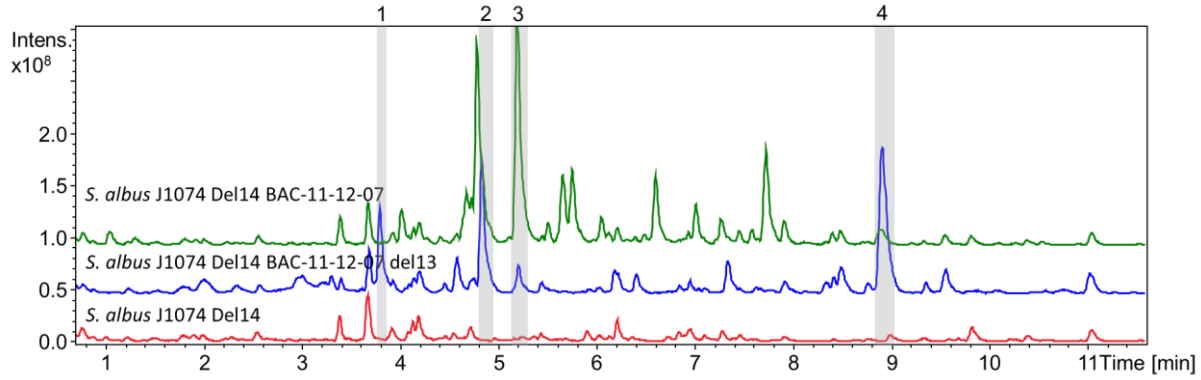

Figure S17: HPLC-MS chromatogram of different *S. albus* J1074 Del14 strains cultivated in DNPM medium: Red *S. albus* J1074 Del14; Blue *S. albus* J1074 Del14 BAC-11-12-07 del13; Green *S. albus* J1074 Del14 BAC-11-12-07. Highlighted masses [M+H] are: 1: 561 Da; 2: 601 Da; 3: 576 Da; 4: 367/530 Da.

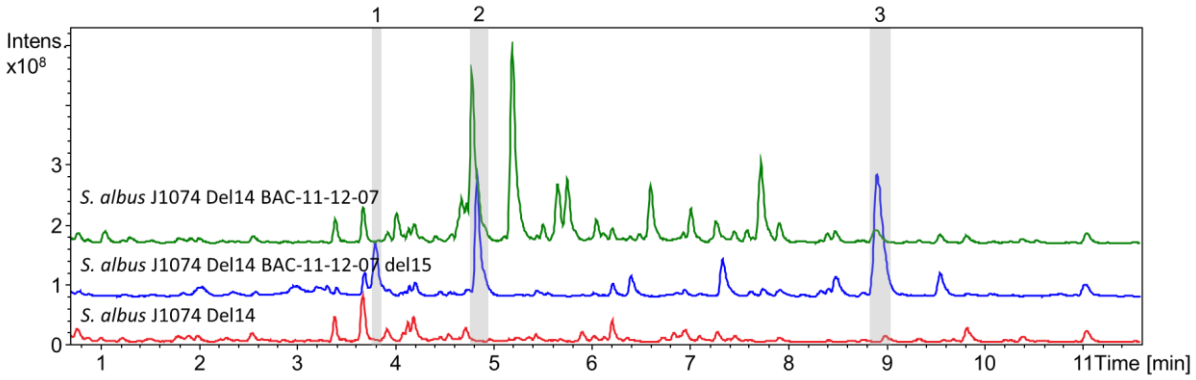

Figure S18: HPLC-MS chromatogram of different *S. albus* J1074 Del14 strains cultivated in DNPM medium: Red *S. albus* J1074 Del14; Blue *S. albus* J1074 Del14 BAC-11-12-07 del15; Green *S. albus* J1074 Del14 BAC-11-12-07. Highlighted masses [M+H] are: 1: 561 Da; 2: 601 Da; 3: 367/530 Da.

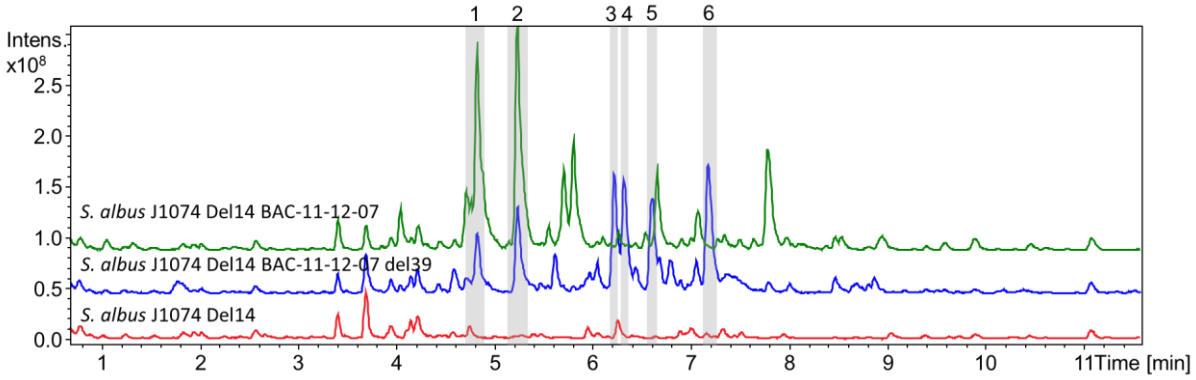

Figure S19: HPLC-MS chromatogram of different *S. albus* J1074 Del14 strains cultivated in DNPM medium: Red *S. albus* J1074 Del14; Blue *S. albus* J1074 Del14 BAC-11-12-07 del39; Green *S. albus* J1074 Del14 BAC-11-12-07. Highlighted masses [M+H] are: 1: 576 Da; 2: 576 Da; 3: 373 Da; 4: 560 Da; 5: 564 Da; 6: 403 Da.

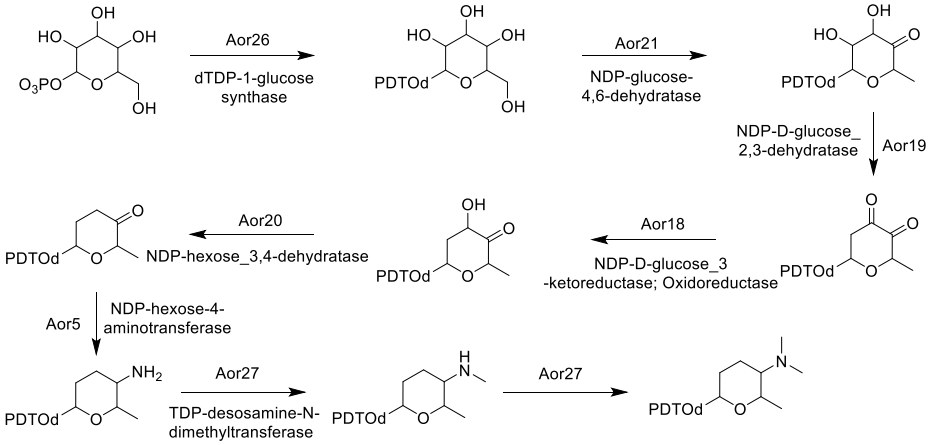

Figure S20: D-forosamine biosynthesis scheme.

References

1. Flett, F., V. Mersinias, and C.P. Smith, *High efficiency intergeneric conjugal transfer of plasmid DNA from Escherichia coli to methyl DNA-restricting streptomycetes.* FEMS microbiology letters, 1997. **155**(2): p. 223-229.
2. Fu, J., et al., *Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting.* Nature biotechnology, 2012. **30**(5): p. 440-446.
3. Myronovskyi, M., et al., *Generation of a cluster-free Streptomyces albus chassis strains for improved heterologous expression of secondary metabolite clusters.* Metabolic engineering, 2018. **49**: p. 316-324.
4. Axenov-Gribanov, D., et al., *The isolation and characterization of actinobacteria from dominant benthic macroinvertebrates endemic to Lake Baikal.* Folia microbiologica, 2016. **61**(2): p. 159-168.
5. Baker Brachmann, C., et al., *Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications.* Yeast, 1998. **14**(2): p. 115-132.

2.3 Discovery and isolation of novel highly active pamamycins through transcriptional gene cluster engineering

Nikolas Eckert, Liliya Horbal, Yurii Rebets, Josef Zapp, Jennifer Herrmann, Rolf Müller, Andriy Luzhetskyy

To be submitted

2.3.1 Abstract

A sustainable supply of complex natural products for their pharmacological evaluation is often a major challenge in the drug discovery process. Herein, we describe the transcriptional engineering of a pamamycin gene cluster, to enable isolation of known and discovery of novel bioactive molecules. This engineering approach includes the generation of a pamamycin gene cluster library with random combinations of synthetic promoters, its incorporation into the advanced host strain and LC-MS screening of recombinant strains for improved production of various pamamycins. To further increase the production level of active molecules, mutants of *Streptomyces albus* with increased resistance to pamamycins were selected. Combination of these two approaches led to the sustainable isolation of known and discovery of three novel pamamycins as well as to their bioactivity profiling. The antimicrobial and anticancer activity of isolated pamamycins was evaluated. New pamamycin pam-663 A shows an extraordinary activity (IC50 2 nM) against hepatocyte cancer cells as well as a strong activity (one digit micromolar range) against the whole range of Gram-positive pathogenic bacteria. Our results demonstrate that these technologies, refactoring a gene cluster via synthetic promoters in combination with an increased resistance of the producing strains, are highly promising strategies to establish a sustainable supply and discovery platform for bioactive natural products.

2.3.2 Introduction

Natural products are a prime source of agents displaying biological activities over the whole spectrum of therapeutically relevant indications, including anticancer, antibacterial, antifungal, antiparasitic, antiviral and many further treatment options, indispensable for modern human and animal medical care. The structural and chemical diversity of natural products is enormous and evolutionarily optimised to exert “drug-like” properties, which can hardly be matched by any synthetic libraries of small molecules [1]. Especially in the area of antiinfectives, natural products have been playing a key role since the beginning of modern pharmaceutical research and systematic screenings for effective drug candidates, which have been isolated mainly from microbial resources. For comparison, the role of natural products in drug discovery in about seventy years, from early 1940s into the beginning of the present decade, yielded about 50 % of approved small molecules for anticancer therapy, and about 70 % of antibiotics being either natural products or directly derived therefrom [2]. However,

mortality caused by AMR (antimicrobial resistance) rises exponentially and is predicted to exceed 10 million deaths per year by 2050 [3]. A similar situation is observed in cancer treatment, where the tumour resistance to chemotherapy is a major problem. Clearly, the development of novel antiinfectives and anticancer drugs is critical to stand a chance in the global fight against AMR and multi drug resistant tumours.

An almost untapped resource of novel bioactive metabolites is hidden in the genomic repertoire of actinomycetes. However, most of the corresponding biosynthetic gene clusters in actinomycetes are either silent under standard laboratory conditions or encode for compounds that are detectable only at nanogram quantities in crude extracts, far below the amounts required for further exploration [4]. A sustainable production of complex natural products in the advanced chassis strains after heterologous expression of the corresponding biosynthetic gene clusters (BGCs) is a promising solution. Moreover, the efficient heterologous expression of known BGCs responsible for the synthesis of high-value compounds enables the generation of chemical diversity by design-based engineering of the encoded pathways. However, the efficiency of complex BGCs expression from various organisms in the chassis strains is rather low. A key reason for this is very likely attributed to highly complex and very often BGC specific regulatory cascades controlling the expression of BGCs in the native producer, which are lacking in the heterologous host. Therefore, “refactoring” of BGCs must be applied to replace all native regulatory elements by synthetic regulation, in order to ensure the successful heterologous expression of the BGCs of interest in ratios that enable the purification of the desired compound in the required amount.

Pamamycins are a group of polyketide type II compounds with molecular masses ranging from 579 Da to 705 Da, pronounced bioactivity and challenging chemical structure [5]. “Light” pamamycins are highly active against gram-positive bacteria, including clinical isolates of *Mycobacterium tuberculosis* within a narrow MIC range of 1.5–2.0 mg/l irrespective of their resistance to isoniazid or rifampicin [6]. Pamamycin molecules with higher MW as 649 Da were only detected, but never chemically or pharmaceutically described, because of miserable yields. Due to the extraordinary bioactivities and challenging molecular structure, pamamycins have stimulated intense synthetic as well as biotechnological efforts [7, 8] in order to provide sufficient amounts and enable pharmacokinetic and pharmacodynamic studies. However, chemical synthesis of the pamamycins is very complex and protracted due to stereo control of the tetrahydrofuran rings

and the adjacent stereocentre in the fragments [9] and the production yield of pamamycins in native or heterologous hosts is extremely low.

Herein, we report the successful heterologous production of pamamycins through the refactoring of the corresponding biosynthetic gene cluster and its subsequent expression in the heterologous host strain *Streptomyces albus* J1074 Del14. The refactoring approach is based on the creation of a library of clusters with random combination of promoters with arbitrary spacer regions and selection for the promoters which led to the increased pamamycin production. Combining this strategy with an enhanced pamamycin resistance of the heterologous production strain led to further increased yields and as a result discovery of three novel highly active pamamycins. Using the above described approach we were able to increase the yield of particular pamamycins up to 26 fold. As a result, sufficient amounts of pamamycins pam-635 G, pam-663 A and homopam-677 A were purified for the first time and their structures could be determined as well as their activities analysed. We anticipate that combinations of above-mentioned generic strategies will facilitate the overproduction and metabolic profile engineering of neglected NPs as well as the discovery of new potentially active NPs in *Actinobacteria*.

2.3.3 Material and methods

Bacterial strains and culture conditions

The bacteria strains used in this study are listed in table S1. *E. coli* strains were grown in liquid lysogeny broth (LB) or solid LB containing 2 % (w/w) agar at 37 °C and at 180 rpm for liquid cultures. Targeted homologous recombination (Red/ET) was performed with the *E. coli* GB05-redCC strain (Table S1). *Streptomyces* strains were grown on mannitol soy agar (MS) [10] for 5-7 days at 29 °C, for sporulation. Pre-cultures were cultivated in 20 ml tryptic soy broth (TSB, Sigma-Aldrich, USA) in a 100 ml flask (4 baffles) with glass beads (Carl Roth; 4 mm) for 1 day at 29 °C and agitation at 180 rpm with the respective antibiotics if necessary. The main culture was inoculated with 1 ml of pre-culture. It was cultivated under the same conditions for 90 h in 50 ml optimised starch glucose glycerol medium (SGG), in a 300 ml flask with one baffle and glass beads. SGG medium is composed of: 3 g calcium carbonate; 2.5 g corn steep powder; 10 g glycerol; 5 g peptone; 1 g sodium chloride; 10 g starch and 2 g yeast extract; dissolved in 1 l demineralised water and the pH was adjusted to 7.2. Antibiotics (Roth, Germany; Sigma, USA) were added to the culture in the following concentrations: 50 µg/ml apramycin, 50 µg/ml kanamycin, 120 µg/ml hygromycin, 50 µg/ml nalidixic acid.

Recombinant DNA techniques

Chromosomal DNA from *Streptomyces* strains and plasmid DNA from *E. coli* were isolated using standard protocols [11, 12]. Restriction enzymes and molecular biology reagents were used according to the manufacturer's protocol (NEB, England; Thermo Scientific, Germany).

Construction of the pTOS-P21pamW plasmid

The putative transporter gene *pamW* was cloned into the vector pTOS via ligation. For this purpose, the vector pTOS was digested with the endonuclease *Sna*BI and ligated with the amplified *pamW* fragment, obtained using primer pair 1 (Table S2). As a result, the pTOS-P21pamW plasmid (Table S1) was obtained. Correct transformants were verified by enzymatic digestion and sequencing using primer pair 2 (Table S2).

The constructed vector was conjugated into *S. albus* J1074 [11]. Exconjugants were used as heterologous host for the screening of the R2 cosmid library.

Construction of the R2 cosmid library with random promoters using Red/ET method

A library of the R2 cosmid with random promoters, containing the genes for the pamamycin biosynthesis, mutants containing degenerated promoters for the two main operons, was constructed. For this purpose a hygromycin resistance cassette (Hyg^R) surrounded with randomly generated outwards oriented promoters was amplified using the degenerate oligonucleotide pair 3 (Table S2), containing 50 bp in length homologous regions to the *pamA* and *pamF* genes (Figure 1). The obtained PCR fragment was used to perform recombination with the R2 cosmid. After recombination, colonies were selected on LB agar plates with 120 μg/ml hygromycin. As a result a library of the R2 cosmid with random promoters was obtained. The validity of some of the constructs was tested by enzymatic digestion and sequencing using primer pair 4 (Table S2).

For further screening the R2 mutant library with random promoters was transferred into *S. albus* J1074 pTOS-P21pamW via triparental conjugation [11].

Determination of relative promoter strength via GUS activity assay

Promoters were analysed with an assay based on the changed β-glucuronidase expression depending on the promoter strength [13]. Promoters of the Constructs R2, R2-67, R2-73 and R2-100 were amplified using primer pairs 7-16 (Table S2) and a hygromycin resistance cassette as selection marker. The obtained cassettes, as well as the vector pGUS (Table S1),

were digested with the restriction enzymes *Spe*I and *Xba*I and ligated. To exclude any influence of the promoter of the resistance gene the final construct was designed in such a way, that the synthetic promoters faced downstream towards the *uidA* gene and the hygromycin resistance upstream. The obtained constructs were sequenced using primer pair 4 (Table S2), to ensure mutation free promoters as well as correct orientation. The constructs were then conjugated into *S. albus* J1074 for further analysis. The strains were cultivated as described before in SGG medium. After 30 h of growth 1 ml culture was harvested, washed once with distilled water, resuspended in 0.5 ml lysis buffer (50 mM phosphate buffer pH 7.0, 5 mM dithiothreitol (DTT), 0.1% Triton X-100, 2 mg/ml lysozyme) and incubated for 15 min at 37 °C. Following the lysate was supersonificated once for 10 s (Fisher scientific; sonicator 3.1 mm). 10 µl of the lysate was transferred into 175 µl dilution buffer (50 mM phosphate buffer (pH 7.0), 5 mM DTT, 0.1% Triton X-100) supplemented with 5 µl of 4-nitrophenyl-β-d-glucopyranoside (61 mg/ml) which was kept on ice. Immediately after adding the lysate the 96 well plate containing the samples was put into a well plate reader (BMG Labtech; POLARstar Omega) and the OD_{415} was measured each 1 min at 37 °C and shaking at 500 rpm for 20 min. The slope of the resulting curve was divided by the dry biomass to calculate the enzyme activity in units per gram dry weight.

Isolation of pamamycins

For pamamycin production, 1 ml of a 1-day-old pre-culture grown in 25 ml of TSB media (Sigma-Aldrich, USA) was inoculated into 50 ml optimised SGG, in a 300 ml flask with one baffle and glass beads and grown for 90 hours at 29 °C with agitation at 180 rpm. Pamamycins were extracted from the supernatant of the culture broth with ethyl acetate and from the biomass with an acetone and methanol mixture (ratio, 1:1). For small scale extractions 30 ml of a bacterial culture were centrifuged at 9000 g for 5 min. 20 ml supernatant was mixed with 20 ml ethyl acetate and the cell pellet was resuspended in 15 ml of the acetone-methanol mixture. Both extractions were shaken for 1 h at 180 rpm (Laboshake, Gerhardt GmBH, Dreieich), afterwards centrifuged at 9000 g for 5 min. Both samples were evaporated, dissolved in 1 ml methanol, mixed together and again evaporated. The obtained crude extract was dissolved in 250 µl mixture of methanol and DMSO (1:1), centrifuged (19000 g, 10 min) and subjected to LC-MS.

HPLC-MS analysis of pamamycin production and novel derivatives thereof

Pamamycin extracts were analysed by HPLC-HRMS analysis (Dionex Ultimate 3000, Thermo Fisher Scientific; AmaZon Speed ETD, Bruker). Samples were eluted with solvent A: Ammonium formate buffer 90 mM and solvent B: 100:20 acetonitrile/100 mM ammonium formate buffer in a multistep gradient: 0.2 min 20% B, 20% B to 97% B in 2.8 min, 97% B to 100% B in 7 min, 1 min at 100%, 100% B to 20% B in 1 min, 3 min equilibration at 20% B. A Waters BEH C18 column (100 mm x 2.1 mm, 1.7 µm) was used, the column temperature was set to 45 °C and the flow was set to 0.55 ml/min. The data were analysed with the software Compass DataAnalysis (Bruker).

Samples from the recombinant strains were additionally analyzed on a maXis 4G hr-ToF ultrahigh-resolution mass spectrometer using the Apollo II ESI source (Bruker Daltonics, Germany). Extracts were separated on UPLC system (Dionex Ultimate 3000, Thermo Fisher Scientific GmbH) with a Waters BEH C18 column (100 mm x 2.1 mm, 1.7 µm, column, temperature 45 °C). Buffers A (MilliQ, 0.1% formic acid) and B (Acetonitrile, 0.1% formic acid) and the gradient of solvent B from 5 to 95% in 18 min were used. The flow rate was set to 0.5 ml/min.

Isolation and purification of pamamycins

The crude extract obtained after ethyl acetate and an acetone-methanol extraction was used for purification of the pamamycin compounds on a 63 cm long column packed with cross-linked dextran polymer beads (LH-20, GE Healthcare Bio-Science AB). Methanol was used as a solvent for elution. 200 fractions were collected and screened for the presence of pamamycins.

The fractions containing pamamycins were pooled together, evaporated, and dissolved in methanol. Afterwards, the prepurified extract was further purified on an Agilent Technologies 1260 Infinity semi-preparative HPLC using a Synergi column TM 4 µm Fusion RP 80 Å (250×10 mm). Samples were eluted with solvent A (MilliQ, 0.1% formic acid) and B (Acetonitrile, 0.1% formic acid) and multistep gradient: 2 min 5% B, 5% B to 60% B in 4 min, 60% B to 95% B in 20 min, 2 min at 95%, 95% B to 5% B in 1 min, 3 min equilibration at 5% B. The fractions with pamamycin were again evaporated and dissolved in methanol. After these two purification steps, the pamamycins were used for NMR analysis.

NMR analysis of pamamycins

NMR spectra were acquired at 298 K on a Bruker Ascend 500 or a Bruker Ascend 700 NMR spectrometer, respectively. Both spectrometers were equipped with a 5 mm TXI cryoprobe. $CDCl_3$ was used as solvent for all samples and for all experiments. The chemical shifts (δ) were reported in parts per million (ppm) relative to $CHCl_3$ at 7.26 ppm for the proton spectra and to $CDCl_3$ at 77.0 ppm for the carbon spectra, respectively. 2D NMR ^{1}H-^{1}H-COSY, edited-HSQC, HSQC-TOCSY and HMBC were recorded using the standard pulse programs from the TOPSPIN v.3.6 software. For compound **3**, 1D selective TOCSY were measured with a mixing time of 75 and 125 ms.

Bioactivity tests

Antimicrobial activity studies were performed against several standard test strains: *Acinetobacter baumannii* DSM-30008, *E. coli* JW0451-2 (ΔacrB), *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. As described in Megahed et al. 2022 minimum inhibitory concentrations (MICs) were determined according to standard procedures [14]. Serial dilutions of pamamycins, dissolved in DMSO, ranging from 1 to 128 µM were prepared in sterile 96-well plates, and the bacterial suspension was added. Growth inhibition was assessed after overnight incubation (16–48 h) at 30 or 37 °C.

Additional activity analysis was performed with human cancer cell lines KB-3.1 (ACC-158) and HepG2 (ACC-180). The cell lines were cultivated according to the collection recommendations. Serial dilutions of pamamycins were added to the cell lines and incubated for 3 days. Viability was determined with photometric measurement of MTT (thiazolyl blue tetrazolium bromide), compared to the DMSO control and the average value was used for determination of the IC_{50} (Standard deviation <10%).

To analyse the effect of pamamycins on whole eukaryotic systems the maximum-tolerated concentration (MTC) was studied in zebra fish as described in Richter et al. 2019 [15]. Therefore, 12 zebra fish eggs for each pamamycin, either 1 or 3 days post-fertilisation, were mixed with 10 µM of pamamycins and monitored for 1 day via a LEICA M205 FA stereo microscope (Leica Mikrosysteme Vertrieb GmbH, Wetzlar, Germany).

Herbicidal tests were performed as well. For this purpose 10 *Agrostis stolonifera* seeds were transferred into a 96 well plate containing required salts as described in Rodríguez et al. 2020 Decreasing concentrations of pamamycins dissolved in DMSO were added (40, 20, 10, 5 and 2.5 mM) and DMSO in the same concentrations without pamamycins was used for the control. The seeds were then incubated for 3 days at room temperature and light radiation (Osram Fluora lamp). Following the lid was removed, the plants incubated for 3 more days and the germinated plants counted.

