# Genome mining in myxobacteria and actinobacteria: Studies of natural products and their biosynthetic pathways

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

von
**Maja Remškar**
Saarbrücken
2020

| | |
|---|---|
| **Tag des Kolloquiums:** | 09.10.2020 |
| **Dekan:** | Prof. Dr. Guido Kickelbick |
| **Berichterstatter:** | Prof. Dr. Rolf Müller<br>Prof. Dr. Andriy Luzhetskyy |
| **Vorsitz:** | Prof. Dr. Alexander Titz |
| **Akad. Mitarbeiter:** | Dr.-Ing. Michael Kohlstedt |

Die vorliegende Arbeit wurde von Juni 2016 bis Mai 2020 unter Anleitung von Herrn Prof. Dr. Rolf Müller am Helmholtz-Institut für Pharmazeutische Forschung Saarland (HIPS) angefertigt.

# Acknowledgments

First, I would like to thank Prof. Dr. Rolf Müller for offering me this PhD position in his group. I am grateful for every entrusted project, his supervision during this time and for his professional advice, especially when correcting my thesis.

I would like to thank Prof. Dr. Hrvoje Petković for giving me an opportunity for doing this PhD abroad, which proved to be an invaluable experience, and for taking time to correct part of the thesis. Thank you to Dr. Chengzhang Fu and Dr. Daniel Krug for reading the thesis and their helpful comments that made it better.

I would like to thank Prof. Dr. Andriy Luzhetskyy for being my second supervisor and reviewing my thesis.

Thanks to Dr. Tadeja Lukežič for introducing me to some of the molecular biology techniques, Dr. Antoine Abou Fayad for teaching me the compound purification procedures, Dr. Ronald Garcia for advice in cultivation of myxobacteria and Dr. Nestor Zaburannyi for his contribution in bioinformatics.

I would especially like to thank Dr. Alexander Popoff for his collaboration in the analytical portion of our project. Without his expertise, a part of this thesis would not be possible.

I would also like to thank all of the former and current group members for collaborations, discussions, professional advice and all the time we spent together. You made my time abroad enjoyable and unforgettable.

I would like to thank my parents, my sisters and my friends for all the support during my time abroad. Last, I would like to thank my boyfriend Thomas for the support and making Germany a nice place.

# Abstract

Bacterial resistance to antibiotics drives the need for incessant research efforts to discover substitutes for ineffective antibiotics. Soil-dwelling Myxobacteria and Actinobacteria provide many biologically active natural products that can be developed into new treatment options. In postgenomic era, the natural product discovery is often based on the genome mining approach combined with genetic engineering used also in this work. This thesis presents two novel natural products, myxopentacin and myxoglucamide, produced by *Cystobacterineae* sp. MCy9003, a representative of a novel myxobacterial genus. The nonribosomal peptide myxopentacin features rare cispentacin and pseudoarginine moieties, incorporated by a unique assembly mechanism. The glycolipopeptide myxoglucamide contains an unusual vinyl moiety and its biosynthesis involves multiple biochemical transfomations. In addition, this thesis describes the identification of the biosynthetic gene cluster in *Streptomyces hagronensis*, responsible for the production of the known antibiotic globomycin, a signal peptidase II inhibitor. Furthermore, the genome of the antibiotic fidaxomicin producer *Actinoplanes deccanensis* was genetically engineered to acquire desired shunt products for further semi-synthesis. Finally, the biosynthetic pathway of the atypical tetracycline chelocardin, regarding cyclization and C6-methylation, was characterized via gene knockout and site-directed mutagenesis in *Amycolatopsis sulphurea*.

# Zusammenfassung

Aufgrund der zunehmenden Resistenz von Bakterien gegen die bekannten Antibiotika, besteht ein ständiger Bedarf an neuen Wirkstoffen. Im Boden lebende Mikroorganismen wie Myxobakterien und Actinobakterien produzieren viele biologisch aktive Naturstoffe, die zu neuen Behandlungsoptionen entwickelt werden können. Um dieser Problematik entgegen zuwirken, wurde in dieser Arbeit der "genome-mining" Ansatz in Kombination mit gentechnischer Manipulationen von *Cystobacterineae* sp. MCy9003, einem Vertreter einer neuartigen myxobakteriellen Gattung, verwendet. Das nichtribosomale Peptid Myxopentacin mit den seltenen Cispentacin- und Pseudoarginin-Einheiten, weist einen einzigartigen Assemblierungsmechanismus auf. Das Lipoglykopeptide Myxoglucamid enthält eine ungewöhnliche Vinyleinheit und die Biosynthese umfasst viele biochemische Prozesse. . Zusätzlich wurde in dieser Arbeit der Biosyntheseweg des bekannten Antibiotikums Globomycin, ein Signalpeptidase II-Inhibitor, nach Identifizierung des entsprechenden Biosynthesegenclusters in *Streptomyces hagronensis* analysiert. Darüber hinaus wurde das Genom des Fidaxomicin-Produzenten *Actinoplanes deccanensis* gentechnisch verändert, um spezielle Shunt-Produkte für weitere semisynthetische Studien zu erhalten. Des weiteren wurde der Biosyntheseweg des atypischen Tetracyclin Antibiotikums Chelocardin hinsichtlich Zyklisierung und C6-Methylierung in *Amycolatopsis sulphurea* charakterisiert.

# Vorveröffentlichungen der Dissertation

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

## Publications

Olatunji, S.; Yu, X.; Bailey, J.; Huang, C-Y.; Zapotoczna, M.; Bowen, K.; **Remškar, M**.; Müller, R.; Scanlan, EM.; Geoghegan, JA.; Olieric, V.; Caffrey, M. Structures of lipoprotein signal peptidase II from *Staphylococcus aureus* complexed with antibiotics globomycin and myxovirescin. Nature Communications, 2020; 11:140.

**Remškar, M.***; Popoff, A.*; Garcia, R.; Müller, R. Genome mining in a novel myxobacterium *Cystobacterineae* sp. MCy9003 reveals two novel natural products myxopentacin and myxoglucamide with rare structural features, to be published

Lukežič, T.*; **Remškar, M.***; Slemc, L.; Groß, S.; Bader, CD.; Šala, M.; Santos Beneit, F.; Müller, R.; Petković, H. Biosynthesis of atypical tetracyclines: Characterization of early steps in chelocardin biosynthetic pathway, to be published

Kiefer, A.; **Remškar, M.**; Rox, K.; Müller, R.; Hirsch, A.$^{+}$ The biosynthesis of globomycin and total synthesis of its derivatives with enhanced biological activities, to be published

Dorst, A.; **Remškar, M.**; Müller, R.; Gademann, K.$^{+}$ Novel semisynthetic derivatives of fidaxomicin, to be published

*these authors contributed equally; $^{+}$ a complete set of authors is not yet known

## Other publication

Hug, JJ.; Bader, CD.; **Remškar, M.**; Cirnski, K.; Müller R. Concepts and Methods to Access Novel Antibiotics from Actinomycetes. Review. Antibiotics (Basel) 2018, 7(2), 44.

## Patent

Gene cluster for the biosynthetic production of tetracycline compounds in a heterologous host Müller, R.; Lukežič, T.; **Remškar, M.**; Zaburannyi, N.; Bader, C.; Sikandar, A.; Köhnke, J. 2019. Publication number: WO2019122400

## Conference Contributions (Posters and Oral Presentations)

**M. Remškar.** Genetic Manipulation Strategies to Uncover Natural Products from Novel Myxobacteria. **Oral Presentation**, VI. Symposium of the Interdisciplinary Graduate School of Natural Products Research, 2017: Saarbrücken, Germany.

**M. Remškar**, R. Müller. Genetic Manipulation Strategies to Uncover Natural Products from Novel Myxobacteria. **Poster Presentation**, International VAAM-Workshop 2017: Tübingen, Germany.

A. Popoff, **M. Remškar**, J. Birkelbach, R. Garcia, R. Müller. Genome mining revealed a new myxobacterial glycolipopeptide containing non-proteinogenic amino acids. **Poster Presentation**, 9$^{th}$ International HIPS Symposium, 2019: Saarbrücken, Germany.

# 1 Introduction

## 1.1 Natural products

Natural products (NPs) are defined as “chemical compounds or substances produced by a living organism found in nature” [1]. Secondary metabolites, on the other hand, are “organic compounds that are not directly involved in the normal growth, development, or reproduction of an organism and unlike primary metabolites, absence of secondary metabolites does not result in immediate death, but rather in long-term impairment of the organism's survivability, fecundity, or aesthetics, or perhaps in no significant change at all” [2]. They can be found in plants, fungi, bacteria and animals to improve their chances for survival [3]. Throughout the ages, humans relied on Nature not only for their basic needs but also for medicines. Plants were the basis for the establishment of traditional medicine that is still used after thousands of years, in some less developed countries as primary health care [4]. First records of NPs as medicines, written on clay tablets in cuneiform, originate from Mesopotamia from about 2600 BCE. They were used before mainly as complex extracts, without knowing the structure of a biologically active compound or its mode of action, but only had the potential to be characterized later on with the observation of a penicillin-producing microorganism in 1928 by Sir Alexander Fleming [5]. Penicillin was the first known antibiotic produced by a living organism – *Penicillium notatum*, a species of fungus in the genus *Penicillium*. It took up until 1941 that it was produced in high enough amounts to be tested in humans. In 1942, penicillin was available for patient treatment [5]. After successful large scale production of penicillin during World War II, the pharmaceutical companies focused their programs on the discovery of new antibiotics [6]. Other than fungi, bacteria turned out to be a gold mine of natural products. Isolation of gramicidin from the soil bacterium *Bacillus brevis* in 1939 launched further discoveries of actinomycin (1940), streptothricin (1942) and first aminoglycoside streptomycin (1943) [7] from *Streptomyces* species [8]. Additional discoveries of gentamicin, tetracycline and other antibiotics motivated the industry for research and development programs for NP discovery [6]. Streptomycin underwent clinical trials, was approved as anti-tuberculosis agent in 1948 [9] and is the first FDA-approved NP of bacterial origin [10]. In next 40 years, more than 20 structural classes of new antimicrobial agents were discovered [11] (Figure 1). Chemical modifications of NPs are sometimes necessary to achieve desirable properties. For instance, the precursor-feeding during the fermentation of a penicillin-producing strain yielded an acid-stable derivative, penicillin V, that could be administered orally unlike intravenously administered penicillin G [7, 12].

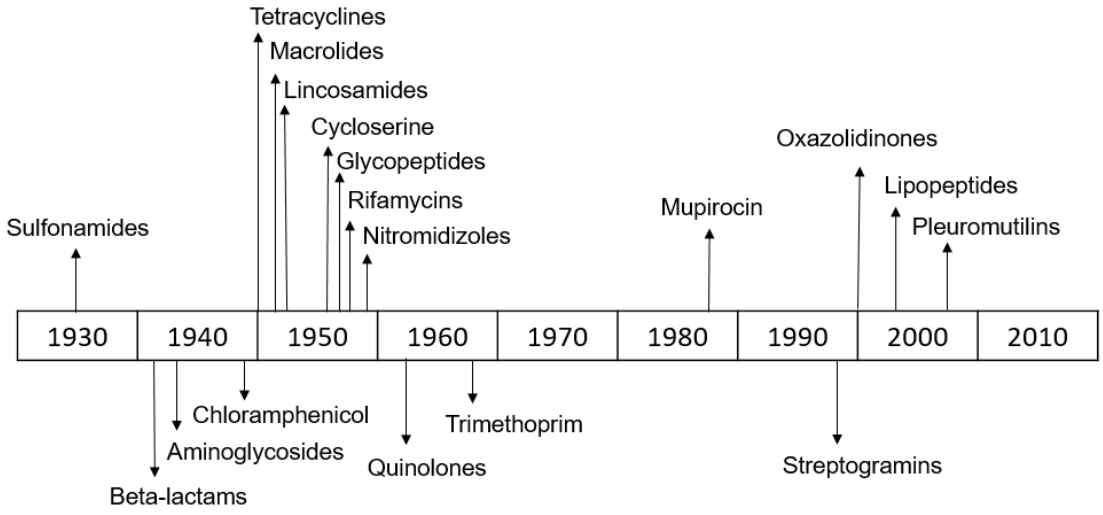



**Figure 1. Structural families of NPs with antibacterial properties introduced into clinical trials over time.** The figure was taken from the publication by Song in 2012 [13].

At first very successful NP discovery became hindered by the rediscovery of known compounds [14]. The classical NP discovery was as well labor-intensive and developments such as automated high throughput screening (HTS) programs and combinatorial chemistry approach promised improvements and new chances for NP discovery. However, NP research programs could not keep up with demands for the number of brand new natural products, so large pharmaceutical companies decided to close their programs in the 1990s and early 2000s [6]. By the end of 2013, over one-fifth (21 %) of all FDA-approved new molecular entities (NMEs) represented not only unmodified NPs (7.5 %) but also semisynthetic derivatives (9.6 %) and synthetic structures derived from a NP (4 %) [10]. The current data show that 53 % of all NMEs are NPs isolated from plants, bacteria, fungi, marine organisms, non-mammalian species [10]. However, many compounds isolated from marine organisms, such as sponges, are actually produced by symbiotic bacteria [15, 16]. Recently, the insect microbiome was shown to hold promise for discovery of novel antimicrobial NPs [17]. Up to 2016, out of all antibacterial agents 69 % originated from NPs, with 97 % isolated or derived from microbes [10]. However, usage of bacterial products to fight against them forces bacteria in turn to find new ways of survival. Bacteria as single-celled microorganisms mutate fast and somehow like in an enforced evolution process can take up resistance genes through horizontal gene transfer (HGT) from the environment [18]. For that reason, we are battling with antibiotic resistance that was first reported in 1948 during clinical trials for streptomycin-resistant *Mycobacterium tuberculosis* [19]. The search of new NPs is therefore focused on discovery of anti-infectives of new structural classes [20]. In a best-case scenario these would be compounds that bacteria cannot develop resistance against. In addition, new medicines may be derived from “rediscovered

compounds", i.e. NPs found in past but never followed up because the antibiotic with the similar biological activities were already on the market. However, the target microorganisms causing infective diseases, not responding anymore to the available treatments due to antimicrobial resistance (AMR), are monitored from May 2015 on by the new Global Antimicrobial Surveillance System (GLASS) that implemented standardized AMR surveillance globally [21, 22]. The survey prioritizes *Salmonella* spp., *Staphylococcus aureus, Neisseria gonorrhoeae* for research and development of new and effective antibiotic treatment, while *Acinetobacter baumanii*, *Escherichia coli* and *Klebsiella pneumoniae* present critical, and *Shigella* spp. as well as *Streptococcus pneumoniae* medium priority [21, 22]. Even after many decades of exploitation of bacteria as a source of NPs, novel bacterial genera are continuously being isolated, offering yet unexplored possibilities for discovery of NPs [23]. The important NP sources, actinobacteria and myxobacteria, are described in more detail below.

### 1.2 Actinobacteria as producers of natural products

The class of actinobacteria consists of Gram-positive, aerobic filamentous bacteria with high content of GC base pairs observed in their genomes. They are found in alkaline to neutral soils [24], fresh water or marine environment [25], depending on the amount of oxygen in the given environment. They all produce the volatile secondary metabolite geosmin, giving soil its characteristic odor. About 45 % of all bioactive microbial metabolites discovered originate from Actinobacteria [25]. The most abundantly represented is the genus *Streptomyces*, belonging to the order *Streptomycetales* [26]**.** Over 500 *Streptomyces* species are characterized to date since they proved to be of industrial importance as producers of broad spectrum of secondary metabolites [24]. Structural classes are divided into aminoglycosides (e.g. kanamycin), peptides (e.g. actinomycin D), glycolipids (e.g. bleomycin), β-lactams (e.g. clavulanic acid), nucleosides (e.g. nikkomycin) and polyketides divided further into macrolides (e.g. rapamycin), tetracyclines (oxytetracycline, chelocardin), polyethers (e.g. monensin), polyene macrolides (e.g. nystatin), anthracyclines (e.g. doxorubicin), just to name a few [27]. Streptomycetes provide not only antibacterial but also antifungal, antiparasitic, anticancer, antiviral, immunosuppressive drugs [27]. Examples are presented in Figure 2. Interactions of Streptomyces with other bacteria are poorly understood. As soil, their main habitat, is densely populated by other bacteria, fungi, nematodes [28], there is lots of competition for nutrients. The production of secondary metabolites harmful for competitors gives Streptomycetes an advantage to successfully finish their development cycle with drought-resistant spores [24].

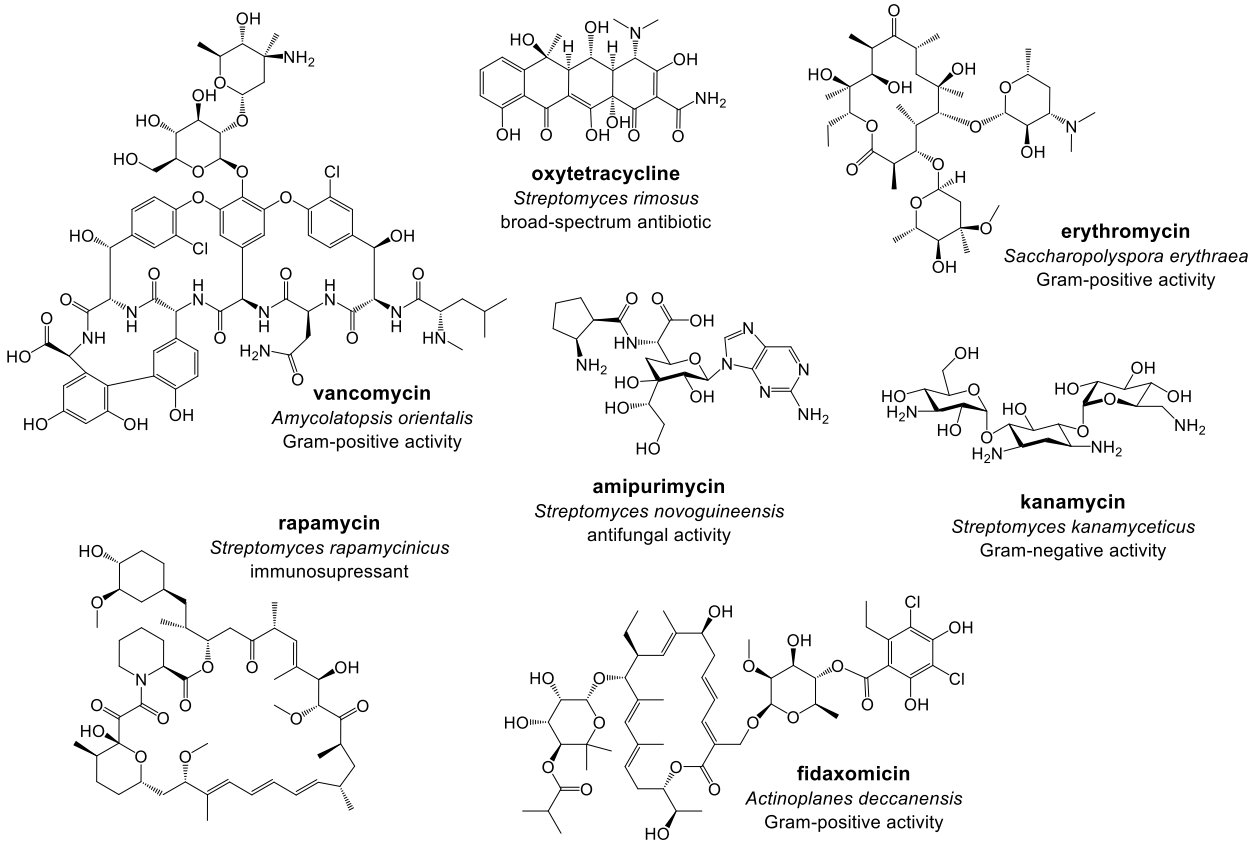



**Figure 2. Selected actinobacterial NPs**

Bacteria are classified into genera based on phenotypic (morphology, physiology, biochemistry) and genotypic traits (e.g. sequence analysis for 16S rRNA) [27]. Three well-classified actinobacterial NP-producing strains (Table 1), each representing a different order, are used in the scope of this thesis. Globomycin producer *Streptomyces hagronensis* belongs to the large family of *Streptomycetaceae*. Classified into the class of *Actinobacteria*, with 14 bacterial orders with a number of genera, are also fidaxomicin and chelocardin producers. Fidaxomicin producer *Actinoplanes deccanensis* [29] is a member of genus *Actinoplanes*. Genus *Actinoplanes* differs slightly from *Streptomyces*, since it produces true mycelium and forms spores that stay inside sporangia [24]. However, genus *Amycolatopsis* was once part of *Streptomyces* genus but got separated in 1986, since its species lack mycolic acid in their cell wall [30, 31]. Well-known members of the *Amycolatopsis* genus are *A. orientalis* and *A. sulphurea*, producers of vancomycin and chelocardin [32], respectively.

The development cycle of *Streptomyces* resembles that of filamentous fungi. They both grown as branching hyphae, form vegetative mycelium, and propagate through spores, which are formed in specialized breeding structures called aerial hyphae [33]. Both streptomycetes and fungi mostly live as saprophytes in the soil and are adapted to similar ecological niches, although *Streptomyces* spp. additionally inhabit a wide range of terrestrial and aquatic niches,

and some strains are plant and animal pathogens [33]. Production of diverse secondary metabolites, antibiotics, and other bioactive molecules, is coordinated with the *Streptomyces* developmental program [33].

**Table 1. Classification of three Actinobacteria species** according to Madigan et al. [24]

| | | | |
|---|---|---|---|
| **Kingdom** | *Bacteria* | *Bacteria* | *Bacteria* |
| **Phylum** | *Actinobacteria* | *Actinobacteria* | *Actinobacteria* |
| **Class** | *Actinobacteria* | *Actinobacteria* | *Actinobacteria* |
| **Order** | *Streptomycetales* | *Micromonosporales* | *Pseudonocardinales* |
| **Family** | *Streptomycetaceae* | *Micromonosporaceae* | *Pseudonocardiaceae* |
| **Genus** | *Streptomyces* | *Actinoplanes* | *Amycolatopsis* |
| **Species** | *Streptomyces hagronensis* | *Actinoplanes deccanensis* | *Amycolatopsis sulphurea* |

When a typical *Streptomyces* spore encounters favorable conditions with sufficient nutrients, it germinates; germ tubes emerge and grow to form hyphae [33] of 0.5–2.0 μm in diameter [24] (Figure 3). These germ tubes grow by tip extension [24] and branch to form a vegetative mycelium, often anchoring deep into the surrounding substrate [33]. The vegetative phase consists of a complex intertwined matrix, resulting in compact mycelium [24].

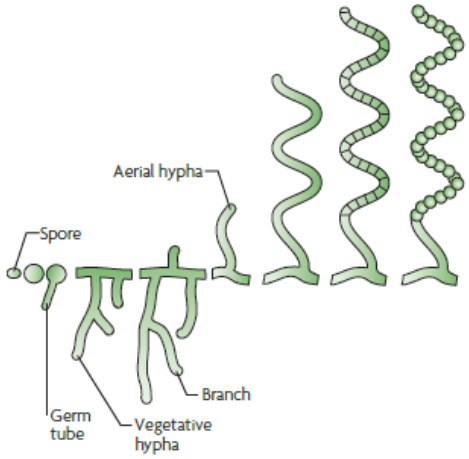



**Figure 3. The developmental life cycle of *Streptomyces coelicolor*.** Spore swells during germination, germ tube sprouts develop into hyphae. The hyphae extend and branch into vegetative mycelium that grows into and out of the surface and forms a vegetative colony. Upon depletion of nutrients, air hyphae grow and differentiate into a long chain of prespore compartments, which develop thick spore wall, synthesize a grey spore pigment and develop other characteristics of mature spores. Dormant spores are a means for dispersal of the organism in the environment, and can germinate, when suitable conditions are met [33]. The figure was taken from the publication by Flärdh *et al*. in 2009 [33].

Nutrient depletion and other signals initiate the production of secondary metabolites and morphological differentiation. Aerial hyphae must break the surface tension to grow into the

air. The aerial hyphae become divided by a developmentally controlled form of cell division into long chains of prespore compartments, which then thicken into spore walls. Mature spores synthesize a grey spore pigment and acquire other characteristics, such as hydrophobic surface of spore wall and relative drought-tolerance due to the abundance of trehalose [27, 34]. The optimal growth temperature for most species is 25-35 °C with some thermophilic and psychrophilic exceptions. The optimum pH for growth is between 6.5 and 8.0 [27].

For decades, it was seen as typical that one bacterial species apparently produces mainly one NP; only in some cases of better studied species more NPs were found. Growth conditions play a pivotal role for the expression of secondary metabolites. Therefore the so-called OSMAC approach (One Strain-MAny Compounds) [35] is being widely used and is providing new compound discoveries until today. However, the first whole-genome sequence of the model microorganism *Streptomyces coelicolor* A3(2) [36] published in 2002, changed the course of an entire research field and the genome era began. Since then it became increasingly evident how little was actually known even about intensely studied strains in terms of secondary metabolism.

### 1.3 Myxobacteria as producers of natural products

A gram of soil can contain $10^9$ bacteria, $10^6$ fungi, $10^3$ protozoa, $10^2$ nematodes, as well as annelids and arthropods [28]. It should not be a surprise that there are bacterial species, other than *Actinobacteria*, fighting for everyday survival with the help of secreted NPs. Recently, it was established that bacteria with genomes bigger than 8 Mb most probably contain biosynthetic gene clusters (BGCs) for secondary metabolites [37]. BGCs were in fact detected in members of the *Firmicutes*, containing classes of *Bacillus* and *Clostridia*, *β-proteobacteria*, *γ-proteobacteria* and *δ-proteobacteria* [37]. The Gram-negative rod-shaped *δ-proteobacteria*, myxobacteria (*Myxococcales*), with the largest bacterial genomes from 9 to 14.8 Mb [38] and high GC base pair content 67-72 % [39], were initially recognized due to their complex life-cycle, and “social behaviour” first reported in 1962 [40]. They are divided into three suborders: *Cystobacterineae*, *Sorangiineae*, and *Nannocystineae* [41] (Table 2). The model organism *Myxococcus xanthus* is known for its predatory lifestyle that explains some of its social interactions, unique developmental life cycle (Figure 4) and production of several NPs. In contact with either Gram-negative or Gram-positive bacteria, it also produces and excretes outer-membrane vesicles (OMVs) containing a combination of antibiotics and hydrolytic enzymes to lyse target prey, releasing nutrients for the growth of *M. xanthus* [42].

**Table 2. Taxonomic survey of the myxobacteria** according to Hartzell *et al*. [43]

Order: Myxococcales
- Suborder: Cystobacterineae
  - Families and genera
    - Myxococcaceae
      - *Myxococcus*
      - *Corallococcus* (formerly *Chondrococcus*)
      - *Pyxicoccus*
    - Cystobacteraceae
      - *Anaeromyxobacter*
      - *Archangium*
      - *Cystobacter*
      - *Melittangium*
      - *Stigmatella*
      - *Hyalangium*
- Suborder: Sorangineae
  - Families and genera
    - Phaselicystidaceae
      - *Phaesicystis*
    - Polyangiaceae
      - *Sorangium*
      - *Polyangium*
      - *Chondromyces*
      - *Byssovorax*
      - *Jahnella*
- Suborder: Nannocystineae
  - Families and genera
    - Nannocystaceae
      - *Enhygromyxa*
      - *Nannocystis*
      - *Plesiocystis*
      - *Pseudohygromyxa*
    - Kofleriaceae
      - *Kofleria*
      - *Haliangium*

At the cellular level, vegetative cells are flexible rods with dimensions of 5 x 0.8 mm that are motile on most solid surfaces but differentiate into non-motile, spherical spores (1-2 $mm^3$) upon nutrient depletion. At the multicellular level, *M. xanthus* forms elaborate biofilms that include macroscopic fruiting bodies where vegetative cells differentiate into spores. Their predatory behaviour has been compared to a “wolfpack” [42] and other microbes have been observed to be the preferred nutrient source [44]. *M. xanthus* utilizes two distinct types of motility systems – social (S), adventurous (A) motility – and four different cooperative, multicellular behaviours, including scouting, branching, rippling and fruiting-body aggregation, all of them observed on agar medium. Both motility and multicellular behaviour are recognized as integral for predation. Scouting refers to behaviour during surface colonization. Bacteria travel at low cell density, utilizing low-energy consuming A-motility, scouting around the area and leaving trails for neighbouring bacteria to join and therefore forming groups. Branching is observed at higher cell densities and facilitates faster migration of larger cell groups using rapid but highly-energy consuming S-motility. Rippling is utilized during predation of high-cell densities of *M. xanthus* traveling in waves back and forth across the available prey to maximize predator-prey contact.

Lastly, upon high cell density and starvation vegetative cells aggregate into stable, dormant fruiting bodies after cell differentiation into spores. Fruiting bodies form also during predatory interactions with *E. coli*, ensuring at least part of the population to go dormant, while peripheral rods remain active and continue the battle with the prey [42]. Myxobacteria play an important role in the global carbon cycle as degraders of plant biomass, and as micropredators they impact carbon flow and composition of microbial communities [28, 45] because of their cellulosic and proteolytic activities [46], respectively.

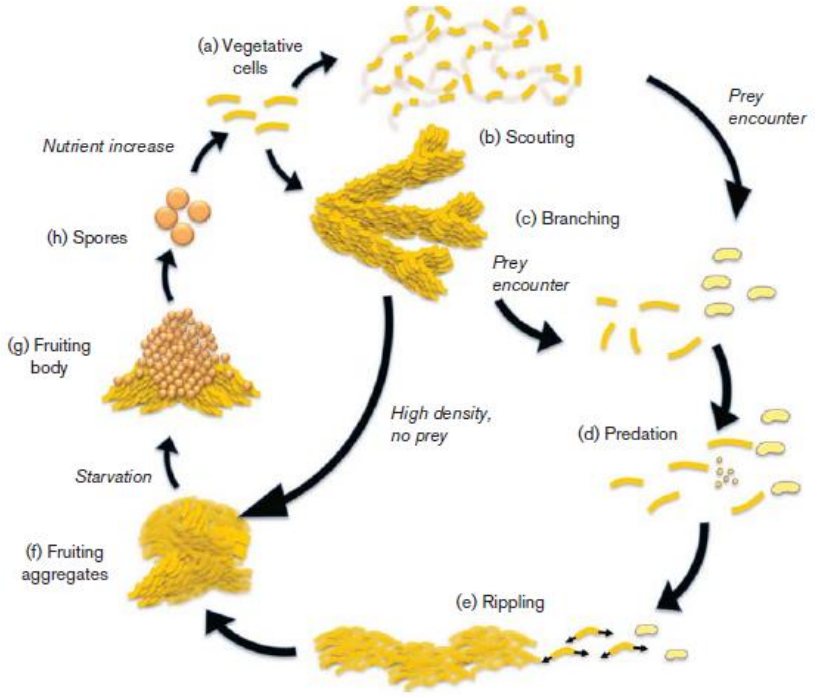



**Figure 4. The developmental life cycle of *Myxoccocus xanthus*.** Vegetative cells (a) depending on availability of nutrients either (b) scout for food or grow to high densities and migrate in a group, or (c) branch towards more nutrients. If no prey is encountered, vegetative cells (f) aggregate. Upon starvation, (g) fruiting bodies are formed storing (h) spores that will germinate when the nutrients are available again. If a vegetative cell encounters prey, they gather and attack the bacteria with an effective (e) rippling behaviour. Higher cell densities even during predation trigger aggregate formation, which is followed by differentiation into fruiting bodies [42]. The figure was taken from the publication by Keane *et al*. in 2016 [42].

Myxobacterial genomes might be larger than genomes of other bacteria due to large proteomes supporting their complex social behaviour [47]. For example, in *Sorangium cellulosum* So ce56 a large percentage of the genome appears to be devoted to regulation, particularly post-translational phosphorylation, with the highest number of eukaryotic protein kinase–like kinases discovered in any living organism [48], definitely supporting their social lifestyle [47]. Large myxobacterial genomes also encode biosynthetic pathways for a wide range of secondary metabolites that might come in handy for defeating prey during predation and potential protection from toxic compounds released by predator organism upon its lysis [42]. Since the

discovery of ambruticin in 1970 [49] more than 100 NPs with distinct core structures were isolated from various species of myxobacteria, some of them exhibiting potent biological activities [50, 51]. Myxobacterial NPs (Figure 5) exhibit biological activities such as antifungal (e.g. ambruticin [52], soraphen [53]), antibacterial (e.g. myxovirescin [54], cystobactamid [55]), antitumor (e.g. epothilone [56]) as well as antiviral (aetheramide [57]), antimalarial (e.g. chlorotonil [58]), immunomodulatory (e.g. argyrin [59]), antifilarial (e.g. corralopyronin [60]) and antioxidative (e.g. soraphinol [61]).

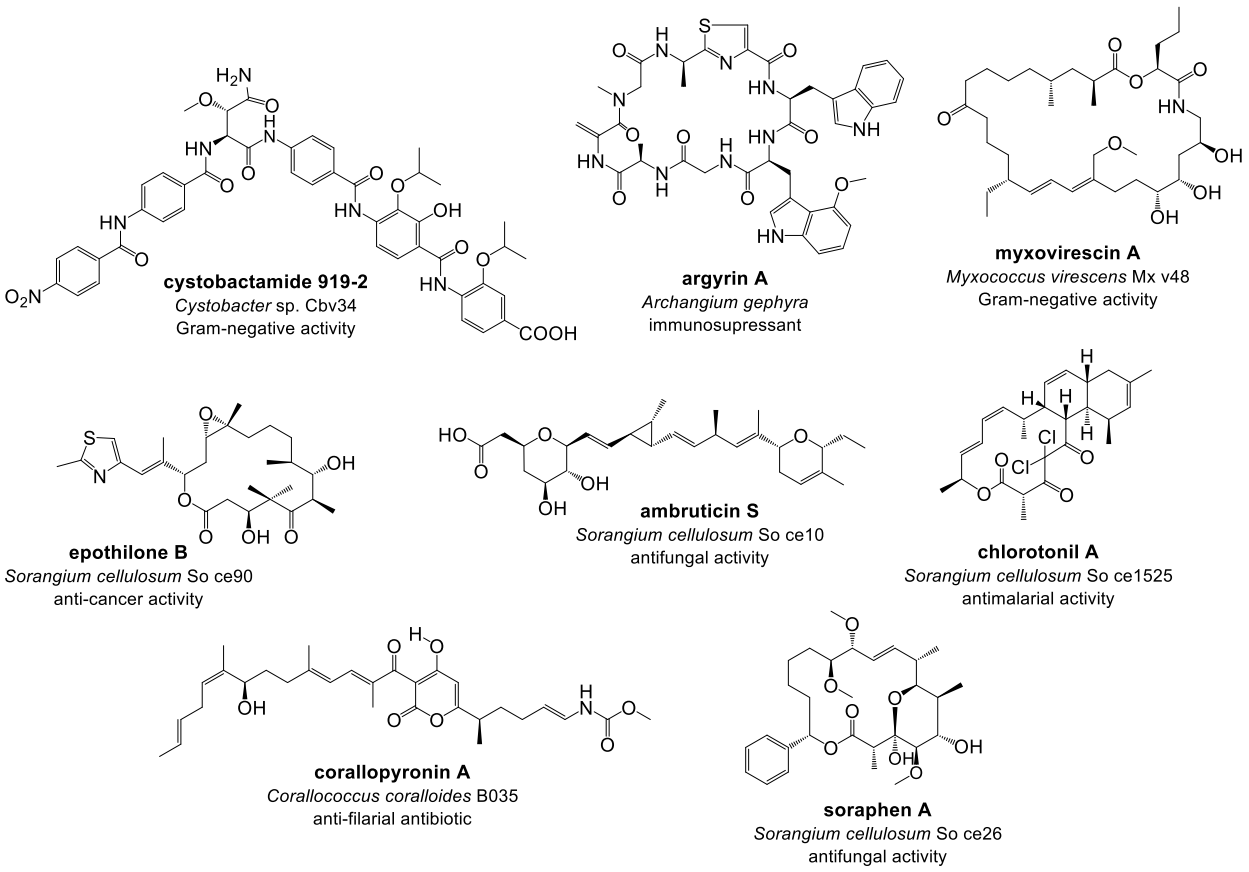



**Figure 5. The selected myxobacterial NPs**

Due to its antitumor activity, epothilone B entered phase II clinical trials [62] and its semisynthetic derivative ixabepilone was in October 2007 FDA-approved for the treatment of metastatic breast cancer [63]. Moreover, derivatives of bengamide B have been tested in clinical trials, however, only until the cardiotoxicity was observed [64], while tubulysin showed promising results in preclinical studies and was tested in phase I clinical trials [65], both compounds exhibiting anticancer properties. Two other promising drugs of myxobacterial origin are to be mentioned. The macrolactone antibiotic chlorotonil was isolated from *Sorangium cellulosum* So ce1525. It has potent bioactivity against Gram-positive pathogens and lacks cytotoxic activity but interestingly exhibits activity against *Plasmodium falciparum*

[51, 58]. Moreover, chlorotonil led to a substantial reduction of parasitemia in *Plasmodium berghei*-infected mice [58]. Another very potent antibacterial agent, cystobactamide, was isolated from *Cystobacter* sp. Cb v34 and was identified as an inhibitor of bacterial topoisomerases [55]. It is interesting for its high inhibitory activity against various Gram-negative bacteria [51] but it is active against a range of tested Gram-positive pathogens as well [55].

### 1.4 Biosynthesis of natural products

Secondary metabolite biosynthetic pathways are encoded on the bacterial genome through genes that are usually clustered, as opposed to the majority of biosynthetic genes in plants [66]. A major part of secondary metabolites known to date consists of polyketides and nonribosomal peptides built by polyketide synthases (PKSs), nonribosomal peptide synthetases (NRPSs), respectively, and hybrids thereof. The biosynthetic route to polyketide-based molecules built by modular PKS shares some similarity with fatty acid synthesis (FAS). Both processes use a common pool of building blocks – acyl-CoA units and apply the same chemical mechanism to link them together, namely repetitive decarboxylative Claisen thioester condensations of an activated acyl starter unit with malonyl-CoA-derived extender units [67]. Nonribosomal peptides are synthesized by NRPS enzymes that as precursors use proteinogenic, nonproteogenic amino acids and carboxylic acids [68]. Both mentioned types of machinery for biosynthesis of secondary metabolites consist of large multimodular enzymes. Each module subsequently activates and adds a new building block to the growing acyl chain, which resembles an assembly line [68]. Modules consist of domains with active sites, each catalyzing a reaction step. A basic module for chain elongation with an activated building block by modular PKS consists of ketosynthase (KS), acyltransferase (AT) and acyl carrier protein (ACP) domain. In NPRS biosynthesis, it consists of condensation (C), adenylation (A) and peptidyl carrier protein (PCP) domain. The above-described modular PKS (type I) mechanism is categorized as noniterative since each KS domain catalyzes only one round of elongation, as opposed to iterative PKS type II where the KS domain catalyzes more elongation rounds [67]. PKS type II requires a “minimal PKS” that consists of two ketosynthase units – KSα and KSβ, and an ACP that is tethering the growing polyketide chain but can contain additional domains like ketoreductases (KR), cyclases (CYC) and aromatases (ARO) that modify polyketide chain [67]. Common features of all three described biosynthesis systems are the carrier proteins. Carrier proteins (acyl- or peptidyl-) are expressed in inactive form and require holo-[acyl-carrier-protein] synthase (AcpS), a 4’-phosphopantetheinyl transferase, for 4'-

phosphopantetheine moiety to be post-translationally attached to the hydroxyl group (-OH) of conserved serine residue of the ACP or PCP (Figure 6) [68, 69].

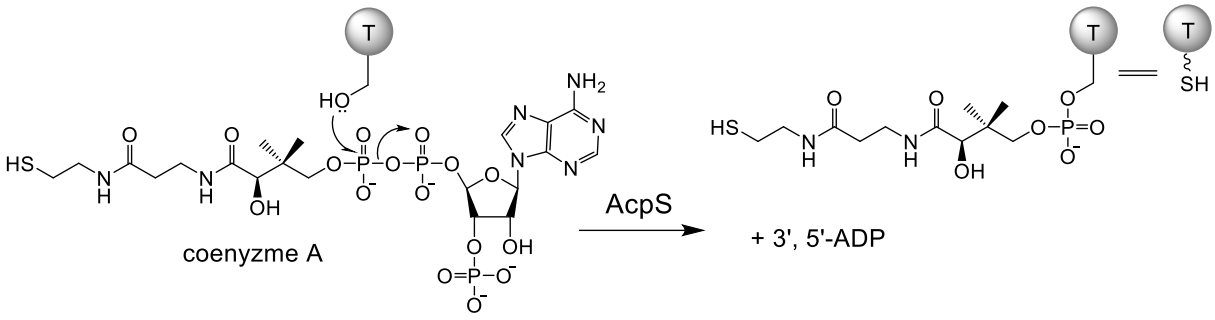



**Figure 6. The activation of thiolation (T) domains (or carrier proteins) in PKS and NRPS biosynthetic genes.** 4'-phosphopantetheinyl transferase AcpS catalyzes the breakdown of coenzyme A into 4'- phosphopantetheinyl moiety and ADP. The former reacts with hydroxyl group of conserved serine and so activates the domain, which is now ready to form thioester bonds with acyl-CoA units.

Activated carrier protein domains contain an additional flexible 4'-phosphopantetheinyl arm that provides a thiol (-SH) group on which substrate and intermediate acyl chains are covalently tethered during chain elongations [68]. Acyl groups are important components of primary metabolism. They are metabolically accessible when covalently bound to a reactive thiol group of coenzyme A (CoA), forming a thioester [68]. Thioesters have high acyl group transfer potential and can donate their acyl groups to a variety of acceptor molecules. Hydrolysis of acyl-CoA thioester bond is highly exergonic but the transthiolation to carrier protein is an energetically neutral reaction. The released energy is used for condensation of acyl group to the elongating acyl chain, with $CO_2$ loss, and for translocation of the intermediate acyl chain from ACP to the cysteine residue of KS [68]. A KS domain in PKS module elongates the acyl chain by forming C-C bonds, while C domain in NRPS catalyzes the formation of peptide bonds (C-N). Both biosynthetic systems are compatible and can therefore form PKS-NRPS as well as NRPS-PKS hybrids [68]. The biosynthetic systems compatibility creates a higher diversity of secondary metabolite core structures. Biosynthesis of epothilone is an example of a hybrid PKS/NRPS natural product (Figure 7). Its main core is biosynthesized in a modular fashion by EpoA, EpoP, EpoB, EpoC, EpoD, EpoE proteins [70]. EpoA is encoding the loading module and the rest of proteins nine other biosynthetic modules. Each module subsequently activates one building block and adds it to the growing acyl chain. Every malonyl-CoA extender unit adds two carbons to the elongating carbon chain. The acyltransferase ($AT_L$) domain in the loading module accepts and activates an acetyl-CoA unit that condensates onto $ACP_L$ domain as acetyl moiety [70]. Adenylation (A) domain in NRPS EpoP selects and activates the amino acid cysteine. Adjacent cyclization (Cy) domain transforms the intermediate chain into a 2-

methylthiazoline that after oxidation (Ox) transforms into a 2-methylthiazole attached to PCP [71]. Acyltransferase of each following module selects a new building block, either malonyl-CoA or methylmalonyl-CoA unit that is after activation attached to the ACP [70]. Ketosynthase (KS) domain is responsible for condensation of the newly activated unit to the rest of an intermediate acyl chain. If no other catalytic domains are present in the module, this intermediate chain is transferred to the KS domain of the next module where the process is repeated.

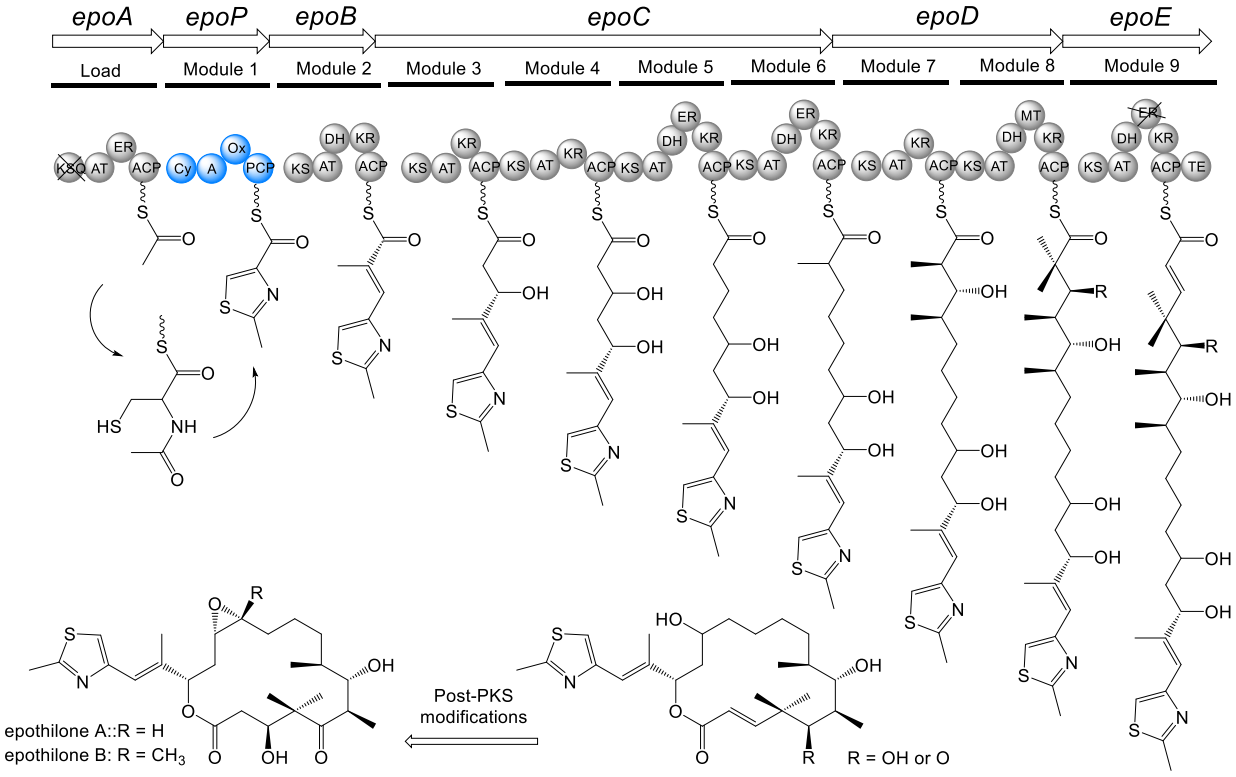



**Figure 7. The biosynthetic gene cluster of epothilone.** Biosynthetic gene cluster consists of PKS and NRPS parts. Acyltransferases of different modules accept only certain building blocks. Each newly condensed acyl moiety is processed by domains present in each module. The β-keto group can be reduced (KR) to hydroxyl group, further dehydratation (DH) yields double bond and after enoyl reductase´s (ER) action double bond is fully saturated. Additional domains like methyltransferase (MT) domains can be present. Some domains do not participate in the modification of the intermediate acyl chain since mutation in their active site inactivates them (e.g. ER in module 9). PKS domains are highlighted in grey and NRPS domains in blue.

Modules of PKS type I frequently contain domains that can modify β-keto group on the growing chain, with reduction and dehydratation steps, either partially or all the way to a saturated acyl chain characteristic for fatty acids. Furthermore, many other domains exist in biosynthetic gene clusters of NPs, additionally increasing their diversity, such as epothilone´s C-methyltransferase (c-MT) domain present in module 8 [70]. Catalytic domains in some cases contain a mutation in their active site and are therefore inactive, e.g. the enoyl reductase (ER)

domain in module 9 of the epothilone pathway. Chain termination of epothilone PKS requires a thioesterase (TE) domain, to release the finished polyketide chain via macrolactonization. The final structure of epothilone is a product of post-PKS modifications that take place after the polyketide chain release [70].

### 1.4.1 Nonribosomal peptides

Nonribosomal peptides are synthesized by giant NRPS enzymes. A minimal NRPS elongation module has two catalytic domains, analogous to AT and KS domains of PKS modules. The adenylation (A) domain selects the amino acid, activates the carboxylate with ATP to make the aminoacyl-O-AMP and installs the aminoacyl group on the thiolate of the adjacent thiolation domain or peptidyl carrier protein (PCP) [68]. The amino acid specificity of A domains is determined by 10 highly conserved core motifs, established on the phenylalanine-activating adenylation domain PheA [72]. Highly conserved Lys517 residue (A10 motif), together with Asp235 (A4), stabilizes α-amino and the α-carboxylate groups of the substrate amino acid by electrostatic interactions [72]. A comparison of functional A domains revealed 3 % variation in amino acid residue at the position 235, while no variation in position 517 was found [72]. The condensation (C) domain catalyzes the chain elongation, joining an upstream peptidyl-S-$PCP_{n-1}$ to the downstream aminoacyl-S-$PCP_n$ [68] onto $PCP_n$ (Figure 8) with a peptide bond. The elongation process reminds of ribosomal peptide synthesis where the amino acid is adenylated, added an AMP, by an aminoacyl-tRNA synthetase and chaperoned to the large subunit of the ribosome for mRNA-instructed protein biosynthesis [73, 74].

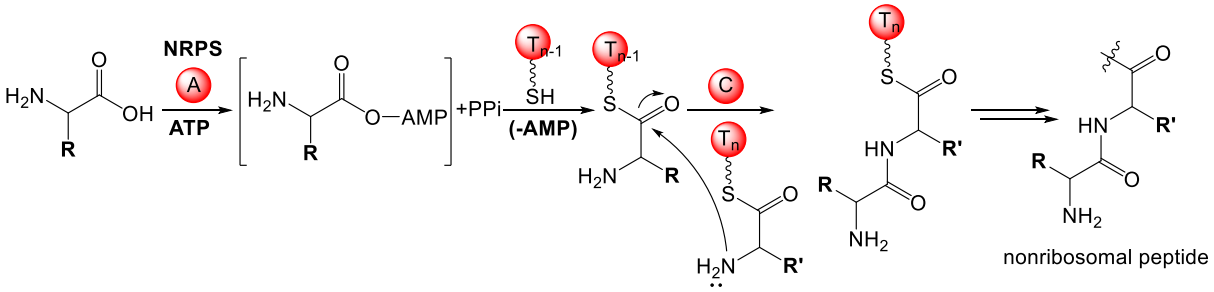



**Figure 8. The elongation of a nonribosomal peptide (NRP).** Selected amino acid is adenylated or activated by an ATP molecule via NRPS A domain. The aminoacyl-O-AMP intermediate binds to following PCP or thiolation (T) domain. The condensation domain catalyzes peptide bond formation between the PCP-bound amino acid (R) and the adenylated amino acid (R') bound to the downstream T' domain. Upon completion of peptide chain assembly, the nonribosomal peptide is released usually by TE domain (not shown).

A wide variety of building blocks is used by NRPS, such as proteinogenic, nonproteogenic amino acids and carboxylic acids [68]. It has been estimated that about 500 different building

blocks have been identified, which makes the 20 proteogenic amino acids a minority with only 4 % [74]. In contrast, only some nonproteinogenic amino acids are utilized as intermediates in primary metabolic pathways (e.g. homoserine, ornithine) but the rest (96 %) is synthesized by microbes and used for the biosynthesis of secondary metabolites [74]. The proteinogenic L-amino acids are available in the common pool of building blocks for primary metabolism. On contrary, the nonproteinogenic amino acids are mostly generated according to the requirements of specific secondary metabolite pathway [74]. They are either modified L-amino acids by additional catalytic domains present in NRPS modules or biosynthesized *de novo* by enzymes encoded in the proximity of the biosynthetic genes [74]. Many NPs contain nonproteogenic D-amino acids. The conversion can be catalyzed by pyridoxal-phosphate (PLP)-dependent or PLP-independent racemases [74] encoded in the BGC or by epimerization (E) domain as part of a biosynthetic module. In vancomycin biosynthesis, there are three NRPS modules with additional E domains that convert L-β-hydroxytyrosine (L-βHt) and L-4-hydroxyphenylglycine (L-Hpg) into the corresponding D-type amino acids [75]. The domain organization in modules containing E domain is $C_n$-$A_n$-$T_n$-$E_n$. The adenylated L-amino acid bound onto the intermediate peptidyl chain by module n is accepted by $C_{n+1}$ domain and only then epimerized by $E_n$. Such $C_{n+1}$ domains are designated as L-acceptors but have D-peptidyl-S-T donor selectivity ($^{D}C_L$) [68, 76]. In addition to common $^{L}C_L$ and $^{D}C_L$ domains, some C domains have heterocyclization activity towards Cys, Ser and Thr amino acids [68]. They catalyze cyclization using thiol (Cys) and hydroxyl (Ser, Thr) groups from their side chains to form a five-membered cyclic adduct, which after dehydratation yields thiazoline or oxazoline rings (Figure 9), respectively [68]. Those C domains are termed as Cy (cyclization) domains and have cyclodehydration activity.

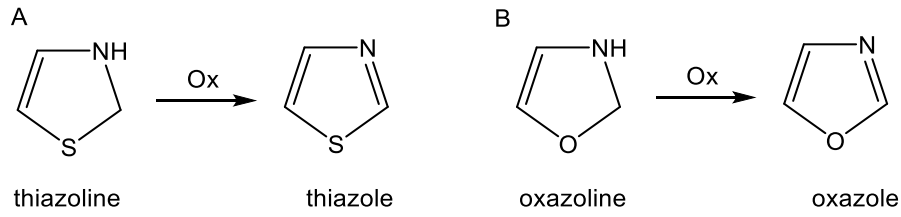



**Figure 9. Reactions of NRPS oxidase domain.** A. Oxidase catalyzes conversion of thiazoline ring to thiazole and B. oxazoline ring to oxazole.

Adjacent to Cy domains, oxidation (Ox) domains are found that catalyze oxidation of the thiazoline ring to thiazole and oxazoline ring to oxazole (Figure 9) [77]. The functional Ox domain can be positioned upstream (Ox-PCP) or downstream (PCP-Ox) of PCP domain, e.g. in EpoB of epothilone [71] or in BlmIII of bleomycin [77], respectively.

Frequently appearing in NRPS modules are methylation (MT) domains that are inserted into A domains. They catalyze the transfer of a methyl group from cosubstrate S-adenosylmethionine (SAM) to the carbon, nitrogen or oxygen atoms to the various positions on the backbones of PKs, NRPs and fatty acids and are classified as c-MT, n-MT and o-MT, depending on their site of methylation [74, 78]. However, many optional catalytic domains can be organized within a module (in cis) [68] or they can also be encoded outside of the biosynthetic gene (in trans) and still modify the growing chain, while it is tethered on the carrier domain of the assembly line [68]. Aminomutases interconvert α- and β-amino acids and so form nonproteinogenic β-amino acids. Two types of microbial enzymes with mutase activity have been observed, e.g. conversions of Phe and Tyr are characteristic for one of the types. The superfamily of the latter contains a covalently attached cofactor 4-methylidene-5-imidazole-5-one (MIO) that acts as an electrophilic center for attack by the amino group that gets modified. The second superfamily of amino acid mutases consists of lysine-2,3-aminomutase, glutamate-2,3-aminomutase, D-ornithine-4,5-aminomutase and lysine-5,6-aminomutase that utilize both PLP and adenosyl-B12 as required cofactors [74]. Commonly found in various biosynthetic gene clusters are hydroxylases that hydroxylate side chains of NPs. They either harbour a redox active iron species generating a high valent oxoiron intermediate or they use the organic coenzyme FAD, which forms a less reactive FAD-OOH complex that acts like oxygen transfer agent. An amino acid chlorinase functions in the same manner as certain oxygenase enzymes. The responsible catalysts are again mononuclear iron enzymes, requiring an α-ketoglutarate, chloride and oxygen, similar to Fe(II)- and α-ketoglutarate-dependent dioxygenases [79]. Such halogenases act on aminoacyl-S-T substrates Pro, Leu and Ile. In case of Tyr with electron-rich phenol side chain and Trp with indole ring a weaker chlorinating agent is required, so the $FADH_2$-containing halogenases are used [74]. Some amino acid modifications are described above to underpin a general idea, while many less common ones exist with some being only single-case specific [74].

### 1.4.2 Polyketide synthase type I as exemplified by tiacumicin B biosynthesis

Modular PKS systems will be further described on the example of tiacumicin B, also called fidaxomicin, biosynthetic gene cluster [80], a secondary metabolite produced by *Dactylosporangium aurantiacum* subsp. *hamdenensis* and few other strains (more in Section 4.1). The reported BGC consists of 31 genes, where only four genes - *tiaA1-tiaA4*, encode multifunctional type I modular PKSs. The core structure is assembled by a loading unit and 8

extending modules (Figure 10). Modules consist of standard domains found in PKS systems, namely KS, AT, DH (dehydratase), KR, ACP and TE domain [80].

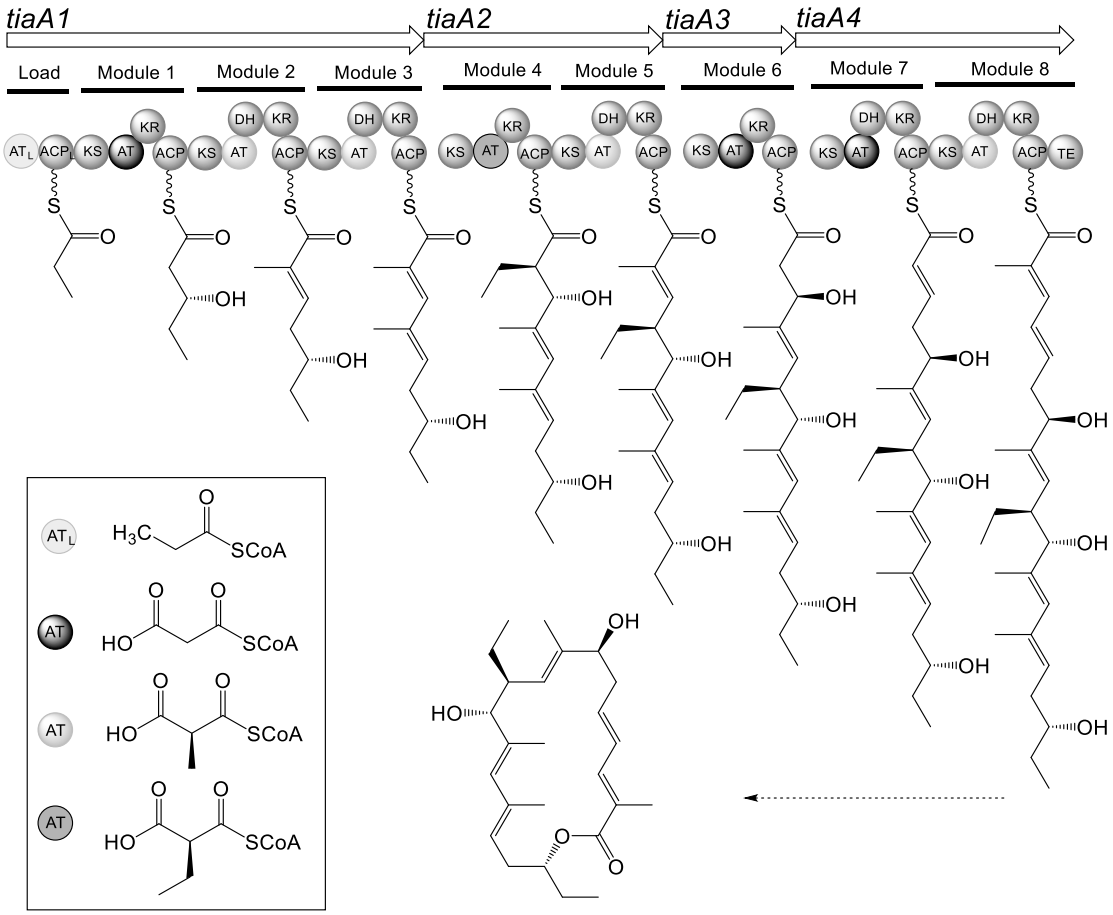



**Figure 10. The biosynthesis of tiacumicin B aglycone.** Different shade of AT domains denotes specific selectivity for building block as shown in the box. The biosynthetic scheme was constructed based on publication by Xiao *et al*. in 2011 [80].

As shown in Figure 10, $AT_L$ domain from the loading module selects a propionyl-CoA unit and loads it onto $ACP_L$. A propionyl moiety is transferred to $KS_1$ domain. Meanwhile, $AT_1$ selects the malonyl-CoA unit, which is attached to $ACP_1$. $KS_1$ condenses propionyl moiety and malonyl-CoA unit, releasing in sum a $CO_2$ molecule in the process. $KR_1$ in the module reduces the β-keto group to a hydroxyl group (Figure 11). The resulting acyl chain is then loaded onto $KS_2$. $AT_2$ selects a methylmalonyl-CoA (Figure 10) and loads it onto $ACP_2$. $KS_2$ condenses the previous acyl chain with new methylmalonyl moiety. Its β-keto group is reduced by $KR_2$ to β-hydroxyl group and further dehydrated by $DH_2$, yielding a double bond. Modified chain by module 2 is transferred to $KS_3$. The process continues until the growing acyl chain arrives in the last module, where it is cleaved off by the TE domain. The tiacumicin B polyketide chain is so released and cyclized into 18-membered macrocyclic aglycone. Functions of main catalytic domains and their products are shown in Figure 11. During the polyketide chain

biosynthesis, a type II thioesterase TiaE is surveilling precursor loading and removing them if not fitting [81]. A type II TE may function as editing enzyme and promote the accuracy and efficiency of biosynthesis.

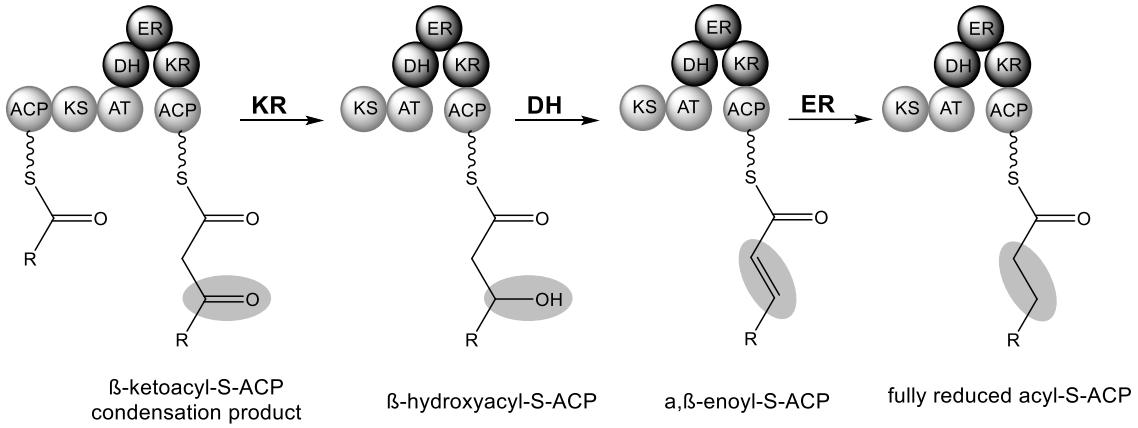



**Figure 11. What reactions do the PKS biosynthetic domains catalyze?** If the module contains only KS, AT and ACP domain, a condensation product is β-ketoacyl-S-ACP. A ketoreductase domain reduces β-keto group to β-hydroxyl group. If a dehydratase domain is present in the module, a hydroxyl group is dehydrated and loss of $H_2O$ molecule yields α,β-enoyl group or double bond. An enoyl reductase (ER) domain further reduces the double bond and saturates it. In the biosynthesis of polyketide metabolites, KR, DH, ER domains are optional [68].

The active sites of catalytic domains in PKS systems are well-characterized, which is fundamental for the accurate prediction of domain's functionality. First, the active sites T209, C221, S315, H346, K379, N381, H381, H384, Q386, N455 and H457 of KS domains were studied on the example of EryA1-KS [82]. The residues C211, H346, K379, H384 are universally conserved but T209, S315, N381, Q386, N455 and H457 are only conserved within either an iterative (LovB-KS) or a non-iterative KS (EryA1-KS) domain subgroup [82]. Second, AT domains can select various moieties from the primary metabolite pool, e.g. malonyl-CoA, methylmalonyl-CoA or ethylmalonyl-CoA. The sequence analysis of AT domains revealed distinct motifs believed to be directly involved in substrate selectivity [83]. In fidaxomicin biosynthesis, "Q...GHSIGE...HAFH" motif indicates malonyl-CoA specificity, while "Q...GHSQGE... YASH" and "Q...GSSQGE...VASH" motifs indicate (2S)-methylmalonyl-CoA and (2S)-ethylmalonyl-CoA specificity, respectively [80]. Third, KR domains found in PKS modules are usually members of a complex short-chain dehydrogenases/reductase (SDR) family [84]. Those members are present in all living organisms as parts of multifunctional enzyme complexes, e.g. fatty acid synthase. They are NADP(H)-binding proteins with the SDR region with a β-ketoacyl reductive function. This group has the unique YxxxN motif at the active site rather than the typical YxxxK [84]. However, they all require "GGxGxxG" NADP-binding motif to function. The active sites of KR domains were studied on the example of

tropinone reductase-II (TR-II), where Tyr159, Ser146 and Lys163 amino acid residues form the catalytic triad [85]. Additionally, the KR domains are predicted to produce a carbon centre with D-configuration in case that Tyr residue, Tyr100 in TR-II, aligns to Asp [85]. Fourth, a DH domain is characterized by four motifs, namely HxxxGxxxxP, GYxYGPxF, LPFxW, and Dxxx(Q/H), which were established based on erythromycin dehydratase [86]. Similarly, conserved residues of ER domains [87] and conserved binding sites of ACP domains [88] are well characterized.

The remaining 26 genes in fidaxomicin biosynthesis are responsible for the supply of building blocks, their attachment to the molecule, tailoring reactions, the export of the final product out of the cell and regulation of the biosynthesis (Figure 12). Functions of some genes will be described below in more detail since fidaxomicin biosynthesis covers many aspects of a secondary metabolite biosynthesis, an important topic in NP research discussed in each of the following chapters.

#### 1.4.2.1 Tailoring reactions in PKS type I biosynthesis on the example of fidaxomicin

Simple building blocks utilized by PKS, like malonyl-CoA, are available from primary metabolism but more complex ones have to be additionally synthesized and are generated by dedicated genes. In the case of tiacumicin B, six genes are necessary to provide building blocks such as starter and extender units for antibiotic biosynthesis. A starter unit propionyl-CoA is generated by TiaC and TiaD, encoding the branched-chain α-keto acid dehydrogenase E1-α and E1-β subunit, respectively. They are proposed to be involved in the generation of isobutyryl-CoA, propionyl-CoA, and other acyl-CoA analogs by oxidative decarboxylation of α-branched-chain fatty acids derived from α-branched-chain amino acids [80, 89]. On the other hand, the extender units are formed by the action of TiaJ, TiaK, TiaL and TiaN. TiaL, a propionyl-CoA carboxylase, should generate the 2-methylmalonyl-CoA extender unit. The genes *tiaJ*, *tiaN* and *tiaK* encode a 3-hydroxybutyryl-CoA dehydrogenase, a crotonyl-CoA hydratase, and a crotonyl-CoA carboxylase/reductase, respectively, and are involved in the ethylmalonyl-CoA pathway [90, 91]. As sometimes different moieties are attached to the polyketide scaffold, those have to be biosynthesized separately. The polyketide backbone of fidaxomicin bears modified D-noviose and D-rhamnose moieties [92]. The attached sugar moieties are biosynthesized from GDP-D-mannose in a two-step procedure catalyzed by GDP-D-mannose dehydratase (Gmd) and GDP-4-keto-6-deoxymannose reductase (Rmd) [93]. However, in *D. aurantiacum* three putative Gmds – TiaS1, TiaS3 and TiaS4 are sufficient to carry out this reaction [80], while in *Actinoplanes deccanensis* only FadS1 and FadS4 seem to suffice [94]. The homo-orsellinic acid

moiety is biosynthesized by an iterative type I PKS TiaB (Figure 12). TiaB contains KS-AT-DH-KR-ACP domains but its DH and KR domains are not active. Usually, acetyl-CoA and three malonyl-CoA units would make up a homo-orsellinic acid moiety [95] but in the case of tiacumicin B a propionyl-CoA is used instead of an acetyl-CoA unit. These building blocks are iteratively processed by TiaB and so condensed into a chain that is cyclized [80].

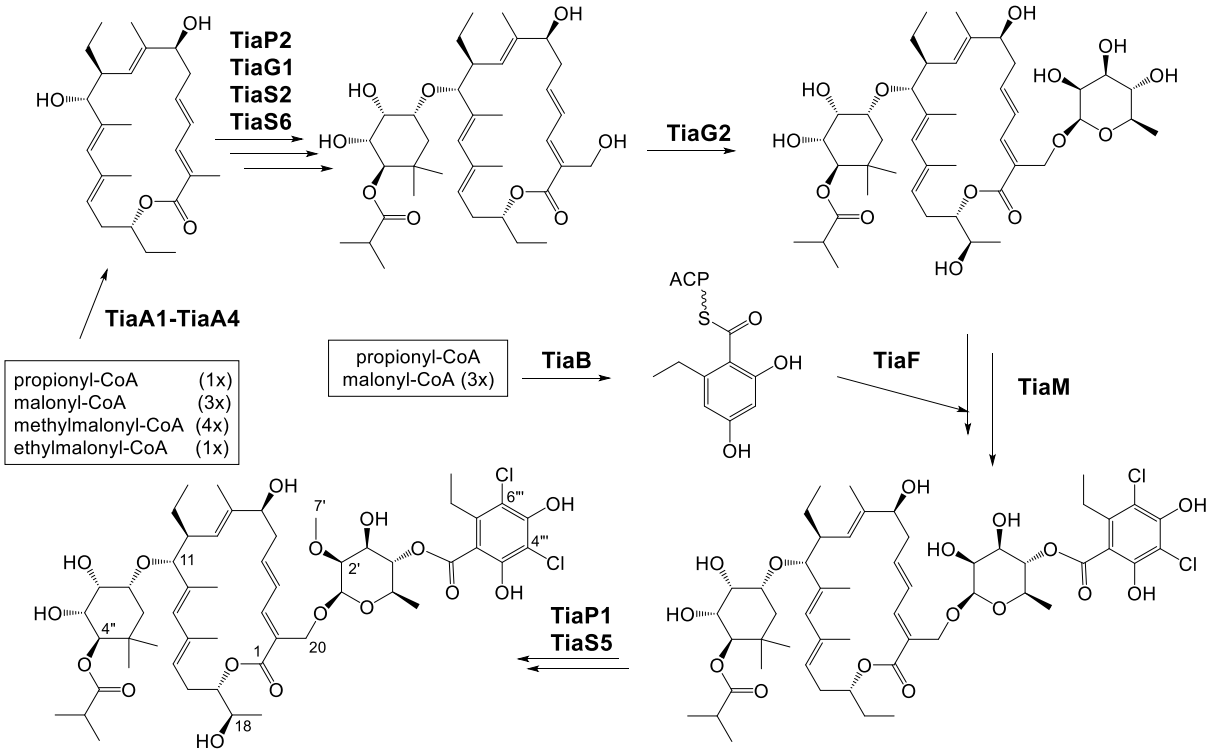



**Figure 12. Tiacumicin B biosynthesis: Tailoring steps.** A polyketide macrocycle is subjected to a number of tailoring reactions. TiaP2 adds a hydroxyl group to C-20 position, TiaG1 transfers a D-noviose moiety, TiaS2 modifies a D-noviose moiety into a 5-methyl-D-noviose and TiaS6 adds an isobutyryl moiety to C-4'' position onto the sugar. The D-rhamnose is transferred by TiaG2 to C-20 hydroxyl group. Meanwhile, a homo-orsellinic acid moiety is built, incorporated by TiaF and further halogenated by TiaM [80]. TiaP1 adds a hydroxyl group to C-18 position [96] and TiaS5 methylates a C-2' hydroxyl group, forming a 2-O-methyl-D-rhamnose [97]. Building blocks used in the biosynthesis are generated by *tiaC*, *tiaD*, *tiaL*, *tiaJ*, *tiaN*, *tiaK*, *tiaS1* and *tiaS4* genes.

Following tailoring reactions are catalyzed by different transferases, a halogenase and hydroxylases. TiaG1 and TiaG2 glycosyltransferases transfer D-noviose and D-rhamnose moieties to C-11 and C-20 hydroxyl group, respectively. TiaS2 as C-methyltransferase and TiaS5 as O-methyltransferase modify D-noviose and D-rhamnose moieties into 5-methyl-D-noviose and 2-O-methyl-D-rhamnose, respectively [80]. An O-acyltransferase TiaS6 esterifies a C-4'' hydroxyl group with an isobutyryl moiety (Figure 12). Moreover, an acyltransferase TiaF shares similarity with other AT domains, responsible for the incorporation of an aromatic moiety into corresponding secondary metabolite, and incorporates the homo-orsellinic acid

moiety to the 2-O-methyl-D-rhamnose residue [80]. Further tailoring reactions are catalyzed by a halogenase TiaM that incorporates two chlorines onto the attached homo-orsellinic acid moiety [80] and by two cytochrome P450 (CYP) hydroxylases TiaP1 and TiaP2 that catalyze hydroxylations at C-18 and and C-20, respectively [80], directly modifying the polyketide aglycone.

Additional tailoring enzymes often present in the PKS type I NPs are the aminotransferases (AMTs). They use glutamine as a substrate for the transamination reaction. The α-amino group located on its side chain is transferred to the β-ketoacyl-S-T yielding β-aminoacyl-S-T and α-keto-glutamine as a side product [68, 98]. Same as in the NRPS biosynthesis, many catalytic domains can be found within a module (in cis) [68], e.g. AMT domain in MycA from biosynthesis of mycosubtilin [98] or can be encoded outside of the biosynthetic gene (in trans) [68].

### 1.4.3 Type II PKS

As mentioned in Section 1.4, a polyketide synthase (PKS) type II contains a minimal set of enzymes, a so-called "minimal PKS" that consists of two ketosynthase (KS) units – KSα and KSβ, and an acyl carrier protein (ACP) catalyzing the biosynthesis of nascent polyketide chain [67]. The KSα subunit catalyzes Claisen type C-C bond formations with SCoA-activated acyl and malonyl building blocks. Since a number of reiterative elongation rounds take place, the PKS type II is therefore also called "iterative type PKS" [67]. The KSβ unit with extremely high homology to KSα lacks the catalytic cysteine residue in the active site [99]. The function of KSβ is not yet entirely understood, and two hypotheses exist on the function of that subunit. According to one hypothesis, KSβ acts as the primary determinant of carbon chain length also known as chain length factor (CLF) [99]. However, the role of KSβ in starter unit selection, often involving decarboxylative activity, was also suggested [100]. KSβ domain is in that case referred to as KSQ since Q is an important glutamine amino acid residue involved in such decarboxylative activity. However, the glutamine residue is important also for polyketide synthesis [100]. Both subunits form a heterodimer [67], providing a hydrophobic environment for the assembled poly-β-keto intermediate carbon chain. The minimal PKS can therefore partially control regiochemistry of the first-ring cyclization [99]. The PKS type II systems can contain additional domains like ketoreductases (KR), cyclases (CYC) and aromatases (ARO) that modify the polyketide chain [67]. In chelocardin (CHD) biosynthesis, the minimal PKS consists of three genes: *chdP*, *chdK* and *chdS* [32]. Chelocardin is a tetracycline (TC) antibiotic that differs from other tetracyclines, hence designation as "atypical" TC. It uses malonate or

acetate instead of malonamate starter unit, which is likely used in oxytetracycline (OTC) biosynthesis (Figure 14) [101]. In comparison to OTC, CHD contains an additional methyl group at C9 and an unmethylated amino group at C4 position [32]. Presumably, ten malonyl-CoA units are used by minimal PKS complex to produce full length polyketide chain (Figure 13) [32].

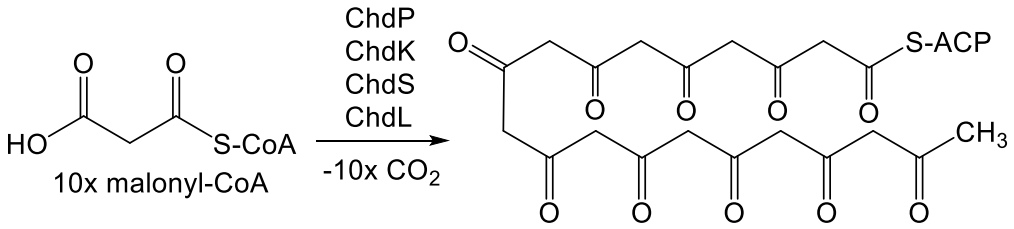



**Figure 13. Chelocardin polyketide chain assembly by minimal PKS.** Presumably, 10 malonyl-CoA units are iteratively assembled by ChdPKS minimal PKS complex.

