Studies towards secondary metabolite formation in Myxobacteria: Elucidation of biosynthetic pathways and yield optimization

Dissertation
zur Erlangung des Grades
des Doktors der Naturwissenschaften
der Naturwissenschaftlich-Technischen Fakultät III
Chemie, Pharmazie, Bio- und Werkstoffwissenschaften
der Universität des Saarlandes

von
Stefan Müller
Saarbrücken
2015

Tag des Kolloquiums:	31.07.2015
Dekan:	Prof. Dr. Dirk Bähre
Berichterstatter:	Prof. Dr. Rolf Müller
	Prof. Dr. Rolf W. Hartmann
Vorsitz:	Prof. Dr. Uli Kazmaier
Akad. Mitarbeiter:	Dr. Judith Becker

Diese Arbeit entstand unter der Anleitung von Prof. Dr. Rolf Müller in der Fachrichtung 8.2 Pharmazeutische Biotechnologie der Naturwissenschaftlich-Technischen Fakultät III der Universität des Saarlandes in der Zeit von Mai 2011 bis April 2015.

Veröffentlichungen der Dissertation

Teile dieser Arbeit wurden vorab mit Genehmigung der Naturwissenschaftlich-Technischen Fakultät III, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen veröffentlicht:

Müller S, Rachid S, Hoffmann T, Surup F, Volz C, Zaburannyi N, & Müller R (2014) Biosynthesis of crocacin involves an unusual hydrolytic release domain showing similarity to condensation domains. *Chemistry & Biology*, 21 (7), 855-865.

Hoffmann T, **Müller S**, Nadmid S, Garcia R, & Müller R (2013) Microsclerodermins from terrestrial myxobacteria: An intriguing biosynthesis likely connected to a sponge symbiont. *Journal of the American Chemical Society*, 135 (45), 16904-16911.

Darüber hinaus konnte zu einer weiteren Publikation, welche nicht Teil dieser Arbeit ist, beigetragen werden:

Jahns C, Hoffmann T, **Müller S**, Gerth K, Washausen P, Höfle G, Reichenbach H, Kalesse M, & Müller R (2012) Pellasoren: Structure elucidation, biosynthesis, and total synthesis of a cytotoxic secondary metabolite from *Sorangium cellulosum*. *Angewandte Chemie International Edition*, 51 (21), 5239-5243.

Table of contents

Zusammenfassung

Von Myxobakterien produzierte Naturstoffe weisen eine enorme strukturelle Vielfalt mit bemerkenswerten biologischen Aktivitäten auf. In diesem Kontext befasst sich die vorliegende Dissertation sowohl mit der Identifizierung und Charakterisierung von Naturstoff-Biosynthesewegen, als auch mit der Optimierung eines Produktionsstammes.

Die Arbeiten zum Sekundärmetaboliten Crocacin zeigen, dass die eingehende Erforschung bisher unbekannter Biosynthesewege großes Potential zur Entdeckung beispielloser Aspekte von naturstoffproduzierenden Multienzymkomplexen birgt. Die erstmalige Beschreibung einer Kondensationsdomäne mit hydrolytischer Aktivität ebnet somit den Weg für Untersuchungen zum Mechanismus der Katalyse.

Ein weiterer Teil dieser Arbeit befasste sich mit der vergleichenden Charakterisierung der Microsclerodermin Gencluster aus *Sorangium cellulosum* So ce38 und *Jahnella* sp. MSr9139. Hierbei konnten die strukturellen Unterschiede der einzelnen Derivate auf Unterschiede in den jeweiligen Genclustern zurückgeführt werden. Des Weiteren wurden terrestrische Produzenten des bisher nur aus marinen Schwämmen bekannten Naturstoffs Microsclerodermin D identifiziert.

Im Rahmen der Optimierung des Ambruticinproduzenten *S. cellulosum* strain #8405 wurden mehrere Faktoren, die im Zusammenhang mit der Regulation der Biosynthese stehen, charakterisiert und modifiziert. Zudem konnten für diese Bakteriengattung neuartige genetische Methoden entwickelt und erfolgreich angewandt werden.

Abstract

Natural products produced by myxobacteria exhibit an enormous structural diversity with noticeable biological activities. In this context, the present dissertation deals with both, the identification and characterization of natural product biosynthetic pathways and the optimization of a production strain.

The work on the secondary metabolite crocacin illustrates that the detailed exploration of so far unknown biosynthetic pathways bears huge potential to reveal unprecedented aspects of multienzyme complexes which assemble natural products. The primal depiction of a condensation domain with hydrolytic activity consequently paves the way for investigations concerning the mechanism of catalysis.

Another part of this work covers the comparative characterization of the microsclerodermin gene clusters from *Sorangium cellulosum* So ce38 und *Jahnella* sp. MSr9139. Hereby, the structural differences among the respective derivatives could be linked to differences in corresponding gene clusters. Furthermore, terrestrial producers of the natural product microsclerodermin D, previously only known from marine sponges, were identified.

In the course of the optimization of the ambruticin producer *S. cellulosum* strain #8405 several factors connected to the regulation of the biosynthesis were characterized and modified. In addition novel genetic methods for this particular genus were developed and successfully applied.

Introduction

Natural products: history, sources, applications, and future potentials

Natural products or so called secondary metabolites are by definition low molecular weight molecules which are produced by a biological source [1]. The tremendous chemical diversity generated throughout millions of years of evolution truly accounts for the certainty that natural products have been one of the most important sources for potential drug leads [2;3]. Besides their therapeutic use as for example antibiotics, antitumor or immunosuppressive medications, natural products are used in agriculture, serving as insecticides, herbicides, or fungicides. The biosynthesis of natural products comes along with notable expenses for the producer's metabolism, thus supporting the current notion that literally all such secondary metabolites have a certain purpose. As a consequence of the metabolic effort that is needed, biosynthesis of those metabolites is strictly regulated. In the microbial world, secondary metabolites are secreted in order to interfere with competitor organisms [4;5], to facilitate interspecies communication [6;7], or to coordinate developmental processes [8-11].

The utilization of natural products by mankind is not a concept of the modern times; in fact the documented use of herbals to cure diseases dates far back to the 3rd millennium BC. The earliest record thereof was found on clay tablets in cuneiform from Mesopotamia dating from around 2600 BC. Those relicts describe the use of oils from trees like cypress or cedar and herbs like licorice and myrrh to treat coughs, colds and inflammations [12]. As plants have been widespread and readily accessible during the centuries, many effective therapies originate from people experimenting with their local flora, aiming to search for a cure to fight different diseases. To date, among the 270,000 terrestrial plants that were taxonomically classified a number of approximately 30,000 is used in medicinal applications [13], with salicin from *Salix alba* L. (analgesic), morphine from *Papaver somniferum* (analgesic, sedative), and digoxin from *Digitalis lanata* (treatment of heart conditions) being only some of the most prominent ones.

In line with this, fungi are as well a noticeable source of important natural products, although their actual potential has long remained undiscovered. With the more recent advantages in microbiology and the ability to isolate and grow fungi they were recognized as valuable suppliers of natural products including pharmaceutically relevant molecules [14]. Without doubt, the discovery of penicillin from the fungus *Penicillium notatum* made by Alexander Fleming in 1929 [15] together with subsequent observations of its broad therapeutic applicability by Ernst Boris Chain and Howard Walter Florey ushered a new era in medicine (The Golden Age of Antibiotics) and won them the Nobel prize in 1945 [16]. Besides penicillin, other important natural products derived from fungi are the cyclosporines from *Tolypocladium inflatum* (immunosuppressant),

asperlicin from *Aspergillus alliaceus* (cholecystokinin receptor CCK antagonist), and mevinolin from *Aspergillus terreus* (cholesterol-lowering agent).

The discovery of penicillin promoted the extensive investigation of microorganisms with the aim to detect more natural products of potential medicinal interest. During this time, mainly soil-derived bacteria came into the focus of natural product research. In 1939, René J. Dubos together with Rollin D. Hotchkiss discovered gramicidin, an antimicrobial peptide produced by *Bacillus brevis* that exhibits activity against gram positive bacteria [17-19]. It was the very first antibiotic that was used in clinics and it gained certain importance in treating infected wounds of soldiers during World War II. In the following decades, the identification of active compounds from microbes progressed rapidly. This impressive development is illustrated best by numbers: In 1940 only 10-20, in 1950 300-400, in 1960 around 800-1,000, and by the year 1970 already 2,500 antibiotics were known [20]. Thereafter the amount of identified bioactive microbial compounds has doubled in every ten years, from about 5,000 in 1980 to more than 22,000 until the end of 2002. Actinobacteria account for the vast majority of these molecules with vancomycin, erythromycin, and streptomycin being popular representatives [21;22]. Besides actinobacteria, the group of bacilli has proven to be an additional prolific resource of structurally diverse and active metabolites [23;24]. In this context, myxobacteria have also gained increased attention more recently. Their exploration significantly contributed to the pool of microbial natural products, because of which they joined the exclusive club of extraordinary secondary metabolite producers [25-27].