Selection of spontaneous resistant mutants

Streptomyces albus J1074 was grown on MS agar plates for 7 days. The spores were scratched with 10 ml water from the plates and a serial dilution was performed (10^{-1} until 10^{-9}). Those dilutions were spread on MS agar plates containing 50 µg mL^{-1} of a pre-purified mixture of pamamycins and grown colonies were selected for further selections with increased concentrations of the pre-purified pamamycin up to 400 µg mL^{-1}. Resistant mutants were obtained from 100, 200 and 400 µg mL^{-1} of pre-purified pamamycins and used for later experiments.

2.3.4 Results

Pamamycin gene cluster refactoring using random semisynthetic promoters

Rational promoter engineering is a widely used strategy to trigger or increase the production of secondary metabolites [16, 17]. Moreover, this strategy allows overcoming regulatory barriers to secondary metabolite production that might exist in the cell. The pamamycin gene cluster contains five operons: two of them comprise key biosynthetic genes and the remaining three encode for a positive and negative regulator, as well as a putative transporter (Figure 1) [8]. The expression of the biosynthetic genes with the semi-synthetic promoters may decouple their expression from the native regulatory network. Since multivariate modular metabolic engineering [18] is rather time-consuming and very laborious, we decided to apply an earlier developed random rational strategy [19]. The idea was also to use this strategy not only to increase the production but to shift the metabolic profile of pamamycins from “light” to “heavy” ones. The latter were only detected but never characterised before, from a chemical or pharmaceutical point of view. Taking into account that PamW is involved in the excretion of the pamamycins from the cell and thus might have an important function for the resistance (Figure S1), the strain with additional copies of *pamW* was used as a heterologous host.

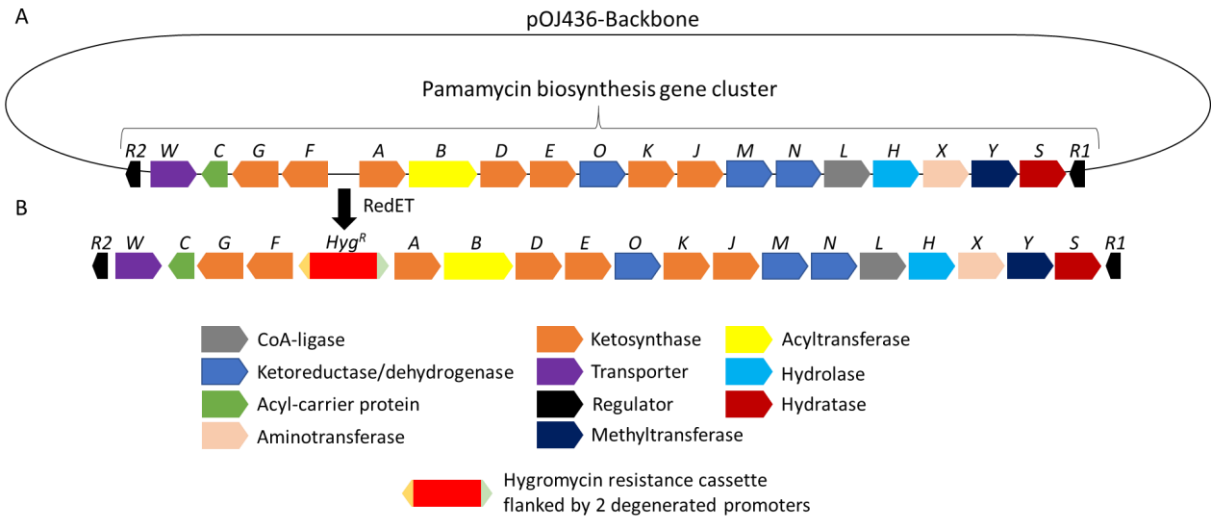

Figure 1: Engineering of the pamamycin gene cluster. (A) Schematic representation of the R2 cosmid and pamamycin gene cluster (B) Schematic depiction of the pamamycin cluster with the insertion of the cassette with two outward oriented semisynthetic promoters

To this aim, a library of the pamamycin gene clusters carrying random synthetic promoters of unknown strengths inserted between the two main operons (Figure 1) was constructed. Random promoters were created based on the consensus -35 and -10 sequences of the erythromycin resistance gene promoter from *Saccharopolyspora erythraea* [20] (Figure 1). The use of the degenerate primer pair 3 (Table S2) and an amplified hygromycin resistance gene facilitated the rapid and easy Red/ET cloning of a library of different semi-synthetic promoters between the two main operons in the pamamycin gene cluster (Figure 1).

106 recombinant pamamycin BGCs containing the semi-synthetic promoter cassettes were selected for the expression in the *S. albus* pTOS-P21pamW strain. Analysis of the corresponding extracts revealed 2 strains (1.8%) with a substantially shifted and increased metabolic profile towards “high” molecular weight pamamycins (Table S3, Figure 2).

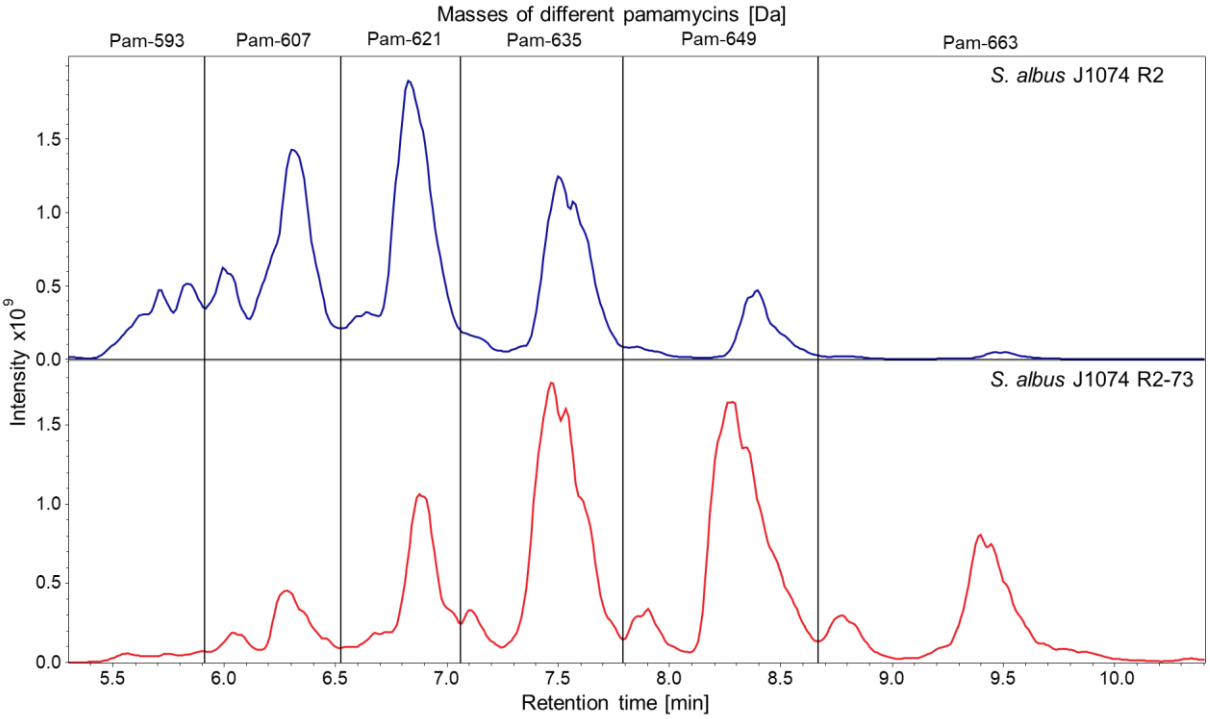

Figure 2: Comparison of the MS-chromatograms of *S. albus* J1074 containing the R2 and the R2-73 cosmid.

The production level of the pamamycins with a MW of 663 Da was in average 13 fold increased in comparison to the strain containing the unmodified cluster (Figure 3).

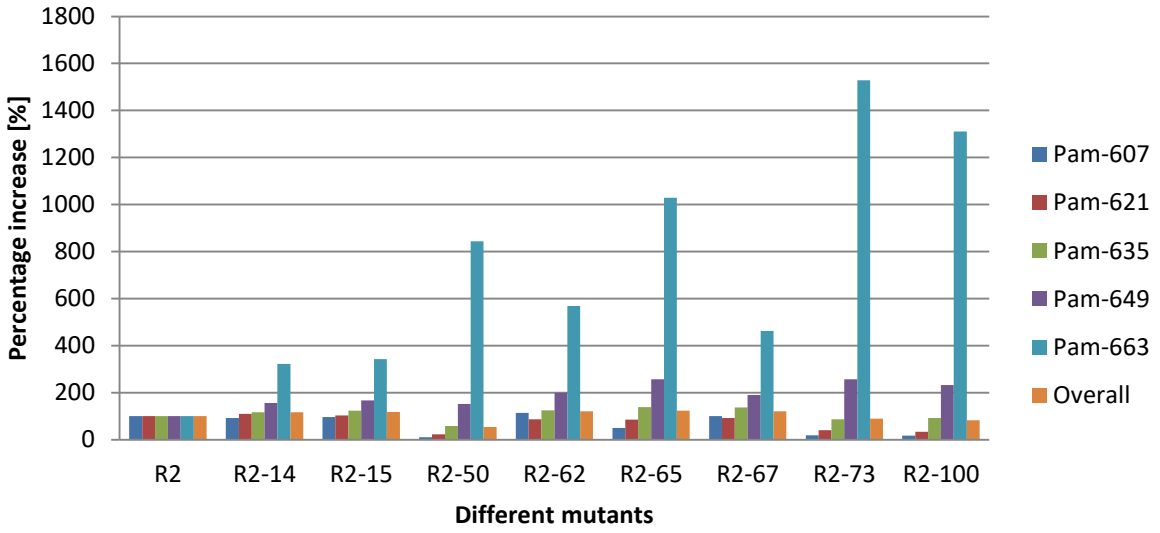

Figure 3: Pamamycin production of the rational random engineered R2 cosmids in the heterologous host *S. albus* J1074 pTOS-P21pamW.

Furthermore, the yield of pamamycins with the MW 677, 691 and 705 Da, which are produced in miserable amounts in the parental strain, was also increased. 11 out of 106 strains (10 %) are characterised by the same metabolic profile and similar production levels as the

S. albus R2 strain (Table S3). The remaining 95 strains are characterised by a low production level or no pamamycin production at all (Table S3).

Semi-synthetic promoters cause a delay in the biosynthesis

To get more insight into the factors leading to the pamamycin overproduction and the shifted metabolic profile, an analysis of the pamamycin production at different time points was performed. For this purpose, two strains, namely *S. albus* R2-73 and *S. albus* R2-100, with a metabolic profile shifted towards pamamycin with high MW and the R2-67 mutant, which is characterised by a similar metabolic profile as the control strain, were used. Four strains, including the control one, were cultivated for 4 days in SGG medium at 29 °C. The analysis of the pamamycin biosynthesis was performed after 13, 22 h and afterwards each 12 h of growth. The control R2 strain starts to produce pamamycins 13 h after the culture has been inoculated (Figure 4, A). Pamamycin production is in all tested recombinant strains strongly delayed and starts about 22 h after the culture was inoculated. (Figure 4, B, C, D).

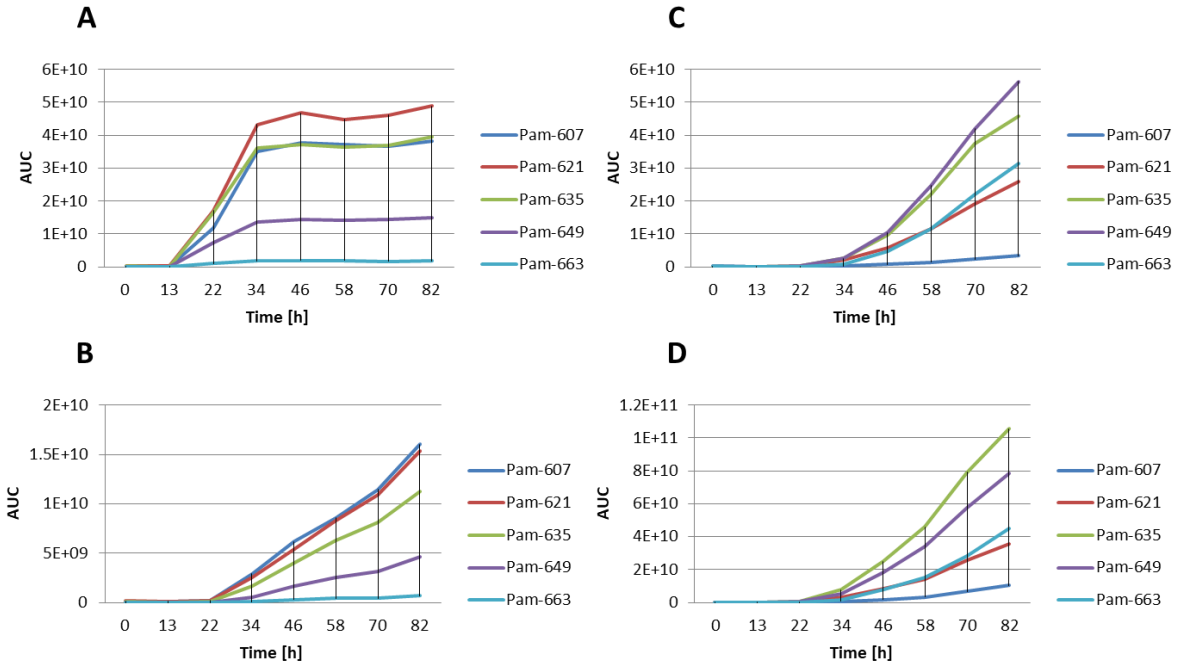

Figure 4: Pamamycin production over time in the different *S. albus* strains. (A) *S. albus* J1074 R2, (B) *S. albus* J1074 R2-67, (C) *S. albus* J1074 R2-73, (D) *S. albus* J1074 R2-100. Shown is the area under the curve (AUC) for different pamamycins pam-607-663.

Furthermore, the R2 control strain reaches a pamamycin production plateau at 34 h after the inoculation, whereas the strains with semi-synthetic promoters show a continuous growing production until about 82 h (Figure 4). Thus, the biosynthesis of the pamamycins in these strains is extended over time. Taking the obtained results into account it might be that the

generated promoters are activated a bit later than the native promoters from the cluster. Therefore, the delay in the biosynthesis allows the recombinant strains to accumulate the biomass first and then start the production avoiding a severe toxic effect of pamamycins.

Analysis of the relative promoter strengths via GUS activity assay

To get further insights into the changed expression levels of the cluster a GUS activity assay was performed. This assay allows the relative determination of promoter strengths by cloning them in front of a reporter gene, thereby resulting in different conversion rates of a colourless substrate into a photometric detectable substance [13]. Each of the 2 introduced promoters from the constructs R2, R2-67, R2-73, R2-100 was cloned in front of the reporter gene and analysed after 30 h of growth in SGG medium (Other time points were measured as well and showed similar results, data not shown).

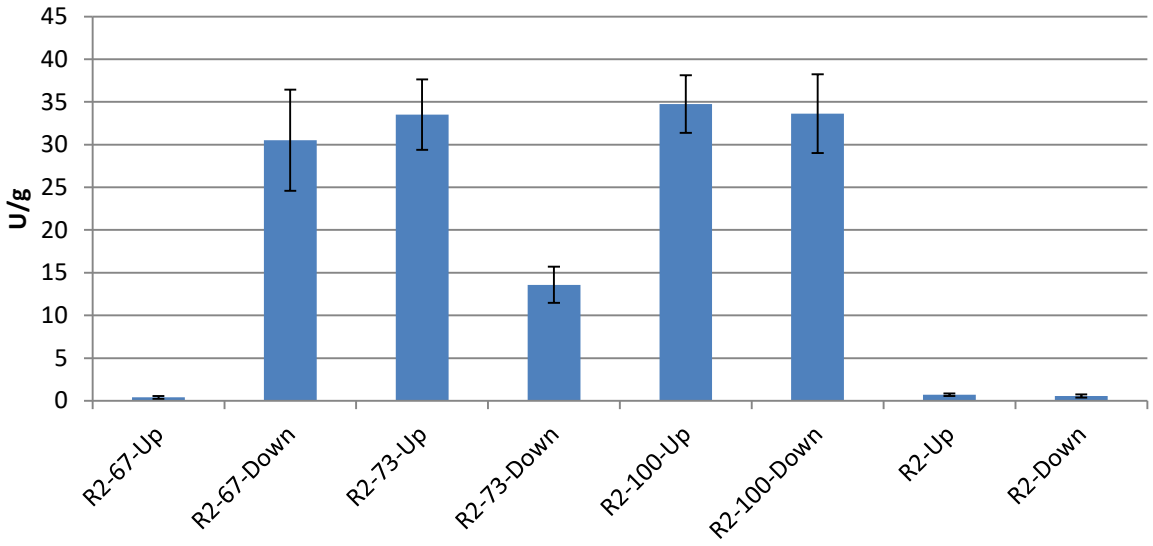

Figure 5: Glucuronidase activity in cell lysates of *S. albus* J1074 containing pGUS constructs with different synthetic promoters and native promoters of R2, R2-67, R2-73 and R2-100 cloned in front of the reporter gene. The strain was grown for 30 h in SGG medium and *S. albus* J1074 pGUS was used for the blank sample.

As it can be seen in figure 5 the upstream native promoter of the strain R2 and the upstream semi-synthetic promoter of R2-67 show similar low activities of 2 U/g. Even though they have substantially different expression levels of the downstream promoter 2 U/g for R2 compared to 30 U/g for R2-67, they show compared to the other two cosmids a similar pamamycin production spectrum.

When comparing the 2 mutants with a shifted production spectrum with each other it can be seen that the upstream oriented promoter is in both cases strong with 30 U/g. Unlike the

upstream promoter the downstream promoter can vary in strength, as the production is shifted similar for a promoter with strengths of 13 and 33 U/g. From this data it can be concluded that an overexpression of the smaller left operon is crucial for a shifted pamamycin production. Similar results were obtained with 5'-end enriched RNA seq analysis (Figure S2).

Effect of separate engineered promoters on the pamamycin production

To determine whether the shift in the metabolic profile is indeed caused by the overexpression of one of the two operons, additional experiments with the overexpression of the left or right operons in the pamamycin cluster separately from the upstream or downstream semi-synthetic promoters were performed. The R2-73 and R2-100 strains are both characterised by an increased production and a metabolic profile shifted towards heavy pamamycins. The average pamamycins production is similar in both strains, thus we took the R2-73 construct for further experiments. To obtain the sequence of the p73 promoter pair (Table S4), R2-73 was partially sequenced. Two constructs R2-p73Down and R2-p73Up, containing left or right operons under the control of the one of the p73 promoters, were constructed. Therefore, two hygromycin cassettes flanked with the outward-oriented p73Downstream or p73Upstream promoter and containing a region of homology to the pamamycin cluster were amplified using primer pairs 5 and 6 (Table S2). The obtained fragments were incorporated into the R2 cosmid via Red/ET recombination to generate R2-p73Down and R2-p73Up (Table S1). To exclude potential mutations from the cloning procedure, the expression constructs were verified via partial sequencing. The obtained cosmids were transferred into the *S. albus* strain via conjugation. Both recombinant strains and a control strain with the unmodified R2 cosmid were cultivated in SGG medium at 29 °C for 90 h. The cultures were extracted and the pamamycin production profile was analysed.

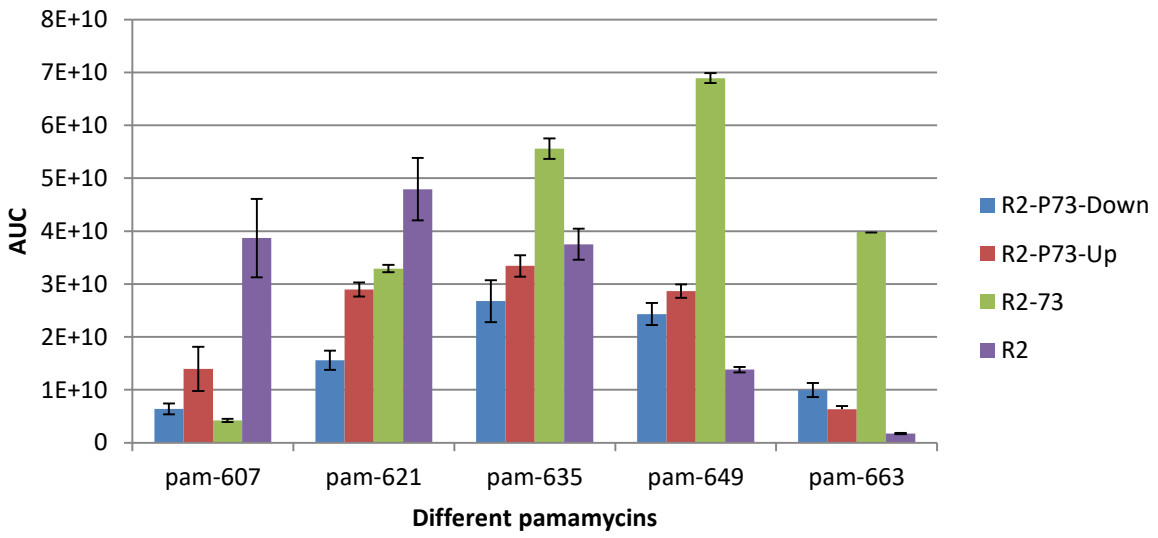

Figure 6: Pamamycin production in the *S. albus* J1074 strains containing different variants of the R2 cosmid. Completed with either one pair of the R2-73 cosmid and the R2 native promoter and the R2 and R2-73 cosmid as a control

HPLC-MS analysis revealed that both strains show an increased production of pamamycins 649 and 663 compared to R2 (Figure 6), but still less than the strain with the R2-73 cluster. Additionally, the production of lower molecular weight pamamycins is lower than the strain with R2. Therefore, it is most likely a synergetic effect of both promoters and the 2 operons need to be highly tuned to reach a high production as well as a shift.

Structure elucidation of new pamamycins by NMR

Because the R2-73 strain is constructed on the basis of the *S. albus* J1074 strain, which contains more than 20 secondary biosynthetic gene clusters, we decided to express the modified cluster in the *S. albus* J1074 Del14 chassis strain that was recently generated in our lab. It lacks most of the biosynthetic gene clusters [21], and thus has a simplified metabolic profile. The R2-73 cosmid was transferred into the *S. albus* J1074 Del14 strain via conjugation. The obtained *S. albus* J1074 Del14 R2-73 recombinant strain was cultivated in 10 l of SGG medium at 29°C for 4 days. The obtained culture broth was extracted with ethyl acetate and the biomass with a mixture of acetone-methanol, evaporated and dissolved in methanol. The crude extract was purified using size-exclusion chromatography on a dextran polymer bead LH20 column and reverse-phase chromatography on a preparative C18 column. After this purification step fractions of pamamycins with a MW of 635, 663 and 677 Da, with 90 % purity, were obtained. These were used for further NMR analysis.

Pamamycins show a large number of homologues with various substituents in certain positions [22]. All structures known so far consist of a 16-membered macrodiolide composed

of two hydroxy acids in a longer upper and a shorter lower fragment. Structurally interesting features include three cis-2,5-disubstituted tetrahydrofurans with adjacent methyl, ethyl and propyl substituted stereogenic centers and a mono or dimethyl amine in the upper fragment (Figure 7).

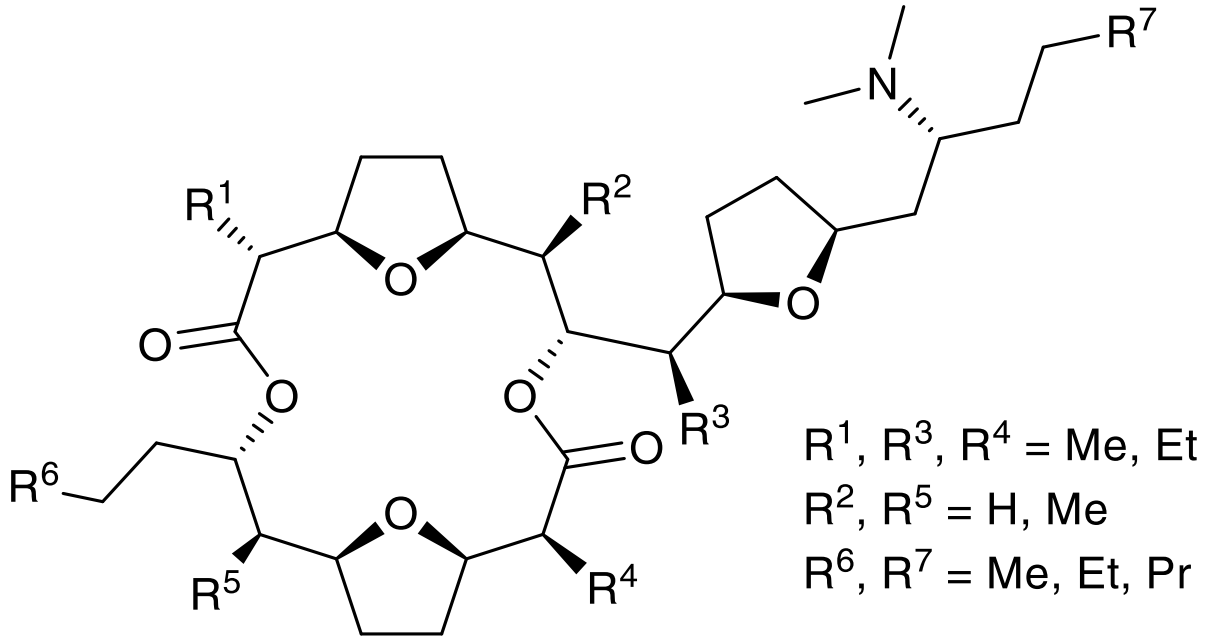

Figure 7: Basic structure of pamamycin with its substitution pattern known so far

In addition to the determination of the molecular mass by means of MS, the analysis of the number of methyl groups and their coupling pattern by means of proton NMR provides a first insight into the structure of the investigated pamamycins.