A putative aromatic PKS ketoreductase (KR) ChdT reduces the carbonyl group at the C8 position of CHD (Figure 14). According to the current understanding of CHD biosynthesis, two cyclases/aromatases (CYC/ARO) ChdQI and ChdQII are responsible for cyclization and aromatization of the first ring [32]. In contrast, only one CYC/ARO homolog (OxyK) is neccesary to fullfil this task in OTC biosynthesis [101]. Moreover, second ring formation in OTC biosynthesis is catalyzed by OxyN [101], however, in CHD biosynthesis no protein homolog was initially found and cyclization was assumed to occur spontaneously [32]. Cyclization of the third ring is spontaneous and cyclization of the fourth is not yet understood in any TC system [32]. It is thought that tailoring reactions in CHD biosynthesis take place after the chelocardin backbone is formed. Although, there are some indications that methylation of the C6 position in OTC might happen even earlier [102]. The post-PKS tailoring steps occur in all the TCs, however, to a different extent, which affords the diversity of TC natural products. One of the most studied TCs in that regard is oxytetracycline with an additional hydroxyl group at C5 position, when compared to TC (Figure 14). The oxytetracycline backbone was in *in vitro* experiments shown to be modified by several enzymes such as a methyltransferase OxyF that methylates C6, an oxygenase OxyL that hydroxylates C4 and C12a positions with help of an ancillary oxygenase OxyE and an aminotransferase OxyQ that exchanges a keto group at C4 position with an amino group, which is then dimethylated by a methyltransferase OxyT (Figure 14) [101]. Further hydroxylation of C6 position is believed to be assisted by an oxygenase OxyS but the following reactions such as hydroxylation of C5 position and reduction of C5a-C11 double bond, leading to the final product of oxytetracycline, are not yet understood [101]. Another post-PKS tailoring step observed in TC biosynthetic pathways is halogenation of C7 position catalyzed by a halogenase CtcP of a typical TC chlorotetracycline (Figure 14) [103].

An atypical TC chelocardin compared to just described OTC biosynthetic pathways lacks the methylation of the amino group at C4 position and contains a methyl group at C9 position [32]. Moreover, other post-PKS modifications such as O-glycosylation of the C10 position were found in TC biosynthetic pathways [104]. The final chemical structures of some fully modified tetracyclines backbones are shown in Figure 14.

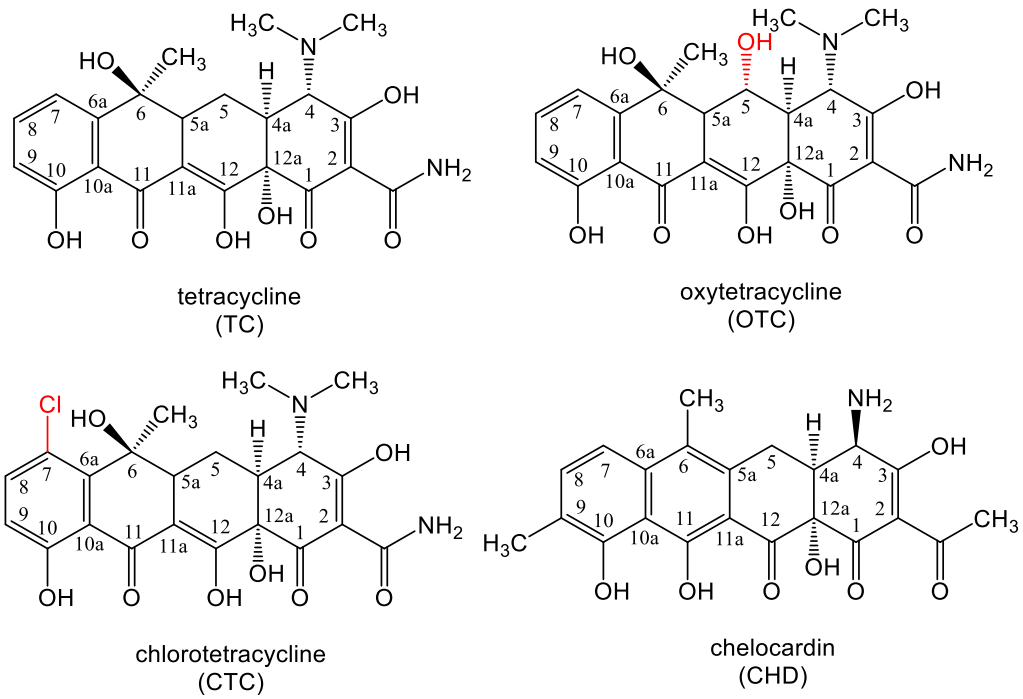



**Figure 14. The structures of tetracyclines.** TC, OTC and CTC are a typical tetracyclines in contrast to atypical CHD. Although OTC and CTC scaffolds are additionally modified at positions C5 and C7, respectively, their starting unit is an amide group opposed to an acetyl group in CHD [101]. Moreover, CHD differs in the modifications at C4, C6, C9, C11 and C12 positions [32].

### 1.4.4 Ribosomally synthesized and post-translationally modified peptides

The addition to the classes of natural product with antibacterial activities described above are ribosomally synthesized and post-translationally modified peptides (RiPPs). This group of antimicrobial agents often exhibits a narrow activity spectra, making them especially suitable for clinical applications since the off-target effects can be so avoided [105]. RiPPs are produced or at least encoded by many bacterial phyla, also anaerobic bacteria [105], and observed to most often target near relatives of the producing organism. Regardless, they have diverse targets either interrupting the cell membrane or the cellular processes within the cell, such as DNA, RNA and protein metabolism [106]. RiPPs are a large group of structurally diverse NP classes, encompassing lanthipeptides, bacteriocins, thiopeptides, lasso peptides, just to name a few, with representatives such as Gram-positive antibiotics nisin and bottromycin [107] or most recently reported a Gram-negative antibiotic darobactin [108]. Although they are diverse, their

biosynthetic logic is mostly uniform. First, a precursor peptide consisting of an N-terminal leader sequence and a C-terminal core sequence, encoded by a single gene, is translated on the ribosome. The leader sequence is cleaved by peptidases and the remaining peptide is further modified by other enzymes, usually encoded in the same BGC, resulting in a biologically active NP [105]. RiPPs are also ideal for engineering since the precursor peptides are encoded by one single gene and point mutations in their sequences can easily lead to new derivatives. Lately, the focus is on the incorporation of noncanonical amino acids into those NPs to alter their biological activity [109, 110].

### 1.5 Genome mining of 'silent' biosynthetic gene clusters

An important milestone for natural product discovery in bacteria, that influenced the guidelines of research for almost last two decades, was the first whole-genome sequence of *Streptomyces coelicolor* A3(2) in 2002 [36]. By the time, S. *coelicolor* was known to produce actinorhodin, methylenomycin, calcium-dependent antibiotic (CDA), undecylprodigiosin and a grey spore pigment [36, 111–114]. The last of those was reported in 1985. The secondary metabolites, found in bacteria, are produced by organized groups of genes known as biosynthetic gene clusters (BGCs). One BGC produces one or several similar compounds that structurally vary and have therefore different strengths of biological activity. The *S. coelicolor* genome sequence revealed further 18 BGCs [36] that were termed 'silent' or cryptic gene clusters because they were not connected to any known NP at the time of sequencing. In selected production conditions, the corresponding NPs might be produced in very small amounts or not at all for various reasons: the building blocks are not provided or available in sufficient amounts, the transcription could be repressed by other produced secondary metabolites, the produced NP is required in small amounts (e.g. siderophores) or the bacterium in certain conditions does not require the secondary metabolite (e.g. spore pigment in liquid culture). Since the development of the antiSMASH online tool in 2011 [115], last updated in 2019 [116], potential BGCs could be automatically identified and analyzed in any strain with available whole or partial genome sequence. The facilitated recognition of BGCs, indicating the potential for the production of unknown metabolites, is a stepping stone for genome mining. *In silico* analysis of selected BGC can be very thorough due to existing bioinformatic tools. The NRPS precursors and modular PKS assembly lines can be predicted by NRPSpredictor2 [117] or the Stachelhaus code [72], by domain organization and their predicted substrate specificity [118], respectively, with both algorithms being integrated into antiSMASH [21]. Building block predictions and their potential assembly gives us useful information but the structure prediction is still not feasible

in many cases. Moreover, predictions might be more reliable analyzing a genome of *Streptomycetes* than of myxobacterium since many more NPs are known for the former. The antiSMASH can also recognize the precursor peptides of ribosomally synthetized and post-translationally modified peptide (RiPP) NPs but cannot predict their structure after the post-translational modifications took place [116]. Other bioinformatic tools for RiPPS and bacteriocins are available, namely RiPPMiner [119] and BAGEL4 [120]. The targeted genome modification, such as overexpression of the biosynthetic genes of cryptic BGC or its inactivation, can result in the production of a novel NP or its abolishment, respectively. Unfortunately, not every genetic modification of cryptic BGC results in changes in the secondary metabolome. Such BGC could be unfunctional if it was of no benefit to a bacterium, so the active site mutated to spare building blocks for other secondary metabolites. The increasing number of bacterial genome sequences led to the high number of cryptic BGC present in public databases but with no certainty on their functionality. The BGC could also be only recently acquired from the environment, perhaps only partially, and might not have any benefit to a bacterium. Furthermore, not every BGC in bacteria encodes a secondary metabolite with biological activity since bacteria utilize also other molecules like iron chelating agents siderophores [121] for daily survival or produces a spore pigment [122]. To direct the search for NPs with biological activity, genomes can be screened for existence of resistance genes in BGCs. The identified resistance gene would also hint on the mode of action of a potential secondary product. The cases of resistance genes located within the BGC are known like *griR* resistance gene from griselimycin gene cluster, encoding an additional copy of the DNA polymerase III beta subunit [123] or a putative self-resistance protein CysO found in cystobactamid gene cluster belonging to the pentapeptide repeat protein family, which suggests that the cystobactamids act as inhibitors of bacterial type IIa topoisomerases [55]. Prediction tools such as PRISM3 [124] and the Antibiotic Resistant Target Seeker (ARTS) [125] are available online but have not yet delivered anything new. Notably, the focus of research is currently the discovery of novel scaffolds of antibiotics [21] and if very unusual, they might not be possible to detect using bioinformatic tools. On the other hand, lots of sequencing data was acquired in the last years but not all the knowledge is stored in one online database. Rediscovery of known NPs became a real issue in NP research and is therefore important to use dereplication tools to avoid it [21]. Once the BGC of interest is selected, the native promoter upstream of the biosynthetic operon is exchanged with a constitutive or an inducible promoter if the strain is genetically manipulable [126]. The overexpression of the functional biosynthetic genes causes increased production yields of a corresponding NP. If the secondary metabolite was produced

in the wild-type strain all along in detectable amounts, the inactivation of the biosynthetic gene would interrupt its biosynthesis. The LC-MS spectra of the crude extracts from parallel cultivation of wild-type and mutant strain are compared and screened for differences of their secondary metabolomes. The targeted "activation" and inactivation of the biosynthetic genes approach can be used in any genetically manipulable bacteria. However, unlike myxobacteria [127–129], actinobacterial BGCs besides the biosynthetic genes harbor also genes involved in the regulation of the encoded NP. A constitutive overexpression of a LAL family regulator that acts as an activator in *S. ambofaciens* enabled the discovery of stambomycins [130]. Similarly, the knock-out of negative regulator genes, e.g. *tetR* repressor in *Streptomyces* sp. PGA64 and *gbnR* repressor in *S. venezuelae*, resulted in novel angucyclinone metabolite UVM6 [131] and discovery of gaburedin A [132], respectively. Unfortunately, some bacterial species are not genetically manipulable. In such a bacterium the correlation of the BGC to the encoded natural product is only possible via heterologous expression, which is a time-consuming process with no guarantee for successful expression of the encoded NP. Cryptic BGCs can be heterologously expressed in host strain such as *Myxococcus xanthus* DK1622 or *Streptomyces* host strains, some with minimalized genomes for increased genetic stability of unknown BGCs [133, 134].

### 1.6 Outline of the present work

The work for the thesis was done in a research group focusing on the discovery, characterization and improvement of microbial natural products. Genetic engineering was used for the derivatization of known antibiotics fidaxomicin and chelocardin to prepare desired shunt products and partially characterize the biosynthetic pathway, respectively. Genome mining approach combined with genetic engineering was used for the identification of biosynthetic gene cluster of globomycin, and discovery of two novel natural products myxopentacin and myxoglucamide from a novel myxobacterial isolate. The biosyntheses of those were proposed based on the analysis of their biosynthetic genes and catalytic domains combined with the chemical structure.

The predatory myxobacteria are an established source of natural products, often exhibiting a range of biological activities [50, 51]. Furthermore, novel myxobacterial strains isolated from soil or marine samples in this lab offer even higher possibilities to uncover interesting and new natural products [23]. In chapter 2, titled "Connecting biosynthetic gene clusters to their products: *Cystobacterineae* strain MCy9003", the genome mining of a novel myxobacterium, isolated from a soil sample collected in the Philippines, a representative of a new genus, is described. Genome mining combined with the genetic engineering resulted in the discovery of

two novel myxobacterial NPs, myxopentacin and myxoglucamide. The genome sequence analysis using antiSMASH tool [115] revealed 45 biosynthetic gene clusters and potential for 39 new natural products. The overexpression of selected biosynthetic genes in a few cases resulted in changes of secondary metabolome of those mutant strains, detected via LC-MS profiling, when compared to the wild-type strain. The chemical structures of myxopentacins as well as myxoglucamides contain rare building blocks. Moreover, their biosyntheses are of special interest. The nonribosomal peptides myxopentacins contain rare cispentacin and pseudoarginine moieties [135, 136] but their assembly, e.g. polymerization of cispentacin moieties as observed in the chemical structure, raises many questions that remain unanswered so far. The myxoglucamides, glycolipopeptides of hybrid NRPS/PKS type I origin, contain an unusual α-vinyl moiety. Moreover, their biosynthetic route comprises many oxidation domains not yet correlated to the specific biosynthetic or tailoring step and many different types of C domains [137]. The myxoglucamides exhibit moderate cytotoxicity against the HCT 116 cell line.

Chapter 3 focuses on the known antibiotic globomycin found in a few bacterial species, including *Streptomyces hagronensis* strain 360 [138]. Globomycin is a cyclic depsipeptide that acts specifically against Gram-negative bacteria by inhibition of a signal peptidase II LspA, involved in lipoprotein PTMs in many pathogenic bacteria [139]. Although LspA is not essential for viability of some bacteria such as methicillin-resistant *Staphylococcus aureus* (MRSA), it is vital for survival of that bacterium in human blood [140]. The globomycin binding to the LspA from MRSA was proven with a complex structure [140]. However, nothing was known about the biosynthesis of globomycin. In this work, the genome sequence of *S. hagronensis* strain 360 was obtained and the globomycin BGC was identified. The disruption of the main biosynthetic gene by a plasmid insertion via homologous recombination abolished the antibiotic production. The biosynthetic protein domains were analyzed and the biosynthetic pathway was proposed. Globomycin scaffold is biosynthesized mainly by a large NRPS but requires also an additional in trans-acting NRPS module and a TE type II.

Fidaxomicin is a powerful antibiotic with broad biological activity against Gram-positive bacteria that acts through the inhibition of bacterial RNAP [141]. It was approved by FDA as a drug for *Clostridium difficile* associated infection (CDI) [142] since it displays bactericidal antibiotic properties, while so far used vancomycin acts bacteriostatically [141]. After FDA approval, it was further developed against vancomycin-resistant *Enterococci* and as methicillin-resistant *Staphylococcus aureus* cure [141] but has potential also as an anti-tuberculosis [143]

and an anticancer agent [144]. Fidaxomicin is an 18-membered macrocyclic glycoside that contains an unsaturated macrolide core scaffold, modified D-noviose and D-rhamnose moieties, and a dichorinated homo-orsellinic acid moiety [80]. Moreover, an actinomycete marine natural product mangrolide A contains an almost identical 18-membered polyketide ring structure as found in fidaxomicin but was claimed to have selectivity for Gram-negative bacteria like *Acinetobacter baumanii* and *E. coli*. Mangrolide A lacks any modification on the primary alcohol at C-20 but has a disaccharide attached at C-11 [145]. The moieties attached to fidaxomicin aglycone potentially influence the biological activity. The fidaxomicin aglycone that could be further modified semisynthetically as desired, could be achieved either by a total synthesis [146] or by a gene knock-out of the glycosylation enzymes in its biosynthetic gene cluster [80]. The biosynthetic genes and pathway of fidaxomicin were known to a major extent already in 2010. The deletion of genes encoding glycosyltransferases that directly attach sugar moieties onto the macrocyclic scaffold was known to leave us with certain shunt products [80]. The project is based on collaboration with Prof. Gademann's research group from ETH Zürich, specialized in semisynthesis. The preparation of the glycosyltransferase deletion mutants, in our case using *Actinoplanes deccanensis* ATCC 21983 strain, is described in chapter 4.

Moreover, the biosynthesis of an atypical tetracycline chelocardin (CHD) with the focus on cyclization/aromatization of the first two rings, cyclization of the second and fourth ring and methylation of C6 position was studied in chapter 5. Chelocardin's biosynthesis is under scrutiny due to its broad-spectrum antibiotic activity [32] and promising results in a small phase II clinical study for patients with urinary tract infections caused by Gram-negative pathogens in 1977 [147]. The CHD biosynthesis was compared with well-studied oxytetracycline biosynthesis [148] and their common points and differences were described. Sequencing of *Amycolatopsis sulphurea*'s genome resulted in a complete CHD gene cluster and enabled correcting the proposed biosynthetic pathway of CHD reported previously [32]. Gene knock-out mutants and mutants with point mutations in putative aromatases/cyclases, cyclases and C6-methyltransferase were generated. Their secondary metabolome was analyzed and screened for potential shunt products. Mutant strains were complemented with putative complementary protein homologs from oxytetracycline biosynthesis [101]. The enzymes necessary for cyclization and aromatization of CHD scaffold were assigned.

# 2 Connecting BGCs to their products: *Cystobacterineae* sp. MCy9003

## 2.1 Introduction

The search for new antibiotics can be approached in a number of ways. The advantage is definitely in finding a niche of microorganisms that differentiates itself from all the known ones and is able to produce NPs. As *Streptomyces* and other actinobacteria are being worked on for almost 80 years, the chance for finding a new secondary metabolite seems to be getting smaller and the approaches used have to be more innovative. The work in this research group mostly focuses on myxobacteria [149], belonging among *δ-proteobacteria* distributed in 3 suborders, 6 families and 21 genera [43]. The efforts to find a new antibiotic are diverse but they are all based on newly isolated myxobacterial species being cultivated in for myxobacteria optimized media (Figure 15) [23].

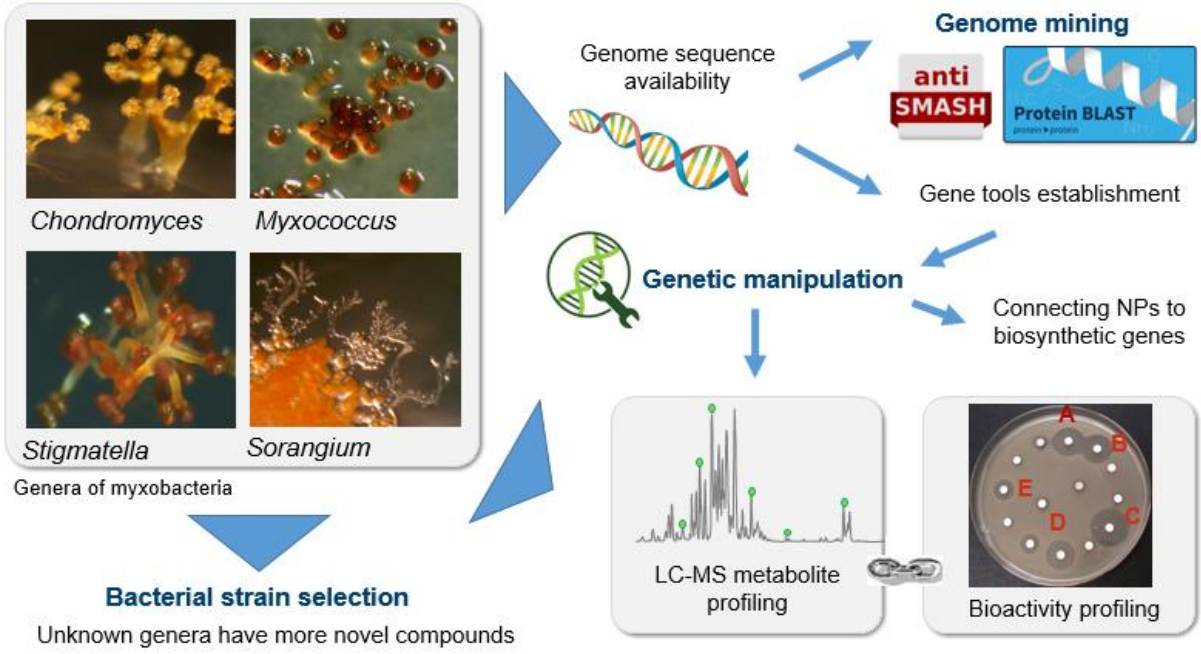



**Figure 15. Novel myxobacterial isolates offer the possibility for discovery of novel NPs.** Various approaches in the discovery of novel NPs are available. The LC-MS metabolite profiling together with bioactivity screening is the basis for the bioactivity-guided approach. Genome sequence availability opens a window to many more opportunities regarding genome mining and reviving of "silent" BGC. Gene tools establishment allows genetic manipulation, providing a different outlook in the bacterial secondary metabolome and correlating NPs with its biosynthetic machinery.

A bioactivity-guided approach consists of crude extract screening for biological activities. If promising biological activity is observed, the crude extract is further fractionated and fractions

tested against a standard library of microorganisms to narrow down the compound responsible for the bioactivity, which is represented in a LC/MS chromatogram by a peak with a specific molecular mass. The in-house database of mainly myxobacterial secondary metabolites, each with identified exact mass and retention time, ensures that known NPs are readily detected. Such dereplication approach is essential in NP discovery to prevent re-isolation and re-characterization of known bioactive compounds [150]. Dereplication or discharge of known compounds comprises analytic techniques, databases and combined procedures from various scientific fields to accurately and efficiently determine the novelty status of a certain NP [150]. Methods such as genomics (e.g. antiSMASH, BLAST), metabolomics (LC/DAD, LC/MS, NMR, LC/NMR) highly depend on the amount of information in used database [151]. On the other hand, dereplication strategy based on MS/MS data involves molecular networking [152–154] and the comparison of MS/MS spectra results in clustered networks of structurally related compounds [155, 156]. The selected compound that passes the dereplication step, is either novel or a derivative of known NP. The isolation, purification and structure elucidation by NMR follow. Regardless of the method used, the characterized secondary metabolite is still to be connected with the biosynthetic genes clustered in the bacterial genome. Certain features, e.g. halogenases, are easier to find but if unusual moieties are present in the NP, the corresponding genes are harder to recognize. A confirmation of BGC being responsible for the biosynthesis of certain secondary metabolites is achieved either by inactivation of corresponding genes in the native producer strain or by heterologous expression of BGC in the host strain. Firstly, the strategy applied depends on the microorganism or more precisely, on the ability of a microorganism to accept foreign DNA by methods used in the laboratory. Secondly, the availability of selection markers for genetic manipulations can be a huge limitation, especially observed in bacterial strains bearing all up to 60 BGCs with some of them certainly containing resistance genes protecting bacterium against its own products. However, the NP discovery can be approached from the other direction. Genome mining became an important and indispensable approach since genome sequencing is affordable and lots of genome data available in the databases. Among a few others, the myxobacterial strain MCy9003 was selected for genome mining in this study. The strain, isolated from a soil sample from the Philippines, belongs to a new genus within the suborder of *Cystobacterinae*. As NPs were found to be genera-specific, a strain from a novel genus offers high chances for discovery of unique NPs [23]. The genome of MCy9003 strain has 11.3 Mb, 71.2 % GC content and contains 45 BGCs according to antiSMASH and manual analysis, of which 8 are of NRPS origin, 1 PKS type I, 1 PKS type II, 1 PKS type III, 12 hybrid NRPS-PKS type I BGCs and the remaining 22 are two

siderophore BGCs, lanthipeptides, bacteriocins, terpenes and two BGCs with some NRPS or PKS features. Among all those BGCs, only half of them, according to publicly available genome database, can also be found in other myxobacteria.

In the initial bioactivity screening weak *E. coli* TolC and *B. subtilis*, *S. aureus* and *M. luteus* activities were observed that later after re-cultivation disappeared. Bacteria produce NPs under various circumstances in response to the stimuli from the environment. Those conditions are hard to reproduce in the laboratory therefore the metabolome changes are inevitable [157]. Even though the desired bioactivity was not observed again, the representative of a new genus with 45 detected BGCs implied on the potential for the discovery of novel NPs. Further analysis of the genome revealed some potential self-resistance genes within the BGCs that could protect the producer bacterium against the biosynthesized NPs [125]. Moreover, it was previously shown that the upregulation of "silent" biosynthetic genes via genome modification sometimes results in changes of the secondary metabolome [158]. Similarly, the biosynthetic gene inactivation proved successful in the past [159]. In this chapter, the information obtained using genome mining approach was used for further genetic modification and subsequent identification of new NPs in MCy9003 strain.

## 2.2 Materials and methods

### 2.2.1 Cultivation of MCy9003 and other myxobacteria used in this study

Wild-type strain MCy9003 and mutant strains were cultivated in autoclaved CYHv3 medium (0.2 % Bacto™ soytone, 0.3 % Bacto™ casitone, 0.2 % glucose, 0.3 % soluble starch, 0.15 % Bacto™ yeast extract, 0.05 % $CaCl_2 \times 2H_2O$, 0.1 % $MgSO_4 \times 7H_2O$, 50 mM HEPES, 8 mg $L^{-1}$ Fe-EDTA, $dH_2O$; pH 7.2), the autoclaved CYS medium (0.25 % Bacto™ casitone, 0.1 % Bacto™ yeast extract, 0.1 % $CaCl_2 \times 2H_2O$, 0.25 % soluble starch, 10 mM HEPES, $dH_2O$; pH 7.0) was used as production medium. For testing the production of myxopentacins in alternative producers, MCy10644 and MCy11578, M7/s4 medium (0.5 % soy flour, 0.5 % corn starch, 0.2 % glucose, 0.1 % yeast extract, 0.1% $MgSO_4 \times 7H_2O$, 0.1 % $CaCl_2 \times 2\ H_2O$, 1 % HEPES, with final pH 7.4 and supplemented with 0.1 mg $L^{-1}$ of vitamin B12 and 5 mg $L^{-1}$ of $FeCl_3$ after autoclaving) was used as vegetative and production medium. For GNPS analysis, the strains of interest were cultivated in RG5 production medium (0.05 % soya peptone (Roth), 0.05 % Bacto™ soytone, 0.2 % soya flour (Hensel), 0.1 % corn steep solids (Sigma), 0.05 % Bacto™ yeast extract, 0.8 % soluble starch (Roth), 0.5 % baker's yeast, 0.2 % gluten from wheat (Sigma), 0.1 % $MgSO_4 \times 7H_2O$, 0.1 % $CaCl_2 \times 2H_2O$, 25 mM HEPES, $dH_2O$; pH 7.2 adjusted

with KOH). Strains in liquid cultures were cultivated on an orbital shaker (30 °C, 180 rpm) for 1-2 days in CYHv3 medium (depending on the cell density), inoculated (2 %) into CYS or RG5 production media and were cultivated as 50 mL cultures in 300 mL baffled Erlenmeyer flasks for 10-12 days. For large-scale cultivations, 10 % inoculum was added to 1.5 L production medium in 5-L Erlenmeyer flask. The 2 % XAD-16 resin is added into a culture broth after completion of production. For genetic modification purposes, the susceptibility of MCy9003 strain to different antibiotics was determined. The bacterium is resistant to hygromycin (100 μg $ml^{-1}$) and sensitive to kanamycin (75 μg $ml^{-1}$) and oxytetracycline (10 μg $ml^{-1}$). Kanamycin (75 μg $ml^{-1}$) antibiotic was added into solid or liquid media for selection of transformants and cultivation of mutant strains, respectively. If required, the bacterial cultures were induced with 1 mM vanillate at the beginning of the cultivation.

### 2.2.2 Construction of plasmids for activation or inactivation of selected biosynthetic genes

Routine handling of nucleic acids, such as isolation of plasmid DNA, restriction endonuclease digestions, DNA ligations, and other DNA manipulations, was performed according to standard protocols [160]. *E. coli* DH10β (Invitrogen) was used as the host for standard cloning experiments. *E. coli* strains were cultured in 2TY medium or on 2TY agar (1.6 % tryptone, 1 % yeast extract, 0.5 % NaCl, (1.8 % agar), $dH_2O$) at 30-37 °C, at 200 rpm overnight. Antibiotics were used at the following final concentrations: 100 μg $ml^{-1}$ ampicilin, 50 μg $ml^{-1}$ kanamycin. Transformation of *E. coli* strains was achieved via electroporation in 0.1 cm-wide cuvettes at 1350 V, 200 Ω, 25 μF. Plasmid DNA was purified by standard alkaline lysis [160] or by using the GeneJet Plasmid Miniprep Kit (Thermo Fisher Scientific). Restriction endonucleases, alkaline phosphatase (FastAP) and T4 DNA ligase were obtained from Thermo Fisher Scientific. Oligonucleotides used for PCR and sequencing were obtained from Sigma-Aldrich and are listed in Table S1- 1. PCR reactions were carried out in Mastercycler® pro (Eppendorf) using Phusion™ High-Fidelity or Taq DNA polymerase (Thermo Fisher Scientific) according to the manufacturer's protocol. For Taq: Initial denaturation (2 min, 95 °C); 30 cycles of denaturation (30 s, 95 °C), annealing (30 s, -5 °C of lower primer's $T_m$) and elongation (based on PCR product length 1 kb/min, 72 °C); and final extension (10 min, 72 °C). For Phusion™: Initial denaturation (2 min, 98 °C); 30 cycles of denaturation (15 s, 98 °C), annealing (30 s, $T_m$ of the lower primer) and elongation (based on PCR product length 0.5 kb/min, 72 °C); and final extension (10 min, 72 °C). PCR products or DNA fragments from restriction digests were purified by agarose gel electrophoresis and isolated using NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel). The digested PCR products were cloned into suitable plasmids

pSBtn5Kan, pSB2tn5Kan and pFPVan, verified by restriction analysis and sequenced using sequencing primers (Table S1- 1). The PCR products cloned into pCR2.1 vector were adenylated prior to use of the TOPO TA Cloning Kit (Invitrogen). Details on the construction of all plasmids used and generated in this study are given in Table S1- 2. Bacterial strains prepared in this study are presented in Table S1- 3.

### 2.2.3 Transformation of MCy9003

The MCy9003 wild-type strain was cultivated in CYHv3 medium for 2-3 days. 2 mL of bacterial culture were transferred in 2 mL Eppendorf tube, depending on the density of the culture, centrifuged at 6000 rpm for 2 min, washed twice with autoclaved ultrapure water and resuspended in 30 µL of it. So prepared cells were added 1 µg of constructed plasmid, mixed and transferred to a cuvette with 1-mm gap. The transformation mixture was electroporated using the following parameters: 650 V, 25 µF, 400 Ω [161]. Transformed cells were resuspended in 1 mL of CYHv3 medium and incubated in 2 mL Eppendorf tube with punctured lid at 30 °C, shaking at 800 U $min^{-1}$ for 6 h. The CYHv3 agar plates containing kanamycin (75 µg $ml^{-1}$) were overlayed with 3 mL of CYHv3 soft agar (0.75 %) added incubated electroporated cells and kanamycin to the final concentration of 75 µg $ml^{-1}$. After incubation at 30 °C for 8-10 days, successfully transformed colonies were observed and re-plated onto a new agar plate.

### 2.2.4 Verification of transformants and their preservation

After 3-4 days of incubation on agar plates at 30 °C, some cells were collected for isolation of genomic DNA with Gentra® Puregene® Cell Kit (Qiagen) for further verification of correct plasmid integration. For activation mutants three primer sets were used. Integration on the left and right sides of the pSBtn5Kan-based plasmid were confirmed by a successful DNA amplification with one primer on the genome, on either left (g001-bgc#-F) or right (g001-bgc#-R) side of the homology region, and the other on the plasmid, M13-24R or M13-21F, respectively. The integration of the pFPVan-based plasmid was confirmed similarly, however, primer vanseq-F was used instead of M13-21F when confirming the right side of the homology. If the plasmid was integrated correctly, no PCR product was expected using primers aligning on the genome just outside the homology region (g001-bgc#-F and g001-bgc#-R) (Table S1-1). Two sets of primers were used on inactivation mutants for pCR2.1-based plasmid integration verification. The homology itself was PCR-amplified as a PCR control and the integration into desired gene by G001-inact#o# primers, located outside of homology region, in combination with PCR21R primer aligning on the plasmid (Table S1- 1). For storage purposes, the

transformant with correct PCR profile was grown in 50 mL liquid medium CYHv3 with kanamycin (75 μg $ml^{-1}$) and centrifuged at 8000 rpm for 6 min. The cell pellet was resuspended in 20 % glycerol and stored at -80 °C.

### 2.2.5 Extraction and LC-MS analysis of myxobacterial crude extracts

After completion of production, culture broth with 2 % XAD-16 is transferred into 50 mL plastic falcons and centrifuged for 10 min at 8000 rpm. The supernatant is discarded but the pellet is frozen at -20 °C or in liquid nitrogen, and lyophilized. For extraction of 50 mL falcon tube content a mixture of 50 mL MeOH/acetone 1:1 is used and incubated on a magnetic stirrer for 3 h. Extracts are further filtered through glass wool into round bottom flasks. The solvent is evaporated on a rotary evaporator (Rotavapor R210, Büchi) at 40 °C at reduced pressure. Each sample is dissolved in 2 mL methanol and measured in 20x dilution on UHPLC coupled to maXis 4G MS system (Bruker Daltonics; Q-ToF). Separation of 1 μl sample, with 1.7 μm Acquity UPLC BEH C-18 Column (100 x 2.1 mm), was achieved by a linear gradient from (A) $H_2O$ + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 0.600 ml/min at 45 °C. The gradient was initiated by a 0.5 min isocratic step at 5 % B, followed by an increase to 95 % B in 18.5 min to end up with a 2 min step at 95 % B before re-equilibration to initial conditions in 2 min. The detection was usually performed in the positive MS mode, where mass detection range reached from 150 to 2.500 m/z units for maXis 4G MS system.

## 2.3 Results and discussion

### 2.3.1 Known BGCs from MCy9003

Out of 40 BGCs predicted by antiSMASH, and five found subsequently since its algorithm cannot separate the BGCs located next to each other, some of them are known or characteristic for myxobacteria. The strain encodes enzymes for biosynthesis of terpenes geosmin [162] and carotenoids [163], a PKS type III product alkylpyrone [164], a siderophore myxochelin [165] and fatty acids VEPE/TG-1 [166], all detected by antiSMASH online tool. After successful activation of a certain BGC, we detected NRPS product chloromyxamide [167] that was known at the time but not yet present in databases.

### 2.3.2 Genome mining strategy

The selection of BGCs was first based on the presence of known resistance genes in the vicinity of predicted biosynthetic genes (Figure 16). Firstly, a pentapeptide repeat protein [168] was found encoded in geosmin BGC. Geosmin is known for its odorous and volatile properties. It

is reported as cytotoxic to human cells HepG2 in concentrations above 75 µl $ml^{-1}$ since it causes cytoplasmic membrane damage but it does not induce DNA damage and complex genomic alterations [169]. Geosmin is therefore not a DNA gyrase inhibitor as pentapetide repeat protein would suggest. Secondly, we observed the gene encoding an elongation factor Tu (MCy9003_22480) and elongation factor Tu GTP binding domain (MCy9003_82640) in the vicinity of BGC-12 and BGC-39B, respectively, which could or could not be connected with the biosynthesis of NP. The BGC-12 consisted of a NPRS gene containing four modules predicted to be producing a short peptide with A domain specificity roughly for Leu, Val, Pro and Ile amino acids. The BGC-39B is a PKS type I/NRPS hybrid, which according to antiSMASH prediction consists of 7 modules, starting with CAL (Co-enzyme A ligase) domain, condensing a C-methylated methylmalonyl-CoA moiety, followed by Ser, Gly, Ser, Gln, Ile, Lys amino acids predicted by the Stachelhaus code [72]. The elongation factor Tu (EF-Tu) is responsible for delivering aminoacyl-tRNA molecules to mRNA-programmed ribosomes and plays an essential role in the elongation cycle of protein biosynthesis [170]. EF-Tu as a part of ternary complex EF-Tu·GTP·aa-tRNA binds to the ribosome and allows interaction between aminoacyl-tRNA molecule and mRNA codon in the ribosomal A site. Binding of a selected aa-tRNA leads to GTP hydrolysis, where an inorganic phosphate $P_i$ dissociates and EF-Tu undergoes conformational changes [170]. The antibiotics that target EF-Tu consequently interfere with the protein biosynthesis, either by blocking EF-Tu · GDP on the ribosome or inhibition of the interaction of EF-Tu · GTP with aa-tRNA [171]. The examples are so-called elfamycins [172]: kirromycin [173], enacyloxin [174], pulvomycin [175], GE2270 A [175] and factumycin [176]. Three copies of EF-Tu are encoded in the genomes of the kirromycin producer *S. ramocissimus* as well as the GE2270 producer *Planobiospora rosea*. In each strain, one of the copies was shown to provide resistance to kirromycin, pulvomycin, GE2270 [172] or GE2270 [177], respectively. However, the EF-Tu inhibitor-producer bacteria were mostly found to protect themselves in a different manner, e.g. with precursor secretion (enacyloxin) or by export through ABC transporter (factumycin) [172], rather than with an additional copy of EF-Tu.

Although only a few potential resistance genes were detected, MCy9003 strain with many unknown BGCs of NRPS and NRPS/PKS type I origin still held promise to find something new. Therefore, we later focused on BGCs that contain biosynthetic genes orientated in one direction and are preferably in one operon (Figure 16), due to convenience as well as the limitation of only one selection marker that could be used for genetic modification. We successfully overexpressed biosynthetic genes in BGC-33 and BGC-34B, resulting in the

production of linear peptides (Figure S1- 1) with no notable antibacterial activity tested so far (Table 3). Next, we chose the BGC-1, with 4-5 biosynthetic genes (38.2 kb), containing 9 modules spanning from MCy9003_00740 to MCy9003_00780 gene. According to antiSMASH prediction, the initial CoA ligase (CAL) domain selects a specific fatty acid-CoA building block and continues with the condensation of amino acids D-Gly, Ile, Hpg, N-methylated Thr, Aad, Thr, D-Ser, Hpg and Phe. The Hpg stands for 4-hydroxyphenylglycine and Aad for δ(L-α-aminoadipic acid) [72]. The selected BGC-33 is predicted to assemble a fatty acid-CoA, Gly, D-Gln, Ile, Val, Ser, Val, malonyl-CoA, Val, Val, Glu and Ala on a hybrid NRPS/PKS type I encoded by a 43.9-kb MCy9003_77130 gene with 11 modules. Moreover, a flavin-utilizing monooxygenase is positioned in between the modules incorporating malonyl-CoA and Val units. The BGC-34B contains biosynthetic genes MCy9003_78320 and MCy9003_78330 (29 kb) with 6 modules that are predicted to assemble malonyl-CoA, Phe, Ile, D-Val, D-Leu, Ala, Gly units on a hybrid NPRS/PKS type I.

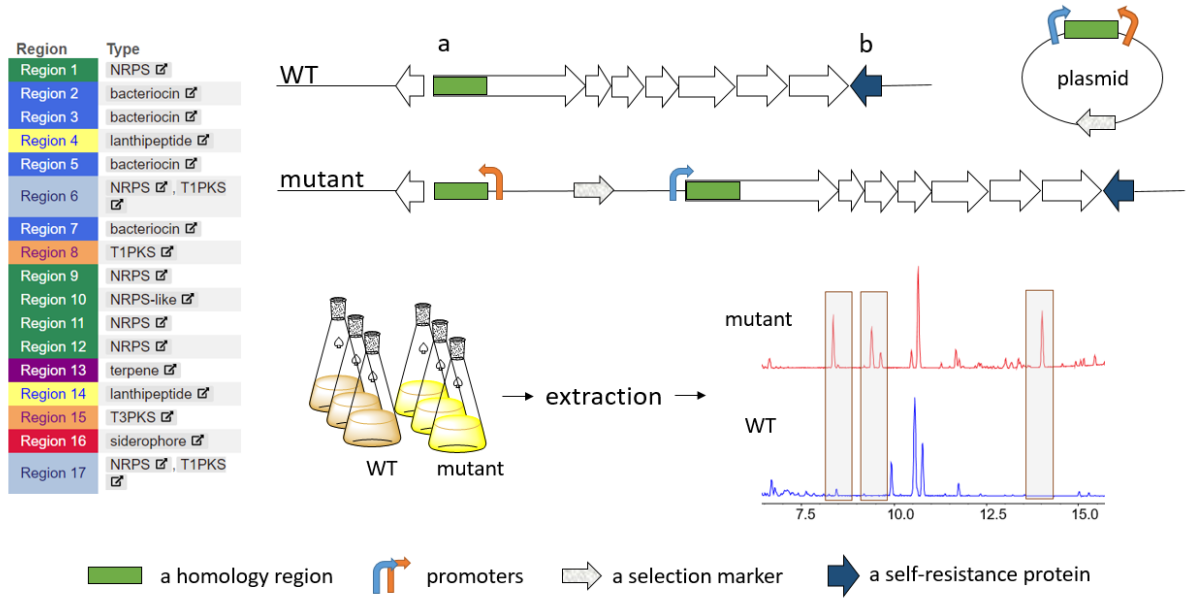



**Figure 16. Genome mining strategy in MCy9003.** The genome sequence of MCy9003 strain was analyzed with antiSMASH and manual inspection. The potential BGCs were at first selected if they contained a potential self-resistance protein (b). Later, the focus was on BGCs encoded in one operon since only one selection marker was available, however, also the bilateral gene overexpression was possible (a). After the discovery of myxopentacin (see below), produced as well by two alternative producers, the BGC was selected only if present in all three genomes. Constructed plasmids were electroporated into WT strain. Mutant was identified and cultivated in parallel with its parent strain. Culture broth was extracted and subjected to LC-MS profiling. Dereplication was made before compound isolation, the structure elucidation and bioactivity testing followed.

The biosynthetic gene MCy9003_80180 (26.3 kb) with 6 modules from BGC-35B encodes a hybrid NRPS/PKS type I predicted to condense fatty acid-CoA, Phe, malonyl-CoA, Gly, Gln,

Val, Val units. The BGC-35B as well contains a tailoring TE type II, and a flavin-utilizing monooxygenase encoded in the biosynthetic gene between the modules incorporating a malonyl-CoA and Gly unit. The selection of further BGCs for activation or inactivation was based on the presence of those in the genomes of myxopentacin alternative producers due to the search of its biosynthetic genes (see below). The selected BGC are shown in Table 3, together with the experimental outcome.

**Table 3. The modified BGCs in MCy9003**

| The selection rationale | BGC# | Genetic modification | Output |
|---|---|---|---|
| Pentapeptide repeat protein | BGC-32 | X | Geosmin |
| Elongation factor Tu GTP binding domain | BGC-39B | inactivation | X |
| Elongation factor Tu | BGC-12 | OE<br>inactivation | Myxopentacins type A: 762, 776a, 776b<br>Changes in myxopentacins' conc. ratio |
| One operon | BGC-33 | OE | Linear peptide |
| One operon | BGC-34B | OE | Linear peptide |
| One operon | BGC-1 | OE<br>inactivation | Myxopentacins type B: 913, 926<br>Changes in myxopentacins' conc. ratio |
| One operon | BGC-35B | OE | Increased production of myxopentacins |
| In both alternative producers | BGC-18 | OE | Myxoglucamides (Section 2.3.7) |

OE – overexpression

### 2.3.3 Discovery of myxopentacins

The inactivation by a plasmid insertion into MCy9003_82470 gene, encoding a NRPS from BGC-39B, showed no significant differences in LC-MS chromatogram analysis and did not imply on the disappearance of a NP-molecular mass potentially produced by BGC-39B. The overexpression of MCy9003_22370 gene (13.8 kb), encoding the main NRPS in BGC-12 with strong constitutive promoter P*tn5*, was followed by comparison of secondary metabolomes of mutant and wild-type strain MCy9003. In the mutant strain, certain target peaks in LC-MS spectrum were visibly increased, especially target peaks 4 and 5 (Figure S1- 2). The dereplication analysis confirmed that compounds are generally unknown, with the exception of compound number 10 that belongs to the family of myxochelins. 10 out of 11 target masses were new and in MS-MS analysis all contained a fragment ion with molecular mass 223.14 m/z (dimer of cispentacin; explained further in the text) and a difference of 111.068 m/z between fragments (Figure 17). The NPs exhibiting that characteristic MS-MS pattern were named myxopentacins. The compounds represented by peaks 4 and 5 in LC-MS chromatogram (Figure S1- 2), with molecular masses 381.7477 $[M+2H]^{2+}$ or 762.4868 $[M+H]^{+}$ and 388.7557 $[M+2H]^{2+}$ or 775.5114 $[M+H]^{+}$, respectively, were isolated, structurally elucidated using NMR

and stereochemically assigned by Alexander Popoff [178]. The isolated compounds were named myxopentacin-762, myxopentacin-776a and b (Figure 18).

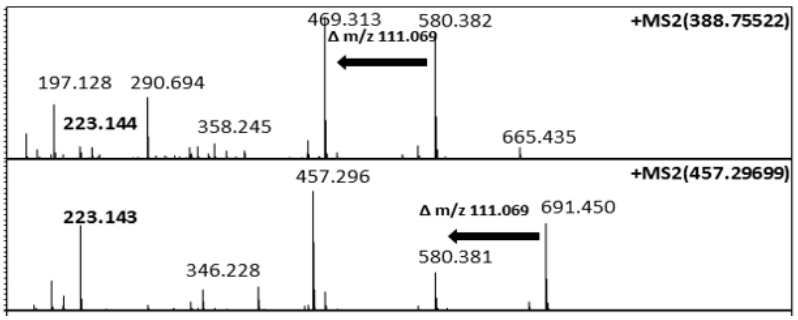
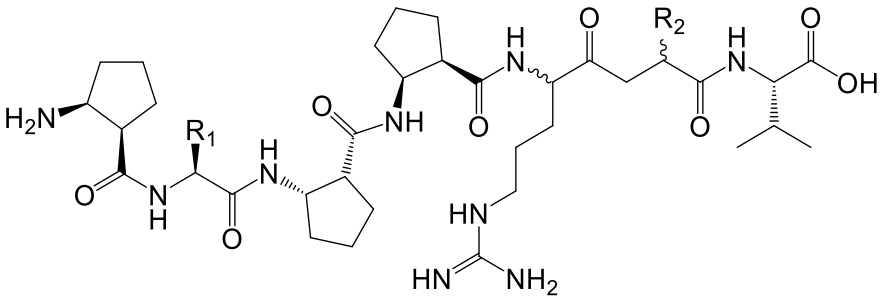



**Figure 17. The myxopentacin-specific MS-MS fragmentation pattern.** The common fragment 223.14 m/z and the common difference between the fragments of 111.069 Da. Spectra originate from the fragmentation of the double charged mass $[M+2H]^{2+}$ ion and were extracted from HRMS measurements of crude extracts of MCy9003 [178]. The figure was taken from the PhD thesis of A. Popoff [178].

The overexpression of MCy9003_00740 gene, located at the beginning of a putative BGC-1, with strong constitutive promoter P*tn5* as well resulted in changes in secondary metabolome of MCy9003 strain (Figure S1- 3). Surprisingly, 7 out of 8 target molecular masses showed the above mentioned characteristic pattern in MS-MS analysis and therefore belong into the compound family of myxopentacins. The compounds with molecular masses 457.29 $[M+2H]^{2+}$ or 913.58 $[M+H]^{+}$ and 464.3 $[M+2H]^{2+}$ or 925.51 $[M+H]^{+}$, respectively, with most abundant concentrations were isolated and their structure elucidated [178], yielding myxopentacin-913 and myxopentacin-926 (Figure 19).



(1) $R_1 = CH_3$ $R_2 = CH_2CH_3$
(2) $R_1 = CH_2CH_3$ $R_2 = CH_2CH_3$
(3) $R_1 = CH_3$ $R_2 = CH_2CH_2CH_3$

**Figure 18. Myxopentacin (1) 762, (2) 776a and (3) 776b (type A) isolated from MCy9003/pSBtn5Kan001nrps12 strain**

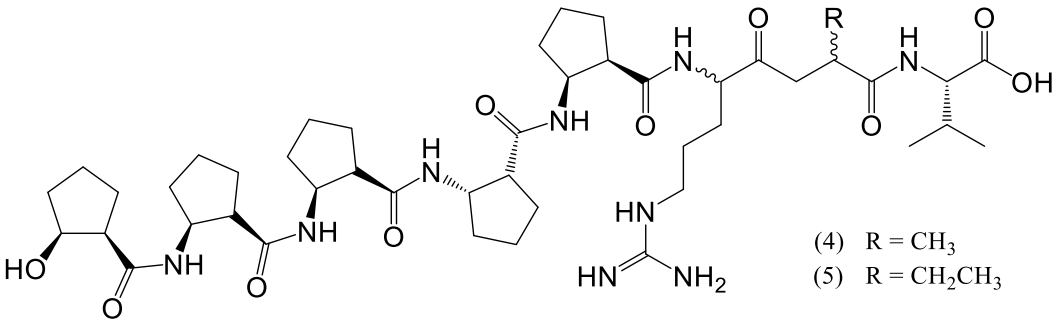



**Figure 19. Myxopentacin (4) 913 and (5) 927 (type B) isolated from MCy9003/pSBtn5Kan001nrps1 strain**

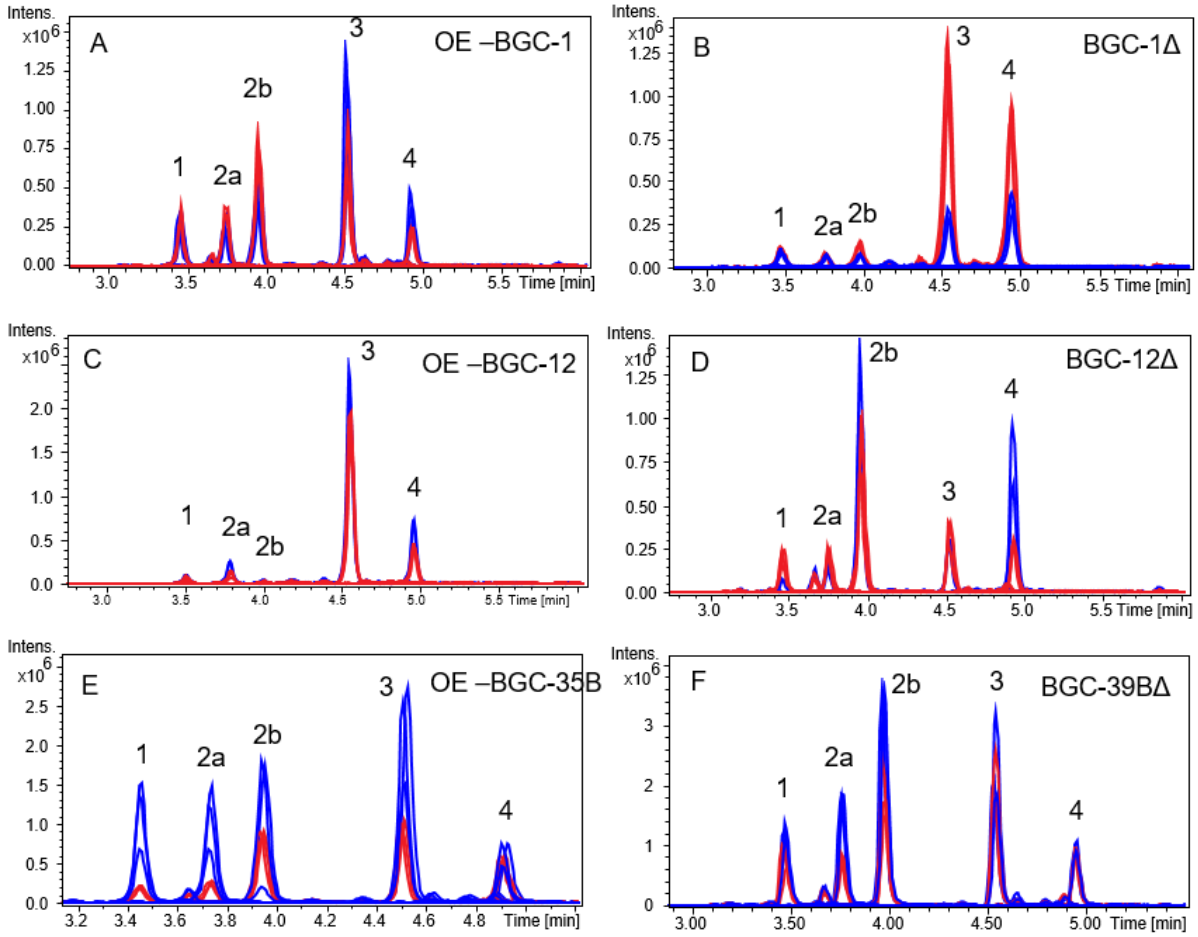



**Figure 20. The secondary metabolome comparisons between WT and mutant strains**. The overexpressions of BGCs 1 (M41) and 12 (M23) are presented in A and C, and the inactivations of BGC 1 (M127), 12 (M37) and 39B (M137) in B, D and F chromatograms, respectively. In case of BGC-35B, the secondary metabolome of mutant strains prior and after the vanillate induction was compared. The EICs of WT and uninduced strain M58 are marked with red, while the EICs of mutants and vanillate-induced strain M58 are blue. The numbers correspond to EICs (m/z) of: 1- 381.7477 (±0.02), 2- 388.7557 (±0.02), 3- 457.29 and 4- 464.3 (±0.02). The information about the mutated strains M23, M37, M41, M58, M127 can be found in Table S1- 3. The data for each chromatogram were obtained in a separate experiment.

Although the building block prediction of BGC-1 and BGC-12 did not fit with isolated compounds (Section 2.3.2), the inactivation mutants of MCy9003_00740 and MCy9003_22370 gene, respectively, were prepared by a plasmid insertion via homologous recombination. The secondary metabolomes of both mutant strains exhibited changes of the concentration ratio of compounds from myxopentacin family, in comparison to the wild-type strain (Figure 20). The myxopentacins were still produced and we so confirmed that those BGCs are not responsible for their assembly. Moreover, the inactivation of the BGC-39B prepared by a plasmid insertion via homologous recombination into MCy9003_82470 gene similarly exhibited changed production yields of myxopentacin derivatives (Figure 20F). In the course of genome mining in MCy9003 strain, many other activation and inactivation mutants were prepared (data not shown). The secondary metabolome of the BGC-35B activation mutant (Figure 20E, Table 3) has shown increased production of all 5 isolated myxopentacins. In general, the inactivation of BGC-1 (Figure 20B) and BGC-39B (Figure 20F) caused the decrease of myxopentacins type B, while the inactivation of the BGC-12 changed the distribution of produced myxopentacin regardless of their type. The BGCs 1, 12, 35B and 39B are somehow involved in yet not understood regulation of their biosynthesis but many more BGCs are present in MCy9003 strain and their potential involvement is still completely unknown.

### 2.3.4 Biosynthesis of myxopentacins

The newly discovered myxopentacin family contains a cyclopentane ring that is a unique feature in the world of NPs. Only three secondary metabolites are known with such feature, namely borrelidin [179], cispentacin [180] and amipurimycin [135, 181]. The angiogenesis inhibitor borrelidin, isolated from *Streptomyces parvulus* Tü4055, contains a *trans*-cyclopentane-1,2-dicarboxylic acid incorporated into the macrolide scaffold as a staring unit [179]. Cispentacin, FR109615 or 2-aminocyclopentane-1-carboxylic acid is a non-proteinogenic amino acid, isolated from the culture broth of *Bacillus cereus* strain, L450-B2 in 1989 [180] and in 1990 from *Streptomyces setonii* [182]. It exhibited weak antifungal activity against *Candida albicans* A9540 *in vitro* but demonstrated strong protection of mice from lethal systemic infection with *C. albicans* [180]. On the other hand, antimycotic amipurimycin [183] was isolated from *Streptomyces novoguineensis* sp. nov in 1977 [184] and has shown remarkable activity against *Pyricularia oryzae*, a fungus causing the rice blast disease [185]. Due to the novelty of myxopentacin structure, the biosynthetic genes were hard to predict but after the biosynthetic pathway of amipurimycin [135] was published, the BGC-39A was an obvious choice. Additionally, the production of myxopentacins was detected in two

myxobacterial alternative producers MCy11578 and MCy10644 with help of the in-house database MyxoBase. All three strains were grown in parallel and their crude extracts were measured with LC-MS and analyzed (Figure S1- 4). The protein homologs responsible for formation of cispentacin moiety in amipurimycin, encoded by *apmA1-apmA7* [135] or *amcB-amcH* [181], were found in the MCy9003 genome using BLAST and named MxpB-MxpH (Figure 23, Table 4). The inactivation mutant MCy9003/*mxpA*Δ (Table S1- 3: M215) further proved the involvement of a NRPS MxpA in the myxopentacin biosynthesis. The alternative producers harbour myxopentacin BGC as well but only the genes encoded in BGC-39A of MCy9003 are presented in Table 4. Analysis of MXP gene cluster suggests that proteins from MxpA to MxpL are necessary for the biosynthesis of myxopentacin's scaffold, except for MxpI, a potential self-resistance gene.

**Table 4. Annotated genes found in the myxopentacin BGC-39A**

| *Gene* | *Size*[a] | *Protein homolog*[b] | *Proposed function* | *Catalytic domains* |
|---|---|---|---|---|
| *orf(-8)* | 123 | WP_047856686.1 (77/92) | hypothetical protein | |
| *orf(-7)* | 219 | WP_002625481.1 (92/99) | hydrolase | |
| *orf(-6)* | 621 | WP_073560195.1 (97/93) | amidohydrolase | |
| *orf(-5)* | 541 | WP_071897078.1 (87/96) | MFS transporter | |
| *orf(-4)* | 75 | WP_169344977.1 (85/97) | DUF1427 family protein | |
| *orf(-3)* | 78 | WP_095989583.1 (92/100) | DUF1427 family protein | |
| *orf(-2)* | 283 | WP_161666456.1 (83/92) | endonuclease | |
| *orf(-1)* | 226 | WP_164017081.1 (82/95) | hypothetical protein | |
| *mxpA* | 2579 | WP_084609656.1 (69/88) | non-ribosomal peptide synthetase | C-A-PCP-C-A-PCP-C |
| *mxpB* | 469 | WP_169343210.1 (84/94), ApmA1, AmcH, Cfa5 | acyl--CoA ligase | A |
| *mxpC* | 194 | WP_043392319.1 (78/92), ApmA2, AmcG, Cfa4 | Unknown, coronafacic acid synthetase | |
| *mxpD* | 374 | WP_164002924.1 (85/93), ApmA3, AmcF, Cfa3 | beta-ketoacyl-ACP synthase | KS |
| *mxpE* | 174 | WP_120205330.1 (79/93), ApmA4, AmcE, Cfa2 | beta-hydroxyacyl-ACP dehydratase | DH |
| *mxpF* | 148 | WP_043392327.1 (75/96), ApmA5, AmcD | 4-hydroxybenzoyl-CoA thioesterase | TE |
| *mxpG* | 465 | WP_073559291.1 (83/95), ApmA6, AmcC | aminotransferase class III-fold pyridoxal phosphate-dependent enzyme | AMT |
| *mxpH* | 89 | WP_043392332.1 (79/97), ApmA7, AmcB, Cfa1 | ACP | ACP |
| *mxpI* | 153 | OJT27517.1 (85/94) | YbaK/prolyl-tRNA synthetase | |
| *mxpJ* | 282 | WP_120625560.1 (77/92), KtmA | aldolase | |
| *mxpK* | 410 | WP_002634949.1 (78/95), KtmB | amino acid C-acyltransferase | AMT |
| *mxpL* | 178 | WP_073559295.1 (86/95), KtmC | dehydratase | DH |

| | | | |
|---|---|---|---|
| *orf1* | 529 | WP_120546462.1 (89/97) | MBL fold metallo-hydrolase |
| *orf2* | 534 | WP_120601942.1 (81/94) | MBL fold metallo-hydrolase |
| *orf3* | 293 | WP_120557531.1 (61/81) | CPBP family intramembrane metalloprotease |
| *orf4* | 272 | WP_120557532.1 (67/86) | aspartyl/asparaginyl beta-hydroxylase domain-containing protein |
| *orf5* | 243 | WP_120619789.1 (68/88) | FkbM family methyltransferase |
| *orf6* | 435 | WP_147439182.1 (60/85) | MFS transporter |
| *orf7* | 303 | WP_120558049.1 (76/91) | hypothetical protein |

*[a] Sizes are given in amino acids; [b] Accesson numbers and percentage of identity/similarity are given in parentheses.*

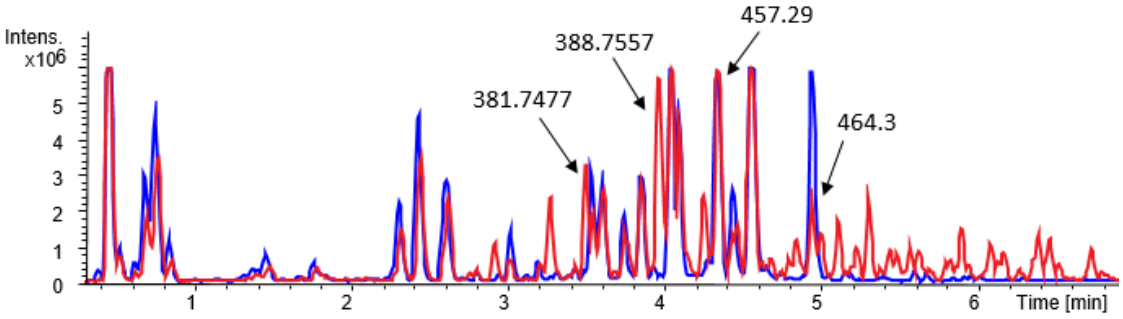



**Figure 21. The comparison of secondary metabolome of MCy9003 wild-type and the mutant strain with disrupted *mxpA* gene from BGC-39A.** Base peak chromatogram (BPC) of WT strain is shown in red, whereas BPC of the MCy9003/*mxpA*Δ mutant strain is shown in blue. The peaks corresponding to the respective exact masses are marked. However, myxopentacin peaks are often overlapping with other compounds therefore it might appear as their production is not changed or is even increased in the mutant strain. The myxopentacin production in both strains is better shown in Figure 22.

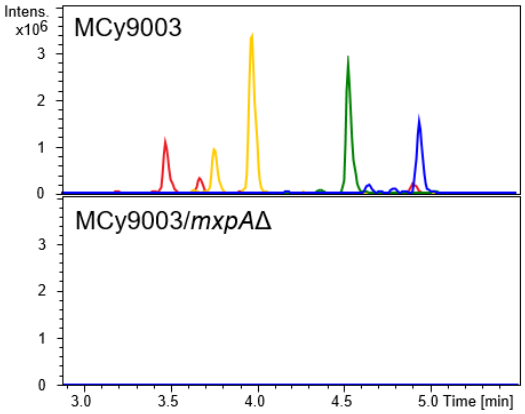



**Figure 22. Myxopentacin production in MCy9003 and MCy9003/*mxpA*Δ strains.** EICs of m/z 381.7477 (±0.02) (red), m/z 388.7557 (±0.02) (yellow), m/z 457.29 (±0.02) (green) and m/z 464.3 (±0.02) (blue) are shown (chromatograms adapted from DataAnalysis (available from Bruker Daltonics)).

The inactivation mutant MCy9003/*mxpA*Δ was prepared by a transformation of a plasmid bearing 1 kb-homology to the first A domain of *mxpA* gene into MCy9003 wild-type strain. The plasmid pCR2.1-001inact39Ao1 (Table S1- 2) was incorporated into the genome through homologous recombination, which disrupted the modular NRPS MxpA. Triplicates of mutant strain and wild type were cultivated in production medium in parallel, their culture broths were extracted and subjected to LC-MS analysis. No myxopentacins were detected in crude extracts of a mutant strain MCy9003/*mxpA*Δ in contrast to the wild-type strain (Figure 21, Figure 22). The transcription of MxpA was therefore successfully interrupted and it can be concluded that this NRPS protein plays an essential role in myxopentacin biosynthesis and is least partly responsible for its assembly as explained in Section 2.3.4.3. A summary of the proposed biosyntheses of cispentacin and pseudoarginine moieties is presented in Figure 23. More detailed information on their biosynthesis and the possible myxopentacin building block assembly strategies are presented below.

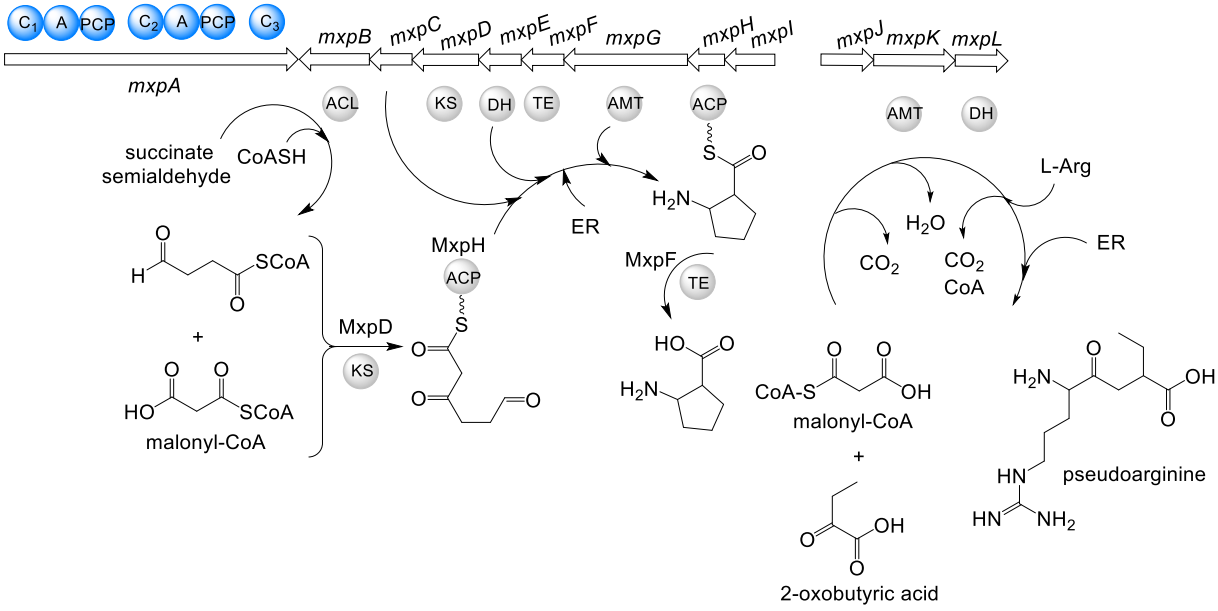



**Figure 23. A summary of the biosynthesis of cispentacin and pseudoarginine moieties.** The assembly of pseudoarginine and valine are proposed in Figure 27, whereas the biosynthesis of cispentacin and pseudoarginine are explained in more detail in Figure 24 and Figure 25, respectively. Two ER domains, necessary for the biosynthesis of the building blocks, are not located in the myxopentacin BGC.

### 2.3.4.1 The biosynthesis of cispentacin moiety

The proposal for the biosynthesis of cispentacin moiety (Figure 24) is based on the *in silico* protein analysis [135, 181, 186]. Recent publications [135, 181] propose cispentacin

biosynthesis on the basis of comparison to coronafacic acid (CFA), a polyketide component of the phytotoxin coronatine, a virulence factor of the plant pathogen *Pseudomonas syringae* [186]. The heterologous expression and gene disruption experiments in amipurimycin BGC confirmed the involvement of a cluster of genes in the biosynthesis of cispentacin moiety [135, 181]. However, no *in vitro* function of any protein hypothetically involved was shown. According to the hypothesis by Romo *et al*., *amcA-amcH* genes encoded in amipurimycin BGC are proposed for its biosynthesis [181] (Figure 23). The necessary starting components are suggested to originate from primary metabolism.

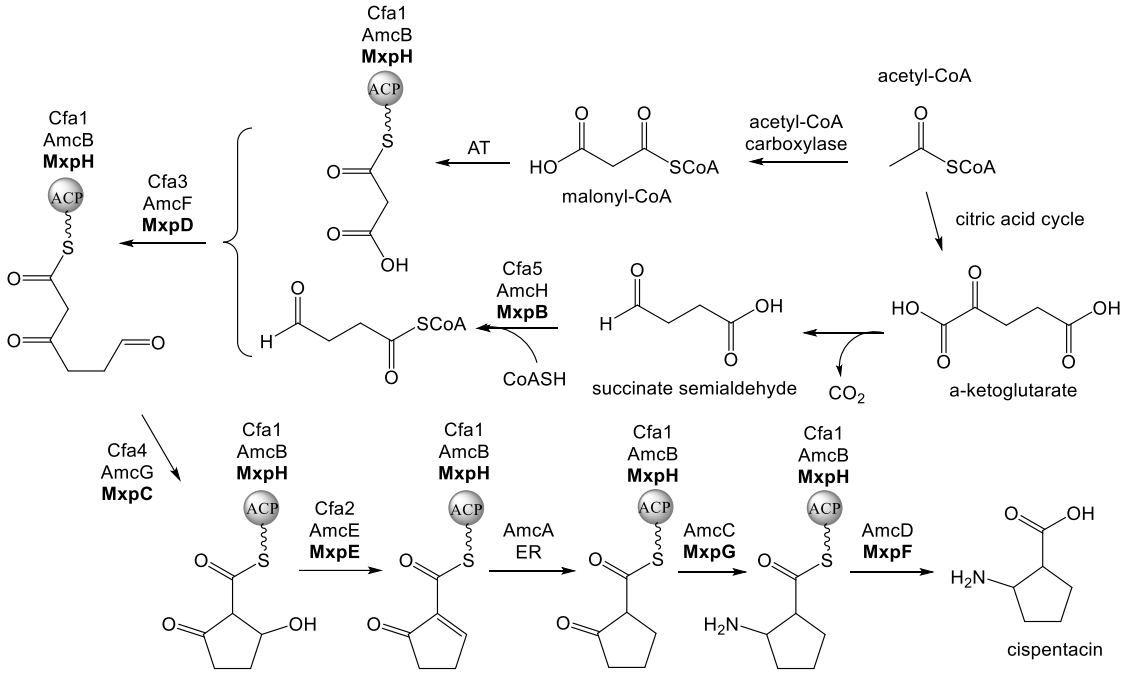



**Figure 24. The suggested biosynthesis of the cispentacin moiety in myxopentacins according to the hypothesis by Romo *et al*..** Building blocks are derived from primary metabolism, from acetyl-CoA [135, 181]. An ACP MxpH loads a malonyl-CoA moiety. The CoA activated succinate semialdehyde is condensed onto malonylated-ACP by MxpD, a KS protein. A hypothetical protein MxpC cyclizes an acyl chain into a cyclopentane ring. A dehydratase MxpE removes a hydroxyl group and an unknown ER domain reduces the newly formed double bond. An aminotransferase MxpG installs amino group and a TE MxpF releases formed cispentacin moiety from MxpH ACP.

A versatile acetyl-CoA is transformed by an acetyl-CoA carboxylase into a malonyl-CoA or enters the citric acid cycle and exits as an α-ketoglutarate [135, 181]. After the dissociation of $CO_2$ from α-ketoglutarate, a succinate semialdehyde is formed [181], which is further activated by an acyl-CoA ligase AmcH. A malonyl-CoA moiety is selected by an AT domain and loaded onto ACP AmcB. A KS protein AmcF catalyzes the condensation of an activated succinate semialdehyde and a thiolated malonyl-CoA moiety. A hypothetical protein AmcG was proposed to cyclize the ACP-attached acyl chain into a cyclopentane ring. A dehydratase AmcE

removes the hydroxyl group and a double bond is formed. In amipurimycin biosynthesis, Romo *et al.* [181] reports on the presence of 3-oxoacyl-(ACP) reductase (KR) AmcA in the BGC and proposes that AmcA is responsible for saturation of formed double bond but no experimental proof exists [181]. AmcC encodes an aminotransferase that exchanges a keto group with an amino group on the scaffold of the cyclopentane ring. A TE AmcD releases the final product – a cispentacin moiety (Figure 24).

Based on the above presented hypothesis [181], paralells can be drawn for the biosynthesis of cispentacin moiety present in myxopentacins. Proteins encoded by *mxpB-mxpH* genes are homologous to AmcB-AmcH from amipurimycin BGC (Table 5). However, no homolog of 3-oxoacyl-(ACP) reductase (KR) AmcA [181] that would saturate the double bond of the intermediate is found directly in MXP gene cluster but there are many homologs encoded in the rest of the genome that could fullfil the task.

**Table 5. Similarities of protein homologs responsible for the biosynthesis of the cispentacin moiety**

| Protein from MXP BGC | Protein homolog (Identity/Similarity) |
|---|---|
| MxpB | AmcH (42/68)<br>Cfa5 (39/68) |
| MxpC | AmcG (19/53)<br>Cfa4 (25/54) |
| MxpD | AmcF (50/72)<br>Cfa3 (46/71) |
| MxpE | AmcE (56/78)<br>Cfa2 (49/74) |
| MxpF | AmcD (49/74) |
| MxpG | AmcC (58/83) |
| MxpH | AmcB (48/81)<br>Cfa1 (42/72) |

### 2.3.4.2 The biosynthesis of pseudoarginine moiety

In addition to the above described cispentacin moiety, myxopentacins contain a pseudoarginine moiety (Figure 25), another rare feature among NPs, found also in arphamenines and ketomemicins [136]. The biosynthesis of a pseudopeptide moiety was elucidated *in vitro* and is in ketomemicin biosynthesis catalyzed by KtmA, KtmB, KtmC and KtmF enzymes [136]. Three protein homologs corresponding to KtmA-KtmC are present in myxopentacin BGC and were named MxpJ-MxpL (Table 4). MxpJ, KtmA homolog (57 % identity), shows significant

homology to CitE family enzymes. CitE is a β-subunit of the bacterial ATP-independent citrate lyase that otherwise consists of α, β and γ subunits [187]. A citryl-S-ACP, CitE, catalyzes a retro-aldol reaction from citryl-ACP to oxaloacetate and acetyl-ACP [136], while other similar lyases like L-malyl-CoA lyase/β-methylmalyl-CoA lyase [188] utilize acyl-CoA thioesters as substrates [136]. MxpK, KtmB homolog (59 % identity), shares characteristics with PLP-dependent amino acid Cα-acyltransferases, catalyzing C-C bond formation between the α-carbon of an amino acid and the carbonyl carbon of an acyl-CoA thioester, followed by decarboxylation to generate an α-oxoamine structure [136]. PLP-dependent amino acid Cα-acyltransferases are important enzymes in various processes such as threonine degradation, heme, biotin and sphingolipid biosynthesis [136]. MxpL, KtmC homolog (63 % identity), shows homology to MaoC dehydratases. Those are classified into a thioesterase/thiol ester dehydrase-isomerase superfamily with characteristic Hotdog fold. In *E. coli*, MaoC protein contains an N-terminal short-chain dehydrogenase domain that is known to catalyze dehydrogenation of a variety of aliphatic and aromatic aldehydes using NADP as a cofactor. The C-terminal is assumed to be responsible for its enoyl-CoA hydratase activity [189].

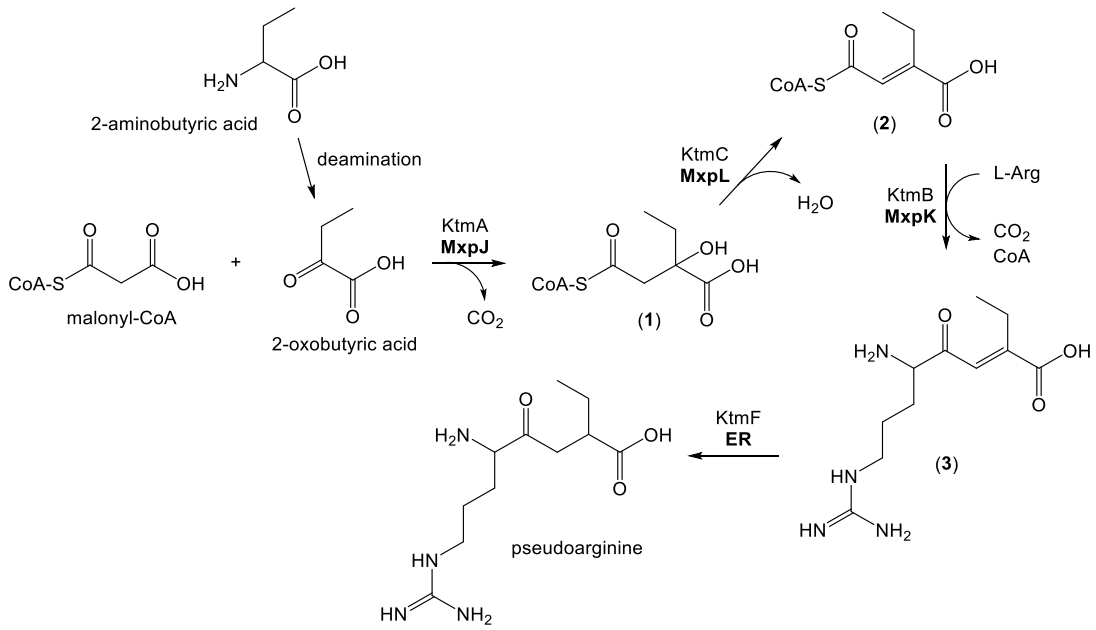



**Figure 25. A proposed mechanism for the biosynthesis of the pseudoarginine moiety.** After deamination of 2-aminobutyric acid to 2-oxobutyric acid or 2-aminovaleric acid to 2-oxovaleric acid, MxpJ catalyzes the reaction between malonyl-CoA and 2-oxocarboxylic acid to yield (**1**). A dehydratase MxpL catalyzes the removal of water, which results in the formation of the double bond in (**2**). MxpK, an arginine Cα-acyltransferase, condenses L-Arg to (**2**) and meanwhile removes CoA and forms (**3**). After the action of an unknown ER domain, the pseudoarginine moiety found in myxopentacins is biosynthesized.