The analysis of terrestrial organisms as a source of bioactive agents has a rather long tradition while the marine environment remained mostly untouched. But after the introduction of modern SCUBA diving in the mid-1970s things changed [28]. Consequently, the continuous exploration of the oceans resulted in the discovery of several thousand compounds by the turn of the century, thereby firmly establishing this sub-discipline in the large field of natural product research [29]. The marine habitats are a tremendously rich source for natural products owing to the variety of organisms found in the oceans. Alongside with algae, tunicates, snails, and bryozoans, sponges have frequently proven to be notable sources of chemical entities with versatile activities [30;31]. In their natural habitats sponges are often closely associated with bacteria of different kinds [32;33], thus some of the compounds are hypothesized to originate from these microbes rather than from the sponge itself. This assumption is supported by the fact that in a comprehensive phylogenetic study of marine sponge-associated microorganisms indicative DNA fragments that can be linked to actinobacterial and myxobacterial genomes were detected [34]. Since marine isolates derived from both bacterial groups have already shown

promising results, the efforts to further explore the marine environment for microbial diversity appears justified [35;36].

The facts depicted above all together show the great impact natural product research had on drug discovery in the last 80 years. Furthermore it should become apparent, that secondary metabolite producers are found among all domains of life and in a huge variety of habitats. In 2005, it was estimated that only less than 1 % of bacterial and 5 % of fungal species are known so far, respectively [12]. Thus, these numbers suggest that nature's treasure chest is still far from being exploited and that research in the area of secondary metabolites still bears great potential for the identification of unprecedented bioactive molecules. However, it will get increasingly difficult to detect novel chemical scaffolds from already well investigated sources. For example, approximately 1 % of known soil actinomycetes are producing streptomycin, whereas daptomycin producers are found only at a frequency of 10^{-7} amongst those bacteria [37]. This scenario does not apply to myxobacteria yet, because they have not been as intensively studied as most actinobacteria or bacilli. In addition to the fact that novel taxa are still being discovered as a potential source of novel compounds, enormous advances in all areas of myxobacterial natural product research were made and consequently this division is well positioned in order to tackle the future tasks of modern drug discovery.

Myxobacteria: producers of interesting secondary metabolites

Myxobacteria are Gram-negative, rod-shaped, and non-motile microorganisms characterized by high GC rich DNA that belong to the class of Deltaproteobacteria. Besides their fascinating appearance (see below), one distinctive characteristic is their complex developmental life cycle which undergoes stages from unicellularity, cellular aggregations and multicellular fruiting bodies bearing the myxospores (Figure 1) [38]. Similar to endospores from other bacteria, they are built to outlive harsh environmental conditions [39]. Another remarkable morphological feature of some myxobacteria is their ability to move in a collective effort, called gliding, thereby leaving their "footprints" as ripples on solid surfaces (Figure 1B). The main and most common source of myxobacteria is the soil where they can colonize the dung of herbivores and decaying plant materials, but metagenomic data together with actual isolates identified the costal and marine environments as other yet unexplored habitats [40].

By far the most popular and best studied representative of this order is *Myxococcus xanthus* DK1622, which became a model organism for the investigation of social interactions and development among the domain of bacteria [41]. To facilitate a better understanding of the mechanisms behind these cellular processes, the genome of *M. xanthus* was sequenced (9.14 Mb) and published in 2006, making it the first publicity available myxobacterial genome

[42]. Around 600 genes encode for putative regulatory functions illustrating the complexity of this organism. Even more impressively, a portion of 8.6 % of the whole genome is dedicated to the production of secondary metabolites (18 gene clusters), thereby showing a capacity for secondary metabolite production which is twice as high as for *Streptomyces coelicolor* [43]. A year later, the genome of *Sorangium cellulosum* So ce56 comprising 13.03 Mb was published [44]. Today it is the second biggest myxobacterial genome after that of *S. cellulosum* So0157-2 (14.78 Mb) [45]. Similar to *M. xanthus*, both strains encode a huge amount of regulatory kinases (So ce56: 498, So0157-2: 648) and sigma factor dependent regulators (So ce56 σ^{54}-dependent regulators: 80, So0157-2: 188), again displaying the enormous regulatory network that acts "behind the scenes". In terms of secondary metabolism, both organisms dedicate around 10 % of their genomes for this purpose.

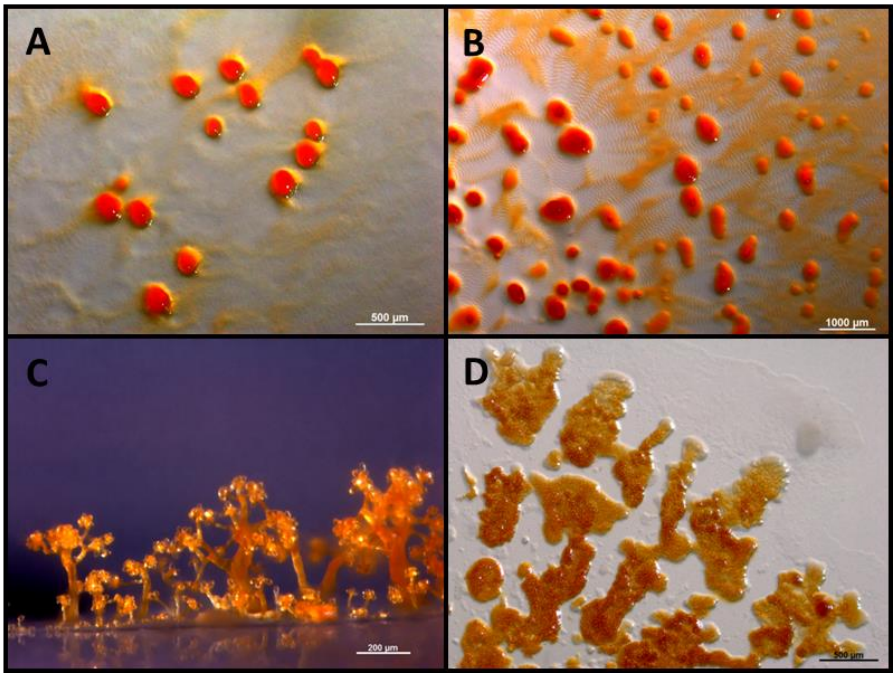


Figure 1: Myxobacterial cells and fruiting bodies of different representative genera. A: Knob-shaped fruiting bodies of *Myxococcus* spec. on agar. **B:** Rippling and fruiting body formations of *Myxococcus* on agar. C: Tree-like fruiting bodies of *Chondromyces crocatus* Cm c5 on edge of a petri dish. **D:** *Sorangium cellulosum* clusters of fruiting bodies on agar. (All pictures by courtesy of Dr. Ronald Garcia)

Although the myxobacterial molecules isolated so far exhibit a wide range of bioactivities (e.g. cytotoxic, anti-viral, anti-inflammatory, or metal-chelating), most of these substances inhibit the growth of fungi or bacteria [46]. This likely reflects the highly competitive ecosystem they live in; cellulose degraders like *Sorangium* species compete against fungi for their wooden substrates while proteolytic microbes like *Myxococcus* species are fighting against other degrading microorganisms in the soil [47]. Among the myxobacteria, the genus *Sorangium* has

proven to be the richest source of interesting bioactive metabolites, such as epothilone [48], ambruticin [49], ripostatin [50], and soraphen [51], but also other genera like *Chondromyces* or *Myxococcus* produce agents of interest, e.g. crocacin [52], tuggacin [53], myxovirescin [54], or saframycin [55] (Figure 2).