Compound **1**, *m/z* 635.4761 (Figure 8, 1) shows only five triplet methyls in addition to the two singlet N-methyl signals in its proton spectrum (Figure SX1a). Two of them represent the chain ends methyl H-18 and methyl H-11' of the two hydroxyl acid fragments (Figure SX1b). This leaves three methyl groups, each of which is presumably attached as part of an ethyl group to the previously known stereogenic centres C-2, C-7, C-9, C-2' or C-7', respectively. Since all known pamamycins-635, pam-635 A, pam-635 F [23, 24] and bis-homo-pamamycin-635 A [25], bear at least two doublet methyls, compound **1** must be a new pamamycin. 2D NMR measurements support this assumption. According to the analysis of ^{1}H-^{1}H COSY, HSQCED, HSQC-TOCSY and HMBC (Figure SX3- SX6), the three ethyl residues are found at C-2, C-9 and C-2' (Figure 8, 1). In analogy to the previous naming, we call this new compound **1** pamamycin-635 G.

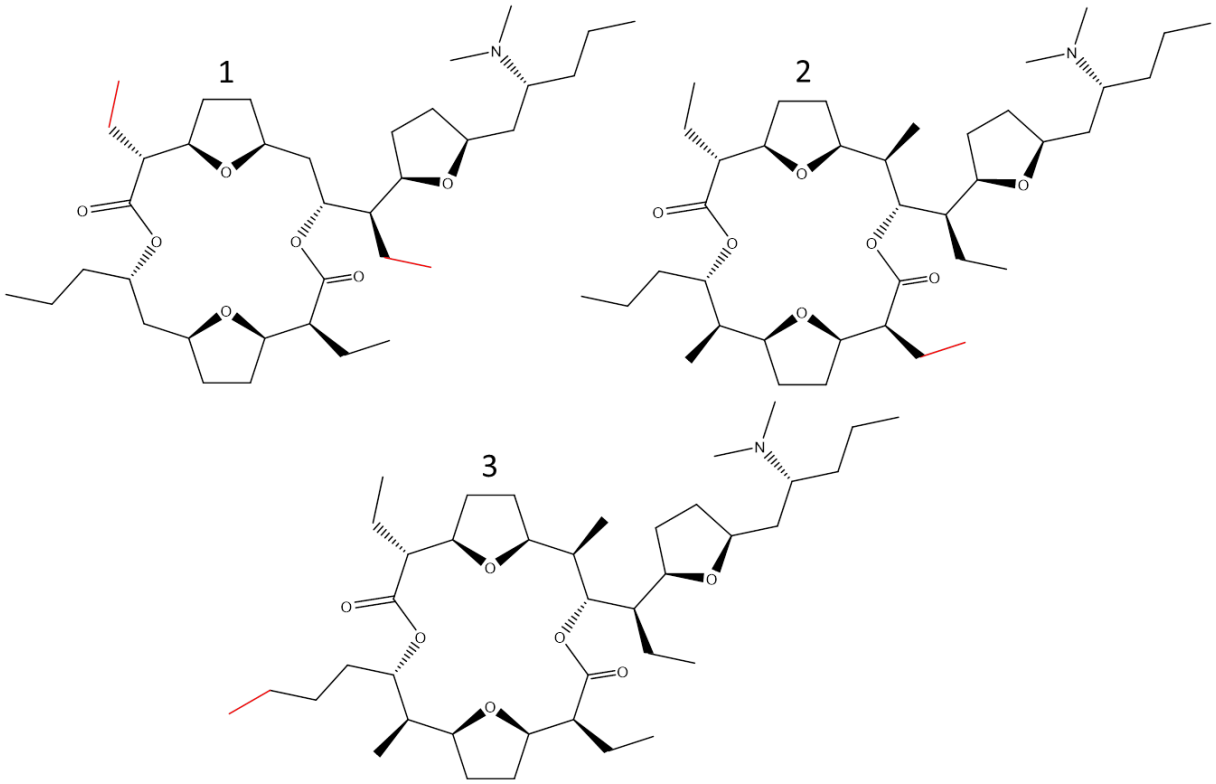

Figure 8: Structures of pamamycin 635 G (1), pamamycin 663 A (2) and homopamamycin 677 A (3). Red marked are the differences to already described pamamycins.

The data of compound **2** are close to those of pam-635 G. However, its MW (*m/z* 663.5074 $[M+H]^+$) is 28 units higher and two doublet methyl newly appear in the proton spectrum (Figure SY1a and SY1b). Their positions at C-7 and C-7' together with the complete assignment (Table SY) can be deduced from the relevant 2D NMR spectra (Figure SY3-SY6). As far as we know, no decided structure for a pamamycins with a MW of 663 has been published yet. However, since further structural variants are conceivable, we call this compound pamamycin-663 A (Figure 8, 3).

Despite multi-step purification and repeated chromatography by HPLC, compound **3** (*m/z* 677.5231 $[M+H]^+$) still turns out to be a 3 to 1 mixture of two structurally closely related pamamycins. Nevertheless, after extensive NMR analysis and comparison of the data obtained with those of pamamycin-635 G and pamamycin-663 A, a proposed structure can be presented at least for the main component. Compared to pam-663 A, neither the number of methyl groups nor the chemical shift nor the coupling pattern of the main compound has changed significantly in the proton NMR (Figure SZ1a and SZ1b). This finding is consistent only with a chain extension of one of the two hydroxy acid fragments. A first indication of this position is provided by the ^{13}C NMR data of compound **3** (Figure SZ2) and the

comparison with those of pamamycin-635 G (Figure SY2). Only the former methyl group C-11' shows a small high-field shift of 0.2 ppm. All other methyl shifts remain almost the same. If we now examine the connectivities to its two nearest methylene groups, we detect significant differences in their ^{13}C shifts to the comparative groups at pam-635 G. All this leads to the conclusion that the main compound **3** is a homologue of pam-663 A, extended by C-12´ in the lower fragment. We therefore name it homopamamycin 677 A (Figure 8, 3). It should be noted that the assignments for homopamamycin 677 A were only possible after extensive selective 1D TOCSY measurements (Figure SZ7a- SZ7g).

Activity analysis of different pamamycins

The activity of a range of "small" and "heavy" pamamycins was tested against different gram-positive and gram-negative strains, as well as cell lines, a fish test systems and the plant *Agrostis stolonifera* (Table S5). The minimum inhibitory concentration (MIC) values of pamamycins with MW's of 649 Da and pam-663 A for *Bacillus subtilis*, *Enterococcus faecium*, *Mycobacterium smegmatis*, *Streptococcus pneumoniae*, *Staphylococcus aureus* and methicillin resistant *S. aureus* (MRSA) is in the range of 1 to 8 µM. Whereas the MICs of pamamycins with 607 Da for the same strains is in the range of 8-16 µM. Thus, pamamycins with higher MW seems to be more active against gram-positive bacteria, than the ones with lower MW. However, all pamamycins are not highly potent against gram-negative bacteria, which is in accord with the earlier published data [6].

Interestingly, all tested pamamycins showed high activity against both carcinoma (KB-3.1) and hepatocyte (HepG2) cancer cell lines. Just like in the case with gram-positive bacteria, heavy pamamycins are more active than pamamycins with an MW of 607 Da. The difference in the activity is in average 100-300 times. Two pamamycin mixtures pam-649 and pam-663 A showed to be very potent against HepG2 cancer cells. The half maximal inhibitory concentration values are in the nanomolar range, 10 and 2 nM, respectively. The pamamycin sample pam-663 A is in average 300 times more active against HepG2 cells than the pamamycin sample pam-607. All tested pamamycins are toxic against zebrafish on different stages of development, and have only some minor activity against the plant *Agrostis stolonifera*.

Elevation of the pamamycin resistance of the production strain led to further increase in the production

To determine whether increased resistance improves the production rate and tolerance of the strain to the compound of interest, *S. albus* J1074 pamamycin resistance mutants were selected. For this purpose, a *S. albus* R2 strain was used. A pamamycin fraction was prepurified with the help of size-exclusion chromatography and used to select pamamycin-resistant *S. albus* J1074 strains.

Mutant P100I was used for the expression of the R2-73 cosmid. All tested P100I R2-73 exconjugants were viable on MS and other tested media. They did not show any significant differences in growth and morphology in comparison to the wild-type strain. HPLC-MS analysis of the extracts from this recombinant strain and the *S. albus* R2-73 control strain revealed the presence of pamamycins in both extracts (Figure 9). Furthermore, the production level of pamamycins in the resistant mutant was in average two times higher than that of the control strain and the other selected mutants.

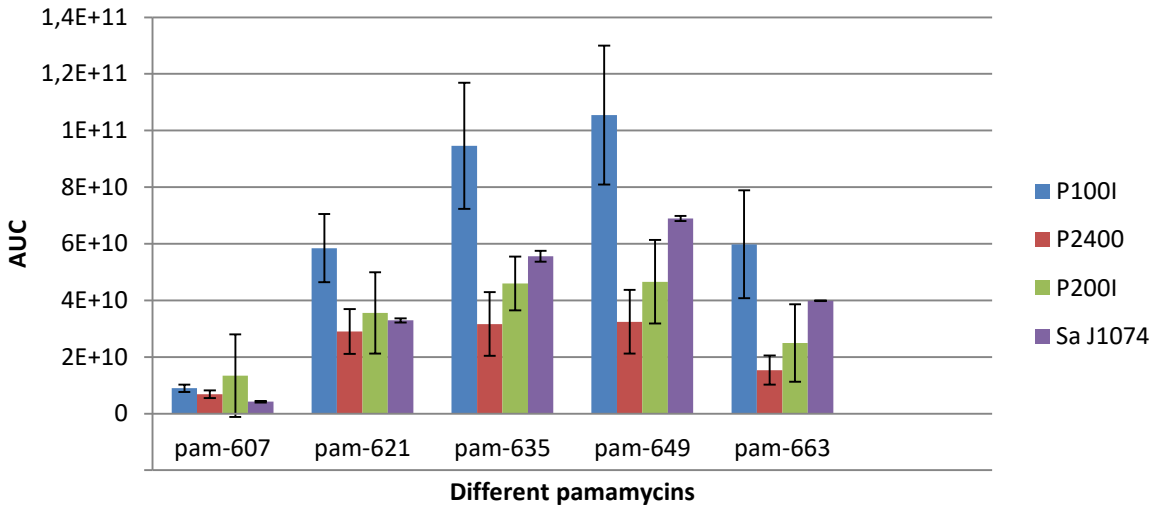

Figure 9: Pamamycin production in the *S. albus* mutants strains with enhanced resistance to pamamycin containing the modified R2-73 cosmid. The resistant mutants were selected on rising pamamycin concentrations. *S. albus* P1001 was selected on 100 µg/ml, *S. albus* P2400 on 400 µg/ml and *S. albus* P200I on 200 µg/ml pamamycin.

To determine the putative genetic reasons for the elevation of the resistance to pamamycins the whole genome sequencing of the P100I strain was performed and showed mutations in several drug resistance transporters.

2.3.5 Discussion

With the availability of fast and cheap sequencing techniques it was clear that the full secondary metabolite potential of *Streptomyces* has not been exhausted yet. Therefore, the huge potential of silent or barely expressed gene clusters could hide the capability for new natural drug leads. For this reason, methods for the activation or increased production of those gene clusters are in urgent need [26]. Here we demonstrated that decoupling of the transcription of pamamycin biosynthetic genes from the natural regulatory network as well as modulating its strength which led to the significant production of novel pamamycins.

By introducing synthetic derivatives of the *ermEP1* promoter the transcription of two main operons of the pamamycin cluster were rebalanced, which led to a 13 fold increased production of pamamycins. The changed transcription strength and timing can be seen by comparing the analysis of the synthetic promoters with the GUS assay (Figure 5). *S. albus* J1074 strains containing the pGUS constructs with the upstream promoters of R2-73 and R2-100 show higher activity than the upstream promoters of R2 and R2-67. Contrary to this, the downstream promoters of R2-67 and R2-100 are showing a high activity and the one of R2-73 only medium activity. As mentioned above, the native up- and downstream R2 promoters only show low activity. When comparing the above mentioned observations with the production spectrum of the respective strains (Figure 4), it can be said that the enhanced transcription of the upstream operon led to the shifted production spectrum of pamamycins. This operon contains three upstream oriented genes. The first two genes are coding for ketosynthases and the third gene is coding for the acyl-carrier protein (ACP). Overexpression of ACP's has already previously led to increased production of tetracenomycin and intermediates thereof [27]. For this experiment we can assume that the overexpression of the ACP led to increased binding of malonyl-CoA's. Once the ACP competes with other enzymes for the substrate there is a lack of malonyl-CoA's and the cell increases the production of the missing substrates. When analysing the spectrum of produced pamamycins an increase in high MW pamamycins can be seen. From those data it can be inferred that the ACP preferably binds to ethylmalonyl-CoA. Previous deletion experiments of *pamC* led to a a shifted spectrum with only lower molecular weight pamamycins, which is supporting this idea [8].

Referring to this data, upregulation of the genes from the left operon should result in an increased production of pamamycins 649 and 663. Therefore, 2 modified R2 clusters were constructed to check for the influence of the created upstream and downstream promoters of

R2-73. A hygromycin cassette was introduced into the R2 cluster by homologous recombination with only one defined promoter p73Up or p73Down. The spacer-region between *pamF* and *pamA* was left intact, allowing the native promoter the expression of the left or right operon. When comparing those 2 constructs to the strains containing R2 or R2-73 it can be seen that the overall production is about half the production of the control strains. In contrast to the lower production it can be seen that both constructs either with the optimised upstream or downstream promoter show higher production levels of pamamycins with a MW of 649 and 663 Da, respectively. Taking the previous observations into account, the increased expression of the ACP should lead to a shifted production. When the big downstream operon is overexpressed with the R2-73 promoter an increased presence of higher MW pamamycins is detectable. This can be explained by a higher downstream processing of the malonyl-CoA's and thereby a higher availability of the free ACP's. However, if the cascade of enzymes for the processing of the substrate is not balanced, the shift is not as big as in the balanced system. Therefore, overexpression of the acyl-carrier-protein *pamC*, the ketosynthases *pamF*, *pamG* or a combination of them leads to a shifted production spectrum. The shift is visible but the amount of pamamycins with higher MW is not as high as with the simultaneous use of the modified downstream and upstream promoter. This indicates that a balanced expression of the left and right operon is required and the R2-73 downstream promoter also has an important role for the expression. Another interesting observation is that the spacer sequences of the promoters for the R2-73 and R2-100 cluster vary. Therefore, even though they vary in their sequence they result in a similar production shift and should have a similar strength. It can also be seen that the production of pamamycins starts at a later time point and keeps on increasing continuously for strains containing the cluster with modified *ermEP1* promoters. In many biotechnological processes it has been found beneficial to have at first a growth phase and after a certain biomass is reached the production phase is initiated. Often the product itself is toxic and by shifting the production to a later time point higher amounts could be obtained.

In addition improved production could be obtained by expression of the optimised cluster in a strain with improved resistance. The overexpression of the putative resistance gene *pamW* with the strong p21 promoter did not lead to an increased production but to less pamamycins in the biomass. Thereby it can be assumed that PamW indeed exports pamamycins but the pamamycin production is still limited by the toxicity of this compound. This assumption was further examined with the help of activity tests. In general it can be seen that pamamycins are

active against Gram-positive bacteria, immortalised cell lines, zebra fish as well as plants. Against Gram-negative bacteria the demonstrated activity is low. But once an important resistance factor in *E. coli* is knocked out it shows improved activity against it (*E. coli* ΔacrB). *AcrB* codes for an efflux pump and is located in the periplasmic space and inner membrane. Pamamycins are lipophilic and could potentially induce pore formation in the membrane. This would explain the higher resistance of Gram-negative bacteria and the high activity against eukaryotic cells as well as against Gram-positive bacteria. However, this would need further investigations. The activity of pamamycin 663 A is in the lower nanomolar range against cancer cell lines. This could be quite interesting as far as some anticancer compounds have a similar activity range [28].

Also a rational improved resistant strain was created by cultivation of *S. albus* J1074 on medium with rising concentrations of pamamycin. The pamamycin production with the optimised R2-73 cluster increased about 2-fold compared to the not optimised strain resulting in a not 13 but 26 times increased production of pamamycins with a MW of 663 Da. The target of pamamycins has not been described yet. The obtained strains with increased resistance showed several mutations, which indicate an accumulation of several resistance factors. One mutation was in the repressor of a multidrug resistance factor, leading to an in general increased resistance.

Due to the increased production, the isolation of pamamycins with a MW of 663 and 677 Da was possible. Pamamycins with a MW of 691 Da have also been purified but the structure could not be elucidated due to mixtures with compounds of the same MW. Sufficient amounts of pamamycins were collected to solve their structure and perform activity tests, which revealed new interesting potential for pamamycins. Previously only structures and activities of pamamycins up a MW of 649 Da were described. So far, 18 structures have been described but taking into account the general structure and proposed biosynthesis, at least 288 structures are possible. According to the activity data, pamamycins with an increased MW seem to be more active than lower MW pamamycins. This contradicts the so far observed trend discussed in the literature that smaller pamamycins are more active [23, 29]. However, different pamamycins were tested and different test strains were used, preventing direct comparability of these tests. Further analysis needs also to be performed on the position of the different side-chains, due to a possible dependence of the activity on the position of the methyl, ethyl and propyl -groups.

Summarising the above listed data, a previously described method [19] has been utilised to effectively shift the production of the low yield secondary metabolite pamamycin towards higher amounts of a barely detectable derivative. This shifted production could also be obtained in a closely related strain for optimised production. The increased amounts enabled the structure elucidation of 3 new derivatives. Additionally, activity tests of selected derivatives were performed, to investigate the potential of pamamycins. Further strain optimisation resulted in increased production. Therefore, this investigation successfully demonstrates that random rational promoter optimisation is a fast method for the increase of production. Potentially, this method could also be used for the activation of not expressed clusters with few operons in a heterologous host, for uncoupling the cluster from the strict regulation network. Furthermore, pamamycins exhibit a broad range of activities. Pamamycins with a MW of 663 Da are 300 times more active against HepG2 cells than pamamycins with a MW of 607 Da and has 8 times higher activity against Gram-positive bacteria. Pamamycins demonstrate good activity against immortalised cell lines and therefore could be an interesting drug lead for further development in this direction. However, the mode of action of this antibiotic still needs to be uncovered and thereby offers potential for future studies.

2.3.6 References

1. Hutchings, M.I., A.W. Truman, and B. Wilkinson, *Antibiotics: past, present and future.* Current opinion in microbiology, 2019. **51**: p. 72-80.
2. Schneider, Y.K., *Bacterial Natural Product Drug Discovery for New Antibiotics: Strategies for Tackling the Problem of Antibiotic Resistance by Efficient Bioprospecting.* Antibiotics, 2021. **10**(7): p. 842.
3. Renwick, M.J., D.M. Brogan, and E. Mossialos, *A systematic review and critical assessment of incentive strategies for discovery and development of novel antibiotics.* The Journal of antibiotics, 2016. **69**(2): p. 73-88.
4. Ren, H., B. Wang, and H. Zhao, *Breaking the silence: new strategies for discovering novel natural products.* Current opinion in biotechnology, 2017. **48**: p. 21-27.
5. Chou, W.G. and B.M. Pogell, *Mode of action of pamamycin in Staphylococcus aureus.* Antimicrobial agents and chemotherapy, 1981. **20**(4): p. 443-454.
6. Lefèvre, P., et al., *Antimycobacterial activity of synthetic pamamycins.* The Journal of antimicrobial chemotherapy, 2004. **54**: p. 824-7.
7. Fischer, P., et al., *A General Sultone Route to the Pamamycin Macrodiolides—Total Synthesis of Pamamycin-621A and Pamamycin-635B.* Angewandte Chemie International Edition, 2005. **44**(38): p. 6231-6234.
8. Rebets, Y., et al., *Insights into the Pamamycin Biosynthesis.* Angewandte Chemie International Edition, 2014. **54**.
9. Kang, E.J. and E. Lee, *Total synthesis of oxacyclic macrodiolide natural products.* Chemical reviews, 2005. **105**(12): p. 4348-4378.

10. Hobbs, G., et al., *Dispersed growth of Streptomyces in liquid culture.* Applied Microbiology and Biotechnology, 1989. **31**(3): p. 272-277.
11. Kieser, T., et al., *Practical streptomyces genetics*. Vol. 291. 2000: John Innes Foundation Norwich.
12. J.F, S. and D. Russell, *Molecular Cloning: A Laboratory Manual (3-Volume Set)*. Vol. 1. 2001.
13. Siegl, T., et al., *Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes.* Metabolic engineering, 2013. **19**.
14. Megahed, S.H., et al., *Novel 2, 4-disubstituted quinazoline analogs as antibacterial agents with improved cytotoxicity profile: Modification of the benzenoid part.* Bioorganic & Medicinal Chemistry Letters, 2022: p. 128531.
15. Richter, L.H., et al., *Tools for studying the metabolism of new psychoactive substances for toxicological screening purposes–a comparative study using pooled human liver S9, HepaRG cells, and zebrafish larvae.* Toxicology Letters, 2019. **305**: p. 73-80.
16. Myronovskyi, M. and A. Luzhetskyy, *Native and engineered promoters in natural product discovery.* Natural product reports, 2016. **33**(8): p. 1006-1019.
17. Palazzotto, E., et al., *Synthetic biology and metabolic engineering of actinomycetes for natural product discovery.* Biotechnology advances, 2019. **37**(6): p. 107366.
18. Biggs, B., et al., *Multivariate modular metabolic engineering for pathway and strain optimization.* Current opinion in biotechnology, 2014. **29C**: p. 156-162.
19. Horbal, L., et al., *Secondary metabolites overproduction through transcriptional gene cluster refactoring.* Metab Eng, 2018. **49**: p. 299-315.
20. Strohl, W., *Compilation and analysis of DNA sequences associated with apparent Streptomyces promoters.* Nucleic acids research, 1992. **20**: p. 961-74.
21. Myronovskyi, M., et al., *Generation of a cluster-free Streptomyces albus chassis strains for improved heterologous expression of secondary metabolite clusters.* Metabolic engineering, 2018. **49**: p. 316-324.
22. Hanquet, G., X. Salom-Roig, and S. Lanners, *New insights into the synthesis and biological activity of the pamamycin macrodiolides.* CHIMIA International Journal for Chemistry, 2016. **70**(1): p. 20-28.
23. Natsume, M., et al., *Structure-activity relationship of pamamycins: effects of alkyl substituents.* The Journal of antibiotics, 1995. **48**(10): p. 1159-1164.
24. Natsume, M., et al., *The structures of four new pamamycin homologues isolated from Streptomyces alboniger.* Tetrahedron letters, 1991. **32**(26): p. 3087-3090.
25. Kozone, I., et al., *Structure-activity Relationship of Pamamycins: Effect of Side Chain Length on Aerial Mycelium-inducing Activity.* The Journal of antibiotics, 2008. **61**(2): p. 98-102.
26. Ochi, K. and T. Hosaka, *New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters.* Applied microbiology and biotechnology, 2013. **97**(1): p. 87-98.
27. DECKER, H., R.G. SUMMERS, and C.R. HUTCHINSON, *Overproduction of the acyl carrier protein component of a type II polyketide synthase stimulates production of tetracenomycin biosynthetic intermediates in Streptomyces glaucescens.* The Journal of antibiotics, 1994. **47**(1): p. 54-63.
28. Liston, D.R. and M. Davis, *Clinically relevant concentrations of anticancer drugs: a guide for nonclinical studies.* Clinical Cancer Research, 2017. **23**(14): p. 3489-3498.
29. Lefèvre, P., et al., *Antimycobacterial activity of synthetic pamamycins.* Journal of antimicrobial chemotherapy, 2004. **54**(4): p. 824-827.