On the basis of the *in vitro* experiments on ketomemicins, we are able to propose possible biosynthesis of a pseudoarginine moiety found in myxopentacins. MxpJ is speculated to

catalyze the reaction between malonyl-CoA and 2-oxobutyric or 2-oxopentanoic acid, which results in (**1**) (Figure 25). The 2-oxobutyric or 2-oxopentanoic acid could originate from 2-aminobutyric and 2-aminovaleric acids (an isomer of valine), respectively, after a deamination reaction common in protein degradation [190]. The reaction product (**1**) is further transformed by a dehydratase MxpL and results in loss of water and (**2**) (Figure 25). MxpK catalyzes C-C bond formation between L-Arg and (**2**), removing CoA moiety and $CO_2$. The product of that reaction resembles a pseudoarginine moiety (**3**) present in myxopentacins. The only necessary step still missing is the reduction of the double bond, which is usually catalyzed by an enoyl reductase enzyme such as NAD(P)H-dependent reductase KtmF (H340_08821) from ketomemicin biosynthesis [136]. No KtmF homologs are present in the BGC-39A but there are 4 homologs encoded in the genome of producer strain MCy9003.

### 2.3.4.3 The assembly of myxopentacins

The building block sequence in myxopentacins of type A and B partly differs but all the family members isolated so far share a common part: two subsequent cispentacin moieties, a pseudoarginine moiety and a valine. Moreover, all the building blocks in myxopentacins are bound via peptide bonds (Figure 26).

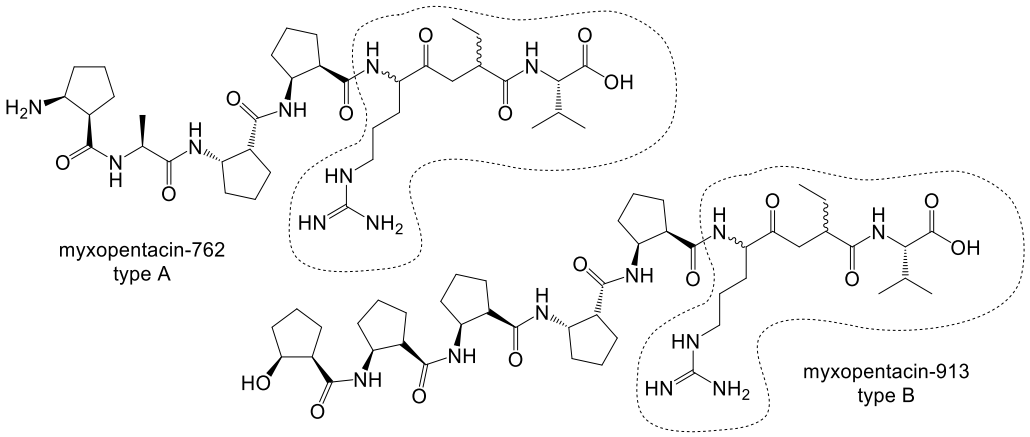



**Figure 26. Myxopentacins type A and B contain pseudoarginine moiety and valine**

As the inactivation of *mxpA*, encoding a NRPS with domain sequence of C-A1-PCP-C-A2-PCP-C, resulted in abolished production of myxopentacins of both types (Figure 22), we can propose that MxpA is essential for the myxopentacin biosynthesis. The selectivity of A domain is determined by its amino acid sequence in certain positions [72]. According to the prediction tools incorporated into antiSMASH program, Leu, Ile or Val are selected by A1 domain and Cha (β-cyclohexyl-L-alanine), Val or Ile by A2 domain. Due to common pseudopeptide moiety and the terminal valine residue in both types of myxopentacins (Figure 26), we suggest that the

MxpA module 1 selects the 2-aminobutyric acid-originating part of pseudoarginine moiety, which is condensed with valine incorporated by module 2 (Figure 27).

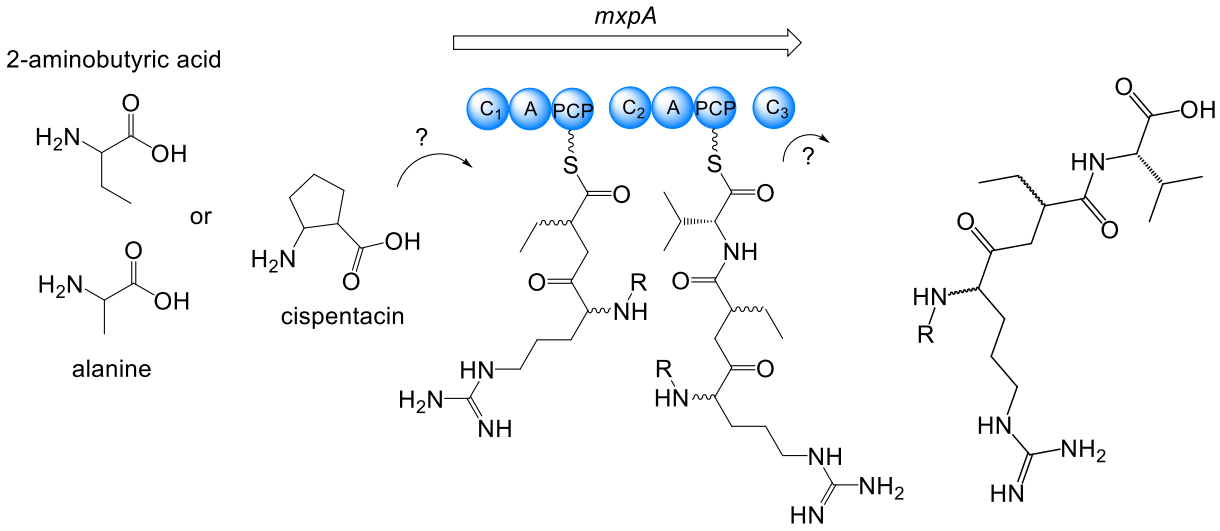



**Figure 27. The proposed partial assembly of myxopentacin building blocks on NRPS MxpA.** The module 1 could select the 2-aminobutyric acid-originating part of pseudoarginine moiety and further condense it with valine, selected by module 2. The polymerization of cispentacin moieties and 2-aminobutyric acid or alanine building blocks in type A myxopentacins is not yet understood, neither the release mechanism. R denotes the remaining of the myxopentacin molecule.

The assembly of cispentacin moieties, and alanine or 2-aminobutyric acid found in myxopentacins type A, to pseudoarginine moiety is not yet understood (Figure 26). It might be catalyzed by some unknown enzyme other than modular NRPS or by another BGC encoded in the MCy9003 genome. The latter possibility could be evaluated after the heterologous expression of myxopentacin BGC in host strain. Importantly, the role of other BGCs, such as BGC-1, BGC-12 and BGC-35B that interfered in myxopentacin production yields (Table 3), in this biosynthesis is unclear. An enzyme such as KtmD, a homolog of peptide ligase PGM1 [191] with an ATP-grasp domain [192], responsible for the peptide bond formation between the amidino amino acid and pseudodipeptide in the final step of ketomemicin biosynthesis [136] that would somehow explain the assembly of myxopentacin building blocks, does not exist in the genome of MCy9003.

The partial assembly proposal suggests that only pseudoarginine moiety and valine are selected and tethered by NRPS MxpA. However, the final product contains not only two but 6 or 7 moieties (Figure 26) and is therefore too large to be assembled strictly noniteratively. Moreover, the myxopentacins are a much larger family of compounds and consist of many more variants [178] that could change our opinion about biosynthesis once known. The biosynthesis could

also be at least partly iterative. The iterative NRPS mechanism was reported in case of pyrrolamide antibiotics: congocidine, distamycin and new analog, revealing that amidohydrolase Pya25 is necessary for the polymerization of pyrrole moieties [193]. The congocidine consist of two 3-aminopropionamidine and two 4-aminopyrrole-2-carboxylate parts (Figure 28A). After the analysis of the culture extract, the acetylated 4-aminopyrrole-2-carboxylic acid was detected. Consequently authors proposed that the 4-acetamido-pyrrole-2-carboxylic acid attaches to both PCP domains Pya19 and Pya21 but only the precursor attached to Pya21 can be deacetylated by an amidohydrolase Pya25 prior to condensation with the intact acetoamidopyrrole residue (Figure 28B) [193, 194].

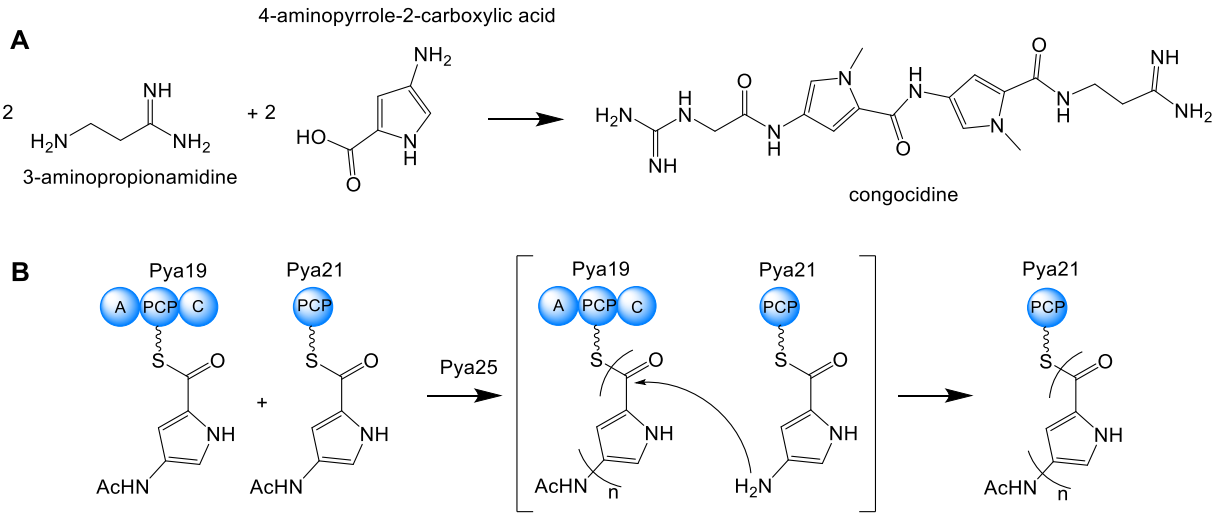



**Figure 28. The iterative mechanism of congocidine biosynthesis.** A) The precursors 3-aminopropionamidine and 4-aminopyrrole-2-carboxylic acid used for congocidine biosynthesis. B) The proposed congocidine assembly mechanism. The acetylated 4-aminopyrrole-2-carboxylic acid attaches to both PCP domains in Pya19 and Pya21. It remains intact when attached to Pya19 before condensation with deacetylated precursor from Pya21. The deacetylation is catalyzed by an amidohydrolase Pya25.

If an acetylated cispentacin moiety is detected and if a responsible N-acetyltransferase would be found, the assembly mechanism similar to pyrrolamide's could be assessed. However, no N-acetyltransferase was detected in the myxopentacin BGC. The 4-aminopyrrole-2-carboxylic acid from congocidine biosynthesis originates from fructose-6-phosphate and the acetyl group is installed by GlmU, encoding a fused N-acetylglucosamine-1-phosphate uridyltransferase and glucosamine-1-phosphate acetyltransferase from primary metabolism [194]. There are a hydrolase (Orf(-7)), an amidohydrolase (Orf(-6)) and two MBL fold metallo-hydrolases (Orf1, Orf2) present next to myxopentacin biosynthetic genes (Table 4). However, their role in myxopentacin polymerization cannot be proposed without further experiments.

### 2.3.4.4 The gene cluster borders of myxopentacin

As no other inactivation, gene knock-outs or heterologous expression mutants were prepared, except for the inactivation of *mxpA* gene, nothing certain can be said about the cluster borders. The *orf(-8), orf(-1)* and *orf2* genes (Table 4) encode hypothetical proteins, whereas *orf(-4)* and *orf(-3)* belong among DUF1427 family proteins. The latter contain domains with unknown function and are probably not involved in biosynthesis process. Two MFS transporters encoded by *orf(-5)* and *orf6* are present, one on each side of the biosynthetic genes (Table 4). They belong into the family of macrolide efflux protein A (MefA) [195] and might be multidrug resistance (MDR) transporters, the drug/$H^+$ antiporters (DHAs) that mediate the efflux of a variety of drugs and toxic compounds through the membranes, conferring resistance to these compounds [196]. They could be involved in the transport of myxopentacin out of the cell. The amidohydrolase (*orf(-6)*) mentioned above (Table 4) might play a role in myxopentacin assembly. The involvement of hydrolase and MBL fold metallo-hydrolases in myxopentacin biosynthetic pathway is unknown. Orf3 belongs to family of intramembrane metalloproteases, proteolytically removing the C-terminal three residues of certain prenylated proteins. Orf4 is predicted to be an aspartyl/asparaginyl β-hydroxylase and Orf5 a methyltransferase. Orf(-2), encoded upstream of MxpA, belongs to an endonuclease/exonuclease/phosphatase family. The family includes $Mg^{2+}$-dependent endonucleases and many phosphatases involved in cell-cycle regulation intracellular signaling [197]. Orf(-2) is therefore believed not to be involved in myxopentacin biosynthesis. Orf(-1) is recognized to be a hypothetical protein with no domain detection. Search for Orf(-1) and Orf(-2) in other myxobacterial genomes revealed that they are always found right next to each other and might therefore be in vicinity of the myxopentacin BGC only by chance. The genes downstream of MxpL, except for *orf1* and *orf2* genes in some cases, were not found in the BGC vicinity in other myxobacterial strains described in Section 2.3.5. The genome information about the other strains harbouring myxopentacin biosynthetic genes revealed that only *orf(-6)* gene, encoding an amidohydrolase is present in all of them (see below).

### 2.3.5 Other myxobacterial strains harbour parts of MXP gene cluster

The genome collection of novel myxobacterial strains isolated in this group offers the possibility to search for the investigated BGC in other strains although they were not found to be producers of myxopentacins. We were interested in whether any of the strains contain biosynthetic genes for cispentacin or pseudoarginine moieties. After a thorough analysis, a trend as summed up in Table 6 was observed.

**Table 6. Pseudoarginine and cispentacin biosynthetic genes-containing myxobacterial BGCs**

| Strain | MxpI | KtmF homolog | NRPS | A domain selectivity | Orientation | Figure 29 |
|---|---|---|---|---|---|---|
| MCy9003[b] bgc39a | Yes | No | C-$A_1$-PCP-C-$A_2$-PCP-C | $A_1$ – Leu<br>$A_2$ – Cha[c] | →←→ | Group A |
| MCy10644[b] | Yes | No | C-$A_1$-PCP-C-$A_2$-PCP-C | | →←→ | |
| MCy11578[b] | Yes | No | C-$A_1$-PCP-C-$A_2$-PCP-C | | →←→ | |
| MCy8383 bgc47 | Yes | Yes | C-$A_1$-PCP-C-$A_2$-PCP-C | $A_1$ – Leu<br>$A_2$ – Leu | →←← | Group B |
| MCy8375 bgc8 | Yes | Yes | C-$A_1$-PCP-C-$A_2$-PCP-C | | →←← | |
| MCy8337 bgc23 | Yes | Yes | No info[a] | | →←← | |
| MCy10585 bgc16 | No | Yes | C-$A_1$-PCP-C | $A_1$ – Leu | →→→ | Group C |
| MCy9280 bgc6 | No | Yes | C-$A_1$-PCP-C | | →→→ | |
| MCy9171 bgc4 | No | Yes | C-$A_1$-PCP-C | | →→→ | |
| MCy8401 bgc44 | No | Yes | C-$A_1$-PCP-C | | →→→ | |
| MCy8396 bgc67 | No | Yes | C-$A_1$-PCP-C | | →→→ | |
| MCy8286 bgc7 | No | Yes | C-$A_1$-PCP-C | | →→→ | |
| MCy8278 bgc55 | No | Yes | No info[a] | | →→→ | |
| MCy5730 bgc64 | No | Yes | C-$A_1$-PCP-C | | →→→ | |
| MCy10653 bgc18 | No | Yes | C-$A_1$-PCP-C | | →→→ | |
| MCy9101 | Yes | / | / | | /→/ | / |

[a]The genomes of analyzed bacteria are assembled in contigs, which can split the BGCs of interest
[b]The confirmed myxopentacin producers
[c]β-cyclohexyl-L-alanine

A *Cystobacter* sp. strain MCy9101 contained biosynthetic genes encoding only enzymes for biosynthesis of a cispentacin moiety. In this BGC an additional gene was observed, namely zinc-binding dehydrogenase (WP_108069954.1 (89/98)), that could act as the original missing ER domain in cispentacin moiety biosynthesis (Figure 24). However, no homology with proposed AmcA from amipurimycin BGC [181] was detected. In 14 other strains, three major differences among found BGC were observed (Figure 29).

1. NRPS contains 2 modules in C-A-PCP-C-A-PCP-C sequence only when encoded in different operon than cispentacin and/or pseudoarginine part. In case all three parts are encoded in the same operon, an NRPS gene contains only one module in C-A-PCP-C sequence
2. The pseudoarginine biosynthetic part differs according to the content of an additional ER domain containing protein, a KtmF homolog
3. The cispentacin biosynthetic part differs according to the content of MxpI prolyl-tRNA synthetase

DNA sequence, stretching from *orf(-8)* to *orf7*, encompassing *mxpA-mxpL* genes, in MCy9003 (Table 4) is in 96.8 % and 99.5 % identical in the genomes of MCy10644 and MCy11578, respectively. Two representatives of the group B, MCy8375 and MCy8383, chosen since their *mxpA* sequence is fully assembled (Table 6), both share 62 % identity to *mxpA-mxpI* (NRPS and cispentacin part) and 82 % to *mxpJ-mxpL* - pseudoarginine part (Figure 29). Regarding the

genes surrounding the MXP biosynthetic genes, *orf(-8)-orf(-3)* and *orf1* genes are present in all three strain but not in proximity of myxopentacin BGC. The latter is also true for *orf2*, however, this gene is present only in MCy8383 strain. The group C representatives, MCy9280, MCy10653 and MCy10585, show 39-40 % identity to *mxpA*, 48-50 % to *mxpB-mxpH* cispentacin biosynthetic genes and 79-82 % to *mxpJ-mxpL* pseudoarginine biosynthetic genes (Figure 29). Only the amidohydrolase, encoded by *orf(-6)* gene, is present in all three strains.

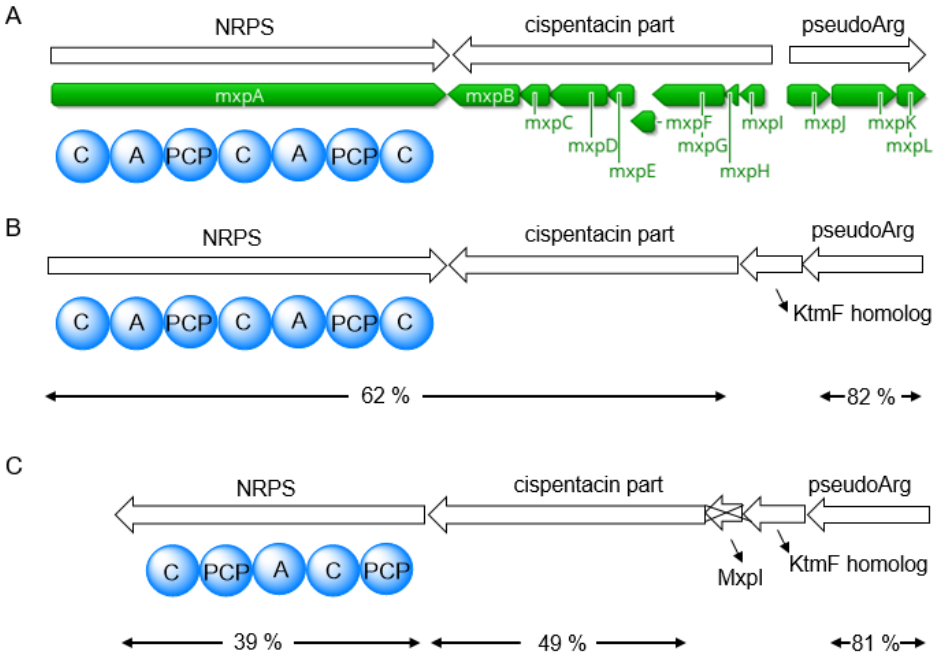



**Figure 29. The orientation of MXP and MXP-like BGCs in MCy9003 and other myxobacterial strains.** A. The investigated BGC from strain MCy9003 contains two adenylation domains and MxpI but no KtmF homolog with ER domain. B. The BGC is arranged in two operons. The genes responsible for pseudoarginine moiety biosynthesis have reverse orientation and additionally contain a KtmF homolog. C. All biosynthetic genes are organized in one operon. A pseudoarginine biosynthetic part has an additional KtmF homolog, the cispentacin biosynthetic genes are short for MxpI and the NRPS gene has only one A domain but a pair of PCP and C domains. Additionally, the identity percentage of a selected region of the MXP-like BGCs to either NRPS and/or cispentacin, or pseudoarginine biosynthetic genes of MCy9003's MXP BGC is shown for each group.

The BGC arrangement as shown in Figure 29B seems to be characteristic for myxobacteria from genus *Archangium* [43] as the analyzed strains were novel isolates of *Archangium violaceum*, *A. gephyra* and *Archangium* sp.. The BGC arrangement from Figure 29C seems to be less restrained to one genus and is therefore found in few members of *Myxococcaceae* family, namely *Myxococcus* sp., *Pyxidicoccus* sp. and other novel isolate *Cystobacterineae* sp. not yet classified into a beloning family or genus. The search for KtmF homologs found in BGCs in groups B and C results in up to 13 BLAST hits in MCy9003, whereas there are only

4 BLAST hits when looking for protein homologs of KtmF from *Streptomyces mobaraensis*. The strains containing myxopentacin-like BGCs from groups B and C (Figure 29) were not detected as myxopentacin alternative producers. However, no LC-MS data of crude extracts was available for strains MCy8383, MCy8375 and MCy8337 (Group B). They might be even better myxopentacins producers, due to different operon arrangement and direct presence of KtmF homolog in the BGC but that cannot be confirmed at the moment. Although, it could also be possible that the myxopentacin BGC is not expressed in WT when cultivated in selected growth conditions, as observed on the example of myxoglucamide (Section 2.3.7). The MS-MS spectra of MCy10644 and MCy9003 crude extracts were subjected for Global Natural Product Social Molecular Networking (GNPS) [156]. The derivatives of myxopentacins type A and type B (Figure 18, Figure 19) were clustered separately under selected parameters (Figure S1- 5). However, none of those molecular masses were found in MS-MS spectra of MCy10585, MCy9280 and MCy10653 (Group C). As most probably not only the known biosynthetic genes (*mxpA*-*mxpL*) are responsible for the myxopentacin biosynthesis and assembly, it cannot be said for sure if the B and C BGCs are complete and capable of producing a secondary metabolite without further experiments.

### 2.3.6 Potential biological activity of myxopentacins

Additionally, the MXP BGC contains MxpI that encodes a prolyl-tRNA synthetase from YbaK/EbsC family, found in MCy9003 genome in two copies in total. An aminoacyl-tRNA synthetases catalyze an essential process of the amino acids attachment to the 3' termini of cognate tRNAs and ensure the accuracy of genetic information transfer from DNA to protein. Due to their importance in organisms, the antibiotics selectively targeting them are being developed [198]. Cispentacin is a weak inhibitor of *C albicans*' prolyl-tRNA synthetase [199, 200], therefore MxpI probably acts as a resistance protein and protects MCy9003 against the produced compound. A synthetic cispentacin derivative icofungipen, tested in clinical trials in phase II [200], is an isoleucyl-tRNA synthetase inhibitor in *C. albicans* [201]. On the other hand, NPs related to ketomemicin and myxopentacins, arphamenines are known as specific inhibitors of aminopeptidase B and enhance immune responses [202]. Due to structure similarity (Figure 30), myxopentacins could also have protease inhibitory properties but were not tested for such activity.

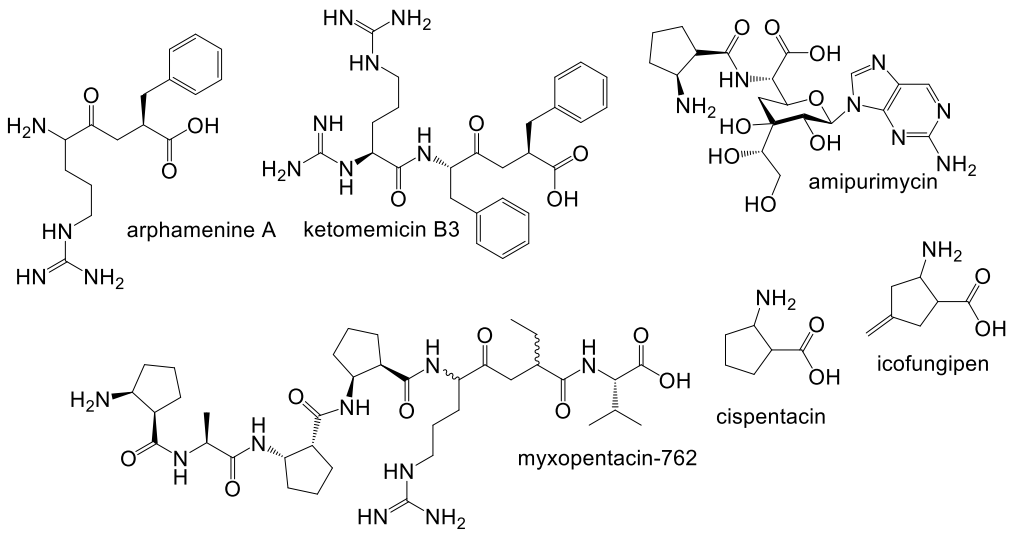



**Figure 30. A list of NPs with features found also in myxopentacin family**

### 2.3.7 Discovery of myxoglucamides

The discovery of myxoglucamide came in connection with the quest after myxopentacin BGC. Due to rare features in myxopentacin structure, its biosynthetic genes were hard to predict. Two myxobacterial alternative producers MCy11578 and MCy10644 with accessible genome sequences were available. Strain MCy11578 shares a very high identity with MCy9003 but more differences can be found when compared to MCy10644. In total 14 common BGCs were identified between the three strains as candidates for myxopentacin BGCs, among them also BGC-18. A plasmid with two strong constitutive promoters P*tn5*, both directed towards the homology region, constructed to upregulate genes in both directions, was integrated into the MCy9003 genome via homologous recombination (Figure 31). The production profile of mutant strain MCy9003/pSB2tn5Kan001bgc18 (Table S1- 3: M172) showed significant differences; three new peaks with exact masses 1- 631.3795 $[M+H]^+$, 2- 645.3955 $[M+H]^+$ and 3- 659.4112 $[M+H]^+$ appeared (Figure 32). The large-scale cultivation of this mutant strain enabled isolation of four derivatives from peaks 2 and 3. The compound production yield from peak 1 was too low to be isolated. The structures of novel derivatives were elucidated (Figure 33) in 2018 by Alexander Popoff as well as the other structural features like stereochemistry [178].

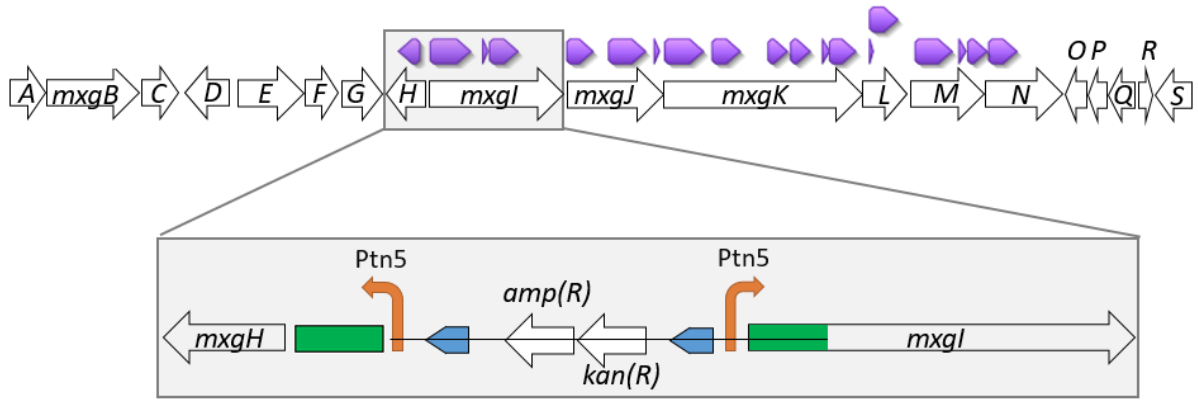



**Figure 31. The promoter integration into BGC-18 of MCy9003.** The two P*tn5* promoters are marked with orange arrows and were integrated to upregulate *mxgH* and *mxgI* gene expression. The catalytic domains for the genes *mxgA-mxgS* are shown in purple. The integrated plasmid pSB2tn5Kan001bgc18, harbouring two *Ptn5* promoters, the homology arm 001bgc18 (green), origins of replication (blue) and selection markers for ampicillin and kanamycin, is marked with the black line.

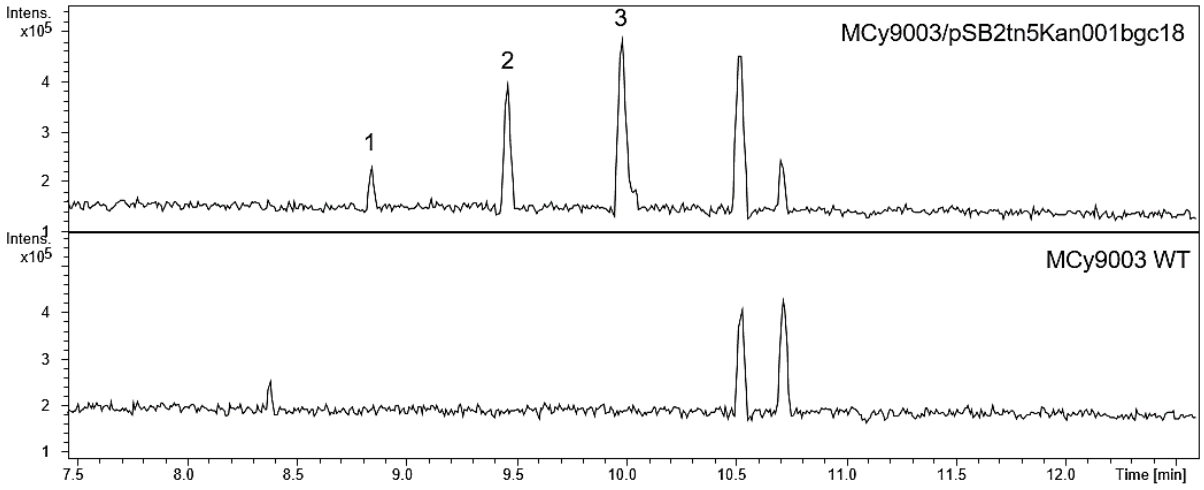



**Figure 32. The production profile of WT strain and mutant strain with upregulated BGC-18.** The base peak chromatograms (BPC) of wild type and mutant strain M172 (Table S1- 3) crude extracts show increase of 3 peaks. The peak numbers correspond to molecular masses: 1- 631.3795 m/z $[M+H]^+$; 2 - 645.3955 m/z $[M+H]^+$; 3 – 659.4112 m/z $[M+H]^+$. Only peaks 2 and 3 were produced in sufficient amounts for isolation and structure elucidation. The analysis confirmed that these compounds are most likely unknown structures.

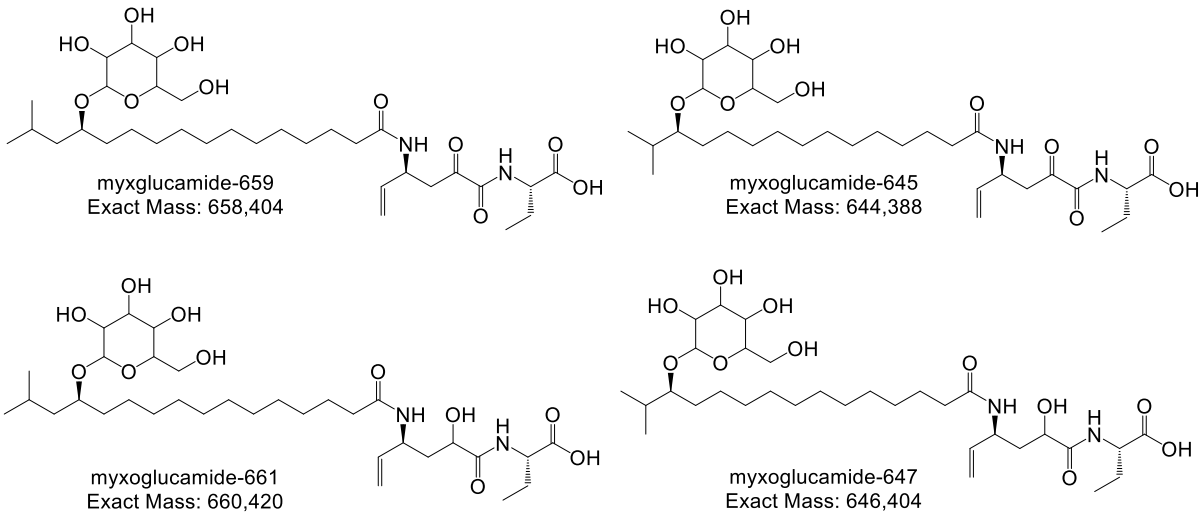



**Figure 33. Myxoglucamides**

### 2.3.8 Biosynthesis of myxoglucamides

Myxoglucamides are hybrid PKS type I and NRPS secondary metabolites of MCy9003. The myxoglucamide biosynthetic genes are listed in Table 7 and the proposed biosynthesis is presented in Figure 34. First, the catalytic domains were analyzed. The alignments ACP, PCP, C, DH and KS domains are presented in Figure S1- 7 to Figure S1- 11.

**Table 7. Annotated genes found in the myxoglucamide BGC-18**

| *Gene* | *Size*[a] | *Protein homolog*[b] | *Proposed function* | *Catalytic domains* |
|---|---|---|---|---|
| *mxgA* | 393 | WP_081468166.1 (78/94) | efflux RND transporter periplasmic adaptor subunit | |
| *mxgB* | 1048 | WP_013374292.1 (88/98) | efflux RND transporter permease subunit | |
| *mxgC* | 420 | WP_145679605.1 (58/82) | glycosyltransferase | GT |
| *mxgD* | 492 | WP_120584626.1 (66/81) | unknown | |
| *mxgE* | 745 | WP_043394394.1 (71/88) | S9 family peptidase | |
| *mxgF* | 368 | WP_169345882.1 (87/97) | aldo/keto reductase | |
| *mxgG* | 464 | WP_120613650.1 (87/97) | cytochrome P450 | Ox |
| *mxgH* | 455 | WP_120613649.1 (73/87) | monoxygenase/C-domain (hypothetical protein) | C |
| *mxgI* | 1509 | WP_015347655.1 (63/84) | PKS/NRPS | FAAL; ACP; $^{L}C_{L}$; Ox |
| *mxgJ* | 1082 | WP_015347653.1 (58/82) | non-ribosomal peptide synthetase | $^{L}C_{L}$; A; PCP |
| *mxgK* | 2242 | WP_120613648.1 (81/92) | polyketide synthase | KS; AT; KR; DH; ACP; E |
| *mxgL* | 490 | WP_120643780.1 (77/90) | non-ribosomal peptide synthetase | COM; $^{D}C_{L}$ |
| *mxgM* | 827 | WP_120643780.1 (83/93) | non-ribosomal peptide synthetase | A; PCP; TE |
| *mxgN* | 877 | WP_120615411.1 (76/90) | luciferase-like monooygenase class flavin-dependent oxidoreductase | $^{D}C_{L}$; Ox |
| *mxgO* | 235 | WP_052518234.1 (76/92) | lipase | |
| *mxgP* | 209 | WP_164016063.1 (77/89) | CoA pyrophosphatase | |

| | | | | |
|---|---|---|---|---|
| *mxgQ* | 297 | WP_141640493.1 (86/97) | undecaprenyl-diphosphate phosphatase | |
| *mxgR* | 156 | WP_120205724.1 (90/96) | unknown | |
| *mxgS* | 415 | WP_121718591.1 (85/96) | glycosyltransferase | GT |

*[a]Sizes are given in amino acids; [b]Accession numbers and percentage of identity/similarity are given in parentheses.*

The backbone assembly is assumed to start with MxgI, an NRPS, with 4 biosynthetically relevant domains, namely fatty acid AMP-ligase (FAAL), ACP, C and Ox domains (Figure 34). As there are no existing genes in the BGC-18 to actually build the methylpentadecenoic (iso 15:0) or methylhexadecenoid acid (iso 16:0) selected by FAAL domain in MxgI, they probably originate from primary metabolism [203]. The selected fatty acid is transferred to ACP. The remaining condensation and oxidation domains in MxgI are a part of module 1, which continues with neighbouring MxgJ protein containing an unfunctional C domain, followed by an A domain, predicted to select glutamine (Figure 34) [72]. Module 1 incorporates an unknown amino acid (more in Section 2.3.8.1), which is after a tailoring step transformed into an α-vinyl moiety (Figure 34). This tailoring step could be catalyzed by an oxidation domain (MxgI-Ox), an α-ketoglutarate-dependent taurine dioxygenase (TauD) from TfdA family reported to be involved in the biosynthesis of some natural products like fungal meroterpenoids [204].

MxgK comprises a full PKS module with a terminal E domain. Such E domain arrangement is unusual since they catalyze the transformation of L-amino acids to D-amino acids [68] and are usually located in NPRS modules. A sequence analysis of the AT domain reveals a distinct motif "QTALTQ…GHSVGE…HAFH'' matching with sequence requirements for malonyl-CoA selectivity [83]. The module 2 encodes as well DH and KR catalytic domains. The MxgK-DH domain was compared to known and functional erythromycin dehydratase [86]. Many of the amino acid residues in the presumably conserved DH domain motifs deviate from the expected residues (Figure S1- 8), therefore it was found to be likely non-functional. The MxgK-KR contains the NADP-binding motif GGxGxxG and has additional conserved residues Ser1299, Tyr1312 and Asp1316 (Figure S1- 9), aligning to Ser146, Tyr159 and Lys163 from tropinone reductase-II (TR-II), respectively [205], which suggests that the domain is functional. The ACP domain, bearing the intermediate acyl-peptidyl chain with now three building blocks: a fatty acid, an unknown amino acid with a PKS part, is followed by an epimerization domain. The stereochemical assignment reveals that the amino acid containing an α-vinyl moiety, incorporated into myxoglucamide by MxgJ and MxgK, is D- configured [178], which confirms the functionality of E domain. This E domain is assumed to act on the α-carbon of the α-vinyl moiety [178]. However, the conserved His-motif HHxxxDG characteristic for C and E domains was found mutated into HAxxxDG in this E domain, which could also mean that the domain is

not functional and the observed D- configuration of the α-vinyl moiety is catalyzed by a different mechanism. However, the sequence of HVxxxDG was found in E domains of MtbF from mycobactin biosynthesis, which is located in the C-terminal of the protein and is proposed to act as a TE domain [206]. The protein BLAST of the E domain-containing MxpK found homologs in a few strains of δ- and γ-proteobacteria but were not connected to any known NP. The location of the E domain in this case is definitely unusual since it could be, according to its function, located just after the module 1.

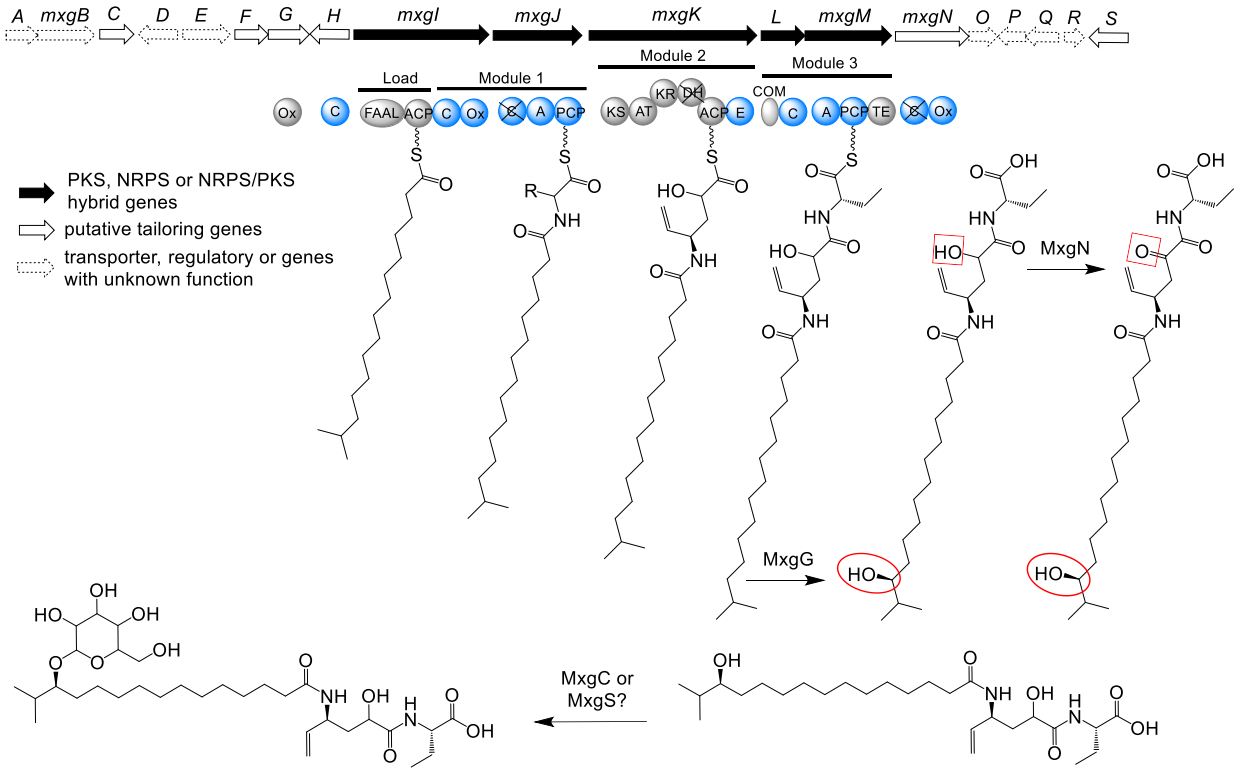



**Figure 34. Putative biosynthesis of myxoglucamides.** PKS genes, as well as stand-alone Ox domain from tailoring enzymes, are coloured grey. The NRPS domains present in the BGC are coloured blue. The R represents the side chain of an unknown amino acid. MxgN, a flavin-dependent monooxygenase, could catalyze the transformation of the hydroxyl group to the keto group in myxoglucamide derivatives. The hydroxyl group addition to the fatty acid could be catalyzed by a cytochrome P450 MxgG. The attachment of the glucose moiety is probably catalyzed by MxgC glycosyltransferase.

MxgL and MxgM comprise another NRPS module (Figure 34). MxgL has a communication (COM) domain and a $^{D}C_{L}$ domain designated as L-acceptor but has D-peptidyl-S-T donor selectivity. The module continues with A, PCP and TE domains encoded in MxgM. The A domain, predicted to select valine, incorporates L-aminobutyric acid on the C-terminal part of myxoglucamide. The following MxgN has an unfunctional $^{D}C_{L}$ domain (Figure S1- 10) and a flavin-dependent monooxygenase domain.

Additionally, MxgH contains a functional condensation domain. Tailoring enzymes present in the BGC-18 are MxgG cytochrome P450, MxgF aldo/keto reductase and two glycosyltransferases, MxgC and MxgS (Table 7). First, a flavin-dependent monooxygenase MxgN could be involved in the α-carbon oxidation [159, 207], and could catalyze the transformation of the hydroxyl group to the keto group in myxoglucamide derivatives (Figure 34). Second, the cytochrome P450 MxgG could oxidize the fatty acid in myxoglucamide since cytochrome P450 enzymes in *Candida tropicalis* also catalyze the corresponding transformation of fatty acids [208]. MxgG might add the hydroxyl group at C-4 or C-5 position of fatty acid in myxoglucamide 645 and 647 or myxoglucamide 659 and 661, respectively (Figure 33). The timing of that reaction is unknown. Last, the final secondary metabolite contains a glucose moiety attached to a hydroxyl group of the fatty acid chain. Two glycosyltransferases MxgC and MxgS are present in the MXG BGC but no knock-out or inactivation mutants were prepared to confirm which of them catalyzes the reaction. The MxgC is a glycosyltransferase family 1 protein and shares 31 % identity with GtfB, a UDP-glucose transferase from vancomycin biosynthesis [209]. Best BLAST hits are found in *Azospirillum brasilense*, a nitrogen fixing and plant growth promoting bacteria [210]. On the other hand, the glycosyltransferase MxgS is more likely involved in cell wall biosynthesis and might be important for the myxobacteria in general also because the best BLAST hits were found in *Corallococcus* sp.. The proposed glycosyltransferase attaching the glucose moiety could therefore be MxgC since it was additionally found in other myxobacterial strains with partial myxoglucamide BGCs (Section 2.3.11). The specific function and importance of MxgF, MxgH and other genes in BGC for the biosynthesis of myxoglucamide is still unknown.

#### 2.3.8.1 The α-vinyl group

The biosynthesis of myxoglucamides can be predicted *in silico* with a thorough analysis of biosynthetic genes and the final chemical structure (Figure 33). An extraordinary feature such as an α-vinyl group in 4-amino-2-hydroxyhex-5-enoic acid is present in the final myxoglucamide structure. Due to the rare occurrence in NPs, its origin and biosynthesis are of interest. Examples of NPs containing such feature are rhizobitoxine from symbiotic bacteria *Rhizobium japonicum*, *Bradyrhizobium elkanii* [211] and a plant pathogen *Pseudomonas andropogonis* [212], the fungal metabolite D-vinylglycine [212], aminoethoxyvinylglycine (AVG) from *Streptomyces* sp. NRRL 5331 [213] and methoxyvinylglycine (MVG) from *Pseudomonas aeruginosa* [214].

Some information about the origin of the α-vinylglycine component in biosynthetic pathways of rhizobitoxine, AVG and MVG is available. The α-vinylglycine amino acid in rhizobitoxine and AVG is derived from homoserine (Figure 35A) [211, 213]. However, the *in vitro* reconsitution of MVG biosynthesis has shown that L-glutamic acid is necessary for the formation of the α-vinylglycine (Figure 35A) [214]. Based on that, the feeding experiments of the myxoglucamide producer with isotope labelled L-glutamic acid, L-glutamine and L-aspartic acid, a precursor of L-homoserine [211], were conducted but no incorporation was observed (Figure 35B) [178].

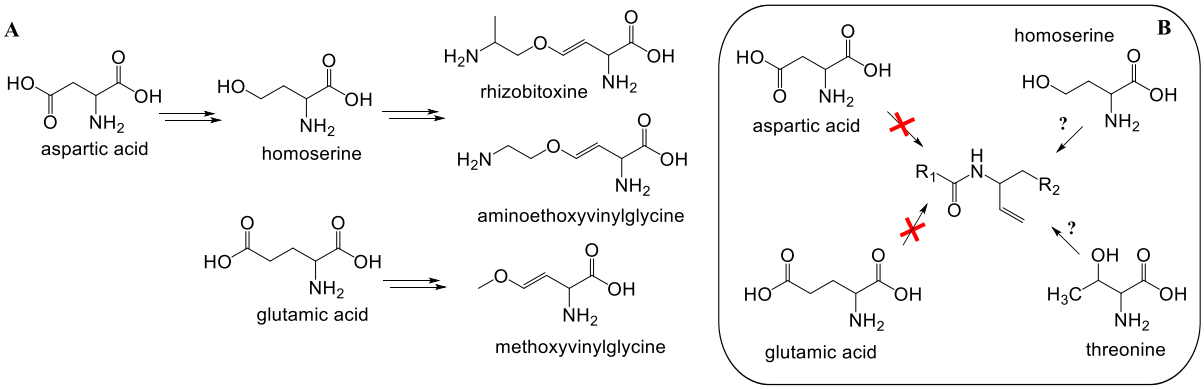



**Figure 35. The precursors for α-vinylglycine containing NPs.** A) Aspartic acid is a precursor of homoserine, which is incorporated into rhizobitoxine and aminoethoxyvinylglycine. On the other hand, the α-vinylglycine from methoxyvinylglycine is derived from glutamic acid. B) Feeding of isotope labelled aspartic acid and glutamic acid showed no incorporation in the myxoglucamide scaffold.

The origin of the α-vinyl group in myxoglucamide so remains elusive. For that reason, the protein homologs of enzymes involved in metabolic pathway encompassing L-aspartate, L-homoserine and L-threonine were sought after in MCy9003 (Figure S1- 12). The protein homologs of aspartate kinase/homoserine dehydrogenase and aspartate semialdehyde dehydrogenase from *M. xanthus* that usually catalyze the metabolism of L-aspartic acid to L-homoserine were found encoded in the genome of MCy9003, each in two copies. Further, we found one protein homolog of homoserine kinase and two of threonine synthase. The metabolic pathway from L-homoserine to L-threonine is therefore possible. The conversion of threonine to homoserine does not occur in primary metabolism of living organisms. However, the secondary metabolism is mostly unpredictable and might contain enzymes that can 59atalyse the transformation of threonine to homoserine, and therefore this possibility cannot be excluded. The feeding experiments of the isotope labelled L-threonine and L-homoserine should be conducted, and depending on the result obtained, the selected genes encoding the metabolic pathway from L-homoserine to L-threonine would require inactivation to evaluate their

involvement in the biosynthesis of myxoglucamide. Additionally, proteins such as MxgD and MxgR (see below) encoded in myxoglucamide BGC might play a role in biochemical transformation of the potential precursor of the selected amino acid before its modification into an α-vinyl group. Their role is though unknown.

### 2.3.9 Other genes encoded in the myxoglucamide BGC

The *mxgA* and *mxgB* encode an efflux RND transporter, a member of the resistance-nodulation-cell division family (RND) [215], that could likely serve to transport the myxoglucamide out of the bacterial cell. The *mxgD* and *mxgR* gene encode hypothetical proteins with no predicted function. The MxgE is a prolyl oligopeptidase PreP predicted to specifically hydrolyze oligopeptides with less than 30 amino acids [216]. The MxgF is an oxidoreductase from aldo/keto reductase family. Those enzymes catalyze aldehyde or ketone reduction [217]. The *mxgO* gene is predicted to encode a pimeloyl-ACP methyl ester carboxylesterase involved in biotin metabolism. The *mxgP* gene encodes a CoA pyrophosphatase that hydrolyzes the pyrophosphate moiety of coenzyme A [218]. Lastly, the *mxgQ* gene encodes an undecaprenyl phosphate, a key lipid intermediate involved in the synthesis of various bacterial cell polymers [219], which is also a protein target of bacitracin [220]. The direct involvement of above mentioned enzyme in the myxoglucamide biosynthesis has not been explored.

### 2.3.10 The myxoglucamide biological activity

Myxoglucamides share some similarities with biologically active ieodoglucomides A, B and C (Figure 36), secondary metabolites of marine-derived *Bacillus licheniformis* strain 09IDYM23 [221, 222].

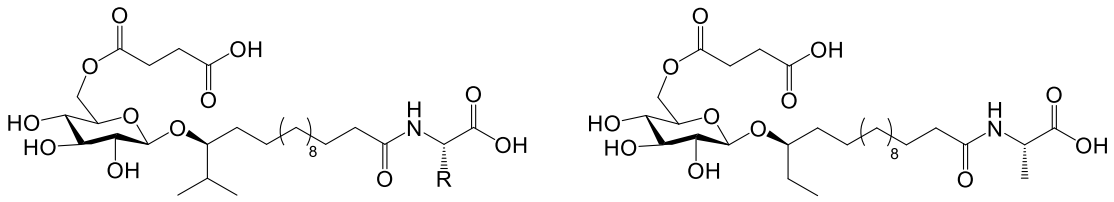



**Figure 36. Ieodoglucamides A-C**

Ieodoglucomide A and B exhibit moderate *in vitro* antimicrobial activity, while ieodoglucomide B additionally shows cytotoxic activity against lung cancer and stomach cancer cell lines [221]. Ieodoglucomide C exhibited good antibiotic properties against Gram-positive and Gram-

negative bacteria as well as antifungal activity against some plant pathogenic fungi [222]. Their BGC is unknown since the genome sequence of producer strain 09IDYM23 is not available. Furthermore, the α-vinyl group in aminoethoxyvinylglycine [213] is reported to have irreversible inhibitory activity against the 1-aminocyclopropane-1-carboxylate synthase involved in the biosynthesis of the senescence hormone ethylene in plants [213, 214]. The same mode of action was reported for rhizobitoxine [214]. On the other hand, methoxyvinylglycine inhibits the growth of bacteria, including *Bacillus subtilis*, *Escherichia coli*, phytopathogen *Erwinia amylovora*, and the human pathogen *Staphylococcus aureus* [214]. Myxoglucamides with α-vinyl group are based on the chemical structure expected to have some inhibitory biological activities against yet unrevealed target. They show no antimicrobial properties but have moderate cytotoxic activity against HCT 116 cell line (Table 8).

**Table 8. Bioactivity testing results for glycolipopeptides**

| | myxoglucamide-645 | myxoglucamide-647 | myxoglucamide-659 | myxoglucamide-661 |
|---|---|---|---|---|
| **Cell type** | **Concentration [µg/ml]** | | | |
| KB 3.1 $IC_{50}$ | >111 | n.d. | 111 | n.d. |
| HCT 116 $IC_{50}$ | 100 | 4.15 | 100 | 85.13 |

n.d. – nor determined

### 2.3.11 Alternative producers of myxoglucamide

Myxoglucamide gene cluster, activated in MCy9003, is present in strains MCy11578 and MCy10644 as well but the myxoglucamides cannot be detected in crude extracts of these wild-type strains. The BGC with partial similarity to MXG BGC is found in genomes of *Pyxidicoccus* sp. MCy8408 and *Cystobacterineae* MCy10597, including only *mxgC*, *mxgG*, *mxgH*, *mxgK*, *mxgL*, *mxgM* and *mxgN* genes (Table 7). No additional information is known about those BGCs and it can be hypothesized that they would most probably be producing different secondary metabolites or only part of myxoglucamide molecule.

## 2.4 Conclusion

Two novel myxobacterial natural products, myxopentacins and myxoglucamides, were identified in *Cystobacterineae* strain MCy9003 using genome mining approach. Myxopentacins were discovered through overexpression of BGC-12 and BGC-1 in the MCy9003 genome that might have an effect on their regulation, however, their effect on myxopentacin biosynthesis cannot be excluded. Myxopentacin production was successfully correlated to BGC-39A after the inactivation of NRPS encoding *mxpA* gene via a plasmid insertion. Myxopentacins are a family of many compounds, most of them still unknown,

containing rare cispentacin and pseudoarginine moieties. Cispentacin biosynthesis is explained based on a phytotoxin coronatine component coronafacic acid and gene knock-out experiments of amipurimycin biosynthetic genes. Cispentacin itself exhibited antifungal activities and is proposed to be a prolyl-tRNA synthetase inhibitor. A pseudopeptide moiety is found in other natural products, namely ketomemicins and arphamenines, recognized as protease and aminopeptidase B inhibitors, respectively. The assembly of myxopentacin's building blocks could be directed by (amido)hydrolases present in the BGC but no further proof is provided. No biological activity against a standard testing panel, covering Gram-negative and positive bacteria as well as fungi, was observed. On the other hand, the myxoglucamides were discovered after the simultaneous double promoter insertion in front of *mxgH* and *mxgI* genes via homologous recombination. Three target masses were visibly increased in the mutant strain M172. The isolated compounds are glycolipopeptides with a rare α-vinyl group. Myxoglucamide biosynthesis is proposed but the identity of incorporated amino acids, the exact function of each oxidation domain and a responsible glycosyltransferase are not yet confirmed. They exhibit moderate cytotoxicity but no antibacterial or antifungal biological activities were detected. In cooperation with Alexander Popoff, we discovered two novel myxobacterial natural products with rare features using genome mining approach. Even though their antibacterial biological activities are not determined yet, their biosynthesis raises many questions and will enlighten us when understood.

## 2.5 Supplementary information

### Table of Tables



### Table of Figures

**Table S1- 1. Oligonucleotides used in that study**

| Primer | 5'-Sequence-3' | Restr. sites | Construct |
|---|---|---|---|
| tn5-HindIII-F | ATATAAAGCTTGCCCTTTGGACAGCAAGCGAAC | *Hin*dIII | pSB2tn5Kan |
| tn5-R | ATATAGAATTCATGCTGTACCTCCTTATCCTGTCTCTTGATC | *Eco*RI | |
| 001bgc18-NdeI-F | ATATATCATATGAGCCCCAACCTGATCGCG | *Nde*I | pSB2tn5Kan001bgc18 |
| 001bgc18-EcoRI-R | TATATGAATTCTTGCCTCCGGTGGTGATGCA | *Eco*RI | |
| 001-inact-nrps39A-F | GTGTCGGGCGCTCGGTC | / | pCR2.1 001inact39Ao1 |
| 001-inact-nrps39A-R | CACGGATCTTCACCTGGAGGTC | / | |
| 001nrps12-NdeI-F | ATATACATATGGTGCCGGTGCCCC | *Nde*I | pSBtn5Kan001nrps12 |
| 001nrps12-EcoRI-R | ATATAGAATTCGAGAGGTCGTTCCTCGCCACAC | *Eco*RI | |
| 001-inact-nrps12F | ACGAGTCGCTGCGCACG | / | pCR2.1 001 inact-nrps12 |
| 001-inact-nrps12R | CAGCAGCCGCGTGTAATGC | / | |
| 001nrps1-NdeI-F | ATATACATATGAAGCCCTCCGTTGCCGC | *Nde*I | pSBtn5Kan001nrps1 |
| 001nrps1-EcoRI-R | ATATAGAATTCAGAGACGATGAGGGTGCCCTC | *Eco*RI | |
| 001-inact-nrps1F | CAGCTCTTCCAGCATCAGACG | / | pCR2.1 001-inact-nrps1 |
| 001-inact-nrps1R | CAGCTCTTCCAGCATCAGACG | / | |
| 001nrpst1pks33-NdeI-F | ATATACATATGTCCACCGCGCGC | *Nde*I | pSBtn5Kan001nrps33 |
| 001nrpst1pks33-EcoRI-R | ATATAGAATTCAGACGATCAGCGTCGCCTC | *Eco*RI | |
| 001nrps34B-NdeI-F | ATATACATATGGATGGGAACATCGCTTCTC | *Nde*I | pSBtn5Kan001nrps34B |
| 001nrps34B-EcoRI-R | ATATAGAATTCAGGTGGCCGATGTTCGTC | *Eco*RI | |
| 001-inact-nrpst1pks39B-F | GAGCTGATGGACGCCGAGC | / | pCR2.1 001-inact-nrpst1pks39B |
| 001-inact-nrpst1pks39B-R | GAGCGCTGGGCCCAGTAGC | / | |
| 001nrps35BGrsT-NdeI-F | ATATACATATGGTGGCCCAACCCTGG | *Nde*I | pSBtn5Kan001nrps35BGrsT |
| 001nrps35BGrsT-EcoRI-R | ATATAGAATTCGTGCACAGCTCGCCTGAC | *Eco*RI | |
| 001bgc25-NdeI-F | ATACATATGGCCGCGGAGTCGTCCAC | *Nde*I | pSBtn5Kan001bgc25 |
| 001bgc25-EcoRI-R | TATATGAATTCTCCAGCACGGCGCACAGC | *Eco*RI | |

| Primer | 5'-Sequence-3' | Usage |
|---|---|---|
| M13-21F | TGTAAAACGACGGCCAGT | For general plasmid integration verifications |
| M13-24R | CGGATAACAATTTCACACAGG | |
| vanseq-F | CATTCGTAACGCCAAGGTCTTC | |
| g001-12F | CTCCCACTCGACATGAA | Alignment on the genome, outside of 001nrps12 homology |
| g001-12R | GTCGAAGACATCGGTGTTGTA | |
| g001-1F | CTCCACCAACATCCATTGGAG | Alignment on the genome, outside of 001nrps1 homology |
| g001-1R | GAGCGAGTGTCCGCAG | |
| g001-bgc18-F | GTCGATCTGCCAGAAGAGA | Alignment on the genome, outside of 001bgc18 homology |
| g001-bgc18-R | CGACGATGCGCAATTCATG | |

| | | |
|---|---|---|
| g001-33F<br>g001-33R | CTTCCTTTATTCCACGTTGATCG<br>CAGCCGATGAGGAGCTG | Alignment on the genome, outside of 001nrpst1pks33 homology |
| g001-34B-F<br>g001-34B-R | CTTGAATAAGCGGCCTGTC<br>GAATCTGCCGGTGCTTGAG | Alignment on the genome, outside of 001nrps34B homology |
| g001-35BGrsT-F<br>g001-35BGrsT-R | GACCTTCTAGCACGTTCACT<br>GAAGAGGCAGCCGAAGAAG | Alignment on the genome, outside of 001nrps35BGrsT homology |
| g001-bgc25-F<br>g001-bgc25-R | AATGATGGTTCTCCAGGTTCTC<br>TTCTGTGTGTTCGCGACGAT | Alignment on the genome, outside of 001bgc25 homology |
| PCR21R | ATAGGGCGAATTGGGCCCT | Reverse primer on the backbone of pCR2.1 vector |
| G001-inact12Fo2<br>G001-inact1Fo2<br>G001-inact39Ao1<br>G001-inact39Bo2 | CTGGTTCCACTCCACCAGCAC<br>GTGGCTCAGGGCGAAGAGA<br>CCATCTCCAGCGCGATGAC<br>GATGTTCGGGTACGCATGCAG | Aligning on the genome located in front of pCR2.1-000inact-based plasmids |

**Table S1- 2. Genetic constructs used in that study**

| Plasmid name | Details | Reference |
|---|---|---|
| pCR2.1-TOPO | ori(pUC), ori(f1), kanamycin resistance (kanR), ampicilin resistance (ampR) | TOPO cloning kit |
| pSBtn5Kan | ori(pUC), ori(f1), kanR, ampR, P*tn5* – for gene activation by single homologous recombination, homology arm is inserted through *Nde*I and *Eco*RI restriction sites | [158] |
| pSB2tn5Kan | ori(pUC), ori(f1), kanR, ampR, P*tn5* (2x) - pSBtn5Kan vector with additional *tn5* promoter, for activation of genes in both directions, homology arm is inserted through *Nde*I and *Eco*RI restriction sites | This study |
| pSB2tn5Kan001bgc18 | for activation of *mxgH* and *mxgI* genes; 1.2 kb fragment from the start of *mxgI* gene is cloned into pSB2tn5Kan | This study |
| pCR2.1 001inact39Ao1 | 1 kb homology from the first A domain of *mxpA* gene cloned into pCR2.1 | This study |
| pSBtn5Kan001nrps12 | 1.2 kb fragment from the start of MCy9003_22370 gene cloned into pSBtn5Kan | This study |
| pCR2.1 001inact-nrps12 (o2) | 1.1 kb fragment starting at the second C domain in MCy9003_22370 gene cloned into pCR2.1 vector in reverse orientation | This study |
| pSBtn5Kan001nrps1 | 1 kb fragment from the start of MCy9003_00740 gene cloned into pSBtn5Kan | This study |
| pCR2.1 001-inact-nrps1 | 1.1 kb fragment comprising E domain from the second module of MCy9003_00740 gene cloned into pCR2.1 | This study |
| pFPVan001nrps35BGrsT | 1.1 kb fragment starting with MCy9003_80190 gene cloned into pFPVan | This study |
| pSBtn5Kan001nrps33 | 1 kb fragment from the start of MCy9003_77130 gene cloned into pSBtn5Kan | This study |
| pSBtn5Kan001nrps34B | 1.1 kb fragment from the start of MCy9003_78330 gene cloned into pSBtn5Kan | This study |
| pCR2.1 001inactnrps39B | 1 kb fragment comprising first AT domain in MCy9003_82470 gene cloned into pCR2.1 vector in reverse orientation | This study |
| pSBtn5Kan001bgc25 | 0.75 kb fragment from the start of the MCy9003_57920 gene cloned into pSBtn5Kan | This study |

**Table S1- 3. Myxobacterial strains used or prepared in this study**

| Organism | Genotype or usage purpose | Reference |
|---|---|---|
| *Cystobacterineae* sp. SBAr001 | 45 biosynthetic gene clusters | MCy9003 |
| *Cystobacterineae* sp. SBCy024 | alternative producer of myxopentacin | MCy10644 |
| *Myxococcus* sp. SBMx268 | alternative producer of myxopentacin | MCy11578 |
| M23-M25 | MCy9003/pSBtn5Kan001nrps12 | This study |
| M37 | MCy9003/pCR2.1 001inact-nrps12 | This study |
| M41-M42, M67 | MCy9003/pSBtn5Kan001nrps1 | This study |
| M58, M59 | MCy9003/pFPVan001nrps35BGrsT | This study |
| M108-M112 | MCy9003/pSBtn5Kan001nrps33 | This study |
| M115-M117 | MCy9003/pSBtn5Kan001nrps34B | This study |
| M127, M128, M130 | MCy9003/pCR2.1 001inact-nrps1 | This study |
| M137, M138 | MCy9003/pCR2.1 001inactnrps39B | This study |
| M172-M175 | MCy9003/pSB2tn5Kan 001bgc18; myxoglucamide producer | This study |
| M208 | MCy9003/pSBtn5Kan001bgc25; chloromyxamide producer | This study |
| M215 | MCy9003/pCR2.1 001inact39Ao1; Δmyxopentacins | This study |

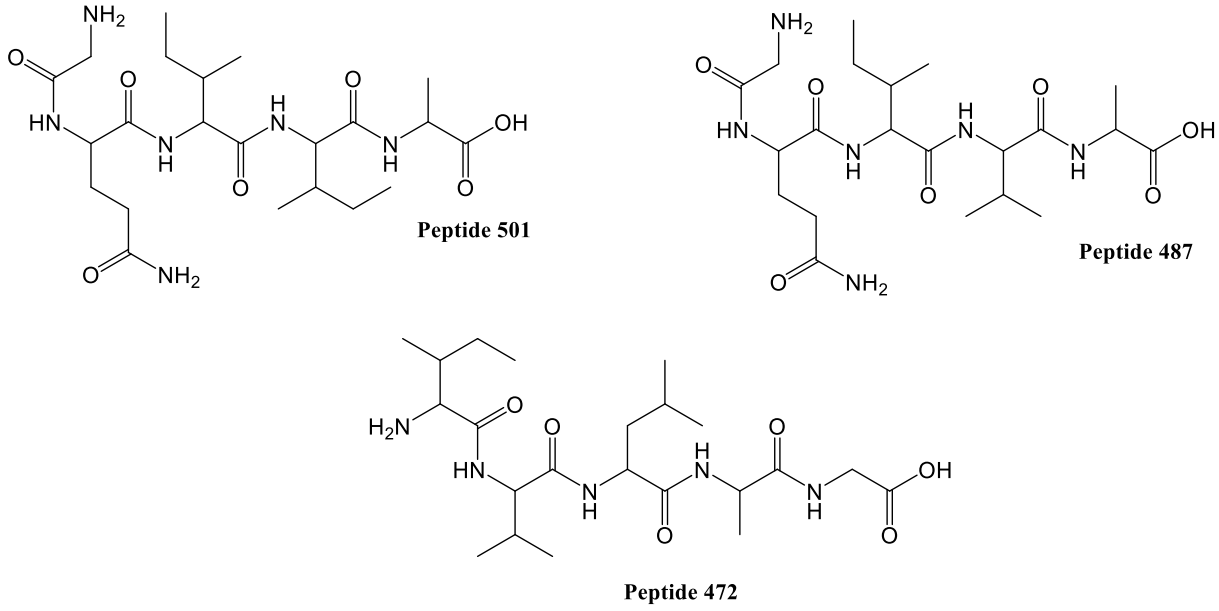



**Figure S1- 1. Plane structures of peptides 501, 487 and 472 isolated from MCy9003 mutant strains.** The upregulation of biosynthetic genes encoded in BGC-33 and 34B (M108 and M115: Table S1- 3) resulted in production of peptides 501 and 487, and peptide 472, respectively. More peptides were produced by those two mutant strains but the structure elucidation of peptides 619, 691, 917 and 743 did not succeed due to their poor solubility [178]. According to the A domain selectivity predictions found in the respective BGCs, none of the above shown peptides are complete.

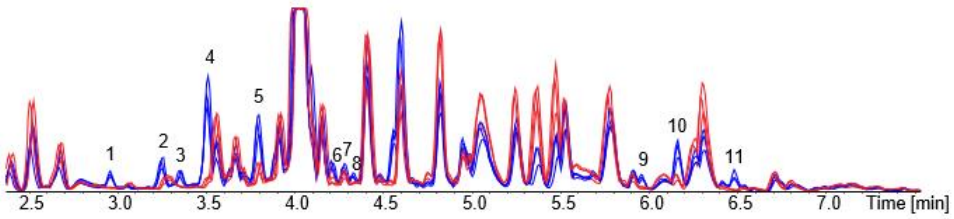



**Figure S1- 2. The secondary metabolome of WT strain and mutant strain with upregulated BGC-12.** The base peak chromatograms (BPC) of wild type (red) and mutant strain MCy9003/pSBtn5Kan001nrps12 (blue) crude extracts show increase of 11 numbered peaks. The peak numbers correspond to target masses: 1- 290.6950 m/z $[M+2H]^{2+}$; 2- 296.6952 m/z $[M+2H]^{2+}$; 3- 311.7003 m/z $[M+2H]^{2+}$; 4- 381.7477 m/z $[M+2H]^{2+}$; 5- 388.7557 m/z $[M+2H]^{2+}$; 6- 403.7608 m/z $[M+2H]^{2+}$; 7- 395.7636 m/z $[M+2H]^{2+}$; 8- 407.7636 m/z $[M+2H]^{2+}$; 9- 648.4336 m/z $[M+H]^{+}$; 10- 460.2383 m/z $[M+H]^{+}$; 11- 522.3126 m/z $[M+2H]^{2+}$.

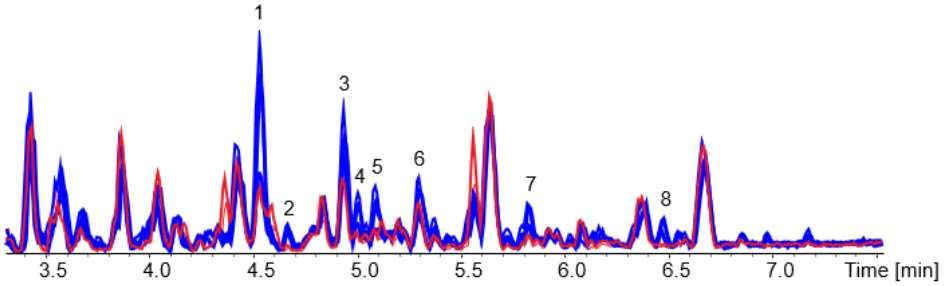



**Figure S1- 3. The secondary metabolome of WT strain and mutant strain with upregulated BGC-1.** The base peak chromatograms (BPC) of wild type (red) and mutant strain MCy9003/pSBtn5Kan-001nrps1 (blue) crude extracts show increase of numbered 8 peaks. The peak numbers correspond to target masses: 1- 457.29 m/z $[M+2H]^{2+}$; 2- 677.42 m/z $[M+H]^{+}$; 3- 464.3 m/z $[M+2H]^{2+}$; 4- 512.83 m/z $[M+2H]^{2+}$; 5- 443.27 m/z $[M+2H]^{2+}$; 6- 568.37 m/z $[M+2H]^{2+}$; 7- 554.35 m/z $[M+2H]^{2+}$; 8- 522.32 m/z $[M+2H]^{2+}$.

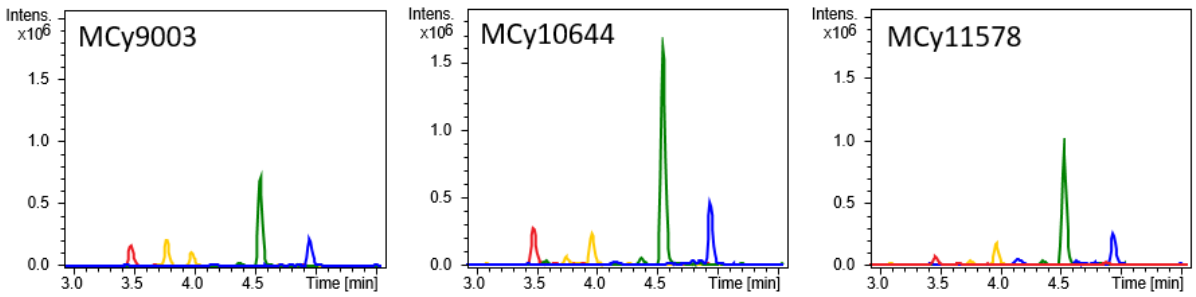



**Figure S1- 4. The production of myxopentacins in MCy9003 and alternative producers MCy10644 and MCy11578.** EICs of m/z 381.7477 (±0.02) (red), m/z 388.7557 (±0.02) (yellow), m/z 457.29 (±0.02) (green) and m/z 464.3 (±0.02) (blue) are shown (chromatograms adapted from DataAnalysis (available from Bruker Daltonics)).

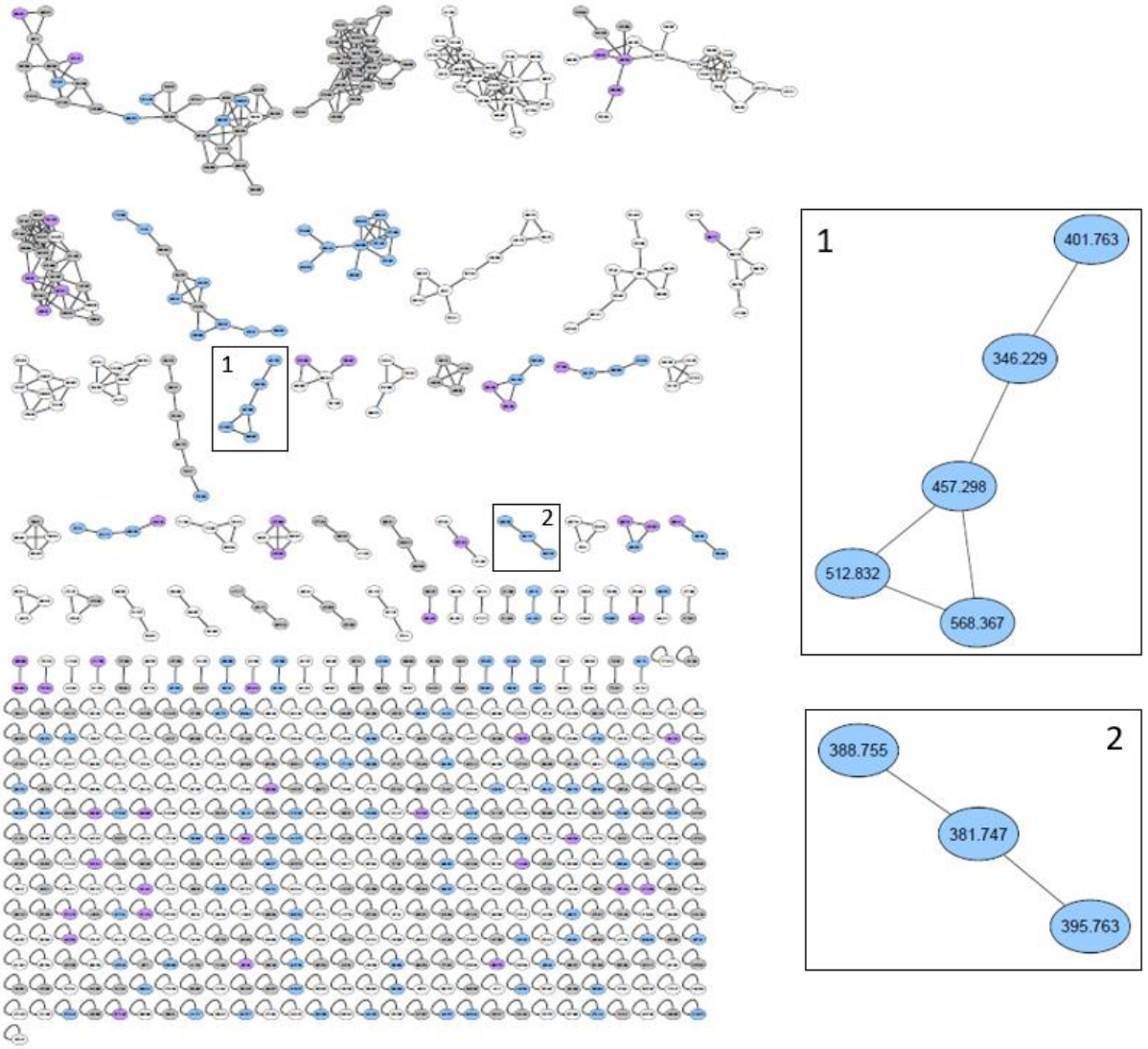



**Figure S1- 5. GNPS molecular networking for MCy9003, MCy10585 and MCy10653 strains.** Confirmed myxopentacin producer MCy9003 was compared with two myxobacterial strains with different arrangement of myxopentacin biosynthetic genes (Table 6: Group C; Figure 29). Marked grey are molecular masses found in medium crude extract, blue are molecular masses found only for MCy9003 strain and purple denotes shared molecular masses between all three analyzed strains. White are molecular masses found in MCy10585 and MCy10653, either only in one or in both. The LC-MS peaks of compounds with molecular masses shown in cluster 2 were increased after the upregulation of BGC-12 (Figure S1- 2), while the cluster 1 contains some increased when upregulating BGC-1 (Figure S1- 3). Only the parameters that were modified will be listed: Precursor Ion Mass Tolerance: 0.05 Da; Fragment Ion Mass Tolerance: 0.1 Da; Min Pairs Cos: 0.75; Minimum Cluster Size: 1.