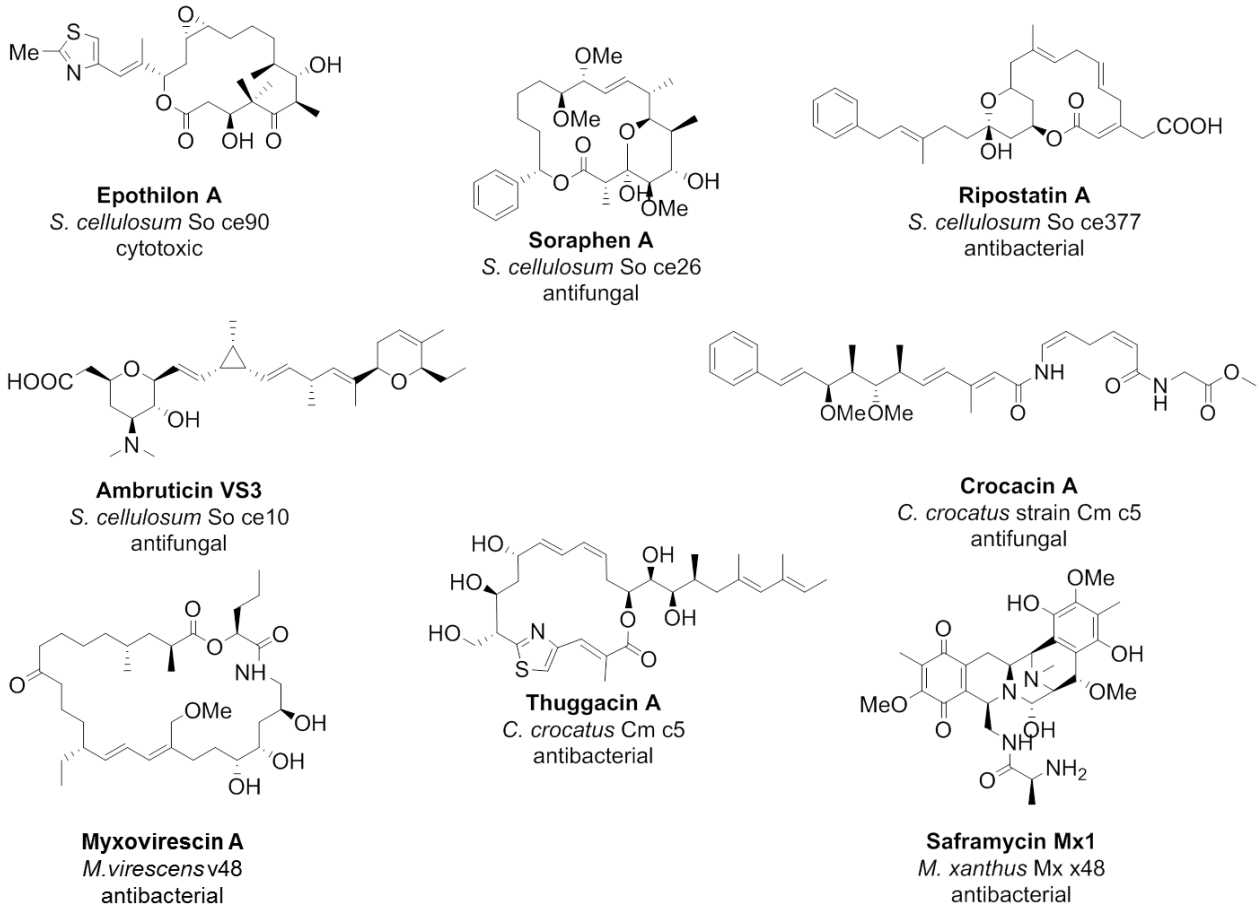


Figure 2: Selected structures of secondary metabolites from different *Sorangium and Myxococcus* species together with some compounds produced by *C. crocatus* Cm c5.

Of all the above depicted compounds, epothilon is currently regarded as the most important one as a chemically modified version (ixabepilone or Ixempra®) has been approved for clinical use to treat cancer in humans [56]. Additional compounds, such as ambruticin or soraphen, as well have great industrial potential and might be prospective candidates for fungicide development in agriculture. Most myxobacterial secondary metabolites published to date are synthesized by large multienzyme complexes named polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs), or hybrids thereof (see below). Furthermore, small bioactive peptides can be made directly at the ribosome and it is assumed that they are produced by nearly all organisms, from bacteria over plants to animals [57]. In fact, recent

studies about the origin of the cittilins in strains of *M. xanthus* have shown that this compound is made by the ribosome, thereby making it the first suchlike myxobacterial natural product [58].

Polyketide synthases

Polyketides are a group of very diverse natural products, comprising polyphenol, macrolide, polyene, enediyne, or polyether moieties, mainly build up by the sequential assembly of small building blocks like malonyl-CoA and methylmalonyl-CoA [59]. According to their architecture and the way they act, polyketide synthases (PKS) are categorized into different types [60;61]. Bacterial type I PKSs, such as 6-deoxyerythronolide synthase [62], are large multimodular enzymes with different catalytic activities, which assemble the respective compounds in a stepwise fashion. In particular, a minimal set of catalytic domains comprises a biosynthetic module which is essential for a single extension (Figure 3). This one module to one extension cycle principle is named colinearity. Hence, the individual modules are used just once during biosynthesis, but more and more pathways have been characterized in which particular units are working iteratively, i.e. a module is used more than once [63]. Contrary to this, fungal type I PKSs like the lovastatin synthase [64] usually work iteratively and co-linearity is not observed. A second type of PKS, named PKS type II, occurs, which is restricted to bacteria. These systems consist of discrete and usually monofunctional entities that are encoded by distinct genes. Apart from that, a third kind of polyketide synthesizing enzymes denoted as type III, or chalcone synthase-like, PKSs exist. Those are mainly employed by plants and even if their mechanism apparently resembles that of type I and II PKSs, there are some dissimilarities [60].

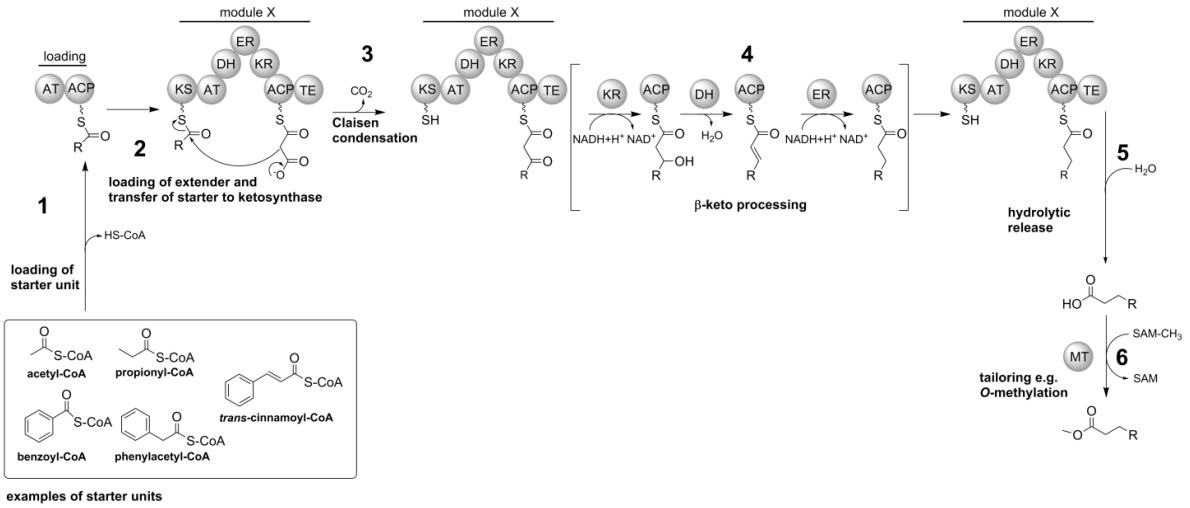


Figure 3: Schematic illustration of polyketide synthesis using the example of a hypothetical compound X. AT: acyltransferase, ACP: acyl carrier protein, DH: dehydratase, ER: enoyl reductase, KR: ketoreductase, KS: ketosynthase, MT: methyltransferase, TE: thioesterase. Further explanations are given in the text.

1: PKS assembly is initiated by transferring an activated starter unit onto the 5'-phosphopantetheine (Ppant) arm of the loading module's acyl carrier protein (ACP). This first step is performed by an intrinsic acyltransferase domain (AT), or alternatively a CoA-ligase that activates the precursor and further transfers it onto the ACP [65]. The fact that different loading modules utilize different starter molecules constitutes a lot to the diversity of secondary metabolites found in nature. Such starters can be rather simple like acetyl-CoA (chivosazol [66]), or propionyl-CoA (erythromycin [62]), but also more complex like benzoyl-CoA (soraphen [67]), phenylacetyl-CoA (microsclerodermin [68]), or *trans*-cinnamoyl-CoA (crocacin [69]).

2: The starter unit is then forwarded to a conserved cysteine residue of the ketosynthase domain (KS) from the next module. Simultaneously, the second AT selects the extender unit and hands it over to the Ppant moiety of the ACP. In addition to the most common extender units malonyl-CoA (as shown in the present example) or methylmalonyl-CoA, the use of ethylmalonyl-CoA, hydroxymalonyl-ACP, methoxymalonyl-ACP, and aminomalonyl-ACP has been reported [70].

3: The enzyme bound malonate undergoes decarboxylation followed by nucleophilic attack of the KS-bound starter unit, eventually giving rise to C-C bond formation between these molecules. This process is referred to as Claisen condensation.