2.3.7 Supplementary

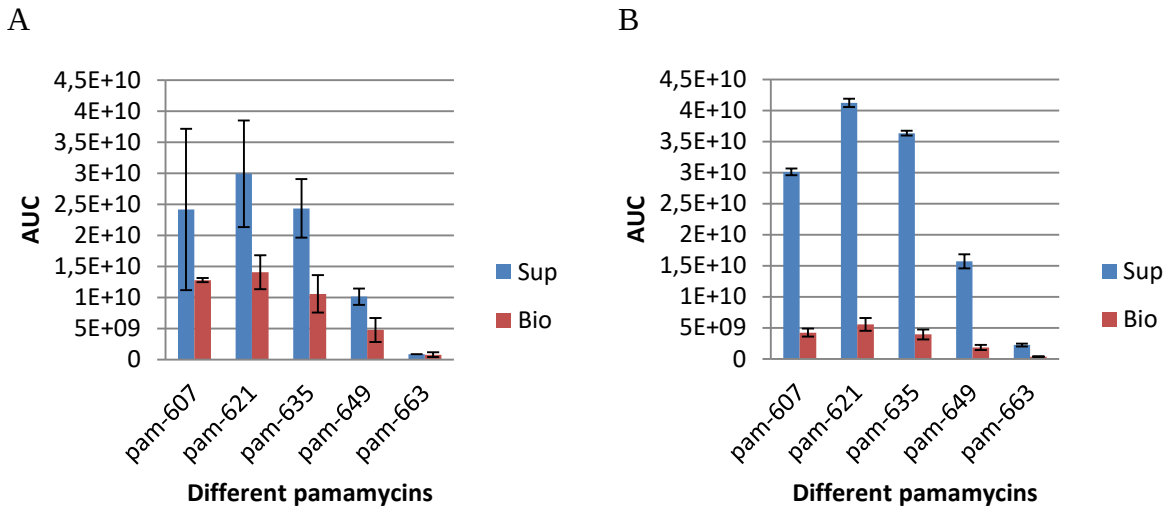

Figure S1: Comparison of the pamamycin production in the *S. albus* J1074 R2 strain and *S. albus* 1074 pTOS-P21pamW R2 with overexpression of *pamW*. (A) Pamamycin content of *S. albus* J1074 R2 in the supernatant and biomass. (B) Pamamycin content of *S. albus* J1074 pTOS-P21pamW R2 in the supernatant and biomass.

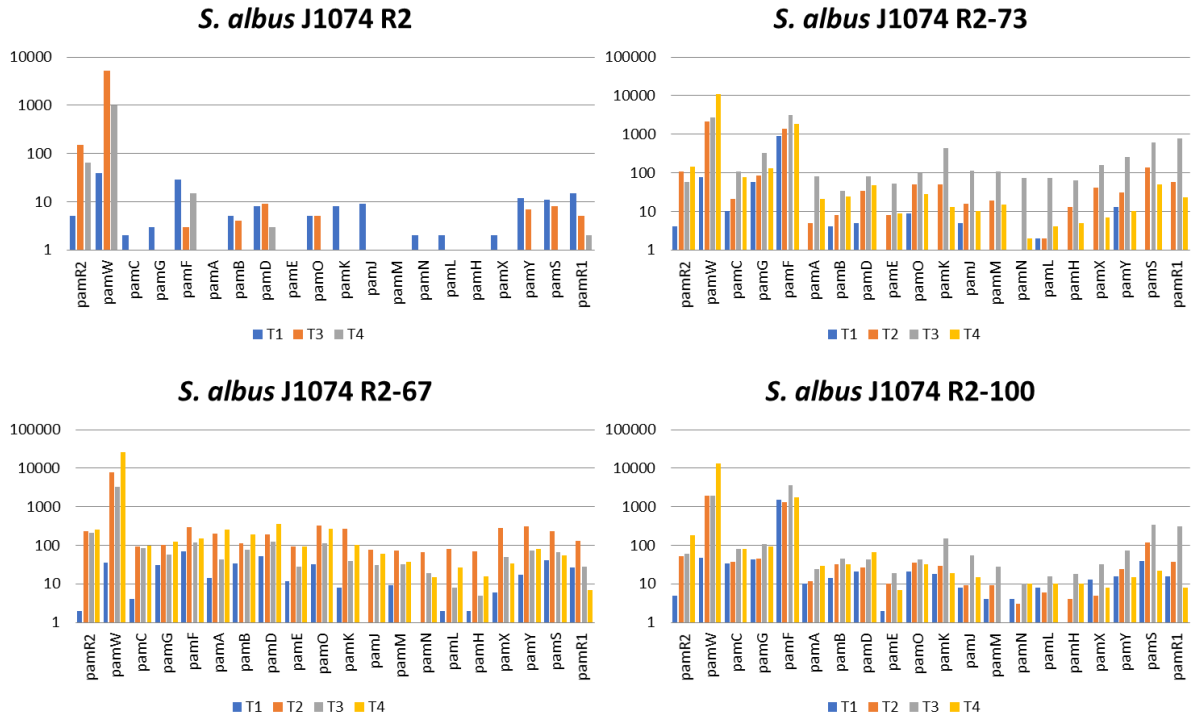

Figure S2: The expression levels of the genes in the R2 control cluster and engineered R2-73, R2-67 and R2-100 gene clusters in *S. albus* J1074.

Table S1: Bacterial strains and plasmids

Bacterial strains	Description	Source or reference
E. coli ET12567 (pUB307)	Conjugative transfer of DNA	[1]
E. coli GB05-redCC	Derivative of GB2005 containing integration of PBAD-ETgA operon	[2]
E. coli GB2005	*E. coli* strain used for the overexpression of vectors	[3]
S. albus J1074	*Streptomyces* strain used for heterologous expression	[4]
S. albus J1074 R2	*S. albus* J1074 containing the R2 cosmid which contains the genes for pamamycin synthesis	This study
S. albus J1074 pTOS-P21pamW	*S. albus* J1074 with overexpressed putative transport protein PamW	This study
S. albus J1074 pTOS-P21pamW R2	*S. albus* J1074 pTOS-P21pamW containing the R2 cosmid	This study
S. albus J1074 pTOS-P21pamW R2 library	*S. albus* J1074 pTOS-P21pamW with derivatives of the R2 cluster	This study
S. albus J1074 Del14	*S. albus* J1074 mutant for heterologous expression with 14 deleted secondary metabolite clusters	[5]
S. albus J1074 R2-67/R2-73/R2-100	*S. albus* J1074 strain with the R2-67 or R2-73 or R2-100 cluster	This study
S. albus J1074 del14 R2-73	*S. albus* J1074 del14 strain with the R2-73 cluster	This study
S. albus J1074 pGUS	*S. albus* J1074 strain with the pGUS plasmid	[6]
S. albus J1074 pGUS-p67up/p73up/p100up/pR2up	*S. albus* J1074 strain with the pGUS plasmid with different upstream promoters of R2 constructs cloned in front of the reporter gene	This study
S. albus J1074 pGUS-p67down/p73down/p100down/pR2down	*S. albus* J1074 strain with the pGUS plasmid with different downstream promoters of R2 constructs cloned in front of the reporter gene	This study
S. albus J1074 R2-p73up	*S. albus* J1074 strain with the R2 cluster and the semisynthetic promoter p73up	This study
S. albus J1074 R2-p73down	*S. albus* J1074 strain with the R2 cluster and the semisynthetic promoter p73down	This study
S. albus J1074 P100I	*S. albus* J1074 strain resistant to 100 µg/ml of a pamamycin mixture	This study
Plasmids/Cosmids	**Description**	**Source or reference**
pTOS	Apramycin resistance (Am^R), VWB-based Streptomyces integrative vector	[7]
pTOS-P21pamW	pTOS derivative containing *pamW* and Am^R	This work
R2	Derivative of pOJ436 containing the cluster for pamamycin biosynthesis	[8]
R2 mutant library	Random mutated R2 cosmids, containing a Hyg^R cassette flanked by 2 random promoters between the *pamF* and *pamA* genes	This work
R2-p73Down	R2 cosmid containing a Hyg^R cassette with the downstream oriented promoter of the R2 mutant R2-73	This work
R2-p73Up	R2 cosmid containing a Hyg^R cassette with the upstream oriented promoter of the R2 mutant R2-73	This work
pGUS	pSET152 derivative with *aadA* flanked by T4 terminators upstream of promoterless *gusA* gene	[9]
pGUS-pR2Up/p67Up/p73Up/p100Up	Derivative of pGUS containing the upstream promoter of The R2/R2-67/R2-73 or R2-100 cosmids cloned in front of the reporter gene	This work
pGUS-pR2Down/p67Down/p73Down/p100Down	Derivative of pGUS containing the downstream promoter of The R2/R2-67/R2-73 or R2-100 cosmids cloned in front of the reporter gene	This work

Table S2: Oligonucleotides used in this study

Pair Nr.	Name	Sequence (5'→3')	Purpose
1	PamWIntro-F	ACGATCTGTTCACGGCTCAGCGGCTGCCTCGCCGGCCGCTGCTTC	Amplification of *pamW*
	PamWIntro-R	GGCTGCTTCCTGTGGTCATGTGCGGGCTCTAACACGTCCTAGTAT	
2	pTOS-F	AAAATCCTGTATATCGTGCG	Primer pair for checking for introduction of *pamW*
	pTOS-R	ACTCAGACTCACTGAGGCTC	
3	PamPrandomHyg^R-F	GATCAGTGCGTCGCGACCGGGATCCTGCTGCTGATGCGCCATGCGACTTCCCTCCTTCANNNNNNATCCTANNNNNNNNNNNNNNNNNNAGCCNNNNNNCGCGGGATGTATCAGGCGCC	Primer pair for homologous recombination and introduction of degenerated promoter sequences
	PamPrandomHyg^R-R	CCCCGCGGTTTTTCCGTGCAGCGGATTCAGCCTGTGCAGCCATAACGATTCCTCCGATANNNNNNATCCTANNNNNNNNNNNNNNNNNNAGCCNNNNNAAATACTTGACATATCACTG	
4	Hyg^R-Seq-Down	AGGCTCGCGTAGGAATCATCC	Primer pair for partially sequencing of the introduced promoters attached to the Hyg^R cassette
	Hyg^R-Seq-Up	AACTGCATCTCAACGCCTTCC	
5	R2-P73-Down-F	GGATTCTCCCGGGTACCGGTGGGCCGATGTCCCTCGAATTCCCGGCTCAGGCGCCGGGGGCGGTGT	Primer for amplification of the downstream p73Down promoter and introduction with an Hyg^R cassette into R2
	R2-P73-Down-R	CCCCGCGGTTTTTCCGTGCAGCGGATTCAGCCTGTGCAGCCATAACGATTCCTCCGATACCTCGTATCCTATATGAATATCACGGGGATGAGCCGGGATAAATACTTGACATATCACTG	
6	R2-P73Up-F	TGCGTCGCGACCGGGATCCTGCTGCTGATGCGCCATGCGACTTCCCTCCTTCACAACTCATCCTAACTCATCGCTAGCCGTTGGAGCCGGTACTCGCGGGATGTATCAGGCGCC	Primer for amplification of the upstream p73Up promoter and introduction with an Hyg^R cassette into R2
	R2-P73Up-R	CGACCCGCACCGCTCTAGCGTCGGCGGCCTGTCGACAGCGGCACGCAAATACTTGACATATCACTG	
7	R2-P67Up-GU-R	AAAAGGTACCGCGACTTCCCTCCTTCAGCTGGTATCCTAATTTCTGGGTTTTTCTTCGAGCCCGGTGAAAATACTTGACATATCACTGT	Primer for the amplification of the P67-Up promoter and a Hyg^R cassette for the cloning into pGUS with restriction sites *Spe*I and *Xba*I
	R2-GU-F-all	AAAATCTAGATCAGGCGCCGGGGGCGGTGT	
8	R2-P67Down-GU-R	AAAAGGTACCAACGATTCCTCCGATAGCGA	Primer for the amplification of the P67Down promoter and a Hyg^R cassette for the cloning into pGUS with restriction sites *Spe*I and *Xba*I
	R2-GU-F-all	AAAATCTAGATCAGGCGCCGGGGGCGGTGT	
9	R2-P73Up-GU-R	AAAAGGTACCGCGACTTCCCTCCTTCACAACTCATCCTAACTCATCGCTAGCCGTTGGAGCCGGTACTAAATACTTGACATATCACTGT	Primer for the amplification of the p73Up promoter and a Hyg^R cassette for the cloning into pGUS with restriction sites *Spe*I and *Xba*I
	R2-GU-F-all	AAAATCTAGATCAGGCGCCGGGGGCGGTGT	
10	R2-P73Down-GU-R	AAAAGGTACCAACGATTCCTCCGATACCTC	Primer for the amplification of the p73Down promoter and a Hyg^R cassette for the cloning into pGUS with restriction sites *Spe*I and *Xba*I
	R2-GU-F-all	AAAATCTAGATCAGGCGCCGGGGGCGGTGT	
11	R2-P100Up-GU-R	AAAAGGTACCGCGACTTCCCTCCTTCATCTGCGATCCTAGCACGAAACAGGGTATGACAGCCTGTGTTAAATACTTGACATATCACTGT	Primer for the amplification of the P100Up promoter and a Hyg^R cassette for the cloning into pGUS with restriction sites *Spe*I and *Xba*I
	R2-GU-F-all	AAAATCTAGATCAGGCGCCGGGGGCGGTGT	
12	R2-P100Down-GU-R	AAAAGGTACCAACGATTCCTCCGATAGAGG	Primer for the amplification of the P100Down promoter and a Hyg^R cassette for the cloning into pGUS with restriction sites *Spe*I and *Xba*I
	R2-GU-F-all	AAAATCTAGATCAGGCGCCGGGGGCGGTGT	
13	PR2Down-GU-R	AAAAGGTACCACGGGAAAGCTATCCGCCGG	Primer for the amplification of the native downstream promoter and a Hyg^R cassette for the cloning into pGUS with restriction sites *Spe*I and *Xba*I
	R2-GU-F-all	AAAATCTAGATCAGGCGCCGGGGGCGGTGT	

14	R2-PL-Nat-R	AAAAGGTACCGAGACTCCCTGTGTTGCGTG	Primer for the amplification of the native upstream promoter for the cloning into pGUS with restriction site *Spe*I
	R2-PL-Nat-F	TATGTGAATCACAGTGATATGTCAAGTATTTGCCGGG AATTCGAGGGACA	
15	R2-PL-Nat-HygR-R	GGTGGGCCGATGTCCCTCGAATTCCCGGCAAATACTT GACATATCACTGT	Primer for the amplification of a Hyg^R cassette for the cloning into pGUS with restriction site *Xba*I
	R2-GU-F-all	AAAATCTAGATCAGGCGCCGGGGGCGGTGT	
16	R2-GU-F-all	AAAATCTAGATCAGGCGCCGGGGGCGGTGT	Primer for the amplification of the native upstream promoter and a Hyg^R cassette for the cloning into pGUS with restriction sites *Spe*I and *Xba*I
	R2-PL-Nat-R	AAAAGGTACCGAGACTCCCTGTGTTGCGTG	

Table S3: Screening for optimised pamamycin production of the mutanted R2 constructs in *S. albus* J1074 pTOS-P21pamW strain.

Name	Production	Production	Name
R2-1			R2-28
R2-2			R2-29
R2-3			R2-30
R2-4			R2-31
R2-5			R2-32
R2-6			R2-33
R2-7			R2-34
R2-8			R2-35
R2-9			R2-36
R2-10			R2-37
R2-11			R2-38
R2-12			R2-39
R2-13			R2-40
R2-14			R2-41
R2-15			R2-42
R2-16			R2-43
R2-17			R2-44
R2-18			R2-45
R2-19			R2-46
R2-20			R2-47
R2-21			R2-48
R2-22			R2-49
R2-23			R2-50
R2-24			R2-51
R2-25			R2-52
R2-26			R2-53
R2-27			

Name	Production	Production	Name
R2-54			R2-81
R2-55			R2-82
R2-56			R2-83
R2-57			R2-84
R2-58			R2-85
R2-59			R2-86
R2-60			R2-87
R2-61			R2-88
R2-62			R2-89
R2-63			R2-90
R2-64			R2-91
R2-65			R2-92
R2-66			R2-93
R2-67			R2-94
R2-68			R2-95
R2-69			R2-96
R2-70			R2-97
R2-71			R2-98
R2-72			R2-99
R2-73			R2-100
R2-74			R2-101
R2-75			R2-102
R2-76			R2-103
R2-77			R2-104
R2-78			R2-105
R2-79			R2-106
R2-80			

Table S4: Sequences of the semisynthetic promoters (Sequence (5'→3'))

Promoter Name	Spacer Region	RBS	SR	Random SR	-10 Region	Random SR	-35 Region	Random SR
Promoter up	GCGACTTC	CCTCC	TTCA	NNNNNN	ATCCTA	NNNNNNNNNNNNNNNNNNN	AGCC	NNNNNN
P67Up	GCGACTTC	CCTCC	TTCA	GCTGGT	ATCCTA	ATTTCTGGGTTTTTCTTCG	AGCC	CGGTGA
P73Up	GCGACTTC	CCTCC	TTCA	CAACTC	ATCCTA	ACTCATCGCTAGCCGTTGG	AGCC	GGTACT
P100Up	GCGACTTC	CCTCC	TTCA	TCTGCG	ATCCTA	GCACGAAACAGGGTATGAC	AGCC	TGTGTT
Promoter Down	AACGATT	CCTCC	GATA	NNNNNN	ATCCTA	NNNNNNNNNNNNNNNNNNN	AGCC	NNNNN
P67Down	AACGATT	CCTCC	GATA	GCGAAG	ATCCTA	AGACTGTACCACCTGAGCA	AGCC	GATCA
P73Down	AACGATT	CCTCC	GATA	CCTCGT	ATCCTA	TATGAATATCACGGGGATG	AGCC	GGGAT
P100Down	AACGATT	CCTCC	GATA	GAGGCT	ATCCTA	ACACTCGGTCCAGGTTGAC	AGCC	GTTAA

Table S5: Results of activity tests of pamamycins with different molecular weight against different test strains

Indicator strain	MIC [µM]		
	Pam-607	Pam-649	Pam-663
E. coli WT	> 128	32	32
E. coli ΔacrB	32	32	32
P. aeruginosa PA14	> 128	128	128
P. aeruginosa PA14 ΔmexAB	> 128	64	128
B. subtilis DSM-10	8	2	2
E. faecium DSM-20477	16	8	4
M. smegmatis mc^2155	16	8	8
S. aureus Newman	8	2	1
S. aureus N315 (MRSA)	16	2	1
S. pneumoniae DSM-20566	4-8	1	1
Cell line	IC_{50} [µM]		
	Pam-607	Pam-649	Pam-663
KB-3.1	0.48	0.06	0.02
HepG2	0.62	0.01	0.002
Danio rerio line and age	MTC [µM]		
	Pam-607	Pam-649	Pam-663
TL; 1 dpf (embryos)	< 10	< 10	< 10
TL; 3 dpf (larvae)	< 10	10	< 10

Agrostis stolonifera	Different pamamycins and their Concentration [µM] / Number of grown plants			
Dilution	Pam-621	Pam-607	Pam-663	Pam-649
0	40.25 / >5	41.2 / 0	37.7 / 0	38.5 / 0
1	20.125 / >5	20.6 / 2	18.85 / 4	19.25 / 3
2	10.0625 / >5	10.3 / 5	9.425 / 5	9.625 / >5
3	5.03125 / >5	5.15 / >5	4.7125 / >5	4.8125 / >5
4	2.515625 / >5	2.575 / >5	2.35625 / >5	2.40625 / >5

Table SX: NMR data (500 MHz, $CDCl_3$) for Pamamycin-635 G

	δ_C m	□□□$_H$ m (J in Hz)
1	172.80 C	----------
2	54.86 CH	2.13 m
3	81.77 CH	3.61 td (10.5, 5.0)
4	29.55 CH_2	1.29 m and 1.93 m
5	30.59 CH_2	1.47 m and 2.03 m
6	74.47 CH	3.79 m
7	38.90 CH_2	1.44 m and 2.13 m
8	70.06 CH	5.32 ddd (12.5, 3.0, 1.5)
9	49.33 CH	1.30 m
10	79.65 CH	3.42 dt (10.5, 7.0)
11	29.03 CH_2	1.58 m and 2.05 m
12	30.89 CH_2	1.52 m and 2.06 m
13	78.65 CH	3.80 br m
14	33.73 CH_2	1.73 m and 2.18 m
15	67.91 CH	3.40 m
16	28.72 CH_2	1.46 m and 1.72 m
17	19.56 CH_2	1.40 m and 1.50 m
18	13.86 CH_3	0.99 t (7.0, 3H)
1´	173.44 C	----------
2´	49.89 CH_2	2.34 dt (11.3, 2.5)
3´	79.15 CH	3.89 ddd (9.3, 7.5, 2.5)
4´	27.35 CH_2	1.70 m and 1.83 m
5´	31.44 CH_2	1.36 m and 1.97 m
6´	74.38 CH	3.72 m
7´	38.26 CH_2	1.83 m and 1.58 m
8´	71.38 CH	4.80 m
9´	36.62 CH_2	1.47 m and 1.62 m
10´	18.02 CH_2	1.30 m (2H)
11´	13.96 CH_3	0.89 t (7.5, 3H)
2-Et	22.04 CH_2	1.43 m and 1.54 m
	11.74 CH_3	0.89 t (7.5, 3H)
9-Et	18.92 CH_2	1.21 m and 1.52 m
	13.81 CH_3	0.95 (3H)
15-N(Me)$_2$	36.31 CH_{3a}	2.86 d (5.0, 3H)
	43.33 CH_{3b}	3.08 d (5.0, 3H)
2´-Et	16.02 CH_2	1.38 m and 1.75 m
	12.69 CH_3	0.85 t (7.5, 3H)
NH	----------	8.87 br s

Table SX1: HMBC Key correlations for Pamamycin-635 G

Position	HMBC correlations
H-2	C-1, C-3, C-2-Et
H-3	C-1, C-2, C-4, C-5, C-2-Et
H-6	C-3, C-8
H-8	C-6, C-7, C-9, C-10, C-1´, C-9-Et
H-10	C-8, C-9, C-9- C-2-Et
H-13	C-11, C-15
H-15	C-12, C-14, C-17, 15-N-CH_{3a}
H-18	C-16, C-17
H-2´	C-1´, C-3´, C-2´-Et
H-3´	C-1´, C-2´, C-4´, C-5´, C-2´-Et
H-6´	C-3´, C-8´
H-8´	C-1, C-6´, C-7´, C-10´
H-11´	C-9´, C-10´
CH_3-2 (from Et)	C-2
CH_3-9 (from Et)	C-9
CH_3-2´(from Et)	C-2´
CH_{3a}-N	C-15, 15-N-CH_{3b}
CH_{3b}-N	C-15, 15-N-CH_{3a}