**Figure S1- 6. The alignment of C and PCP domains from MxpA with key residues for catalytic activity.** The C domains in NRPS modules of myxopentacin were compared to amino acid sequence SrfA-C from PDB: 2VSQ [223] with thoroughly annotated core motives for each of the domains. The core motif in C domains is HHXXXDG [224] and [I/L]GG[D/H]SL in PCP domains [225]. The key residue is marked with an asterix (*).

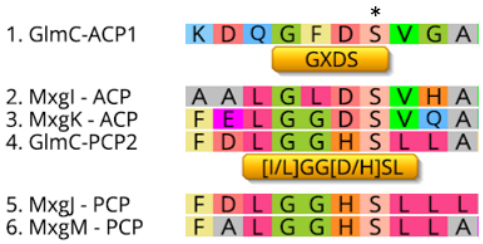



**Figure S1- 7. The alignment of ACP and PCP domains of myxoglucamide BGC with key residues for catalytic activity.** The core motif in ACP domains is GXDS [88] and [I/L]GG[D/H]SL in PCP domains [225]. The key residue is marked with an asterix (*).


**Figure S1- 8. The alignment of DH domain from MxgK with erythromycin dehydratase.** Erythromycin dehydratase was found in PDB under accession code 3EL6 [226]. The conserved motifs are listed in orange boxes. The deviation of some residues suggests that the domain is not functional.

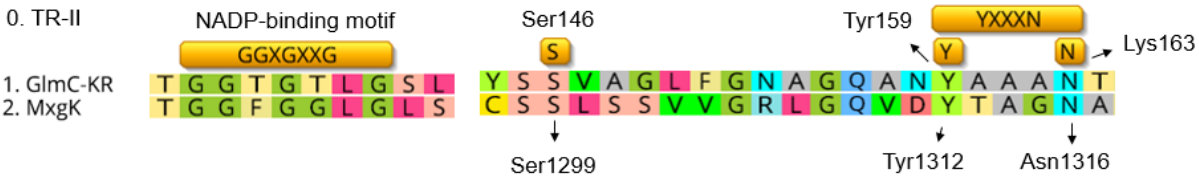



**Figure S1- 9. The catalytic sites of myxoglucamide KR domain.** The conserved motifs and amino acid residues forming catalytic triad of tropinone reductase-II (TR-II) (PDB: 2AE2) [205] are listed in orange boxes. Note that TR-II belongs among classical SDRs that contain the catalytic Lys163 residue, which is often exchanged by Asn residue in modular KR domains from the complex SDR family [84, 85]. The Lys to Asn exchange retains the domains activity [85]. The important amino acid residues from TR-II were found in KR domain of globomycin BGC and then aligned with myxoglucamide MxgK, containing a KR domain.

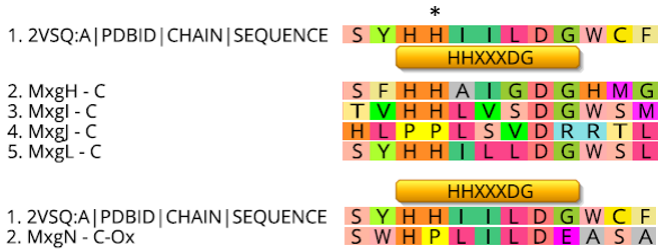



**Figure S1- 10. The alignment of C domains of myxoglucamide BGC with key residues for catalytic activity.** The C domain in NRPS modules of myxoglucamide were compared to amino acid sequence SrfA-C from PDB: 2VSQ [223] with thoroughly annotated core motives for each of the domains. The key residue is marked with an asterix (*). There is no functional C domain in MxgJ and MxgN proteins [224].


**Figure S1- 11. The important catalytic residues of KS domains.** The active sites T209, C211, S315, H346, K379, N381, H384, Q386, N455, H457 from non-iterative EryA1 and iterative LovB KS domain [82] are marked with the orange rectangles. Additionally marked is the residue where a difference in MxgK-KS domains is observed. The Q386 residue is believed to enhance the basicity of H384 [82]. As negatively charged Glu (E) residue can be found in its place in iterative KS (LovB), presence of other negatively charged amino acid Asp (D) should not change the catalytic environment and therefore the MxgK-KS is proposed to be functional.

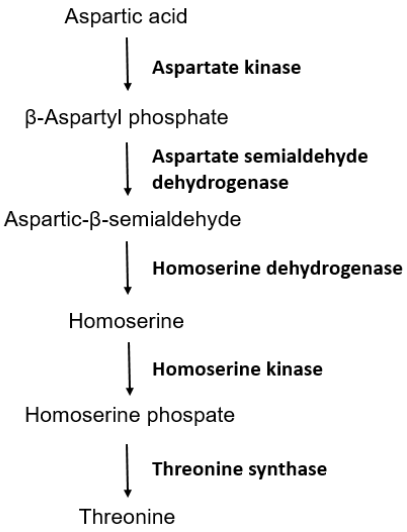



**Figure S1- 12. The aspartic acid, homoserine and threonine metabolism**

# 3 Characterization of the biosynthesis of the antibiotic globomycin active against LspA

## 3.1 Introduction

Globomycin is an antimicrobial compound, effective against Gram-negative bacteria [227], discovered in 1977 in Japan [228]. It is produced by a few *Streptomyces* strains [228], namely *S. halstedii* 13912, *Streptoverticillium cinnamoneum* 15037, *S. neohygroscopicus* subsp. *globomyceticus* 15631, *S. hagronensis* 17834, *Streptomyces hygroscopicus* SF-1902 [229] and *Streptomyces hagronensis* strain 360, NRRL 15064 [230]. A cyclic depsipeptide consists of six building blocks: L-serine, L-*allo*-threonine, glycine, N-methylleucine, L-*allo*-isoleucine and 3-hydroxy-2-methylnonanoic acid, with L-*allo*-threonine, L-*allo*-isoleucine and 3-hydroxy-2-methylnonanoic acid at the time of discovery being novel components of NPs [231]. Its total synthesis was achieved in 2000 [232] but globomycin biosynthetic machinery is not described in the literature. Globomycin is known to inhibit signal peptidase II LspA, a key enzyme involved in post-translational modifications of lipoproteins in many pathogenic bacteria but has no equivalent in humans [139], which makes it a potential drug target. Although LspA is not essential for viability of some bacteria such as a methicillin-resistant *Staphylococcus aureus* (MRSA), it is vital for survival of that bacterium in human blood [140].

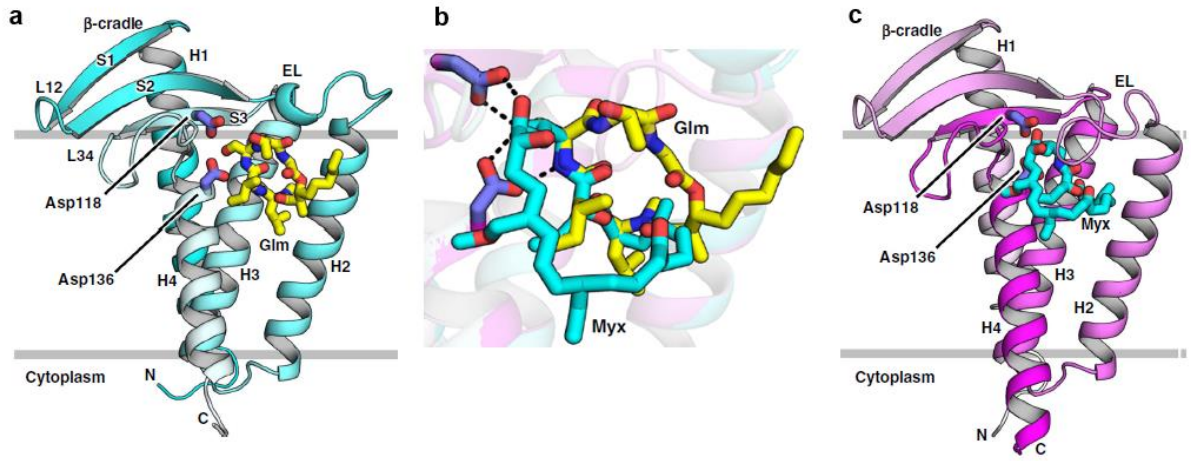



**Figure 37. Structures of LspA from methicillin-resistant *S. aureus* (MRSA) in complex with globomycin and myxovirescin.** The views into the binding pocket containing globomycin (**a**) and myxovirescin (**c**). A superpose of both antibiotics in the binding pocket is shown in **b**. The figure was taken from the publication by Olatunji *et al.* in 2020 [140].

The globomycin binding to the LspA from MRSA was shown with a complex structure [140]. The antibiotic sterically blocks the LspA active site, which leads to the accumulation of

prolipoprotein [233] and compromises the integrity of the bacterial cell envelope. The same mode of action is characteristic of the antibiotic myxovirescin produced by *Myxococcus xanthus* DK1622 [54]. Even though myxovirescin is of PKS origin [234] and globomycin of NRPS origin, they both contain a lipophilic part, important for insertion into the membrane where the LspA target protein is located (Figure 37) [140].

Globomycin is special for its exclusive Gram-negative biological activity [228] but it was shown that its derivatives with longer alkyl side chain have improved Gram-negative activity and additionally show moderate activity against Gram-positive bacteria including methicillin-resistant *Staphylococcus aureus* (MRSA) [227]. The natural analog SF-1902 $A_5$ is produced by *Streptomyces hygroscopicus* SF-1902 as a minor component [235] while a synthetic analog was prepared by total synthesis [227] (Figure 38).

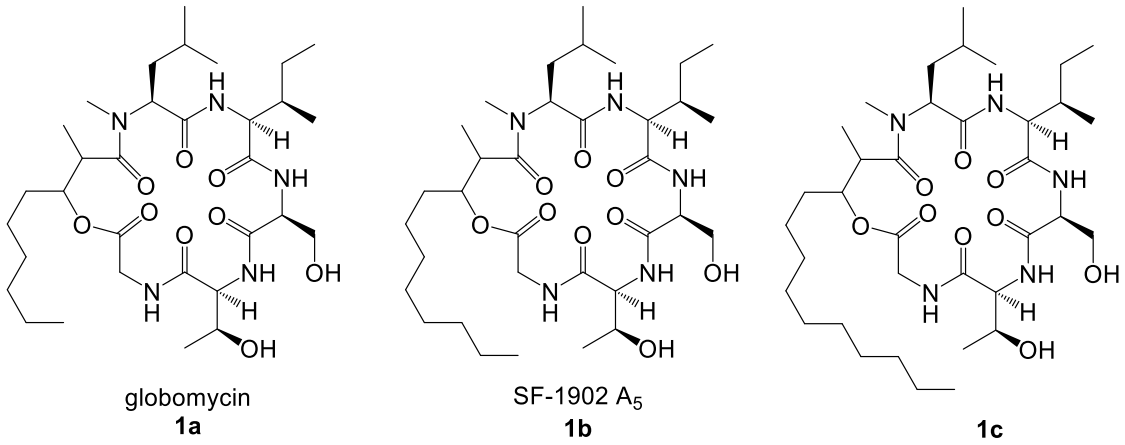



**Figure 38. The structure of globomycin and its derivatives.** The derivatives of globomycin (**1a**): SF-1905 $A_5$ (**1b**) and a synthetic derivative (**1c**) exhibit improved Gram-negative activity, while a synthetic derivative (1c) shows additional antibacterial activity against Gram-positive bacteria [227]. They differ only in the length of the fatty acid chain. The fatty acid chain of **1b** and **1c** have additional two and four carbon atoms, respectively, more than the globomycin's.

## 3.2. Materials and methods

### 3.2.1 Cultivation of *Streptomyces hagronensis*

*Streptomyces hagronensis* strain 360, NRRL 15064 [236] was used to produce globomycin. Spores were generated on ISP2 medium (1 % malt extract, 0.4 % yeast extract, 0.4 % glucose, 1.8 % agar, $dH_2O$, pH 7.3) agar plates at 30 °C for 10–14 days, and were preserved in 20 % glycerol at −80 °C. For globomycin production, a 1 % spore inoculum was prepared in SHG-V medium (1 % glycerol, 1 % Bacto™ peptone, 1 % Bacto™ yeast extract, $dH_2O$, pH 7.0),

cultivated at 28 °C, at 220 rpm for 3 days [236]. The seed culture was further inoculated (5 %) into SOYF6 medium (4.5 % glycerol, 2 % Bacto™ beef extract, 1 % soy flour (Hensel), 0.2 % $CaCO_3$, $dH_2O$, pH 6.0 adjusted with KOH; adapted from [228]) and incubated at 26 °C, at 220 rpm for 5 days. Smaller-scale globomycin production was done in 20 mL or 50 mL medium in 100 mL round Erlenmeyer flasks or in 300 mL baffled Erlenmeyer flasks, respectively. For genetic modification purposes, the susceptibility of *S. hagronensis* to different antibiotics was determined. Bacterium is resistant to nalidixic acid (25 µg $mL^{-1}$) and still grows on erythromycin (25 µg $mL^{-1}$), spectinomycin (25 µg $mL^{-1}$) and kanamycin (50 µg $mL^{-1}$). It is sensitive to apramycin (25 µg $mL^{-1}$), streptomycin (50 µg $mL^{-1}$) and thiostrepton (30 µg $mL^{-1}$).

### 3.2.2 Extraction and quantification of globomycin

To 500 µL of culture broth of *S. hagronensis*, 1 mL of methanol was added. The sample was rotated for 1 h and then centrifuged at 15000 rpm for 10 min at 15 °C. An UHPLC system UltiMate 3000 LC (Dionex) with 1.7 µm Acquity UPLC BEH C-18 Column (2.1 mm × 100 mm) coupled with an electrospray ionization (ESI) source linked to an amaZon speed MS (Bruker Daltonics; 3D ion trap)) was used for quantification. Separation of a 1 µL sample was achieved using a linear gradient (56–57 % B in 7.5 min) from (A) ultrapure water + 0.1 % FA to (B) acetonitrile + 0.1 % FA at flow rate of 0.600 mL $min^{-1}$ at 45 °C. For precise mass determination (HRMS) samples were measured on a UHPLC coupled to maXis 4G MS system (Bruker Daltonics; Q-ToF). Separation of a 1 µL sample, with 1.7 µm Acquity UPLC BEHC-18 Column (2.1 mm × 50 mm), was achieved using a linear gradient (5-95 % B in 6 min) from (A) ultrapure water + 0.1 % FA to (B) acetonitrile + 0.1 % FA at a flow rate of 0.600 mL $min^{-1}$ at 45 °C. Absorption spectra (200-600 nm) were recorded by a diode array detector. Mass spectra were acquired in centroid mode ranging from 200 to 2000 m/z (amaZon) or from 150 to 2500 m/z (maXis 4G) in positive ionization mode with auto MS-MS fragmentations.

### 3.2.3 Globomycin purification

After the completion of fermentation, 5 % XAD-16 resin was added to the culture broth and left shaking on rotary shaker for a day. The culture broth was centrifuged at 8300 × g for 10 min at RT and DCM was added to the cell pellet with XAD-16. The DCM extract was concentrated on a rotary evaporator (Rotavapor R210, Büchi) at 50 °C at reduced pressure and washed once with 500 mL 0.05 M HCl and 2 % $NaHCO_3$ and twice with 200 mL of brine followed by drying with $Na_2SO_4$ [237]. The organic phase was dried, fractionated on silica by flash chromatography using a linear gradient from 0 to 100 % B using (A) hexane + 0.1 %

formic acid (FA) and (B) 80 % EtOAc, 20 % MeOH + 0.1 % FA. Fractions containing globomycin were subjected to flash chromatography once again, this time using (A) EtOAc + 0.1 % FA and (B) MeOH + 0.1 % FA with a linear gradient from 1 % to 20 % B. Obtained samples were purified using preparative LC-MS (Dionex Ultimate 3000 coupled with Bruker High capacity trap mass spectrometer). Separation of the sample was carried out on a 5 µm EVO C18 100 Å LC column (250 mm × 10.0 mm, Kinetex) using a linear gradient (5–95 % B in 31 min) from (A) ultrapure water + 0.1 % FA to (B) acetonitrile + 0.1 % FA and on a 5 µm Biphenyl 100 Å LC column (250 mm × 10.0 mm, Kinetex) with a linear gradient of 42–43 % B in 22 min at a flow rate of 5 mL $min^{-1}$ and 45 °C.

### 3.2.4 Sequencing of *S. hagronensis* genome and assembly of globomycin BGC

Genomic DNA of *Streptomyces hagronensis* strain 360, NRRL 15064 [236] was isolated using phenol/chloroform according to the standard protocol [160] and was sent for Illumina sequencing. The sequencing reads were assembled by Nestor Zaburannyi into 87 contigs. The encoded proteins were compared to the BLAST database (Table 9), which uncovered the genome similarity to bacterial strain *Streptomyces* sp. DvalAA-43 with partial genome sequence. Due to different contigs from *S.* sp. DvalAA-43 the NRPS gene of globomycin was elongated. As the globomycin BGC was still not complete, *S. hagronensis* cells were sent to PacBio sequencing. The received sequencing reads were again assembled by Nestor Zaburannyi and completed the sought BGC.

### 3.2.5 Conjugation of *S. hagronensis*

The vectors pKC1132, pKC1139 and pSET152 were used for intergeneric conjugation between *E. coli* and *S. hagronensis*. Methylation-deficient strain *E. coli* ET12567/pUZ8002 was used as the *E. coli* donor strain. Its overnight culture was reinoculated (5 %) into 20 mL fresh 2TY medium containing Cm (10 µg $ml^{-1}$), Kn (25 µg $ml^{-1}$), Apr (50 µg $ml^{-1}$) antibiotics. After cultivation at 37 °C the $OD_{600}$ = 0.4 was reached and the culture was washed twice and finally resuspended in 1 mL. Spores of *S. hagronensis*, previously stored at -80 °C (approx. CFU = $10^8$), were washed twice with 2TY medium and centrifuged for 5 min at 6000 rpm. Spores resuspended in 0.5 mL 2TY medium are incubated 10 min at 52 °C and 5 min on ice. A total of 400 µL of spores and 1 mL of the *E. coli* donor strain is sufficient for 4 reactions. The 350 µL of conjugation mixture is spread on well dried GYM MgCl agar plates (0.4 % glucose, 0.4 % yeast extract, 1 % malt extract, 0.2 % $CaCO_3$, 1.8 % agar, $dH_2O$, pH 7.2; 10 mM MgCl after autoclaving; according to DSMZ). After 16 h, the plates were overlaid with nalidixic acid (25

μg $ml^{-1}$) and after an additional 8 h with apramycin (25 μg $ml^{-1}$). After 6-10 days at 30 °C, the number of exconjugants was determined.

### 3.2.6 Construction of plasmids for genetic manipulation of *S. hagronensis*

Routine handling of nucleic acids, such as isolation of plasmid DNA, restriction endonuclease digestions, DNA ligations, and other DNA manipulations, was performed according to standard protocols [160]. *E. coli* DH10β (Invitrogen) was used as host for standard cloning experiments and *E. coli* ET12567/pUZ8002 as a donor strain for bacterial transformation [27]. *E. coli* strains were cultured in 2TY medium or on 2TY agar (1.6 % tryptone, 1 % yeast extract, 0.5 % NaCl, (1.8 % agar), $dH_2O$) at 30-37 °C, (200 rpm) overnight. Antibiotics were used at the following final concentrations: 50 μg $ml^{-1}$ apramycin, 25 μg $ml^{-1}$ kanamycin, 10 μg $ml^{-1}$ chloramphenicol. Transformation of E. coli strains was achieved via electroporation in 0.1 cm-wide cuvettes at 1350 V, 200 Ω, 25 μF. Plasmid DNA was purified by standard alkaline lysis [160] or by using the GeneJet Plasmid Miniprep Kit (Thermo Fisher Scientific). Restriction endonucleases, alkaline phosphatase (FastAP) and T4 DNA ligase were obtained from Thermo Fisher Scientific. Oligonucleotides used for PCR and sequencing were obtained from Sigma-Aldrich and are listed in Table S2- 1-Table S2- 2. PCR reactions were carried out in Mastercycler® pro (Eppendorf) using Phusion™ High-Fidelity or Taq DNA polymerase (Thermo Fisher Scientific) according to the manufacturer's protocol. For Taq: Initial denaturation (2 min, 95 °C); 30 cycles of denaturation (30 s, 95 °C), annealing (30 s, -5 °C of lower primer's $T_m$) and elongation (based on PCR product length 1 kb/min, 72 °C); and final extension (10 min, 72 °C). For Phusion™: Initial denaturation (2 min, 98 °C); 30 cycles of denaturation (15 s, 98 °C), annealing (30 s, $T_m$ of the lower primer) and elongation (based on PCR product length 0.5 kb/min, 72 °C); and final extension (10 min, 72 °C). PCR products or DNA fragments from restriction digests were purified by agarose gel electrophoresis and isolated using NucleoSpin Gel and PCR Clean up kit (Macherey-Nagel). The PCR products were cloned into suitable plasmids pCR2.1 and pCK1132, selected by restriction analysis and sequenced using sequencing primers (Table S2- 2). Details on the construction of all plasmids used and generated in this study are given in Table S2- 3.

### 3.2.7 Disruption and overexpression of globomycin biosynthetic genes

Plasmid pKC1132 is a suicide vector for *Streptomyces* containing *oriT* RK2 for conjugation from *E. coli* to *Streptomyces*, here used for generating *S. hagronensis* mutants with disrupted selected genes. Firstly, the 0.84-kb DNA fragment containing n-MT domain from GlmC was

amplified by PCR with globo-inact-F and globo-inact-R primer pair, adenylated and ligated into pCR2.1. The sequencing confirmed plasmid pCR2.1 inact-globo1 was digested with *Pvu*II and ligated into pKC1132 through the same restriction site, generating pKC1132 inact-globo1. Sequencing confirmed plasmid was used for conjugation of *S. hagronensis*. No exconjugants were obtained, therefore a larger homology was selected. The 2.3-kb DNA fragment as well containing n-MT domain from GlmC was amplified by PCR with inact-globo-HindIII-F and inact-globo-EcoRV-R primer pair. The *Hin*dIII/*Eco*RV digested DNA fragment was ligated into pKC1132 plasmid to generate pKC1132 inact-globo2.3. The plasmid was sequencing confirmed using PKC1132-F, PKC1132-R and check-inact-globo-R primers. The former was conjugated into *S. hagronensis* and incorporated into the genome via homologous recombination. Similarly, the 2.1-kb DNA fragment from GlmF was PCR amplified and cloned in between *Hin*dIII/*Eco*RV sites into pKC1132 plasmid, generating pKC1132 SHG-inactA8. The plasmid sequence correctness was confirmed with PKC1132-F, PKC1132-R, check-inactA8-F and check-inactA8-R primers. Conjugation and secondary metabolite production profile comparison with WT strain followed. Used and prepared bacterial strain are described in Table S2- 4.

### 3.3 Results and discussion

#### 3.3.1 Analysis of globomycin biosynthetic gene cluster

To verify that the assembled globomycin BGC is responsible for globomycin biosynthesis, the *glmC* gene encoding a NRPS was inactivated. A plasmid pKC1132inact-globo2.3 was inserted via homologous recombination to disrupt the NRPS gene since the NRPS catalyzes peptide biosynthesis as an assembly-line similarly to modular PKS system [68]. The disruption was successful and yielded 17 exconjugants of which 8 sporulated. Four mutant strains and the wild-type strain were analysed for globomycin production. All *S. hagronensis* pKC1132inact-globo2.3 strains failed to produce globomycin as shown in Figure 39. The genes from globomycin BGC cluster are presented in Table 9.

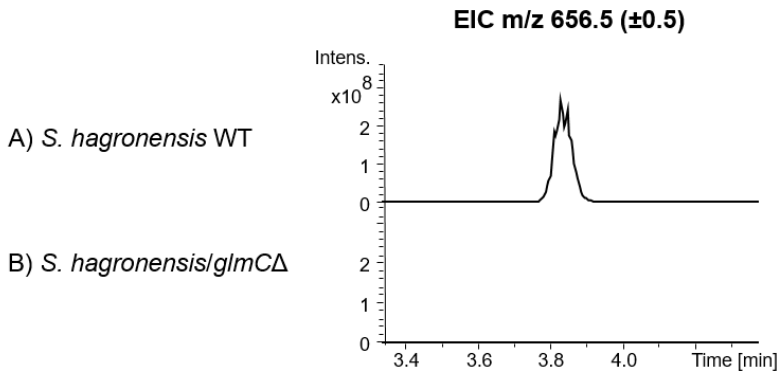



**Figure 39. The inactivation of globomycin BGC.** EIC chromatograms for m/z 656.50 (±0.2) represent globomycin production in *S. hagronensis* WT (A) and lack of it in *S. hagronensis*/pKC1132inact-globo2.3 strain with disrupted *glmC* gene encoding for NRPS (B).

**Table 9. The annotated genes found in globomycin BGC**

| *Gene* | *Size*[a] | *Protein homolog*[b] | *Proposed function* | *Catalytic domains* |
|---|---|---|---|---|
| *orf(-7)* | 305 | WP_093538839.1 (98/99) | DUF4232 domain-containing protein | |
| *orf(-6)* | 408 | WP_093538840.1 (100/100) | ParA family protein | |
| *orf(-5)* | 408 | WP_093538841.1 (98/98) | ParB/RepB/Spo0J family partition protein | |
| *orf(-4)* | 358 | SCD30600.1 (99/99) | bifunctional DNA primase/polymerase, N-terminal | |
| *orf(-3)* | 513 | WP_093538842.1 (100/100) | DNA primase | |
| *orf(-2)* | 128 | WP_093538843.1 (100/100) | transcription factor | |
| *orf(-1)* | 323 | WP_107405752.1 (98/99) | alpha/beta fold hydrolase | |
| *glmA* | 198 | WP_093538844.1 (99/99) | TetR/AcrR family transcriptional regulator | |
| *glmB* | 186 | WP_093538866.1 (99/100) | signal peptidase II | |
| *glmC* | 8938 | WP_141718495.1 (99/99) | amino acid adenylation domain-containing protein, partial | KS-AT-ACP-KS-AT-DH-KR-ACP-C-A-nMT-PCP-C-A-PCP-C-A-PCP-C-A-PCP-C-A-PCP-TE |
| *glmD* | 470 | WP_093539563.1 (99/100) | PLP-dependent aminotransferase family protein | AMT |
| *glmE* | 207 | WP_093539562.1 (98/100) | AraC family transcriptional regulator | |
| *glmF* | 1008 | WP_107405771.1 (98/99) | non-ribosomal peptide synthetase | A-PCP |
| *glmG* | 274 | WP_093539561.1 (100/100) | mycofactocin-coupled SDR family oxidoreductase | KR |
| *glmH* | 253 | WP_093539560.1 (99/100) | thioesterase | TEII |
| *glmI* | 128 | WP_093539559.1 (99/100) | ketosteroid isomerase | |
| *glmJ* | 211 | WP_093539558.1 (99/100) | 4'-phosphopantetheinyl transferase superfamily protein | ACPS |
| *orf1* | 396 | WP_141718395.1 (99/100) | hypothetical protein | |
| *orf2* | 247 | WP_093539556.1 (97/97) | hypothetical protein | |

| | | | |
|---|---|---|---|
| *orf3* | 545 | WP_093539555.1 (99/100) | GH3 auxin-responsive promoter family protein |
| *orf4* | 780 | WP_093539554.1 (99/100) | phosphotransferase |
| *orf5* | 262 | WP_093539553.1 (100/100) | hypothetical protein |

[a]*Sizes are given in amino acids;* [b]*Accession numbers and percentage of identity/similarity are given in parentheses.*

### 3.3.1.1 The core globomycin structure and its biosynthesis

The biosynthesis of globomycin is assumed to start with GlmC, a hybrid PKS/NRPS, containing 7 modules (Figure 40). The first two modules are modular PKS (type I) and the remaining 5 modules are of NRPS type. All the domains found in the biosynthetic gene cluster were analyzed in order to get a better idea about the biosynthesis.

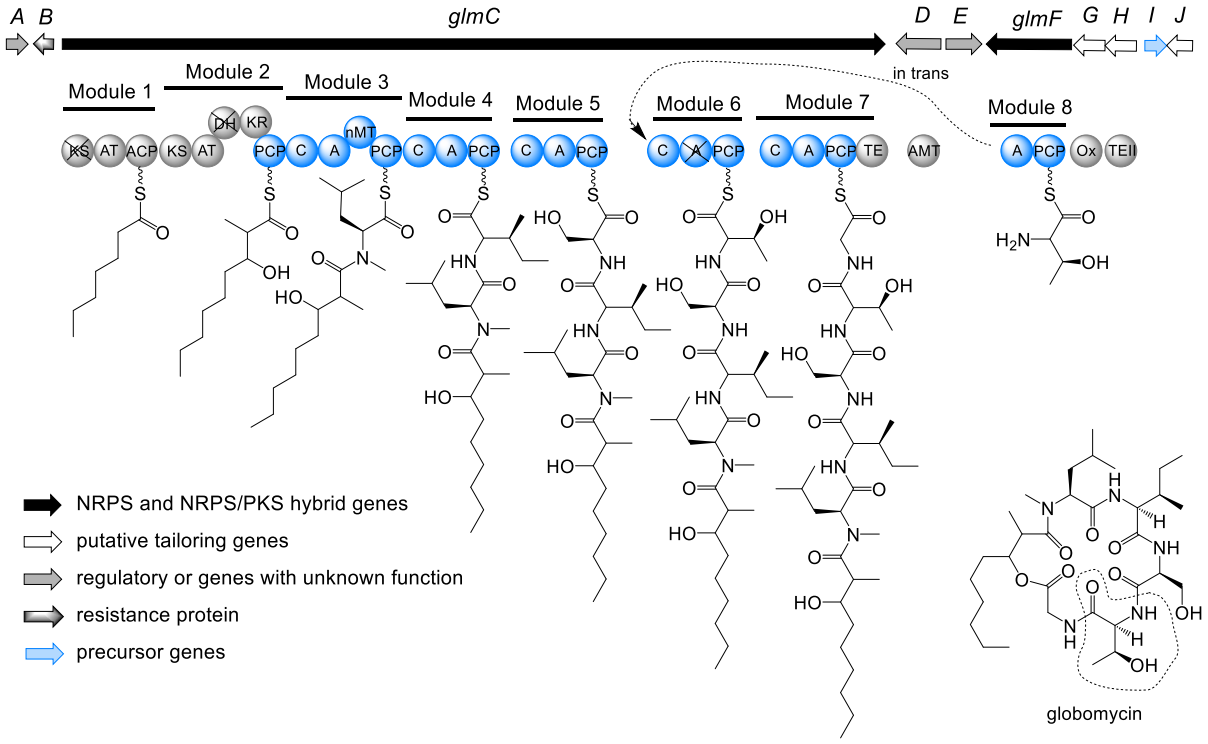



**Figure 40. Globomycin biosynthetic gene cluster and the biosynthesis proposal.** Genes *glmA* to *glmJ* stretch over 37 kb DNA. The module 1 is assumed to load a heptanoyl-CoA moiety from primary metabolism onto ACP1. Module 2 selects methylmalonyl-CoA, the KR reduces the keto group to hydroxyl group, yielding 3-hydroxy-2-methylnonanoyl-ACP. The DH domain is not functional. Module 3 selects L-leucine, the n-MT domain methylates the amino group and so N-methylleucine is incorporated into the growing chain. Module 4 incorporates L-*allo*-isoleucine, module 5 L-serine, module 6 with an unfunctional A domain accepts L-*allo*-threonine selected by module 8 and condenses it into the growing chain. Lastly, module 7 selects glycine. The TE domain in module 7 catalyzes intermolecular capture by one of the –OH groups in the acyl-peptidyl chain, releasing a cyclic depsipeptide globomycin [68]. The incorporated L-*allo*-threonine in globomycin is marked with a dashed curve.

First, two KS domains are present in modules 1 and 2. The putative active sites T209, C221, S315, H346, K379, N381, H381, H384, Q386, N455 and H457 are well characterized in EryA1

KS1 domain [82]. The residues C211, H346, K379, H384 are universally conserved but T209, S315, N381, Q386, N455 and H457 are only conserved within either an iterative or a non-iterative KS domain subgroup. The alignment of GlmC-KS1 and GlmC-KS2 domains to a non-iterative EryA1 and an iterative LovB KS domains had shown differences in T209, C211, S315 and Q386 residues, in one or both domains. The GlmC-KS1 contains mutations of residues T209A, C211Q, S315M and Q386E[1], while in GlmC-KS2 the only difference observed is Ser315 residue exchange to Thr (Figure 41).

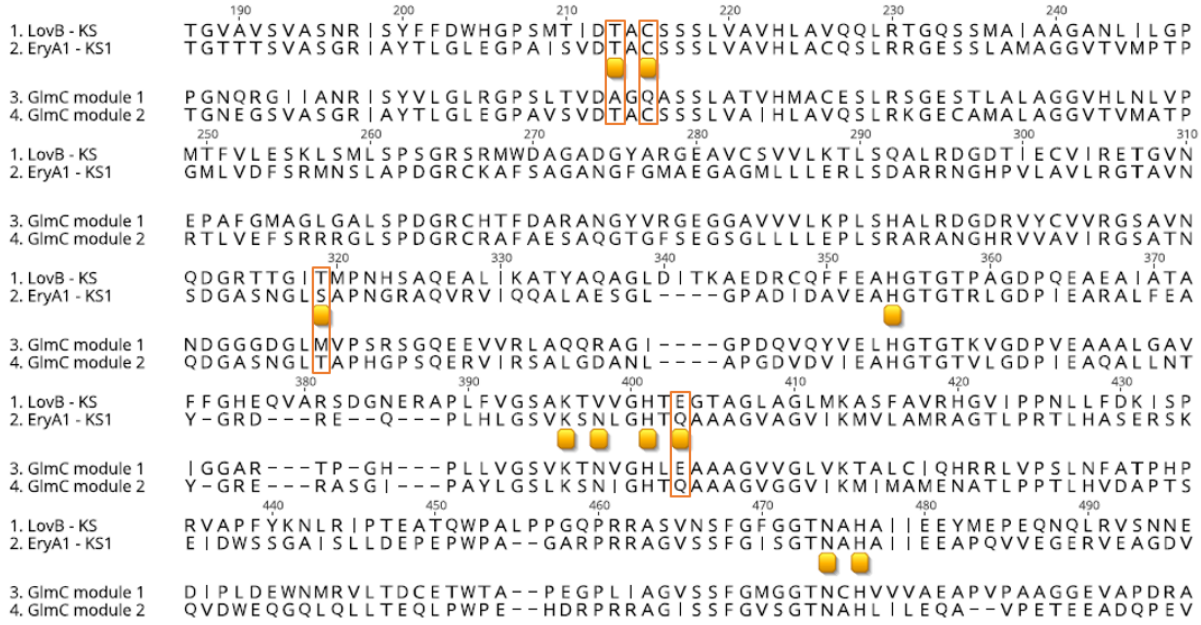



**Figure 41. The important catalytic residues of KS domains.** The active sites T209, C211, S315, H346, K379, N381, H384, Q386, N455, H457 from non-iterative EryA1 and iterative LovB KS domain [82] are marked with the orange rectangles. Additionally marked are the residues where a difference in either of the globomycin KS domains is observed.

At the position 386, a non-iterative KS contains Gln (Q) residue and an iterative KS Glu (E), the same as in GlmC-KS2 and GlmC-KS1, respectively. The Ser315 is characteristic for a non-iterative KS and Thr (T) for an iterative KS from LovB. The latter is found also in GlmC-KS2, whereas GlmC-KS1 contains Met (M) in that position. The T209 and C211 residues in a non-iterative EryA1 KS are present at the same positions as in an iterative LovB and GlmC-KS2. However, T209 and C211 residues in GlmC-KS1 are exchanged by Ala (A) and Gln (Q), respectively. GlmC-KS2 domain seems perfectly functional but its KS domain-type is hard to assign since T315 hints on an iterative mechanism, whereas Q386 on the non-iterative. Moreover, the active sites in GlmC-KS1 show many differences: the Glu (E) in place of Q386

[1] EryA1 KS1 domain numbering

hints on iterative mechanism, while A209, Q211 and M315 appear to be mutations that restrict its catalytic activity. The main role of C211 is to provide an exceptionally nucleophilic active site for anchoring the growing polyketide chain before the decarboxylative Claisen condensation takes place [82]. The Gln is about 5000-fold less reactive than Cys [238], which hints on GlmC-KS1 inability for the intermolecular translocation of the growing polyketide chain [82]. However, the GlmC-KS1 domain as a part of a loading module is not required and does not need to be functional. Second, the AT domains are predicted to be involved in (2S)-methylmalonyl-CoA selectivity [83]. According to the globomycin chemical structure, we propose that GlmC-AT1 loads an odd-chain heptanoic acid (or nonanoic acid in SF-1905 $A_5$), obtained from the pool of primary metabolites [239], and the GlmC-AT2 selects methylmalonyl-CoA as predicted. Third, the KR domain in module 2 (GlmC-KR) contains a NADP-binding motif GGxGxxG (not shown) and the active site YxxxN (Figure S2- 1) [84]. It also contains conserved residues Tyr2616, Ser2603 and Asn2620 in the active site (Figure S2-1) [85], aligning to Tyr159, Ser146 and Lys163 from tropinone reductase-II (TR-II), respectively [205], all proving its functionality. It is proposed to convert the intermediate 2-methyl-3-oxononanoyl-ACP to 3-hydroxy-2-methylnonanoyl-ACP (Figure 40). Additionally, the GlmC-KR domain is predicted to produce a carbon centre with D-configuration [85]. Fourth, the GlmC-DH domain lacks the whole LPFxW motif (Figure S2- 2). The highly conserved surface residue Phe227 from LPFxW motif is, together with an invariant Arg275 from erythromycin dehydratase, believed to help with the formation of an ACP docking site [86]. The Arg2241 is present but the "Phe227" residue is not. Therefore, the globomycin DH domain cannot interact with ACP and is because of that not functional [86], which also explains the retained hydroxyl group in 3-hydroxy-2-methylnonanoic acid part of the globomycin structure. Fifth, the adenylation domain in module 3 (A3) is specific for L-leucine and the following n-MT domain methylates its amino group, resulting in N-methylleucine building block. The A4 domain incorporates L-*allo*-isoleucine and A5 domain L-serine. The prediction of A6 for the selection of hydrophobic aromatic amino acid like Phe, Tyr or Trp is unreliable. The A domain analysis reveals the absence of amino acids in core motif A5 and mutation K7492T highly conserved in core motif A10 (Figure S2- 3) [72]. The A6 domain is therefore unable to select any amino acid into an elongating acyl chain. Module 7 is predicted to incorporate L-glycine which fits to the structure. Sixth, all the NRPS modules (3-8) have functional PCP domains containing Ser residue in the core motif [I/L]GG[D/H]SL [225]. Suprisingly, the PKS type I module 2 contains a functional PCP domain instead of the expected ACP domain (Figure S2- 4). Last, the TE domain positioned at the end of module 7 catalyzes

intermolecular capture by one of the –OH groups in the polyketide-peptidyl chain, generating a cyclic depsipeptide [68]. If the A6 domain is unfunctional as assumed, the globomycin cannot be produced. However, GlmF protein contains A domain (module 8) that is predicted to select threonine. The module 8 lacks a condensation domain [224] (Figure S2- 5) but its A domain appears essential for globomycin biosynthesis since it could select the missing L-*allo*-threonine [240]. The GlmF NRPS protein could then work *in trans* and transfer activated amino acid to C domain of module 6 (Figure 40). For that reason, an inactivation mutant *S. hagronensis*/*glmF*Δ, with interrupted *glmF* gene achieved by a plasmid insertion via a homologous region, was prepared (Table S2- 4). The mutant strain and WT strain were cultivated in parallel. Crude extracts were measured via LC-MS and analyzed. The globomycin production was significantly decreased to 13 % in the mutant strain, when compared to the WT strain production, but not completely abolished (Figure S2- 6). As the strain allows homologous recombination only with homologies longer than 2 kb, the clean inactivation mutant was hard to prepare. Gene disruptions with >2-kb homology region are efficient in multidomain PKS- or NRPS-encoding genes such as *glmC*. However, in shorter genes the major part of the gene is still retained and in case of *S. hagronensis*/*glmF*Δ this could potentially lead to partially preserved activity. Moreover, the WT conjugation efficiency was low as only 2 exconjugants were obtained. However, the substantial drop in production shows that the *glmF* gene is most likely involved in the biosynthesis.

The correctness of incorporated amino acid on GlmC is probably assured by a type II TE GlmH since the incorrect amino acids selected by $A_n$ domain cannot be processed by $C_n$ domain, consequently the peptide bond cannot be formed. Such module is stalled until the incorrect amino acid is hydrolyzed [241]. There are plenty of examples of NRPSs not following the rules of "textbook" modular biosynthesis, e.g. iterative and non-linear biosynthetic strategies (see Discussion). The modular NRPSs encoded in bacterial genomes constantly evolve, subject to the "optimization objective", to give bacteria an advantage in competition with all the other microorganisms present in their natural habitat [28]. In some cases the catalytic domains are inactive such as ER domain in epothilone [70], KS and DH domains in globomycin (Figure 40) and A domain in cystobactamide BGC [55]. A high flexibility of NRPS protein was observed in case of myxochromides, where an entire module is skipped [242]. The incorporation of L-*allo*-threonine into an intermediate acyl-peptidyl chain assembling on GlmC NRPS by the proposed mechanism would as well demand certain flexibility of a main NRPS protein in order to provide spatial access to the GlmF protein. Taken together, the building block assembly of globomycin appears mainly linear with a non-linear component.

### 3.3.1.2 The origin of building block L-*allo*-isoleucine and L-*allo*-threonine

The biosynthesis of L-*allo*-isoleucine from L-isoleucine requires two enzymes, namely a pyridoxal 5′-phosphate (PLP)-linked aminotransferase and an unprecedented isomerase acting synergistically [243]. Enzyme pairs DsaD/DsaE from the desotamide and MfnO/MfnH from the marformycin biosynthetic pathway were shown to be responsible for the biosynthesis of L-*allo*-isoleucine present in respective natural products [243]. Only one branched-chain amino acid aminotransferase (NRRL15064_48630) homologous to DsaD (identity 52 %; similarity 65 %) and MfnO (identity 49 %; similarity 65 %) was found in the genome of *S. hagronensis* in contig 14 outside of globomycin BGC. On the other hand, DsaE and MfnH enzymes belong to the nuclear transport factor 2 (NTF2) superfamily with members including $\Delta^5$-3-ketosteroid isomerases and share 53 % and 39 % identity and 63 % and 55 % similarity to the homolog GlmI found in globomycin gene cluster [243]. The L-*allo*-threonine could originate from the interconversion of L-*allo*-threonine and glycine by L-*allo*-threonine aldolase [244]. The responsible enzyme is not found in the globomycin BGC but is present elsewhere in *S. hagronensis* genome, therefore the L-*allo*-threonine selected by A8 domain is probably taken from the pool of primary metabolites.

### 3.3.1.3 The signal peptidase II

In globomycin BGC, GlmB encodes a signal peptidase II. The LspA homolog is involved in post-translational modifications of lipoproteins in many pathogenic bacteria [139]. Globomycin and myxovirescin act by sterically blocking the LspA active site, which results in accumulation of prolipoprotein [233] and compromises the integrity of the bacterial cell envelope. In myxovirescin producer strain *Myxococcus xanthus*, four *lspA* genes are present. LspA1 and LspA2 serve as housekeeping proteins but only one of them is necessary for the viability of the bacterial cell [245]. LspA3 and LspA4 are located in the myxovirescin BGC and are proposed to play a role in myxovirescin resistance and regulation [245]. In *Streptomyces hagronensis*, two LspA homologs are present, here named GlmB and LspShg. They share 42 % identity and 73 % similarity. GlmB is located in the globomycin BGC and is most alike LspA4 (MXAN_3930) with 26 % identity and 62 % similarity. Due to its location, it is proposed to act as a self-resistance protein. On the other hand, LspShg is located elsewhere in the genome, is the most alike LspA2 (MXAN_0369) with 28 % identity and 54 % similarity and most probably serves as a housekeeping protein. The protein identity between *M. xanthus* and *S. hagronensis* is rather low, which is also due to the fact that the bacteria belong to different phyla. Both protein sequences originating from *S. hagronensis* were compared to the LspA from *Bacillus*

*subtilis* [246] and LspA from MRSA [140] and were found to contain all 14 highly conserved residues characteristic for signal peptidases II. However, the mutual comparison of the GlmB and LspShg sequences does not elucidate why would one act as a housekeeping protein and the other as a self-resistance protein, neither if they increase the bacterial resistance towards globomycin. The gene knock-out study, as conducted for the myxovirescin producer *M. xanthus* [245], or the protein expression in globomycin-susceptible bacteria, followed by testing of the minimal inhibitory concentrations (MICs), would be required.

#### 3.3.1.4 Other genes encoded in the globomycin biosynthetic gene cluster

GlmA is a TetR/AcrR family transcriptional regulator potentially involved in the negative regulation of globomycin production [247]. Gene *glmD* is predicted to encode a DNA-binding transcriptional regulator from MocR family which contains a PLP-dependent aminotransferase domain. MocR-TFs are a subfamily of GntR family of bacterial transcriptional factors involved in the regulation of various biological processes [248, 249]. These proteins catalyze the reversible transfer of an amino group from the amino acid substrate to an acceptor α-keto acid. They require pyridoxal 5´-phosphate (PLP) as a cofactor to catalyze this reaction. Transamination reactions are of central importance in amino acid metabolism and linked to carbohydrate and fat metabolism [250]. In *Streptomyces* strains only few members of GntR family are known but their role is mostly unknown [250]. WhiH encodes a sporulation transcription factor [251] and DevA a developmental factor [252] in *Streptomyces coelicolor*, PdxR from *Streptomyces venezuelae* is involved directly in the regulation of pyridoxal phosphate synthesis [253]. Some of known MocR-TFs are involved in central biosynthetic pathways, while others regulate the catabolism of compounds that bacteria come across only in particular circumstances [248] but are not reported to be directly involved in antibiotic biosynthesis. The GlmE protein contains type 1 glutamine amidotransferase (GATase1)-like domain, also found in transcriptional regulators, and belongs to the DJ-1/PfpI family with a member PfpI, an intracellular cysteine protease which may hydrolyze small peptides to provide a nutritional source. For example, a member of ThiJ/DJ-1/PfpI family in *Saccharomyces cerevisiae*, Hsp31 protein plays an indispensable role in regulation of redox homeostasis [254]. The protein GlmE might have a role in bacteria in general but might not have an effect on globomycin production. The *glmG* gene encodes a classical mycofactocin-coupled SDR family oxidoreductase [84]. The reactions catalyzed within the SDR family include isomerization, decarboxylation, epimerization, C=N bond reduction, dehydratase activity, dehalogenation, enoyl-CoA reduction, and carbonyl-alcohol oxidoreduction. The role of GlmG for the

biosynthesis of globomycin is not clear. The GlmJ is one of two ACPSs in *S. hagronensis*, necessary for activation of PKS and NRPS biosynthetic genes [68], belonging to the 4'-phosphopantetheinyl transferase superfamily.

#### 3.3.1.5 Gene cluster borders

The *orf(-7)* encodes a hypothetical protein. The *orf(-6)* and *orf(-5)* encode ParA family protein and ParB/RepB/Spo0J family partition protein, respectively, known to participate in segregation of newly replicated chromosomes in bacterial cells [255]. The DNA replication requires as well DNA polymerase and DNA primase [256] encoded by *orf(-4)* and *orf(-3)* genes. Orf(-4) is a N-terminal domain of bifunctional DNA primase/polymerase and Orf(-3) a C-terminal of DNA primase. As transcriptional regulators play a pivotal role in DNA replication [257], Orf(-2) encoding an unspecified transcription factor could be involved as well. The Orf(-4)-Orf(-2) protein homologs were found also upstream, in the vicinity of a butyrolactone BGC, similarly followed by a GlmA homolog, a TetR/AcrR family transcriptional regulator, and are therefore believed not to be involved in the globomycin biosynthesis. However, the butyrolactone is essential for *Streptomyces* development and streptomycin production in *S. griseus* [258], and may regulate the transcriptional expression of globomycin biosynthetic genes. The Orf(-1) an α/β fold hydrolase predicted to be dienelactone hydrolase [259] and is assumed to be redundant for the globomycin biosynthesis. Downstream of GlmJ, five genes (*orf1-orf5*) encoding for homologs of SepM-SepQ are observed. They were previously reported to be found in proximity of biosynthetic genes of septacidin but were not involved in its biosynthesis [260]. We assume the same in this case. Comparing the genes located outside of GlmA-GlmJ to the NCBI Nucleotide collection database, does not return any relevant information concerning the boundaries of the globomycin BGC.

### 3.3.2 Yield improvement attempts

*Streptomyces hagronensis* strain 360 was primarily described as producer of U-56,407 antibiotic [236] but produces globomycin as well. Reported strain used for fermentation of globomycin, *Streptomyces halstedii No*. 13912, produced 10 mg $L^{-1}$ [237]. Globomycin yield was initially improved by optimization of bioreactor production medium published together with globomycin discovery [228]. The production medium here required 1 % feather meal for the nitrogen source but was in this laboratory exchanged for fish meal, corn steep solids, corn steep liquor or soy flour. The production of all the media variants in parallel showed that best

and most efficient production was achieved when medium contained soy flour (SOYF) (Figure 42).

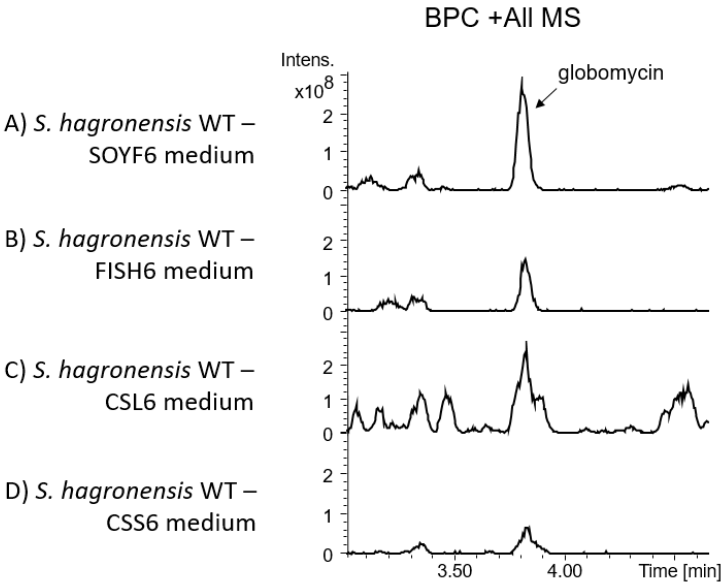



**Figure 42. The comparison of globomycin production in selected media.** The globomycin peak with a molecular mass 656.56 m/z $[M+H]^+$ is visible at 3.8 min in BPC chromatograms. In media other than SOYF6 (A), smaller production (B, C) or peak overlay (C, D) is observed. The information about the production yields is depicted in Figure 43.

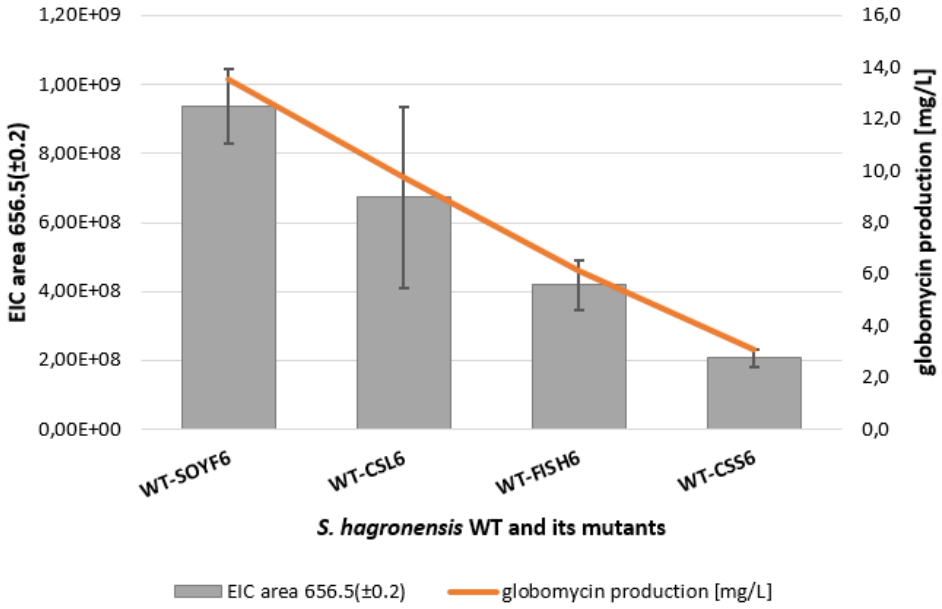



**Figure 43. Globomycin production of *S. hagronensis* wild-type (WT) strain in different media.** The SOYF6 medium shows best production yields compared to FISH6 (45 %), CSL6 (72 %) and CSSL6 (22 %) media. The globomycin calibration curve is shown in Figure S2- 7. The average globomycin production yields are depicted with the orange line.

As the only information about pH was the value after sterilization [228], the pH range was tested from 6.0 to 8.0. The best production was achieved in case of pH 6.0, therefore medium SOYF6 was used for production of globomycin [140].The globomycin yield in SOYF6 medium is 13.5 ± 1.5 µg/ml (Figure 43), while the production yields achieved in FISH6, CSL6 and CSS6 media reach 45 %, 72 % and 22 %, respectively, compared to SOYF6 medium (Figure 43). Since the aim was purification of several milligrams, sufficient for crystallography studies and bioactivity testing, performed by the research group of Prof. Martin Caffrey [140], the highest possible globomycin yield was desired.

### 3.4 Conclusions

The *S. hagronensis* whole-genome sequence was obtained. The putative globomycin BGC was identified and the conjugation method was established. The inactivation mutant *S. hagronensis*/*glmC*Δ produced no globomycin, therefore the identified BGC was confirmed. The biosynthetic genes were analyzed *in silico* and the biosynthesis pathway was proposed. The NRPS module 8 in GlmF is assumed to provide L-*allo*-threonine to the GlmC in trans. Two copies of a signal peptidase II are present in the genome. Due to their location in the genome, GlmB is might encode a self-resistance protein and LspShg a housekeeping protein. After a simple medium optimization attempt, SOYF6 was chosen as best globomycin production medium.

## 3.5 Supplementary information

### Table of Tables



### Table of Figures



**Table S2- 1. List of primers used in this study**

| Primer | 5'-Sequence-3' | R. sites | Construct |
|---|---|---|---|
| globo-inact-F | GGACATCTACGACTCCGTCTACGC | / | pCR2.1 inact-globo |
| globo-inact-R | TCCTGGCCCCAGACCAGTT | / | |
| inact-globo-HindIII-F | TATAAGCTTCCACGCTGGTGGAGTCCTTCCAG | *Hin*dIII | pKC1132 inact-globo2.3 |
| inact-globo-EcoRV-R | ATATAGATATCGTGCCTCCAGGCTGTCGAGGTC | *Eco*RV | |
| SHG-inactA8-HindIII-F | ATAAAGCTTCTTCGGCGGGTGTCACGGTCTG | *Hin*dIII | pKC1132 SHG-inactA8 |
| SHG-inactA8-EcoRV-R | TATATGATATCGACCCGTCCGGAATCCCGAATTC | *Eco*RV | |

**Table S2- 2. List of sequencing and other primers**

| Primer | 5'-Sequence-3' | Usage |
|---|---|---|
| M13-21F | TGTAAAACGACGGCCAGT | For general plasmid integration verifications |
| M13-24R | CGGATAACAATTTCACACAGG | |
| check-inact-globo-F | AGTGCCTCCAGGCTGTCGA | Orientation of inact-globo fragment in pKC1132 inact-globo |
| PKC1132-R | ATGCTTCCGGCTCGTATGTTGTGTG | |
| PKC1132-F | CGATTAAGTTGGGTAACGCCAGG | |
| check-inact-globo-R | CCGAACAGGTCGGCGAATG | |
| PKC1132-F | CGATTAAGTTGGGTAACGCCAGG | Verification of pKC1132 inact-globo2.3 plasmid |
| PKC1132-R | ATGCTTCCGGCTCGTATGTTGTGTG | |
| check-inact-globo-R | CCGAACAGGTCGGCGAATG | |
| PKC1132-F | CGATTAAGTTGGGTAACGCCAGG | Verification of pKC1132 SHG-inactA8 plasmid |
| PKC1132-R | ATGCTTCCGGCTCGTATGTTGTGTG | |
| check-inactA8-F | CAGGACTTCGGCGGCCTCGC | |
| check-inactA8-R | GGTGAGCCTGCTGAACACGG | |

**Table S2- 3. Genetic constructs used in that study**

| Plasmid name | Details | Reference |
|---|---|---|
| pCR2.1 | TA cloning vector; $kan^r$, $amp^r$ | Thermo Scientific |
| pKC1132 | *aac(3)IV lacZα oriT*$_{RK2}$; suicide vector for *Streptomyces* containing *oriT* RK2 for conjugation from *E. coli* to *Streptomyces* | [261] |
| pCR2.1 inact-globo1 | 0.84 kb homology from 3$^{rd}$ module (nMT domain) in GlmC cloned into pCR2.1 | This study |
| pKC1132 inact-globo1 | Homology from pCR2.1 inact-globo was transferred into pKC1132 vector | This study |
| pKC1132 inact-globo2.3 | 2.3-kb DNA fragment containing GlmC nMT domain was cloned into pKC1132 plasmid | This study |
| pKC1132 SHG-inactA8 | 2.071 kb DNA fragment (*glmF*: 58 –2110 bp) was cloned into pKC1132 plasmid | This study |

**Table S2- 4. Bacterial strains used or prepared in that study**

| Organism | Genotype or usage purpose | Reference |
|---|---|---|
| *Escherichia coli* | | |
| DH10β | F$^-$ *endA*1 *recA*1 *galE*15 *galK*16 *upG rpsL* Δ*lacX*74 Φ80*lacZ*ΔM15 *araD*139 Δ(*ara-leu*)7697 *mcrA* Δ(*mrr-hsdRMS-mcrBC*) λ$^-$ | Invitrogen |
| ET12567/pUZ8002 | F$^-$ *dam*13::Tn*9*, *dcm*6, *hsdM*, *hsdR*, *recF*143::Tn*1I*, *galK*2, *galT*22, *ara*14, *lacY*1, *xyl*5, *leuB*6, *thi*1, *tonA*31, *rpsL*136, *hisG*4, *tsx*78, *mtl*1 *glnV*44 pUZ8002 | [27, 262] |
| *Streptomyces hagronesis* strain 360, NRRL 15064 | Globomycin producer; WT | [236] |
| *S. hagronensis*/pKC1132inact-globo2.3 | Inactivation of *glmC* | This study |
| *S. hagronensis*/pKC1132inactA8 | Inactivation of *glmF* | This study |


**Figure S2- 1. The catalytic sites of globomycin KR domain.** The alignment of GlmC-module 2 and TR-II (PDB: 2AE2) [205] was generated using Geneious Prime® 2020.0.5 [263].

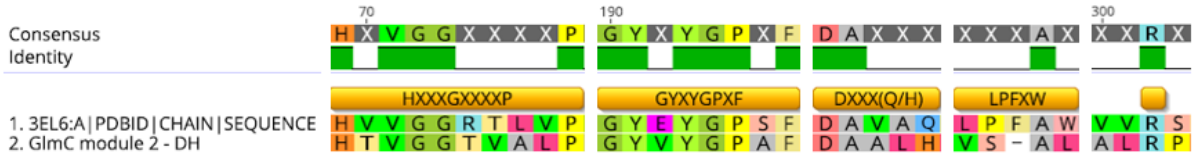



**Figure S2- 2. The alignment of DH domain from GlmC module 2 with erythromycin dehydratase.** Erythromycin dehydratase was found in PDB under accesion code 3EL6 [226]. The conserved motifs are listed in orange boxes.

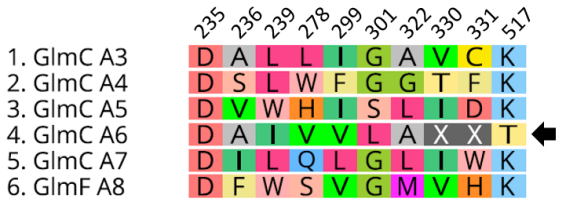



**Figure S2- 3. The important amino acid residues for selectivity and functionality of A domains of globomycin BGC.** The amino acid denoted by X is in this case a missing amino acid residue. The GlmC A6 domain is due to the K517T mutation predicted to be unfunctional [72].

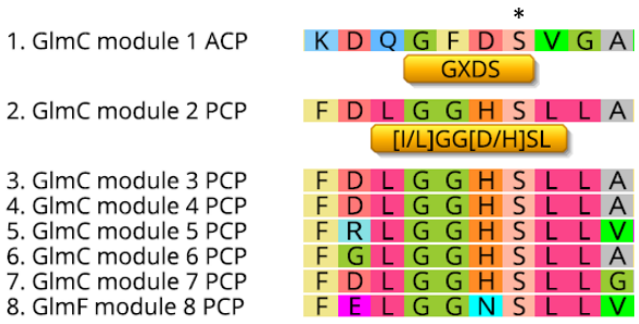



**Figure S2- 4. The alignment of ACP and PCP domains of globomycin BGC with key residues for catalytic activity.** The core motif in ACP domains is GXDS [88] and [I/L]GG[D/H]SL in PCP domains [225]. The key residue is marked with an asterix (*). Surprisingly, the module 2 contains PCP domain at the end of PKS-type module.

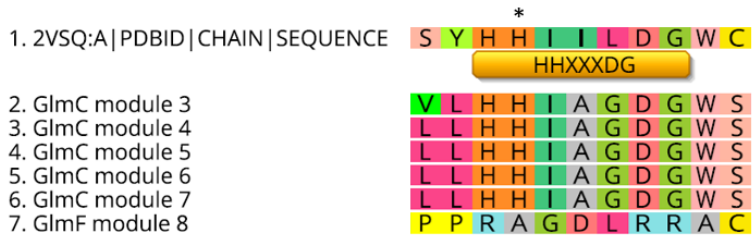



**Figure S2- 5. The alignment of C domains with key residues for catalytic activity.** The C domains of globomycin NRPS modules were compared to amino acid sequence SrfA-C from PDB (2VSQ) [223] with thoroughly annotated core motives for each of the domains. The key residue is marked with an asterix (*). There is no functional C domain in module 8 [224].

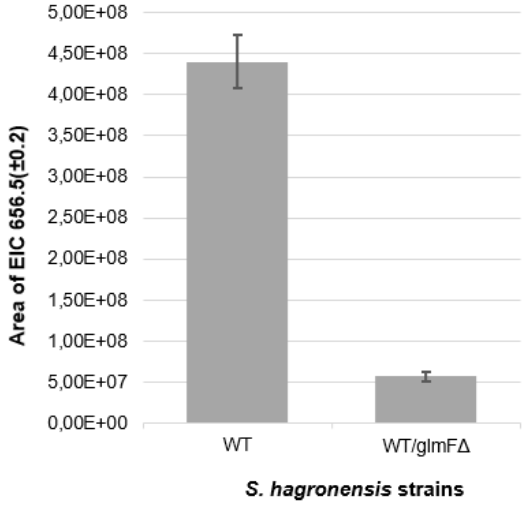



**Figure S2- 6. Globomycin production of *S. hagronensis* WT strain and *S. hagronensis*/*glmF*Δ mutant.** The globomycin production is greatly reduced in the mutant strain, with inactivated GlmF NRPS protein, reaching only 13 % compared to the WT production. Decreased yields suggest the involvement of GlmF in the globomycin biosynthesis.

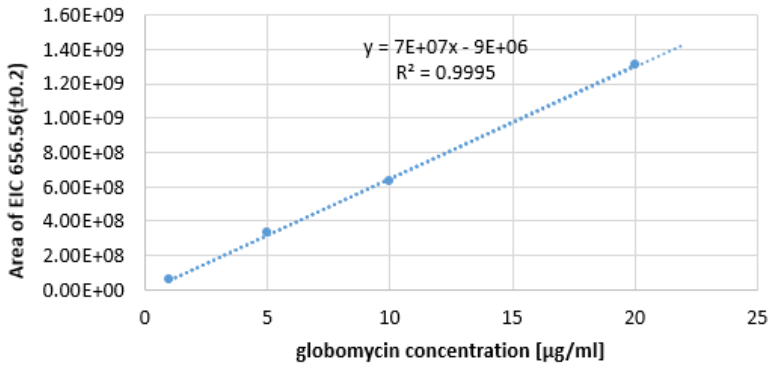



**Figure S2- 7. Globomycin calibration curve.** The measured globomycin concentration points are 1 µg/ml, 5 µg/ml, 10 µg/ml and 20 µg/ml. The areas of EIC 656.56 (±0.2) are used as values for the standard curve (y=7E+07x-9E+06). Low concentrations are chosen since the measured samples reach such concentration range.

# 4 Fidaxomicin: Creating a platform for semi-synthesis

## 4.1 Introduction

Fidaxomicin (Figure 44) was first discovered in 1975 [29] and recognized as an inhibitor of bacterial RNA polymerase (RNAP) [264]. In 1987, it was reported that fidaxomicin [29] actually consisted of two compounds named lipiarmycin A3 and A4 in ratio 3:1 [265]. In the literature, it appears as lipiarmycin A3 [29], tiacumicin B [80], clostomicin B1, difimicin, OPT-80 and PAR-101 [141, 266]. It inhibits bacterial transcription by trapping an open-clamp state of RNAP and so preventing interaction with promoter −10 and −35 elements [267]. Due to its powerful and broad biological activity, much is known about the antibiotic. It exhibits biological activity against some *Staphylococcus aureus* strains and other Gram-positive bacteria, a good antiplaque activity against strains of cariogenic *Streptococcus mutans*. It shows inhibition of bacteriophage growth in *Bacillus subtilis*, a good activity against different aerobic bacteria but weaker activity against anaerobic bacteria. It is also active against *Staphylococcus epidermidis* biofilms and on some drug-resistant strains of *Mycobacterium tuberculosis.* It is a broad-spectrum antibiotic against Gram-positive pathogenic bacteria with the lowest minimal inhibitory concentration (MIC) against diarrhea-associated *Clostridium difficile* [268]. In 2008, it was rediscovered as an effective anti-tuberculosis agent [143] and was also shown to exhibit good anticancer activities against human breast or cervical cancer cells below the cytotoxicity threshold [144]. Fidaxomicin was developed as a drug against *Clostridium difficile* associated infection (CDI) by Optimer Pharmaceuticals, Inc. and was approved by FDA in 2011 [142] under trade name Dificid [141]. *Clostridium difficile*, a Gram-positive anaerobic bacterium, is the causative agent of 20 % to 25 % of all cases of antibiotic-dependent diarrheas. Usual treatment of these infections involves two broad-spectrum antibiotics: metronidazole and/or vancomycin. Metronidazole is easily absorbed along the gastrointestinal tract, resulting in use of huge doses of antibiotic and it is not as effective as vancomycin. However, they both promote the development of vancomycin-resistant *Enterococci* and the risk of relapse is 15-35 % [141]. On the other hand, fidaxomicin shows good *in vitro* bioactivity in CDI treatment. It is orally consumed, stays in the gastrointestinal tract, is only detectable in the blood in nM range and shows a clinical cure almost identical to vancomycin (91.7 % to 90.6 %) but a lower rate of recurrence of CDI (12.8 % compared to 25.3 %) [141]. Fidaxomicin acts bactericidal to *C. difficile*, while vancomycin as bacteriostatic agent only inhibits its development. Besides, it kills less non-pathogenic bacterial species in guts because of its selective activity against Gram-positive bacteria [141]. After FDA approval, fidaxomicin was further developed against

vancomycin-resistant *Enterococci* and as methicillin-resistant *Staphylococcus aureus* cure [141]. Fidaxomicin belongs to the family of novel 18-membered macrocyclic glycosides and contains an unsaturated macrolide core scaffold, modified D-noviose and D-rhamnose moieties, and a dichorinated homo-orsellinic acid moiety [80, 92]. It is produced by *Actinoplanes deccanensis* ATCC 21983 [29], *Micromonospora echinospora* subsp. *armeniaca* [269], *Dactylosporangium aurantiacum* subsp. *hamdenensis* NRRL 18085 [270] and *Catellatospora* sp. Bp3323-81 [143], with all bacterial species being members of the order *Micromonosporales* [24]. The project is based on the collaboration with Prof. Gademann's research group from ETH Zürich, specialized in semisynthesis of natural products. A marine natural product called mangrolide A, produced by actinomycete similar to *Actinoalloteichus spitensis* from genus *Pseudonocardinales*, contains an almost identical 18-membered polyketide ring structure as found in fidaxomicin (Figure 44) but was claimed to have selectivity for Gram-negative bacteria like *Acinetobacter baumanii* and *E. coli* [145]. However, its reported bioactivity could never be reproduced [271]. Mangrolide A lacks any modification on the primary alcohol and has a disaccharide attached at C-11. The moieties attached to fidaxomicin aglycone could therefore have an influence on the biological activity.

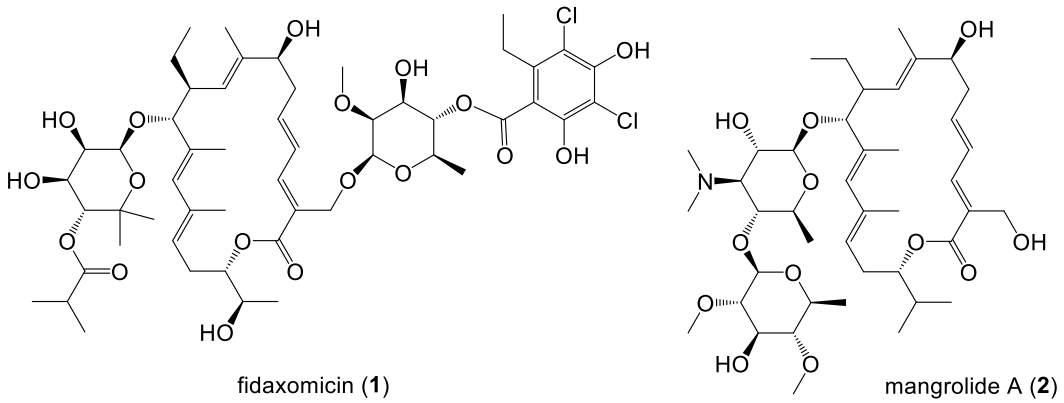



**Figure 44. Structures of fidaxomicin and mangrolide A**

Various novel derivatives could be prepared by attachment of structurally diverse fragments to the fidaxomicin shunt products via semisynthesis, which are not accessible via semi-synthetic modifications of the natural product (Figure 45). The starting material for such experiment could be obtained by total synthesis since the procedures are already established for fidaxomicin [92] as well as for its aglycone [146, 272]. However, the chemical synthesis' yields are usually low. An alternative approach to acquire fidaxomicin aglycone, and shunt products without noviose or rhamnose moieties, is by gene knock-out of glycosyltransferases FdxG1 and FdxG2 from the fidaxomicin biosynthetic gene cluster as performed already by Xiao *et al*. in *Dactylosporangium aurantiacum* subsp. *hamdenensis* [80]. The plan of my part of the project

was therefore to prepare the gene knock-out mutants of glycosyltransferases in *Actinoplanes deccanensis* (DSM 43806, ATCC 21983).

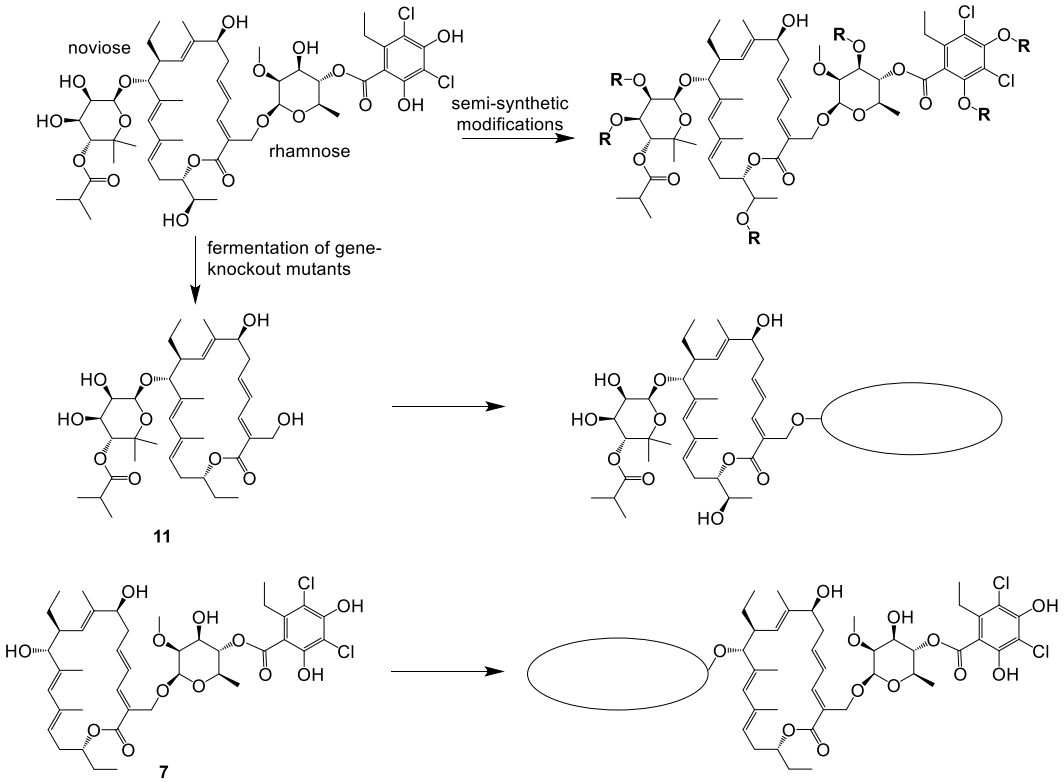



**Figure 45. The general idea for the project.** Shunt products isolated from fermentation cultures of the mutant strains Δ*fdxG1* and Δ*fdxG2* will be used for a semi-synthetic attachment of structurally diverse fragments to generate novel derivatives, not available in nature, with potentially improved antibacterial activities.

## 4.2. Materials and methods

### 4.2.1 Cultivation of *Actinoplanes deccanensis*

Fidaxomicin producer *Actinoplanes deccanensis* is cultivated on ISP3 [273] or MS agar (2 % soy flour, 2 % D-mannitol, 1.8 % agar; $dH_2O$) medium for 7-10 days at 30 °C. For pre-culture, 1 $cm^2$ agar piece is inoculated into 50 mL V-medium (0.3 % meat extract (Merck), 0.5 % Bacto® tryptone, 0.5 % yeast extract (Sigma), 0.1 % sucrose (AppliChem Panreac), 0.4 % calcium malate (ChemCruz); $dH_2O$) [29, 274], with media occupying 1/5 V Erlenmeyer flask, for 3-5 days on 30 °C in rotary shaker at 200 rpm. Due to cell clump formation, baffled flasks are recommended. 10 % pre-culture is reinoculated to fresh V-medium and incubated at 28 °C in round Erlenmeyer flask for a day at 200 rpm [29]. 10 % of *A. deccanensis* vegetative culture is inoculated into E-medium (0.4 % meat extract (Merck), 0.4 % Bacto® peptone, 0.1 % yeast extract, 0.25 % NaCl, 1 % soybean flour (Sigma), 0.5 % $CaCO_3$ (AppliChem Panreac), 0.9 V $dH_2O$, pH 7.6; after autoclaving 0.1 V filtered 50 % glucose (w/v) was added) [29] in baffled

Erlenmeyer flask for 6 days at 28 °C and 200 rpm. After 1-2 days, cellulose [138] to final concentration of 20 g/L and 2 % Amberlite XAD-16 were added to the fermentation broth. For storage purposes, *A. deccanensis* is cultivated in TSB (Tryptone Soy Broth) medium at 28 °C, 180 rpm for 2-3 days. After centrifugation, the cell pellet is resuspended in 50 % (w/v) glycerol and stored at -80 °C. For genetic modification purposes, we determined the susceptibility of *A. deccanensis* to different antibiotics. Bacterium is resistant to erythromycin (25 µg $mL^{-1}$), kanamycin (50 µg $mL^{-1}$), spectinomycin (50 µg $mL^{-1}$); less abundant growth is observed on erythromycin (50 µg $mL^{-1}$) and nalidixic acid (25 µg $mL^{-1}$). Bacterium is sensitive to gentamycin (50 µg $mL^{-1}$), hygromycin (50 µg $mL^{-1}$), streptomycin (50 µg $mL^{-1}$) and apramycin (50 µg $mL^{-1}$).