4: The ACP-bound β-keto ester is further processed by the subsequent action of a ketoreduction domain (KR), a dehydratase domain (DH), and an enoylreductase domain (ER) yielding the alcohol, enoyl or methylene moiety. Notably, not every PKS module contains all of those domains and some may be extended by a methyltransferase domain that acts on carbon or oxygen atoms. Accordingly β-keto processing would stop at earlier stages. Is the β-keto ester just reduced to the alcohol, a chiral center is introduced whose stereochemistry is highly dependent on the nature of the KR [71].

5: Polyketide synthesis usually ends when the growing chain reached the end of the assembly line. The release of the molecule is mediated via a thioesterase domain (TE) and occurs either by hydrolysis or by the attack of an intrinsic nucleophile, yielding the linear carboxylic acid or the macrocycle [72].

6: The polyketide scaffold might be furthermore modified in one or more tailoring reactions such as methylation, oxygenation, halogenation, or glycosylation, thereby creating the final product [59].

As outlined earlier, the repertoire of polyketides made by nature is huge. This is mainly attributed to the modular architecture of the respective pathways, what allowed the evolution of a variety of assembly lines by intra- and intergenomic recombination events, thereby causing this wealth of molecules [73;74]. Consequently, there are certain levels of diversification within

polyketide biosynthesis. First, the building blocks together with the number of extension cycles and the mechanisms of chain release are primarily determining the polyketide scaffold. Moreover, β-keto processing is another important mechanism to the introduce diversity, and lastly the tailoring reactions are put the finishing touches to the compound. Moreover, hybrid pathways consisting of PKS and NRPS features exist [75]. This combination permits the construction of an even greater diversity than one of the systems alone would be capable of (see below).

Nonribosomal peptide synthetases

Nonribosomal peptide synthetases (NRPSs) are further production facilities of bioactive secondary metabolites. First discovered in 1970 by Lipmann *et al.* [76], these large enzyme complexes have been extensively studied [77], revealing structural and mechanistic insights [78-80].

Although the collectivity of secondary metabolites produced by NRPS systems is strikingly diverse, the basic principles of their biosynthesis are the same. As the name already implies, these types of molecules are not made at the ribosome, but at large multifunctional enzyme complexes, which assemble their products in a stepwise manner, likewise to PKS systems. Notably, NRPSs are capable of incorporating much more (~ 500) than just the limited set of the 22 proteinogenic amino acids, what contributes to the wide range of their products [81]. Figure 4 illustrates the main steps of nonribosomal peptide synthesis.

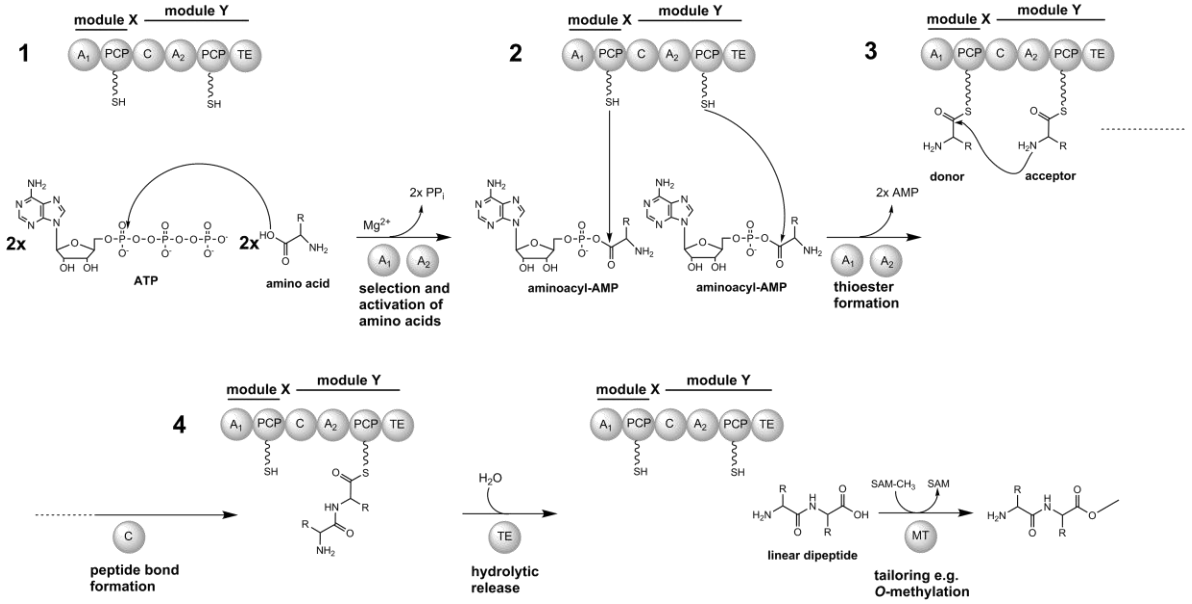


Figure 4: Schematic illustration of nonribosomal peptide synthesis. A: adenylation domain, C: condensation domain, MT: methyltransferase domain, PCP: peptidyl carrier protein, TE: thioesterase domain. Further explanations are given in the text.

1: Biosynthesis usually starts with the selection and activation of a dedicated building block by a specialized domain, the adenylation domain (A). This process resembles the activation of amino acids performed by aminoacyl-tRNA synthetases. However, both enzyme families share no sequence or structural similarities [79;82]. As the respective A domains only recognize one or a few amino acids, this step mainly determines the scaffold of the molecule. The substrate specificity of an A domain can either be predicted *in silico* [83], or actually determined using the heterologously produced domain in an equilibrium ATP-PP_i exchange assay [84].

2: In the second step, the activated amino acids are covalently attached to the Ppant moiety of the individual peptidyl carrier proteins (PCP). Similar to polyketide synthesis, the Ppant arm is used as a transportation system to shuttle substrates among the active sites of the assembly line.

3: Following the activation of the amino acids and their subsequent attachment to the assembly line, they get connected, yielding the dipeptide. This reaction is mediated by a condensation domain (C). In particular, C domains contain two sites, one for the donor and one for the acceptor molecule. Upon binding of both substrates, they catalyze the nucleophilic attack of the α-amino functionality from the acceptor to the thioester of the donor [85]. In typical NRPS assembly lines, C domains are usually only present in elongation modules and their number normally equals the number of peptide bonds formed. Nevertheless, special cases exist, where this general observation does not apply. In such a context these domains may be involved in other procedures, e.g. chain termination (see below).

4: After complete synthesis of the compound's backbone, release from the last PCP is performed. The easiest scenario is the hydrolysis of the thioester by a thioesterase domain (TE). However, there are other ways how the final product can be released. Such possibilities are discussed below.

5: Tailoring enzymes like methyl- or glycosyltransferases, mostly acting in *trans*, can add more diversity to the scaffold [86].

In addition to the basic domains shown in Figure 4 NRPS assembly lines often contain editing domains which are part of a distinct module. These entities act in close concert with the rest of the assembly line and account for the formation of D-amino acids or *N*-methylations in the growing peptide chain. One reason for the introduction of a D-amino into a natural product is stability, because this protects the compounds from being degraded by ordinary proteases. Furthermore the presence of D-amino acids may favor unusual conformations within the molecule, thus resulting in special activities. The most common strategy to introduce a D-amino acid into the growing peptide chain is the use of an epimerization domain (E). E domains catalyze the epimerization of the α-carbon of the PCP tethered donor amino acid resulting in a

D/L equilibrium [87]. Specificity regarding the incorporation of the D-amino acid is conferred by the downstream C domain, whose donor site promotes only incorporation of the D-enantiomer [85]. A second strategy is known from the cyclosporine pathway where an A domain solely loads the D-enantiomer, which is beforehand supplied by a pyridoxal phosphate dependent alanine-racemase [88].

N-methylation of a nascent molecule is carried out by *N*-methyltransferases that are usually located between the A and the PCP domain of a certain module. Examples for *N*-methylation of the nascent peptide chain are found in microsclerodermin [68] and enniatin [89] biosynthesis. Further studies of the enniatin synthetase have identified S-adenosylmethionine (SAM) as methyl group donor and these results suggest that methylation might take place prior to amide bond formation [90].

Another sort of modification represents the heterocyclization of amino acid side chains from cysteine, serine, or threonine catalyzed by specialized cyclization domains (Cy). This kind of NRPS domain was first identified in bacitracin biosynthesis, where it accounts for the condensation of isoleucine and cysteine building blocks as well for formation of the thiazole ring [91]. Finally, such heterocycles may be further process by oxidation domains (Ox) as reported for epothilon [92], myxothiazol [93], or tubulysin [94] assembly.