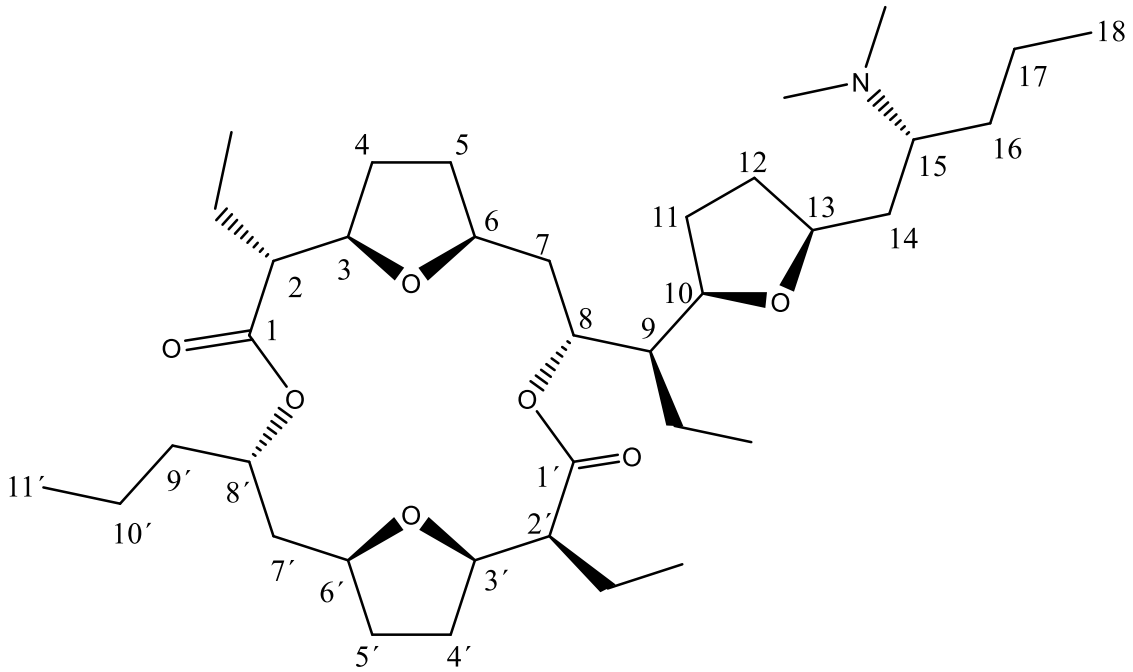

Figure SX: Structure of Pamamycin-635 G

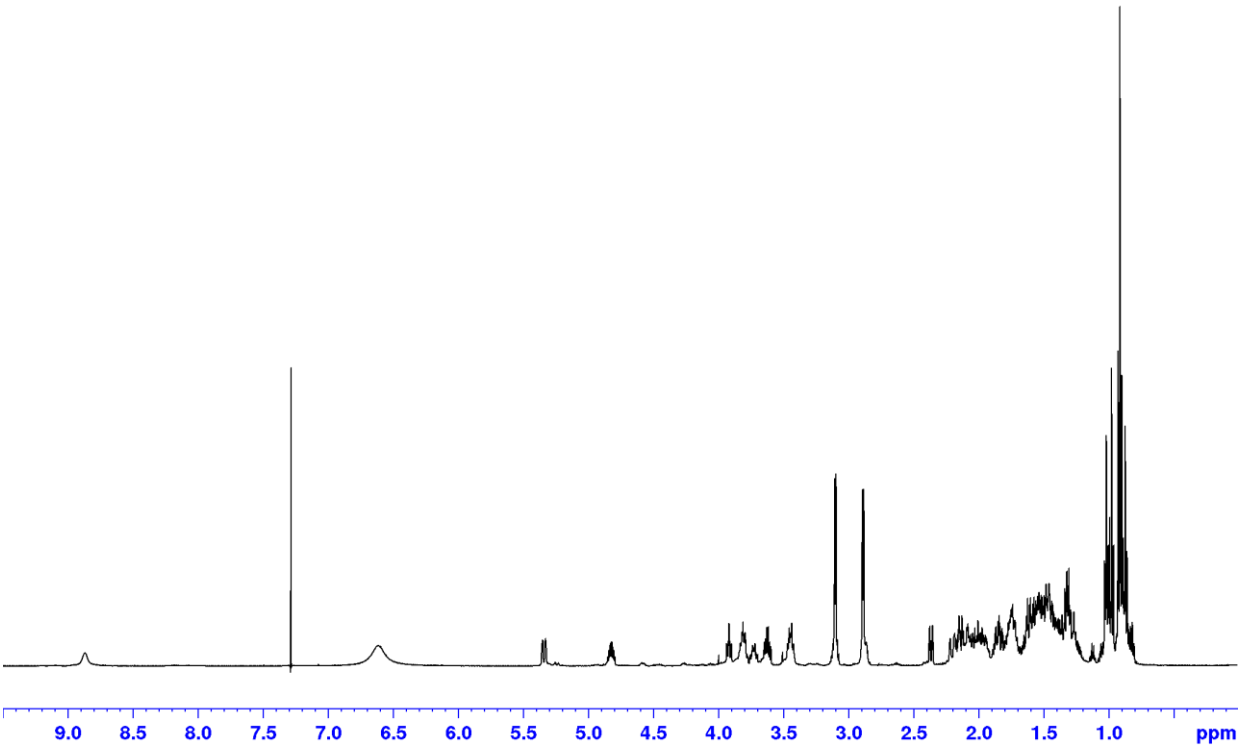

Figure SX1a: ^{1}H NMR spectrum (500 MHz, $CDCl_3$) of Pamamycin-635 G

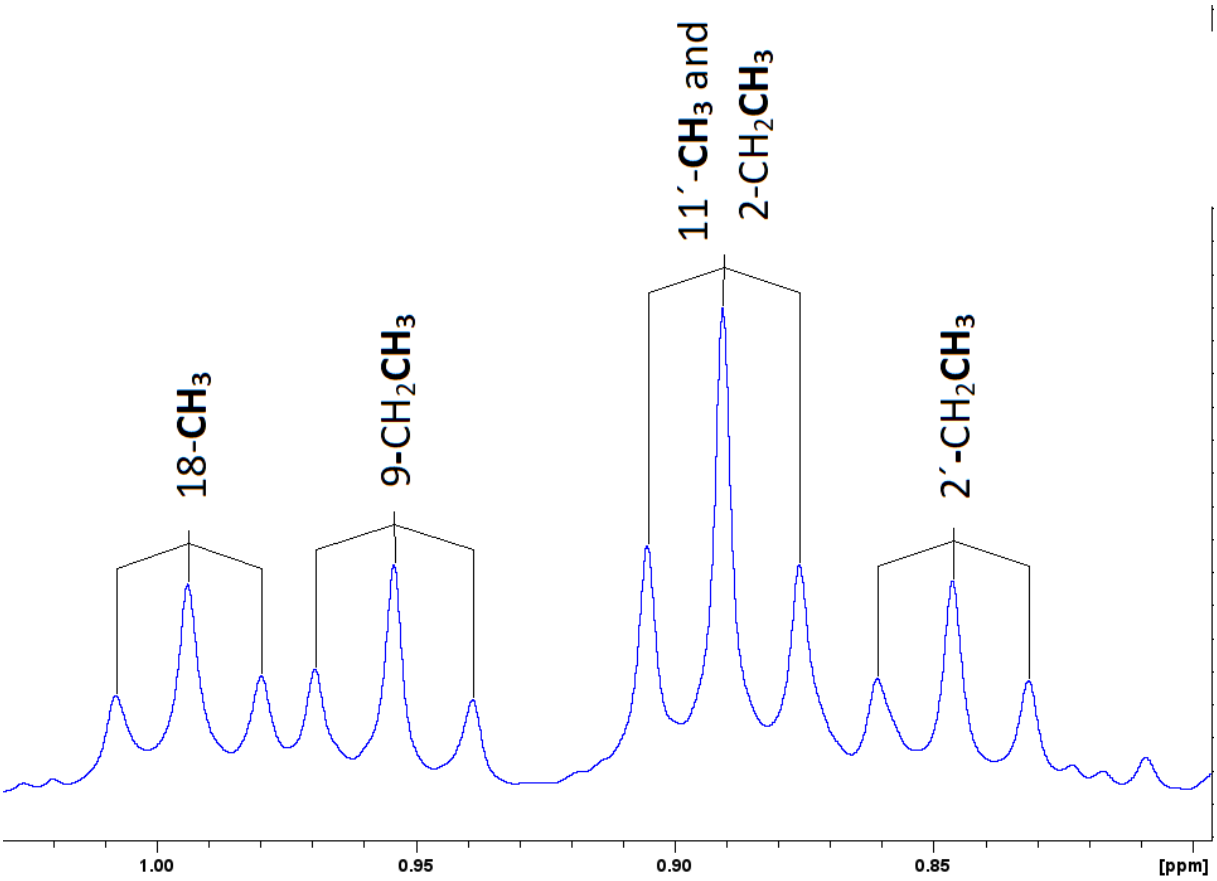

Figure SX1b: ^{1}H NMR spectrum (500 MHz, $CDCl_3$) of Pamamycin-635 G, methyl region

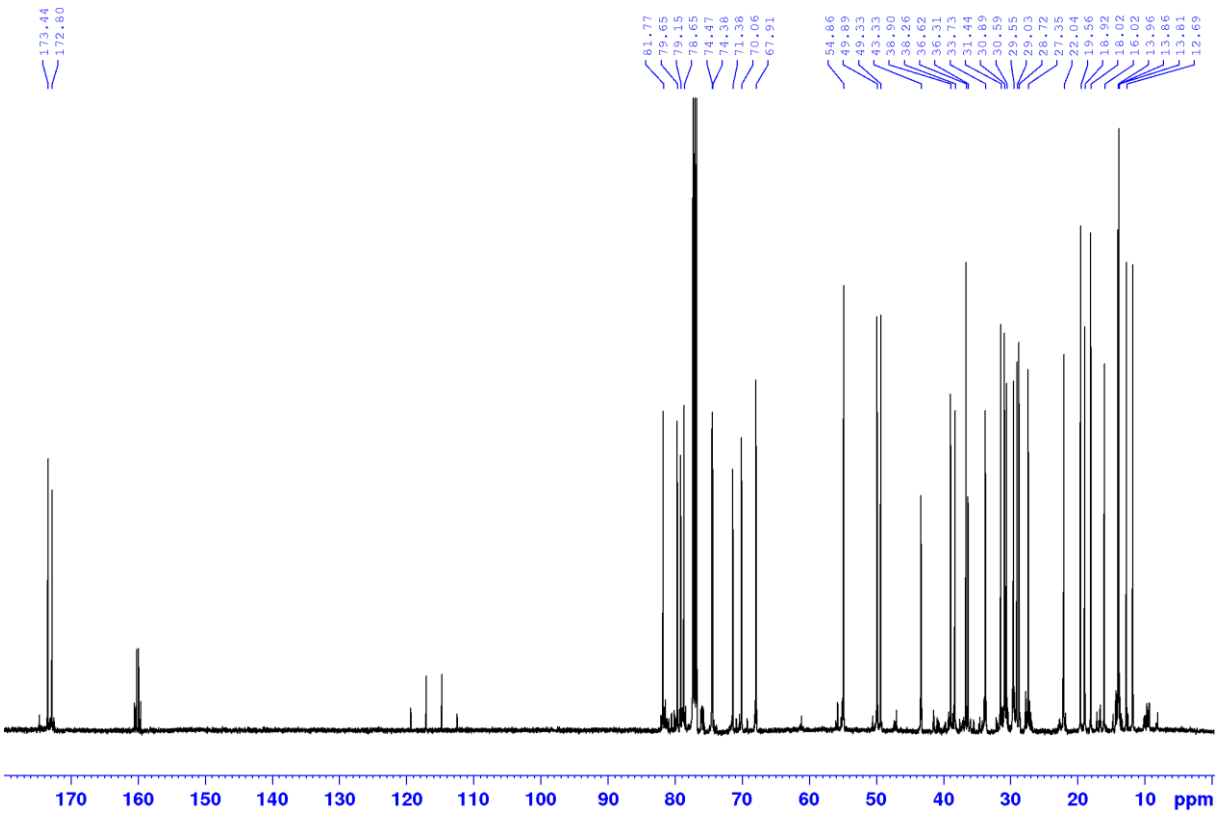

Figure SX2: ^{13}C NMR spectrum (125 MHz, $CDCl_3$) of Pamamycin-635 G

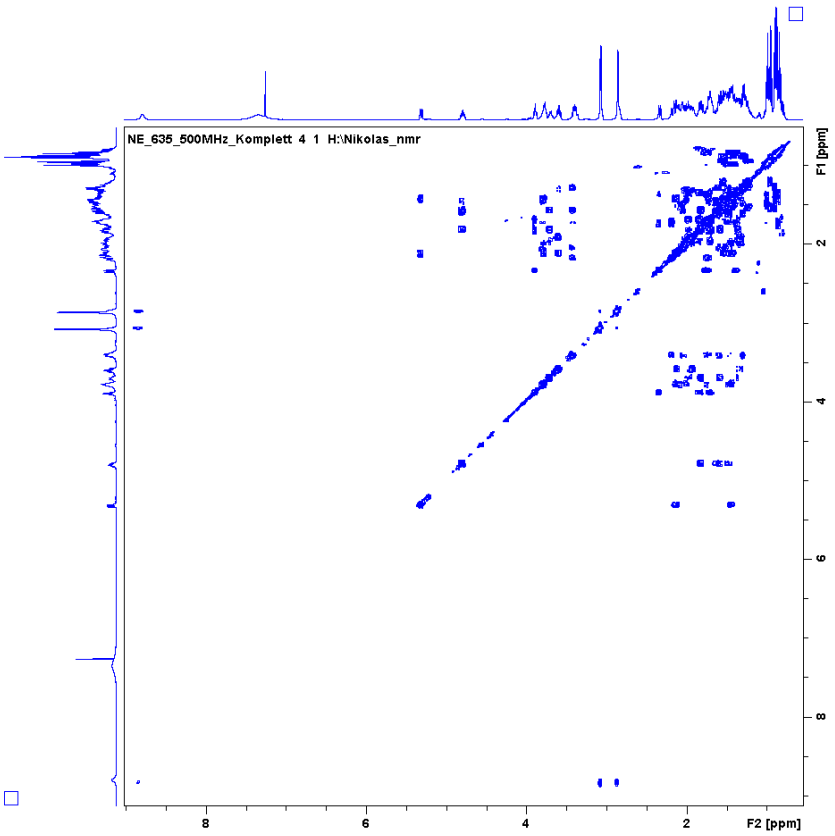

Figure SX3: ^{1}H-^{1}H COSY spectrum (CDCl3) of Pamamycin-635 G

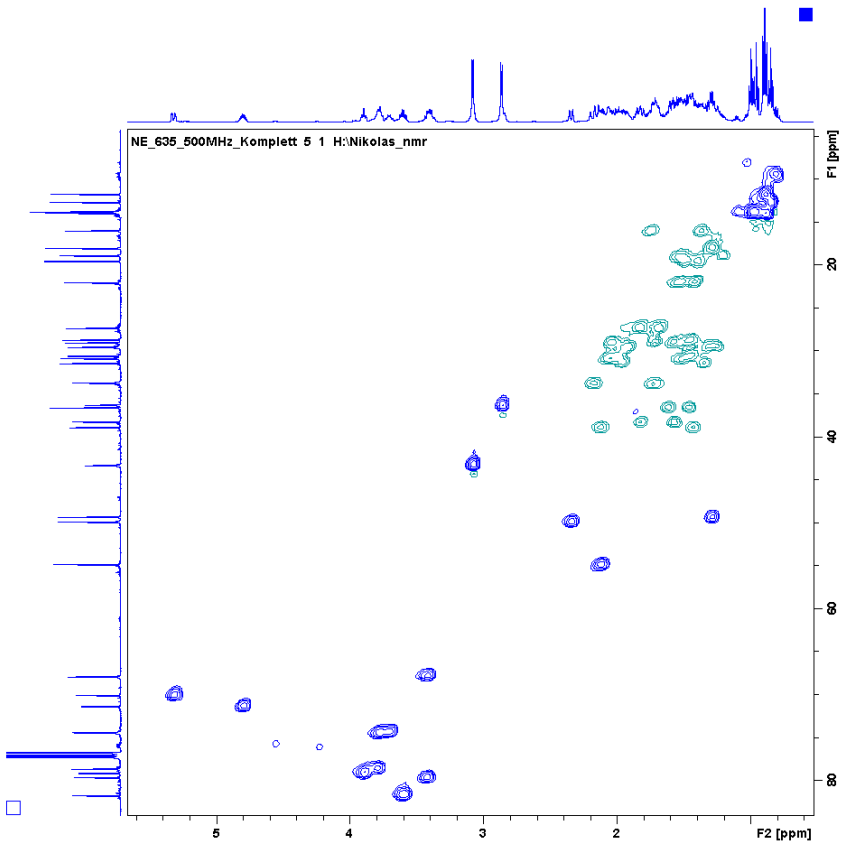

Figure SX4: Edited HSQC spectrum ($CDCl_3$) of Pamamycin-635 G

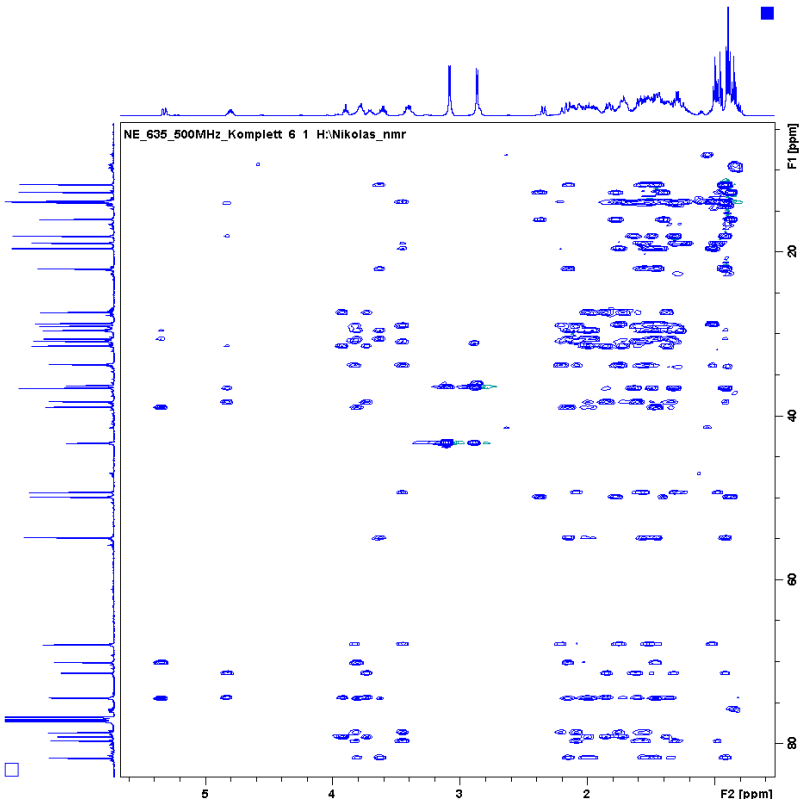

Figure SX5: HSQC-TOCSY spectrum (CDCl3) of Pamamycin-635 G

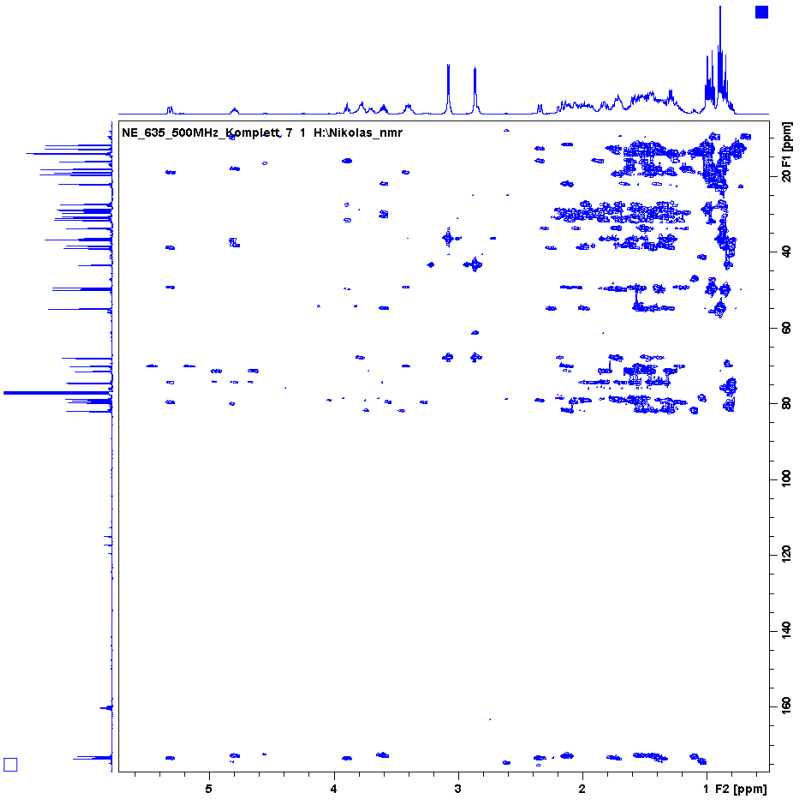

Figure SX6: HMBC spectrum (CDCl3) of Pamamycin-635 G

Table SY: NMR data (700 MHz, $CDCl_3$) for Pamamycin-663 A

	δ_C m	δ_H m (J in Hz)
1	172.31 C	----------
2	55.31 CH	2.11 ddd (14.0, 10.0, 4.0)
3	81.34 CH	3.66 ddd (11.0, 10.0, 4.2)
4	30.90 CH_2	1.20 m and 1.94 m
5	27.51 CH_2	1.71 m and 1.94 m
6	76.19 CH	4.21 ddd (9.0, 5.3, 2.2)
7	37.51 CH	1.93 m
8	75.20 CH	4.91 d (11.0)
9	46.41 CH	1.53 m
10	79.72 CH	3.36 dt (10.0, 7.0)
11	29.18 CH_2	1.61 m and 2.06 m
12	30.99 CH_2	1.53 m and 2.09 m
13	78.64 CH	3.75 br s
14	33.81 CH_2	1.71 m and 2.21 m
15	68.13 CH	3.39 br t (11.0)
16	28.72 CH_2	1.46 m and 1.72 m
17	19.61 CH_2	1.38 m and 1.52 m
18	13.81 CH_3	1.00 t (7.0, 3H)
1´	173.07 C	----------
2´	50.12 CH_2	2.36 dt (11.5, 2.5)
3´	78.13 CH	3.83 ddd (9.2, 6.0, 2.5)
4´	27.85 CH_2	1.62 m and 1.81 m
5´	26.70 CH_2	1.64 m and 1.77 m
6´	76.25 CH	3.94 td (7.6, 2.8)
7´	36.85 CH	1.87 m
8´	75.77 CH	4.57 dt (11.0, 4.0)
9´	34.24 CH_2	1.55 m and 1.68 m
10´	16.79 CH_2	1.25 m and 1.34 m
11´	14.21 CH_3	0.88 t (7.5, 3H)
2-Et	22.22 CH_2	1.42 m and 1.56 m
	11.75 CH_3	0.90 t (7.5, 3H)
7-Me	10.06 CH_3	0.82 d (7.0, 3H)
9-Et	18.98 CH_2	1.18 m and 1.54 m
	13.86 CH_3	0.95 (3H)
15-N(Me)$_2$	35.97 CH_{3a}	2.86 d (5.0, 3H)
	43.59 CH_{3b}	3.10 d (5.0, 3H)
2´-Et	15.99 CH_2	1.37 m and 1.77 m
	12.70 CH_3	0.83 t (7.5, 3H)
7´-Me	9.34 CH_3	0.82 d (7.0, 3H)
NH	----------	8.74 br s

Table SY1: HMBC Key correlations for Pamamycin-663 A

Position	HMBC correlations
H-2	C-1, C-3, C-4, C-2-Et
H-3	C-1, C-2, C-5, C-2-Et
H-6	C-7, C-8, -7-Me
H-7	C-5, C-6, C-8, C-9, C-7-Me
H-8	C-6, C-7, C-9, C-10, C-1´, C-7-CH_3, C-9-Et
H-10	C-8, C-9, C-9- C-2-Et
H-13	C-15
H-18	C-16, C-17
H-2´	C-1´, C-3´, C-2´-Et
H-3´	C-1´, C-5´, C-2´-Et
H-6´	C-7´, C-8´, 7´-Me
H-7´	C-8´, C-9´, 7´-Me
H-8´	C-1, C-6´, C-7´, C-9´, C-10´, 7´-Me
H-11´	C-9´, C-10´
CH_3-2 (from Et)	C-2
CH_3-7	C-7
CH_3-9 (from Et)	C-9
CH_3-2´ (from Et)	C-2´
CH_3-7´	C-7´
CH_{3a}-N	C-15, 15-N-CH_{3b}
CH_{3b}-N	C-15, 15-N-CH_{3a}