#### 4.2.1.1 Cultivation of *Actinoplanes deccanensis* Δ*fdxG1* mutant

Mutant strain *A. deccanensis* Δ*fdxG1* was cultivated in 4.8 L of E-medium with 4.8 g/L cellulose. Strain was cultivated in 300 mL baffled flasks containing 50 mL medium with 5 % inoculum prepared in V-medium as described above. Additional 500 mL baffled flasks containing 100 mL medium with 5 % inoculum were also used for the cultivation for 7 days at 28 °C, 180-200 rpm. 4 % Amberlite XAD-16 was added and cultures continued shaking for 2 more days. Culture broth was centrifuged for 10 min at 6000 rpm. Mainly XAD-16 with some remaining cells was used for extraction described in Section 4.2.2.

### 4.2.2 Extraction of fidaxomicin and shunt products

Upon completion of fermentation and incubation with XAD-16 resin, 50 mL culture broth was centrifuged at high speed. Due to fidaxomicin insolubility in water [275], supernatant was discarded. The pellet with XAD-16 was lyophilized and extracted with 1-butanol on magnetic stirrer for 3 h and filtered through glass wool into a round bottom flask. Solvent was evaporated on a rotary evaporator (Rotavapor R210, Büchi) at 50 °C at reduced pressure. Sample was dissolved in 2.5 mL methanol, centrifuged at 15000 rpm for 10 min at 15 °C and measured in dilutions for quantification purposes. The fermentation material of *A. deccanensis* Δ*fdxG1* was extracted with ethanol [80] and stirred on magnetic stirrer for few hours. The solvent, cells and XAD-16 were centrifuged for 10 min at 6000 rpm. Liquid layer was filtered through glass wool and dried on a rotary evaporator. Ethyl acetate and water were added. Ethyl acetate layer was evaporated, crude extract remains were dissolved in methanol and dried under $N_2$ flow. When the crude extract was dried, 2 mL ultrapure water was added into 4 mL glass vial, vortexed, frozen with liquid $N_2$ and lyophilized. The vial was sent to collaborators at ETH Zürich to

evaluate produced shunt products and for their purification. All following cultivations of mutant strains and culture broth extractions were performed by Andrea Dorst (ETH Zürich), therefore the extraction procedure was not optimized in this lab.

### 4.2.3 Quantification of fidaxomicin

An UHPLC system (UltiMate 3000 LC (Dionex) with 1.7 µm Acquity UPLC BEH C-18 Column (2.1 mm × 100 mm) coupled with an electrospray ionization (ESI) source linked to an amaZon speed MS (Bruker Daltonics; 3D ion trap)) was used for quantification. Separation of 1 µl sample was achieved by a linear gradient (5-95 % B in 6.5 min) from (A) $H_2O$ + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 0.600 ml $min^{-1}$ at 45 °C. For precise mass determination (HRMS) samples were measured on a UHPLC coupled to maXis 4G MS system (Bruker Daltonics; Q-ToF). Separation of 1 µl sample, with 1.7 µm Acquity UPLC BEH C-18 Column (2.1 mm x 100 mm), was achieved by a linear gradient (5-95 % B in 6 min or 18 min) from (A) $H_2O$ + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 0.600 ml $min^{-1}$ at 45 °C. Absorption spectra (200-600 nm) were recorded by a diode array detector. Mass spectra were acquired in centroid mode ranging from 200 to 2000 m/z (amaZon) or from 150 to 2500 m/z (maXis 4G) in positive ionization mode.

### 4.2.4 Sequencing of *A. deccanensis* genome

Genomic DNA of *A. deccanensis* was isolated using phenol/chloroform according to the standard protocol [160] and was sent for Ilumina sequencing. The sequencing reads were aligned with tiacumicin B gene cluster from *Dactylosporangium aurantiacum* subsp. *hamdenensis* (NRRL 18085) [80] and also *de novo* assembled by Nestor Zaburannyi.

### 4.2.5 Construction of plasmids for genetic manipulation of *A. deccanensis*

Routine handling of nucleic acids, such as isolation of plasmid DNA, restriction endonuclease digestions, DNA ligations, and other DNA manipulations, was performed according to standard protocols [160]. *E. coli* DH10β (Invitrogen) was used as host for standard cloning experiments and *E. coli* ET12567/pUZ8002 as a donor strain for bacterial transformation [27]. *E. coli* strains were cultured in 2TY medium or on 2TY agar (1.6 % tryptone, 1 % yeast extract, 0.5 % NaCl, (1.8 % agar), $dH_2O$) at 30-37 °C, (200 rpm) overnight. Antibiotics were used at the following final concentrations: 50 µg $ml^{-1}$ apramycin, 25 µg $ml^{-1}$ kanamycin, 10 µg $ml^{-1}$ chloramphenicol. Transformation of *E. coli* strains was achieved via electroporation in 0.1 cm-wide cuvettes at 1350 V, 200 Ω, 25 µF. Plasmid DNA was purified by standard alkaline lysis [160] or by using

the GeneJet Plasmid Miniprep Kit (Thermo Fisher Scientific). Restriction endonucleases, alkaline phosphatase (FastAP) and T4 DNA ligase were obtained from Thermo Fisher Scientific. Oligonucleotides used for PCR and sequencing were obtained from Sigma-Aldrich and are listed in Table S3- 1. PCR reactions were carried out in Mastercycler® pro (Eppendorf) using Phusion™ High-Fidelity or Taq DNA polymerase (Thermo Fisher Scientific) according to the manufacturer's protocol. For Taq: Initial denaturation (2 min, 95 °C); 30 cycles of denaturation (30 s, 95 °C), annealing (30 s, -5 °C of lower primer's $T_m$) and elongation (based on PCR product length 1 kb/min, 72 °C); and final extension (10 min, 72 °C). For Phusion™: Initial denaturation (2 min, 98 °C); 30 cycles of denaturation (15 s, 98 °C), annealing (30 s, $T_m$ of the lower primer) and elongation (based on PCR product length 0.5 kb/min, 72 °C); and final extension (10 min, 72 °C). PCR products or DNA fragments from restriction digests were purified by agarose gel electrophoresis and isolated using NucleoSpin Gel and PCR Clean up kit (Macherey-Nagel). The PCR products were cloned into suitable plasmids pCK1132 and pSET152, selected by restriction analysis and sequenced using sequencing primers (Table S3-1). Details on the construction of all plasmids used and generated in this study are given in Table S3- 2.

### 4.2.6 Conjugation of mycelium *Actinoplanes deccanensis*

The vectors pKC1132 and pSET152 were used for intergeneric conjugation between *E. coli* and *A. deccanensis*. The conjugation protocol reported for *Actinoplanes friuliensis* by Heinzelmann [276] was used. Methylase-negative strain *E. coli* ET12567/pUZ8002 was used as the *E. coli* donor strain. Its overnight culture was re-inoculated (5 %) into 20 mL fresh 2TY medium containing Cm (10 μg $ml^{-1}$), Kn (25 μg $ml^{-1}$), Apr (50 μg $ml^{-1}$) antibiotics. After cultivation at 37 °C, the $OD_{600}$ = 0.4 was reached and the culture was washed twice and finally resuspended in 1 mL. *Actinoplanes deccanensis* was inoculated from agar plate into 20 ml V-medium and cultivated for 3-4 days at 30 °C, 180 rpm. One mL of homogenized culture was inoculated into 100 mL TSB into 500 ml baffled Erlenmeyer flask and cultivated for 4 days. Culture was diluted in ratio 1:10 (50 ml TSB media, 5 mL culture) and incubated over-night. On the following conjugation day, homogenized culture was diluted in ratio 1:5 (10 ml TSB, 2.5 ml mycelium) and incubated at 180 rpm for 1 to maximum 5 hours. Each aliquot of 12.5 ml was centrifuged for 10 min at 3500 rpm before usage and resuspended in 2 mL of TSB. Mixture of 200 µl *A. deccanensis* mycelium and 250 µl *E. coli* was spread on well dried GYM $MgCl_2$ plate (0.4 % glucose, 0.4 % yeast extract, 1 % malt extract, 0.2 % $CaCO_3$, 1.8 % agar, $dH_2O$, pH 7.2; 10 mM MgCl after autoclaving; according to DSMZ). After 16 h, the plates were overlaid with

spectinomycin (50 μg $ml^{-1}$) according to the weight of the plate (30 g = 1 ml sterile water + 30 μl spectinomycin with concentration 50 μg $ml^{-1}$) and after additional 8 h with apramycin (50 μg $ml^{-1}$). After 6-10 days at 30 °C, the number of exconjugants was determined.

### 4.2.7 Genetic modification of glycosyltransferases

Genes *fdxG1* or/and *fdxG2* encoding glycosyltransferases were deleted from the *A. deccanensis* genome. The *fdxG1* gene (1392 bp) and *fdxG2* gene (1416 bp), both containing glycosyltransferase domain, were disrupted by allelic replacement through a double-crossover event. A suicide plasmid pKC1132-KOfdxG1 (Table S3- 2) was constructed by insertion of two PCR amplified DNA fragments, 1.1 kb and 0.94 kb in length, homologous to sequences upstream and downstream of the core of the gene for glycosyltransferase FdxG1. Similarly, pKC1132-KOfdxG2 was constructed with homology arms 1 kb in length. Left homologies were cloned using *PstI*, *Xba*I and the right ones using *Xba*I and *Eco*RV restriction sites (Table S3-1), yielding disrupted FdxG1 with only 66 amino acids and FdxG2 with 50. Resulting sequencing confirmed plasmids, using PKC1132-F and PKC1132-R primers, were transformed into *E. coli* ET12567/pUZ8002 to obtain a non-methylated plasmid, which was then introduced into *A. deccanensis* via conjugation of mycelium. For the construction of a double mutant strain with deleted both genes, strain *A. deccanensis* Δ*fdxG1* was additionally transformed with pKC1132-KOfdxG2 plasmid. Exconjugants were further repatched on SM Spec (25 μg $mL^{-1}$) Apr (50 μg $mL^{-1}$) medium and subcultivated in V-medium without selection pressure. After a few subcultivation steps, Apr-sensitive colony was isolated (secondary recombinant), confirmed by colony PCR using checkG1F, checkG1R or checkG2F, checkG2R primers (Table S3- 1). All used and prepared bacterial strains are described in Table S3- 3.

## 4.3 Results and discussion

### 4.3.1 Fermentation and data analysis

The chemical formula of fidaxomicin is $C_{52}H_{74}Cl_2O_{18}$ and its molecular weight 1058.0406 g/mol. The observed mass on amaZon speed MS system is 1039.4225 $[M\text{-}H_2O+H]^+$, whereas on maXis 4G MS system exact mass of 1056.4252 g/mol is detected. Fidaxomicin was successfully produced in E-medium and its concentration determined with a calibration curve. The best yields achieved were 237.5 ± 8.1 mg/L.

### 4.3.2 Biosynthetic gene cluster of fidaxomicin

Fidaxomicin BGC from *Actinoplanes deccanensis* in comparison to tiacumicin B BGC [80] differs in gene arrangement and contains 6 additional genes but lacks four (Figure 46, Table S3- 4). The gene arrangement was confirmed by a PCR using primers from Table S3- 1. Fidaxomicin BGC begins with *fdxJ* gene and keeps the same gene order as in tiacumicin B BGC until *fdxN* gene. An insertion of 2 genes with oxidoreductase function (*fdxY1-Y2*) and 3 genes with methylmalonyl-CoA mutase-associated function (*fdxW1-W3*) were observed in between *fdxN* and *fdxT2* genes. The *fdxT2* gene is followed by *fdxH* gene encoding a transposase, which is not present in tiacumicin BGC. Next, the sequence continues with *fdxG1* up until *fdxI* gene. In the FDX gene cluster, no genes homologous to *tiaR2*, *tiaS3*, *tiaT3* and *tiaT4* are observed (Figure 46). As both bacteria *A. deccanensis* and *Dactylosporangium aurantiacum* subsp. *hamdenensis* produce the same antibiotic, the differences in biosynthetic gene clusters therefore have no influence on the structure of the final natural products, which cannot be said for production yield dependent on sufficient supply of precursor moieties. A part of the pathway for biosynthesis and modifications of deoxysugar moieties was previously not completely clarified [80] but due to the absence of TiaS3 homolog in *A. deccanensis* genome, we can propose that only FdxS1 and FdxS4 GDP-D-mannose 4,6-dehydratases are necessary for the biosynthesis. FdxS1 shares 88 % identity to TiaS1, while FdxS4 shares 60 % and 94 % identity to TiaS3 and TiaS4, respectively. An ArsR family transcriptional regulator homolog TiaR2, absent in fidaxomicin BGC, may function as a negative regulator in antibiotic biosynthesis [80]. Both TiaT3 and TiaS4 are homologous to members of the ABC transporter superfamily involved in the transport of a wide variety of different compounds but are not observed in FDX gene cluster. However, there are genes marked *orf3* and *orf4* that could be fidaxomicin's transporter systems (Figure 46). The additional six ORFs, named FdxY1-Y2, FdxW1-W3 and FdxH by Li *et al*. [94], are present in FDX BGC. FdxY1 and FdxY2 are most likely not involved in fidaxomicin's biosynthesis, while FdxW1-W3 probably catalyze the reversible isomerization of methylmalonyl-CoA to succinyl-CoA, providing a precursor moiety [94]. The genome of *Actinoplanes deccanensis* ATCC 21983 strain was sequenced, assembled, analysed for this work but the genome sequence of the same strain was unfortunately published by Li *et al.* in 2019 [94].

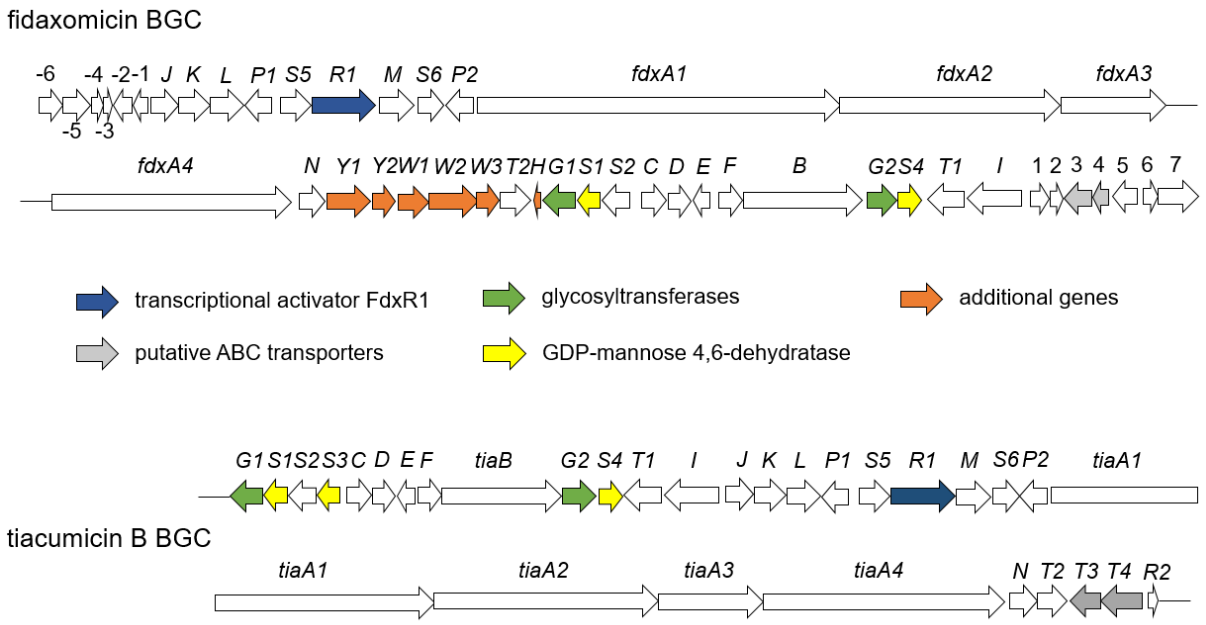



**Figure 46. Fidaxomicin and tiacumicin B gene cluster arrangement with highlighted differences and important proteins for the project.** FdxR1 is a transcriptional activator, which upon overexpression significantly improves production yield [94]. FdxG1 and FdxG2 are glycosyltransferases decorating the fidaxomicin aglycone. FdxY1-Y2, FdxW1-W3 and FdxH are not present in tiacumicin B gene cluster. Only two GDP-mannose 4,6-dehydratases, FdxS1 and FdxS4, are observed in BGC compared to tiacumicin B gene cluster that harbours an additional GDP-mannose 4,6-dehydratase TiaS3.

### 4.3.3 Glycosyltransferase-deficient *Actinoplanes deccanensis*

Two glycosyltransferases are present in tiacumicin B biosynthetic gene cluster [80]. TiaG1 and TiaG2 are responsible for the addition of modified D-noviose and D-rhamnose moieties on C-11 and C-20 positions in the molecule, respectively (Section 1.4.2.1, Figure 12). The glycosyltransferase homologs from *A. deccanensis*, FdxG1 and FdxG2, share 83 % identity with its homolog TiaG1 and 88 % identity with TiaG2, respectively. The primary mutant strain was prepared via *A. deccanensis* transformation of mycelium using the conjugation method with a plasmid constructed in a way that homology arms dictated the desired gene deletion. Upon successful incorporation of the pKC1132-KOfdxG1 and/or pKC1132-KOfdxG2 plasmid via homologous recombination, the mutant strains were cultivated in the absence of the selection marker, which caused the loss of the plasmid, converting the strain either into a mutant with a gene deletion or back into the wild type. The mutant strains *A. deccanensis* Δ*fdxG1* and *A. deccanensis* Δ*fdxG2* were prepared just like described above. On the other hand, *A. deccanensis* Δ*fdxG1*Δ*fdxG2* was prepared using *A. deccanensis* Δ*fdxG1* mutant strain for the deletion of the *fdxG2* gene. All mutant strains were cultivated in the production medium,

screened for the absence of fidaxomicin production and analyzed for the production of shunt products.

#### 4.3.3.1 Fermentation of *A. deccanensis* Δ*fdxG1*

According to Xiao *et al*. [80], five shunt products **3**-7 (Figure 47) were expected to be present in the crude extract of a mutant strain Δ*fdxG1*. They all lack a D-noviose moiety at C-11 and a hydroxyl group at C-18 position, whose attachments are catalyzed by glycosyltransferase FdxG1 and by cytochrome P450 FdxP1, respectively. It was reported that the TiaP1-catalyzed C-18 hydroxylation is the last step in the biosynthesis of fidaxomicin and requires the substrate with 'left' sugar moiety at C-11, the 'right' sugar moiety at C-20 and the aromatic ring coupled to C-4′, preferably all fully decorated by tailoring enzymes [96].

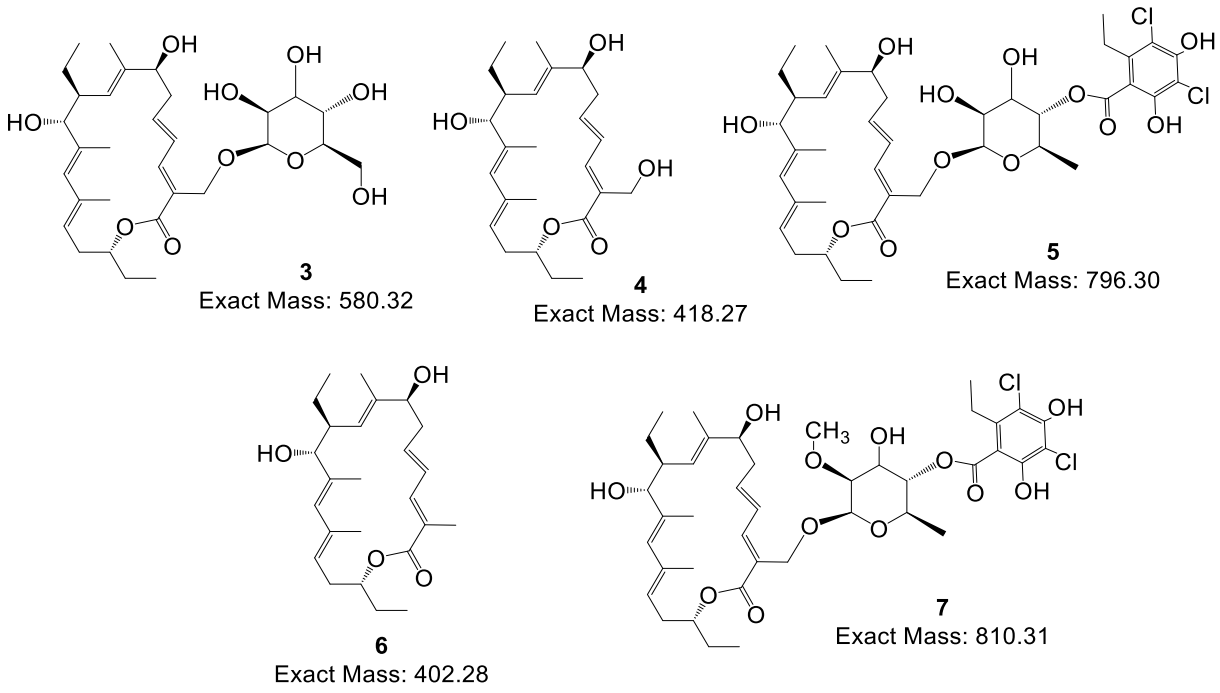



**Figure 47. Shunt products observed in *D. aurantiacum* subsp. *hamdenensis* Δ*tiaG1* mutant.** The shunt products were reported in a publication by Xiao *et al*. [80].

The *A. deccanensis* Δ*fdxG1* mutant was cultivated as described in Section 4.2.1.1 but further cultivations and purifications were done by Andrea Dorst from ETH Zürich. The UHPLC-MS analysis revealed the presence of three derivatives: the denoviosylated fidaxomicin **7**, denoviosylated and demethylated compound **5** and traces of rhamnosylated aglycone **3** (Figure 48). The macrolides **4** and **6** were not detected.

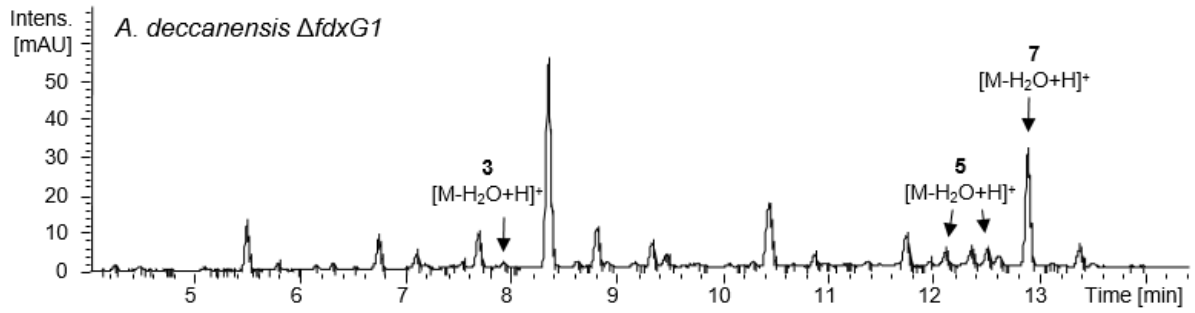



**Figure 48. The UHPLC-MS analysis of crude extract of *A. deccanensis* Δ*fdxG1* fermentation broth.** UV chromatogram at a detection wavelength of 270 nm is shown (chromatograms adapted from DataAnalysis (available from Bruker Daltonics)). The marked target peak was found through EIC-search of $[M\text{-}H_2O+H]^+$ of expected exact masses: m/z 563.317 (±0.02), 401.267 (±0.02), 779.297 (±0.02), 385.277 (±0.02) and 793.307 (±0.02).

### 4.3.3.2 Fermentation of *A. deccanensis* Δ*fdxG2*

The fermentation of *A. deccanensis* Δ*fdxG2* was performed using the same conditions as for the wild-type strain. According to Xiao *et al*. [80], four shunt products **8-11** (Figure 49) were expected to be present in the crude extract of a mutant strain Δ*fdxG2*, namely the desrhamnosylated compound **11** with isobutyric ester attached to noviose and the desrhamnosylated compounds with propionic ester **10**, acetic ester **9** and hydrolyzed ester **8**. They all lack the D-rhamnose-orsellinate moiety and the C-18 hydroxyl group attached by FdxP1 [96] as previously discussed.

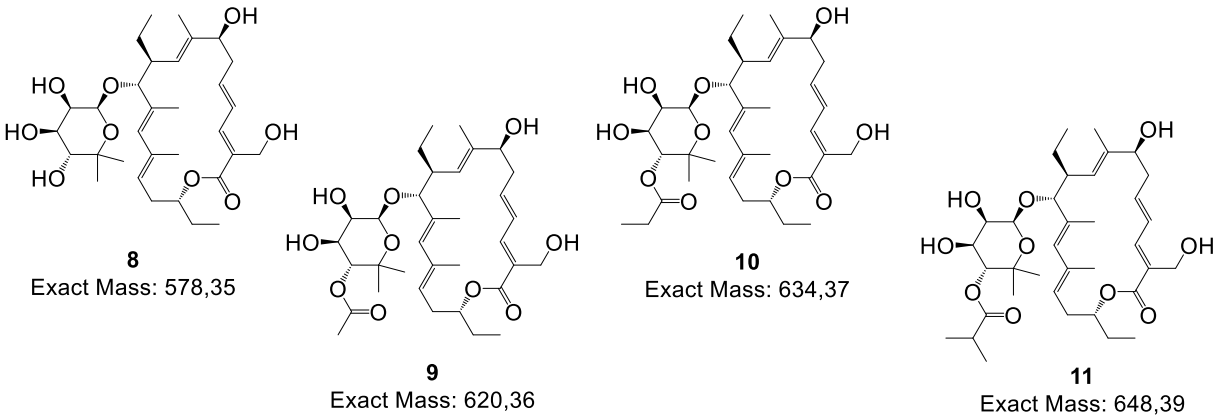



**Figure 49. Shunt products observed in *D. aurantiacum* subsp. *hamdenensis* Δ*tiaG2* mutant.** The shunt products were reported in a publication by Xiao *et al*. [80].

The UHPLC-MS analysis of the *A. deccanensis* Δ*fdxG2* crude extract by Andrea Dorst (ETH Zürich) revealed all four derivatives but the compound with the desired isobutyric ester **11** was present only in small amounts. The analysis of our crude extract showed the production of compounds **8** and **11** (Figure 50). The differences in crude extracts are nothing unusual. The fermentation yields of *A. deccanensis* wild-type strain are not strictly reproducible and they

depend on the growth of bacterium in the production medium. Baffled flask used for fermentation provide additional aeration required for actinobacteria but during the incubation of rotary shakers some bacterial cells attach to the flask´s wall. Those bacterial cells are no longer in production medium and have no access to nutrients, which influences the antibiotic yields. Moreover, the upscaling of the fermentation process usually results in lower yields.

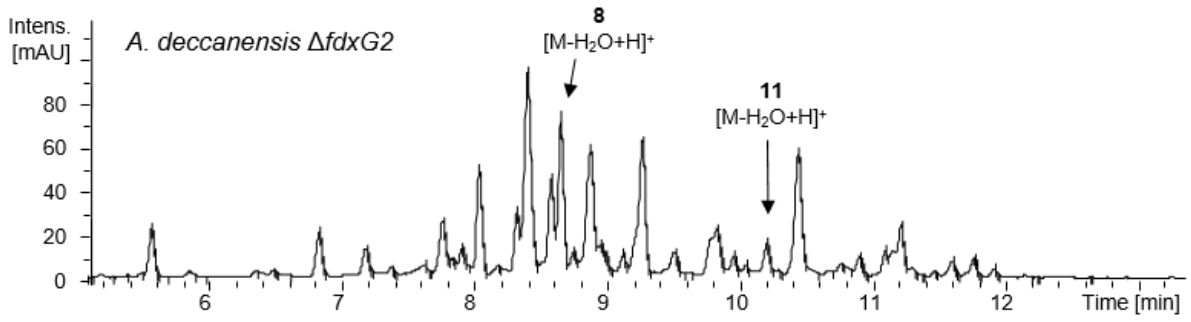



**Figure 50. The UHPLC-MS analysis of crude extract of *A. deccanensis* Δ*fdxG2* fermentation broth.** UV chromatogram at a detection wavelength of 270 nm is shown (chromatograms adapted from DataAnalysis (available from Bruker Daltonics)). The marked target peak was found through EIC-search of $[M\text{-}H_2O+H]^+$ of expected exact masses: m/z 561.347 (±0.02), 603.357 (±0.02), 617.367 (±0.02) and 631.387 (±0.02).

### 4.3.3.3 Fermentation of *A. deccanensis* Δ*fdxG1*Δ*fdxG2*

The mutant *A. deccanensis* Δ*fdxG1*Δ*fdxG2* was prepared the first time in this study. According to the functions of the deleted genes, compounds **4** and **6** (Figure 47) are expected to be present in the crude extract.

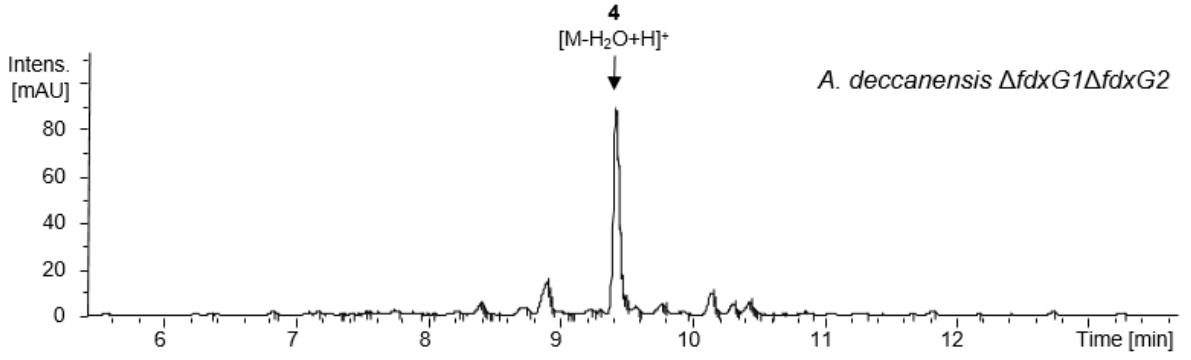



**Figure 51. The UHPLC-MS analysis of crude extract of *A. deccanensis* Δ*fdxG1*Δ*fdxG2* fermentation broth.** UV chromatogram at a detection wavelength of 270 nm is shown (chromatograms adapted from DataAnalysis (available from Bruker Daltonics)). The marked target peak was found through EIC-search of $[M\text{-}H_2O+H]^+$ of expected exact masses: m/z 401.26853 (±0.02) and 385.277 (±0.02).

The UHPLC-MS analysis of the *A. deccanensis* Δ*fdxG1*Δ*fdxG2* crude extract revealed presence of only compound **4**. The compound **6** was not detected (Figure 51).

## 4.4 Conclusions

We successfully prepared two *Actinoplanes deccanensis* strains with deleted glycosyltransferase FdxG1 or FdxG2 and a double knock-out mutant lacking both of the enzymes. Our aim was a production of certain shunt products, if possible in high quantities. The mutant strains were sent to the collaborators for cultivation and purification of desired shunt products that would serve as substrates for semi-synthesis. The denoviosylated fidaxomicin **7**, denoviosylated and demethylated compound **5** were successfully purified from the *A. deccanensis* Δ*fdxG1* and all four derivatives (**8-11**) were found in *A. deccanensis* Δ*fdxG2* (ETH Zürich). The double mutant *A. deccanensis* Δ*fdxG1*Δ*fdxG2* produces the macrolide **4**. However, those shunt products all lack C-18 hydroxyl group, which might affect the biological activity of derivatives prepared by semi-synthesis. In hands of our collaborators, the production yields of different shunt products isolated from *A. deccanensis* Δ*fdxG2* reached 66-195 mg $L^{-1}$. Although the yields are quite high, the fermentation must be performed in small volumes since the upscaling causes the decrease in production. However, the overexpression of the transcriptional activator FdxR1 was shown to significantly improve fidaxomicin yield and is expected to improve yield of shunt products 8-30-fold [94]. Up until now, a derivative **10** was used for attachment of rhamnose bound to the orsellinic acid with alkenated hydroxyl groups but more will be done in future. The semisynthetic modifications were mainly conducted on fidaxomicin (**1**), such as selective modification on the hydroxyl groups, modification on the phenol, chlorine substitutions, rhamnose-resorcylate cleavage and preparation of hybrid antibiotics.

## 4.5 Supplementary information

### Table of Tables



**Table S3- 1. List of primers used in this study**

| Primer | 5'-Sequence-3' | Restriction sites | Construct |
|---|---|---|---|
| fdxG1-PstI-LF | ATACTGCAGGTGCCCTCTGTCGAGTAAGAGG | *Pst*I | pKC1132fdxG1-left |
| fdxG1-XbaI-LR | ATATATTCTAGAGTTGATCAGCCCGGTGCTC | *Xba*I | |
| fdxG1-XbaI-RF | ATATATTCTAGAGAGGACGGTTTCCGCGAAC | *Xba*I | pKC1132-KOfdxG1 |
| fdxG1-EcoRV-RR | ATATGATATCACCTGCGGCAAGCTGATC | *Eco*RV | |
| fdxG2-PstI-LF | ATACTGCAGGTCCTGCTCGACGACGTG | *Pst*I | pKC1132fdxG2-left |
| fdxG2-XbaI-LR | ATATATTCTAGACAGCAGCGGATTGATCAGG | *Xba*I | |
| fdxG2-XbaI-RF | ATATATTCTAGACTCATCCAACGCCGGCTG | *Xba*I | pKC1132-KOfdxG1 |
| fdxG2-EcoRV-RR | ATATGATATCCGTCATCAGCTCCTCGAACTTCAG | *Eco*RV | |

| Primer | 5'-Sequence-3' | Usage |
|---|---|---|
| gAD-F | GTACCGGATCTCGGCGTGCT | Gene cluster arrangement |
| gAD-R | GATCGCCAGCGAGATCGTCTC | |
| gAD-F1 | CACCGCATCGGCATCTCG | |
| gAD-R1 | CGTAGCCGCCTCACCCATC | |
| gAD-R2 | GTGGTCGTCAGGAGCCGCTC | |
| gAD-R3 | GATCAGCATGCGGATCAGGTCG | |
| PKC1132-F | CGATTAAGTTGGGTAACGCCAGG | Anneals before *Pst*I site in pKC1132 |
| PKC1132-R | ATGCTTCCGGCTCGTATGTTGTGTG | Anneals after *Eco*RV site in pKC1132 |
| gfdxG1-LF | CGGTGTTCTTCATCGTCACCG | Screening of a plasmid orientation in first recombinants in combination with pKC1132-F or pKC1132-R primers |
| gfdxG1-RR | GGATACCAGGGCAAGTTCATCGTC | |
| gfdxG2-LF | ACAGCACGTACGCCGTCGTC | |
| gfdxG2-RR | CAATCGACGTGCGCGGAGTACT | |
| checkG1F | CTACAACATGGTCCGCAACTATC | Sequencing of pKC1132-KOfdxG1 and screening of *A. deccanensis* Δ*fdxG1* mutant |
| checkG1R | GTTCACGGCACTGACGAGGAC | |
| checkG2F | GTGGATCCGAGCCTGCTGT | Sequencing of pKC1132-KOfdxG2 and screening of *A. deccanensis* Δ*fdxG2* and *A. deccanensis* Δ*fdxG1*Δ*fdxG2* mutants |
| checkG2R | CTTGCCGAACATCTCCGACGACGAC | |

**Table S3- 2. Genetic constructs used in that study**

| Plasmid name | Details | Reference |
|---|---|---|
| pKC1132 | *aac(3)IV lacZα oriT*$_{RK2}$; suicide vector for *Streptomyces* containing *oriT* RK2 for conjugation from *E. coli* to *Streptomyces* | [261] |
| pKC1132 KOfdxG1 | Homologies 1.1 kb and 0.94 kb cloned in-frame containing first 63 bp and last 132 bp of gene *fdxG1* | This study |
| pKC1132 KOfdxG2 | Homologies of 1 kb cloned in-frame containing first 72 bp and last 75 bp of gene *fdxG2* | This study |

**Table S3- 3. Bacterial strains used or prepared in that study**

| Organism | Genotype or usage purpose | Reference |
|---|---|---|
| *Escherichia coli* | | |
| DH10β | F$^-$ *endA*1 *recA*1 *galE*15 *galK*16 *upG rpsL* Δ*lacX*74 Φ80*lacZ*ΔM15 *araD*139 Δ(*ara-leu*)7697 *mcrA* Δ(*mrr-hsdRMS-mcrBC*) λ$^-$ | Invitrogen |
| ET12567/pUZ8002 | F$^-$ *dam*13::Tn*9*, *dcm*6, *hsdM*, *hsdR*, *recF*143::Tn*1I*, *galK*2, *galT*22, *ara*14, *lacY*1, *xyl*5, *leuB*6, *thi*1, *tonA*31, *rpsL*136, *hisG*4, *tsx*78, *mtl*1 *glnV*44 pUZ8002 | [27, 262] |
| *Actinoplanes deccanensis* ATCC 21983 | Fidaxomicin producer; WT | [29] |
| *A. deccanensis* Δ*fdxG1* | Δ*fdxG1* | This study |
| *A. deccanensis* Δ*fdxG2* | Δ*fdxG2* | This study |
| *A. deccanensis* Δ*fdxG1* Δ*fdxG1* | Δ*fdxG1* Δ*fdxG2* | This study |

**Table S3- 4. Protein homologs from fidaxomicin gene cluster with proposed functions**

| *Gene* | *Size*[a] | *Protein homolog*[b] | *Proposed function* |
|---|---|---|---|
| *orf(-6)* | 348 | WP_091298862.1 (97/99) | hypothetical protein |
| *orf(-5)* | 386 | WP_091298811.1 (96/99) | cytochrome P450 |
| *orf(-4)* | 159 | WP_091298807.1 (99/100) | peptidylprolyl isomerase |
| *orf(-3)* | 133 | WP_018728983.1 (89/96) | DNA-binding protein |
| *orf(-2)* | 273 | TWG21195.1 (84/96) | metallophosphoesterase |
| *orf(-1)* | 218 | ASV46775.1 (73/86) | 4'-phosphopantetheinyl transferase |
| *fdxJ* | 408 | AYW03306.1 | putative 3-hydroxybutyryl-CoA dehydrogenase |
| *fdxK* | 450 | AYW03305.1 | putative crotonyl CoA reductase |
| *fdxL* | 478 | AYW03304.1 | putative propionyl-CoA carboxylase |
| *fdxP1* | 392 | AYW03303.1 | putative cytochrome P450 hydroxylase |
| *fdxS5* | 440 | AYW03302.1 | putative sugar O-methyltransferase |
| *fdxR1* | 912 | AYW03301.1 | putative LuxR class regulator |
| *fdxM* | 495 | AYW03300.1 | putative tryptophan halogenase |
| *fdxS6* | 379 | AYW03299.1 | putative O-acyltransferase |
| *fdxP2* | 399 | AYW03298.1 | putative cytochrome P450 |
| *fdxA1* | 5254 | AYW03297.1 | putative modular polyketide synthase |
| *fdxA2* | 3195 | AYW03296.1 | putative modular polyketide synthase |
| *fdxA3* | 1516 | AYW03295.1 | putative modular polyketide synthase |
| *fdxA4* | 3462 | AYW03294.1 | putative modular polyketide synthase |
| *fdxN* | 385 | AYW03293.1 | putative enoyl-CoA hydratase/isomerase |
| *fdxY1* | 620 | AYW03292.1 | 2-oxoacid:acceptor oxidoreductase subunit alpha |

| | | | |
|---|---|---|---|
| *fdxY2* | 333 | AYW03291.1 | 2-oxoacid:ferredoxin oxidoreductase subunit beta |
| *fdxW1* | 450 | AYW03290.1 | methylmalonyl-CoA mutase |
| *fdxW2* | 714 | AYW03289.1 | methylmalonyl-CoA mutase |
| *fdxW3* | 332 | AYW03288.1 | methylmalonyl CoA mutase-associated GTPase MeaB |
| *fdxT2* | 443 | AYW03287.1 | putative membrane antiporter |
| *fdxH* | 98 | AYW03286.1 | transposase |
| *fdxG1* | 463 | AYW03285.1 | putative glycosyltransferase |
| *fdxS1* | 333 | AYW03284.1 | putative GDP-D-mannose dehydratase |
| *fdxS2* | 402 | AYW03283.1 | putative C-methyltransferase |
| *fdxC* | 359 | AYW03282.1 | putative branched-chain alpha keto acid dehydrogenase E1 alpha subunit |
| *fdxD* | 325 | AYW03281.1 | putative transketolase central region alpha-ketoacid dehydrogenase subunit beta |
| *fdxE* | 248 | AYW03280.1 | thioesterase |
| *fdxF* | 343 | AYW03279.1 | putative 3-oxoacyl-ACP synthase III |
| *fdxB* | 1715 | AYW03278.1 | putative iterative type I polyketide synthase |
| *fdxG2* | 417 | AYW03277.1 | putative glycosyltransferase |
| *fdxS4* | 344 | AYW03276.1 | GDP-mannose 4,6 dehydratase |
| *fdxT1* | 528 | AYW03275.1 | putative membrane transport protein |
| *fdxI* | 783 | AYW03274.1 | putative ATP-dependent DNA helicase |
| *orf1* | 274 | SFE86170.1 (87/96) | S-adenosyl methyltransferase |
| *orf2* | 192 | WP_143234489.1 (87/95) | hypothetical protein |
| *orf3* | 387 | WP_097319205.1 (94/98) | ABC transporter permease |
| *orf4* | 223 | SNY26207.1 (93/98) | putative ABC transport system ATP-binding protein |
| *orf5* | 369 | WP_097319206.1 (83/92) | hypothetical protein |
| *orf6* | 231 | WP_097319207.1 (92/98) | DNA-binding response regulator |
| *orf7* | 587 | WP_097319208.1 (89/97) | sensor histidine kinase |

[a] *Sizes are given in amino acids;* [b] *Accession numbers and percentage of identity/similarity are given in parentheses.*

# 5 Biosynthetic engineering of enzymes involved in chelocardin biosynthesis: Cyclization and C6 methylation reactions

## 5.1 Introduction

Considering the wide spread of antibiotic resistance among pathogens, the biosynthesis and its regulation of medicinally still important antibiotics are studied in-depth, for either development of new analogs or improvement of production yield, respectively. This work focuses on chelocardin (CHD), a broad-spectrum atypical tetracycline, discovered in 1962, produced by a Gram-positive bacterium *Amycolatopsis sulphurea* [277]. The biosynthetic pathway of CHD was proposed based on other tetracyclines in 2013 [32]. Tetracyclines are aromatic polyketides assembled by an iterative PKS type II using simple building blocks from primary metabolism. Their characteristic tetracyclic scaffold is cyclized by for that responsible enzymes and subsequently decorated by tailoring enzymes with oxygenase, methyltransferase and aminotransferase activities [99]. The cyclization of the poly-β-keto intermediate is believed to be a critical step in the biosynthesis of natural products and a potential diversification point for engineered biosynthesis [278]. Poly-β-keto intermediates are highly unstable and prone to spontaneous cyclization [99], so the multienzyme PKS type II is believed to provide an adequate environment for directed and controlled cyclization [278]. The described first-ring aromatases/cyclases reported in the biosynthesis of aromatic polyketides catalyze cyclization either between C9–C14 or C7–C12 carbon atoms (Figure 52). They are further divided into two classes: reducing PKS systems, where C9 ketoreductase (KR) is present, and non-reducing PKS systems, lacking KR activity [99, 279]. The example for cyclization between C9–C14 is TcmN from tetracenomycin biosynthesis, is carried out by a mono-domain aromatase/cyclases (ARO/CYC) from non-reducing PKS system, which also catalyzes the second ring cyclization between C7-C16 [278, 280]. On the other hand, the cyclization between C7–C12 is commonly associated with the di-domain ARO/CYCs from reducing PKS systems such as BexL from biosynthesis of anticancer agent BE-7585A (Figure 52) [278, 281]. However, C7–C12 cyclization may also be accomplished in non-reducing systems by mono-domain ARO/CYC such as ZhuI from biosynthesis of R1128 or di-domain ARO/CYCs such as CmmQ from chromomycin biosynthesis (Figure 52) [278, 282]. In non-reducing systems, the di-domain ARO/CYC cyclizes C7-C12 cyclization, followed by aromatization [281], while in reducing systems, the timing of the first ring cyclization was not clear at first and was thought to occur after C9 ketoreduction of the polyketide chain [283].

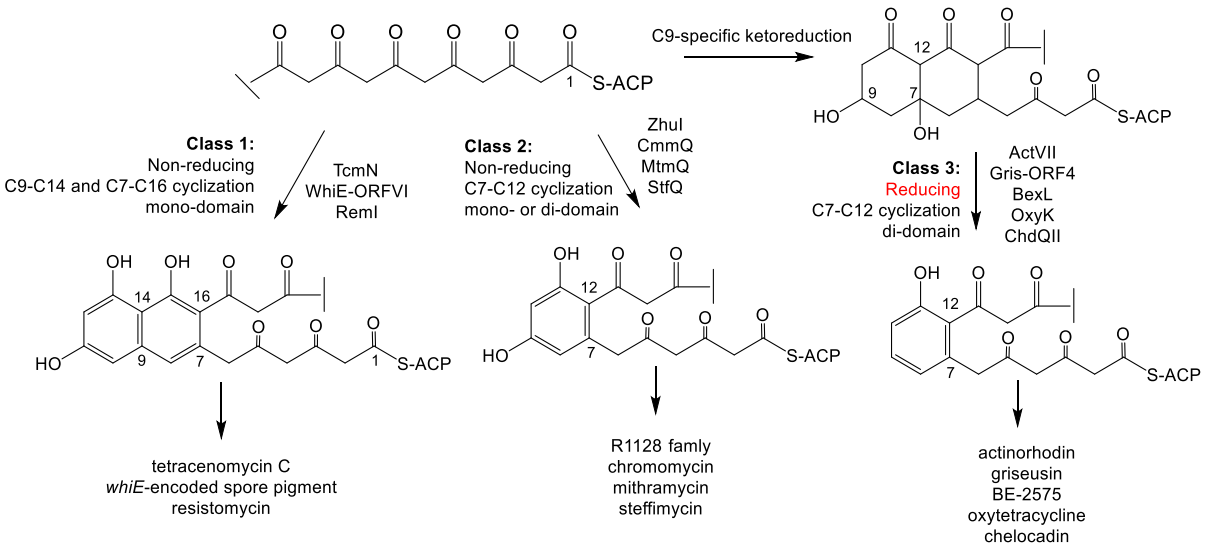



**Figure 52. Three classes of ARO/CYC in type II PKS systems.** In class 1 belong the monodomain ARO/CYCs that promote C9-C14 and C7-C16 cyclization of an unreduced polyketide intermediate. The members of 2nd class are mono- and di-domain ARO/CYCs that promote C7-C12 cyclization of an unreduced polyketide intermediate as well. The 3rd class contains di-domain ARO/CYCs that act only after the C9-specific ketoreduction and promote C7-C12 cyclization [284].

Later it was described that KR first regiospecifically cyclizes the linear poly-β-keto intermediate between C12-C7, followed by C9 ketoreduction. A di-domain ARO/CYC then catalyzes the dehydration of the C9 hydroxyl group, followed by aromatization of pre-cyclized first ring (Figure 52) [281]. The second-ring (C) cyclization requires distinct enzymes such as DpsY from anthracycline biosynthesis catalyzing cyclization between C5-C14, followed by aromatization [278, 285]. The third ring (B) cyclization between C2-C15 occurs spontaneously, while the fourth ring cyclization is still not clear due to many opposing observations in different biosynthetic pathways. The studies of fourth-ring (A) cyclization propose the involvement of specific enzyme like proposed for MtmX from mithramycin biosynthesis [286]. It was also suggested to occur spontaneously in tetracycline biosynthesis due to the terminal amide moiety [101, 287]. This hypothesis was later disproven since the presence of OxyH in the expression construct was overlooked [288]. On the other hand, it was proposed by Pickens *et al*. that acyl-CoA-ligase SsfL2, homologous to OxyH from the *oxy* gene cluster and ChdL from the *chd* gene cluster, is involved in the formation of the fourth ring of the TC-like compound SF2527 [148], which was tested *in vitro* and concluded that SsfL2 and OxyH directly adenylate the tricyclic carboxylic acid to facilitate the final C1-C18 bond formation (Figure 53) [288]. Unrelated to cyclization of chelocardin is well-characterized tailoring reaction of methylation of C6 position [287]. The 6-methyl group in tetracyclines is not essential for its biological activity but 6-desmethyl tetracycline exhibits better stability in basic conditions [289].

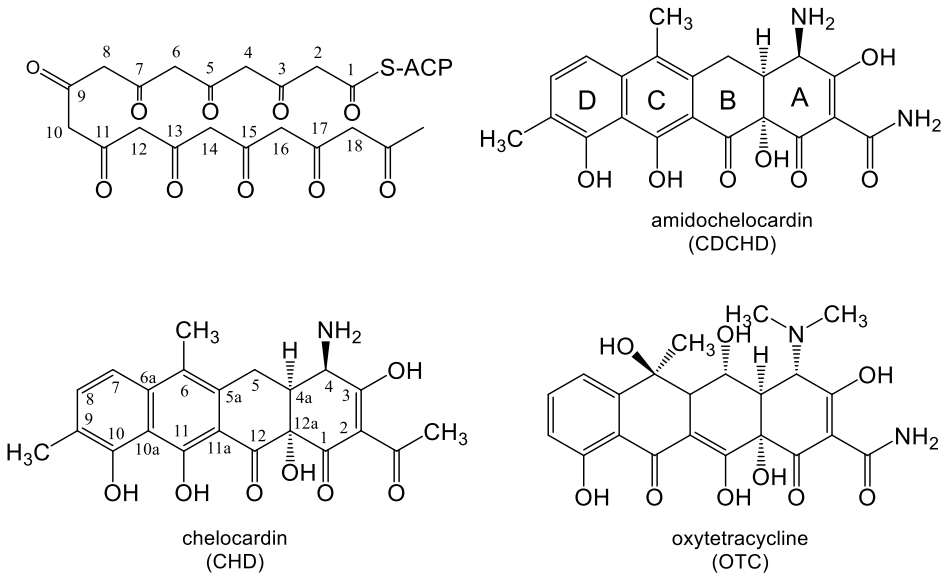



**Figure 53. The numbered acetate-primed decaketide chain and chemical structures of amidochelocardin, chelocardin and oxytetracycline.** The stereochemistry of chemical structures is shown as previously reported.

## 5.2 Materials and methods

### 5.2.1 Bacterial strains used in this study

*Amycolatopsis sulphurea* NRRL2822 (ARS Culture Collection) was used for the production of CHD and as a source of DNA for microbiological manipulations. *Streptomyces rimosus* M4018 [290] was used as a source of DNA. *Escherichia coli* DH10β (Invitrogen) was used for standard cloning procedures [160], *E. coli* ET12567 [262] and SCS110 (Stratagene) strains for isolation of non-methylated plasmid DNA, suitable for transformation of *A. sulphurea*.

### 5.2.2 Media and cultivation conditions for *A. sulphurea*

Soya mannitol (MS) agar and tryptone soy broth (TSB) [27] with incubation at 30 °C were used for sporulation and cultivation of actinobacteria in liquid medium, respectively. For CHD and CHD analogues production, *A. sulphurea* was cultivated in CH-V seed medium (1.5 % soy flour, 0.1 % yeast extract, 1.5 % glucose, 0.5 % NaCl, 0.1 % $CaCO_3$, pH 7.0) and CH-F2 production medium (2 % soy flour, 0.5 % yeast extract, 0.2 % $CaCO_3$, 0.05 % citric acid, 5 % glucose, pH 7.0) (adapted from [277, 291, 292]). Cultivations were performed in Falcon tubes at 30 °C on a rotary shaker at 220 rpm for 36 h in seed medium with 15 % (v/v) used to inoculate CH-F2 production medium and cultivated for further 7 days under the same conditions. For transformation of *A. sulphurea*, S27M and R2L media were used [293]. Apramycin (Apr; 200 μg $mL^{-1}$), erythromycin (Erm; 20 μg $mL^{-1}$) or kanamycin (Kan; 300 μg $mL^{-1}$) was used for selection of *A. sulphurea* transformants on S27M. For further subcultivation of *A. sulphurea*

transformants, MS was supplemented with Apr (400 μg $mL^{-1}$), Erm (20 μg $mL^{-1}$) or kanamycin (400 μg $mL^{-1}$). For selection of *E. coli* transformants, ampicillin (Amp; 100 μg $mL^{-1}$), Apr (50 μg $mL^{-1}$), Kan (25 μg $mL^{-1}$) or chloramphenicol (Cm; 10 μg $mL^{-1}$) were added into LB medium.

### 5.2.3 Competent cell preparation and transformation of *Amycolatopsis* species

*A. sulphurea* spores from MS agar plates were inoculated into 5 mL of TSB medium and incubated at 30 °C and 200 rpm for 48 h. After 48 h of incubation, 100 mL of fresh TSB medium was inoculated with 1 % of pre-culture into a baffled 1 L Erlenmeyer flask and incubated again at 30 °C and 200 rpm for another 48 h. The culture was then centrifuged at 5000 rpm for 15 min. Every 5 g of pellet was resuspended in 4 ml of 20 % glycerol. Aliquots of 500 μl were stored at -20 °C. Transformation of *A. sulphurea* NRRL2822 was carried out by the protocol for transformation of mycelium of *A. mediterranei* [293], using vectors pAB03 and pNV18 already described previously [32], described below. For transformation, the aliquot of competent cells was thawed and washed three times in 1.5 ml TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0). Cells were centrifuged for 2 min at 5000 rpm. After the last centrifugation, the mycelium was resuspended in 125 μl TE buffed and divided into aliquots of 50 μl, to which we added 31.3 μl of 4 M CsCl, 2 μl of thymus DNA (10 mg $mL^{-1}$ calf thymus DNA resuspended in TE, sheared to low viscosity, centrifuged, incubated at 100 °C for 20 min and stored at -20 °C), 2 μl of plasmid DNA (isolated with the GeneJet Miniprep kit) and 100 μl of 70 % PEG1000 in TE buffer [293]. Transformation mixture was homogenized and incubated first at 30 °C for 40 min and later at 42 °C for 5 min. To each 0.2 ml of transformation mixture 2 ml of R2L medium was added and plated on S27M plates, which were previously dried for 2 h. After 16 h of incubation at 30 °C, the plates were overlaid with 1 ml of the antibiotic solution in autoclaved distilled water and incubated at 30 °C for 10 days.

### 5.2.4 Sequencing of genomic DNA

Genomic DNA of *A. sulphurea* was isolated using phenol/chloroform according to the standard protocol [160] and was sent for Illumina sequencing. The sequencing reads of chelocardin biosynthetic gene cluster were assembled by Nestor Zaburannyi.

### 5.2.5 Construction of plasmids for genetic manipulation of *A. sulphurea*

Routine handling of nucleic acids, such as isolation of plasmid DNA, restriction endonuclease digestions, DNA ligations, and other DNA manipulations, was performed according to standard protocols [27, 160]. *E. coli* DH10β (Invitrogen) was used as host for standard cloning

experiments. *E. coli* strains were cultured in 2TY medium or on 2TY agar (1.6 % tryptone, 1 % yeast extract, 0.5 % NaCl, (1.8 % agar), $dH_2O$) at 30-37 °C, at 200 rpm overnight. Transformation of *E. coli* strains was achieved via electroporation in 0.1 cm-wide cuvettes at 1350 V, 200 Ω, 25 μF. Plasmid DNA was purified by standard alkaline lysis [160] or by using the GeneJet Plasmid Miniprep Kit (Thermo Fisher Scientific). Restriction endonucleases, alkaline phosphatase (FastAP) and T4 DNA ligase were obtained from Thermo Fisher Scientific. Oligonucleotides used for PCR and sequencing were obtained from Sigma-Aldrich and are listed in Table S4- 1. PCR reactions were carried out in Mastercycler® pro (Eppendorf) using Phusion™ High-Fidelity or Taq DNA polymerase (Thermo Fisher Scientific) according to the manufacturer's protocol. For Taq: Initial denaturation (2 min, 95 °C); 30 cycles of denaturation (30 s, 95 °C), annealing (30 s, -5 °C of lower primer's $T_m$) and elongation (based on PCR product length 1 kb/min, 72 °C); and final extension (10 min, 72 °C). For Phusion™: Initial denaturation (2 min, 98 °C); 30 cycles of denaturation (15 s, 98 °C), annealing (30 s, $T_m$ of the lower primer) and elongation (based on PCR product length 0.5 kb/min, 72 °C); and final extension (10 min, 72 °C). PCR products or DNA fragments from restriction digests were purified by agarose gel electrophoresis and isolated using NucleoSpin Gel and PCR Clean up kit (Macherey-Nagel). The digested PCR products were cloned into suitable plasmids pNV18Erm and pAB03, verified by restriction analysis and sequenced using sequencing primers (Table S4- 1). Details on the construction of all plasmids and bacterial strains used and generated in this study are given in Table S4- 2.

### 5.2.6 Genetic engineering of *A. sulphurea*

#### 5.2.6.1 *chdQI* and *chdMI* gene disruption in *A. sulphurea*

A putative di-domain aromatase/cyclase and methyltransferase encoded by *chdQI* and *chdMI* genes, respectively, were disrupted in order to assign their function in the chelocardin biosynthetic pathway. First, the *chdQI* gene was disrupted by allelic replacement through a double-crossover event. Plasmid pNV18chdQIErm was constructed by cloning *ermE* [294], flanked by two DNA fragments, 1.1 kb in length, homologous to sequences upstream and downstream of the core of the gene for a putative aromatase *ChdQI*. The *ermE* gene was amplified by PCR using primers FSB01C and FSB02C (Table S4- 1), the 1.6-kb DNA fragment was digested with *EcoR*I/*Xba*I and ligated into pNV18 to obtain pNV18Erm. The left arm of homology was amplified by PCR using primers FSB05 and FSB06 and digested with *Spe*I/*Sph*I, while the right arm by FSB03 and FSB04 primers and digested with *EcoR*I/*Nde*I. Both homology arms were gradually ligated into appropriately digested pNV18Erm. Similarly, the

homology arms were amplified on both sides of *chdMI* (left: chdMI-KO-LF/chdMI-KO-LR, 1 kb; right: chdMI-KO-RF/chdMI-KO-RR, 1.1 kb) and cloned left and right of *ermE* gene in pNV18Erm after digestion with *Sph*I/*Xba*I and *Nde*I/*Eco*RI, respectively, generating pNV18ErmKOchdMI. The resulting plasmids was transformed into *E. coli* ET12567 [262] to obtain the non-methylated plasmid, which was then introduced into *A. sulphurea*. S27M plates were overlaid with Erm after overnight incubation. Each transformant colony was further repatched onto MS agar containing Erm and subcultivated. After few subcultivations on MS plates or in liquid TSB medium, Erm-resistant/Kan-sensitive ($Erm^R/Kan^S$) colonies (secondary recombinants) of *A. sulphurea* Δ*chdQI* and *A. sulphurea* Δ*chdMI* could be isolated and the double-crossover confirmed by PCR.

#### 5.2.6.2 Site-directed mutagenesis of *chdQI, chdQII, chdY* and *chdX* in *A. sulphurea*

The site-directed mutagenesis approach was used to inactivate the catalytic residues of putative aromatases/cyclases ChdQI and ChdQIIa and putative cyclases ChdY and ChdX. Catalytic residues of *chdQI, chdQII, chdY* and *chdX* genes were independently replaced by Ala or Ser using site-directed mutagenesis. The mutations were introduced using the double cross-over approach to replace the target gene with the mutated gene. To mutate residue Arg72 in ChdQI, 0.7 kb upstream and 0.7 kb downstream fragments were PCR amplified, using chdQIalaLF/chdQIalaLR and chdQIalaRF/chdQIalaRR primer pairs (Table S4- 1), respectively. Primers labeled with LR (left reverse) and RF (right forward) were designed to anneal to the region containing the catalytic residue Arg72 and introduce the desired mutation. Third PCR was performed with outer set of primers, chdQIalaLF and chdQIalaRR, using previous two PCR products as template, which overlapped in the region where the mutation was introduced, yielding 1.4 kb fragment. Resulting DNA fragment was digested with *Sph*I/*Spe*I and ligated into pNV18Erm to obtain pNV18ErmchdQIala. The latter was transformed into *E. coli* SCS110 (Stratagene) to obtain the non-methylated plasmid, which was then introduced into *A. sulphurea*. S27M plates were overlaid with Erm after overnight incubation. Each transformant colony was further re-patched onto MS agar containing Erm and subcultivated. After three or more subcultivations in TSB without antibiotic, Erm-sensitive ($Erm^s$) colonies (secondary recombinants) were isolated. To confirm that secondary recombinants contain the introduced mutation and are not revertants to wild-type, colony PCR using the outer pair of primers labeled with LF (left forward) and RR (right reverse) - chdQIalaLF/chdQIalaRR was performed, followed by DNA sequencing. The same approach

was used in all the following site-directed mutagenesis experiments unless otherwise indicated. The details are presented in Table 10.

**Table 10. Details on constructs for site-directed mutagenesis of *chdQI, chdQII, chdY* and *chdX* genes**

| Mutation | Fragment length | Primers | Generated plasmid |
|---|---|---|---|
| ChdQI-Arg72Ala | 1.424 kb | chdQIalaLF, chdQIalaLR<br>chdQIalaRF, chdQIalaRR | pNV18ErmchdQIala |
| ChdQI-His115Ala | 1.396 kb | chdQIH115ALF,<br>chdQIH115ALR<br>chdQIH115ARF,<br>chdQIH115ARR | pNV18ErmchdQIH115A |
| ChdQI-Arg220Ala | 1.384 kb | chdQICalaLF, chdQICalaLR<br>chdQICalaRF, chdQICalaRR | pNV18ErmchdQICala |
| ChdQII-Arg71Ala | 1.427 kb | chdQIIalaLF, chdQIIalaLR<br>chdQIIalaRF, chdQIIalaRR | pNV18ErmchdQIIala |
| ChdQII-Arg224Ala | 1.374 kb | chdQIICalaLF, chdQIICalaLR<br>chdQIICalaRF, chdQIICalaRR | pNV18ErmchdQIICala |
| ChdY-Gly176Ser | 1.230 kb | chdYserLF, chdYserLR<br>chdYserRF, chdYserRR | pNV18ErmchdYser |
| ChdX-Asp126Ala | 1.362 kb | chdXalaLF, chdXalaLR<br>chdXalaRF, chdXalaRR | pNV18ErmchdXala |

### 5.2.6.3 "In trans" expression of *chdQI, chdQI*-R72A and *chdQI*-R220A genes in *A. sulphurea* Δ*chdQI*

For complementation of *A. sulphurea* Δ*chdQI* mutant, we constructed plasmids pAB03chdQI, pAB03chdQIala (containing mutation Arg72Ala) and pAB03chdQICala (containing mutation Arg220Ala) (Table S4- 2). Fragments with native *chdQI* gene, *chdQIala* gene with mutation Arg72Ala and *chdQICala* gene with mutation Arg220Ala were PCR amplified using chdQIF and chdQIR primers (Table S4- 1) while the templates were genomic DNA of *A. sulphurea* and pNV18ErmchdQIala (ChdQI-R72A), pNV18ErmchdQICala (ChdQI-R220A) plasmids, respectively. Obtained PCR products were digested with *Nde*I/*Xba*I and separately ligated into an integrative pAB03 plasmid, downstream of the *PactI* promoter. Constructs were confirmed by DNA sequencing and used to transform *E. coli* SCS110 and introduced into *A. sulphurea* Δ*chdQI* strain.

### 5.2.6.4 "In trans" expression of wild-type *chdQI, chdQII, chdY, chdOII-chdY* and *chdX* genes in *A. sulphurea* engineered strains

*A. sulphurea* engineered strains obtained through previously described site-directed mutagenesis approach (Section 5.2.5.2) were *in trans* complemented with wild-type genes *chdQI, chdQII, chdY, chdOII-chdY* and *chdX* from *A. sulphurea* via integrase-mediated

chromosomal integration [295]. Gene for ChdQI (aromatase) was obtained as described above and ligated into pAB03. Genes for ChdQII (aromatase), ChdY (cyclase), ChdOII-ChdY (oxygenase-cyclase fusion) and ChdX (cyclase) were amplified by PCR using chdQIIF/chdQIIR, chdYF/chdYR, chdOIIF/chdYR or chdXF/chdXR sets of primers (Table S4-1), respectively, using genomic DNA of *A. sulphurea* as a template. PCR products were digested with *Nde*I/*Xba*I and separately cloned into pAB03, resulting in pAB03chdQII, pAB03chdY, pAB03chdOII-chdY and pAB03chdX, respectively. Constructs were confirmed by sequencing, transformed into *E. coli* SCS110 and introduced into *A. sulphurea* via transformation of its mycelium. Plasmid pAB03chdQI was used for expression of wild-type gene *chdQI* in *A. sulphurea* ChdQI-H115A, plasmid pAB03chdQII was integrated into genomes of *A. sulphurea* ChdQII-R71A and *A. sulphurea* ChdQII-R224A, pAB03chdY and pAB03chdOII-chdY separately into *A. sulphurea* ChdY-G176S, and pAB03chdX into *A. sulphurea* ChdX-D126A.

#### 5.2.6.5 Heterologous expression of *oxyK, oxyN, oxyI* and *oxyDP* in engineered strains of *A. sulphurea*

Different *A. sulphurea* engineered strains were complemented *in trans* with genes from OTC gene cluster encoding an aromatase/cyclase OxyK, a cyclase OxyN, a putative cyclase OxyI [296] from *S. rimosus* M4018 [290]. Genes *oxyK*, *oxyN*, *oxyI* and *oxyDP* were amplified by PCR using oxyKFw/oxyKRv, oxyNFw/oxyNRv and oxyIFw/oxyIRv primers (Table S4- 1), respectively, digested with *Nde*I/*Xba*I and separately cloned into pAB03 vector, downstream of the *PactI* promoter, resulting in pAB03oxyK, pAB03oxyN and pAB03oxyI. Constructs were confirmed by sequencing, transformed into *E. coli* SCS110 and introduced via transformation of mycelium into engineered strains of *A. sulphurea.* Plasmid pAB03oxyK was integrated into genome of strains *A. sulphurea* Δ*chdQI, A. sulphurea* ChdQI-H115A, *A. sulphurea* ChdQII-R72A and *A. sulphurea* ChdQII-R224A, plasmid pAB03oxyN into *A. sulphurea* ChdY-G176S and plasmid pAB03oxyI into *A. sulphurea* ChdX-D126A via integrase-mediated chromosomal integration [295]. Genes for OxyD and OxyP were introduced through integration of pAB03oxyDP [297] into *A. sulphurea* ChdX-D126A and *A. sulphurea* Δ*chdMI* mutant strains.

### 5.2.7 Extraction, isolation and quantification of CHD and derivatives

To confirm production of CHD, CDCHD and intermediate products, *A. sulphurea* culture broth was acidified to pH 1-2 with 50 % TFA, followed by extraction with 2V of MeOH. The extract was centrifuged and analyzed by LC-MS. All measurements were performed on a Dionex

Ultimate 3000 LC system using a Luna C-18 (2) HST, 100 × 2.0 mm, 2.5 µm column (Phenomenex) coupled with an electrospray ionization (ESI) source linked to an amaZon speed MS (Bruker Daltonics; 3D ion trap)). Separation of 1 µl sample was achieved by a linear gradient from (A) $H_2O$ + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 500 µl $min^{-1}$ and 45 °C. For determination of exact mass of shunt products, UltiMate 3000 LC System with a 1.7 µm Acquity UPLC BEH C-18 column (100 x 2 mm) was coupled to an ESI source and hyphenated to maXis 4G MS system (Bruker Daltonics; Q-ToF). Separation of 1 µL sample took place at 45 °C, at a flow rate 0.6 ml $min^{-1}$ of eluents ACN and ultrapure water, both with 0.1 % HPLC grade FA. A standardized HPLC linear gradient was used (5-95 % ACN in 17.5 min). Absorption spectra (200-600 nm) were recorded by a diode array detector. Capillary voltage in the ESI source was set to 4.000 V and ESI source temperature was 200 °C. Dry gas flow rate was 5 l $min^{-1}$ at a nebulizer gas pressure of 1 bar. Mass spectra were acquired in centroid mode ranging from 200 to 2000 m/z (amaZon) or from 150 to 2500 m/z (maXis 4G) in positive ionization mode.

#### 5.2.7.1 Isolation of CHD-369 shunt product

CHD-369 was isolated from *A. sulphurea*-ChdYG176S broth using flash chromatography and semi-preparative HPLC method. Broth was acidified with HCl to pH 1-2 and CHD analogue was extracted with ethyl acetate. The extract was evaporated, dissolved in methanol and fractionated with Biotage Isolera system using a CHROMABOND Flash BT 40 SiOH Cartridge. Separation of the sample was achieved isocratically with 91 % (A) $CHCl_3$ + 0,1 % FA and 9 % (B) MeOH + 0,1 % FA after 16 min at a flow rate of 50 ml/min. UV spectra were recorded by a diode array detector at 280 and 310 nm. Fractions, containing CHD analogue, were evaporated, dissolved in MeOH and then loaded on a Biphenyl 100 Å, 250 x 10.0 mm column with 5 µm particles (Kinetex, Phenomenex). A Dionex Ultimate 3000 UHPLC system was used. Separation of 20 µL sample was achieved by a linear gradient from (A) $H_2O$ to (B) ACN at a flow rate of 5 mL $min^{-1}$ and 45 °C. The gradient was increased from 50 % B to 95 % B in 23 min followed by a 2 min step at 95 % B before equilibration to the initial conditions. UV spectra were recorded by a diode array detector at 220 and 280 nm.

#### 5.2.7.2 Isolation of CHD-398 shunt product

CHD-398 derivative was isolated from 50 mL culture broth of *A. sulphurea* Δ*chdMI*. The culture broth was acidified with 2 mL 50 % TFA, to reach pH around 1-2, and extracted with ethyl acetate. The extract was dried under $N_2$ and dissolved in methanol to be used for semi-

prep HPLC purification. CHD389 derivative was separated on column Jupiter 4u Proteo 90 Å, 250 x 10.0 mm (Phenomenex). Separation of each 20 µL of sample, with Dionex Ultimate 3000 UHPLC system, was achieved by a isocratic gradient from (A) H2O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 5 ml $min^{-1}$ and 45 °C. A 31-minute gradient was set for 3 min to 10 % B, increased to 25 % in $5^{th}$ minute and kept at 25 % B for 20 minutes. It reached 95 % B in a minute, kept at the same solvent mixture for another 2 min, lowered to 10 % B in a minute and kept there for 2 min. UV spectra were recorded by a DAD at 220 and 280 nm.

## 5.3 Results and discussion

### 5.3.1 Chelocardin biosynthetic gene cluster

Sequencing of the entire genome of *A. sulphurea* ATTC 27624 revealed that previously published CHD biosynthetic gene cluster [32] lacked three putative genes in total, due to the chimeric cosmid, which was originally sequenced. One of those, *chdY* gene encoding a putative second ring cyclase, a homolog of OxyN in OTC biosynthesis [101], is essential for the CHD biosynthesis. We have also corrected the original sequence of *chdOII* gene, revealing even higher similarity to a putative oxygenase OxyL involved in hydroxylation of the C4 and C12a positions of the oxytetracycline scaffold [101]. ChdOII and ChdY are encoded as a fusion protein unlike separately encoded homologs OxyL and OxyN found in OTC biosynthetic gene cluster, while *chdL* and *chdOIII* form a fusion protein just as their homologs *oxyH*, *oxyG*. The sequencing of *A. sulphurea* also revealed two regulatory genes, *chdB* and *chdC*, encoding SARP and LuxR, respectively, whose homologs are also found to regulate OTC and CTC biosynthesis [298, 299].

### 5.3.2 Revised biosynthesis of CHD

Although atypical TC, chelocardin can be directly compared to OTC since all *oxy* genes [101] responsible for the generation of basic tetracyclic scaffold have homologs in CHD biosynthetic gene cluster. Furthermore, 4-keto-anhydrotetracycline (4-keto-ATC) intermediate from OTC biosynthesis resembles putative CHD intermediate, 4-keto-9-desmethyl-CHD (**8**, Figure 54), differing only in the moiety at C2 position, due to incorporation of a different starter unit. Tetracyclines are typically primed by an amide starter unit [101], which also makes an acetate-primed chelocardin atypical. The polyketide skeleton of CHD is synthesized by a type II minimal polyketide synthase (minimal PKS) genes, consisting of ketosynthase α, ketosynthase β and acyl carrier protein (ACP), designated as ChdP, ChdK and ChdS, respectively, presumably catalyzing condensation of 10 malonate-derived building blocks into acetate-

primed decaketide (**1**, Figure 54). As already suggested earlier, and it is the case in CHD biosynthesis, the malonyl-CoA:ACP malonyltransferase, required for the transfer of the extender unit malonyl-CoA to ACP in actinorhodin biosynthesis, was proposed to be shared with fatty acid biosynthesis [300]. It is thought that the acyl chain is "protected" in a pocket formed by the minimal PKS with a suitable environment for following enzymatic reactions [284]. The initial folding of the growing polyketide chain is possibly directed by a ketoreductase ChdT, OxyJ homolog [101], reducing the keto group at C9 (**2**-**3**, Figure 54). The parallels drawn between CHD and OTC biosynthesis suggest CHD cyclization is directed by aromatases/cyclases ChdQI, ChdQII, ChdY and ChdX, first two being similar to OxyK and the last two homologs to OxyN and OxyI, respectively [101]. The protein homology comparisons of the di-domain aromatases ChdQI and ChdQII suggest that these two cyclase/aromatase homologs are responsible for first ring (D) formation (**4**, Figure 54), while a monodomain cyclase ChdY catalyzes the closure of the second ring (C). The formation of the third ring (B) is assumed to be spontaneous (**5**, Figure 54), similarly to OTC biosynthesis. The last ring (A) cyclization is proposed to be catalyzed by a putative cyclase ChdX (**6**, Figure 54) based on chromomycin and mithramycin biosynthesis [282]. However, it is currently believed that an important role in the cyclization of the last ring most probably has ChdL, an ATP-dependent acyl-CoA ligase, since its homologs SsfL2 and OxyH function as a fourth ring cyclases in the biosynthesis of tetracyclic SF2575 and oxytetracycline, respectively [288].

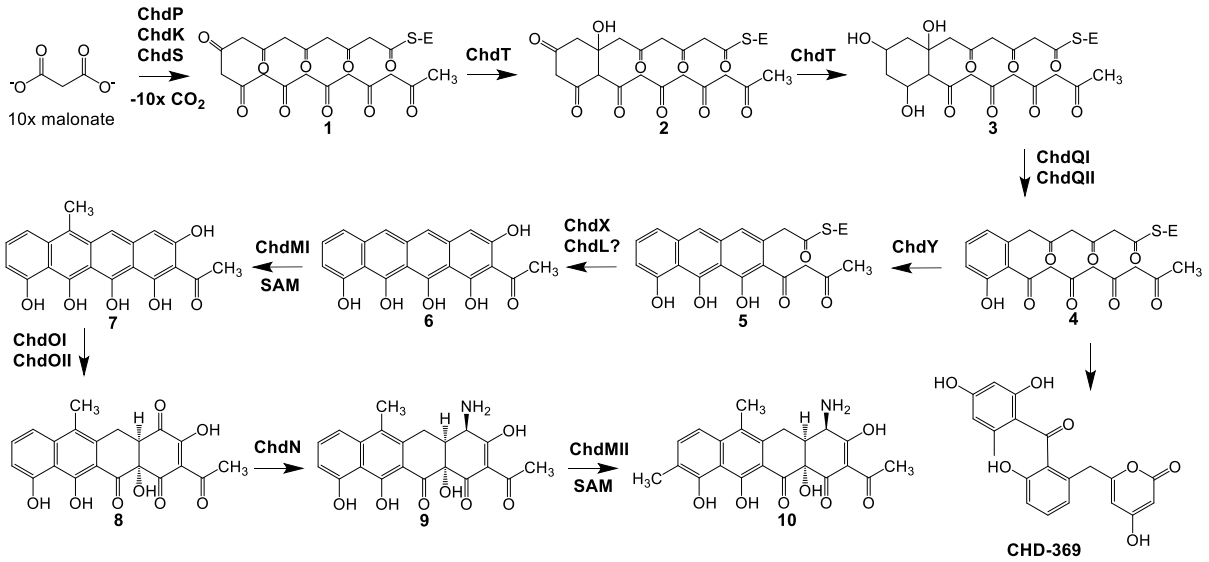



**Figure 54. Updated CHD biosynthetic pathway and gene annotation.** The previously published biosynthesis [32] was changed according to recents findings [104], including those described in this thesis.

The generated tetracyclic scaffold further undergoes various post-PKS tailoring reactions, with the last two steps being unique in CHD biosynthesis (Figure 54). ChdMI, OxyF homolog from OTC biosynthesis [101], methylates C6 position in CHD intermediate (**7**, Figure 54), which is followed by presumed double hydroxylation of ring A at C4/C12a (**8**, Figure 54) by oxygenase pair ChdOII and ChdOI, homologs of OxyL and OxyE, respectively [101]. Hydroxylation at C4 is followed by transamination by ChdN (**9**, Figure 54) [104], a PLP-dependent aminotransferase distantly related to OxyQ, responsible for the incorporation of an amino group at C4 in OTC biosynthesis [101]. One of the last tailoring steps in CHD biosynthesis is likely C9-methylation catalyzed by ChdMII (**10**, Figure 54) [104], homologous to C9-methyltransferases from chromomycin and mithramycin biosynthesis [282].