All these biosynthetic options allow the modification of the molecule's core scaffold, which also contributes to the product diversity. As discussed earlier, PKSs pathways as well contain specialized domains which modify their core compounds in order to expand the product variety. However, the combination of PKSs and NRPSs machineries within a concrete biosynthetic route facilitates the construction of an even greater compound spectrum than one of the systems could produce solely.

NRPS/PKS hybrid pathways

Bleomycin, the potent antitumor drug produced by *Streptomyces verticillus* ATCC15003, is a very good example of a natural product derived by a NRPS/PKS hybrid pathway. As reasoned from numerous feeding experiments, bleomycin is composed out of nine amino acids, one molecule of acetate, and two methyl groups that are provided by SAM [95]. Accordingly, in 1999 Shen *et al.* proposed that bleomycin is made by an assembly line comprising features of both, NRPSs and PKSs systems. Hereupon, they cloned and sequenced the putative *ble* gene cluster, thereby confirming their initial assumption [95;96]. Around the same time, Paitain *et al.* investigated the biosynthesis of the antibacterial compound TA made by *M. xanthus* [97]. In the course of their studies, they sequenced and analyzed the initial part of the corresponding gene cluster, and revealed that the first gene (*ta1*) encodes a NRPS/PKS hybrid protein which is

responsible for loading and extension of the glycine derived starter unit. Another myxobacterial NRPS/PKS hybrid pathway, as well published in 1999, is that of myxothiazol [93]. Besides a detailed scheme of how the compound is assembled, this study describes the first NRPS/PKS hybrid gene cluster in its entire sequence. These primal investigations firmly established the natural existence of NRPS/PKS hybrid systems. Today, a huge number of those hybrid pathways is known, which among others account for the production of medically important molecules such as epothilon [92], curacin [98], FK506 [99], and rapamycin [100].

The fact that NRPS and PKS systems use compatible buildings blocks, that are bound to the Ppant arm of the respective carrier proteins, promoted the evolution of hybrid pathways, featuring units of both machineries [75]. The impact of such interplay for the evolution of natural products is illustrated best by some numbers. As a pathway consisting of three basic and distinct randomly arranged NRPS modules can yield in one out of 500^3 (125,000,000) possible products (~ 500 amino acids are available to NRPSs [81]), a likewise PKS assembly line would generate one out of 6^3 (216) possible molecules (when three different out of the six most common building blocks are used). Notably, these numbers are considered without any modification of the building blocks. The addition of only one PKS or NRPS module at the very end of the respective pathways would increase the amount of putative molecules by the number of six (750,000,000) and 500 (108,000), respectively. Comparable considerations are the foundation of so called combinatorial biosynthesis approaches, which aim to generate tailor made molecules by combining distinct modules into novel assembly lines [101-103].

Chain termination in NRPS and PKS pathways: generating diversity through different release mechanisms

An essential step in the biosynthesis of every NRPSs or PKSs derived secondary metabolite is the release of the final product from the last carrier protein of the assembly line. Not only that this process makes the compound bioavailable and allows the assembly of further molecules, it also displays an opportunity to significantly alter the chemical nature of the molecule [72;104]. Most of the biosynthetic pathways described to date employ a terminal thioesterase domain (TE) in order to set free their final products (Figure 5). Such domains contain a catalytic triad (Ser-His-Asp) essential for product release [105]. In particular, the Asp stabilizes the His residue which accepts a proton from the Ser.

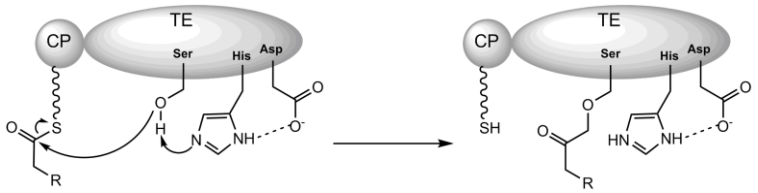


Figure 5: Schematic illustration showing the first step of TE mediated chain termination: The formation of the peptidyl- or acyl ester with the conserved serine from the TE. CP: carrier protein, TE: thioesterase.

Hereupon, the nucleophilic oxygen from the Ser attacks the carbonyl functionality from the carrier bound intermediate and forms a peptidyl or acyl ester. Depending on the characteristics of the TE domain, the respective ester is either attacked by an external (H_2O) or an internal (usually NH_2 or OH) nucleophile, which results in the formation of the linear carboxylic acid or the macrocycle, respectively. Figure 6 summarizes those possibilities by giving some examples. Formation of the linear carboxylic acid is observed for bacillaene [106] and the penicillin precursor, the ACV tripeptide [107]. The process of macrocyclization can yield the macrolacton as presented for thuggacin [108], or as in the case of microsclerodermin the macrolactam [68].

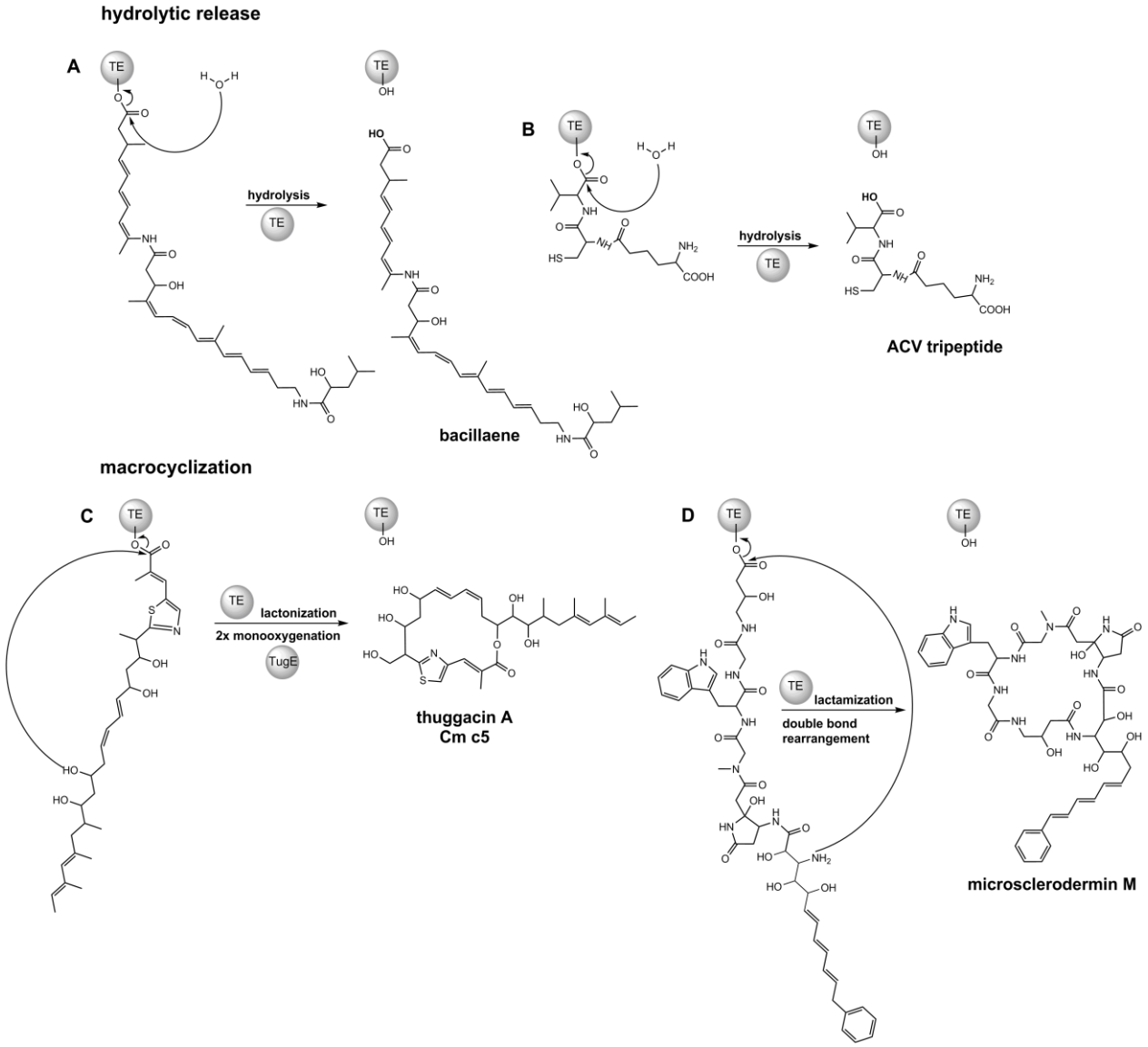


Figure 6: Illustration of TE catalyzed chain termination. Hydrolytic release as found in **A:** bacillaene [106] and **B:** ACV tripeptide biosynthesis results in the linear carboxylic acid [107], whereas macrocyclization can yield **C:** the lacton as observed in thuggacin biosynthesis [108], or **D:** the lactam as shown for microsclerodermin [68]. TE: thioesterase

Noteworthy, macrocycles can also originate from a thiol group attacking the carbonyl moiety. This reaction leads to the formation of a thiolacton and depicts the last step in the biosynthesis of thiocoraline [109]. Another instance of a TE catalyzed chain termination was recently described by Gavalda *et al.* [110]. They reported that the thioesterase-like domain of FAAL32-Pks13 from *Mycobacterium tubeculosis* exhibits acyltransferase activity *in vitro*. Accordingly, it is believed that compound release is accomplished by directly transferring the assembled products onto a freely diffusible molecule of trehalose thereby forming the respective

ester. The last-mentioned examples are, however, rather exotic but nevertheless they illustrate the broad spectrum of activities certain TEs can exert.