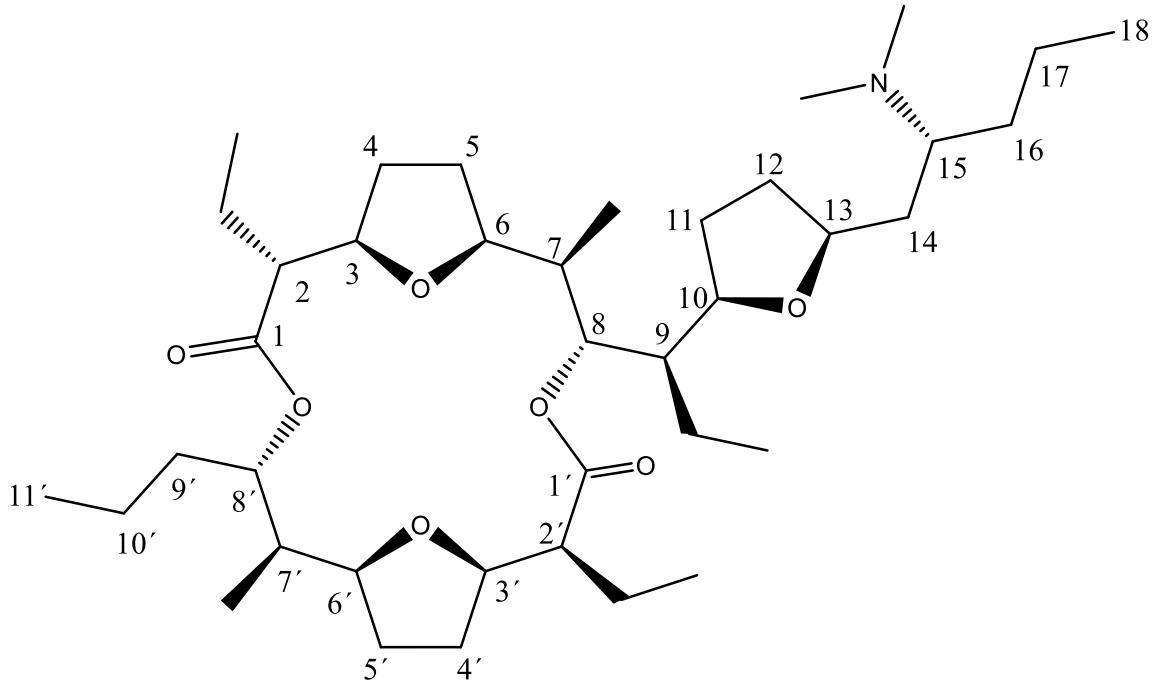

Figure SY: Structure of Pamamycin-663 A

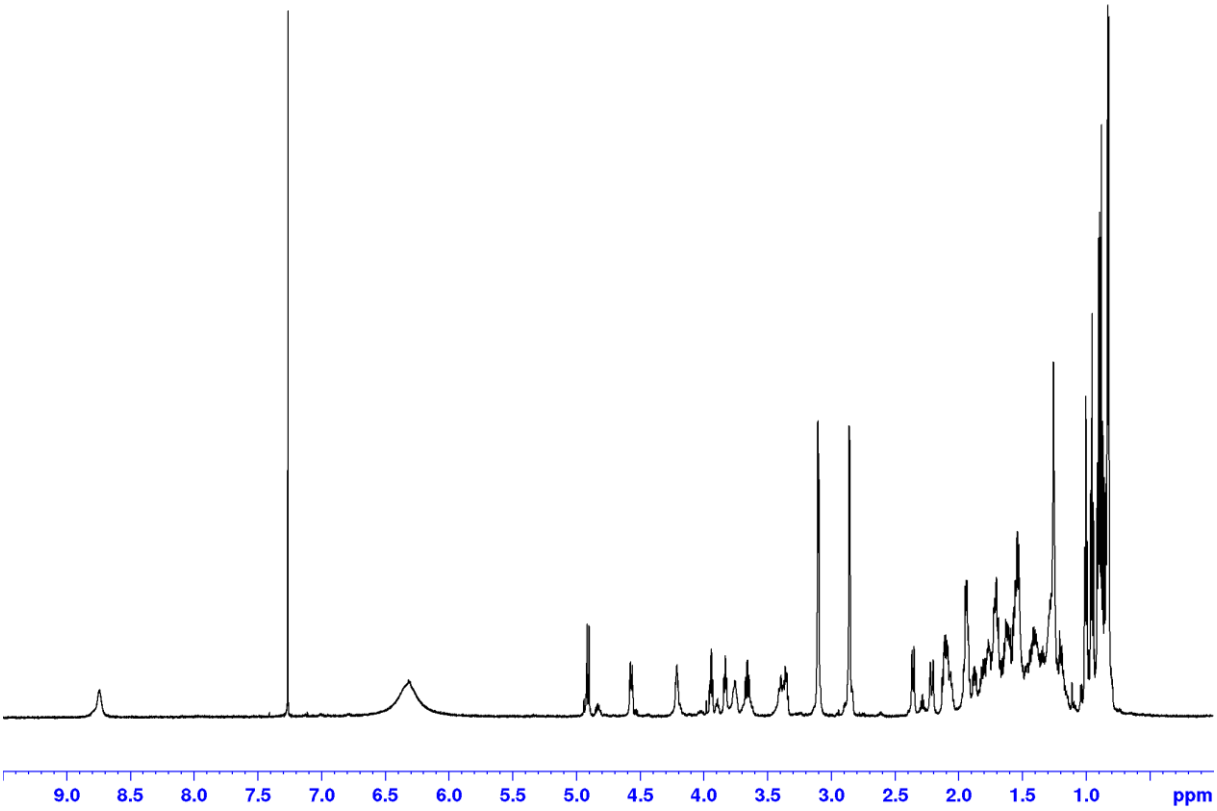

Figure SY1a: ^{1}H NMR spectrum (700 MHz, $CDCl_3$) of Pamamycin-663 A

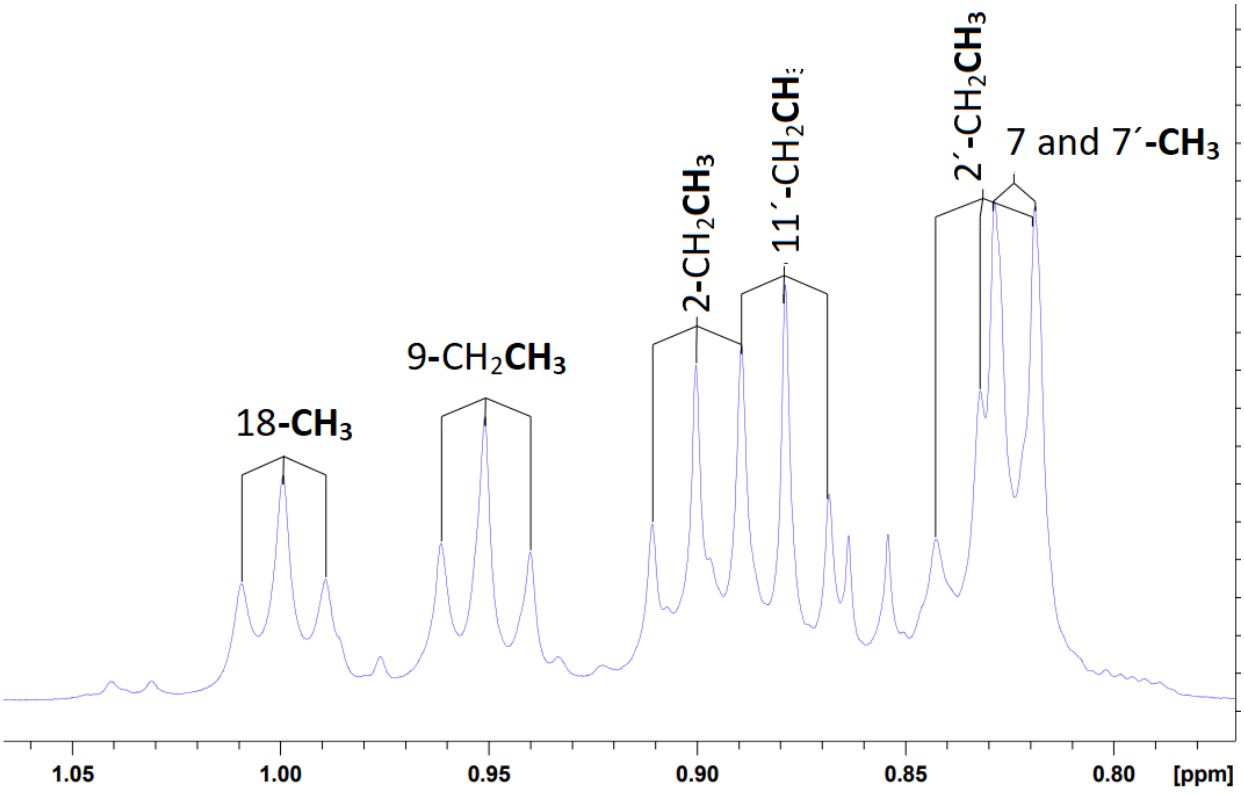

Figure SY1b: ^{1}H NMR spectrum (700 MHz, $CDCl_3$) of Pamamycin-663 A, Methyl region

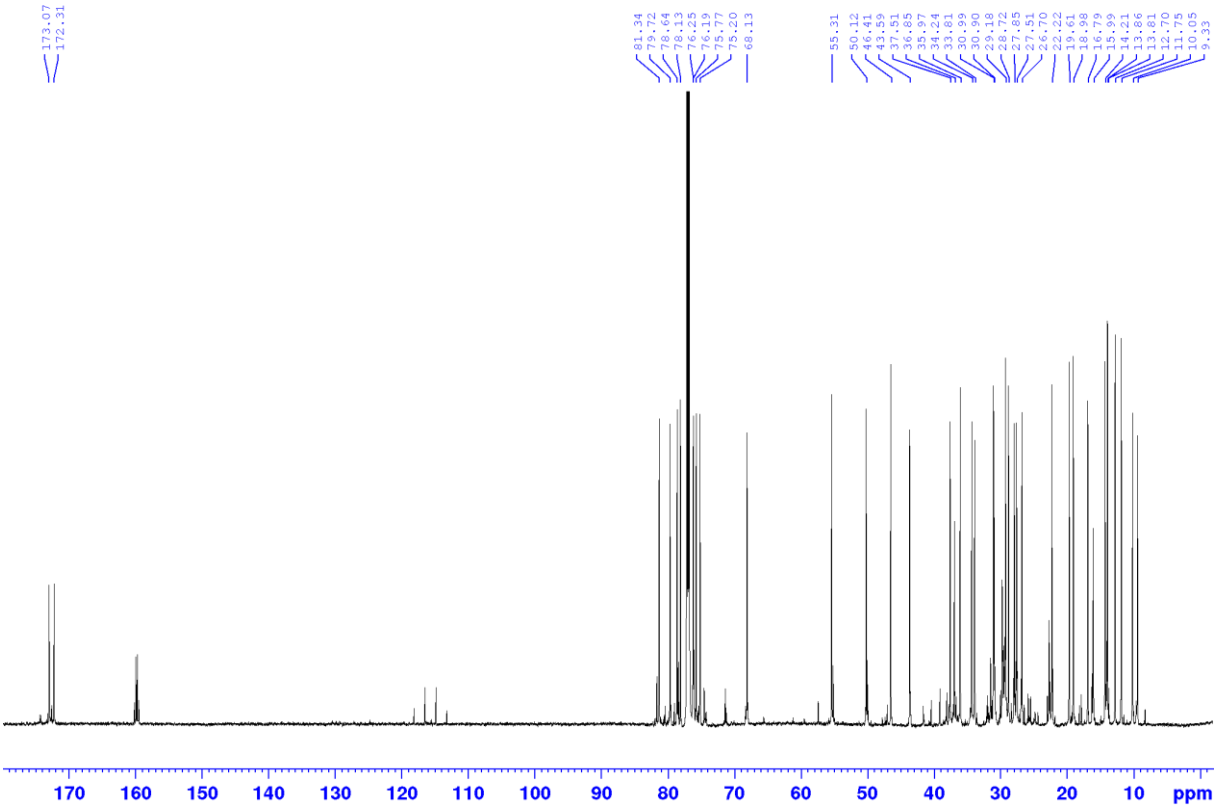

Figure SY2: ^{13}C NMR spectrum (175 MHz, $CDCl_3$) of Pamamycin-663 A

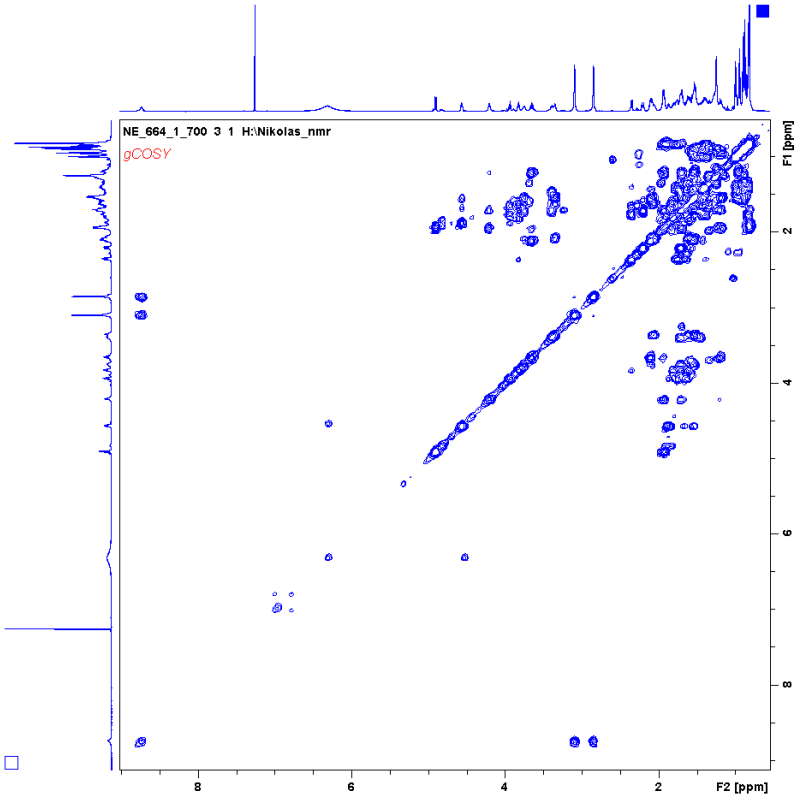

Figure SY3: ^{1}H-^{1}H COSY spectrum (CDCl3) of Pamamycin-663 A

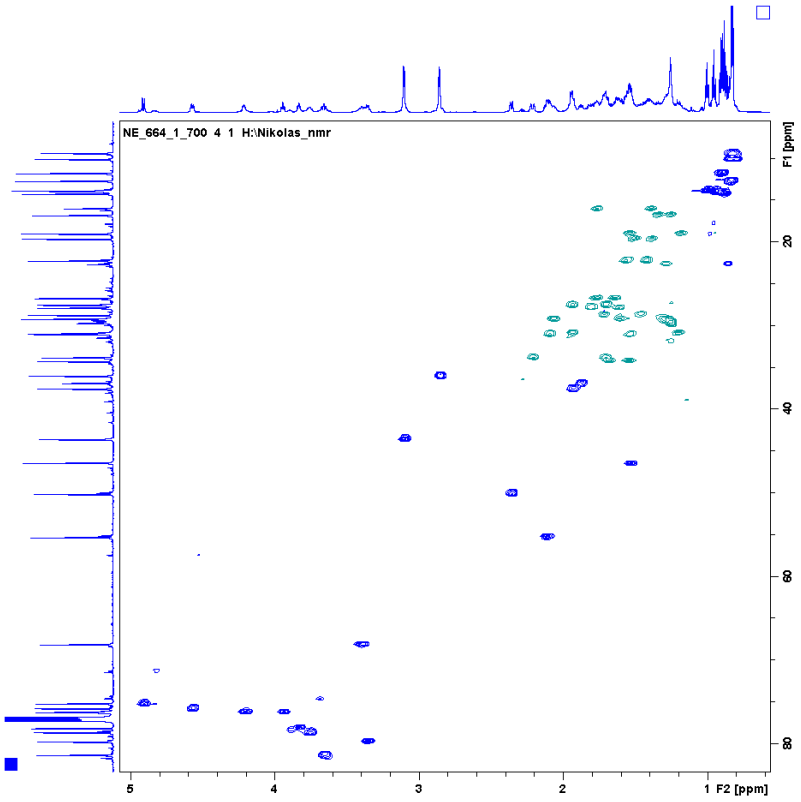

Figure SY4: Edited HSQC spectrum (CDCl3) of Pamamycin-663 A

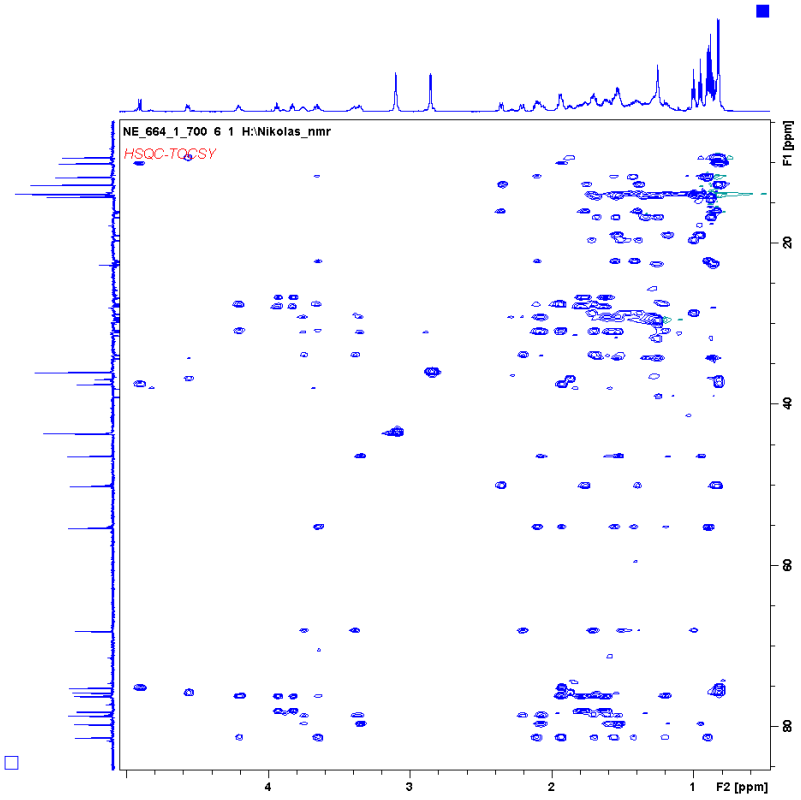

Figure SY5: HSQC-TOCSY spectrum (CDCl3) of Pamamycin-663 A

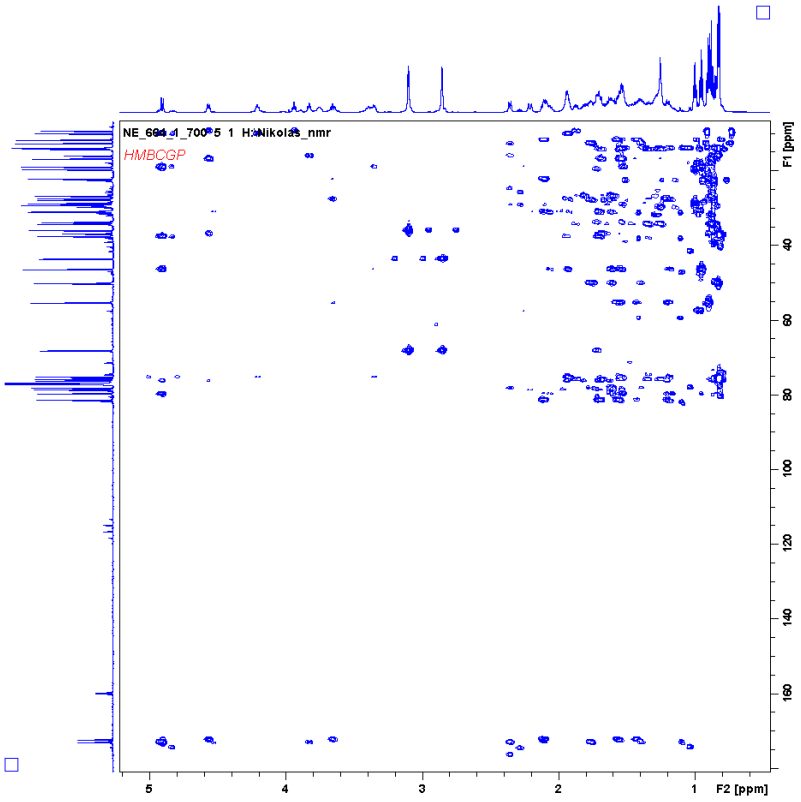

Figure SY6: HMBC spectrum (CDCl3) of Pamamycin-663 A

Table SZ: NMR data (700 MHz, $CDCl_3$) of Homopamamycin-677 A

	δ_C m	δ_H m (J in Hz)
1	172.30 C	----------
2	55.28 CH	2.11 m
3	81.34 CH	3.65 td (11.0, 4.0)
4	30.99 CH_2	1.20 m and 1.94 m
5	27.51 CH_2	1.70 m and 1.94 m
6	76.15 CH	4.22 m
7	37.46 CH	1.93 m
8	75.15 CH	4.92 d (11.0)
9	46.34 CH	1.53 m
10	79.65 CH	3.34 m
11	29.14 CH_2	1.61 m and 2.07 m
12	30.94 CH_2	1.53 m and 2.09 m
13	78.59 CH	3.76 m
14	33.71 CH_2	1.71 m and 2.21 m
15	68.05 CH	3.38 br t (11.0)
16	28.67 CH_2	1.46 m and 1.72 m
17	19.61 CH_2	1.38 m and 1.50 m
18	13.81 CH_3	0.99 t (7.0, 3H)
1´	173.04 C	----------
2´	50.07 CH_2	2.35 dt (11.5, 2.5)
3´	78.12 CH	3.82 td (6.5, 2.5)
4´	27.83 CH_2	1.61 m and 1.81 m
5´	26.66 CH_2	1.65 m and 1.77 m
6´	76.22 CH	3.91 td (7.6, 2.8)
7´	36.54 CH	1.89 m
8´	75.84 CH	4.59 dt (11.0, 4.0)
9´	34.12 CH_2	1.50 m and 1.79 m
10´	25.34 CH_2	1.25 m and 1.34 m
11´	22.89 CH_2	1.28 m (2H)
12´	14.01 CH_3	0.87 t (7.5, 3H)
2-Et	22.25 CH_2	1.43 m and 1.57 m
	11.77 CH_3	0.90 t (7.5, 3H)
7-Me	10.03 CH_3	0.82 d (7.0, 3H)
9-Et	18.98 CH_2	1.18 m and 1.54 m
	13.91 CH_3	0.95 t (7.5, 3H)
15-N(Me)$_2$	35.87 CH_{3a}	2.84 m
	43.50 CH_{3b}	3.09 m
2´-Et	15.99 CH_2	1.39 m and 1.77 m
	12.65 CH_3	0.83 t
7´-Me	9.27 CH_3	0.82 m
NH	----------	8.74 br s

Table SZ: HMBC Key correlations for Homopamamycin-677 A

Position	HMBC correlations
H-2	C-1, C-3, C-2-Et
H-3	C-1
H-6	C-7-Me
H-8	C-7, C-9, C-10, C-1´, C-7-Me, C-9-Et
H-10	C-8, C-9, C-9-Et
H-13	C-15
H-18	C-16, C-17
H-3´	C-1´, C-2´-Et
H-6´	C-7´-Me
H-8´	C-1, C-6´, C-7´, C-7´-Me
CH_3-2 (from Et)	C-2
CH_3-7	C-7
CH_3-9 (from Et)	C-9
CH_3-2´ (from Et)	C-2´
CH_3-7´	C-7´
CH_{3a}-N	C-15, 15-N-CH_{3b}
CH_{3b}-N	C-15, 15-N-CH_{3a}