### 5.3.3 Study of the enzymes presumably involved in the cyclization of the CHD backbone

The cyclization reactions of CHD backbone was studied for various reasons. CHD has promising clinical applications, therefore the knowledge on its biosynthesis might play an important role in drug optimization. In contrast to OTC biosynthesis, two first ring aromatase/cyclase candidate genes are present in CHD biosynthetic gene cluster, while in typical tetracyclines only one aromatase/cyclase is required. We have recently identified a putative second ring cyclase ChdY, which provided evidence on its importance in formation of the CHD backbone and proved essential in our heterologous expression experiments, as presented in this work. Furthermore, we also investigated importance of the putative fourth ring cyclase ChdX, OxyI homolog in OTC, considering contradictory observations reported in the literature in different TC-like biosynthetic pathways.

#### 5.3.3.1 Cyclization of the first ring

In CHD biosynthesis two di-domain proteins ChdQI and ChdQII share identity of 31 % and 58 %, respectively, to aromatase OxyK from OTC biosynthesis [101] (Figure S4- 1). These enzymes are putatively involved in the aromatization of the first ring of an intermediate after C9-specific ketoreduction thus promoting C7-C12 cyclization [284]. The deletion of *chdQI* gene encoding a putative first ring aromatase ChdQI, through insertional mutagenesis by replacement with *ermE* resistance cassette, resulted in decrease in CHD production to about 1 % of production yield compared to the yield of CHD in WT strain (Figure 55). The *A. sulphurea* Δ*chdQI* strain was complemented with homologous *oxyK* gene, which failed to restore or increase the CHD production (Figure 55). The aromatase OxyK, previously designated as OtcD1 [301], was shown to play a crucial role in cyclization of the first ring of OTC but its

disruption caused minimal PKS to synthesize polyketides with shorter chain lengths [301]. As we expected that both aromatases ChdQI and ChdQII may play role in the cyclization of the first ring, the obtained result raised a question of the experimental strategy that we used in the *chdQI* gene replacement experiment. The protein structure of mutated ChdQI should change due to the insertion of another nucleotide sequence (*ermE* insertion; see Section 5.2.5.1) and could therefore spatially obstruct the catalytic environment inside the multienzyme PKS complex [284].

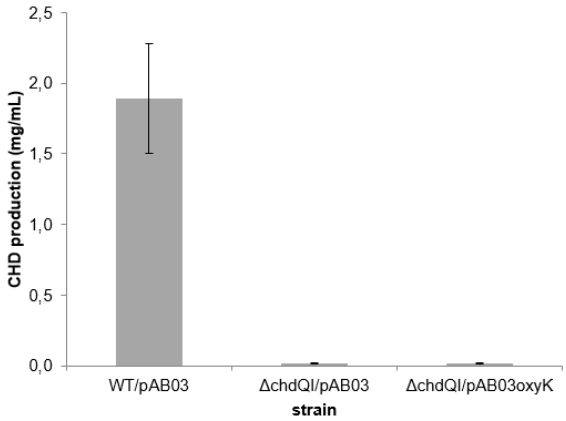



**Figure 55. CHD production yields of ChdQI knock-out mutant (Δ*chdQI*/pAB03) and after complementation experiment** with OxyK (Δ*chdQI*/pAB03oxyK) was compared to *A. sulphurea* WT with integrated empty plasmid pAB03 (WT/pAB03)

A precise site-directed mutagenesis was therefore carried out to generate strains with inactivated CHD aromatases and cyclases, which would cause a minimal disturbance to the entire CHD biosynthetic protein complex. This way, the structure of the mutated enzyme is mostly preserved and only its catalytic function should be affected. On the other hand, the selected catalytic sides have distinct importance for enzyme's activity and in many cases do not provide total inactivation as it was demonstrated in *in vitro* experiments for aromatases BexL and StfQ from BE-7585A and steffimycin, respectively [281]. The catalytic residues of the first-ring aromatases/cyclases from CHD biosynthetic pathway were modified based on *in vitro* assays reported in the literature, which caused decreased aromatase activity [278, 281]. The CHD aromatases were compared to homologs from biosynthetic pathways of other aromatic polyketides such as ZhuI from the biosynthetic pathway of R1128 family of polyketides [278] and BexL from the biosynthesis of an anticancer agent BE-7585A [281]. ZhuI is a mono-domain aromatase/cyclase in a non-reducing PKS system catalyzing aldol condensation between C7 and C12, cyclizing the first ring formation in R1128 biosynthesis [278]. The mutation R66A completely abolished ZhuI activity, while H109A (corresponding to H115 in

ChdQI) mutation caused 20-fold activity decrease in *in vitro* experiment [278]. On the other hand, BexL is a di-domain aromatase/cyclase in a reducing PKS system also catalyzing C7–C12 cyclization/aromatization, analogously to CHD aromatases ChdQI and ChdQII. The mutations in conserved residues in N-terminal domain caused a decrease in BexL activity in *in vitro* experiment, namely mutations R66A and H190A decreased the activity to 48.8 % and 2.7 %, respectively [281]. Interestingly, it was shown that selected mutations in C-terminal domain, for example R218A in BexL, had no effect on the enzyme's activity [281]. For site-directed mutagenesis of putative di-domain aromatases ChdQI and ChdQII, we chose to mutate residues present in N- and C-terminal domain in both enzymes. In ChdQI, residues R72 and H115 in N-terminal and R220 in C-terminal domain were selected (Figure 56, Figure 57, Figure 58). Similarly, R71 in N-terminal and R224 in C-terminal were chosen in ChdQII, each expected to inactivate one of the catalytic domains (Figure 56, Figure 58).

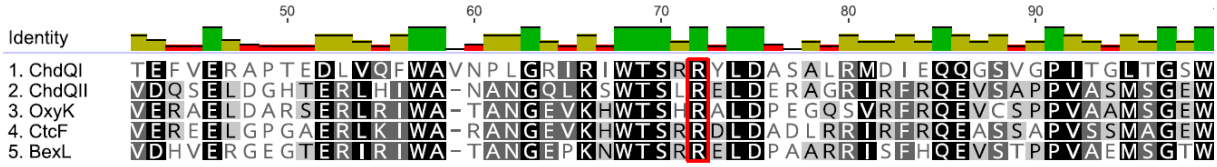



**Figure 56. Protein alignment of N-terminal parts of aromatases ChdQI and ChdQII with homologs from OTC, CTC and BE-7585A biosynthesis**, OxyK, CtcF, and BexL, respectively (ChdQI residue R72 and ChdQII R71 in red frame).

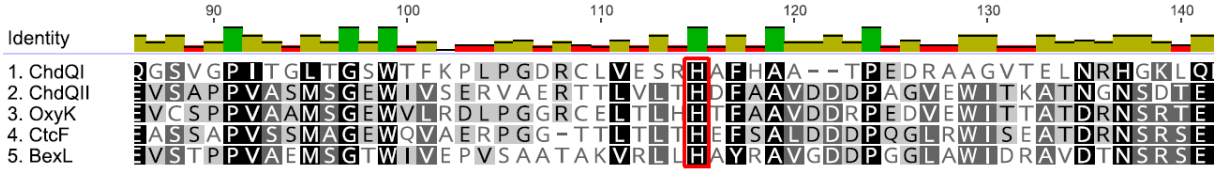



**Figure 57. Protein alignment of N-terminal parts of aromatases ChdQI and ChdQII with homologs from OTC, CTC and BE-7585A biosynthesis**, OxyK, CtcF, and BexL respectively (ChdQI residue H115 in red frame).


**Figure 58. Protein alignment of C-terminal parts of aromatases ChdQI and ChdQII with homologs from OTC, CTC and BE-7585A biosynthesis**, OxyK, CtcF, and BexL respectively (ChdQI residue R220 and ChdQII R224 in red frame).

The strain *A. sulphurea* Δ*chdQI* produced only 1 % of CHD compared to the WT strain (Table 11). Firstly, we complemented the knock-out strain with pAB03, an empty plasmid, which served as a negative control in case of any unspecific effect the plasmid might have on the CHD

production. The complementation with heterologous gene *oxyK* did not restore the production of CHD (Figure 55). Due to substantial decrease of CHD production (to 1 %), ChdQI appears essential for chelocardin production. The ChdQI does not function as a first-ring aromatase/cyclase, however, it probably assists in that process. A knock-out strain Δ*chdQI* was further complemented with a native *chdQI* gene and two other genes bearing mutations in N- and C-terminal domain. The former restored CHD production to 80 %, while the c*hdQI* with R72A and R220A site-specific mutation in ChdQI restored production to 36 % and 88 % compared to the wild-type strain, respectively (Table 11). Based on these results, the N-terminal domain seems more important for the activity of ChdQI than its C-terminal domain. The catalytic importance inequality of cyclization domains was beforehand reported for aromatase BexL [281].

**Table 11. A summary of *A. sulphurea* engineered strains regarding first ring cyclization/aromatization, and corresponding CHD yields**

| Protein | Mutation | Complemented with | Expression of | Mutated domain | CHD yield* | ZhuI[a] | BexL[b] |
|---|---|---|---|---|---|---|---|
| ChdQI | gene disrupted by *ermE* | / | ΔChdQI | / | < 1% | / | / |
| | | ChdQI | ChdQI | / | 80% | / | / |
| | | ChdQI-R72A | ChdQI-R72A | N-domain | 36% | 0% | 48% |
| | | ChdQI-R220A | ChdQI-R220A | C-domain | 88% | / | 100% |
| | | OxyK | OxyK | / | < 1% | / | / |
| | H115A | / | ChdQI-H115A | N-domain | 3% | 5% | 3% |
| | | ChdQI | ChdQI[c] | / | 9% | / | / |
| | | OxyK | OxyK[c] | / | 2% | / | / |
| ChdQII | R71A | / | ChdQII-R71A | N-domain | < 1% | 0% | 48% |
| | | ChdQII | ChdQII[c] | / | 6% | / | / |
| | | OxyK | OxyK[c] | / | 29% | / | / |
| | R224A | / | ChdQII-R224A | C-domain | < 1% | / | 100% |
| | | ChdQII | ChdQII[c] | / | 12% | / | / |
| | | OxyK | OxyK[c] | / | 17% | / | / |

*compared to WT
[a]Activity of ZhuI (mono-domain ARO/CYC in a non-reducing PKS system) with a mutation of the corresponding residue [278].
[b]Activity of BexL (di-domain ARO/CYC in a reducing PKS system) with a mutation of the corresponding residue [281].
[c]Engineered strains still express the mutated protein.

The subsequent experiments were carried out using the site-directed mutagenesis, where the generated engineered strains contained only mutations in selected catalytic sites, which retained the protein structure but inactivated them. Such approach would cause only minimal changes to the entire CHD biosynthetic protein complex opposed to the gene knock-out, where a missing protein would most probably cause its disturbance. The engineered strain would be further

complemented with gene encoding a protein homolog and would therefore express the mutated inactive catalytic protein and the native one, both having entirely native expression profile to participate in the biosynthesis. The yields of CHD therefore would not be expected to reach high values as observed in the example with the ChdQI knock-out mutant in Figure 55. *A. sulphurea* ChdQI-H115A mutant produced less than 3 % of CHD compared to WT strain. The complementation with a native *chdQI* gene increased production to around 9 %, whereas heterologous expression of a homologous *oxyK* gene further reduced the CHD production to 2 % (Table 11).

The characterization of ChdQII was carried out using the site-directed mutagenesis approach. The CHD production in both strains bearing R71A in N-terminal or R224A mutation in C-terminal of ChdQII was less than 1 % (Table 11). Complementation of N- and C-terminal-domain engineered strains with a native *chdQII* gene restored yield of CHD to 6 % and 12 %, respectively, whereas the complementation with *oxyK* gene further increased the production to 29 % and 17 %, respectively (Table 11). In contrast to domains in ChdQI, both catalytic domains (N and C) seem to be equally important for ChdQII activity. The OxyK functionally complements the activity of ChdQII, therefore the ChdQII is indeed the first-ring aromatase/cyclase in CHD biosynthesis. However, it does not catalyze the first-ring aromatization/cyclization independently since it requires a functional ChdQI since the mutation of the latter reduces the yield of CHD as shown in Figure 55 and Table 11. Similarly, the biosynthetic gene cluster of tetracycline SF2575 contains two OxyK homologs, SsfY1 and SsfY3, but a putative aromatase SsfY3 cannot substitute for an inactivated first ring aromatase SsfY1 [148]. The protein identities are shown in Figure S4- 1. Moreover, no shunt products were detected in crude extracts of the engineered strains bearing mutations in either ChdQI or ChdQII enzymes as observed in the biosyntheses of oxytetracycline [101] or SF2575 [148].

#### 5.3.3.2 Cyclization of the second ring

Due to the fact that the absence of a single enzyme in the multienzyme PKS complex might influence the catalytic function of the entire PKS complex, site-directed mutagenesis approach, where entire protein is maintained in its native form, was also applied for the characterization of a putative second ring cyclase ChdY (Figure 54). The catalytic residue of ChdY was chosen based on the reported disruption of daunomycin biosynthesis caused by a point mutation of G191S in DpsY [285], which is homologous to ChdY (Figure 59).

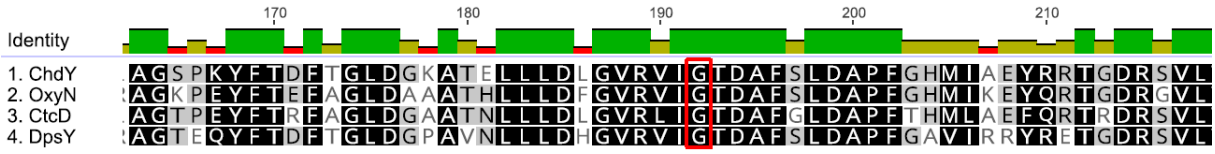



**Figure 59. Protein alignment of cyclase ChdY with homologs from OTC, CTC and daunomycin biosynthesis**, OxyN and CtcD, respectively (ChdY residue G176 in red frame).

Analogously as in daunomycin, the mutation of G176S in ChdY decreases CHD production to less than 0.3 % (Table 12), and thus practically inactivates catalytic activity of ChdY. The mutant strain mainly produces a shunt product designated as CHD-369 with molecular mass 369.09 $[M+H]^+$ (Figure 60). The compound CHD-369 ($C_{20}H_{16}O_7$, Figure 61) was isolated and its structure elucidated by HRMS and NMR analysis (Table S4- 3). It has an identical structure as shunt product (SEK43) observed in daunomycin and jadomycin biosynthesis lacking DpsY or JadI homologs [302], respectively, and is probably a result of spontaneous cyclization followed by the cyclization and aromatization of the first ring (Figure 61). On the other hand, in the anthraquinone actinorhodin biosynthetic pathway a different shunt product mutactin was isolated from *S. coelicolor* B40 mutant strain with "blocked" gene encoding a second ring cyclase ActVII [303, 304] since actinorhodin's nascent polyketide chain is shorter than the chelocardin's – other spontaneous cyclization pattern is expected to be observed (Figure 61).

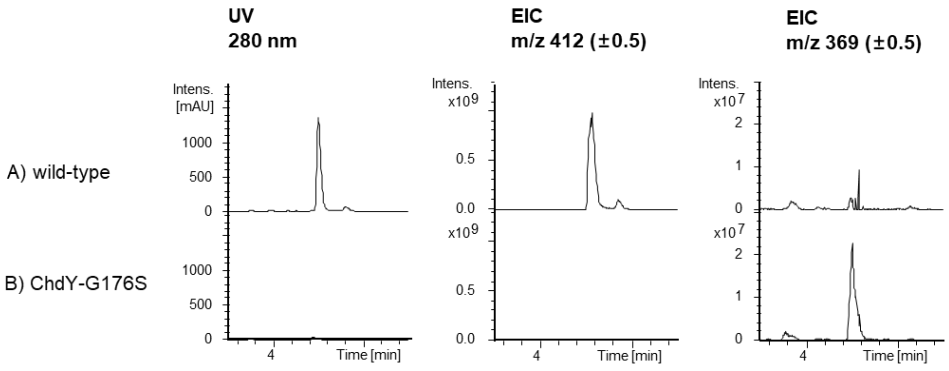



**Figure 60. LC-MS analysis of culture extracts of *A. sulphurea* wild-type (A) and *A. sulphurea*/ChdY-G176S strain (B).** UV chromatograms at detection wavelength of 280 nm and EICs for m/z 412 (±0.5) and 369 (±0.5) are shown (chromatograms adapted from DataAnalysis (available from Bruker Daltonics)).

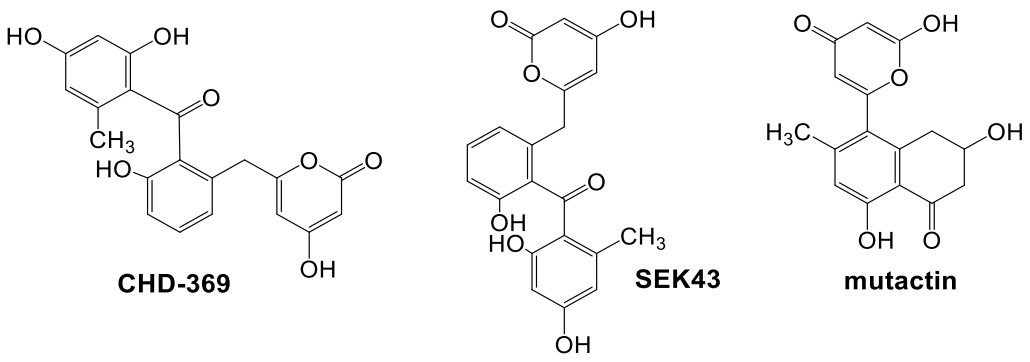



**Figure 61. The polyketide derivatives upon the inactivation or absence of the second ring cyclases:** ChdY, JadI and DspY, and ActVII in the biosynthetic pathways of chelocardin (CHD-369), jadomycin and daunorubicin (SEK43) and actinorhodin (mutactin).

The *A. sulphurea*/ChdY-G176S strain was complemented with homologous OxyN, which restored CHD production to 6 % (Table 12). The complementation with native ChdY and with the whole fusion protein ChdOII-ChdY restored CHD production to 3 % and 5 %, respectively. The highlighted box in Figure 62 shows increased production yields of complemented strains but the overall yields are still very low, when compared to the parent strain. However, the abolished CHD production in *A. sulphurea*/ChdY-G176S strain and structure of a shunt product CHD-369 together with the result of OxyN complementation confirm that ChdY is indeed a second-ring cyclase and is essential for the biosynthesis of CHD. Therefore, the heterologous expression of CHD biosynthetic gene cluster in the absence of the second ring cyclase ChdY would not be successful [305]. Complementation with a native ChdY partly restores CHD production but complementation with a native fusion protein ChdOII-ChdY appears more effective, which could be due to the presence of two functional ChdOII oxygenases.

**Table 12. A summary of *A. sulphurea* engineered strains regarding second ring cyclization, and corresponding CHD yields**

| Protein | Mutation | Complemented with | Expression of | Mutated domain | CHD yield* |
|---|---|---|---|---|---|
| ChdY | G176S | / | ChdY-G176S | / | < 0.3% |
| | | ChdY | ChdY [a] | / | 3% |
| | | ChdOII-Y | ChdOII-Y [a] | / | 5% |
| | | OxyN | OxyN [a] | / | 6% |

* compared to WT
[a] Engineered strains still express the mutated protein.

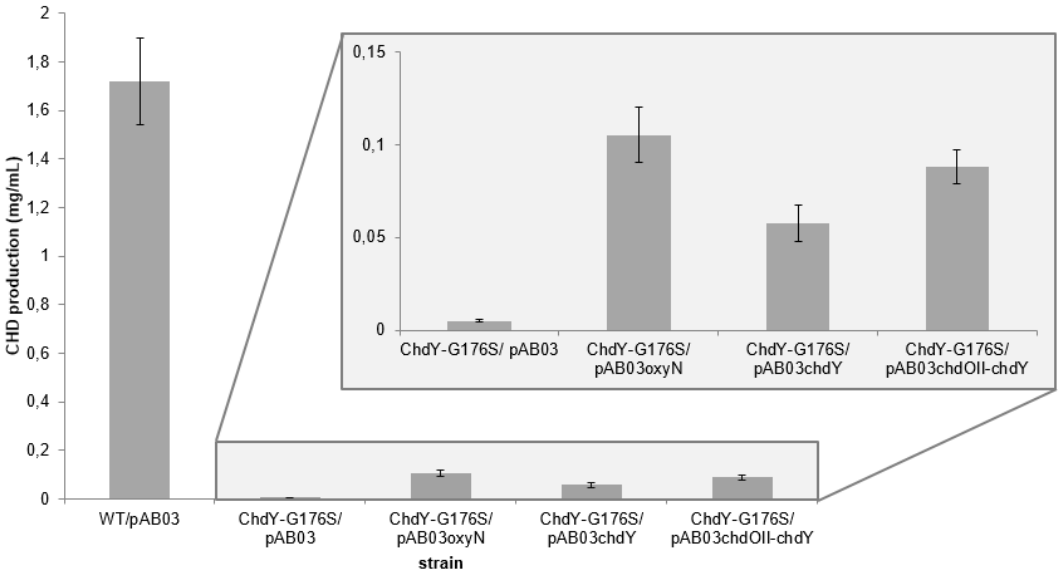



**Figure 62. CHD production in *A. sulphurea*/ChdY-G176S mutant before (ChdY-G176S/pAB03) and after complementation experiments** with genes for either OxyN (ChdY-G176S/pAB03oxyN), ChdY (ChdY-G176S/pAB03chdY) or ChdOII-ChdY (ChdY-G176S/pAB03chdOII-chdY), compared to *A. sulphurea* WT with integrated empty plasmid pAB03 (WT/pAB03)

### 5.3.3.3 Cyclization of the fourth ring

The cyclization of the fourth ring is in aclacinomycin and nogalamycin biosynthesis catalyzed by small polyketide cyclases AknH [306] and SnoaL [307], respectively. The conversion of the Nogalonic Acid Methyl Ester (NAME) intermediate, present in biosynthesis of both anthracyclines, into nogalaviketone or auraviketone occurs via an intramolecular aldol condensation (Figure 63), however, without the formation of a Schiff base intermediate or a metal ion, as usual for the class I and class II aldolases [306]. The catalytic residue Asp121 in SnoaL acts as the catalytic base in the aldol condensation reaction that removes the proton, leading to the enol(ate) intermediate [307]. The inactivation of Asp121 residue in AknH and Asp111 residue in SnoaL in *in vitro* experiment led to decreased [306] or abolished catalytic activity [307], respectively.

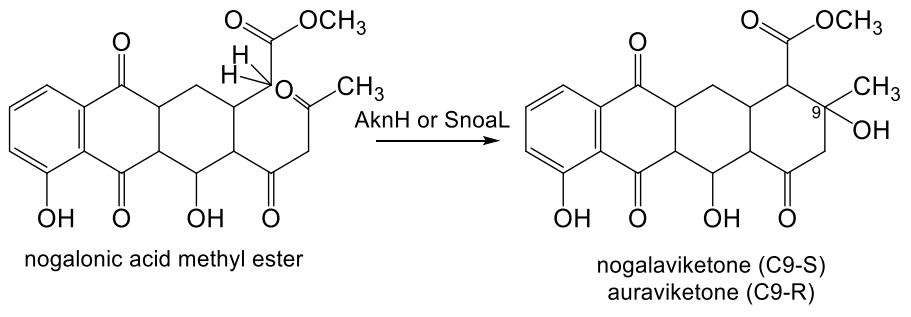



**Figure 63. AknH and SnoaL catalyze the closure of the fourth ring of nogalonic acid methyl ester (NAME).**

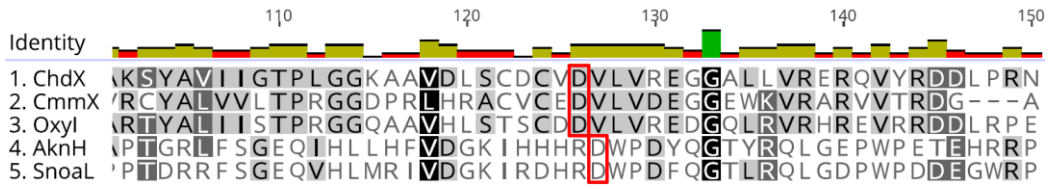



**Figure 64. Protein alignment of cyclase ChdX with homologs.** Close homologs from chromomycin and oxytetracycline biosynthesis, CmmX and OxyI, respectively, and more distant homologs from aclacinomycin and nogalamycin biosynthesis, AknH and SnoaL, respectively (ChdX residue D126 in red frame), were chosen.

Moreover, the cyclization of the fourth ring was proposed to be cyclized by ChdX homologs, MtmX (46 % identity) and CmmX (42 % identity) from mithramycin [308] and chromomycin [282] biosyntheses, respectively, all belonging into nuclear transport factor 2 superfamily and containing a SnoaL-like domain according to BLAST tool. There is no additional protein crystal structures of fourth ring cyclases, or information about their catalytic activities, besides the anthracycline polyketide cyclases AknH and SnoaL that share only 18 % and 20 % identity with ChdX, respectively. The catalytic residue in a putative fourth ring cyclase ChdX from CHD biosynthesis was therefore chosen based on AknH and SnoaL (Figure 64). The selected mutation of residue D126A in ChdX reduced CHD production to 8 % compared to the wild-type strain. Moreover, the complementation of the mutant strain with a native *chdX* gene restored production to 32 % and complementation with *oxyI* gene to 39 % (Table 13).

**Table 13. A summary of *A. sulphurea* engineered strains regarding fourth ring cyclization, and corresponding CHD yields**

| Protein | Mutation | Complemented with | Expression of | Mutated domain | CHD yield* |
|---|---|---|---|---|---|
| ChdX | D126A | / | ChdX-D126A | / | 8% |
| | | ChdX | ChdX [a] | / | 32% |
| | | OxyI | OxyI [a] | / | 39% |
| | | OxyDP | OxyDP [a] | / | < 3% |

* compared to WT
[a] Engineered strains still express the mutated protein.

The production in *A. sulphurea*/ChdX-D126A is not decreased as much as in other mutant strains. The selected catalytic site mutation might not entirely inactivate the enzyme, some activity might be still retained. Moreover, no shunt products were observed that would indicate its essentiality for CHD biosynthesis. Importantly, OxyI complements the missing activity of ChdX, so we can conclude that these homologs have the same function. As CHD is primed by an acetate moiety, resulting in an acetyl moiety at C2, it differs from typically malonamate-primed tetracyclines with carboxyamido moiety at C2 position instead [101]. In 2009, Zhang

and collaborators [287] proposed that OxyI from OTC biosynthesis, sharing 64 % identity with ChdX, is redundant for OTC biosynthesis since the carboxyamido starter unit would contribute to the spontaneity of cyclization of the fourth ring [101, 287]. This proposal was in 2015 disproven by Tang and collaborators [288] since they found out that in *S. coelicolor* CH999 heterologously expressed pWJ83 plasmid harboured an additional *oxyH* gene, encoding an acyl-CoA ligase, next to reported *oxyABCDJKN* genes [288], necessary for OTC biosynthesis. The *oxyH* gene was later deleted from *S. lividans* K4-114/pWJ83 strain and so engineered strain was not able to produce oxytetracycline [288]. We do not have experimental data on the function of an acyl-CoA ligase ChdL from CHD biosynthesis, OxyH homolog, however, we assume based on the findings from the literature stated above that it might be involved in the cyclization of the last ring of the atypical TC [288]. Nevertheless, we complemented *A. sulphurea*/ChdX-D126A mutant with pAB03oxyDP [297], creating a carboxyamido-primed derivative to evaluate potential spontaneity of cyclization of the fourth ring. The heterologous expression of OxyDP decreased *A. sulphurea*/ChdX-D126A's production of chelocardins to less than 3 % but the decrease of CHDs' production upon heterologous expression of OxyD and OxyP in *A. sulphurea* WT was previously reported [298].

Our experimental set up was unique in a way since all the experiments were carried out on the native producer, which produced relatively high yield of CHD. Due to highly reactive keto intermediates [99], spontaneous cyclization might be possible to some degree. However, in high producer strains cyclases function as chaperones to help in directing nascent polyketide intermediates into particular reaction channels and are in those systems pivotal [99]. On the other hand, in *in vitro* and heterologous expression experiments the function of potential auxiliary proteins cannot be evaluated to the same extent. The mutation of D126A in ChdX caused the decrease of production to 8 %, which shows that ChdX has an important role in CHD production, although the CHD production was not completely abolished. ChdX, OxyI homolog, might contribute to the last ring cyclization but we do not have any evidence. However, according to the literature, an acyl-CoA ligase ChdL, OxyH homolog, could have an important role in the cyclization of the last ring but this is not yet confirmed. Further characterization of ChdL would be necessary before drawing any conclusions about the fourth-ring cyclization mechanism in CHD biosynthesis.

### 5.3.4 Evaluation of the selected approaches for characterization of enzymes presumably involved in cyclization of CHD backbone

The yields of CHD after *in trans* complementation experiments were substantially lower in strains with mutated catalytic sites in the target gene, when compared the *in trans* complementation experiments in *A. sulphurea* Δ*chdQI* strain, which had *chdQI* deleted. Complementation of the *A. sulphurea* Δ*chdQI* strain was very efficient, and resulted in CHD yield almost comparable to the parent strain (Table 11). However, this was not the case, when we complemented strains with mutated catalytic sites in ChdY and ChdX. The reason is most probably the presence of two copies of the same gene under study: one functional, which was introduced in the chromosome "in trans" and the second copy containing a single amino acid mutation with significantly decreased activity. We can expect that both copies, the mutated and the native one, can participate in the formation of the CHD's multienzyme PKS complex. It is likely that in our experiments this approach resulted in drastic reduction of CHD production after complementation was completed (Table 11). Additionally, the integrative pAB03 plasmid integrates into the attB site in the chromosome of *A. sulphurea* [295], located far away from the CHD gene cluster. The *in trans* integrated genes are therefore translated elsewhere in the genome, which might influence the efficiency and coordination of CHD PKS complex formation and its activity. Moreover, the obtained results on CHD production in engineered strains cannot be directly compared to the percentage of enzymes' catalytic activity. The *in vitro* approach as reported for BexL [281] or ZhuI [278], and *in vivo* approach presented in this study are methodically substantially different. The former is carried out in test tubes and evaluates only one reaction step, compared to the latter that measures the performance of certain enzymes in a complex biological system. Nevertheless, the mutation of all catalytic amino acid residues selected based on the *in vitro* experiments for the characterization of ChdQI, ChdQII, ChdY and ChdX resulted in substantial decrease in CHD production (Table 11-Table 13).

### 5.3.5 6-desmethyl-CHDs

In heterologous expression study carried out by Zhang *et al*. [287], OxyF was shown to catalyze C6-methylation of the oxytetracycline backbone. ChdMI from CHD biosynthetic gene cluster shares 65 % identity with OxyF and was therefore assumed to have the same role in CHD biosynthetic pathway. To characterize ChdMI, a knockout mutant *A. sulphurea* Δ*chdMI* was engineered, which produced CHD-398 derivative with molecular mass 398.14 $[M+H]^+$ (Figure 65), corresponding to 6-desmethyl CHD. The CHD-398 was isolated and its structure elucidated revealing the expected 6-desmethyl CHD (Table S4- 4, Figure S4- 7). Furthermore, we prepared

the 6-desmethyl CDCHD derivative with a simple integration of pAB03oxyDP plasmid [297] into *A. sulphurea* Δ*chdMI* genome, followed by selection of best producer strain. The *A. sulphurea* Δ*chdMI*/pAB03oxyDP as predicted produced 6-desmethyl CDCHD, which corresponds to the molecular mass 399.10 $[M+H]^+$ (Figure 65). In Figure 65, *A. sulphurea* Δ*chdMI* appears to produce a derivative with molecular mass matching 6-desmethyl CDCHD but it is only an isotope of 398.14 $[M+H]^+$. The biological activity of 6-desmethyl CHD (CHD-398) was evaluated in biological assays against *S. aureus* Newman, *E. faecium* DSM-20477, *E. coli* DSM-1116 with and without additives, either increasing the permeability of outer membrane (PMBN: polymyxin B nonapeptide) and/or an efflux pump inhibitor (PAβN: phenylalanine arginyl β-naphthylamide dihydrochloride), and *P. aeruginosa* PA14 Δ*mexAB* with and without PMBN. Surprisingly, no biological activity against all the tested strains was observed.

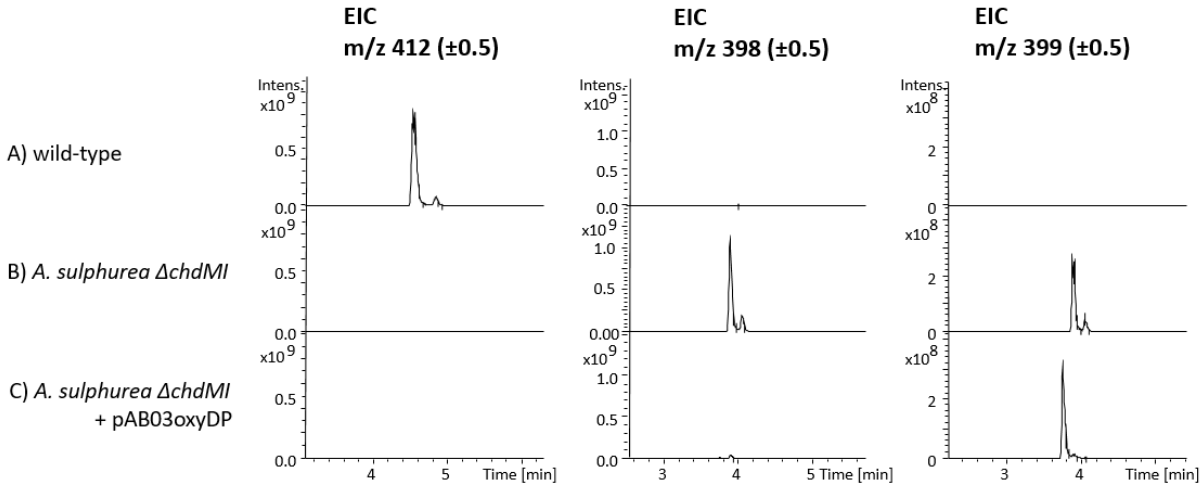



**Figure 65. LC-MS analysis of culture extracts of *A. sulphurea* wild-type (A), *A. sulphurea* Δ*chdMI* (B) and *A. sulphurea* Δ*chdMI* complemented with pAB03oxyDP plasmid.** EICs for m/z 412 (±0.5), 398 (±0.5) and 399 (±0.5) are shown (chromatograms adapted from DataAnalysis (available from Bruker Daltonics)).

## 5.4 Conclusion

We have demonstrated that ChdQII functions as a first ring aromatase/cyclase since it can be replaced by its homolog OxyK from OTC biosynthesis that was shown to function as a first ring cyclase in heterologous expression experiment [287]. Interestingly, ChdQI sharing homology to other aromatases/cyclases was proven essential for the biosynthesis of CHD, however, we could not confirm clearly its function in the first ring cyclization. Furthermore, its N-terminal domain demonstrated to be more important than the C-terminal domain, unlike in ChdQII, where both domains importantly contribute to its activity. Moreover, we could not

detect any shunt products with *chdQI* or *chdQII* inactivated strains. The newly identified ChdY is proposed to cyclize the second ring due to the isolated CHD-369 derivative, identical to SEK43 observed in daunomycin and jadomycin biosynthesis lacking respective protein homologs [302], and its identity to well-characterized OxyN [287]. We have shown that ChdX cyclase is a functional homolog of OxyI, however, it was not confirmed as a fourth ring cyclase since no shunt products were observed in *A. sulphurea*/ChdX-D126A strain. We have not characterized an acyl-CoA ligase ChdL, OxyH homolog, that is currently believed to be involved in the cyclization of the last ring. All enzymes that were inactivated in this study are functionally essential in the CHD biosynthetic pathway. Additionally, we have shown that ChdMI, OxyF homolog, is indeed C6-methyltransferase, methylating CHD at C6 position.

## 5.5 Supplementary information

### Tables of Tables



### Table of Figures



**Table S4- 1. Sequences of oligonucleotide primers for PCR experiments used in this study[a]**

| Primers | 5'-Sequence-3' |
|---|---|
| FSB01C | AGTCGAATTCGCACCATATGAGACCAAGCGCGTCCGGGTG |
| FSB02C | CGACTCTAGAGGATCACTAGTTACCAGCCCGACCCGAGCACGC |
| FSB03 | CCGGGAATTCCGCGACATTCGACCCCGGC |
| FSB04 | GACGCATATGGCCGGTTCGTCCGCATCTG |
| FSB05 | GCTCGACTAGTGGGTGTGACATGGGTTCTC |
| FSB06 | GTGACGCATGCGAAGCTGCACGAAATCATGC |
| chdQIalaLF | TATATAGCATGCAAGAACTACGAGCGCAACGG |

| | |
|---|---|
| chdQIalaLR | CAGGTACG**CG**CGGGAGG |
| chdQIalaRF | CTCCCG**CG**CGTACCTGG |
| chdQIalaRR | TATATAACTAGTCTTCACGGACTCGTCTCGC |
| chdQIH115ALF | ATATATGCATGCTGTTGCGCAGGCTGTTCGA |
| chdQIH115ALR | GGAAGGCG**GC**CCGGGACTC |
| chdQIH115ARF | AGTCCCGG**GC**CGCCTTCCA |
| chdQIH115ARR | ATATATACTAGTACCACGCCGGGAGCG |
| chdQICalaLF | TATATAGCATGCGCCCTCTTTCCCACACC |
| chdQICalaLR | AAGCAGATG**GC**GACGAACCG |
| chdQICalaRF | GGTTCGTC**GC**CATCTGCTTG |
| chdQICalaRR | TATATAACTAGTCTGTCGACTGCGCTGGTG |
| chdQIIalaLF | TATATAGCATGCACGAGCTGTGGCAGGACG |
| chdQIIalaLR | GAGTTCC**GC**GAGCGAGGTCC |
| chdQIIalaRF | CTCGCTC**GC**GGAACTCGAC |
| chdQIIalaRR | TATATAACTAGTCTTGGCATGACGCAGCG |
| chdQIICalaLF | TATATAGCATGCCTCACGGAAAGGCACAGCC |
| chdQIICalaLR | GCACACC**GC**CACCGACAC |
| chdQIICalaRF | GTCGGTG**GC**GGTGTGCTTC |
| chdQIICalaRR | TATATAACTAGTTCGAACCAGGTCAGCTCGAC |
| chdYserLF | TATATAGCATGCGCATCATCGACC |
| chdYserLR | GCGTCGGT**G**C**T**GATGACCC |
| chdYserRF | GGTCATC**A**G**C**ACCGACGCG |
| chdYserRR | TATATAACTAGTCGTCCAGCTGCAGCAGATAAC |
| chdQIF | TATATACATATGTCACACCCCGAGGC |
| chdQIR | TATATATCTAGACGTTCAGCGGGTTTCTGC |
| chdQIIF | ATATACATATGCCCGCCGCAGC |
| chdQIIR | TATATATCTAGATCAGGCATCGGTCGCCTC |
| chdYF | ATATACATATGCGCATCATCGACCTGTC |
| chdYR | TATATATCTAGACTAGTCCAGCAGGGCAACGG |
| chdOIIF | ATATACATATGCCTGAGGACTCCGGC |
| oxyKFw | TATATACATATGCCCGCACCCACATCC |
| oxyKRv | ATATATTCTAGACTCAGGCGGCATGCG |
| oxyNFw | TATATACATATGCGCATCATCGATCTGTCGA |
| oxyNRv | ATATATCTAGACTACTCCTCCACCACCGCC |
| chdXalaLF | TATATAGCATGCTCCTTCACCGTGGAAACGAC |
| chdXalaLR | CAGCACG**G**CCACGCAGTCAC |
| chdXalaRF | GACTGCGTGG**C**CGTGCTGG |
| chdXalaRR | TATATAACTAGTTTCTCGTCGCAAGTGGTGG |
| chdXF | ATATACATATGTCACAGGCGGTACAGGCA |
| chdXR | TATATATCTAGATCAGTTCCTCGGCAGGTCG |
| oxyI Fw | TATATACATATGACCGCCCCGGAACAGC |
| oxyI Rv | ATATATTCTAGACTACTCGGGCCGCAGGTC |
| chdMI-KO-LF | ATATAGCATGCGATCAACGGCGACTACCGG |
| chdMI-KO-LR | ATATATATCTAGAGCGCGGTGAACTCCTACCTG |
| chdMI-KO-RF | ATATACATATGAGCTCGCGGACTCGCA |
| chdMI-KO-RR | ATATATGAATTCGTCCAAGGGTACGGACTTCAGG |

[a] Restriction sites are underlined, introduced point-mutations are in bold

**Table S4- 2. Bacterial strains and constructs used and prepared in this study**

| Strain or plasmid | Relevant characteristics | Reference or source |
|---|---|---|
| *Escherichia coli* | | |
| DH10β | F$^-$ *endA*1 *recA*1 *galE*15 *galK*16 *upG rpsL* Δ*lacX*74 Φ80*lacZ*ΔM15 *araD*139 Δ(*ara-leu*)7697 *mcrA* Δ(*mrr-hsdRMS-mcrBC*) λ$^-$ | Invitrogen |
| ET12567 | F$^-$ *dam*13::Tn*9*, *dcm*6, *hsdM*, *hsdR*, *recF*143::Tn*1I*, *galK*2, *galT*22, *ara*14, *lacY*1, *xyl*5, *leuB*6, *thi*1, *tonA*31, *rpsL*136, *hisG*4, *tsx*78, *mtl*1 *glnV*44 | [262] |
| SCS110 | *rpsL* (Str$^r$) *thr leu endA thi-1 lacY galK galT ara tonA tsx dam dcm supE44* Δ*(lac-proAB)* [F´ *traD36 proAB lacI$^q$ZΔM15*] | Stratagene |
| *Amycolatopsis sulphurea* | | |
| NRRL 2822 | Wild-type producer of chelocardin | ARS Culture Collection |
| *Streptomyces rimosus* | | |
| M4018 | Producer of oxytetracycline | [290] |

| Plasmids | | |
|---|---|---|
| pNV18 | Kan$^r$, *lacZα* | [309] |
| pNV18Erm | Kan$^r$, Erm$^r$, *lacZα* | This study |
| pAB03 | pSET152-derived, containing ΦBT, Apr$^r$ | [32] |
| pAB03oxyDP | *oxyD* and *oxyP* cloned into pAB03 | [297] |
| pNV18ErmKOchdMI | 0.9 kb internal to *chdMI* replaced with *ermE* and surrounded by 1 kb and 1.1 kb homology regions in pNV18 | This study |
| pNV18chdQIErm | 0.6 kb internal to *chdQI* replaced with *ermE* and surrounded by two 1.1 kb homology regions in pNV18 | This study |
| pNV18ErmchdQIala | Fragment containing aromatase gene *chdQI* with mutation Arg72Ala cloned into pNV18Erm | This study |
| pNV18ErmchdQICala | Fragment containing aromatase gene *chdQI* with mutation Arg220Ala cloned into pNV18Erm | This study |
| pNV18ErmchdQIH115A | Fragment containing aromatase gene *chdQI* with mutation His115Ala cloned into pNV18Erm | This study |
| pNV18ErmchdQIIala | Fragment containing aromatase gene *chdQII* with mutation Arg71Ala cloned into pNV18Erm | This study |
| pNV18ErmchdQIICala | Fragment containing aromatase gene *chdQII* with mutation Arg224Ala cloned into pNV18Erm | This study |
| pNV18ErmchdYser | Fragment containing cyclase gene *chdY* with mutation Gly176Ser cloned into pNV18Erm | This study |
| pNV18ErmchdXala | Fragment containing cyclase gene *chdX* with mutation Asp126Ala cloned into pNV18Erm | This study |
| pAB03chdQI | Aromatase gene *chdQI* cloned into pAB03 | This study |
| pAB03chdQIala | Aromatase gene *chdQI* with mutation Arg72Ala cloned into pAB03 | This study |
| pAB03chdQICala | Aromatase gene *chdQI* with mutation Arg220Ala cloned into pAB03 | This study |
| pAB03chdQII | Aromatase gene *chdQII* cloned into pAB03 | This study |
| pAB03chdY | Cyclase gene *chdY* cloned into pAB03 | This study |
| pAB03chdOII-chdY | Gene *chdOII-chdY* with oxygenase and cyclase domain cloned into pAB03 | This study |
| pAB03oxyK | Aromatase gene *oxyK* cloned into pAB03 | This study |

| | | |
|---|---|---|
| pAB03oxyN | Cyclase gene *oxyN* cloned into pAB03 | This study |
| pAB03chdX | Cyclase gene *chdX* cloned into pAB03 | This study |
| pAB03oxyI | Cyclase gene *oxyI* cloned into pAB03 | This study |

| | ChdQI | ChdQII | OxyK | BexL | CtcF | SsfY1 | SsfY3 |
|---|---|---|---|---|---|---|---|
| ChdQI | | 30% | 31% | 30% | 33% | 30% | 23% |
| ChdQII | 30% | | 58% | 52% | 57% | 60% | 23% |
| OxyK | 31% | 58% | | 50% | 56% | 56% | 23% |
| BexL | 30% | 52% | 50% | | 49% | 50% | 23% |
| CtcF | 33% | 57% | 56% | 49% | | 59% | 24% |
| SsfY1 | 30% | 60% | 56% | 50% | 59% | | 22% |
| SsfY3 | 23% | 23% | 23% | 23% | 24% | 22% | |

**Figure S4- 1. The identity between ChdQI, ChdQII and homologs** from OTC, BE-7585A, CTC and SF2575 biosyntheses, OxyK, BexL, CtcF, and SsfY1 and SsfY3, respectively.

**Table S4- 3. NMR data for compound CHD-369 in MeCN-$d_3$ + 0.05% TFA-$d_1$**

| Pos. | $\delta_H$, mult (*J* in Hz) | $\delta_C$ | HMBC correlations |
|---|---|---|---|
| 1 | - | 165.7 | 2 |
| 2 | 5.27, d (*2.1*) | 90.4 | 1, 4 |
| 3 | - | 171.0 | 2, 4 |
| 4 | 5.69, d (*2.1*) | 101.9 | 2, 3 |
| 5 | - | 164.7 | 4, 8 |
| 6 | 3.62, bs | 37.6 | 4 |
| 7 | - | 133.4 | 4, 8, 9, 10 |
| 8 | 6.90, d (*8.2*) | 123.5 | 9, 10 |
| 9 | 7.28, t (*8.0*) | 131.4 | 8, 10 |
| 10 | 6.85, d (*7.6*) | 116.0 | 8, 9 |
| 11 | - | 154.0 | 8, 9, 10 |
| 12 | - | 122.8 | 8, 9, 10 |
| 13 | - | 201.8 | 8, 9, 10, 16, 18, 20 |
| 14 | - | 116.2 | 16, 18, 20 |
| 15 | - | 168.3 | 16, 20 |
| 16 | 6.23, d (*2.6*) | 101.9 | 18, 20 |
| 17 | | 164.4 | 16, 18, 20 |
| 18 | 6.12, d (*2.6*) | 113.0 | 16, 10 |
| 19 | | 145.4 | 20 |
| 20 | 1.78, s | 22.3 | |

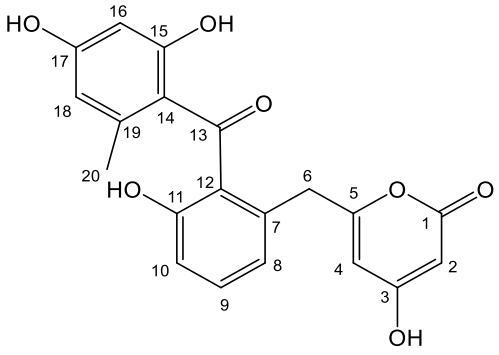



**Figure S4- 2. Structure of CHD-369**

$^{1}$H NMR (501 MHz, ACETONITRILE-$d_3$) δ ppm 1.77 (s, 3 H) 3.64 (br d, $J$=7.91 Hz, 2 H) 5.27 (d, $J$=2.14 Hz, 1 H) 5.69 (d, $J$=2.14 Hz, 1 H) 6.12 (dd, $J$=2.56, 0.75 Hz, 1 H) 6.23 (d, $J$=2.56 Hz, 1 H) 6.85 (br d, $J$=0.75 Hz, 1 H) 6.85 - 6.95 (m, 3 H) 7.28 (br d, $J$=7.91 Hz, 5 H)

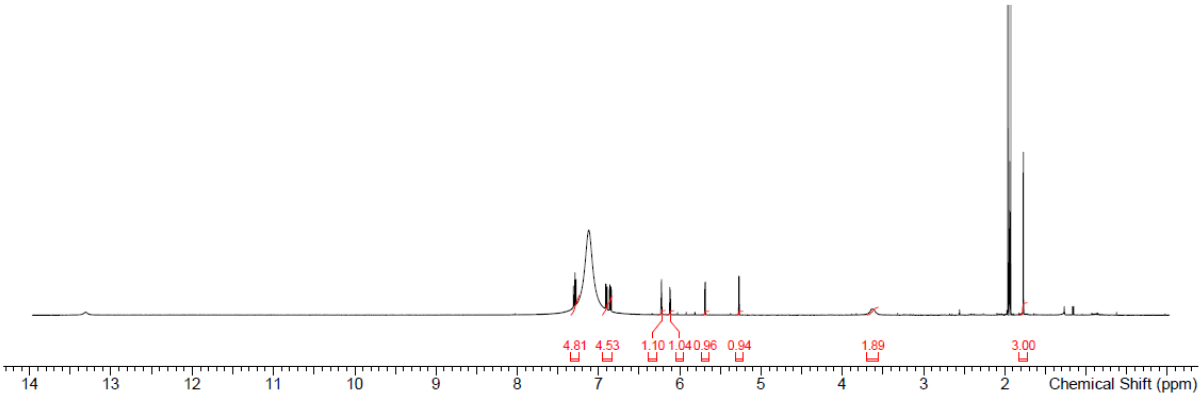



**Figure S4- 3. $^{1}$H spectrum of CHD-369 acquired in MeCN-$d_3$ + 0.05% TFA-$d_1$ at 500 MHz**

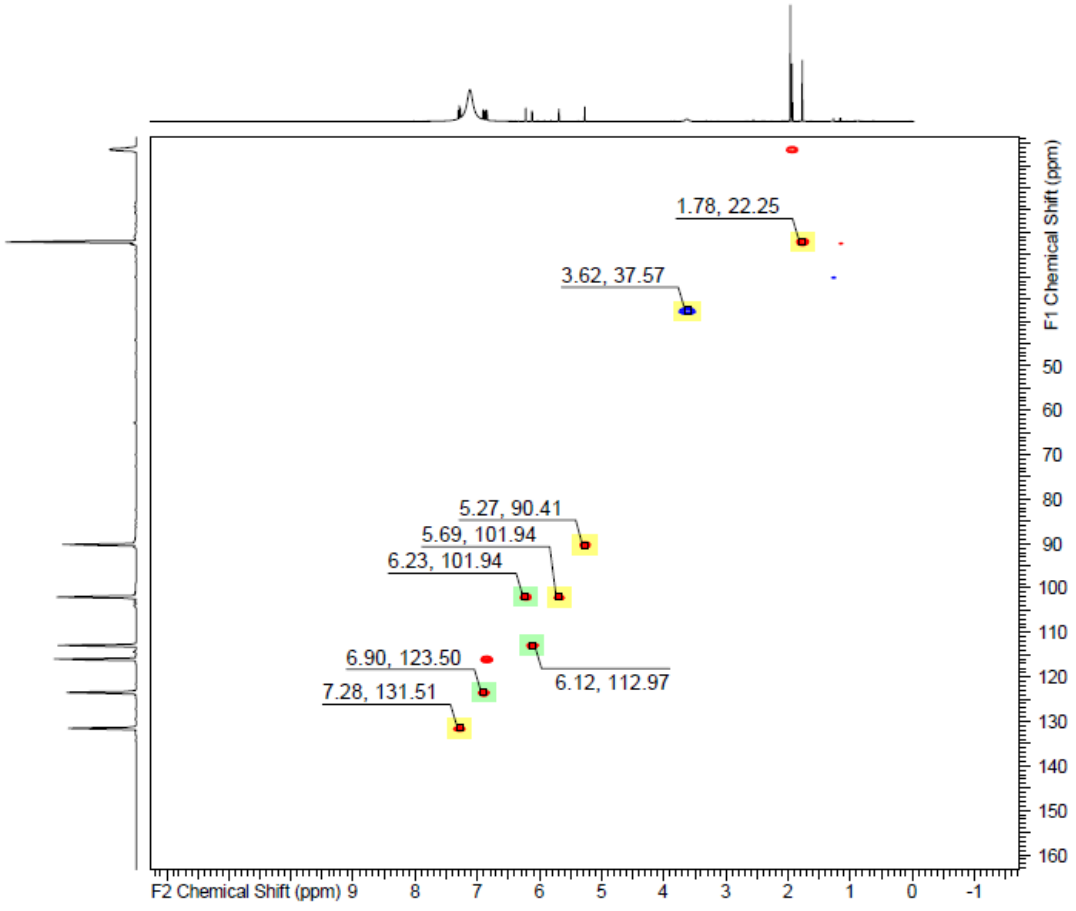



**Figure S4- 4. HSQC spectrum of CHD-369 acquired in MeCN-$d_3$ + 0.05% TFA-$d_1$ at 500/125 MHz**

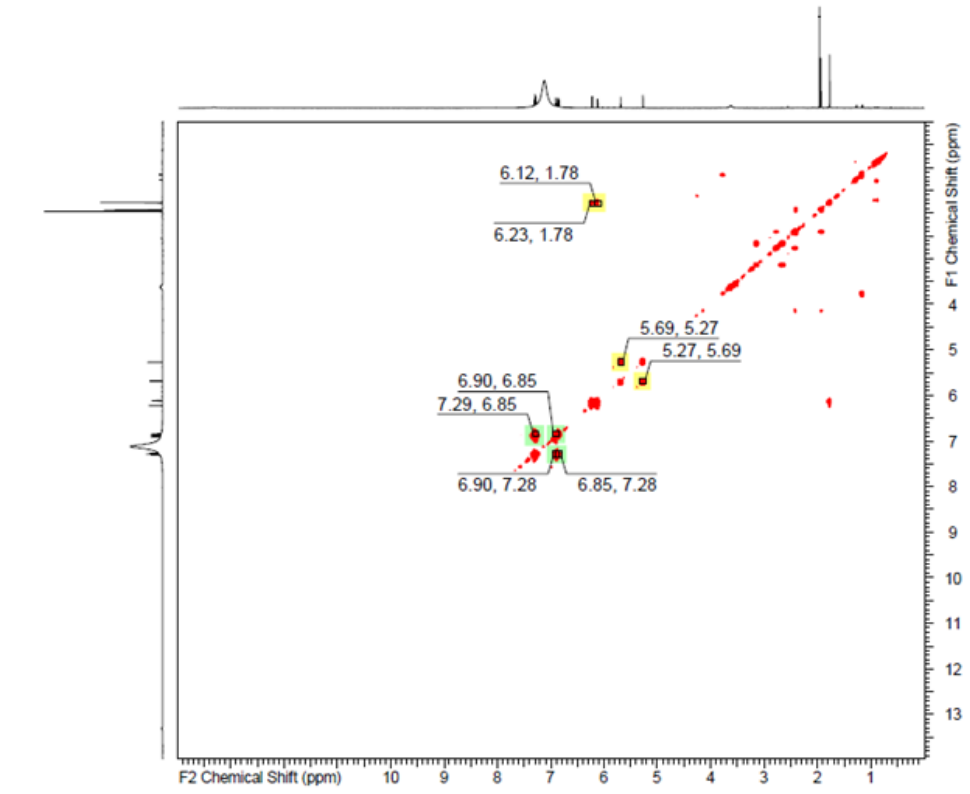



**Figure S4- 5. COSY spectrum of CHD-369 acquired in MeCN-$d_3$ + 0.05% TFA-$d_1$ at 500 MHz**

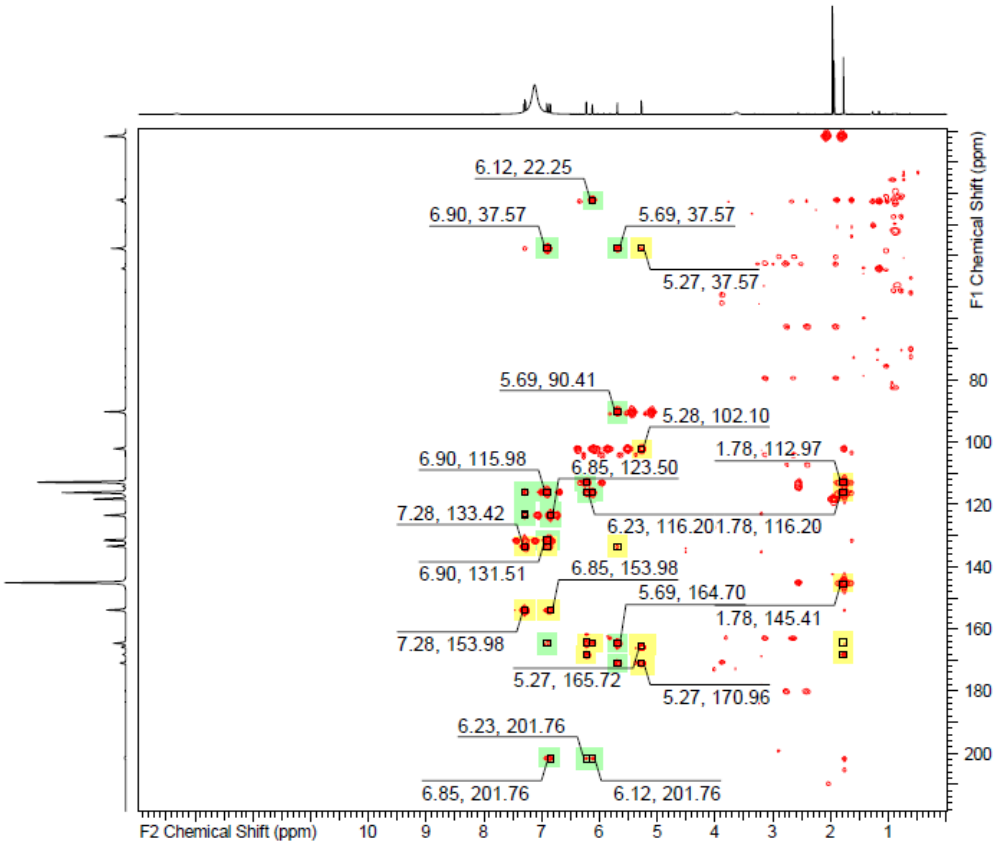



**Figure S4- 6. HMBC spectrum of CHD-369 acquired in MeCN-$d_3$ + 0.05% TFA-$d_1$ at 500/125 MHz**

**Table S4- 4. NMR data for compound CHD-398 in methanol-*d*$_4$**

| Pos. | $\delta_H$, mult (*J* in Hz) | $\delta_C$ | COSY correlations | HMBC correlations |
|---|---|---|---|---|
| 1 | 2.61, s | 26.5 | - | 2 |
| 2 | - | 201.5 | - | - |
| 3 | n.d. | n.d. | n.d. | n.d. |
| 4 | - | 193.1 | - | - |
| 4a | - | 79.7 | - | - |
| 5 | - | 192.1 | - | - |
| 5a | - | 109.6 | - | - |
| 6 | n.d. | n.d. | n.d. | n.d. |
| 6a | - | 113.1 | - | - |
| 7 | - | 155.5 | - | - |
| 8 | - | 121.4 | - | - |
| 9 | 7.47, m | 136.8 | 10, 15 | 6a, 7, 10a, 15 |
| 10 | 7.15, d (8.09) | 119.0 | 9 | 5a, 6a, 8, 10a, 11 |
| 10a | - | 139.5 | - | - |
| 11 | 7.09, s | 119.6 | 12 | 5a, 6a, 10, 10a, 12 |
| 11a | - | 133.0 | - | - |
| 12 | 2.98, 3.25, m | 27.2 | 11, 13 | 4, 4a, 5, 5a, 11, 11a, 12a, 13, 14 |
| 12a | 3.24, m | 42.0 | - | 4, 4a, 5a, 11, 11a, 12, 14 |
| 13 | 4.95, d (4.27) | 54.7 | 12 | - |
| 14 | - | 190.8 | - | - |
| 15 | 2.32, s | 15.6 | 9 | 6a, 7, 8, 9, 10, 10a |

n.d. – not detectable

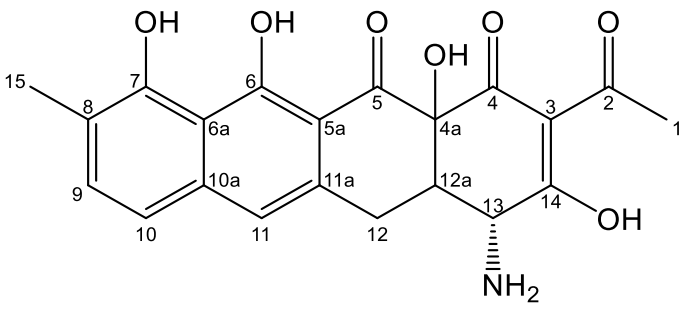



**Figure S4- 7. Structure of CHD-398**

$^{1}$H NMR (501 MHz, METHANOL-$d_4$) δ ppm 2.28 - 2.35 (m, 3 H) 2.56 - 2.65 (m, 2 H) 2.90 - 3.04 (m, 1 H) 3.19 - 3.28 (m, 1 H) 3.20 - 3.30 (m, 1 H) 3.51 - 3.55 (m, 1 H) 4.95 (br d, *J*=4.27 Hz, 1 H) 7.09 (s, 1 H) 7.13 - 7.18 (m, 1 H) 7.45 - 7.50 (m, 1 H)

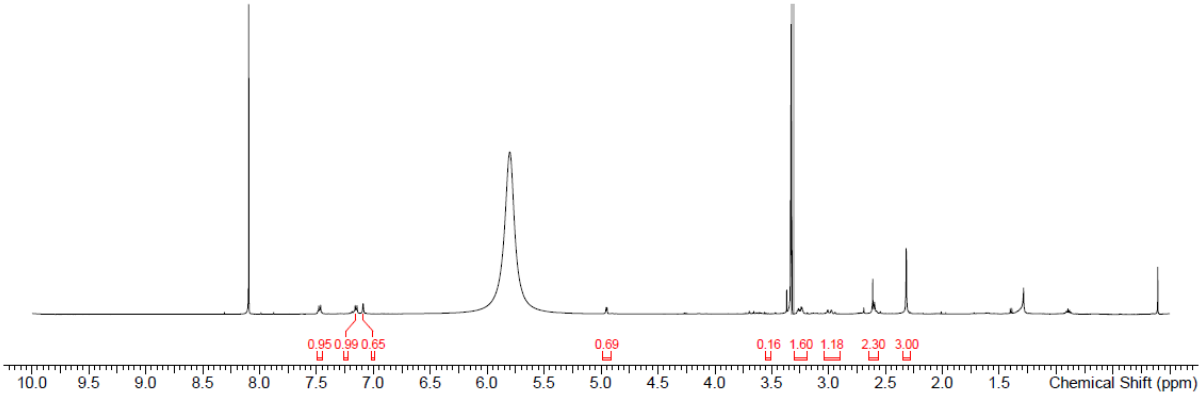



**Figure S4- 8. $^{1}$H spectrum of CHD-398 acquired in $CD_3OD$ at 500 MHz**

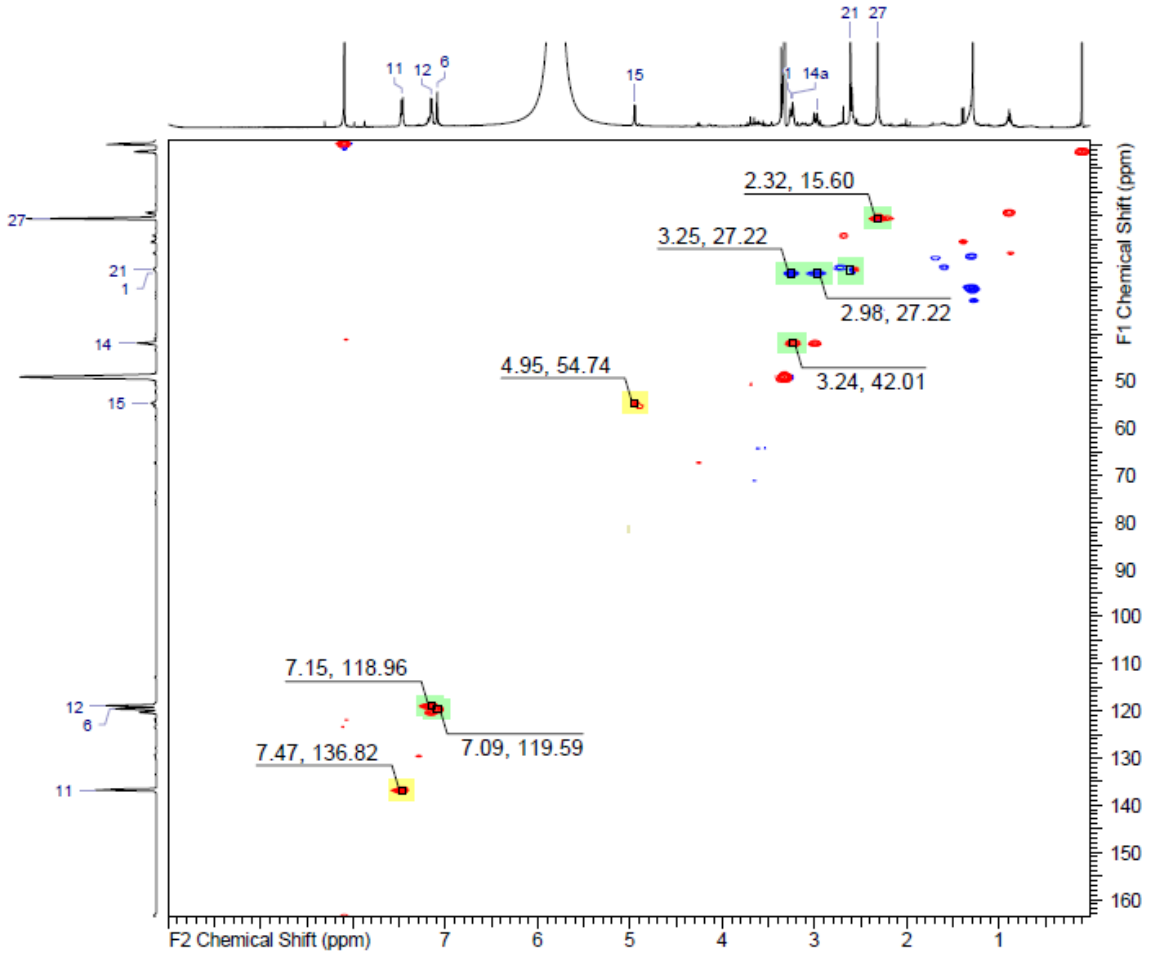



**Figure S4- 9. HSQC spectrum of CHD-398 acquired in $CD_3OD$ at 500/125 MHz**

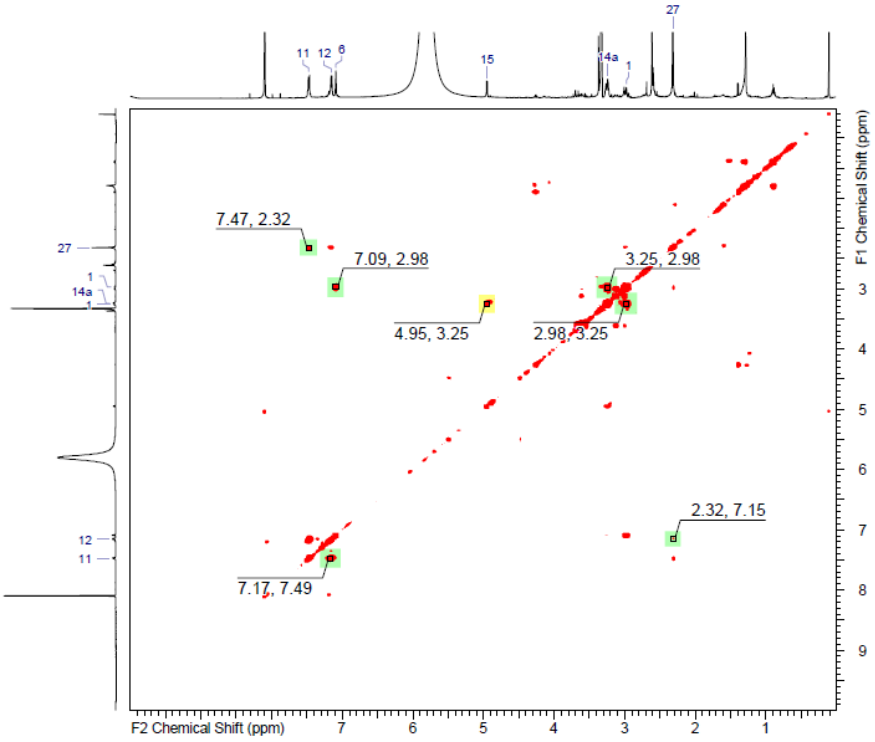



**Figure S4- 10. COSY spectrum of CHD-369 acquired in $CD_3OD$ at 500 MHz**

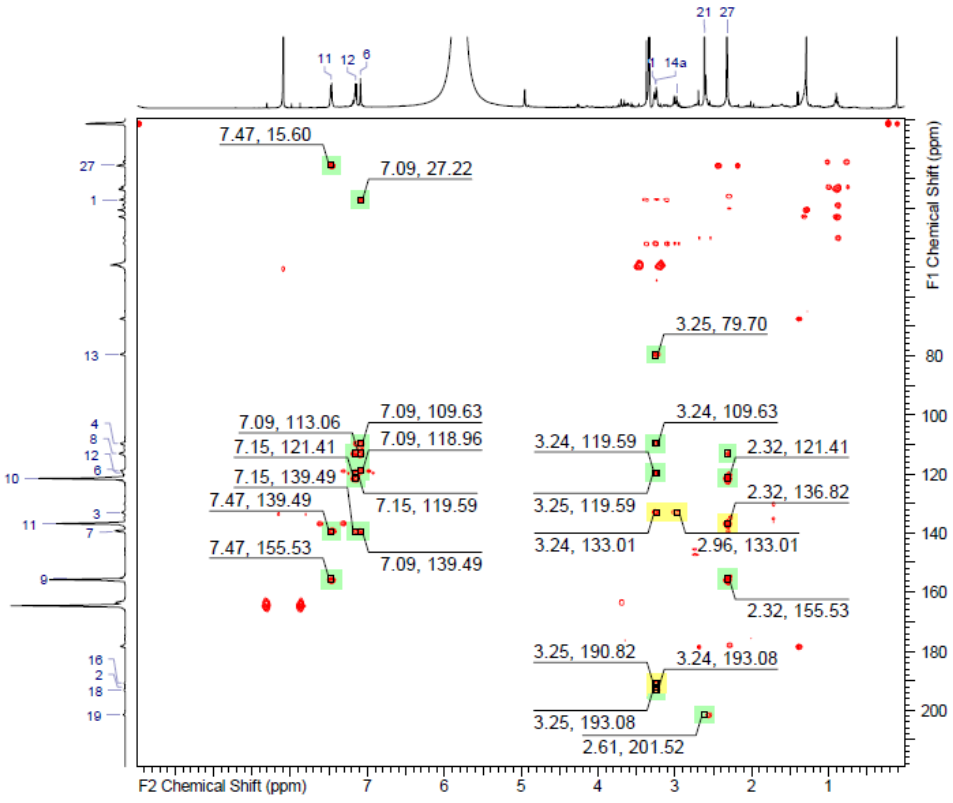



**Figure S4- 11. HMBC spectrum of CHD-369 acquired in $CD_3OD$ at 500/125 MHz**

# 6 Discussion

The projects presented in this thesis focus on diverse aspects of natural product (NP) research, with the main emphasis on deciphering the biosynthesis of novel and known bacterial secondary metabolites, mostly through genetic manipulations of the native producer strain. Chapter 2 focused on genome mining in novel myxobacterial strain *Cystobacterineae* sp. MCy9003. Among other findings, two structurally and biosynthetically interesting NP families of myxopentacins and myxoglucamides were discovered. Their biosynthetic gene clusters (BGCs) were identified and their biosynthesis was proposed. Chapter 3 described the genetic manipulation procedures for the native globomycin producer *Streptomyces hagronensis* strain 360, which led to the identification of the globomycin BGC, allowing for the first time a biosynthesis proposal and supporting yield improvement attempts. Similarly, the genetic manipulation procedures devised for the fidaxomicin producer *Actinoplanes deccanensis*, permitted the generation of gene knockout mutants and their analysis was described in chapter 4. The fifth chapter in major part presented a study of the cyclization of the known atypical tetracycline antibiotic chelocardin in the native producer, the actinobacterium *Amycolatopsis sulphurea*. Gene knockout and site-directed mutagenesis approaches were used to generate mutants, which led to the isolation and structure elucidation of novel chelocardin derivatives. The putative aromatases and cyclases proved essential for chelocardin biosynthesis although the function of each was not complementable by the homologs from the oxytetracycline BGC.

## 6.1 The discovered natural product as a platform for further modification

Besides other living organisms, bacteria are an important source of natural products (NPs). Their indispensability for our everyday lives has been consolidating ever since the 1940s to the present time, encompassing our health care improvement, more efficient food production and conservation. Considering only antibacterial agents, 69 % of antibiotics originate from NPs, with 97 % isolated from microbes but since the year 2000, 77 % of approved antibiotics are NPs, all of them derived from microbes [10]. The unmodified NPs and their derivatives represented 43 % and 14 %, respectively, of FDA-approved new molecular entities (NMEs) before 1940 but since the year 2000, their shares represent 5.3% and 28%, respectively [10]. By the end of 2013, the unmodified microbial NPs represented 7.5 %, while 9.6 % of them were semisynthetic derivatives and 4.0 % synthetic derivatives [10]. The NPs exhibiting potent biological activities that might be potential drug candidates undergo structure-activity relationship (SAR) studies, in order to find the correlation between the chemical structure of

the molecule and its biological activity. Such analysis identifies the chemical groups responsible for the target biological effect in the organism. Methods such as semisynthesis or synthesis inspired by the NP are often used for further improvement of the structure to meet the pharmaceutical requirements for medical applications. The semisynthesis approach was successfully used for the approval of antitumor epothilone semisynthetic derivative ixabepilone [63] and the lipoglycopeptide antibiotic oritavancin [310]. Preferably, native producers are used for NP-derivatization and studying of its biosynthesis. If they can be cultivated under laboratory conditions and genetically manipulated, their genomes can be engineered. The deletion of individual biosynthetic genes results in mutated bacterial strains that produce structural derivatives of NP, which consequently explains the NP biosynthesis. Such approach was used to confirm the functions of cyclases and methyltransferase in chelocardin biosynthesis in native producer *Amycolatopsis sulphurea* (Chapter 5). The growth conditions and genetic manipulation protocols were established prior to the study [32]. The actinobacteria produces chelocardin in high yields, up to 2 mg/ml, and therefore provides a good system to study gene deletion effects on its structure. The enzymes that proved essential for the biosynthesis were confirmed, which is necessary for future heterologous expression studies. Additionally, their function was characterized to a certain extent. Moreover, gene knock-outs were performed in fidaxomicin native producer *Actinoplanes deccanensis*. Adaptation of the culture media and the conjugation protocol from the existing literature were key to success. Although fidaxomicin is a medically approved antibiotic against Gram-positive bacterium *Clostridium difficile* [142], its macrolide scaffold was found as well in mangrolide A, originally reported to act as a Gram-negative antibacterial agent [145]. Along this lines, the macrolide scaffold therefore shows promise for further modifications. The in-frame deletion of glycosyltransferases as reported in chapter 4, which attach noviose and rhamnose moiety to each side of the macrolide scaffold, resulted in shunt products [80] that can be used for attachment of synthetically generated fragments. The semisynthetic fidaxomicin derivatives might exhibit a different panel of biological activities, preferably against Gram-negative bacteria. The literature on studied actinobacterial species *A. sulphurea*, *A. deccanensis* and *S. hagronensis* contained valuable information about the growth and antibiotic production conditions as well as genetic manipulation protocols that could be readily used or adapted for each bacterial strain, which is not the usual scenario for novel bacterial isolates.

## 6.2 Genetic manipulation and gene tools for new myxobacteria and actinobacteria species

The cultivation of actinobacteria under laboratory conditions is well established, due to many decades of research and the International Streptomyces Project (ISP) that provided media and methods for characterization of *Streptomyces* species [311]. In contrast, only selected myxobacteria are well studied in regard to cultivation [312, 313] although many advances have been recently made such as successful cultivation of bacteria previously believed to be "unculturable" [314]. Gene tools for genetic manipulation of actinobacteria, encompass various plasmid types [27], e.g. suicide [309], replicative, integrative plasmids [295], and promoters of different strengths [315–317]. The report on bacteria-derived CRISPR-Cas9 system that enables precise and efficient genome editing in living eukaryotic cells [318–320] encouraged scientists to adapt the technology to *Streptomyces* bacteria. They achieved a faster and more effective method for creating and selecting gene knockout mutants [321]. The usual transformation methods in actinobacteria are conjugation of spores [27] or mycelium [276], protoplast transformation [27], transformation of mycelium [293] and electroporation [322]. There are several gene tools also available for myxobacteria, albeit not the full repertoire named above. Widely used is the pMycoMar plasmid containing *Himar1* transposase [323], which helped with the identification of a few myxobacterial biosynthetic gene clusters [324]. Nonreplicative plasmids pSBtn5 and pFPVan, containing *tn5* and vanillate promoters, respectively, proved to be useful tool for genome mining [158]. The gene deletion is possible in model organism *M. xanthus* by counter-selection with *sacB* gene, conferring sucrose sensitivity [325]. The replicative plasmid pMF1 was successfully used in *M. xanthus* [326]. The plasmid integration is possible using integrase genes and attachments sites for bacteriophages Mx8 [327] and Mx9 [328] as well as the heterologous expression of BGC in *M. xanthus* using origins of replication achieving a single plasmid copy per bacterial cell [329, 330]. The model organism *M. xanthus* is electroporatable [161] as well as some novel myxobacterium strains from the suborder *Cystobacterineae* (Table 2) but there are many other novel isolates where transformation is not possible [167, 331]. Conjugation methods are readily available for *Sorangium cellulosum* So ce56 [332, 333], *S. cellulosum* So ce12 [334] and *Chondromyces crocatus* Cm c5 [335], from suborder of *Sorangineae*, such as biparental or triparental mating. Up until today, the protocols for genetic manipulation in *Sorangium* strains have not yet been adequately developed and genetic methods established in one *Sorangium* strain often cannot be applied to others [336].