In addition to TEs, the so called reductase domains (Red) are as well able to discharge the full length product from its corresponding carrier protein. First revealed by examining lysine biosynthesis in *Saccharomyces cerevisiae*, Red domains catalyze the NAD(P)H+H^+ dependent reductive release of a dedicated product from its cognate carrier protein [111] (Figure 7A).

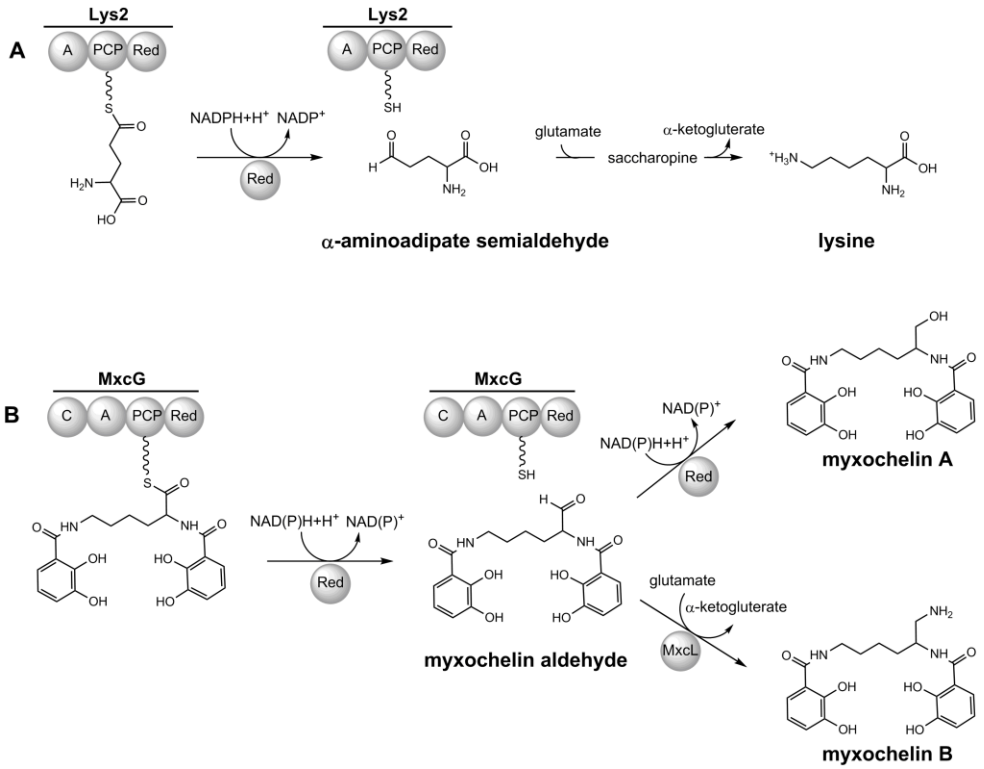


Figure 7: Examples of Red domain catalyzed chain termination. A: The release of α-aminoadipyl-PCP results in formation of the aldehyde. Subsequently, this intermediate undergoes reductive transamination to form the final product lysine [111]. **B:** The assembled compound is released from the corresponding PCP. Thereafter the aldehyde is further reduced to yield myxochelin A, or reductively transaminated giving rise to myxochelin B [112;113]. A: adenylation domain, C: condensation domain, PCP: peptidyl carrier protein, Red: reductase domain.

The first case of a Red domain catalyzed chain termination in natural product synthesis was described for myxochelin biosynthesis [112;113] (Figure 7 B). Here, the assembled product is released through a two electron reduction of the thioester, thereby forming the aldehyde. Next, this intermediate is either further reduced to yield the alcohol (myxochelin A), or reductively transaminated by MxcL, generating myxochelin B. More examples of a Red domain mediated compound release are found in xenovulene [114], myxalamid [115] and lyngbyatoxin [116]

biosynthesis. Common to all these biosynthetic pathways is the fact that the arising aldehyde is further processed in order to avoid undesirable side reactions.

Besides TEs and Red domains, C domains can as well discharge a molecule from a carrier protein (Figure 8). As discussed above, C domains typically promote the extension of biosynthetic intermediates by catalyzing amide bond formation between the growing chain and an extender amino acid. Exactly this ability enables those domains to take part in the release of numerous compounds. In particular, they facilitate intra- or intermolecular amide bond formation, thereby cleaving the respective thioester eventually resulting in liberation of the product. Figure 8A shows the last step of cyclosporine formation, in which a terminal C domain (C_T) catalyzes the nucleophilic attack of the amino functionality of D-Ala to the carbonyl group of L-Ala to yield the cyclic molecule. This mechanism is usually found in NRPSs from fungi, whereas bacterial machineries utilize TEs to cyclize their products [117]. Noteworthy, the nucleophile is not necessarily a part of the carrier bound intermediate, but may also be provided by an external acceptor substrate, as presented for aeruginosin biosynthesis [118] (Figure 8B). The last module (AerG) of the corresponding assembly line consists of the domain order C_1-A-PCP-C_2-PCP. It is responsible for the incorporation of the 2-carboxy-6-hydroxyoctahydroindole (Choi) building block (C_1-A-PCP) and furthermore it is believed that C_2 couples an external agmatine to the carboxyl group of the Choi moiety, thereby setting free the product. Subsequently, the molecule is glycosylated to yield one of the main derivatives, aeruginoside 126B. Another instance of a likewise chain release mechanism is described for streptothricin biosynthesis, where a distinct C domain condensates PCP bound L-β-lysine oligopeptides to an external substrate in order to liberate the products [119].

Besides amide bond formation, in some rare cases, esterification, catalyzed by C domains, displays a further opportunity to terminate compound assembly. The first example thereof is described in the context of fumonisin production [120]. In particular, this study suggests that the C domain of the bimodular protein Fump14 (PCP-C) mediates the release of aconityl-PCP intermediates by promoting esterification with a reduced acylic 18-carbon chain to form didehydro-fumonisin (Figure 8C). The second and currently last example of a C domain that sets free a PCP-bound molecule upon ester formation is SgcC5, taking part in the biosynthesis of the antitumor drug C-1027 [121]. However, in addition to esterification, *in vitro* experiments with the heterologously produced protein have shown that SgcC5 is still capable of building amide bonds. According to this finding, it is believed that SgcC5 provides an evolutionary link between amide and ester forming C domains.