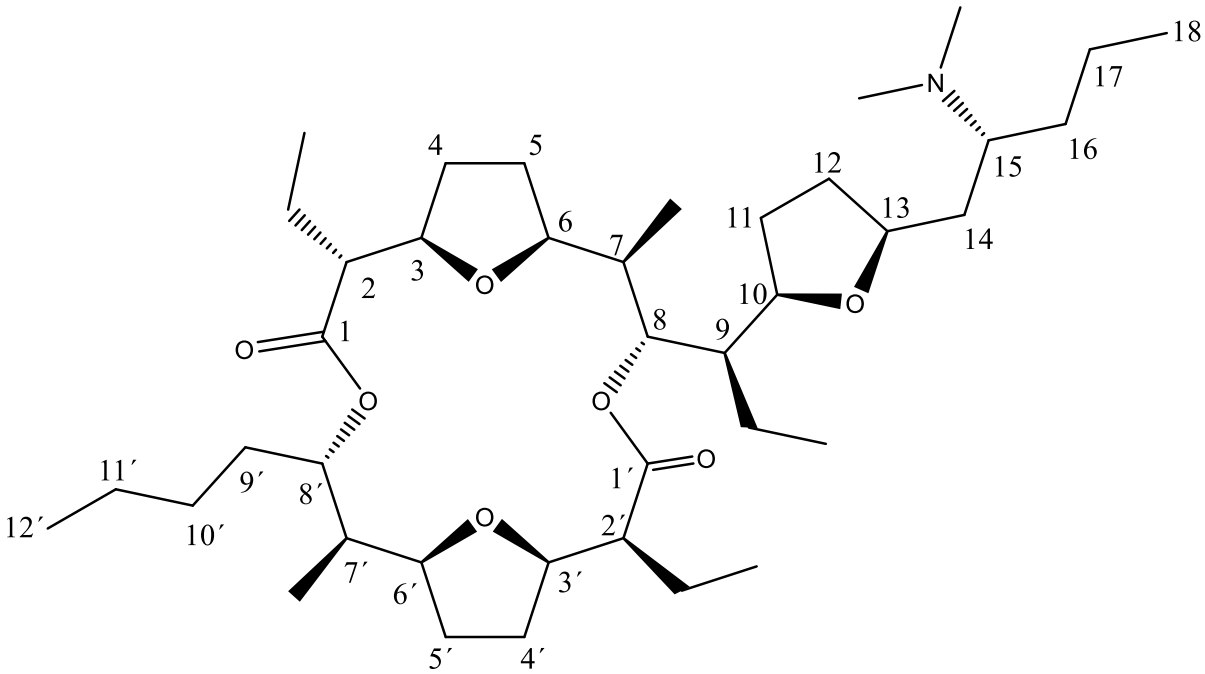

Figure SZ: Structure of Homopamamycin-677 A

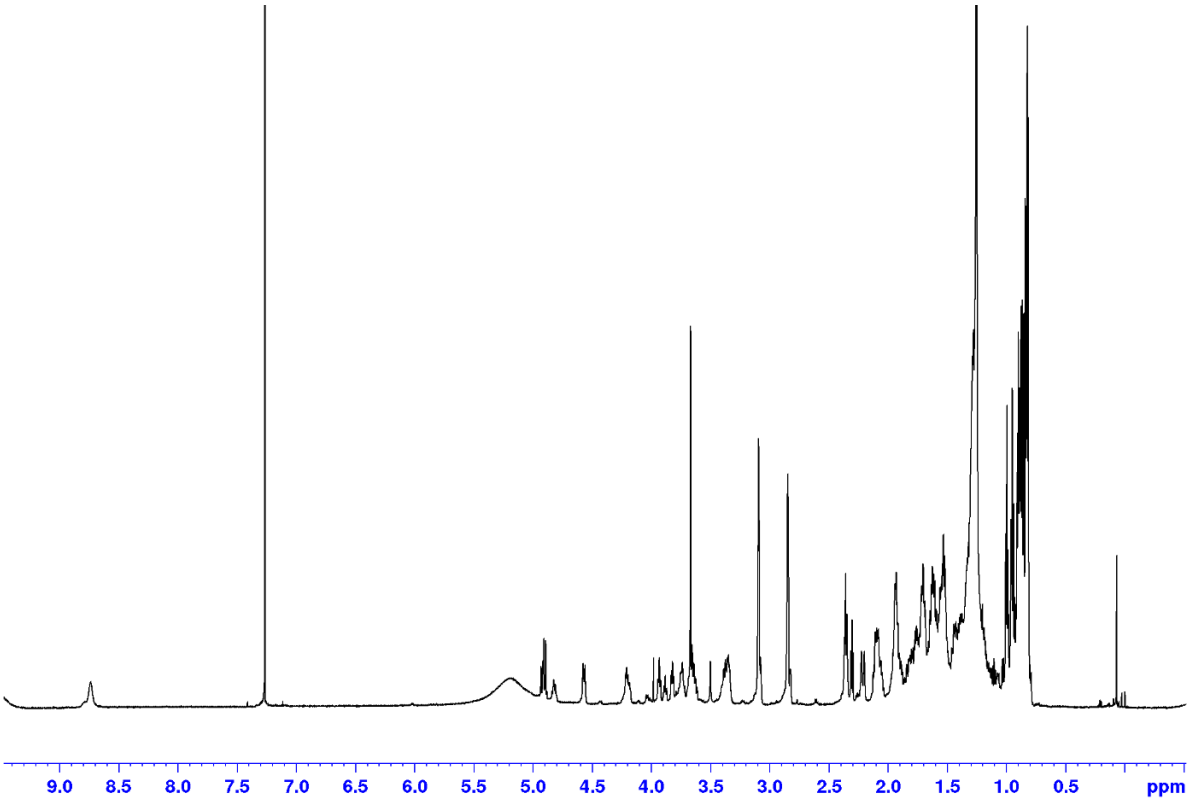

Figure SZ1a: ^{1}H NMR spectrum (700 MHz, $CDCl_3$) of Homopamamycin-677 A

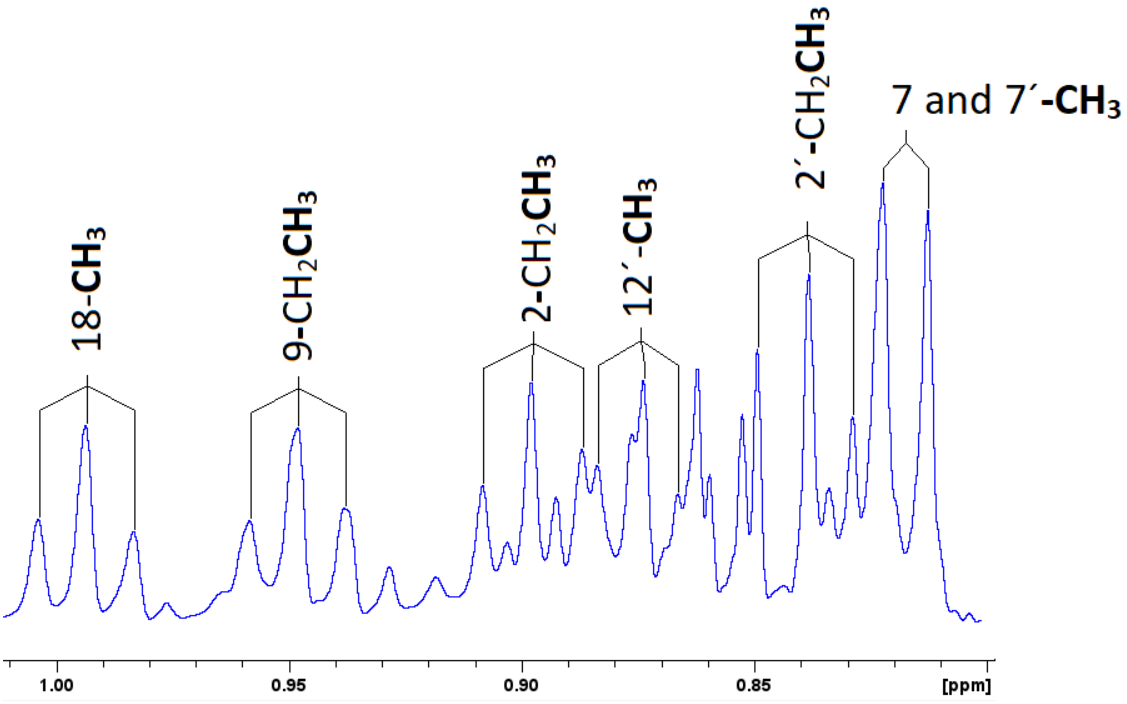

Figure SZ1b: ^{1}H NMR spectrum (700 MHz, $CDCl_3$) of Homopamamycin-677 A, Methyl region