### 6.3 Genome mining as a source of new natural products in myxobacteria

Bacteria, such as actinobacteria and myxobacteria, are known as abundant producers of secondary metabolites, due to their complex habitat and daily struggle to survive [28]. Moreover, the predatory myxobacteria lyse their prey, which releases nutrients for their growth [42] as well as the genomic DNA. Myxobacteria, with the largest bacterial genomes from 9 to 14.8 Mb [38], might obtain part of it through the horizontal gene transfer [337, 338], providing them with an evolutionary advantage against the already met opponents [339]. The biosynthetic gene cluster (BGC) abundance, as a virtual source of NPs, in actinobacteria was especially clear after the first whole-genome sequence of *S. coelicolor* A3(2) was published in 2002 [36]. The whole-genome sequences of myxobacteria were available later, with the model organism *Myxococcus xanthus* in 2006 [338], the model *Sorangium* strain *S. cellulosum* So ce56 in 2007 [48] and *Corallococcus coralloides* DSM 2259 in 2012 [340], as well as revealing the potential for large secondary metabolomes. As only a few NPs were observed to be produced per strain, the rest of the BGC was considered silent or cryptic, which led to a surge in activities to connect these BGCs to small molecules through genome mining [157]. Many genetic approaches were developed in the last 20 years since the genome data became available, most of them leaning onto the bioinformatic tools such as antiSMASH [115], detecting and analyzing the genome information (see Introduction). Genome mining in myxobacteria assisted to reveal the pathways of previously isolated NPs such as myxochromide S [341], myxochelin [342], myxovirescin [234], myxalamid [343], Dkxanthene [344], rhizopodin [345], crocapeptin A [346], haprolid [347], phenalamide [348], crocagins [335], ripostatin [349] and 3,5-dibromo-*p*-anisic acid [350]. Moreover, the genome mining approaches recently explained the origin and biosynthesis of chloromyxamides [167], found by isotope pattern and MS-MS analysis of the novel myxobacterial crude extract, and cytotoxic pyxipyrrolones [331] detected after *in silico* prediction of a novel antibiotic and thorough HPLC-DAD-MS of inactivation mutant analysis. If a native producer is not genetically manipulable and therefore the biosynthetic genes cannot be accessed, the BGC can be assembled or simply introduced onto a plasmid, using TAR cloning [351], RecET direct cloning, Redαβ recombineering [352] or cosmid libraries, and heterologously expressed in an engineered model microorganisms with usually well-defined media per NP-type, which contributes to improved NP yields. The BGC of interest on the plasmid simplifies the manipulation of the biosynthetic genes and consequently grants the access to the biosynthetic genes [353]. The heterologous expression of novel or known BGCs, combined with genome mining approach, is an effective method for either discovery of new NPs, explanation of the biosynthesis and identification of previously undetected derivatives

[354]. On the other hand, the genome modifications in the native producers altered their secondary metabolome and therefore novel NPs were detected via LC-MS chromatography, e.g. myxoprincomide [159], crocadepsins [355], pyxidicyclines [158], alkylpyrones [164], as well as the myxopentacins and myxoglucamides, described in this thesis. Myxopentacins were first observed in *Cystobacterineae* strain MCy9003 after overexpression of an NRPS protein via a single homologous recombination of a nonreplicative plasmid harbouring a *tn5* promoter but were detected afterwards in secondary metabolomes of the WT strain and two other myxobacterial strains as well. The overexpression mutant of another NRPS-encoding gene, similarly, showed significant differences in LC-MS chromatogram compared to the wild-type strain. Isolation and structure elucidation revealed the chemical structure that was not matching the putative NP predictions of the above mentioned NRPSs. As the overexpression of some other BGCs, encoded in the genome of MCy9003, as well had impact on myxopentacin production, we assume that BGCs are somehow connected and regulated by an unrevealed complex regulatory mechanism. The myxopentacin BGC was found later through a protein sequence similarity with amipurimycin BGC [135, 181]. On the other hand, myxoglucamides were discovered by a simple simultaneous "activation" of biosynthetic genes *mxgH* and *mxgI* via a single homologous recombination of a nonreplicative plasmid harbouring two *tn5* promoters, oriented in opposite directions. In this case, the newly isolated and structurally elucidated myxoglucamides could be correlated to the biosynthetic genes (see Chapter 2). The compound was not found in the wild-type MCy9003 strain. Moreover, no alternative producers could be detected.

### 6.3.1 The biological activity aspect

The modifications of the genome, based on the genome sequence analysis, in some cases result in successful discovery of novel NPs. Due to the on-going emergence of bacterial resistance against currently used antibiotics, novel NPs are tested against pathogenic bacteria and cell cultures but some NPs do not exhibit such biological activities. The classical "bioactivity-guided" screening, followed by crude extract fractionation and isolation of secondary metabolites from novel bacterial strains is one of the approaches that ensures the bioactivity of the discovered molecule [356] and it can be also used in combination with genome mining [357]. As also other biological activities are desired in NPs, apart from antimicrobial and anticancer activities, different screening protocols were developed, e.g. focused on protease inhibitors [358] or anti-inflammatory agents [359]. Since the genome mining approach enables the expression of not- or barely-expressed biosynthetic genes, the chances of uncovering a

secondary metabolite less important for bacterial defence and survival might be increased. For the detection of NP with a potential biological activity, bacterial genomes can be screened for resistance genes in the vicinity of the BGCs. The "resistance-guided" approach is based on the observation that the microbial producers of antibiotics must provide themselves with means of self-protection [360], e.g. with transport proteins, molecule or target modifying enzymes [361] or just containing another copy of an essential protein, like a signal peptidase II in globomycin (GlmB) and myxovirescin BGC [140]. The search for BGCs with associated resistance genes increases the possibility that the encoded molecule might be an antibiotic [158, 362]. A few examples confirming the hypothesis are known, among others DNA polymerase III beta subunit GriR as a part of griselimycin BGC [123] and pentapeptide repeat protein CysO in cystobactamids suggesting an inhibitory activity against bacterial type IIa topoisomerases [55]. Moreover, the Antibiotic Resistant Target Seeker (ARTS) tool that identifies duplicated copies of essential housekeeping genes, associated with BGCs [125], was developed for that purpose. The identified resistance protein can also provide information on the potential mode of action and possible target of the resulting molecule [360]. The number of BGCs that are found in the genomes of novel bacteria and have no similarities to already known NPs is overwhelming and host-self-resistance as a prioritization criterion is most welcome.

### 6.3.2 Genome mining in MCy9003

The novel isolate MCy9003 was after 16S rRNA analysis found to represent a novel myxobacterial genus. The strain is a member of the *Myxococcaceae* family but forms a separate branch deviating from the genera *Myxococcus* and *Corallococcus* (Figure 66). As shown by Hoffmann *et al.*, a number of myxobacterial NPs can be produced by bacteria from different genera, whereas a subset of NPs is either unique or highly specific to a certain genus [23]. The selection of MCy9003, a representative of a novel genus, therefore offered higher chances of finding a novel NP scaffold. The whole-genome sequence analysis of 11.3 Mb genome by antiSMASH online tool [115] showed 40 BGCs but an additional manual analysis revealed existence of 5 more, due to the program algorithm limitation that cannot recognize a separate BGC if located next to each other. Out of 45 BGCs, 8 are of NRPS origin, 1 PKS type I, 1 PKS type II, 1 PKS type III, 12 hybrid NRPS-PKS type I BGCs and the remaining 22 are two siderophore BGCs, lanthipeptides, bacteriocins, terpenes and two BGCs of with some NRPS or PKS features (Figure 67). Only six BGCs encode known NPs, namely geosmin [162], carotenoids [163], alkylpyrones [164], myxochelin [165], fatty acids VEPE/TG-1 [166] and chloromyxamide [167].

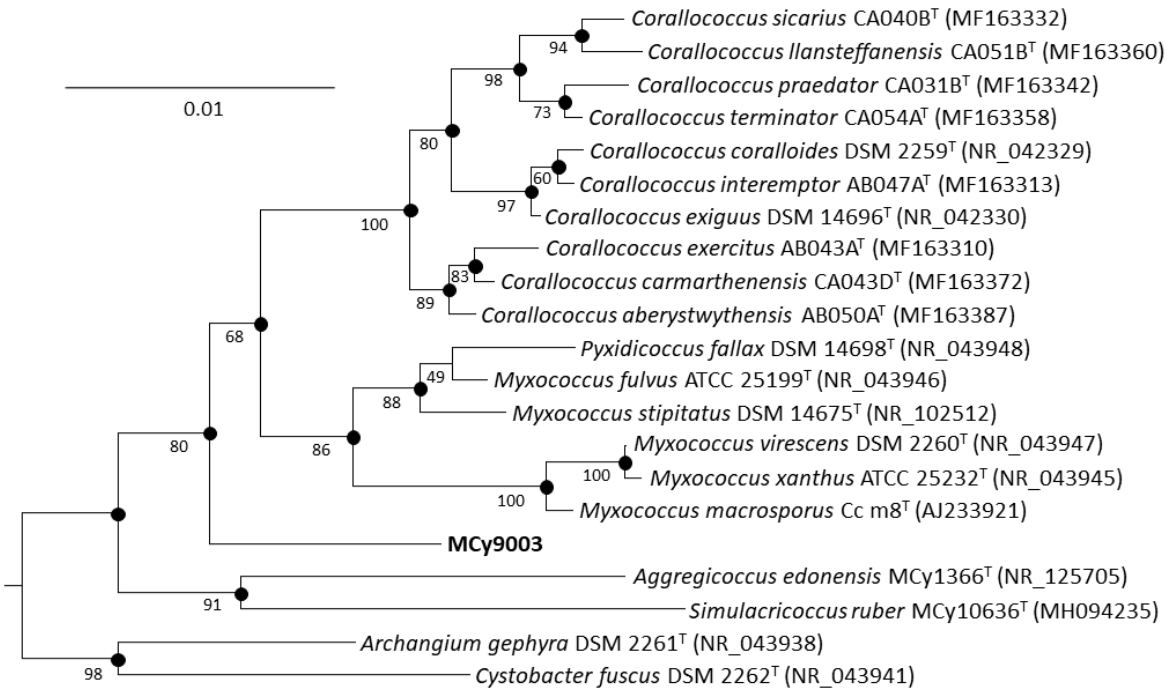



**Figure 66. Neighbor-joining phylogenetic tree inferred from 16S rRNA gene sequences showing the position of the novel isolate MCy9003 in the *Myxococcaceae* family.** GenBank accession numbers are indicated in parentheses. Values at branch points indicate bootstrap support as percentages based on 1000 resamplings. Dots indicate that corresponding nodes were also recovered in the tree generated with MrBayes and the RAxML algorithms. The sequences of *Archangium gephyra* DSM 2261T and *Cystobacter fuscus* DSM 2262T were used as outgroups to root the tree. Bar, 0.01 nucleotide substitution per site. The phylogenetic tree was prepared by Dr. Ronald Garcia.

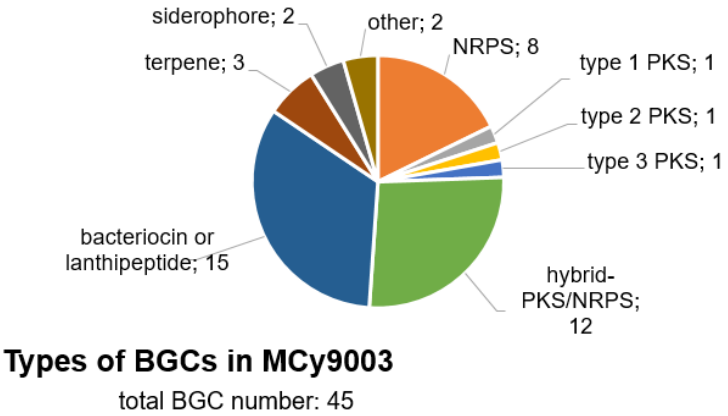



**Figure 67. The biosynthetic potential of *Cystobacterineae* strain MCy9003 for microbial NPs**

Almost half (~44 %) of all the BGCs is intended for the biosynthesis of peptides, even when lanthipeptides and bacteriocins are not taken into account. Moreover, the ratio between C and KS domains in a few secondary metabolite-producing bacteria further emphasizes the MCy9003 as highly specialized for peptide biosynthesis (Table 14). The representation of all

the C and E domains from MCy9003 in Figure 68 shows that the domains of each subtype, except the heterocyclization domains, can be found in MCy9003. The sequence alignments show that in every domain subtype there is further clustering of domains regarding the protein sequence origin. We can observe that in either $^{D}C_{L}$, $^{L}C_{L}$ or E subtype many of the domains from MCy9003 cluster together and are therefore probably "native" to the strain (Figure 68). The rest is clustering elsewhere since those BGCs might originate from other bacteria and were taken up from the environment in the evolution process. Although there is some knowledge on C and E domains regarding the conserved motifs [76] and subtype classification [137], new domains are being found that could play a different role. For example, the terminal E domain in MbtF from mycobactin BGC is proposed to be involved in the intramolecular formation of a peptide bond and NP release [206], similarly as the terminal C domain in MxpA (Figure 27); the function of C domain in MxgH is yet unknown (Figure 68, Section 6.4.3); or the $^{D}C_{L}$ domain, unusual regarding the location, in between two PCP domains, in the biosynthetic enzyme AmbE, and its proposed function – water elimination, which is followed by isomerization and decarboxylation (Figure 69) in methoxyvinylglycine biosynthesis [214], just to name a few. More detailed analysis has shown that a NRPS modules in BGC-35A and in BGC-39B, predicted to incorporate glycine, do not contain E domains upstream of $^{D}C_{L}$ domains. Moreover, the His-motif HHxxxDG characteristic for E and C domains was found inactive in some domains of three BGCs, namely in MxgJ, MxgN C domains and in the E domain of MxgK from BGC-18 (HAxxxDG), in an $^{L}C_{L}$ domain of BGC-35A positioned between E-$^{D}C_{L}$-PCP and A-PCP domains, and in an $^{L}C_{L}$ domain of BGC-34A located between KS-AT-PCP and A-Ox-$^{L}C_{L}$-A-PCP domains. In most of the cases with potentially inactive domains, the biochemical transformation that they assumingly catalyze could be taken over by another domain located in the vicinity.

**Table 14. The comparison of NRPS and PKS biosynthesis potential in various secondary metabolite producers**

| | Number of domains | | Ratio | |
|---|---|---|---|---|
| **Strain** | **C domains** | **KS domains** | **C/KS** | **Genome size** |
| So Ce56 | 16 | 48 | 0.33 | 13.03 Mb |
| *Burkholderia thailandensis* E264 | 19 | 29 | 0.65 | 6.72 Mb |
| *S. coelicolor* A3(2) | 14 | 18 | 0.77 | 8.67 Mb |
| *Bacillus subtilis* subsp. *subtilis* str. 168 | 21 | 16 | 1.31 | 4.25 Mb |
| *Myxococcus xanthus* DK1622 | 77 | 45 | 1.71 | 9.14 Mb |
| MCy6431 | 201 | 67 | 3.00 | 13.2 Mb |
| **MCy9003** | 122 | 21 | **5.81** | 11.3 Mb |

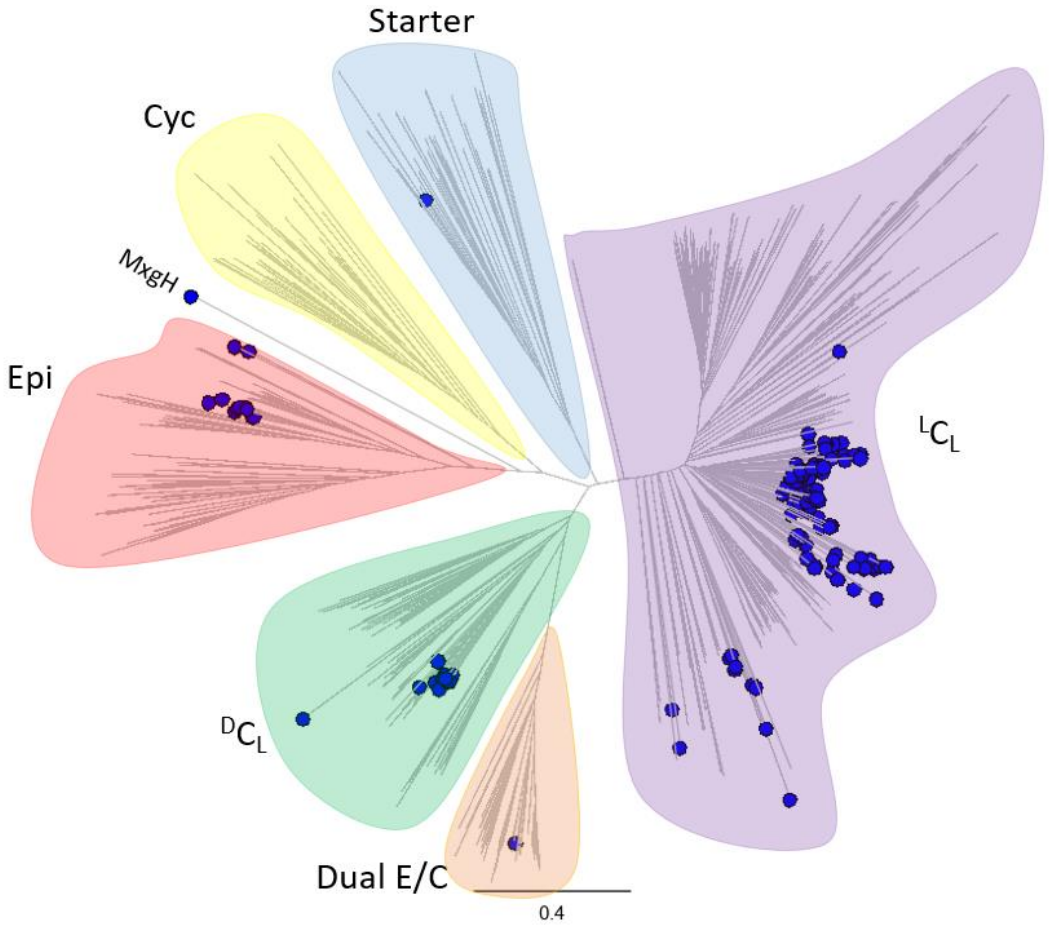



**Figure 68. The phylogenetic tree of MCy9003 C and E domains.** To the 525 C and E domains [137], 122 condensation and 11 E domains from MCy9003 were added. Geneious software's Clustal Omega alignment was made and the tree was build [263]. The domain subtypes are marked. The domains originating from MCy9003 strain are marked with a blue circle.

In the initial bioactivity screening of MCy9003, weak *E. coli* TolC and *B. subtilis*, *S. aureus* and *M. luteus* activities were observed that later after re-cultivation disappeared. The applied bioactivity-guided approach, followed by isolation of weakly bioactive LC-MS fractions, did not provide a desired novel NP. The expression of biosynthetic proteins is thightly correlated to the environmental conditions and stimuli in the bacterial habitat, which result in the production of secondary metabolites that are beneficial for survival of a bacterium [157]. The conditions that trigger the production of such NPs are unknown and hardly possible to reproduce in the laboratory. The closest to the natural conditions might be co-cultivation of a predatory myxobacterium with a prey bacterium on a solid surface such as agar medium [363]. A successful discovery of a novel NP from a novel bacterial isolate using the traditional metabolome screening coupled with bioactivity-guided approach and followed by dereplication is often also a matter of luck. The production of majority of NPs under laboratory conditions reaches very low levels or is below the detection limit since the strain-specific cultivation

requirements for a certain secondary metabolite are unknown. However, an overexpression or an activation of the biosynthetic operon, achieved by targeted genetic manipulation of a bacterium, increases or initiates the expression of biosynthetic proteins responsible for the biosynthesis of the NP. Such genome mining approach offers us access to 'silent BGCs' [21], a common phenomenon in all the identified secondary metabolite producing bacteria, but there is no guarantee that the selected BGC is functional or that the BGC with some inactive catalytic domains does not produce a secondary metabolite. The knowledge we currently have about biosynthetic genes and biosynthesis of NPs is broad but at the same time very limited. The analysis of the MCy9003 genome was initially focused on the prioritization of the BGCs harbouring a second copy of a potential self-resistance gene in the vicinity of the biosynthetic genes as described in chapter 2 that were detected by an in house-developed online tool Clust-o-matic [364]. Only three potential self-resistance proteins were identified. Since MCy9003 is a representative of the novel genus and was found to be different from the rest of bacteria, due to its specialization for peptide biosynthesis, and has shown bioactivity potential in the initial screening, we continued with the BGC activation or inactivation. In the next round, the selected BGCs were prioritized by the complexity of the operon structure. The overexpression of a BGC is achieved by a plasmid insertion via a homologous region encompassing the initial part of the biosynthetic gene preceded by a constitutive or an inducible promoter. After such single-crossover homologous recombination event, the plasmid remains in a mutated bacterium and the latter always requires cultivation in the presence of the antibiotic. The overexpression of a BGC with higher number of biosynthetic operons could be in MCy9003 only done partially since only one selection marker could be used for genetic manipulation of the strain to integrate site-specifically one promoter. However, in some cases two promoters could be integrated simultaneously if their operon structures were directed outwards like in myxoglucamide BGC. The BGC-12 and BGC-1 (Table 3) were chosen for the activation due to proximity of a potential self-resistance gene and a simple one-operon biosynthetic gene structure, respectively. The separate overexpression of those two distinct NRPSs resulted in changes of the secondary metabolome. The increased peaks observed in LC-MS chromatograms of both mutants were isolated and their structures were elucidated. The chemical structure of novel NPs, as well as the characteristic MS-MS fragmentation pattern, revealed that the isolated compounds share certain structural features and belong to the same compound family, that we named myxopentacins. As the myxopentacin structure did not coincide with the predicted functions of the catalytic domains of BGC-1 or BGC-12, the search for the corresponding BGC was initiated. Using the in-house database MyxoBase, two myxopentacin alternative producer

strains with available genome sequences were detected. Their genomes were analyzed and common BGCs were found, eliminating only 7 BGCs. Due to unusuality of the myxopentacin features, the corresponding BGC was impossible to predict with the information obtained. The selected BGCs occurring in all three strains were therefore inactivated or activated in MCy9003 strain. However, the responsible BGC-39A was found only after the biosynthetic pathway of amipurimycin was published [135] since both NPs contain a rare cispentacin moiety. The protein sequence similarity to ketomemicin biosynthetic genes [136] additionally explained the biosynthesis of another rare myxopentacin feature – the pseudoarginine moiety. Many of the BGC activation and inactivation attempts of NRPS and hybrid NRPS/PKS type I BGCs (Figure 67) resulted in no significant changes of the secondary metabolome compared to the WT strain. Based on the finding that in *M. xanthus* an MbtH-like protein encoded by MXAN_3118 gene interacts with NRPSs of at least 7 BGCs, and that it is an essential accessory protein for the function of many NRPSs [365] since it was shown to affect their solubility [366], the co-expression of a MbtH-like protein, encoded by MCy9003_02830 gene in MCy9003, in mutant strains could be a missing link for successful overexpression of the encoded nonribosomal peptides and NRP-containing hybrids. However, one of the previously mentioned activation mutant strains exhibited additional three peaks in LC-MS chromatogram, when compared to the WT strain. The MS-MS and dereplication analysis confirmed that the activation of a novel NP was achieved. The isolated compounds belong to the novel NP family, and were named myxoglucamides. The myxoglucamides structurally resemble ieodoglucomides [221, 222] but they harbour an additional α-vinyl moiety, rarely occurring in NPs. Neither myxopentacins nor myxoglucamides exhibit antibacterial activity, whereas myxoglucamides are weakly cytotoxic. However, only antibacterial and cytotoxicity were tested until now but bacterial NPs in general exhibit a wide range of biological activities (see Introduction), giving them a certain advantage against the competitors. Furthermore, not every BGC in a bacterium encodes a secondary metabolite with biological activity since the bacterium requires also other molecules such as iron chelating agents siderophores [121] for daily survival or it might produce a spore pigment [122]. Those NPs are needed only in smaller quantities and only in certain conditions, e.g. spore pigment production in liquid culture is not required. On the other hand, some of the reported NPs, sharing certain structural features with myxopentacins or myxoglucamides, do in fact exhibit either protein inhibitory, antibacterial, antitumor or antifungal biological activities. Due to the similarity of the chemical structure, myxopentacins as well as myxoglucamides might have inhibitory biological activities but have not been tested for those yet. Similarly, the potential anti-viral bioactivity is not yet known. Taken together, MCy9003 strain did in fact

yield novel NP scaffolds. However, the BGC activation approach proved to be more effective compared to the BGC inactivation. This could be due to the selected production medium since the inactivation mutants grown in richer medium might exhibit some differences in the secondary metabolome. There are many more BGCs encoded in this strain that have not been further explored, e.g. lanthipeptides and bacteriocins, therefore discoveries of novel NPs are still possible.

### 6.4 The natural product biosynthesis: Unanswered questions

The successful activation of so-called silent or less expressed BGCs in our case resulted in NPs without desired potent antimicrobial or cytotoxic activity. Their role for the producer microorganism is not yet uncovered. However, from a structural point of view, some of the identified NPs exhibit uncommon structural features due to uncommon biosynthetic domains or exhibit unusual assembly mechanisms.

#### 6.4.1 The α-vinyl moiety in myxoglucamide

Myxoglucamide chemical structure with an α-vinyl group and the BGC with number of diverse oxygenases as well as condensation domains are of special interest. Such α-vinyl group was observed only as α-vinylglycine in the bacterial NP rhizobitoxine [211, 212], the fungal metabolite D-vinylglycine [212], aminoethoxyvinylglycine (AVG) from *Streptomyces* sp. NRRL 5331 [213] and methoxyvinylglycine (MVG) from *Pseudomonas aeruginosa* [214]. The myxoglucamide α-vinyl group as a part of 4-amino-2-hydroxyhex-5-enoic acid is therefore a novel discovery in nature. The α-vinylglycine moieties observed in above mentioned NPs originate either from L-glutamic acid or L-aspartic acid-derived L-homoserine. The NRPS module 1 in MxgJ is predicted to incorporate L-glutamic acid into the growing acyl-peptidyl chain. The same was shown in the *in vitro* reconstitution of the biosynthetic route of an α-vinylglycine moiety-containing nonproteinogenic amino acid MVG or L-2-amino-4-methoxy-trans-3-butenoic acid (AMB) [214]. In the AMB BGC, three enzymes are proposed to catalyze the transformation of L-glutamic acid into the methoxyvinylglycine, however, the mechanism of the last decarboxylation step, yielding the observed α-vinyl group, is not yet clear (Figure 69). The protein BLAST of AmbC, AmbD and of AmbE-C* domain [214] against the MCy9003 genome sequence yielded hits in the myxoglucamide BGC. The α-ketoglutarate-dependent taurine dioxygenase (TauD) present in MxgI showed 44 % and 42 % identity and 70 % and 73 % similarity to AmbC and AmbD, respectively. Moreover, the protein BLAST search of C* domain (AmbE) in a whole-genome sequence of MCy9003, found the $^{D}C_{L}$ domain MxgL

to be homologous, although with only 28 % identity and 42 % similarity. The α-vinyl group formation mechanism in myxoglucamide based on the AMB biosynthesis and analysis of myxoglucamide catalytic domains (Chapter 2) would require the MxgJ NRPS module for incorporation of L-glutamic acid, the α-ketoglutarate-dependent taurine dioxygenase (TauD) domain located at the terminal of MxgI protein and MxgL $^{D}C_{L}$ domain, eliminating the hydroxyl group, which results in the double bond formation, followed by decarboxylation and isomerization. However, the amino acid precursor for the α-vinyl group in myxoglucamide is still unknown and further feeding experiment are necessary.

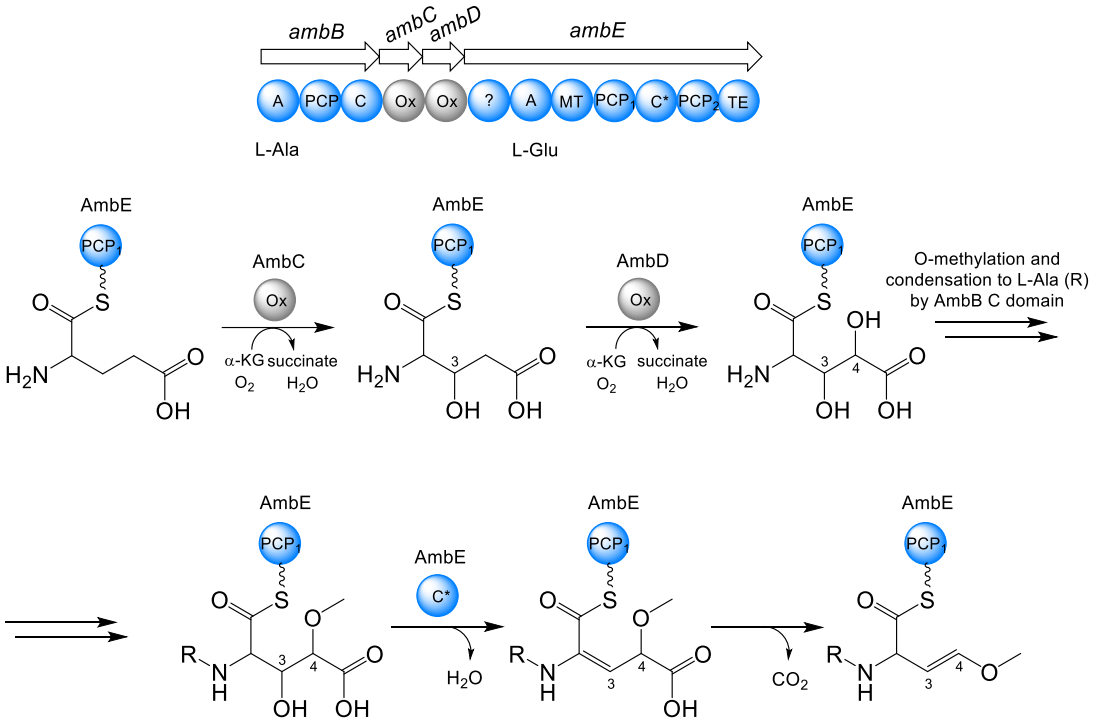



**Figure 69. The proposed AMB biosynthesis.** AmbE A domain is selective for L-glutamic acid. The α-ketoglutarate-dependent taurine dioxygenases AmcC and AmcD catalyze hydroxylation of C3 and C4, respectively. The O-methylation and condensation of L-Ala to the modified L-glutamic acid are not shown here. However, the unusual NRPS C domain (C*) is proposed to catalyze the water elimination of C3 position. The transformed intermediate then undergoes subsequent decarboxylation and isomerization [214].

### 6.4.2 The oxygenases in myxoglucamide BGC

The above mentioned α-ketoglutarate-dependent taurine dioxygenase (TauD) domain encoded by MxgI is a member of TfdA family, which are known to catalyze distinct important biological reactions, including steps in the biosynthesis of collagen and antibiotics, the degradation of xenobiotics, the repair of alkylated DNA, and the sensing of oxygen and response to hypoxia [367]. However, more oxygenase types are encoded in the MXG BGC. A flavin-dependent monooxygenase is encoded by MxgN and a cytochrome P450 by MxgG. The former are

bacterial luciferases or luciferase-like monooxygenases (LLMs) that catalyze the oxidation of a long-chain aldehyde and release energy in the form of visible light. Many BGCs contain LLM-encoding genes but the precise biosynthetic roles are poorly understood [368]. The LLMs that have been reported and studied are tailoring enzymes MsnO2, MsnO4 and MsnO8 in mensacarcin [369], GarO in actagardine [370], Hgc3 in hygrocin biosynthesis [371], OvmO in olimycin [372], DivO in divergolide [373] and AbmE2 in neoabyssomicin biosythesis [368]. The Hgc3, OvmO, DivO, AbmE2 enzymes catalyze the post-PKS O atom insertion on a macrolactam intermediate, which is recognized as Baeyer–Villiger monooxidation reaction mechanism [374]. Baeyer–Villiger monooxygenases catalyze the oxidation of ketones and cyclic ketones to esters and lactones, respectively, by using molecular oxygen and NAD(P)H, as well as sulfoxidations and *N*-oxidations [374]. According to that definition the GarO LLM, catalyzing unique oxidation of S atom in lanthipeptide [370], could as well belong among them. Interestingly, three LLMs enzymes in mensacarcin BGC have a different role. MsnO acts as a dehydratase, MsnO4 introduces two hydroxyl and one epoxy group onto and into the third ring and MsnO8 catalyzes the epoxidation of the side chain [369]. Moreover, the LLM was found in the main biosynthetic gene of myxoprincomide MXAN_3779, which is biosynthetically predicted to introduce an α-keto functional group in the final structure of NP [159]. Similarly, a flavin-dependent monooxygenase (MxgN) could be involved in the α-carbon oxidation [159, 207] of the hydroxyl group present in myxoglucamide-647 and -661 derivatives (Figure 33). Third type of oxygenase, a cytochrome P450, encoded by MxgG, belongs to a group of heme-mono-oxygenase enzymes [375]. They play a role in essential steps in metabolism: bioactivation and detoxification of drugs and xenobiotics, the synthesis of endogenous compounds, and tailoring of secondary metabolites [375–377]. The cytochromes P450 in *Candida tropicalis* catalyze the transformation of fatty acids [208]. The cytochrome P450 MxgG could therefore hydroxylate the fatty acids, methylpentadecenoic (iso 15:0) or methylhexadecenoid (iso 16:0) acid incorporated by loading module in MxgI, in positions C-4 or C-5 in myxoglucamide 645 and 647 or myxoglucamide 659 and 661, respectively (Figure 33). The timing of that reaction is undetermined. The proposed tailoring reaction steps of MxgG and MxgN are presented in Figure 70.

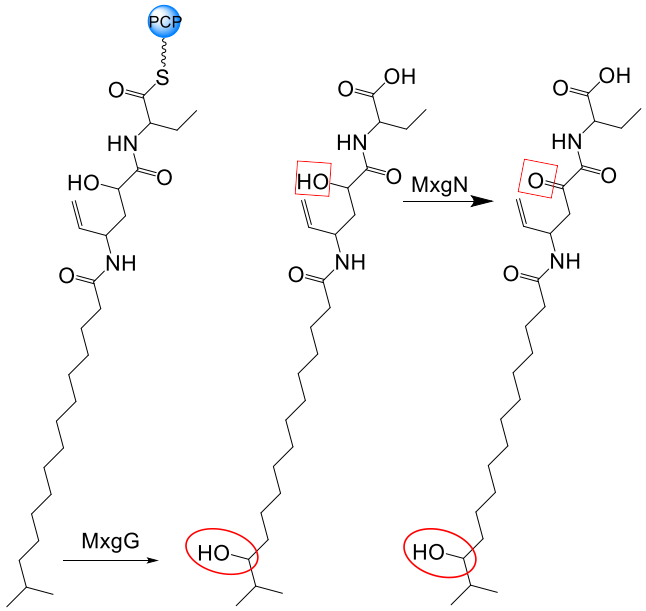



**Figure 70. The oxygenases MxgG and MxgN in myxoglucamide BGC.** A flavin-like monooxygenase MxgN is proposed to oxidize α-carbon position in myxoglucamide-645 and -659. On the other hand, MxgG, a cytochrome P450, could oxidize a fatty acid chain in C-4 or C-5 positions in myxoglucamides, prior to glycosylation.

### 6.4.3 Condensation domain diversity in myxoglucamide BGC

Condensation domains are a part of NRPS modules and catalyze the formation of the peptide bond. Recent analysis of all the known C domains, with an addition of C domains found in glycopeptide-encoding BGCs, performed by Rausch *et al*. [137], deepened the existing knowledge by regrouping them into five functional subtypes. An $^{L}C_{L}$ domain condenses two L-amino acids, a $^{D}C_{L}$ domain catalyzes a peptide bond between an L-amino acid and a D-amino acid, a newly reported Starter C domain acylates the first amino acid with a β-hydroxyl fatty acid, and Heterocyclization (Cyc) domains catalyze the peptide bond formation and subsequent cyclization of cysteine, serine or threonine residues [137]. As known the epimerization domains change the chirality of the last amino acid in the growing chain, whereas the fifth subtype – Dual E/C domains catalyze both the epimerization and condensation [137]. Condensation domains of different types are observed in MXG BGC (Figure 71).

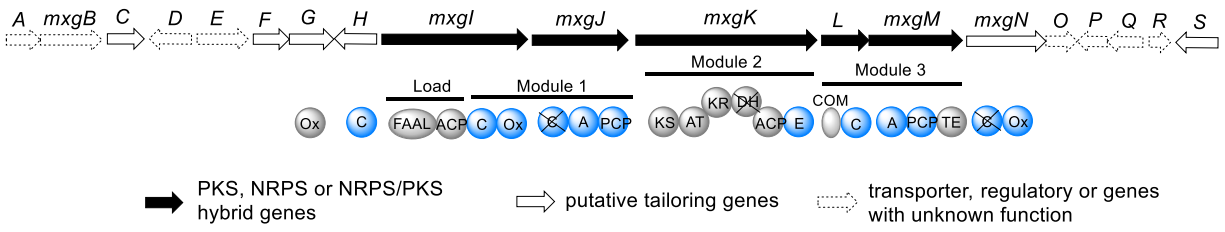



**Figure 71. The myxoglucamide BGC denoting functional domains in biosynthetic genes**

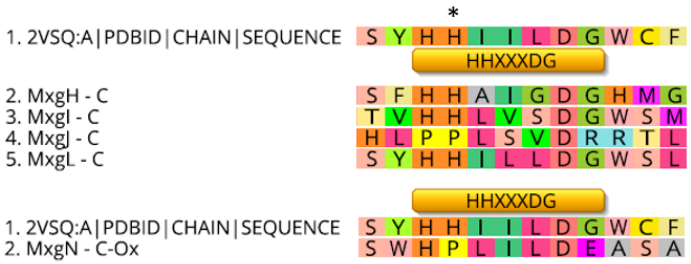



**Figure 72. The alignment of C domains with key residues for catalytic activity.** The C domain in NRPS modules of myxoglucamide were compared to amino acid sequence SrfA-C from PDB: 2VSQ [223] with thoroughly annotated core motives for each of the domains. The key residue is marked with an asterix (*). MxgJ and MxgN proteins are not catalytically active [224].

The MxgI-C domain is assumed to condense the fatty acid loaded by preceding ACP domain with the amino acid selected by A domain in MxgJ. The MxgJ-C domain is lacking a characteristic His-motif (HHxxxDG) essential for the catalytic activity of the domain and we therefore assume that it is inactive (Figure 72). The MxgL-C domain is predicted to be a $^{D}C_{L}$, which fits to the determined D-configuration of the amino acid residue in the chemical structure. The MxgN consists of C and Ox domains. The His-motif, with the second His residue essential for the catalytic activity [137], is not present in MxgN-C domain with HPLILDE residues in the protein sequence (Figure 72). This domain is therefore assumed catalytically inactive. The C domain in MxgH is functional, containing the intact His-motif but its function is unknown.

For better C domain classification, the phylogenetic analysis of E and C domains in myxoglucamide BGC was made using Geneious software [263]. To the 525 protein sequences collected and used by Rausch *et al*. to prove the clustering of C domain by a subtype [137] (Electronic supplementary material: Additional file 6), 6 protein sequences of E and C domains from MXG BGC and C* domain found in AmbE from the methoxyvinylglycine biosynthesis [214] were added. The sequences were aligned using Clustal Omega alignment function and the output was used to build a phylogenetic tree shown in Figure 73. An $^{L}C_{L}$ domain-type is found in MxgH, MxgI and MxgJ (inactive), $^{D}C_{L}$ domains are in MxgL, MxgN (inactive) and AmbE-C*, while MxgK E domain is clustered together with the rest of E domains. According to the antiSMASH prediction, MxgH is of an $^{L}C_{L}$ subtype but in the phylogenetic tree shown in Figure 68, it is clustered elsewhere, in between Cyc and E domains. It almost seems like MxgH-C domain is one of its kind. Its function in the myxoglucamide biosynthetic pathway is unknown and is yet to be determined. Furthermore, the C domain in MxgL and AmbE-C* seem very different (Figure 73) and it cannot be assumed that they would catalyze the same reaction leading to formation of an α-vinyl group found in myxoglucamide.

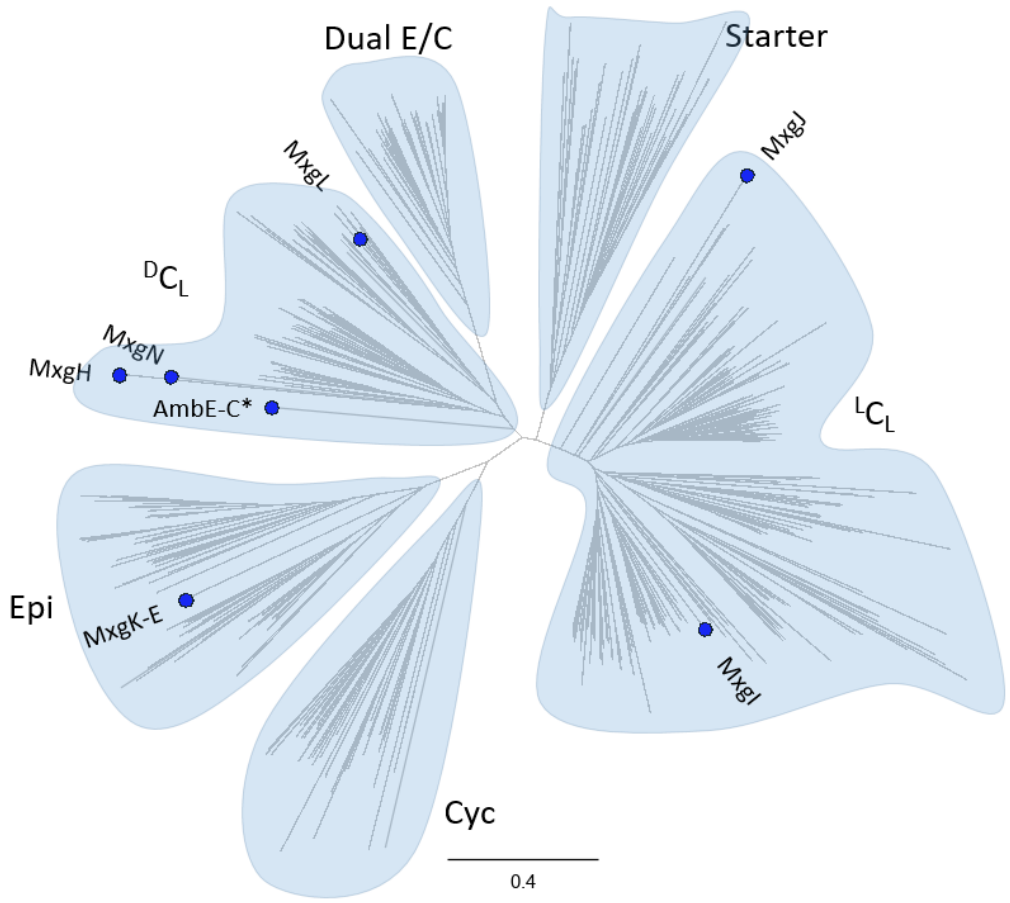



**Figure 73. The phylogenetic tree of myxoglucamide C domains.** To the 525 C domains, including E domains [137], 5 condensation and one E domain from MXG BGC and a C* domain from AMB biosynthesis [214] were added. Geneious software's Clustal Omega alignment was made and the tree was build [263]. The domain subtypes were marked. The domains originating from MCy9003 strain and C domain from AmbE are marked with a blue circle.

### 6.4.4 An iterative NRPS in myxopentacin biosynthesis?

Two different building block sequences were observed upon the structural elucidation of the myxopentacin family members. Myxopentacins 762 and 776 (type A) consist of 6 building blocks, whereas myxopentacins 913 and 926 (type B) contain 7. Type A myxopentacins contain assembled moieties in a certain sequence: the cispentacin moiety is followed by an alanine or a 2-aminobutyric acid, two cispentacin moieties, a pseudoarginine moiety with 2-aminobutyric or 2-aminovaleric acid, and a terminal valine. The pattern observed in type B myxopentacins is different. Five cispentacin moieties are polymerized at the N-terminus and are followed by a pseudoarginine moiety with 2-aminobutyric or 2-aminovaleric acid, and a terminal valine (Figure 74).

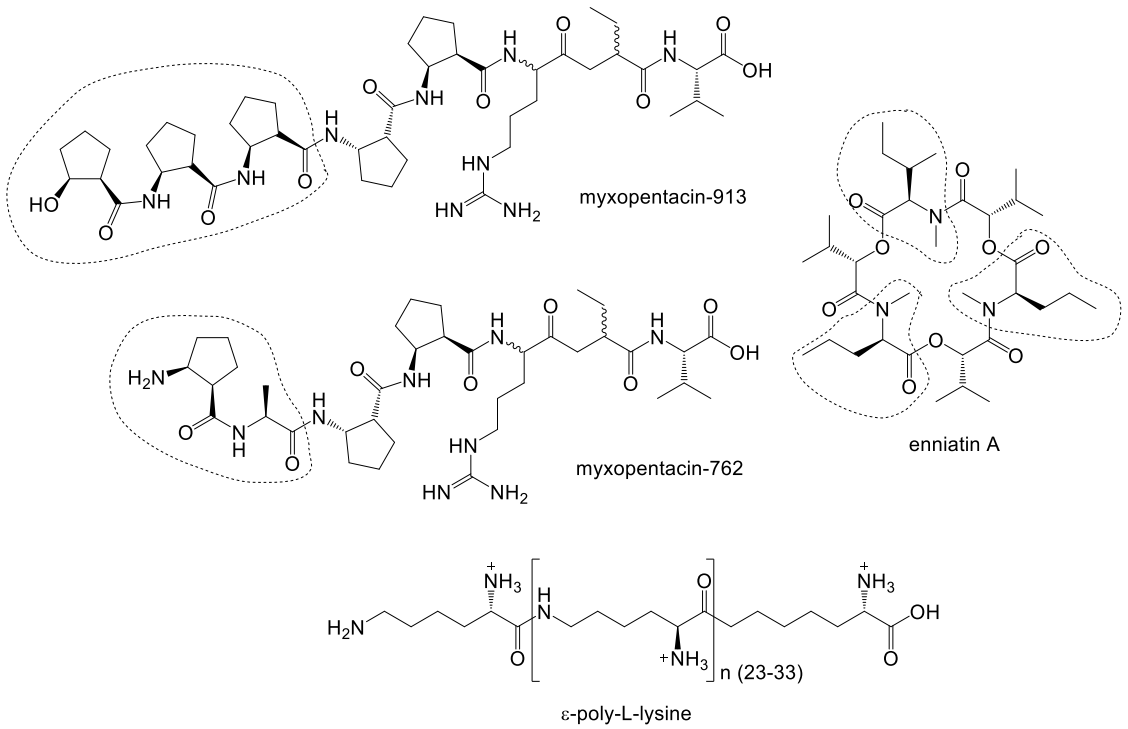



**Figure 74. The molecular structures of NPs assembled with an iterative mechanism.** Myxopentacin-913 partially differs from myxopentacin-762. The difference is marked with dashed curve. A fungal metabolite enniatin A consists out of three sets of branched-chain amino acid and D-hydroxyvaleric acid, forming cyclic depsipeptide. The amino acids in enniatin A are marked with dashed curve. The polymerization of L-lysine amino acids occurs on an ε-PL synthetase, resulting in a peptide chain of 25-35 L-lysines.

However, the inactivation of *mxpA* gene, encoding a NRPS synthetase, resulted in abolished production of myxopentacins of both types. MxpA is thus essential for the myxopentacin biosynthesis. It consists of only two adenylation (A) domains, selecting two distinct moieties, according to antiSMASH prediction- Leu, Ile or Val are activated by an A1 domain and Cha (β-cyclohexyl-L-alanine), Val or Ile by an A2 domain. We assume there is an additional enzyme aiding the building block-assembly in myxopentacins. Nonribosomal peptides (NRPs) are structurally diverse, due to large selection of proteinogenic and nonproteinogenic amino acids and post-translational modifications by tailoring enzymes but their assembly mechanism is highly conserved [74, 209, 378]. They are classified into three categories: linear, iterative or nonlinear with vancomycin, enniatin and vibriobactin, respectively, as the examples. In the linear strategy, the number and the sequence of modules matches with the number and order of amino acids in the peptide. The iterative strategy stands for the modules or domains of the synthetase used more than once to synthesize the peptide, which consists of repeated sequences. The nonlinearly assembled amino acids in NRPs do not correlate to the arrangement of modules on the synthetase template [378].

Enniatins consist of three alternating residues made out of a branched-chain amino acid and D-hydroxyvaleric acid (D-Hiv) (Figure 74) and are synthesized by the two module-enniatin synthase [379]. Instead of as-usual release catalyzed by a final TE domain, in enniatins PCP and C domains catalyze both the transfer and final cyclization [380]. As that kind of pattern or biosynthetic gene cannot be observed in myxopentacins, such assembly seems unlikely. An iterative NRPS module was found in saframycin A BGC from *Streptomyces lavendulae* NRRL 11002. There, a three-module NRPS assembles a tetrapeptidyl scaffold with the last module acting twice per one saframycin A molecule [381]. Iterative hybrid PKS-NRPSs are predominantly found in fungi, with the iterative being the PKS part [382]. A highly unusual NRPS (3.96 kb) was found to catalyze the polymerization of 25-35 L-lysine amino acids into ε-poly-L-lysine (ε-PL), consisting of A domain, PCP domain and three tandem domains C1-C3 [383]. No such domains were observed in myxopentacin BGC. The iterative NRPS mechanism reported for pyrrolamide antibiotics: congocidine, distamycin and new analog is dependent on the amidohydrolase Pya25, which was proven necessary for the polymerization of pyrrole moieties [193] as described in Section 2.3.4.3. A hydrolase (Orf(-7)), an amidohydrolase (Orf(-6)) and two MBL fold metallo-hydrolases (Orf1, Orf2) encoded next to myxopentacin biosynthetic genes would require inactivation in order to evaluate their involvement in the myxopentacin polymerization.

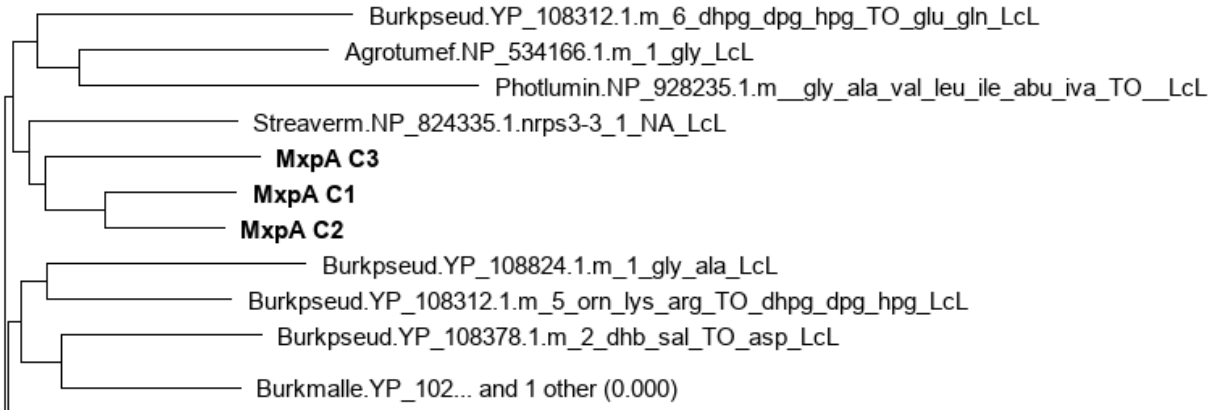



**Figure 75. Phylogenetic tree of MxpA C domains.** To the 525 C and E domains [137], 3 C domains from myxopentacin BGC were added. Geneious software's Clustal Omega alignment was made and the tree was build. All three MxpA domains were classified under $^{L}C_{L}$-type domains.

C domains found in myxopentacin BGC, in MxpA, were added to 525 C and E domains collected by Rausch *et al*. [137]. They were aligned by Clustal Omega (Geneious) and this data was used to build the phylogenetic tree (Figure 75). All three domains are of $^{L}C_{L}$ type. The MxpA C3 domain, potentially involved in a release of myxopentacin from the biosynthetic genes, is not significantly different from other C domains.

### 6.5 Yield improvement of microbial NPs

The production of microbial NP in native producers often reaches low concentrations, which can be plausibly explained due to the limited productivity requirements in their natural habitat [384]. This seems like a bleak picture when contrasted with the much higher requirements of approved antibiotics or their semisynthetic derivatives for human therapy. Yield improvement is therefore a necessary step on the long journey from a newly discovered NP to the medically approved compound. The widely used OSMAC approach (One Strain-MAny Compounds) is based on the observation that changes in growth conditions, which includes media composition, aeration, pH, culture vessel, addition of supplements such as vitamins, metal chelators, rare salts or enzyme inhibitors, etc., influence the expression of secondary metabolites [35]. Higher NP yields are usually achieved by media optimization and strain improvement.

#### 6.5.1 Media optimization

The media optimization is an already traditional yield improvement method also used in this work for minor enhancement of the secondary metabolite profile of *Streptomyces hagronensis* strain 360. Different media, variations of globomycin production medium developed for *Streptomyces halstedii* No. 13912 for fermentations in bioreactor [228], were tested. The production medium required 1 % feather meal [228], containing high protein content of 90 % of its dry weight, which was originally selected probably due to easier accessibility in larger amounts since the feathers are a by-product in the poultry processing industry [385]. In this laboratory no feather meal was available, so other nitrogen sources were tested such as fish meal, corn steep solids, corn steep liquor or soy flour. The best and most efficient production, considering also downstream processing such as isolation and purification of globomycin, was achieved using soy flour and pH 6.0 (Chapter 3). Quick testing of only few different nitrogen sources is accomplished in a span of few days. Moreover, all media optimizations before 1970s were performed by the classical “One-factor-at-a-time” method also used today in the initial stages of research [384]. Such method is simple but also inefficient, expensive and time consuming, when more parameters have to be optimized. Nowadays, the modern statistical approaches save experimental time for the process development. In order to reduce the total price of the product, one must maximize yields and minimize production costs. The appropriate medium components, such as carbon and nitrogen sources, have to be selected based on the metabolic preferences of the producer organism. Enhanced production can usually be achieved by medium optimization or by strain improvement/selection but both approaches are interconnected as they can affect each other [384].

### 6.5.2 Strain improvement

It is clear that biosynthesis of secondary metabolites in microbes is regulated by a variety of environmental and physiological factors such as nutrition, temperature, stress, developmental stage and population density [386–389]. The complex network of secondary metabolism and its regulation are unique for every strain and are generally not well understood, therefore only the trial experiment can help us judge the effect of a certain growth condition on the production of secondary metabolites. The transcriptional regulators can be located in the BGC and have a direct effect on their expression, e.g. the overexpression of transcriptional activator FadR1 in fidaxomicin producer increased the production yield 400-fold [94]. However, the transcriptional regulators of myxobacterial secondary metabolism are typically not co-located with the biosynthetic genes and an additional effort if required to identify them [390]. In other cases, the transcription of this particular NP is directed by global regulators such as γ-butyrolactone regulatory system, important for antibiotic production and morphological differentiation [391], that acts through cluster-situated regulatory genes [386, 392]. The gene knock-outs of the γ-butyrolactone transcriptional repressor ArpA homologs ShbR1 and ShbR3 increased the production of validamycin in *S. hygroscopicus* 5008 for up to 26 % [388]. Similarly, the disruption of ArpA homolog TylP from tylosin biosynthesis, sped up the tylosin production and reached higher yields [393]. However, different outcomes after deletions of ArpA homologs are observed since every gene cluster is regulated by unique pathways [394]. For that reason, the results of deletion experiments cannot be predicted. In order to partially avoid the unknown regulatory pathways in the native producer, the biosynthetic genes clusters can be heterologously expressed in host microorganisms, which come with all the genetic manipulation protocols and basic media for the NP-type. Additionally, strains with minimalized genomes are generated offering increased genetic stability [133, 134]. Simple strain mutagenesis of the essential *rpoB* [395] and *rpsL* genes, achieved by exposing bacteria to high concentrations of rifampin and streptomycin, respectively, was observed to enhance transcription and translation, respectively, which increased production of some NPs [396, 397]. Such attempts are usually followed by selection of best producer strain. Furthermore, the bacteria in some cases contain their own self-resistance mechanism against the produced NP. A strategy for mode of action determination of the antibiotics is by selection of resistant strains among the sensitive pathogens. The resistance is usually developed by a mutation in targeted gene and the comparison of genome sequences with the wild-type strain reveals the target gene [398]. As an example, the resistant *M. smegmatis* mutants were reportedly generated by stepwise increased concentrations of griselimycin. The amplifications of *dnaN* gene were

observed as the griselimycin concentration increased, which provided *M. smegmatis* mutant resistance to griselimycin [399]. Some of the strategies described above were tested on *Streptomyces hagronensis*, the globomycin producer (Chapter 3). The strain is naturally sensitive to streptomycin and therefore contains an intact *rpsL* gene, encoding a 30S ribosomal protein S12 essential for mRNA translation. The mutation of *rpsL* gene was observed to increase the production of secondary metabolites, therefore *S. hagronensis* strain was exposed to 200 μg $ml^{-1}$ of streptomycin. A few colonies grew but due to low number of mutants tested in the production experiment, the globomycin overproducing strain was not achieved. As the exposure to streptomycin causes mutations in different positions in *rpsL* gene, more mutants have to be screened. The overexpression of an additional copy of self-resistance protein signal peptidase II (LspA) in *S. hagronensis* was another strategy for the increase of the production titer of globomycin. A strong, constitutive promoter P*ermE** was used for its overexpression [400] and the plasmid was integrated into the attB site on the bacterial genome [295]. The mutant strain OE-LspA harboured an intact copy of the self-resistance protein GlmB and an additional overexpressed copy elsewhere in the genome. The preliminary results have shown that the production yields decreased. However, two *lspA* genes are present in myxovirescin BGC and were proposed to have an opposite regulatory roles on the production of the natural product [245]. The regulatory system in globomycin producer appears very complex. The global γ-butyrolactone regulatory system might play a role in globomycin regulation, which in the scope of this thesis remains unexplained.

### 6.5.3 Media selection rationalization for cultivation of myxobacterium MCy9003 and actinobacteria

Finding favorable conditions for the cultivation of bacteria is usually time-consuming but of great importance for all the experiments that follow. The novel myxobacterial isolate MCy9003 was initially grown in 10 different media. The extracted cultures were tested for the biological activities against Gram-positive *Bacillus subtilis*, *Staphylococcus aureus*, *Micrococcus luteus*; Gram-negative *Pseudomonas aeruginosa*, *Escherichia coli* TolC, *Escherichia coli*; fast growing *Mycobacterium smegmatis*, fungus *Mucor hiemalis* and yeast *Candida albicans*. The culture grown in CYHv3 medium exhibited only minor biological activities compared to the bioactivities observed when cultivated in other media. We were not interested only in finding the best production medium but also in the medium were bacteria grows fast and reaches high optical density. For transformation via electroporation, the growth in suspension is required, which was not the case at the beginning. Only after the utilization of baffled Erlenmeyer flasks

and a few subcultivations of suspended cells into fresh media, MCy9003 grew in suspension. Moreover, the selected CYS production medium used in our experiments did not provide the most abundant secondary metabolome profile but was chosen due to the simplicity of the medium composition that showed the least background in LC-MS chromatograms. The aim was to easily detect the changes in the secondary metabolite profile of generated BGC-activation and inactivation mutants compared to the wild type. However, screening for the effects of the BGC inactivation in such medium is probably not efficient and other, more complex medium should be chosen for that purpose.

Working with actinobacteria as well requires preparation. Although many information can be found in the literature, each strain should be considered as unique. They require vegetative, production, conjugation and sporulation media. The vegetative medium provides them the environment for rapid growth, while the production medium has to contain carbon and nitrogen sources that initiate the biosynthesis of precursors that regulate the metabolism and influence the production of secondary metabolites [384]. The sporulation medium is of the essence for future genetic manipulation attempts since the conjugation of spores is certainly less time consuming than the conjugation of mycelium. The latter was the only option for *Actinoplanes deccanensis* since the aerial mycelium is absent [29]. The conjugation conditions play a key role in obtaining exconjugants as well and are as well a matter of trial and error.

# 7 Conclusion and outlook

The on-going emergence of bacterial resistance is constantly forcing us to develop effective strategies for the discovery of new antibiotics that would not only fill the gap for a shorter period of time until a new antibiotic is needed again. Scientists try to significantly circumvent this by searching for antibiotics with novel structural scaffolds, ideally inhibiting a target with a completely novel mode of action. Actinobacteria and myxobacteria, two groups of soil-dwelling bacteria, are known producers of natural products with various biological activities, giving them a relative fitness advantages in their highly competitive habitat. As sometimes the natural product produced in highest quantities does not have the desired biological properties, its naturally occurring derivatives are sought. However, the derivatization of a chemical structure can be also achieved via semi-synthesis or by genetic manipulation of the biosynthetic genes in the producer strain. In this study, the combination of both approaches was used for the modification of fidaxomicin antibiotic. The genes encoding glycosyltransferases were knocked-out from the genome of fidaxomicin producer *Actinoplanes deccanensis* and desired shunt products were obtained to be used for further semisynthesis. Such semisynthetic derivatives are

routinely tested for potentially improved biological activities. If the latter are achieved, these compounds would be evaluated for use for clinical treatment in future. Moreover, the genetic modification of chelocardin biosynthetic genes in *Amycolatopsis sulphurea* was used to characterize the enzymes involved in the cyclization and C6-methylation of this atypical tetracycline. The acquired knowledge of the importance of those enzymes for chelocardin biosynthesis enabled a better understanding of chelocardin biosynthetic pathway, and will facilitate its genetic engineering in future. However, sometimes the discovered antibiotics that do not exhibit superior biological activities are soon forgotten but re-discovered later due to their unique features. The Gram-negative antibiotic globomycin inhibits LspA, a signal peptidase II, a protein target of only one other natural product myxovirescin. The BGC of globomycin from *Streptomyces hagronensis*, harbouring an LspA homolog assumed to be a self-resistance protein, was identified. The involvement of the former in the globomycin biosynthesis was confirmed with an abolished production after the inactivation of the main NRPS gene. The activated amino acids in globomycin biosynthetic pathway are therefore proposed to be assembled on modular NRPS enzyme that requires also an additional *in trans*-acting NRPS module and a TE type II. The characterization of an LspA homolog found in globomycin BGC can provide a better outlook on the role of LspA enzymes as self-resistance proteins. In a different line of action, the natural products with novel scaffolds produced by myxobacteria were sought. With genome sequencing becoming more affordable, various genome-based strategies were developed and are successfully used for the discovery of novel NPs. Along these lines, in this thesis two natural product families myxopentacins and myxoglucamides were isolated from the myxobacterium MCy9003, a representative of a novel genus. The genome mining-aimed modifications of the myxobacterial genome led to changes in their secondary metabolome detected by LC-MS profiling. A model for the biosynthesis of myxopentacins and myxoglucamides was proposed. The nonribosomal peptides myxopentacins exhibit a novel natural product scaffold with rare cispentacin and pseudoarginine moieties, and have a unique assembly mechanism that could not be entirely proposed based on the literature. The PKS type I/NRPS myxoglucamides with a rare α-vinyl group have diverse catalytic domains in their BGC, such as different types of oxygenases and condensation domains, or the location of the epimerization domain at the end of the PKS instead of the NRPS module. Both myxobacterial natural products raise many questions regarding their biosyntheses, but at the same time present an opportunity for uncovering of yet unknown ways of Nature.

# 8 References


1 Natural product, 2020. https://encyclopedia.thefreedictionary.com/natural+product (accessed April 17, 2020).

2 Secondary metabolite, 2020. https://encyclopedia.thefreedictionary.com/secondary+metabolite (accessed April 17, 2020).

3 Williams DH, Stone MJ, Hauck PR, Rahman SK: Why are secondary metabolites (natural products) biosynthesized? J Nat Prod 1989;52:1189–1208.

4 Farnsworth NR, Akerele O, Bingel AS, Soejarto DD, Guo Z: Medicinal plants in therapy. Bull World Health Organ 1985;63:965–981.

5 Lobanovska M, Pilla G: Penicillin's Discovery and Antibiotic Resistance: Lessons for the Future? Yale J Biol Med 2017;90:135–145.

6 Baker DD, Chu M, Oza U, Rajgarhia V: The value of natural products to future pharmaceutical discovery. Nat Prod Rep 2007;24:1225–1244.

7 Moellering RC: Past, present, and future of antimicrobial agents. The American Journal of Medicine 1995;99:11s-18s.

8 Schatz A, Bugle E, Waksman SA: Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria.*. Experimental Biology and Medicine 1944;55:66–69.

9 STREPTOMYCIN treatment of pulmonary tuberculosis. Br Med J 1948;2:769–782.

10 Patridge E, Gareiss P, Kinch MS, Hoyer D: An analysis of FDA-approved drugs: Natural products and their derivatives. Drug Discov Today 2016;21:204–207.

11 Silver L, Bostian K: Screening of natural products for antimicrobial agents. Eur J Clin Microbiol Infect Dis 1990;9:455–461.

12 Rolinson GN: 6-APA and the development of the beta-lactam antibiotics. J Antimicrob Chemother 1979;5:7–14.

13 Song YG: The History of Antimicrobial Drug Development and the Current Situation. Infect Chemother 2012;44:263.

14 Baltz RH: Marcel Faber Roundtable: is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? J Ind Microbiol Biotechnol 2006;33:507–513.

15 Mohanty I, Podell S, Biggs JS, Garg N, Allen EE, Agarwal V: Multi-Omic Profiling of Melophlus Sponges Reveals Diverse Metabolomic and Microbiome Architectures that Are Non-overlapping with Ecological Neighbors. Mar Drugs 2020;18.

16 Rust M, Helfrich EJN, Freeman MF, Nanudorn P, Field CM, Rückert C, Kündig T, Page MJ, Webb VL, Kalinowski J, Sunagawa S, Piel J: A multiproducer microbiome generates chemical diversity in the marine sponge Mycale hentscheli. Proc Natl Acad Sci U S A DOI: 10.1073/pnas.1919245117.

17 Chevrette MG, Carlson CM, Ortega HE, Thomas C, Ananiev GE, Barns KJ, Book AJ, Cagnazzo J, Carlos C, Flanigan W, Grubbs KJ, Horn HA, Hoffmann FM, Klassen JL, Knack JJ, Lewin GR, McDonald BR, Muller L, Melo WGP, Pinto-Tomás AA, Schmitz A, Wendt-Pienkowski E, Wildman S, Zhao M, Zhang F, Bugni TS, Andes DR, Pupo MT, Currie CR: The antimicrobial potential of Streptomyces from insect microbiomes. Nat Commun 2019;10:516.

18 Davies J, Davies D: Origins and evolution of antibiotic resistance. Microbiol Mol Biol Rev 2010;74:417–433.

19 CROFTON J, Mitschison DA: Streptomycin resistance in pulmonary tuberculosis. Br Med J 1948;2:1009–1015.

20 Butler MS, Buss AD: Natural products--the future scaffolds for novel antibiotics? Biochem Pharmacol 2006;71:919–929.

21 Hug JJ, Bader CD, Remškar M, Cirnski K, Müller R: Concepts and Methods to Access Novel Antibiotics from Actinomycetes. Antibiotics (Basel) 2018;7.

22 World Health Organization: Global Antimicrobial Resistance Surveillance System (GLASS) Report: Early Implementation 2017-2018. https://apps.who.int/iris/bitstream/handle/10665/279656/9789241515061-eng.pdf?ua=1 (accessed November 11, 2019).

23 Hoffmann T, Krug D, Bozkurt N, Duddela S, Jansen R, Garcia R, Gerth K, Steinmetz H, Müller R: Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. Nat Commun 2018;9:803.

24 Madigan MT, Martinko JM, Bender KS, Buckley DH, Stahl DA: Brock biology of microorganisms, Globakl edition. Boston, Pearson, 2015.

25 Ranjani A, Dhanasekaran D, Gopinath PM: An Introduction to Actinobacteria; in Dhanasekaran D, Jiang Y (eds): Actinobacteria â#x80;#x93; Basics and Biotechnological Applications. [S.l.], InTech, 2016.

26 Kämpfer P: Streptomycetales ord. nov. John Wiley & Sons, Inc. 2015.

27 Kieser T: Practical streptomyces genetics. Norwich, John Innes Foundation, 2000.

28 Traxler MF, Kolter R: Natural products in soil microbe interactions and evolution. Nat Prod Rep 2015;32:956–970.

29 Parenti F, Pagani H, Beretta G: Lipiarmycin, a new antibiotic from Actinoplanes. I. Description of the producer strain and fermentation studies. J Antibiot 1975;28:247–252.

30 Kumari R, Singh P, Lal R: Genetics and Genomics of the Genus Amycolatopsis. Indian J Microbiol 2016;56:233–246.

31 LECHEVALIER MP, PRAUSER H, LABEDA DP, RUAN J-S: Two New Genera of Nocardioform Actinomycetes: Amycolata gen. nov. and Amycolatopsis gen. nov. International Journal of Systematic Bacteriology 1986;36:29–37.

32 Lukežič T, Lešnik U, Podgoršek A, Horvat J, Polak T, Šala M, Jenko B, Raspor P, Herron PR, Hunter IS, Petković H: Identification of the chelocardin biosynthetic gene cluster from Amycolatopsis sulphurea: A platform for producing novel tetracycline antibiotics. Microbiology (Reading, Engl ) 2013;159:2524–2532.

33 Flärdh K, Buttner MJ: Streptomyces morphogenetics: Dissecting differentiation in a filamentous bacterium. Nat Rev Microbiol 2009;7:36–49.

34 Braña AF, Méndez C, Díaz LA, Manzanal MB, Hardisson C: Glycogen and trehalose accumulation during colony development in Streptomyces antibioticus. J Gen Microbiol 1986;132:1319–1326.

35 Bode HB, Bethe B, Höfs R, Zeeck A: Big Effects from Small Changes: Possible Ways to Explore Nature's Chemical Diversity. ChemBioChem 2002;3:619.

36 Bentley SD, Chater KF, Cerdeño-Tárraga A-M, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang C-H, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream M-A, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA: Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature 2002;417:141–147.

37 Baltz RH: Gifted microbes for genome mining and natural product discovery. J Ind Microbiol Biotechnol 2017;44:573–588.

38 Muñoz-Dorado J, Marcos-Torres FJ, García-Bravo E, Moraleda-Muñoz A, Pérez J: Myxobacteria: Moving, Killing, Feeding, and Surviving Together. Front Microbiol 2016;7:781.

39 Dworkin M, Falkow S: The prokaryotes: A handbook on the biology of bacteria, 3rd ed. / Martin Dworkin, editor in chief Stanley Falkow … [et al.], editors. New York, London, Springer, 2006.

40 NOREN B, RAPER KB: Antibiotic activity of myxobacteria in relation to their bacteriolytic capacity. J Bacteriol 1962;84:157–162.

41 Reichenbach H: Order VIII. Myxococcales. Bergey's Manual of Systematic Bacteriology. New York, Springer, 2005.

42 Keane R, Berleman J: The predatory life cycle of Myxococcus xanthus. Microbiology (Reading, Engl ) 2016;162:1–11.

43 Hartzell T: Myxobacteria; in Ltd JW&S (ed): Encyclopedia of life sciences. London, New York, Vols. 21-32, Chichester, West Sussex, U.K., Nature Pub. Group; Wiley, 2002-2010, pp 1–9.

44 Pham VD, Shebelut CW, Diodati ME, Bull CT, Singer M: Mutations affecting predation ability of the soil bacterium Myxococcus xanthus. Microbiology (Reading, Engl ) 2005;151:1865–1874.

45 Lueders T, Kindler R, Miltner A, Friedrich MW, Kaestner M: Identification of bacterial micropredators distinctively active in a soil microbial food web. Appl Environ Microbiol 2006;72:5342–5348.

46 Hyun H, Chung J, Kim J, Lee JS, Kwon B-M, Son K-H, Cho K: Isolation of Sorangium cellulosum carrying epothilone gene clusters. J Microbiol Biotechnol 2008;18:1416–1422.

47 Aravind L, Anantharaman V, Venancio TM: Apprehending multicellularity: Regulatory networks, genomics, and evolution. Birth Defects Res C Embryo Today 2009;87:143–164.

48 Schneiker S, Perlova O, Kaiser O, Gerth K, Alici A, Altmeyer MO, Bartels D, Bekel T, Beyer S, Bode E, Bode HB, Bolten CJ, Choudhuri JV, Doss S, Elnakady YA, Frank B, Gaigalat L, Goesmann A, Groeger C, Gross F, Jelsbak L, Jelsbak L, Kalinowski J, Kegler C, Knauber T, Konietzny S, Kopp M, Krause L, Krug D, Linke B, Mahmud T, Martinez-Arias R, McHardy AC, Merai M, Meyer F, Mormann S, Muñoz-Dorado J, Perez J, Pradella S, Rachid S, Raddatz G, Rosenau F, Rückert C, Sasse F, Scharfe M, Schuster SC, Suen G, Treuner-Lange A, Velicer GJ, Vorhölter F-J, Weissman KJ, Welch RD, Wenzel SC, Whitworth DE, Wilhelm S, Wittmann C, Blöcker H, Pühler A, Müller R: Complete genome sequence of the myxobacterium Sorangium cellulosum. Nat Biotechnol 2007;25:1281–1289.

49 Ringel SM, Greenough RC, Roemer S, Connor D, Gutt AL, Blair B, Kanter G, von Strandtmann: Ambruticin (W7783), a new antifungal antibiotic. J Antibiot 1977;30:371–375.

50 Weissman KJ, Müller R: Myxobacterial secondary metabolites: Bioactivities and modes-of-action. Nat Prod Rep 2010;27:1276–1295.

51 Herrmann J, Fayad AA, Müller R: Natural products from myxobacteria: Novel metabolites and bioactivities. Nat Prod Rep 2017;34:135–160.

52 Vetcher L, Menzella HG, Kudo T, Motoyama T, Katz L: The antifungal polyketide ambruticin targets the HOG pathway. Antimicrob Agents Chemother 2007;51:3734–3736.

53 Bedorf N, Schomburg D, Gerth K, Reichenbach H, Höfle G: Antibiotics from Gliding Bacteria, LIV. Isolation and Structure Elucidation of Soraphen A1α, a Novel Antifungal Macrolide fromSorangium cellulosum. Liebigs Ann. Chem. 1993;1993:1017–1021.

54 Xiao Y, Gerth K, Müller R, Wall D: Myxobacterium-produced antibiotic TA (myxovirescin) inhibits type II signal peptidase. Antimicrob Agents Chemother 2012;56:2014–2021.

55 Baumann S, Herrmann J, Raju R, Steinmetz H, Mohr KI, Hüttel S, Harmrolfs K, Stadler M, Müller R: Cystobactamids: Myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. Angew Chem Int Ed Engl 2014;53:14605–14609.

56 Altmann K-H: Epothilone B and its analogs - a new family of anticancer agents. Mini Rev Med Chem 2003;3:149–158.

57 Plaza A, Garcia R, Bifulco G, Martinez JP, Hüttel S, Sasse F, Meyerhans A, Stadler M, Müller R: Aetheramides A and B, potent HIV-inhibitory depsipeptides from a myxobacterium of the new genus "Aetherobacter". Org Lett 2012;14:2854–2857.

58 Held J, Gebru T, Kalesse M, Jansen R, Gerth K, Müller R, Mordmüller B: Antimalarial activity of the myxobacterial macrolide chlorotonil a. Antimicrob Agents Chemother 2014;58:6378–6384.

59 Sasse F, Steinmetz H, Schupp T, Petersen F, Memmert K, Hofmann H, Heusser C, Brinkmann V, Matt P von, Höfle G, Reichenbach H: Argyrins, immunosuppressive cyclic peptides from myxobacteria. I. Production, isolation, physico-chemical and biological properties. J Antibiot 2002;55:543–551.

60 Schäberle TF, Schiefer A, Schmitz A, König GM, Hoerauf A, Pfarr K: Corallopyronin A - a promising antibiotic for treatment of filariasis. Int J Med Microbiol 2014;304:72–78.

61 Li X, Yu T-K, Kwak JH, Son B-Y, Seo Y, Zee O-P, Ahn J-W: Soraphinol C, a new free-radical scavenger from Sorangium cellulosum. J Microbiol Biotechnol 2008;18:520–522.

62 Forli S: Epothilones: From discovery to clinical trials. Curr Top Med Chem 2014;14:2312–2321.

63 Conlin A, Fornier M, Hudis C, Kar S, Kirkpatrick P: Ixabepilone. Nat Rev Drug Discov 2007;6:953–954.

64 White KN, Tenney K, Crews P: The Bengamides: A Mini-Review of Natural Sources, Analogues, Biological Properties, Biosynthetic Origins, and Future Prospects. J Nat Prod 2017;80:740–755.