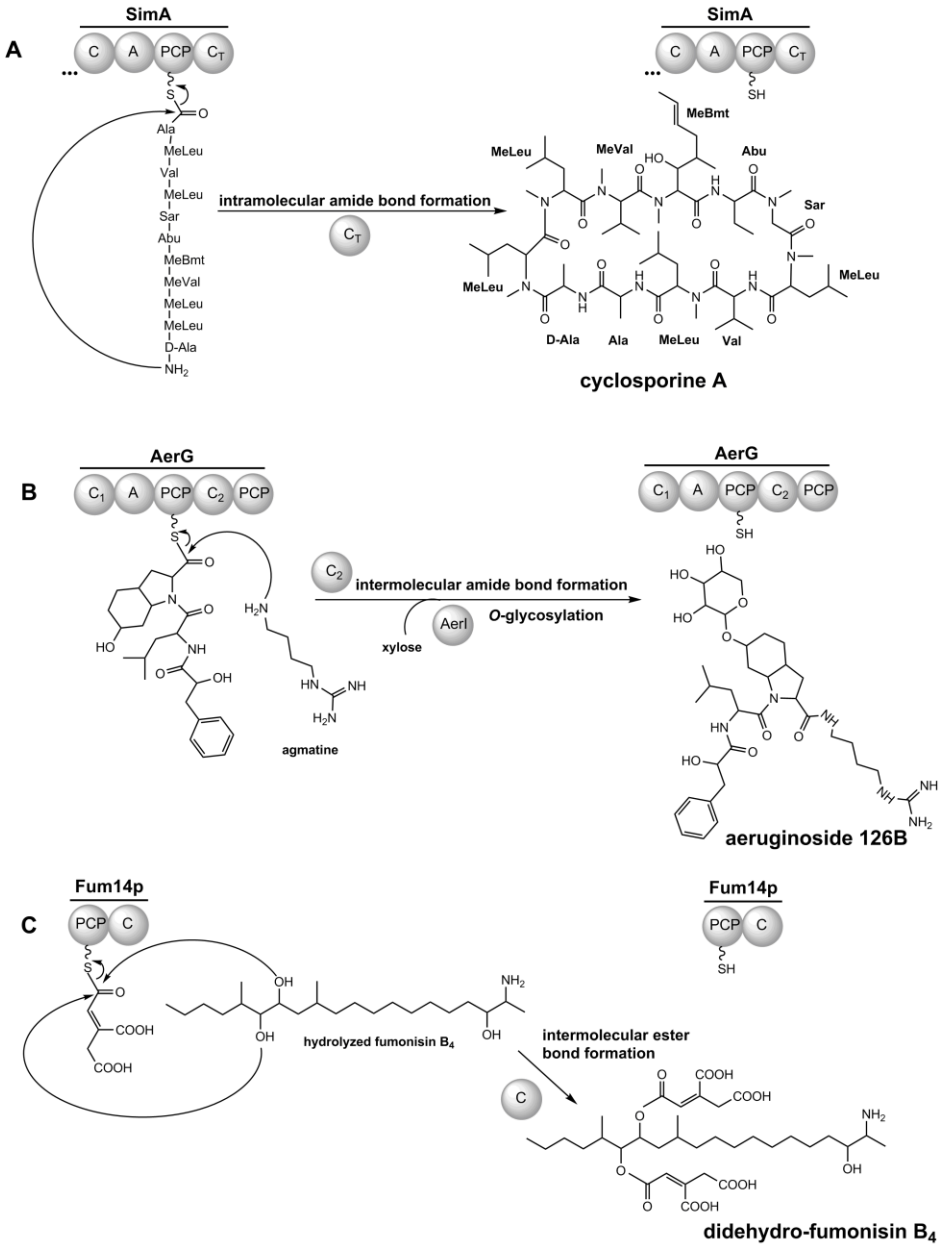


Figure 8: Examples of C domain catalyzed chain termination. A: intramolecular amide bond formation catalyzed by the terminal C domain (C_T) from the cyclosporine synthetase (SimA), resulting in the release of the cyclic product [104]. **B:** intermolecular amide bond formation between the PCP bound intermediate of the aeruginosides and a free agmatine. The released product is further glycosylated to yield aeruginoside 126B [118]. **C:** intermolecular ester bond formation between Fum14p attached aconitic acid and the acceptor substrate [120]. A: adenylation domain, C: condensation domain, PCP: peptidyl carrier protein, Red: reductase domain.

In summary, nature has developed several strategies to terminate PKS and NRPS assembly. Most commonly, TEs are employed which either hydrolyze the thioester or promote macrocyclization of the molecule. An alternative strategy is the reductive cleavage of the thioester bond by the action of Red domains. Finally, C domains and variants of it were shown to promote product liberation via amide or ester bond formation. Though, in the course of one

project presented in this thesis (Chapter I), a terminal C domain variant from the crocacin biosynthetic pathway was examined that shows hydrolytic activity towards the final product. Hence, this study adds a third and unprecedented function to this domain type and although the molecular basis behind the reaction remains elusive, this particular finding may allow engineering of NRPS assembly lines towards a programmed template release.

Strain improvement

In the natural environment, secondary metabolites are synthesized to confer selective advantages for the producer. Since their assembly is often associated with a high degree of energy and resource consumption, an organism tends to produce just as much of it as it needs to maintain these benefits. Under laboratory conditions, typically low production titers are observed which, from a biotechnological point of view, are far too low for large scale industrial production. Accordingly, once a potent compound of potential commercial interest has been identified, huge efforts are undertaken in order to improve the production properties of the particular organism or to find an alternative producer. Often the outcome of those experiments decides if a certain molecule will make it to the market or not. In this context, a very good example for such a progression is given by the development of the antibiotic penicillin. Initially the *P. notatum* strain isolated by Alexander Fleming in 1929 synthesized only poor amounts of penicillin (1.2 mg/l). Still, these amounts were enough to carry out early investigations about its effectiveness, including the vital experiments conducted by Howard Walter Florey, which showed that penicillin could protect mice against infection from deadly *Streptococci*. The isolation of *P. chrysogenum* NRRL-1951 in 1943, which synthesized more than 100 times the initial amount (150 mg/l), paved the way for large scale industrial production [122]. In order to further increase the yield, strain NRRL-1951 was subjected to X-ray mutagenesis resulting in the X1612 mutant (300 mg/l), which in turn was mutagenized again (Q176; 550 mg/l) and again (Wis 54-1255; >600mg/l) [123;124]. Based on Wis 54-1255, laboratories around the world created further optimized versions and recent industrial strains are estimated to yield up to 100,000 times more penicillin than the original producer *P. notatum* [125]. This story impressively illustrates how powerful strain development can be. However, this example should not imply that similar progress can be made with each strain and every compound.

The umbrella term strain improvement covers a variety of methods which can be applied to further increase the ability of a strain to synthesize a certain compound. Generally two fundamental approaches can be distinguished; the random mutagenesis and the rational mutagenesis.

The random approach uses different mutagenic agents such as UV or X-ray irradiation, antibiotics (rifampicin, spectinomycin, or streptomycin), and other chemicals (ethyl methane sulfonat, or nitroso methyl guanidine), to randomly alter the DNA of the strain of interest [126;127]. Subsequently, a panel of mutants is screened for their productivity, and the clones which exhibit the highest production yields are subjected to further rounds of treatment. The big advantage of this approach is that neither knowledge about the genetic and biochemical foundation of the pathway nor about tools to manipulate the organism is needed; it is simply enough to carry out several rounds of mutagenesis, cultivation and screening. Despite that, the whole procedure can be very resource consuming and the chances to obtain a strain with better production yields are estimated to be 1 per 10^4 to 10^{10} [128].

With advances in genetics and the increased understanding of even complex metabolic pathways, today, the rational approach depicts a powerful opportunity to construct strains with a significantly improved metabolite production profile. This may involve diverse strategies, such as increasing the precursor supply, overproducing or increasing the efficiency of bottleneck enzymes, altering the regulation of gene expression, reducing flux toward unwanted by-products or competing pathways, and reconstitution of entire pathways in a heterologous host [129].

The first decision one has to make is whether to optimize the native producer or to transfer the whole pathway into a heterologous host. From an industrial perspective, the second option may be preferred if the native producer is slow growing, genetically intractable, or generally hard to handle. The choice of an appropriate host is often determined by the pathway itself; usually a phylogenetically related organism will be chosen to avoid problems with for example codon usage and/or regulatory elements, but a tailor made synthetic gene cluster might be considered as well. If the native producer is selected for optimization, the procedure should start from the best producer available.

The strategy to increase the supply of precursors can be applied to both the native and the heterologous host. If the precursors needed are rather simple like malonyl-CoA as found in most of the PKS pathways, the overproduction of the acetyl-CoA carboxylase (ACC) complex can promote enhanced compound production. In a similar manner Ryu *et al.* achieved a 6-fold increase of actinorhodin production in *S. coelicolor* [130]. Comparable results were obtained by boosting supply of the methylmalonyl-CoA metabolite node in *Saccharopolyspora erythraea* [131]. Besides that, a study of Han *et al.* reported that heterologous production of propionyl-CoA synthetase from *Ralstonia solanacearum* in *S. cellulosum* So ce90 leads to a 100-fold increase of epothilon B production [132]. In case of NRPS or hybrid pathways, the increase of precursor supply is more difficult. This is owing to the fact, that the amino acid metabolism is more complex than that of short chain carboxylic acids. However, to achieve an enhanced production

of secondary metabolites that consist in large parts of entities derived from the shikimate pathway (e.g. aromatic amino acids, or 2,3-dihydroxybenzoic), the overproduction of 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase, the first enzyme of the shikimate pathway, is a promising option [133;134]. Besides an enhancement of precursor supply through directed overexpression of relevant genes, the deletion of competing metabolic pathways depicts an additional strategy to provide extra quantities of building blocks. After all, a combination of both strategies might be the most promising approach.