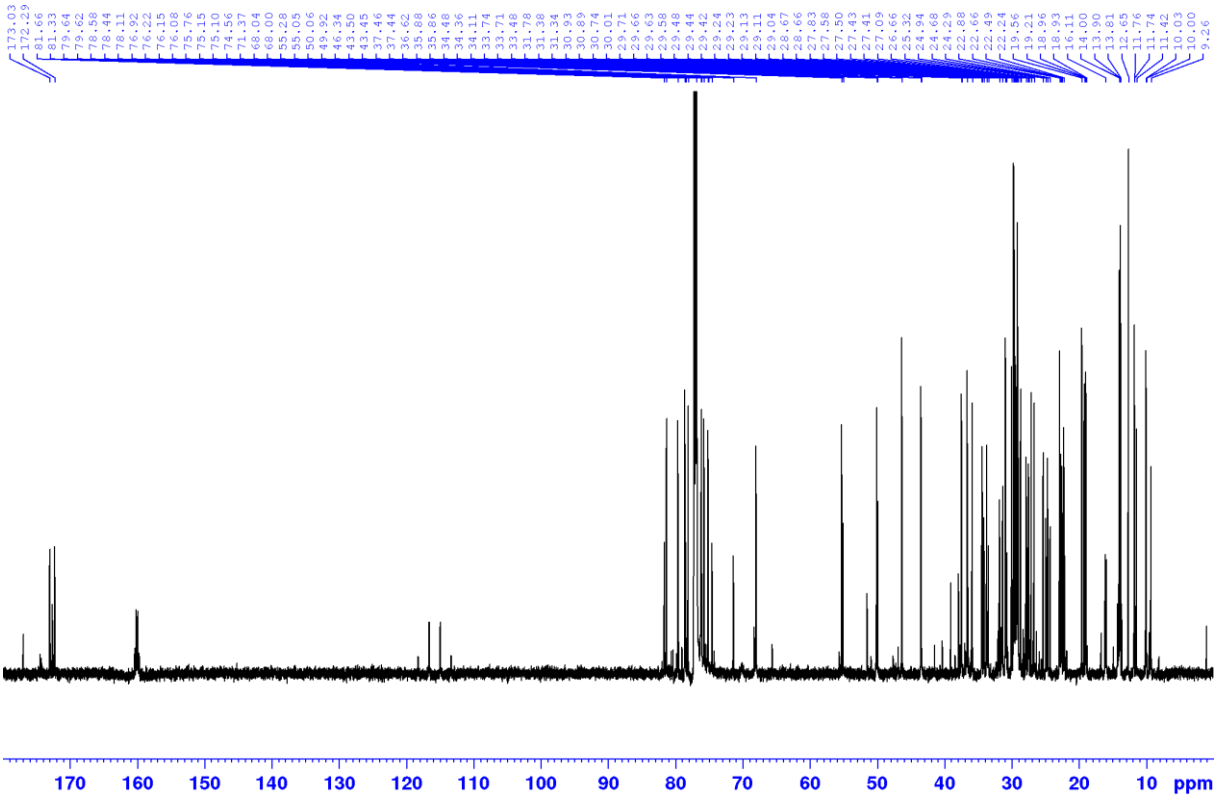

Figure SZ2: ^{13}C NMR spectrum (175 MHz, $CDCl_3$) of Homopamamycin-677 A

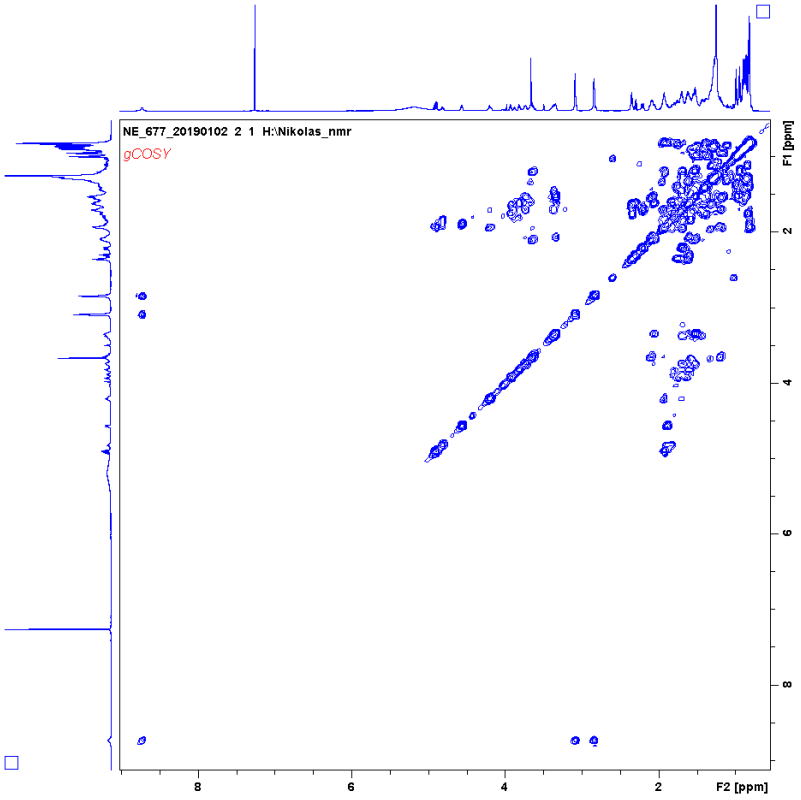

Figure SZ3: ^{1}H-^{1}HCOSY spectrum (CDCl3) of Homopamamycin-677 A

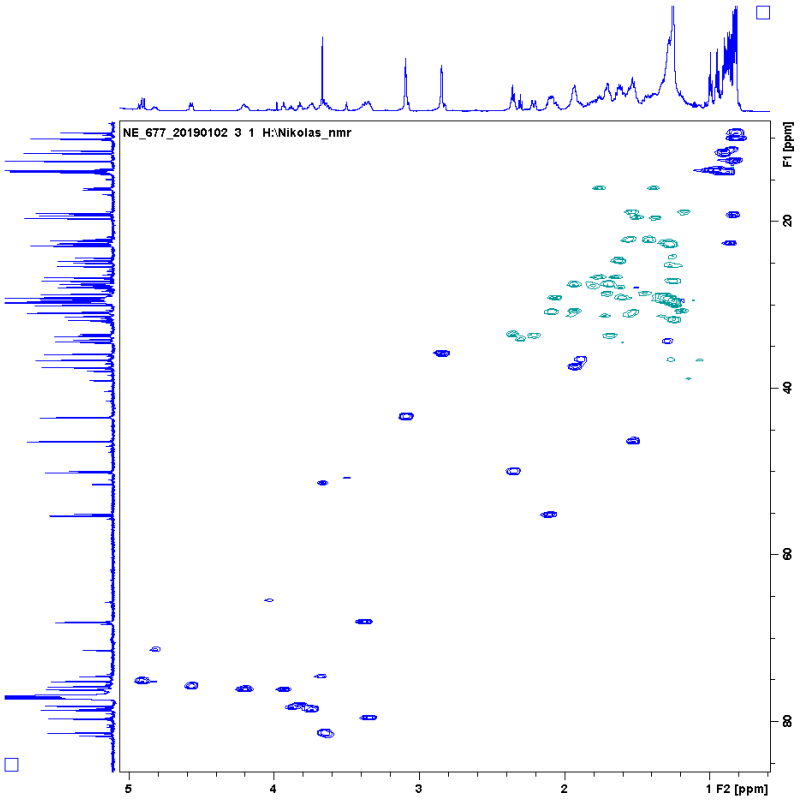

Figure SZ4: Edited HSQC spectrum (CDCl3) of Homopamamycin-677 A

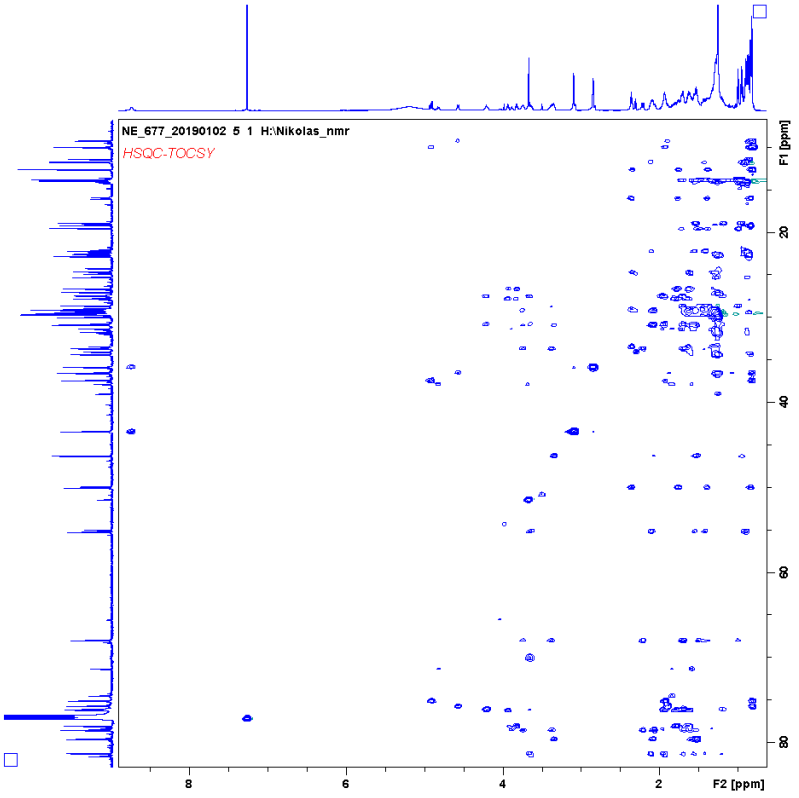

Figure SZ5: HSQC-TOCSY spectrum (CDCl3) of Homopamamycin-677 A

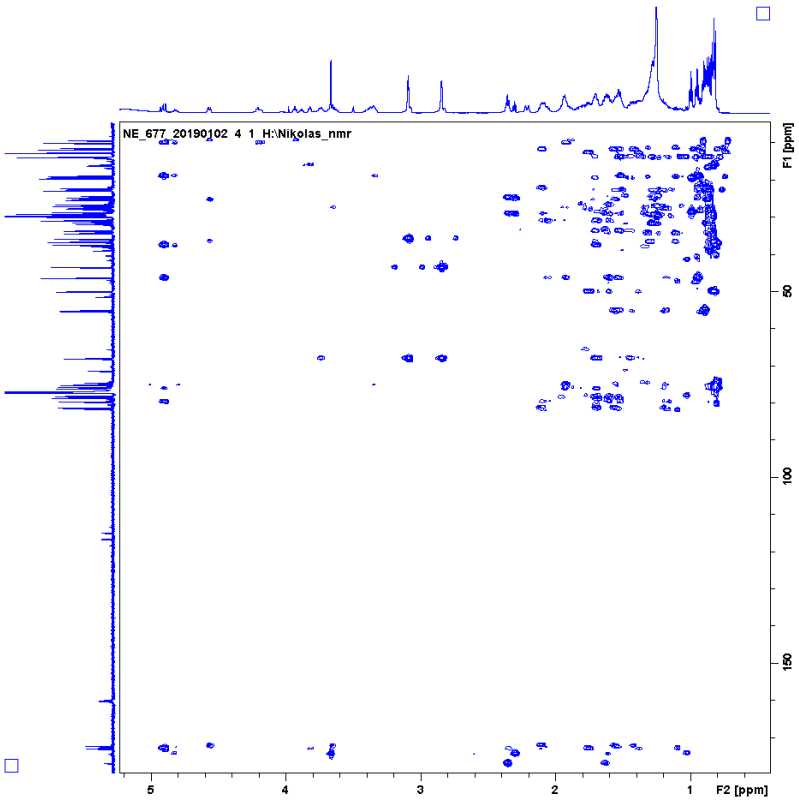

Figure SZ6: HMBC spectrum (CDCl3) of Homopamamycin-677 A

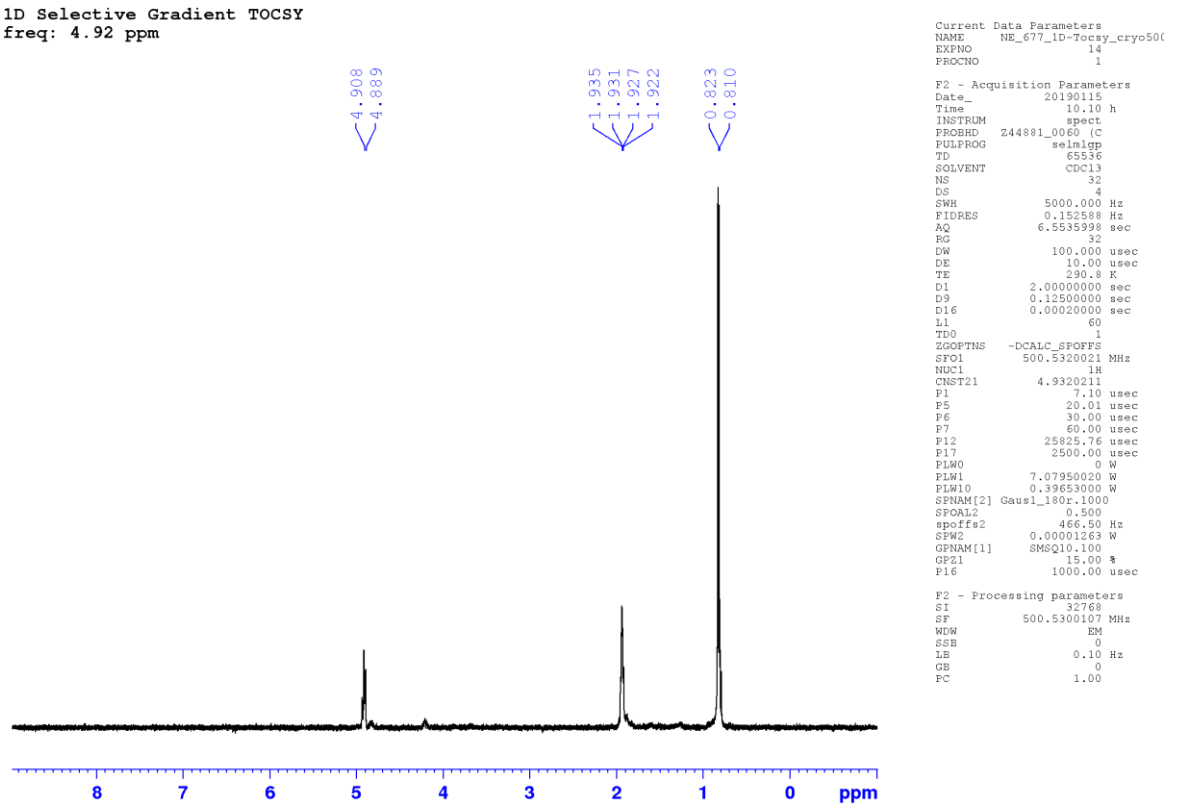

Figure SZ7a: Sel. 1D TOCSY of Homopamamycin-677 A, excitation at 4.92 ppm

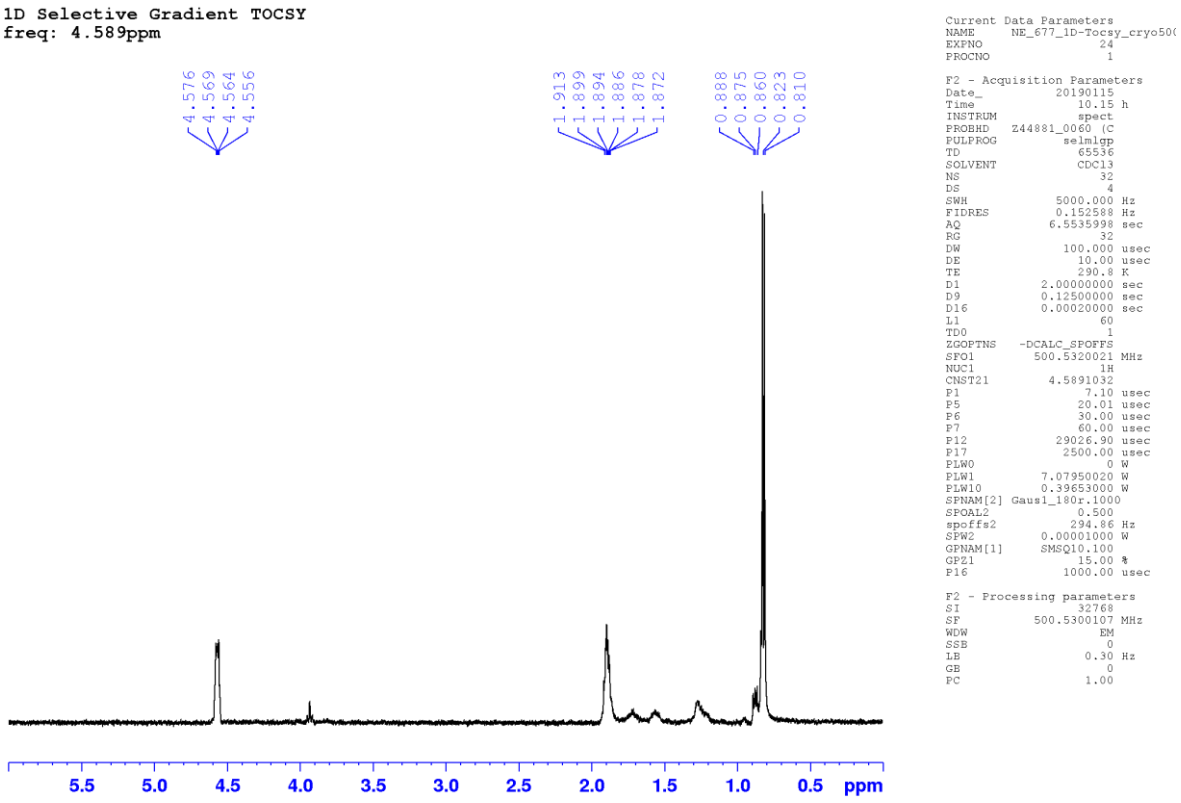

Figure SZ7b: Sel. 1D TOCSY of Homopamamycin-677 A, excitation at 4.59 ppm

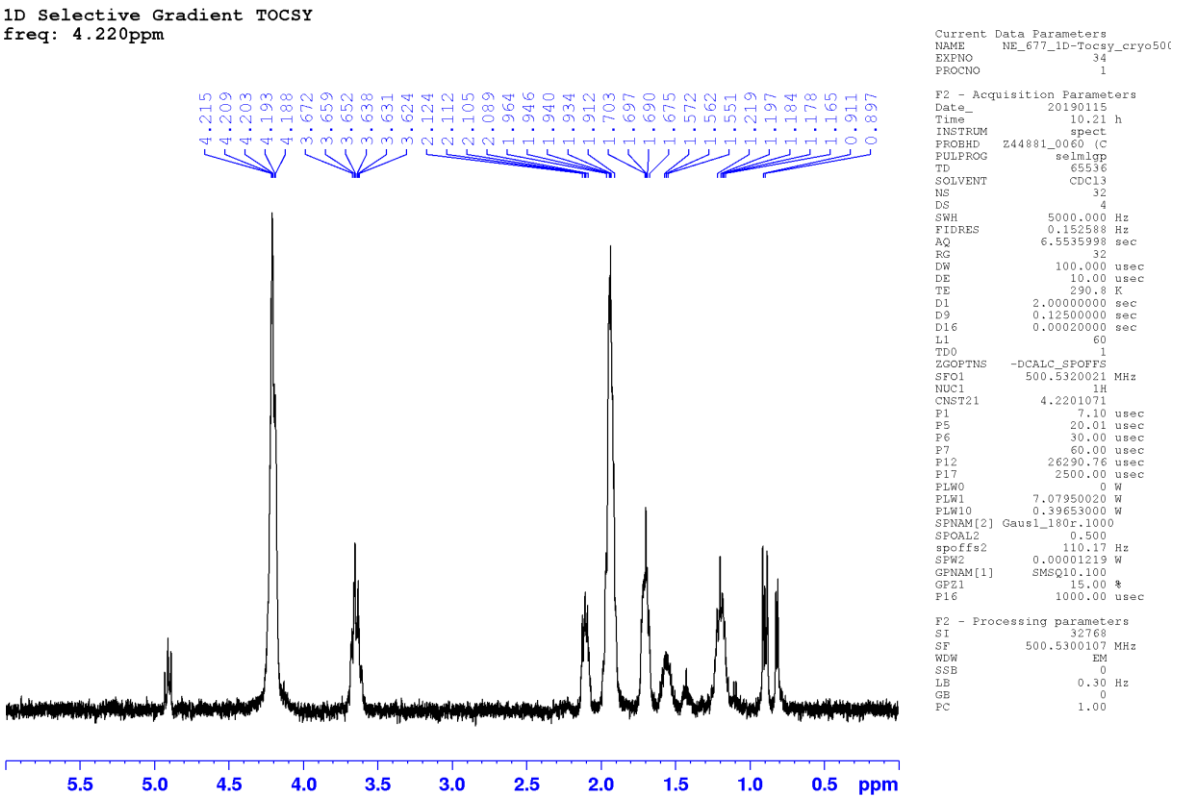

Figure SZ7c: Sel. 1D TOCSY of Homopamamycin-677 A, excitation at 4.22 ppm

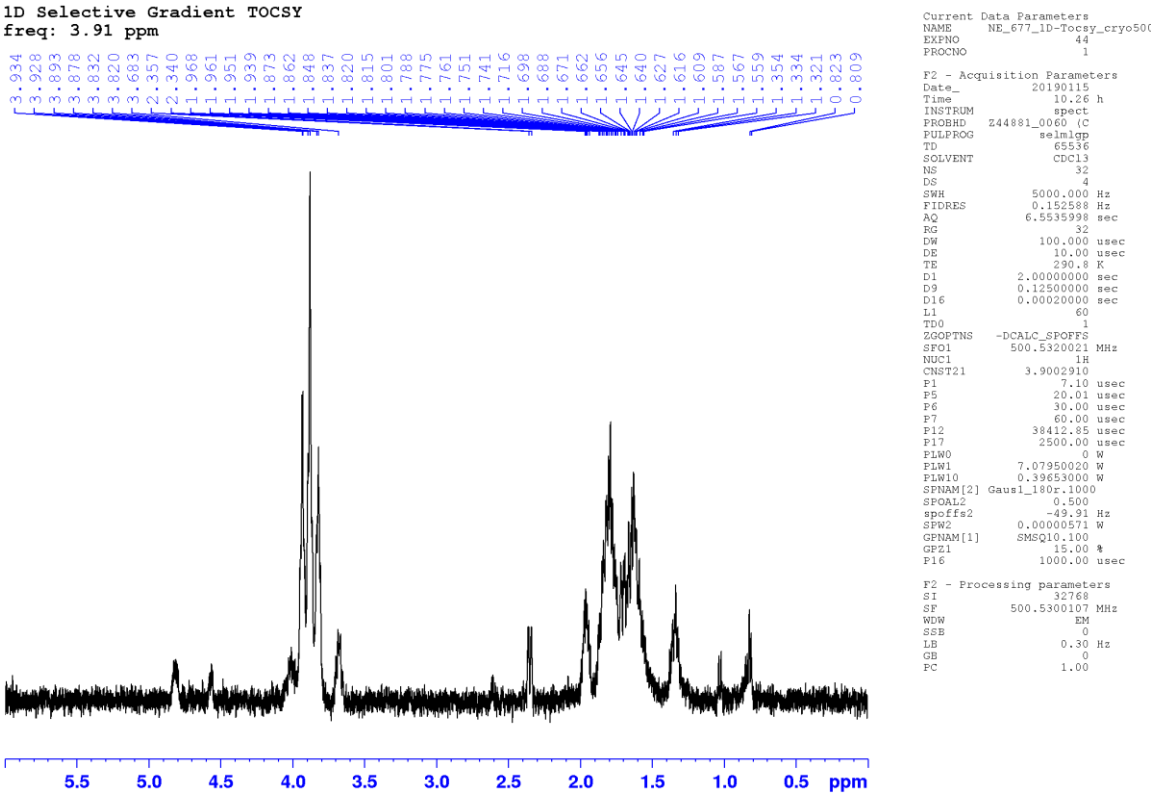

Figure SZ7d: Sel. 1D TOCSY of Homopamamycin-677 A, excitation at 3.91 ppm

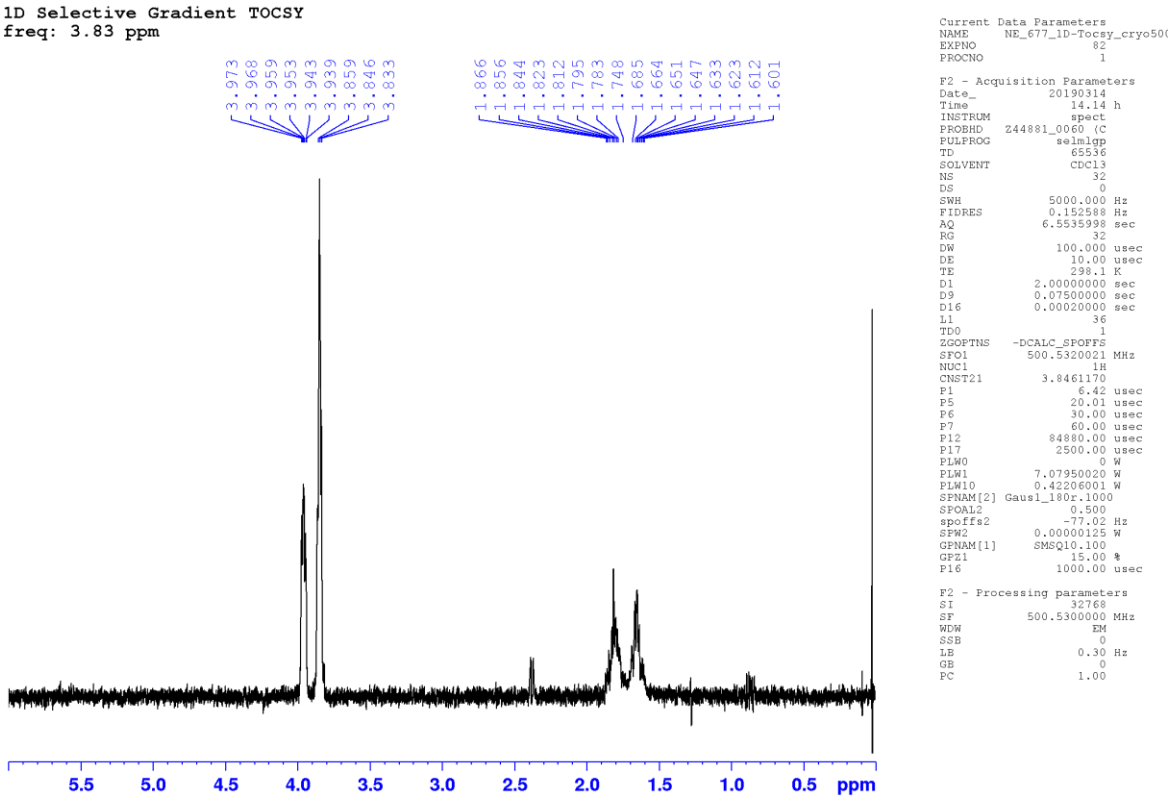

Figure SZ7e: Sel. 1D TOCSY of Homopamamycin-677 A, excitation at 3.83 ppm

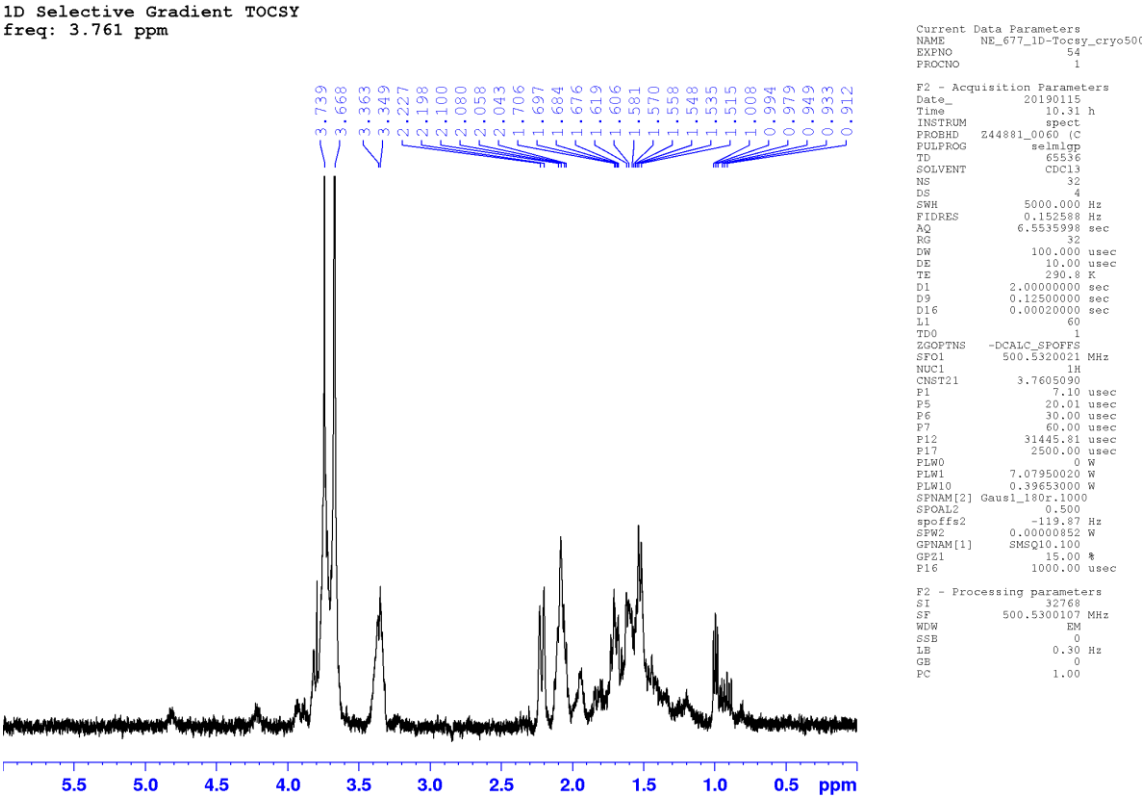

Figure SZ7f: Sel. 1D TOCSY of Homopamamycin-677 A, excitation at 3.76 ppm

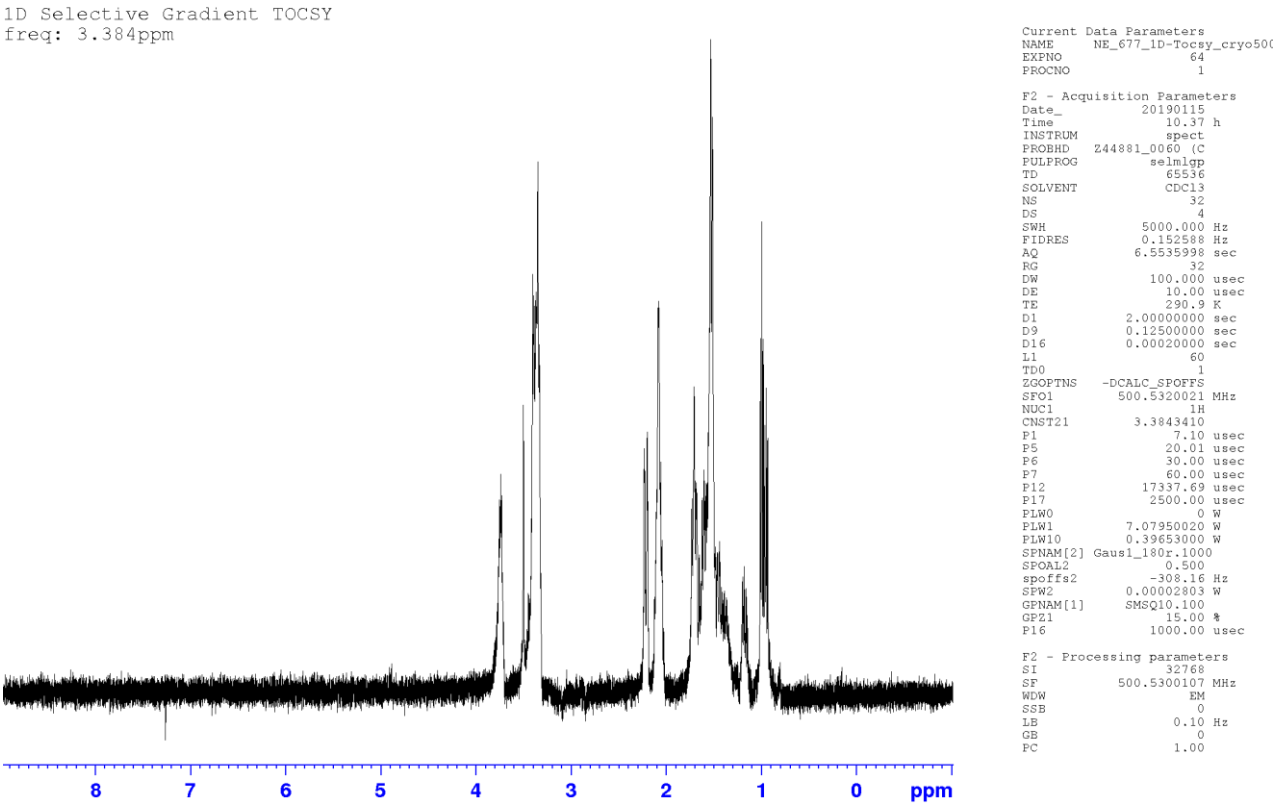

Figure SZ7g: Sel. 1D TOCSY of Homopamamycin-677 A, excitation at 3.38 ppm

References

1. Flett, F., V. Mersinias, and C.P. Smith, *High efficiency intergeneric conjugal transfer of plasmid DNA from Escherichia coli to methyl DNA-restricting streptomycetes.* FEMS microbiology letters, 1997. **155**(2): p. 223-229.
2. Fu, J., et al., *Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting.* Nature biotechnology, 2012. **30**(5): p. 440-446.
3. Fu, J., et al., *Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition.* Nucleic acids research, 2008. **36**(17): p. e113-e113.
4. Chater, K.F. and L.C. Wilde, *Streptomyces albus G mutants defective in the SalGI restriction-modification system.* Microbiology, 1980. **116**(2): p. 323-334.
5. Myronovskyi, M., et al., *Generation of a cluster-free Streptomyces albus chassis strains for improved heterologous expression of secondary metabolite clusters.* Metabolic engineering, 2018. **49**: p. 316-324.
6. Siegl, T., et al., *Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes.* Metabolic engineering, 2013. **19**.
7. Herrmann, S., et al., *Site-specific recombination strategies for engineering actinomycete genomes.* Applied and environmental microbiology, 2012. **78**(6): p. 1804-1812.
8. Rebets, Y., et al., *Insights into the Pamamycin Biosynthesis.* Angewandte Chemie International Edition, 2014. **54**.
9. Myronovskyi, M., et al., *β-Glucuronidase as a sensitive and versatile reporter in actinomycetes.* Applied and environmental microbiology, 2011. **77**(15): p. 5370-5383.

3 Summary and conclusion

In times of rising and higher population densities infective diseases are spreading faster and easier. Consequently the use of medical substances is rising, not only for human use but as well for the increasing demand of agriculture and livestock supplies. This increased use of e.g. antibiotics equals an increased selection pressure which as a result leads to resistant strains that evolve faster than before. A plethora of different antibiotics with different targets is available, seemingly making the problem of resistant strains negligible. However, the real problem is when one strain evolves or gains resistances via horizontal gene transfer to several antibiotics. Strains like this exist already like the methicillin resistant *Staphylococcus aureus* (MRSA). Infections with such multidrug resistant strains prove already quite difficult to treat and pose an ever growing threat to public healthcare, with about 10 million deaths annually expected up until 2050 [1]. Therefore, new antibiotics are needed with desirably new modes of action.

One solution to the ever growing resistance and multidrug resistant bacterial strains is the systematic screening of strain libraries for new active NPs. At first a systematic screening of microorganisms led to the discovery of an abundance of new NPs. After a relatively short time of about 30 years a high rediscovery rate of already known compounds initialised the end of the systematic screening of strain libraries for new NPs, at least for big companies [2]. This approach is by all means not exhausted, though the probability to find new NPs which are active and usable as a commercial product is extremely low. Therefore, this method is mostly used by research facilities, but still leads to the discovery of new interesting scaffolds as shown in Takahashi et al. 2003 and in this thesis as well [3]. To satisfy the need for new active compounds huge libraries of chemical synthesised compounds were screened but the discovery rate of active compounds compared to the discovery rate of active and usable compounds obtained from living organisms was significantly lower [2]. A different approach to obtain new NPs was needed. Thanks to the human genome project the sequencing technologies developed even faster and cheap sequencing of the first sequenced *Streptomyces* genome indicated the presence of more than 20 BGCs whereas each strain was believed to be able to produce only a handful of compounds [4]. This shows the still hidden potential of the cast aside genus and gives opportunities to find new interesting NPs. With the growing knowledge about the biosynthesis of NPs and the construction of databases thereof, the search for new natural products was not anymore focused on the screening of new strains but rather

on the identification of potentially interesting clusters and how to activate them. Based on this discovery already isolated strains, deemed to be not interesting, were thereby proven to be hiding possibly interesting NPs within their genomes. With this approach the search for new NPs with potential applications gained again the focus for the search of unique potential therapeutics. With continuously improving technologies the immense potential of not cultivatable strains can be tapped into, by isolation and cloning of environmental DNA and systematic screening for BGCs as well as activation thereof. For that purpose even more reliable heterologous hosts need to be constructed, to have a higher probability of success, like the optimised strain *S. albus* J1074 Del14 [5].

Within this thesis two problems are addressed: **1. Discovery of new compounds**, which is basically a problem of regulation and the resulting low production. **2. Overproduction** of known but low yielding compounds and derivatives thereof. Firstly, based on genome mining, the NP cyclofaulknamycin was discovered in the well described model organism *Streptomyces albus* and the cyclisation step of the biosynthesis was further characterised. Cyclofaulknamycin contains the interesting non-proteinogenic amino acid D-capreomycidine. However, the needed gene of its synthesis is not present in the cluster. Further screening of the genome revealed a similar gene about 400,000 bp downstream of the cluster. Knockout experiments of this gene, *XNR_1347*, resulted in a significantly reduced cyclofaulknamycin production. This proves that the gene responsible for the biosynthesis is located outside of the cluster and is showing the importance of throughout analysis of the cluster of interest and genes in outside closer proximity. Essential biosynthetic genes could be located outside the cluster in other cases as well, preventing their heterologous expression. With the help of closer analysis and improved bioinformatic software the genes could be identified and BGCs which were previously silent, activated.

Secondly, an already isolated and characterised strain was cultivated in a different medium which resulted in the activation of a previously silent cluster. The compound was purified and its unique structure determined. Via genome mining a potential cluster was identified and validated by knockout experiments. Afterwards, the biosynthesis was further analysed with the help of heterologous expression and previously developed molecular biological tools. The data show that even characterised strains hide the potential for interesting new scaffolds and further research needs to be attempted to find methods on how to activate them more reliably. One possibility is the heterologous expression which worked in this case flawless, even the

modification of the polyketide backbone with a D-forosamine moiety. Therefore, more heterologous hosts would be desirable, to have a higher chance to obtain products from a cluster.

Lastly, experiments were performed to increase the production of pamamycin NPs. Commonly the production of NPs is not only tightly regulated but the genes are balanced to each other as well. Unbalanced expression of genes can thereby lead to a reduced or increased production. Only higher transcription of simply one gene which is believed to be the bottleneck can cause a hindrance at another position of the reaction cascade. Therefore, a previously random rational approach for promoter optimisation was used to balance the two operons of the pamamycin cluster towards a higher yield of pamamycin derivatives with higher molecular weight [6]. Since checking the effect of a promoter set one by one would be labour intense and time consuming the method is designed in such a way that a library of a resistance gene flanked by two degenerated promoters was screened for increased production. The needed strength for the upstream as well as downstream promoter of the cluster for an increased production of derivatives cannot be determined in advance yet, as the effect for each cluster is different. Out of 106 created clusters, two constructs showed an increased production of pamamycins above a molecular weight of 649 Da. As a result of the increased production, several milligrams of intermediates which were not further characterised before could be purified and for the first time activity tests were performed. In comparison to the pamamycins pam-607 the isolated pamamycins (pam-663) showed a 24 times higher activity against the carcinoma cell line KB-3.1. Closer analysis of the promoter strengths revealed that the upstream promoters of the clusters with increased production showed a 30 times higher activity compared to a barely measurable activity of the promoter of the native cluster. The promoters of the downstream operon could vary about 50 % in their strength and still produce a similar shifted production profile. Thereby, it could be deduced that overexpression of the small upstream operon containing the genes *pamC* (acyl-carrier-protein), *pamF* and *pamG* (ketosynthases) leads to a shifted production spectrum. With the help of promoter engineering, new intermediates were identified and tested for their activities. The data proof, that already discovered but due to too low yields not completely characterised NPs and their derivatives are a potent source of NPs and therefore a promising target for further studies and will potentially help in the search for new active usable NPs.

Within the research conducted in the course of this thesis one possible solution to target low production yields of already known NPs was undertaken, via the random rational promoter optimisation. New compounds and intermediates were discovered by genome mining and the help of previously constructed methods and tools. Based on those results it can be seen that there is still a slumbering potential for NPs in *Streptomyces* which is only waiting to be discovered. Often organisms are living in a symbiotic relationship with each other as for example *Streptomyces* sp. IB2014/011-12 isolated from *Trichoptera* sp. larvae. In general those organisms share a mutual benefit, which could be a better substrate uptake or even resistance to other organisms. Especially symbiotic organisms are difficult to cultivate but also have an immense potential for NPs and thereby cultivation of those strains or expression of their clusters could lead to new scaffolds [7]. However, as for other upcoming problems like the reduction of CO_2 emission there is not only one solution to the problems. An awareness of proper usage of e.g.: antibiotics, insecticides, herbicides needs to be risen, as well as other solutions need to be investigated. One interesting branch of treatment methods for bacterial infections, are antimicrobial peptides which could find potential applications [8]. Previously in centre of the research, but nowadays neglected, was the treatment of bacterial infections with bacteriophages which could prove to be an interesting addition to the treatment with antibiotics [9]. Therefore, broadening the field of research could be a good addition to the current lack of new antibiotics in addition to the above mentioned possibilities to screen for new NPs and activation of their biosynthesis.

References

1. de Kraker, M.E., A.J. Stewardson, and S. Harbarth, *Will 10 million people die a year due to antimicrobial resistance by 2050?* PLoS medicine, 2016. **13**(11): p. e1002184.
2. Hutchings, M.I., A.W. Truman, and B. Wilkinson, *Antibiotics: past, present and future.* Current opinion in microbiology, 2019. **51**: p. 72-80.
3. Takahashi, Y. and S. Omura, *Isolation of new actinomycete strains for the screening of new bioactive compounds.* The Journal of General and Applied Microbiology, 2003. **49**(3): p. 141-154.
4. Chung, Y.-H., et al., *Comparative Genomics Reveals a Remarkable Biosynthetic Potential of the Streptomyces Phylogenetic Lineage Associated with Rugose-Ornamented Spores.* Msystems, 2021. **6**(4): p. e00489-21.
5. Myronovskyi, M., et al., *Generation of a cluster-free Streptomyces albus chassis strains for improved heterologous expression of secondary metabolite clusters.* Metabolic engineering, 2018. **49**: p. 316-324.
6. Horbal, L., et al., *Secondary metabolites overproduction through transcriptional gene cluster refactoring.* Metabolic engineering, 2018. **49**: p. 299-315.
7. Adnani, N., S.R. Rajski, and T.S. Bugni, *Symbiosis-inspired approaches to antibiotic discovery.* Natural product reports, 2017. **34**(7): p. 784-814.
8. Lei, J., et al., *The antimicrobial peptides and their potential clinical applications.* American journal of translational research, 2019. **11**(7): p. 3919.
9. Principi, N., E. Silvestri, and S. Esposito, *Advantages and limitations of bacteriophages for the treatment of bacterial infections.* Frontiers in pharmacology, 2019. **10**: p. 513.