65 Kaur G, Hollingshead M, Holbeck S, Schauer-Vukasinović V, Camalier RF, Dömling A, Agarwal S: Biological evaluation of tubulysin A: A potential anticancer and antiangiogenic natural product. Biochem J 2006;396:235–242.

66 Kellner F, Kim J, Clavijo BJ, Hamilton JP, Childs KL, Vaillancourt B, Cepela J, Habermann M, Steuernagel B, Clissold L, McLay K, Buell CR, O'Connor SE: Genome-guided investigation of plant natural product biosynthesis. Plant J 2015;82:680–692.

67 Hertweck C: The biosynthetic logic of polyketide diversity. Angew Chem Int Ed Engl 2009;48:4688–4716.

68 Fischbach MA, Walsh CT: Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. Chem Rev 2006;106:3468–3496.

69 Beld J, Sonnenschein EC, Vickery CR, Noel JP, Burkart MD: The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. Nat Prod Rep 2014;31:61–108.

70 Molnár I, Schupp T, Ono M, Zirkle RE, Milnamow M, Nowak-Thompson B, Engel N, Toupet C, Stratmann A, Cyr DD, Gorlach J, Mayo JM, Hu A, Goff S, Schmid J, Ligon JM: The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from Sorangium cellulosum So ce90. Chemistry & Biology 2000;7:97–109.

71 Chen H, O'Connor S, Cane DE, Walsh CT: Epothilone biosynthesis: assembly of the methylthiazolylcarboxy starter unit on the EpoB subunit. Chemistry & Biology 2001;8:899–912.

72 Stachelhaus T, Mootz HD, Marahiel MA: The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chemistry & Biology 1999;6:493–505.

73 Ibba M, Soll D: Aminoacyl-tRNA synthesis. Annu Rev Biochem 2000;69:617–650.

74 Walsh CT, O'Brien RV, Khosla C: Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. Angew Chem Int Ed Engl 2013;52:7098–7124.

75 Xu L, Huang H, Wei W, Zhong Y, Tang B, Yuan H, Zhu L, Huang W, Ge M, Yang S, Zheng H, Jiang W, Chen D, Zhao G-P, Zhao W: Complete genome sequence and comparative genomic analyses of the vancomycin-producing Amycolatopsis orientalis. BMC Genomics 2014;15:363.

76 Stachelhaus T, Walsh CT: Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. Biochemistry 2000;39:5775–5787.

77 Schneider TL, Shen B, Walsh CT: Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation. Biochemistry 2003;42:9722–9730.

78 Ansari MZ, Sharma J, Gokhale RS, Mohanty D: In silico analysis of methyltransferase domains involved in biosynthesis of secondary metabolites. BMC Bioinformatics 2008;9:454.

79 Neumann CS, Fujimori DG, Walsh CT: Halogenation strategies in natural product biosynthesis. Chemistry & Biology 2008;15:99–109.

80 Xiao Y, Li S, Niu S, Ma L, Zhang G, Zhang H, Zhang G, Ju J, Zhang C: Characterization of tiacumicin B biosynthetic gene cluster affording diversified tiacumicin analogues and revealing a tailoring dihalogenase. J Am Chem Soc 2011;133:1092–1105.

81 Koglin A, Löhr F, Bernhard F, Rogov VV, Frueh DP, Strieter ER, Mofid MR, Güntert P, Wagner G, Walsh CT, Marahiel MA, Dötsch V: Structural basis for the selectivity of the external thioesterase of the surfactin synthetase. Nature 2008;454:907–911.

82 Robbins T, Kapilivsky J, Cane DE, Khosla C: Roles of Conserved Active Site Residues in the Ketosynthase Domain of an Assembly Line Polyketide Synthase. Biochemistry 2016;55:4476–4484.

83 Reeves CD, Murli S, Ashley GW, Piagentini M, Hutchinson CR, McDaniel R: Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations. Biochemistry 2001;40:15464–15470.

84 Kallberg Y, Oppermann U, Jörnvall H, Persson B: Short-chain dehydrogenases/reductases (SDRs). Eur J Biochem 2002;269:4409–4417.

85 Reid R, Piagentini M, Rodriguez E, Ashley G, Viswanathan N, Carney J, Santi DV, Hutchinson CR, McDaniel R: A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. Biochemistry 2003;42:72–79.

86 Keatinge-Clay A: Crystal structure of the erythromycin polyketide synthase dehydratase. Journal of Molecular Biology 2008;384:941–953.

87 Kwan DH, Leadlay PF: Mutagenesis of a modular polyketide synthase enoylreductase domain reveals insights into catalysis and stereospecificity. ACS Chem Biol 2010;5:829–838.

88 Farmer R, Thomas CM, Winn PJ: Structure, function and dynamics in acyl carrier proteins. PLoS ONE 2019;14:e0219435.

89 Denoya CD, Fedechko RW, Hafner EW, McArthur HA, Morgenstern MR, Skinner DD, Stutzman-Engwall K, Wax RG, Wernau WC: A second branched-chain alpha-keto acid dehydrogenase gene cluster (bkdFGH) from Streptomyces avermitilis: its relationship to avermectin biosynthesis and the construction of a bkdF mutant suitable for the production of novel antiparasitic avermectins. J Bacteriol 1995;177:3504–3511.

90 Erb TJ, Berg IA, Brecht V, Müller M, Fuchs G, Alber BE: Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. Proc Natl Acad Sci U S A 2007;104:10631–10636.

91 Eustáquio AS, McGlinchey RP, Liu Y, Hazzard C, Beer LL, Florova G, Alhamadsheh MM, Lechner A, Kale AJ, Kobayashi Y, Reynolds KA, Moore BS: Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from S-adenosyl-L-methionine. Proc Natl Acad Sci U S A 2009;106:12295–12300.

92 Kaufmann E, Hattori H, Miyatake-Ondozabal H, Gademann K: Total Synthesis of the Glycosylated Macrolide Antibiotic Fidaxomicin. Org Lett 2015;17:3514–3517.

93 Kneidinger B, Graninger M, Adam G, Puchberger M, Kosma P, Zayni S, Messner P: Identification of two GDP-6-deoxy-D-lyxo-4-hexulose reductases synthesizing GDP-D-rhamnose in Aneurinibacillus thermoaerophilus L420-91T. J Biol Chem 2001;276:5577–5583.

94 Li Y-P, Yu P, Li J-F, Tang Y-L, Bu Q-T, Mao X-M, Li Y-Q: FadR1, a pathway-specific activator of fidaxomicin biosynthesis in Actinoplanes deccanensis Yp-1. Appl Microbiol Biotechnol 2019;103:7583–7596.

95 Gaucher GM, Shepherd MG: Isolation of orsellinic acid synthase. Biochemical and Biophysical Research Communications 1968;32:664–671.

96 Yu Z, Zhang H, Yuan C, Zhang Q, Khan I, Zhu Y, Zhang C: Characterizing Two Cytochrome P450s in Tiacumicin Biosynthesis Reveals Reaction Timing for Tailoring Modifications. Org Lett 2019;21:7679–7683.

97 Niu S, Hu T, Li S, Xiao Y, Ma L, Zhang G, Zhang H, Yang X, Ju J, Zhang C: Characterization of a sugar-O-methyltransferase TiaS5 affords new Tiacumicin analogues with improved antibacterial properties and reveals substrate promiscuity. ChemBioChem 2011;12:1740–1748.

98 Aron ZD, Dorrestein PC, Blackhall JR, Kelleher NL, Walsh CT: Characterization of a new tailoring domain in polyketide biogenesis: the amine transferase domain of MycA in the mycosubtilin gene cluster. J Am Chem Soc 2005;127:14986–14987.

99 Hertweck C, Luzhetskyy A, Rebets Y, Bechthold A: Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. Nat Prod Rep 2007;24:162–190.

100 Bisang C, Long PF, Cortés J, Westcott J, Crosby J, Matharu AL, Cox RJ, Simpson TJ, Staunton J, Leadlay PF: A chain initiation factor common to both modular and aromatic polyketide synthases. Nature 1999;401:502–505.

101 Pickens LB, Tang Y: Oxytetracycline biosynthesis. J Biol Chem 2010;285:27509–27515.

102 Petković H, Lukežič T, Šušković J: Biosynthesis of Oxytetracycline by Streptomyces rimosus: Past, Present and Future Directions in the Development of Tetracycline Antibiotics. Food Technology and Biotechnology 2017;55:3–13.

103 Zhu T, Cheng X, Liu Y, Deng Z, You D: Deciphering and engineering of the final step halogenase for improved chlortetracycline biosynthesis in industrial Streptomyces aureofaciens. Metab Eng 2013;19:69–78.

104 Lukežič T, Fayad AA, Bader C, Harmrolfs K, Bartuli J, Groß S, Lešnik U, Hennessen F, Herrmann J, Pikl Š, Petković H, Müller R: Engineering Atypical Tetracycline Formation in Amycolatopsis sulphurea for the Production of Modified Chelocardin Antibiotics. ACS Chem Biol 2019;14:468–477.

105 Letzel A-C, Pidot SJ, Hertweck C: Genome mining for ribosomally synthesized and post-translationally modified peptides (RiPPs) in anaerobic bacteria. BMC Genomics 2014;15:983.

106 Cotter PD, Ross RP, Hill C: Bacteriocins - a viable alternative to antibiotics? Nat Rev Microbiol 2013;11:95–105.

107 Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, Camarero JA, Campopiano DJ, Challis GL, Clardy J, Cotter PD, Craik DJ, Dawson M, Dittmann E, Donadio S, Dorrestein PC, Entian K-D, Fischbach MA, Garavelli JS, Göransson U, Gruber CW, Haft DH, Hemscheidt TK, Hertweck C, Hill C, Horswill AR, Jaspars M, Kelly WL, Klinman JP, Kuipers OP, Link AJ, Liu W, Marahiel MA, Mitchell DA, Moll GN, Moore BS, Müller R, Nair SK, Nes IF, Norris GE, Olivera BM, Onaka H, Patchett ML, Piel J, Reaney MJT, Rebuffat S, Ross RP, Sahl H-G, Schmidt EW, Selsted ME, Severinov K, Shen B, Sivonen K, Smith L, Stein T, Süssmuth RD, Tagg JR, Tang G-L, Truman AW, Vederas JC, Walsh CT, Walton JD, Wenzel SC, Willey JM, van der Donk WA: Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat Prod Rep 2013;30:108–160.

108 Imai Y, Meyer KJ, Iinishi A, Favre-Godal Q, Green R, Manuse S, Caboni M, Mori M, Niles S, Ghiglieri M, Honrao C, Ma X, Guo JJ, Makriyannis A, Linares-Otoya L, Böhringer N, Wuisan ZG, Kaur H, Wu R, Mateus A, Typas A, Savitski MM, Espinoza JL, O'Rourke A, Nelson KE, Hiller S, Noinaj N, Schäberle TF, D'Onofrio A, Lewis K: A new antibiotic selectively kills Gram-negative pathogens. Nature 2019;576:459–464.

109 Baumann T, Nickling JH, Bartholomae M, Buivydas A, Kuipers OP, Budisa N: Prospects of In vivo Incorporation of Non-canonical Amino Acids for the Chemical Diversification of Antimicrobial Peptides. Front Microbiol 2017;8:124.

110 Lopatniuk M, Myronovskyi M, Luzhetskyy A: Streptomyces albus: A New Cell Factory for Non-Canonical Amino Acids Incorporation into Ribosomally Synthesized Natural Products. ACS Chem Biol 2017;12:2362–2370.

111 Lakey JH, Lea EJ, Rudd BA, Wright HM, Hopwood DA: A new channel-forming antibiotic from Streptomyces coelicolor A3(2) which requires calcium for its activity. J Gen Microbiol 1983;129:3565–3573.

112 Rudd BA, Hopwood DA: A pigmented mycelial antibiotic in Streptomyces coelicolor: Control by a chromosomal gene cluster. J Gen Microbiol 1980;119:333–340.

113 Tsao SW, Rudd BA, He XG, Chang CJ, Floss HG: Identification of a red pigment from Streptomyces coelicolor A3(2) as a mixture of prodigiosin derivatives. J Antibiot 1985;38:128–131.

114 Wright LF, Hopwood DA: Actinorhodin is a chromosomally-determined antibiotic in Streptomyces coelicolar A3(2). J Gen Microbiol 1976;96:289–297.

115 Medema MH, Blin K, Cimermancic P, Jager V de, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R: antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res 2011;39:W339-46.

116 Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T: antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 2019;47:W81-W87.

117 Röttig M, Medema MH, Blin K, Weber T, Rausch C, Kohlbacher O: NRPSpredictor2--a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Res 2011;39:W362-7.

118 Yadav G, Gokhale RS, Mohanty D: Computational Approach for Prediction of Domain Organization and Substrate Specificity of Modular Polyketide Synthases. Journal of Molecular Biology 2003;328:335–363.

119 Agrawal P, Khater S, Gupta M, Sain N, Mohanty D: RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links. Nucleic Acids Res 2017;45:W80-W88.

120 van Heel AJ, Jong A de, Song C, Viel JH, Kok J, Kuipers OP: BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. Nucleic Acids Res 2018;46:W278-W281.

121 Boukhalfa H, Crumbliss AL: Chemical aspects of siderophore mediated iron transport. Biometals 2002;15:325–339.

122 Yu T-W, Shen Y, McDaniel R, Floss HG, Khosla C, Hopwood DA, Moore BS: Engineered Biosynthesis of Novel Polyketides from Streptomyces Spore Pigment Polyketide Synthases. J Am Chem Soc 1998;120:7749–7759.

123 Kling A, Lukat P, Almeida DV, Bauer A, Fontaine E, Sordello S, Zaburannyi N, Herrmann J, Wenzel SC, König C, Ammerman NC, Barrio MB, Borchers K, Bordon-Pallier F, Brönstrup M, Courtemanche G, Gerlitz M, Geslin M, Hammann P, Heinz DW, Hoffmann H, Klieber S, Kohlmann M, Kurz M, Lair C, Matter H, Nuermberger E, Tyagi S, Fraisse L, Grosset JH, Lagrange S, Müller R: Antibiotics. Targeting DnaN for tuberculosis therapy using novel griselimycins. Science 2015;348:1106–1112.

124 Skinnider MA, Merwin NJ, Johnston CW, Magarvey NA: PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. Nucleic Acids Res 2017;45:W49-W54.

125 Alanjary M, Kronmiller B, Adamek M, Blin K, Weber T, Huson D, Philmus B, Ziemert N: The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery. Nucleic Acids Res 2017;45:W42-W48.

126 Zerikly M, Challis GL: Strategies for the discovery of new natural products by genome mining. Chembiochem 2009;10:625–633.

127 Rachid S, Sasse F, Beyer S, Müller R: Identification of StiR, the first regulator of secondary metabolite formation in the myxobacterium Cystobacter fuscus Cb f17.1. J Biotechnol 2006;121:429–441.

128 Rachid S, Gerth K, Kochems I, Müller R: Deciphering regulatory mechanisms for secondary metabolite production in the myxobacterium Sorangium cellulosum So ce56. Mol Microbiol 2007;63:1783–1796.

129 Rachid S, Gerth K, Müller R: NtcA: a negative regulator of secondary metabolite biosynthesis in Sorangium cellulosum. J Biotechnol 2009;140:135–142.

130 Laureti L, Song L, Huang S, Corre C, Leblond P, Challis GL, Aigle B: Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in Streptomyces ambofaciens. Proc Natl Acad Sci U S A 2011;108:6258–6263.

131 Metsä-Ketelä M, Ylihonko K, Mäntsälä P: Partial activation of a silent angucycline-type gene cluster from a rubromycin beta producing Streptomyces sp. PGA64. J Antibiot 2004;57:502–510.

132 Sidda JD, Song L, Poon V, Al-Bassam M, Lazos O, Buttner MJ, Challis GL, Corre C: Discovery of a family of γ-aminobutyrate ureas via rational derepression of a silent bacterial gene cluster. Chem. Sci. 2014;5:86–89.

133 Zaburannyi N, Rabyk M, Ostash B, Fedorenko V, Luzhetskyy A: Insights into naturally minimised Streptomyces albus J1074 genome. BMC Genomics 2014;15:97.

134 Baltz RH: Genetic manipulation of secondary metabolite biosynthesis for improved production in Streptomyces and other actinomycetes. J Ind Microbiol Biotechnol 2016;43:343–370.

135 Kang W-J, Pan H-X, Wang S, Yu B, Hua H, Tang G-L: Identification of the Amipurimycin Gene Cluster Yields Insight into the Biosynthesis of C9 Sugar Nucleoside Antibiotics. Org Lett 2019;21:3148–3152.

136 Kawata J, Naoe T, Ogasawara Y, Dairi T: Biosynthesis of the Carbonylmethylene Structure Found in the Ketomemicin Class of Pseudotripeptides. Angew Chem Int Ed Engl 2017;56:2026–2029.

137 Rausch C, Hoof I, Weber T, Wohlleben W, Huson DH: Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. BMC Evol Biol 2007;7:78.

138 Malcangi A, Trione, Guido: Procedure for the production of tiacumicin b, 2013.

139 Vogeley L, El Arnaout T, Bailey J, Stansfeld PJ, Boland C, Caffrey M: Structural basis of lipoprotein signal peptidase II action and inhibition by the antibiotic globomycin. Science 2016;351:876–880.

140 Olatunji S, Yu X, Bailey J, Huang C-Y, Zapotoczna M, Bowen K, Remškar M, Müller R, Scanlan EM, Geoghegan JA, Olieric V, Caffrey M: Structures of lipoprotein signal peptidase II from Staphylococcus aureus complexed with antibiotics globomycin and myxovirescin. Nat Commun 2020;11:140.

141 Erb W, Zhu J: From natural product to marketed drug: The tiacumicin odyssey. Nat Prod Rep 2013;30:161–174.

142 Venugopal AA, Johnson S: Fidaxomicin: A novel macrocyclic antibiotic approved for treatment of Clostridium difficile infection. Clin Infect Dis 2012;54:568–574.

143 Kurabachew M, Lu SHJ, Krastel P, Schmitt EK, Suresh BL, Goh A, Knox JE, Ma NL, Jiricek J, Beer D, Cynamon M, Petersen F, Dartois V, Keller T, Dick T, Sambandamurthy VK: Lipiarmycin targets RNA polymerase and has good activity against multidrug-resistant strains of Mycobacterium tuberculosis. J Antimicrob Chemother 2008;62:713–719.

144 Wu M-C, Huang C-C, Lu Y-C, Fan W-J: Derivatives of tiacumicin B as anti-cancer agents, 2009.

145 Jamison MT: Mangrolide A, a novel marine derived polyketide with selective antibiotic activity. PhD Thesis. Dallas, 2013.

146 Glaus F, Altmann K-H: Total synthesis of the tiacumicin B (lipiarmycin A3/fidaxomicin) aglycone. Angew Chem Int Ed Engl 2015;54:1937–1940.

147 Molnar V, Matković Z, Tambić T, Kozma C: Klinicko-farmakolosko ispitivanje kelokardina u bolesnika s infekcijom mokraćnih putova. Lijec Vjesn 1977;99:560–562.

148 Pickens LB, Kim W, Wang P, Zhou H, Watanabe K, Gomi S, Tang Y: Biochemical analysis of the biosynthetic pathway of an anticancer tetracycline SF2575. J Am Chem Soc 2009;131:17677–17689.

149 Gerth K, Pradella S, Perlova O, Beyer S, Müller R: Myxobacteria: proficient producers of novel natural products with various biological activities--past and future biotechnological aspects with the focus on the genus Sorangium. J Biotechnol 2003;106:233–253.

150 Gaudêncio SP, Pereira F: Dereplication: racing to speed up the natural products discovery process. Nat Prod Rep 2015;32:779–810.

151 Hubert J, Nuzillard J-M, Renault J-H: Dereplication strategies in natural product research: How many tools and methodologies behind the same concept? Phytochem Rev 2017;16:55–95.

152 Guthals A, Watrous JD, Dorrestein PC, Bandeira N: The spectral networks paradigm in high throughput mass spectrometry. Mol Biosyst 2012;8:2535–2544.

153 Yang JY, Sanchez LM, Rath CM, Liu X, Boudreau PD, Bruns N, Glukhov E, Wodtke A, Felicio R de, Fenner A, Wong WR, Linington RG, Zhang L, Debonsi HM, Gerwick WH, Dorrestein PC: Molecular networking as a dereplication strategy. J Nat Prod 2013;76:1686–1699.

154 Watrous J, Roach P, Alexandrov T, Heath BS, Yang JY, Kersten RD, van der Voort M, Pogliano K, Gross H, Raaijmakers JM, Moore BS, Laskin J, Bandeira N, Dorrestein PC:

Mass spectral molecular networking of living microbial colonies. Proc Natl Acad Sci U S A 2012;109:E1743-52.

155 Kildgaard S, Subko K, Phillips E, Goidts V, La Cruz M de, Díaz C, Gotfredsen CH, Andersen B, Frisvad JC, Nielsen KF, Larsen TO: A Dereplication and Bioguided Discovery Approach to Reveal New Compounds from a Marine-Derived Fungus Stilbella fimetaria. Mar Drugs 2017;15.

156 Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu W-T, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu C-C, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw C-C, Yang Y-L, Humpf H-U, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, P CAB, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJN, Granatosky EA, Glukhov E, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard P-M, Phapale P, Nothias L-F, Alexandrov T, Litaudon M, Wolfender J-L, Kyle JE, Metz TO, Peryea T, Nguyen D-T, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson BO, Pogliano K, Linington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N: Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol 2016;34:828–837.

157 Ochi K, Hosaka T: New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters. Appl Microbiol Biotechnol 2013;97:87–98.

158 Panter F, Krug D, Baumann S, Müller R: Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. Chem. Sci. 2018;9:4898–4908.

159 Cortina NS, Krug D, Plaza A, Revermann O, Müller R: Myxoprincomide: a natural product from Myxococcus xanthus discovered by comprehensive analysis of the secondary metabolome. Angew Chem Int Ed Engl 2012;51:811–816.

160 Sambrook J, Russell DW: Molecular cloning: A laboratory manual / Joseph Sambrook, David W. Russell. Vol. 3, 3rd ed. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press, 2001.

161 Kashefi K, Hartzell PL: Genetic suppression and phenotypic masking of a Myxococcus xanthus frzF- defect. Mol Microbiol 1995;15:483–494.

162 Gerber NN, Lechevalier HA: Geosmin, an Earthy-Smelling Substance Isolated from Actinomycetes. Appl Microbiol 1965;13:935–938.

163 Liang M-H, Zhu J, Jiang J-G: Carotenoids biosynthesis and cleavage related genes from bacteria to plants. Crit Rev Food Sci Nutr 2018;58:2314–2333.

164 Hug JJ, Panter F, Krug D, Müller R: Genome mining reveals uncommon alkylpyrones as type III PKS products from myxobacteria. J Ind Microbiol Biotechnol 2019;46:319–334.

165 Kunze B, Bedorf N, Kohl W, Höfle G, Reichenbach H: Myxochelin A, a new iron-chelating compound from Angiococcus disciformis (Myxobacterales). Production, isolation, physico-chemical and biological properties. J. Antibiot. 1989;42:14–17.

166 Ring MW, Schwär G, Thiel V, Dickschat JS, Kroppenstedt RM, Schulz S, Bode HB: Novel iso-branched ether lipids as specific markers of developmental sporulation in the myxobacterium Myxococcus xanthus. J Biol Chem 2006;281:36691–36700.

167 Gorges J, Panter F, Kjaerulff L, Hoffmann T, Kazmaier U, Müller R: Structure, Total Synthesis, and Biosynthesis of Chloromyxamides: Myxobacterial Tetrapeptides Featuring an Uncommon 6-Chloromethyl-5-methoxypipecolic Acid Building Block. Angew Chem Int Ed Engl 2018;57:14270–14275.

168 Vetting MW, Hegde SS, Fajardo JE, Fiser A, Roderick SL, Takiff HE, Blanchard JS: Pentapeptide repeat proteins. Biochemistry 2006;45:1–10.

169 Burgos L, Lehmann M, Simon D, Andrade HHR de, Abreu BRR de, Nabinger DD, Grivicich I, Juliano VB, Dihl RR: Agents of earthy-musty taste and odor in water: evaluation of cytotoxicity, genotoxicity and toxicogenomics. Sci Total Environ 2014;490:679–685.

170 Yang H, Perrier J, Whitford PC: Disorder guides domain rearrangement in elongation factor Tu. Proteins 2018;86:1037–1046.

171 Kraal B, Zeef LA, Mesters JR, Boon K, Vorstenbosch EL, Bosch L, Anborgh PH, Parmeggiani A, Hilgenfeld R: Antibiotic resistance mechanisms of mutant EF-Tu species in Escherichia coli. Biochem Cell Biol 1995;73:1167–1177.

172 Prezioso SM, Brown NE, Goldberg JB: Elfamycins: inhibitors of elongation factor-Tu. Mol Microbiol 2017;106:22–34.

173 Wolf H, Chinali G, Parmeggiani A: Mechanism of the inhibition of protein synthesis by kirromycin. Role of elongation factor Tu and ribosomes. Eur J Biochem 1977;75:67–75.

174 Cetin R, Krab IM, Anborgh PH, Cool RH, Watanabe T, Sugiyama T, Izaki K, Parmeggiani A: Enacyloxin IIa, an inhibitor of protein biosynthesis that acts on elongation factor Tu and the ribosome. The EMBO Journal 1996;15:2604–2611.

175 Parmeggiani A, Krab IM, Okamura S, Nielsen RC, Nyborg J, Nissen P: Structural basis of the action of pulvomycin and GE2270 A on elongation factor Tu. Biochemistry 2006;45:6846–6857.

176 Thaker MN, García M, Koteva K, Waglechner N, Sorensen D, Medina R, Wright GD: Biosynthetic gene cluster and antimicrobial activity of the elfamycin antibiotic factumycin. Med. Chem. Commun. 2012;3:1020.

177 Flinspach K, Kapitzke C, Tocchetti A, Sosio M, Apel AK: Heterologous expression of the thiopeptide antibiotic GE2270 from Planobispora rosea ATCC 53733 in Streptomyces coelicolor requires deletion of ribosomal genes from the expression construct. PLoS ONE 2014;9:e90499.

178 Popoff A: Exploiting the biosynthetic potential of myxobacteria for natural product discovery. Doctoral Thesis. Saarbrücken, 2020.

179 Olano C, Moss SJ, Braña AF, Sheridan RM, Math V, Weston AJ, Méndez C, Leadlay PF, Wilkinson B, Salas JA: Biosynthesis of the angiogenesis inhibitor borrelidin by Streptomyces parvulus Tü4055: insights into nitrile formation†. Mol Microbiol 2004;52:1745–1756.

180 Konishi M, Nishio M, Saitoh K, Miyaki T, Oki T, Kawaguchi H: Cispentacin, a new antifungal antibiotic. I. Production, isolation, physico-chemical properties and structure. J. Antibiot. 1989;42:1749–1755.

181 Romo AJ, Shiraishi T, Ikeuchi H, Lin G-M, Geng Y, Lee Y-H, Liem PH, Ma T, Ogasawara Y, Shin-Ya K, Nishiyama M, Kuzuyama T, Liu H-W: The Amipurimycin and Miharamycin Biosynthetic Gene Clusters: Unraveling the Origins of 2-Aminopurinyl Peptidyl Nucleoside Antibiotics. J Am Chem Soc 2019;141:14152–14159.

182 Kawabata K, Inamoto Y, Sakane K, Iwamoto T, Hashimoto S: Synthesis and structure determination of FR109615, a new antifungal antibiotic. J. Antibiot. 1990;43:513–518.

183 Goto T, Toya Y, Ohgi T, Kondo T: Structure of amipurimycin, a nucleoside antibiotic having a novel branched sugar moiety. Tetrahedron Letters 1982;23:1271–1274.

184 Harada S, Kishi T: Isolation and characterization of a new nucleoside antibiotic, amipurimycin. J. Antibiot. 1977;30:11–16.

185 Iwasa T, Kishi T, Matsuura K, Wakae O: Streptomyces novoguineensis sp. Nov., an amipurimycin producer, and antimicrobial activity of amipurimycin. J. Antibiot. 1977;30:1–10.

186 Rangaswamy V, Jiralerspong S, Parry R, Bender CL: Biosynthesis of the Pseudomonas polyketide coronafacic acid requires monofunctional and multifunctional polyketide synthase proteins. Proc Natl Acad Sci U S A 1998;95:15469–15474.

187 Bekal S, van Beeumen J, Samyn B, Garmyn D, Henini S, Diviès C, Prévost H: Purification of Leuconostoc mesenteroides Citrate Lyase and Cloning and Characterization of the citCDEFG Gene Cluster. J Bacteriol 1998;180:647–654.

188 Herter S, Busch A, Fuchs G: L-Malyl-coenzyme A lyase/beta-methylmalyl-coenzyme A lyase from Chloroflexus aurantiacus, a bifunctional enzyme involved in autotrophic CO(2) fixation. J Bacteriol 2002;184:5999–6006.

189 Dillon SC, Bateman A: The Hotdog fold: wrapping up a superfamily of thioesterases and dehydratases. BMC Bioinformatics 2004;5:109.

190 Nshimiyimana P, Liu L, Du G: Engineering of L-amino acid deaminases for the production of α-keto acids from L-amino acids. Bioengineered 2019;10:43–51.

191 Noike M, Matsui T, Ooya K, Sasaki I, Ohtaki S, Hamano Y, Maruyama C, Ishikawa J, Satoh Y, Ito H, Morita H, Dairi T: A peptide ligase and the ribosome cooperate to synthesize the peptide pheganomycin. Nat Chem Biol 2015;11:71–76.

192 Ogasawara Y, Kawata J, Noike M, Satoh Y, Furihata K, Dairi T: Exploring Peptide Ligase Orthologs in Actinobacteria-Discovery of Pseudopeptide Natural Products, Ketomemicins. ACS Chem Biol 2016;11:1686–1692.

193 Hao C, Huang S, Deng Z, Zhao C, Yu Y: Mining of the pyrrolamide antibiotics analogs in Streptomyces netropsis reveals the amidohydrolase-dependent "iterative strategy" underlying the pyrrole polymerization. PLoS ONE 2014;9:e99077.

194 Lautru S, Song L, Demange L, Lombès T, Galons H, Challis GL, Pernodet J-L: A sweet origin for the key congocidine precursor 4-acetamidopyrrole-2-carboxylate. Angew Chem Int Ed Engl 2012;51:7454–7458.

195 Clancy J, Petitpas J, Dib-Hajj F, Yuan W, Cronan M, Kamath AV, Bergeron J, Retsema JA: Molecular cloning and functional analysis of a novel macrolide-resistance determinant, mefA, from Streptococcus pyogenes. Mol Microbiol 1996;22:867–879.

196 Lewinson O, Adler J, Sigal N, Bibi E: Promiscuity in multidrug recognition and transport: the bacterial MFS Mdr transporters. Mol Microbiol 2006;61:277–284.

197 Dlakić M: Functionally unrelated signalling proteins contain a fold similar to Mg2+-dependent endonucleases. Trends in Biochemical Sciences 2000;25:272–273.

198 Ochsner UA, Sun X, Jarvis T, Critchley I, Janjic N: Aminoacyl-tRNA synthetases: essential and still promising targets for new anti-infective agents. Expert Opin Investig Drugs 2007;16:573–593.

199 Capobianco JO, Zakula D, Coen ML, Goldman RC: Anti-Candida activity of cispentacin: the active transport by amino acid permeases and possible mechanisms of action. Biochemical and Biophysical Research Communications 1993;190:1037–1044.

200 Yeates C: Icofungipen (PLIVA). Curr Opin Investig Drugs 2005;6:838–844.

201 Ziegelbauer K, Babczinski P, Schönfeld W: Molecular Mode of Action of the Antifungal β-Amino Acid BAY 10-8888. Antimicrob Agents Chemother 1998;42:2197–2205.

202 Umezawa H, Aoyagi T, Ohuchi S, Okuyama A, Suda H, Takita T, Hamada M, Takeuchi T: Arphamenines A and B, new inhibitors of aminopeptidase B, produced by bacteria. J. Antibiot. 1983;36:1572–1575.

203 Bode HB, Dickschat JS, Kroppenstedt RM, Schulz S, Müller R: Biosynthesis of iso-fatty acids in myxobacteria: iso-even fatty acids are derived by alpha-oxidation from iso-odd fatty acids. J Am Chem Soc 2005;127:532–533.

204 Nakashima Y, Mori T, Nakamura H, Awakawa T, Hoshino S, Senda M, Senda T, Abe I: Structure function and engineering of multifunctional non-heme iron dependent oxygenases in fungal meroterpenoid biosynthesis. Nat Commun 2018;9:104.

205 Yamashita A, Kato H, Wakatsuki S, Tomizaki T, Nakatsu T, Nakajima K, Hashimoto T, Yamada Y, Oda J: TROPINONE REDUCTASE-II COMPLEXED WITH NADP+ AND PSEUDOTROPINE, 1999.

206 Quadri LE, Sello J, Keating TA, Weinreb PH, Walsh CT: Identification of a Mycobacterium tuberculosis gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. Chemistry & Biology 1998;5:631–645.

207 Huijbers MME, Montersino S, Westphal AH, Tischler D, van Berkel WJH: Flavin dependent monooxygenases. Arch Biochem Biophys 2014;544:2–17.

208 Eschenfeldt WH, Zhang Y, Samaha H, Stols L, Eirich LD, Wilson CR, Donnelly MI: Transformation of fatty acids catalyzed by cytochrome P450 monooxygenase enzymes of Candida tropicalis. Appl Environ Microbiol 2003;69:5992–5999.

209 Walsh CT, Chen H, Keating TA, Hubbard BK, Losey HC, Luo L, Marshall CG, Miller DA, Patel HM: Tailoring enzymes that modify nonribosomal peptides during and after chain

elongation on NRPS assembly lines. Current Opinion in Chemical Biology 2001;5:525–534.

210 Tien TM, Gaskins MH, Hubbell DH: Plant Growth Substances Produced by Azospirillum brasilense and Their Effect on the Growth of Pearl Millet (Pennisetum americanum L.). Appl Environ Microbiol 1979;37:1016–1024.

211 Yasuta T, Okazaki S, Mitsui H, Yuhashi K, Ezura H, Minamisawa K: DNA sequence and mutational analysis of rhizobitoxine biosynthesis genes in Bradyrhizobium elkanii. Appl Environ Microbiol 2001;67:4999–5009.

212 Berkowitz DB, Charette BD, Karukurichi KR, McFadden JM: α-Vinylic Amino Acids: Occurrence, Asymmetric Synthesis and Biochemical Mechanisms. Tetrahedron Asymmetry 2006;17:869–882.

213 Fernández M, Cuadrado Y, Aparicio JF, Martín JF: Role of homoserine and threonine pathway intermediates as precursors for the biosynthesis of aminoethoxyvinylglycine in Streptomyces sp. NRRL 5331. Microbiology (Reading, Engl ) 2004;150:1467–1474.

214 Patteson JB, Dunn ZD, Li B: In Vitro Biosynthesis of the Nonproteinogenic Amino Acid Methoxyvinylglycine. Angew Chem Int Ed Engl 2018;57:6780–6785.

215 Wieczorek P, Sacha P, Hauschild T, Zórawski M, Krawczyk M, Tryniszewska E: Multidrug resistant Acinetobacter baumannii--the role of AdeABC (RND family) efflux pump in resistance to antibiotics. Folia Histochem Cytobiol 2008;46:257–267.

216 Venäläinen JI, Juvonen RO, Männistö PT: Evolutionary relationships of the prolyl oligopeptidase family enzymes. Eur J Biochem 2004;271:2705–2715.

217 Barski OA, Tipparaju SM, Bhatnagar A: The aldo-keto reductase superfamily and its role in drug metabolism and detoxification. Drug Metab Rev 2008;40:553–624.

218 Kang L-W, Gabelli SB, Bianchet MA, Xu WL, Bessman MJ, Amzel LM: Structure of a coenzyme A pyrophosphatase from Deinococcus radiodurans: a member of the Nudix family. J Bacteriol 2003;185:4110–4118.

219 Chang H-Y, Chou C-C, Hsu M-F, Wang AHJ: Proposed carrier lipid-binding site of undecaprenyl pyrophosphate phosphatase from Escherichia coli. J Biol Chem 2014;289:18719–18735.

220 Bernard R, El Ghachi M, Mengin-Lecreulx D, Chippaux M, Denizot F: BcrC from Bacillus subtilis acts as an undecaprenyl pyrophosphate phosphatase in bacitracin resistance. J Biol Chem 2005;280:28852–28857.

221 Tareq FS, Kim JH, Lee MA, Lee H-S, Lee Y-J, Lee JS, Shin HJ: Ieodoglucomides A and B from a marine-derived bacterium Bacillus licheniformis. Org Lett 2012;14:1464–1467.

222 Tareq FS, Lee H-S, Lee Y-J, Lee JS, Shin HJ: Ieodoglucomide C and Ieodoglycolipid, New Glycolipids from a Marine-Derived Bacterium Bacillus licheniformis 09IDYM23. Lipids 2015;50:513–519.

223 Tanovic A, Samel SA, Essen L-O, Marahiel MA: Structure of surfactin A synthetase C (SrfA-C), a nonribosomal peptide synthetase termination module, 2008.

224 Stachelhaus T, Mootz HD, Bergendahl V, Marahiel MA: Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. J Biol Chem 1998;273:22773–22781.

225 Weber T, Baumgartner R, Renner C, Marahiel MA, Holak TA: Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. Structure 2000;8:407–418.

226 Keatinge-Clay AT: Crystal Structure of the Erythromycin Dehydratase, 2008.

227 Kiho T, Nakayama M, Yasuda K, Miyakoshi S, INUKAI M, Kogen H: Structure-activity relationships of globomycin analogues as antibiotics. Bioorg Med Chem 2004;12:337–361.

228 INUKAI M, ENOKITA R, TORIKATA A, NAKAHARA M, IWADO S, ARAI M: Globomycin, a new peptide antibiotic with spheroplast-forming activity. I. Taxonomy of producing organisms and fermentation. J. Antibiot. 1978;31:410–420.

229 Omoto S, Suzuki H, Inouye S: Isolation and structure of SF-1902 A5, a new globomycin analogue. J. Antibiot. 1979;32:83–86.

230 Brodasky, Thomas F., Stroman, David W.: Antibiotic compound and process for recovery thereof from a fermentation broth, 1986.

231 Nakajima M, Inukai M, Haneishi T, Terahara A, Arai M, Kinoshita T, Tamura C: Globomycin, a new peptide antibiotic with spheroplast-forming activity. III. Structural determination of globomycin. J. Antibiot. 1978;31:426–432.

232 Kogen H, Kiho T, Nakayama M, Furukawa Y, Kinoshita T, INUKAI M: Crystal Structure and Total Synthesis of Globomycin: Establishment of Relative and Absolute Configurations. J Am Chem Soc 2000;122:10214–10215.

233 Inukai M, Takeuchi M, Shimizu K, Arai M: Mechanism of action of globomycin. J. Antibiot. 1978;31:1203–1205.

234 Simunovic V, Zapp J, Rachid S, Krug D, Meiser P, Müller R: Myxovirescin A biosynthesis is directed by hybrid polyketide synthases/nonribosomal peptide synthetase, 3-hydroxy-3-methylglutaryl-CoA synthases, and trans-acting acyltransferases. Chembiochem 2006;7:1206–1220.

235 Omoto S, Ogino H, Inouye S: Studies on SF=1902 A2 A5, minor components of SF-1902 (globomycin). J. Antibiot. 1981;34:1416–1423.

236 Brodasky TF, Stroman DW, Dietz A, Mizsak S: U-56,407, a new antibiotic related to asukamycin: isolation and characterization. J. Antibiot. 1983;36:950–956.

237 Inukai M, Nakajima M, Osawa M, Haneishi T, Arai M: Globomycin, a new peptide antibiotic with spheroplast-forming activity. II. Isolation and physico-chemical and biological characterization. J. Antibiot. 1978;31:421–425.

238 Brotzel F, Mayr H: Nucleophilicities of amino acids and peptides. Org Biomol Chem 2007;5:3814–3820.

239 Smith S: The animal fatty acid synthase: one gene, one polypeptide, seven enzymes. FASEB j. 1994;8:1248–1259.

240 Scaglione A, Fullone MR, Montemiglio LC, Parisi G, Zamparelli C, Vallone B, Savino C, Grgurina I: Structure of the adenylation domain Thr1 involved in the biosynthesis of 4-chlorothreonine in Streptomyces sp. OH-5093-protein flexibility and molecular bases of substrate specificity. FEBS J 2017;284:2981–2999.

241 Yeh E, Kohli RM, Bruner SD, Walsh CT: Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-S-enzyme that cannot be elongated. Chembiochem 2004;5:1290–1293.

242 Wenzel SC, Meiser P, Binz TM, Mahmud T, Müller R: Nonribosomal peptide biosynthesis: point mutations and module skipping lead to chemical diversity. Angew Chem Int Ed Engl 2006;45:2296–2301.

243 Li Q, Qin X, Liu J, Gui C, Wang B, Li J, Ju J: Deciphering the Biosynthetic Origin of L-allo-Isoleucine. J Am Chem Soc 2016;138:408–415.

244 Liu JQ, Dairi T, Kataoka M, Shimizu S, Yamada H: L-allo-threonine aldolase from Aeromonas jandaei DK-39: gene cloning, nucleotide sequencing, and identification of the pyridoxal 5'-phosphate-binding lysine residue by site-directed mutagenesis. J Bacteriol 1997;179:3555–3560.

245 Xiao Y, Wall D: Genetic redundancy, proximity, and functionality of lspA, the target of antibiotic TA, in the Myxococcus xanthus producer strain. J Bacteriol 2014;196:1174–1183.

246 Prágai Z, Tjalsma H, Bolhuis A, van Dijl JM, Venema G, Bron S: The signal peptidase II (Isp) gene of Bacillus subtilis. Microbiology (Reading, Engl ) 1997;143 (Pt 4):1327–1333.

247 Deng W, Li C, Xie J: The underling mechanism of bacterial TetR/AcrR family transcriptional repressors. Cell Signal 2013;25:1608–1613.

248 Tramonti A, Nardella C, Di Salvo ML, Pascarella S, Contestabile R: The MocR-like transcription factors: pyridoxal 5'-phosphate-dependent regulators of bacterial metabolism. FEBS J 2018;285:3925–3944.

249 Suvorova IA, Korostelev YD, Gelfand MS: GntR Family of Bacterial Transcription Factors and Their DNA Binding Motifs: Structure, Positioning and Co-Evolution. PLoS ONE 2015;10:e0132618.

250 Rigali S, Derouaux A, Giannotta F, Dusart J: Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. J Biol Chem 2002;277:12507–12515.

251 Ryding NJ, Kelemen GH, Whatling CA, Flärdh K, Buttner MJ, Chater KF: A developmentally regulated gene encoding a repressor-like protein is essential for sporulation in Streptomyces coelicolor A3(2). Mol Microbiol 1998;29:343–357.

252 Hoskisson PA, Rigali S, Fowler K, Findlay KC, Buttner MJ: DevA, a GntR-like transcriptional regulator required for development in Streptomyces coelicolor. J Bacteriol 2006;188:5014–5023.

253 Magarvey N, He J, Aidoo KA, Vining LC: The pdx genetic marker adjacent to the chloramphenicol biosynthesis gene cluster in Streptomyces venezuelae ISP5230: functional characterization. Microbiology (Reading, Engl ) 2001;147:2103–2112.

254 Bankapalli K, Saladi S, Awadia SS, Goswami AV, Samaddar M, D'Silva P: Robust glyoxalase activity of Hsp31, a ThiJ/DJ-1/PfpI family member protein, is critical for oxidative stress resistance in Saccharomyces cerevisiae. J Biol Chem 2015;290:26491–26507.

255 Kawalek A, Wawrzyniak P, Bartosik AA, Jagura-Burdzy G: Rules and Exceptions: The Role of Chromosomal ParB in DNA Segregation and Other Cellular Processes. Microorganisms 2020;8.

256 Bird LE, Pan H, Soultanas P, Wigley DB: Mapping protein-protein interactions within a stable complex of DNA primase and DnaB helicase from Bacillus stearothermophilus. Biochemistry 2000;39:171–182.

257 Heintz NH: Transcription factors and the control of DNA replication. Current Opinion in Cell Biology 1992;4:459–467.

258 Ueda K, Umeyama T, BEPPU T, Horinouchi S: The aerial mycelium-defective phenotype of Streptomyces griseus resulting from A-factor deficiency is suppressed by a Ser/Thr kinase of S. coelicolor A3(2). Gene 1996;169:91–95.

259 Holmquist M: Alpha/Beta-hydrolase fold enzymes: structures, functions and mechanisms. Curr Protein Pept Sci 2000;1:209–235.

260 Tang W, Guo Z, Cao Z, Wang M, Li P, Meng X, Zhao X, Xie Z, Wang W, Zhou A, Lou C, Chen Y: d-Sedoheptulose-7-phosphate is a common precursor for the heptoses of septacidin and hygromycin B. Proc Natl Acad Sci U S A 2018;115:2818–2823.

261 Bierman M, Logan R, O'Brien K, Seno ET, Nagaraja Rao R, Schoner BE: Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp. Gene 1992;116:43–49.

262 MacNeil DJ, Gewain KM, Ruby CL, Dezeny G, Gibbons PH, MacNeil T: Analysis of Streptomyces avermitilis genes required for avermectin biosynthesis utilizing a novel integration vector. Gene 1992;111:61–68.

263 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A: Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 2012;28:1647–1649.

264 Sergio S, Pirali G, White R, Parenti F: Lipiarmycin, a new antibiotic from Actinoplanes III. Mechanism of action. J Antibiot 1975;28:543–549.

265 Arnone A, Nasini G, Cavalleri B: Structure elucidation of the macrocyclic antibiotic lipiarmycin. J. Chem. Soc., Perkin Trans. 1 1987:1353.

266 Butler MS, Robertson AAB, Cooper MA: Natural product and natural product derived drugs in clinical trials. Nat Prod Rep 2014;31:1612–1661.

267 Lin W, Das K, Degen D, Mazumder A, Duchi D, Wang D, Ebright YW, Ebright RY, Sineva E, Gigliotti M, Srivastava A, Mandal S, Jiang Y, Liu Y, Yin R, Zhang Z, Eng ET, Thomas D, Donadio S, Zhang H, Zhang C, Kapanidis AN, Ebright RH: Structural Basis of Transcription Inhibition by Fidaxomicin (Lipiarmycin A3). Mol Cell 2018;70:60-71.e15.

268 Goldstein EJC, Babakhani F, Citron DM: Antimicrobial activities of fidaxomicin. Clin Infect Dis 2012;55 Suppl 2:S143-8.

269 Omura S, Imamura N, Oiwa R, Kuga H, Iwata R, Masuma R, Iwai Y: Clostomicins, new antibiotics produced by Micromonospora echinospora subsp. armeniaca subsp. nov. I. Production, isolation, and physico-chemical and biological properties. J Antibiot 1986;39:1407–1412.

270 Hochlowski JE, Swanson SJ, Ranfranz LM, Whittern DN, Buko AM, McAlpine JB: Tiacumicins, a novel complex of 18-membered macrolides. II. Isolation and structure determination. J Antibiot 1987;40:575–588.

271 Hattori H, Roesslein J, Caspers P, Zerbe K, Miyatake-Ondozabal H, Ritz D, Rueedi G, Gademann K: Total Synthesis and Biological Evaluation of the Glycosylated Macrocyclic Antibiotic Mangrolide A. Angew Chem Int Ed Engl 2018;57:11020–11024.

272 Miyatake-Ondozabal H, Kaufmann E, Gademann K: Total synthesis of the protected aglycon of fidaxomicin (tiacumicin B, lipiarmycin A3). Angew Chem Int Ed Engl 2015;54:1933–1936.

273 DSMZ GmbH: Microorganisms: 609. OATMEAL AGAR (A) (ISP-3). Germany, DSMZ GmbH, 2007. https://www.dsmz.de/microorganisms/medium/pdf/DSMZ_Medium609.pdf (accessed February 12, 2020).

274 Wink MJ: Compendium of Actinobacteria: Actinoplanes deccanensis. Braunschweig, Germany, HZI - Helmholtz-Zentrum für Infektionsforschung GmbH, 2012. https://www.dsmz.de/microorganisms/wink_pdf/DSM43806.pdf (accessed January 14, 2020).

275 Coronelli C, White RJ, Lancini GC, Parenti F: Lipiarmycin, a new antibiotic from Actinoplanes. II. Isolation, chemical, biological and biochemical characterization. J Antibiot 1975;28:253–259.

276 Heinzelmann E, Berger S, Puk O, Reichenstein B, Wohlleben W, Schwartz D: A glutamate mutase is involved in the biosynthesis of the lipopeptide antibiotic friulimicin in Actinoplanes friuliensis. Antimicrob Agents Chemother 2003;47:447–457.

277 Oliver TJ, Prokop JF, Bower RR, Otto RH: Chelocardin, a new broad-spectrum antibiotic: I. Discovery and biological properties. Antimicrob Agents Chemother 1962:583–591.

278 Ames BD, Lee M-Y, Moody C, Zhang W, Tang Y, Tsai S-C: Structural and biochemical characterization of ZhuI aromatase/cyclase from the R1128 polyketide pathway. Biochemistry 2011;50:8392–8406.

279 Hopwood DA: Genetic Contributions to Understanding Polyketide Synthases. Chem Rev 1997;97:2465–2498.

280 Ames BD, Korman TP, Zhang W, Smith P, Vu T, Tang Y, Tsai S-C: Crystal structure and functional analysis of tetracenomycin ARO/CYC: implications for cyclization specificity of aromatic polyketides. Proc Natl Acad Sci U S A 2008;105:5349–5354.

281 Caldara-Festin G, Jackson DR, Barajas JF, Valentic TR, Patel AB, Aguilar S, Nguyen M, Vo M, Khanna A, Sasaki E, Liu H-W, Tsai S-C: Structural and functional analysis of two di-domain aromatase/cyclases from type II polyketide synthases. Proc Natl Acad Sci U S A 2015;112:E6844-51.

282 Menéndez N, Nur-e-Alam M, Braña AF, Rohr J, Salas JA, Méndez C: Biosynthesis of the antitumor chromomycin A3 in Streptomyces griseus: analysis of the gene cluster and rational design of novel chromomycin analogs. Chemistry & Biology 2004;11:21–32.

283 Rawlings BJ: Biosynthesis of polyketides (other than actinomycete macrolides). Nat Prod Rep 1999;16:425–484.

284 Tsai S-C, Ames BD: Chapter 2 Structural Enzymology of Polyketide Synthases; in Hopwood DA (ed): Complex enzymes in microbial natural product biosynthesis. Part B, Polyketides, aminocoumarins, and carbohydrates. Methods in enzymology, 0076-6879. Amsterdam, Boston, Elsevier/Academic Press, 2009, v. 459, pp 17–47.

285 Hautala A, Torkkell S, Räty K, Kunnari T, Kantola J, Mantsälä P, Hakala J, Ylihonko K: Studies on a second and third ring cyclization in anthracycline biosynthesis. J. Antibiot. 2003;56:143–153.

286 Rohr J, Méndez C, Salas JA: The Biosynthesis of Aureolic Acid Group Antibiotics. Bioorganic Chemistry 1999;27:41–54.

287 Zhang W, Watanabe K, Wang CCC, Tang Y: Investigation of early tailoring reactions in the oxytetracycline biosynthetic pathway. J Biol Chem 2007;282:25717–25725.

288 Xu W, Raetz LB, Wang P, Tang Y: An ATP-dependent ligase catalyzes the fourth ring cyclization in tetracycline biosynthesis. Tetrahedron 2016;72:3599–3604.

289 Clive DLJ: Chemistry of tetracyclines. Q. Rev., Chem. Soc. 1968;22:435.

290 Rhodes PM, Hunter IS, Friend EJ, Warren M: Recombinant DNA methods for the oxytetracycline producer Streptomyces rimosus. Biochem Soc Trans 1984;12:586–587.

291 Oliver TJ, Sinclair AC: Antibiotic M-319, 1964.

292 Mitscher LA, Swayze JK, Högberg T, Khanna I, Rao GS, Theriault RJ, Kohl W, Hanson C, Egan R: Biosynthesis of cetocycline. J. Antibiot. 1983;36:1405–1407.

293 Madoń J, Hütter R: Transformation system for Amycolatopsis (Nocardia) mediterranei: direct transformation of mycelium with plasmid DNA. J Bacteriol 1991;173:6325–6331.

294 Uchiyama H, Weisblum B: N-methyl transferase of Streptomyces erythraeus that confers resistance to the macrolidelincosamide-streptogramin B antibiotics: amino acid sequence and its homology to cognate R-factor enzymes from pathogenic bacilli and cocci. Gene 1985;38:103–110.

295 Combes P, Till R, Bee S, Smith MCM: The streptomyces genome contains multiple pseudo-attB sites for the (phi)C31-encoded site-specific recombination system. J Bacteriol 2002;184:5746–5752.

296 Zhang W, Ames BD, Tsai S-C, Tang Y: Engineered Biosynthesis of a Novel Amidated Polyketide, Using the Malonamyl-Specific Initiation Module from the Oxytetracycline Polyketide Synthase. Appl Environ Microbiol 2006;72:2573–2580.

297 Lešnik U, Lukežič T, Podgoršek A, Horvat J, Polak T, Šala M, Jenko B, Harmrolfs K, Ocampo-Sosa A, Martínez-Martínez L, Herron PR, Fujs Š, Kosec G, Hunter IS, Müller R, Petković H: Construction of a new class of tetracycline lead structures with potent antibacterial activity through biosynthetic engineering. Angew Chem Int Ed Engl 2015;54:3937–3940.

298 Lešnik U, Gormand A, Magdevska V, Fujs Š, Raspor P, Hunter I, Petković H: Regulatory Elements in Tetracycline-Encoding Gene Clusters: the otcG Gene Positively Regulates the Production of Oxytetracycline in Streptomyces rimosus. Food Technology and Biotechnology 2009;47:323–330. https://hrcak.srce.hr/file/62492.

299 Yin S, Wang W, Wang X, Zhu Y, Jia X, Li S, Yuan F, Zhang Y, Yang K: Identification of a cluster-situated activator of oxytetracycline biosynthesis and manipulation of its expression for improved oxytetracycline production in Streptomyces rimosus. Microb Cell Fact 2015;14:46.

300 Revill WP, Bibb MJ, Hopwood DA: Purification of a malonyltransferase from Streptomyces coelicolor A3(2) and analysis of its genetic determinant. J Bacteriol 1995;177:3946–3952.

301 Petkovic H, Thamchaipenet A, Zhou LH, Hranueli D, Raspor P, Waterman PG, Hunter IS: Disruption of an aromatase/cyclase from the oxytetracycline gene cluster of Streptomyces rimosus results in production of novel polyketides with shorter chain lengths. J Biol Chem 1999;274:32829–32834.

302 Wohlert SE, Wendt-Pienkowski E, Bao W, Hutchinson CR: Production of aromatic minimal polyketides by the daunorubicin polyketide synthase genes reveals the incompatibility of the heterologous DpsY and JadI cyclases. J Nat Prod 2001;64:1077–1080.

303 Rudd BA, Hopwood DA: Genetics of actinorhodin biosynthesis by Streptomyces coelicolor A3(2). J Gen Microbiol 1979;114:35–43.

304 Bartel PL, Zhu CB, Lampel JS, Dosch DC, Connors NC, Strohl WR, Beale JM, Floss HG: Biosynthesis of anthraquinones by interspecies cloning of actinorhodin biosynthesis genes in streptomycetes: clarification of actinorhodin gene functions. J Bacteriol 1990;172:4816–4826.

305 Müller R, Lukežič T, Remškar M, Zaburannyi N, Bader C, Sikandar A, Köhnke J: Gene cluster for the biosynthetic production of tetracycline compounds in a heterologous host, 2019.

306 Kallio P, Sultana A, Niemi J, Mäntsälä P, Schneider G: Crystal structure of the polyketide cyclase AknH with bound substrate and product analogue: implications for catalytic mechanism and product stereoselectivity. Journal of Molecular Biology 2006;357:210–220.

307 Sultana A, Kallio P, Jansson A, Niemi J, Mäntsälä P, Schneider G: Crystallization and preliminary crystallographic data of SnoaL, a polyketide cyclase in nogalamycin biosynthesis. Acta Crystallogr D Biol Crystallogr 2004;60:1118–1120.

308 Lombó F, Blanco G, Fernández E, Méndez C, Salas J: Characterization of Streptomyces argillaceus genes encoding a polyketide synthase involved in the biosynthesis of the antitumor mithramycin. Gene 1996;172:87–91.

309 Chiba K, Hoshino Y, Ishino K, Kogure T, Mikami Y, Uehara Y, Ishikawa J: Construction of a pair of practical Nocardia-Escherichia coli shuttle vectors. Jpn J Infect Dis 2007;60:45–47.

310 Markham A: Oritavancin: first global approval. Drugs 2014;74:1823–1828.

311 Shirling EB, Gottlieb D: Methods for Characterization of Streptomyces Species. International Journal of Systematic Bacteriology 1966;16:313–340.

312 Peterson JE: Chapter IX Isolation, Cultivation and Maintenance of the Myxobacteria; in Norris JR, Ribbons DW (eds): Methods in microbiology. Methods in Microbiology. London, New York, Academic Press, 1969, 3B, pp 185–210.

313 Gerth K, Trowitzsch W, Piehl G, Schultze R, Lehmann J: Inexpensive media for mass cultivation of myxobacteria. Appl Microbiol Biotechnol 1984;19.

314 Garcia R, Gerth K, Stadler M, Dogma IJ, Müller R: Expanded phylogeny of myxobacteria and evidence for cultivation of the 'unculturables'. Mol Phylogenet Evol 2010;57:878–887.

315 Rebets Y, Brötz E, Tokovenko B, Luzhetskyy A: Actinomycetes biosynthetic potential: how to bridge in silico and in vivo? J Ind Microbiol Biotechnol 2014;41:387–402.

316 Siegl T, Tokovenko B, Myronovskyi M, Luzhetskyy A: Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes. Metab Eng 2013;19:98–106.

317 Seghezzi N, Amar P, Koebmann B, Jensen PR, Virolle M-J: The construction of a library of synthetic promoters revealed some specific features of strong Streptomyces promoters. Appl Microbiol Biotechnol 2011;90:615–623.

318 Lander ES: The Heroes of CRISPR. Cell 2016;164:18–28.

319 Le Cong, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F: Multiplex genome engineering using CRISPR/Cas systems. Science 2013;339:819–823.

320 Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM: RNA-guided human genome engineering via Cas9. Science 2013;339:823–826.

321 Wang Q, Xie F, Tong Y, Habisch R, Yang B, Zhang L, Müller R, Fu C: Dual-function chromogenic screening-based CRISPR/Cas9 genome editing system for actinomycetes. Appl Microbiol Biotechnol 2020;104:225–239.

322 Pigac J, Schrempf H: A Simple and Rapid Method of Transformation of Streptomyces rimosus R6 and Other Streptomycetes by Electroporation. Appl Environ Microbiol 1995;61:352–356.

323 Rubin EJ, Akerley BJ, Novik VN, Lampe DJ, Husson RN, Mekalanos JJ: In vivo transposition of mariner-based elements in enteric bacteria and mycobacteria. Proc Natl Acad Sci U S A 1999;96:1645–1650.

324 Wenzel SC, Müller R: The impact of genomics on the exploitation of the myxobacterial secondary metabolome. Nat Prod Rep 2009;26:1385–1407.

325 Wu SS, Kaiser D: Markerless deletions of pil genes in Myxococcus xanthus generated by counterselection with the Bacillus subtilis sacB gene. J Bacteriol 1996;178:5817–5821.

326 Zhao J-Y, Zhong L, Shen M-J, Xia Z-J, Cheng Q-X, Sun X, Zhao G-P, Li Y-Z, Qin Z-J: Discovery of the autonomously replicating plasmid pMF1 from Myxococcus fulvus and development of a gene cloning system in Myxococcus xanthus. Appl Environ Microbiol 2008;74:1980–1987.

327 Salmi D, Magrini V, Hartzell PL, Youderian P: Genetic Determinants of Immunity and Integration of Temperate Myxococcus xanthus Phage Mx8. J Bacteriol 1998;180:614–621.

328 Julien B: Characterization of the integrase gene and attachment site for the Myxococcus xanthus bacteriophage Mx9. J Bacteriol 2003;185:6325–6330.

329 Sucipto H, Pogorevc D, Luxenburger E, Wenzel SC, Müller R: Heterologous production of myxobacterial α-pyrone antibiotics in Myxococcus xanthus. Metab Eng 2017;44:160–170.

330 Pogorevc D, Tang Y, Hoffmann M, Zipf G, Bernauer HS, Popoff A, Steinmetz H, Wenzel SC: Biosynthesis and Heterologous Production of Argyrins. ACS Synth Biol 2019;8:1121–1133.

331 Kjaerulff L, Raju R, Panter F, Scheid U, Garcia R, Herrmann J, Müller R: Pyxipyrrolones: Structure Elucidation and Biosynthesis of Cytotoxic Myxobacterial Metabolites. Angew Chem Int Ed Engl 2017;56:9614–9618.

332 Jaoua S, Neff S, Schupp T: Transfer of mobilizable plasmids to Sorangium cellosum and evidence for their integration into the chromosome. Plasmid 1992;28:157–165.

333 Pradella S, Hans A, Spröer C, Reichenbach H, Gerth K, Beyer S: Characterisation, genome size and genetic manipulation of the myxobacterium Sorangium cellulosum So ce56. Arch Microbiol 2002;178:484–492.

334 Kopp M, Irschik H, Gross F, Perlova O, Sandmann A, Gerth K, Müller R: Critical variations of conjugational DNA transfer into secondary metabolite multiproducing Sorangium cellulosum strains So ce12 and So ce56: development of a mariner-based transposon mutagenesis system. J Biotechnol 2004;107:29–40.

335 Viehrig K, Surup F, Volz C, Herrmann J, Abou Fayad A, Adam S, Köhnke J, Trauner D, Müller R: Structure and Biosynthesis of Crocagins: Polycyclic Posttranslationally Modified Ribosomal Peptides from Chondromyces crocatus. Angew Chem Int Ed Engl 2017;56:7407–7410.

336 Xia Z-J, Wang J, Hu W, Liu H, Gao X-Z, Wu Z-H, Zhang P-Y, Li Y-Z: Improving conjugation efficacy of Sorangium cellulosum by the addition of dual selection antibiotics. J Ind Microbiol Biotechnol 2008;35:1157–1163.

337 Thomas CM, Nielsen KM: Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat Rev Microbiol 2005;3:711–721.

338 Goldman BS, Nierman WC, Kaiser D, Slater SC, Durkin AS, Eisen JA, Eisen J, Ronning CM, Barbazuk WB, Blanchard M, Field C, Halling C, Hinkle G, Iartchuk O, Kim HS, Mackenzie C, Madupu R, Miller N, Shvartsbeyn A, Sullivan SA, Vaudin M, Wiegand R, Kaplan HB: Evolution of sensory complexity recorded in a myxobacterial genome. Proc Natl Acad Sci U S A 2006;103:15200–15205.

339 Cohan FM: Bacterial species and speciation. Syst Biol 2001;50:513–524.

340 Huntley S, Zhang Y, Treuner-Lange A, Kneip S, Sensen CW, Søgaard-Andersen L: Complete genome sequence of the fruiting myxobacterium Corallococcus coralloides DSM 2259. J Bacteriol 2012;194:3012–3013.

341 Wenzel SC, Kunze B, Höfle G, Silakowski B, Scharfe M, Blöcker H, Müller R: Structure and biosynthesis of myxochromides S1-3 in Stigmatella aurantiaca: evidence for an iterative bacterial type I polyketide synthase and for module skipping in nonribosomal peptide biosynthesis. Chembiochem 2005;6:375–385.

342 Gaitatzis N, Kunze B, Müller R: Novel insights into siderophore formation in myxobacteria. Chembiochem 2005;6:365–374.

343 Bode HB, Meiser P, Klefisch T, Cortina NSdj, Krug D, Göhring A, Schwär G, Mahmud T, Elnakady YA, Müller R: Mutasynthesis-derived myxalamids and origin of the isobutyryl-CoA starter unit of myxalamid B. Chembiochem 2007;8:2139–2144.

344 Meiser P, Weissman KJ, Bode HB, Krug D, Dickschat JS, Sandmann A, Müller R: DKxanthene biosynthesis--understanding the basis for diversity-oriented synthesis in myxobacterial secondary metabolism. Chemistry & Biology 2008;15:771–781.

345 Pistorius D, Müller R: Discovery of the rhizopodin biosynthetic gene cluster in Stigmatella aurantiaca Sg a15 by genome mining. Chembiochem 2012;13:416–426.

346 Viehrig K, Surup F, Harmrolfs K, Jansen R, Kunze B, Müller R: Concerted action of P450 plus helper protein to form the amino-hydroxy-piperidone moiety of the potent protease inhibitor crocapeptin. J Am Chem Soc 2013;135:16885–16894.

347 Steinmetz H, Li J, Fu C, Zaburannyi N, Kunze B, Harmrolfs K, Schmitt V, Herrmann J, Reichenbach H, Höfle G, Kalesse M, Müller R: Isolation, Structure Elucidation, and (Bio)Synthesis of Haprolid, a Cell-Type-Specific Myxobacterial Cytotoxin. Angew Chem Int Ed Engl 2016;55:10113–10117.

348 Park S, Hyun H, Lee JS, Cho K: Identification of the Phenalamide Biosynthetic Gene Cluster in Myxococcus stipitatus DSM 14675. J Microbiol Biotechnol 2016;26:1636–1642.

349 Fu C, Auerbach D, Li Y, Scheid U, Luxenburger E, Garcia R, Irschik H, Müller R: Solving the Puzzle of One-Carbon Loss in Ripostatin Biosynthesis. Angew Chem Int Ed Engl 2017;56:2192–2197.

350 Panter F, Garcia R, Thewes A, Zaburannyi N, Bunk B, Overmann J, Gutierrez MV, Krug D, Müller R: Production of a Dibrominated Aromatic Secondary Metabolite by a Planctomycete Implies Complex Interaction with a Macroalgal Host. ACS Chem Biol 2019;14:2713–2719.

351 Bilyk O, Sekurova ON, Zotchev SB, Luzhetskyy A: Cloning and Heterologous Expression of the Grecocycline Biosynthetic Gene Cluster. PLoS ONE 2016;11:e0158682.

352 Wang H, Li Z, Jia R, Hou Y, Yin J, Bian X, Li A, Müller R, Stewart AF, Fu J, Zhang Y: RecET direct cloning and Redαβ recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. Nat Protoc 2016;11:1175–1190.

353 Huo L, Hug JJ, Fu C, Bian X, Zhang Y, Müller R: Heterologous expression of bacterial natural product biosynthetic pathways. Nat Prod Rep 2019;36:1412–1436.

354 Pogorevc D, Panter F, Schillinger C, Jansen R, Wenzel SC, Müller R: Production optimization and biosynthesis revision of corallopyronin A, a potent anti-filarial antibiotic. Metab Eng 2019;55:201–211.

355 Surup F, Viehrig K, Rachid S, Plaza A, Maurer CK, Hartmann RW, Müller R: Crocadepsins-Depsipeptides from the Myxobacterium Chondromyces crocatus Found by a Genome Mining Approach. ACS Chem Biol 2018;13:267–272.

356 Becker H, Scher JM, Speakman J-B, Zapp J: Bioactivity guided isolation of antimicrobial compounds from Lythrum salicaria. Fitoterapia 2005;76:580–584.

357 Kersten RD, Lane AL, Nett M, Richter TKS, Duggan BM, Dorrestein PC, Moore BS: Bioactivity-guided genome mining reveals the lomaiviticin biosynthetic gene cluster in Salinispora tropica. Chembiochem 2013;14:955–962.

358 Oli S, Abdelmohsen UR, Hentschel U, Schirmeister T: Identification of plakortide E from the Caribbean sponge Plakortis halichondroides as a trypanocidal protease inhibitor using bioactivity-guided fractionation. Mar Drugs 2014;12:2614–2622.

359 Rédei D, Kúsz N, Jedlinszki N, Blazsó G, Zupkó I, Hohmann J: Bioactivity-Guided Investigation of the Anti-Inflammatory Activity of Hippophae rhamnoides Fruits. Planta Med 2018;84:26–33.

360 Foulston L: Genome mining and prospects for antibiotic discovery. Curr Opin Microbiol 2019;51:1–8.

361 Sikandar A, Cirnski K, Testolin G, Volz C, Brönstrup M, Kalinina OV, Müller R, Koehnke J: Adaptation of a Bacterial Multidrug Resistance System Revealed by the Structure and Function of AlbA. J Am Chem Soc 2018;140:16641–16649.

362 Tang X, Li J, Millán-Aguiñaga N, Zhang JJ, O'Neill EC, Ugalde JA, Jensen PR, Mantovani SM, Moore BS: Identification of Thiotetronic Acid Antibiotic Biosynthetic Pathways by Target-directed Genome Mining. ACS Chem Biol 2015;10:2841–2849.

363 Lloyd DG, Whitworth DE: The Myxobacterium Myxococcus xanthus Can Sense and Respond to the Quorum Signals Secreted by Potential Prey Organisms. Front Microbiol 2017;8:439.

364 Zaburannyi N: Clust-o-matic, an online tool. In development.

365 Esquilín-Lebrón KJ, Boynton TO, Shimkets LJ, Thomas MG: An Orphan MbtH-Like Protein Interacts with Multiple Nonribosomal Peptide Synthetases in Myxococcus xanthus DK1622. J Bacteriol 2018;200.

366 McMahon MD, Rush JS, Thomas MG: Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in Mycobacterium tuberculosis. J Bacteriol 2012;194:2809–2818.

367 Bollinger JM, Price JC, Hoffart LM, Barr EW, Krebs C: Mechanism of Taurine: α-Ketoglutarate Dioxygenase (TauD) fromEscherichia coli. Eur. J. Inorg. Chem. 2005;2005:4245–4254.

368 Ji X, Tu J, Song Y, Zhang C, Wang L, Li Q, Ju J: A Luciferase-Like Monooxygenase and Flavin Reductase Pair AbmE2/AbmZ Catalyzes Baeyer–Villiger Oxidation in Neoabyssomicin Biosynthesis. ACS Catal. 2020;10:2591–2595.

369 Maier S, Heitzler T, Asmus K, Brötz E, Hardter U, Hesselbach K, Paululat T, Bechthold A: Functional characterization of different ORFs including luciferase-like monooxygenase genes from the mensacarcin gene cluster. Chembiochem 2015;16:1175–1182.

370 Shi Y, Bueno A, van der Donk WA: Heterologous production of the lantibiotic Ala(0)actagardine in Escherichia coli. Chem Commun (Camb ) 2012;48:10966–10968.

371 Li S, Wang H, Li Y, Deng J, Lu C, Shen Y, Shen Y: Biosynthesis of hygrocins, antitumor naphthoquinone ansamycins produced by Streptomyces sp. LZ35. Chembiochem 2014;15:94–102.

372 Zhang C, Zhang H, Ju J: On-PKS Baeyer-Villiger-Type O-Atom Insertion Catalyzed by Luciferase-Like Monooxygenase OvmO during Olimycin Biosynthesis. Org Lett 2020;22:1780–1784.

373 Xu Z, Baunach M, Ding L, Peng H, Franke J, Hertweck C: Biosynthetic code for divergolide assembly in a bacterial mangrove endophyte. Chembiochem 2014;15:1274–1279.

374 Leisch H, Morley K, Lau PCK: Baeyer-Villiger monooxygenases: more than just green chemistry. Chem Rev 2011;111:4165–4222.

375 Khatri Y, Hannemann F, Perlova O, Müller R, Bernhardt R: Investigation of cytochromes P450 in myxobacteria: excavation of cytochromes P450 from the genome of Sorangium cellulosum So ce56. FEBS Lett 2011;585:1506–1513.

376 Malla S, Thuy TTT, Oh TJ, Sohng JK: Identification and characterization of gerPI and gerPII involved in epoxidation and hydroxylation of dihydrochalcolactone in Streptomyces species KCTC 0041BP. Arch Microbiol 2011;193:95–103.

377 Buntin K, Rachid S, Scharfe M, Blöcker H, Weissman KJ, Müller R: Production of the antifungal isochromanone ajudazols A and B in Chondromyces crocatus Cm c5:

biosynthetic machinery and cytochrome P450 modifications. Angew Chem Int Ed Engl 2008;47:4595–4599.

378 Hur GH, Vickery CR, Burkart MD: Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. Nat Prod Rep 2012;29:1074–1098.

379 Hornbogen T, Glinski M, Zocher R: Biosynthesis of depsipeptide mycotoxins in Fusarium; in Logrieco A, Bailey JA, Cooke BM (eds): Mycotoxins in plant disease: Under the aegis of COST Action 835 "Agriculturally important toxigenic fungi 1998-2003," EU project (QLK 1-CT-1998-01380), and ISPP "Fusarium Committee". European journal of plant pathology. Dordrecht, Springer-Science+Business Media, 2002, volume 108, number 7 (Special issue), pp 713–718.

380 Mootz HD, Schwarzer D, Marahiel MA: Ways of Assembling Complex Natural Products on Modular Nonribosomal Peptide Synthetases A list of abbreviations can be found at the end of the text. ChemBioChem 2002;3:490.

381 Li L, Deng W, Song J, Ding W, Zhao Q-F, Peng C, Song W-W, Tang G-L, Liu W: Characterization of the saframycin A gene cluster from Streptomyces lavendulae NRRL 11002 revealing a nonribosomal peptide synthetase system for assembling the unusual tetrapeptidyl skeleton in an iterative manner. J Bacteriol 2008;190:251–263.

382 Fisch KM: Biosynthesis of natural products by microbial iterative hybrid PKS–NRPS. RSC Adv. 2013;3:18228.

383 Yamanaka K, Maruyama C, Takagi H, Hamano Y: Epsilon-poly-L-lysine dispersity is controlled by a highly unusual nonribosomal peptide synthetase. Nat Chem Biol 2008;4:766–772.

384 Singh V, Haque S, Niwas R, Srivastava A, Pasupuleti M, Tripathi CKM: Strategies for Fermentation Medium Optimization: An In-Depth Review. Front Microbiol 2016;7:2087.

385 Chuck-Hernández C, Ozuna C: Protein Isolates From Meat Processing By-Products; in Galanakis CM (ed): Proteins: Sustainable source, processing and applications / edited by Charis M. Galanakis. London, Academic Press, 2019, pp 131–162.

386 Liu G, Chater KF, Chandra G, Niu G, Tan H: Molecular regulation of antibiotic biosynthesis in streptomyces. Microbiol Mol Biol Rev 2013;77:112–143.

387 van Wezel GP, McDowall KJ: The regulation of the secondary metabolism of Streptomyces: new links and experimental advances. Nat Prod Rep 2011;28:1311–1333.

388 Tan G-Y, Peng Y, Lu C, Bai L, Zhong J-J: Engineering validamycin production by tandem deletion of γ-butyrolactone receptor genes in Streptomyces hygroscopicus 5008. Metab Eng 2015;28:74–81.

389 Bibb MJ: Regulation of secondary metabolism in streptomycetes. Curr Opin Microbiol 2005;8:208–215.

390 Wenzel SC, Müller R: Myxobacteria--'microbial factories' for the production of bioactive secondary metabolites. Mol Biosyst 2009;5:567–574.

391 Takano E: Gamma-butyrolactones: Streptomyces signalling molecules regulating antibiotic production and differentiation. Curr Opin Microbiol 2006;9:287–294.

392 Martín J-F, Liras P: Engineering of regulatory cascades and networks controlling antibiotic biosynthesis in Streptomyces. Curr Opin Microbiol 2010;13:263–273.

393 Stratigopoulos G, Gandecha AR, Cundliffe E: Regulation of tylosin production and morphological differentiation in Streptomyces fradiae by TylP, a deduced gamma-butyrolactone receptor. Mol Microbiol 2002;45:735–744.

394 Willey JM, Gaskell AA: Morphogenetic signaling molecules of the streptomycetes. Chem Rev 2011;111:174–187.

395 Tanaka Y, Kasahara K, Hirose Y, Murakami K, Kugimiya R, Ochi K: Activation and products of the cryptic secondary metabolite biosynthetic gene clusters by rifampin resistance (rpoB) mutations in actinomycetes. J Bacteriol 2013;195:2959–2970.

396 Gomez-Escribano JP, Bibb MJ: Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters. Microb Biotechnol 2011;4:207–215.

397 Okamoto-Hosoya Y, Sato TA, Ochi K: Resistance to paromomycin is conferred by rpsL mutations, accompanied by an enhanced antibiotic production in Streptomyces coelicolor A3(2). J. Antibiot. 2000;53:1424–1427.

398 Nair J, Rouse DA, Bai GH, Morris SL: The rpsL gene and streptomycin resistance in single and multiple drug-resistant strains of Mycobacterium tuberculosis. Mol Microbiol 1993;10:521–527.

399 Herrmann J, Lukežič T, Kling A, Baumann S, Hüttel S, Petković H, Müller R: Strategies for the Discovery and Development of New Antibiotics from Natural Products: Three Case Studies. Curr Top Microbiol Immunol 2016;398:339–363.

400 Bibb MJ, White J, Ward JM, Janssen GR: The mRNA for the 23S rRNA methylase encoded by the ermE gene of Saccharopolyspora erythraea is translated in the absence of a conventional ribosome-binding site. Mol Microbiol 1994;14:533–545.