If precursor supply does not seem to be the limiting factor, an increase in the gene dosage of the respective secondary metabolite pathway can be taken into account. Today, many strains originally generated in the course of random mutagenesis and screening campaigns have been examined in order to reveal the basis for their extraordinary production properties. As a result, in several of these organisms, multiple copies (up to 16) of at least parts of the gene clusters have been detected [135-137]. Nowadays, such duplications can be easily introduced into the genome of a strain either randomly by transposition, or site-directed through homologous recombination. Notwithstanding, a simple exchange of the promotor upstream of the gene cluster against a strong and constitutive one can sometimes have a great impact on compound production, as shown for thuggacin and crocapeptin biosynthesis [108;138].

The manipulation of pathway specific regulators has proven to be another suitable method to generate potent overproducers. In *Streptomyces* species, these proteins are often encoded within the corresponding gene clusters and most of them were shown to stimulate gene expression. Consequently an overproduction results in an elevated compound formation [139]. In contrast, regulatory elements that direct secondary metabolite biosynthesis in myxobacteria do not tend to be co-localized with the gene clusters. A straightforward strategy to identify potential regulators represents the method of so called “promotor fishing”. This procedure uses biotinylated DNA fragments bearing the presumable promotor upstream of the cluster to literally fish out proteins which specifically bind to this area. Using this technique, two different regulatory proteins, ChiR and NtcA, that are involved in the regulation of the chivosazole gene cluster in *S. cellulosum* So ce56 could be determined [140;141]. ChiR was shown to be a positive regulator; hence the overexpression of the respective gene in a meroploid mutant caused a 5-fold increase in chivosazol production, notably without affecting the yields of other secondary metabolites. NtcA turned out to negatively affect chivosazol production since disruption of the gene by single cross over led to an improved production of chivosazol (4-fold) and also etnangien (3.5-fold). As described in Chapter III of this thesis, a likewise practice was employed in order to identify regulators of ambruticin production in *S. cellulosum* strain #8405.

Taken together, several methods exist which can be applied in order to generate potential overproducers of a respective compound. However, especially when antimicrobial molecules should be synthesized at high titers in the native producer, the inherent resistance mechanism might not be sufficient anymore to detoxify the elevated intracellular levels of the substance. Such a scenario was observed in the context of doxorubicin biosynthesis, where the overexpression of the resistance genes was essential to allow further yield optimization [142]. As illustrated by the production of the antibiotic bottromycin A2 in *S. coelicolor*, self-resistance needs to be considered as well in heterologous expression systems [143]. In particular, the targeted overexpression of the *botT* gene, encoding an efflux pump, enhanced the production significantly in the heterologous host. These examples clearly show that the issue of self-defense should be taken into account while optimizing a strain towards higher production yields.

Outline of the dissertation

This thesis comprises studies on two fundamental areas of natural product research, the elucidation of biosynthetic pathways underlying secondary metabolite formation and the improvement of a bacterial strain towards the enhanced production of a specific secondary metabolite. While additional knowledge about the molecular basis of compound assembly, significantly accounts for the understanding of how nature generates structural diversity, yield optimization is a necessity in order to make a potent metabolite commercially available. In particular, the biosynthesis of crocacin (Chapter I) and microsclerodermin (Chapter II) was investigated, thereby revealing several unusual features of NRPS/PKS-based biosynthesis. The project depicted in Chapter III dealt with the identification and optimization of strains producing the potent fungicide ambruticin, which is of notable industrial interest.

The crocacins are produced by the myxobacterium *Chondromyces crocatus* Cm c5 and exhibit strong antifungal activity. Although their chemical structure has been elucidated and several total synthesis approaches have been reported, the actual biosynthetic route to the linear crocacin scaffold remained unknown. Based on a retrobiosynthetic suggestion the potential gene cluster bearing the information for crocacin assembly was detected in a draft genome of *C. crocatus* Cm c5 and finally, the targeted inactivation of an associated NRPS gene confirmed this assignment. Hereupon several analytic, genetic and biochemical experiments were conducted which allowed a detailed proposal of how the crocacins are built up in nature. The dedicated biosynthetic pathway features an iteratively acting PKS module, an atypical β-branching cassette and most strikingly a terminal condensation domain releasing the assembled compound from the terminal carrier protein by hydrolysis. Especially the last finding sets the

stage for continuative mechanistic investigations and this knowledge might eventually be used to generate shunt products of assembly lines comprising discrete NRPS modules.

The second part of this thesis covers the identification and analysis of terrestrial producers of the microsclerodermins, a compound class which has been initially isolated from marine sponges of the genus *Microscleroderma*. In particular, the evaluation of LC-MS profiles from approximately 800 myxobacterial extracts led to the identification of 15 distinct strains, all of which produce known and unknown derivatives of this compound class. While all *Sorangium* species synthesize the new derivative microsclerodermin M, representatives of the genera *Jahnella* and *Chondromyces* were found to produce, among others, the “marine” derivative microsclerodermin D. This discovery reflects one of the still rare cases, in which an identical natural product was isolated from a terrestrial and marine a source. Moreover, the corresponding biosynthetic gene clusters from *S. cellulosum* So ce38 and *Jahnella* sp. MSr9139 were identified and a comprehensive analysis of both displayed that variations found in the derivatives can be directly correlated to differences in the biosynthetic machinery. With the identification of bacterial producers we provided a sustainable source for large scale isolation and further exploration of the microsclerodermin compound class.

The last project aimed to generate a rationally engineered strain that produces enhanced amounts of the fungicide ambruticin. To begin with the best producer available to us, a search across a high resolution LC-MS data set comprising the profiles of 757 extracts from various myxobacterial genera was conducted. As a result, it turned out that *S. cellulosum* strain #8405 is the best producer of VS- and *S. cellulosum* strain #7282 of F-ambruticin, two slightly different metabolites differing only by the substitution of an OH (F ambruticin) with an NH_2 moiety (VS ambruticins). Subsequently, genetic methods for the deletion and overexpression of genes were developed. Using these techniques the function of several genes belonging to the ambruticin gene cluster of *S. cellulosum* strain #8405 were evaluated. This approach led to the identification of a putative AtoC-like regulatory protein (Amb04) and a putative hybrid sensor kinase response regulator (Amb07), which are involved in directing ambruticin biosynthesis. In addition, the consensus motif for DNA binding of the AtoC-like regulatory protein was determined, which allowed the postulation of its regulon. Moreover, another potential regulatory element was detected using promotor fishing (NtcA). In summary, these experiments provide a promising basis for the rational optimization of *S. cellulosum* strain #8405 towards the enhanced production of ambruticins.

References

[1] Editorial (2007) *Nat. Chem. Biol.,* 3, 351.
[2] Mishra BB & Tiwari VK (2011) Natural products: An evolving role in future drug discovery. *Eur. J. Med. Chem.*, 46, 4769-4807.
[3] Cragg GM & Newman DJ (2013) Natural products: A continuing source of novel drug leads. *Biochim. Biophys. Acta*, 1830, 3670-3695.
[4] Cordero OX, Wildschutte H, Kirkup B, Proehl S, Ngo L, Hussain F, Le RF, Mincer T, & Polz MF (2012) Ecological populations of bacteria act as socially cohesive units of antibiotic production and resistance. *Science*, 337, 1228-1231.
[5] Williams ST & Vickers JC (1986) The ecology of antibiotic production. *Microb. Ecol.*, 12, 43-52.
[6] Goh EB, Yim G, Tsui W, McClure J, Surette MG, & Davies J (2002) Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics. *Proc. Natl. Acad. Sci. U SA*, 99, 17025-17030.
[7] Linares JF, Gustafsson I, Baquero F, & Martinez JL (2006) Antibiotics as intermicrobial signaling agents instead of weapons. *Proc. Natl. Acad. Sci. USA*, 103, 19484-19489.
[8] Meiser P, Bode HB, & Müller R (2006) The unique DKxanthene secondary metabolite family from the myxobacterium *Myxococcus xanthus* is required for developmental sporulation. *Proc. Natl. Acad. Sci. USA*, 103, 19128-19133.
[9] Hoffman LR, D'Argenio DA, MacCoss MJ, Zhang Z, Jones RA, & Miller SI (2005) Aminoglycoside antibiotics induce bacterial biofilm formation. *Nature*, 436, 1171-1175.
[10] Plaga W, Stamm I, & Schairer HU (1998) Intercellular signaling in *Stigmatella aurantiaca*: Purification and characterization of stigmolone, a myxobacterial pheromone. *Proc. Natl. Acad. Sci. USA*, 95, 11263-11267.
[11] Davies J & Ryan KS (2012) Introducing the parvome: Bioactive compounds in the microbial world. *ACS Chem. Biol.*, 7, 252-259.
[12] Cragg GM & Newman DJ (2005) Biodiversity: A continuing source of novel drug leads. *Pure Appl. Chem.*, 77, 7-24.
[13] Kinghorn AD, Pan L, Fletcher JN, & Chai HY (2011) The relevance of higher plants in lead compound discovery programs. *J. Nat. Prod.*, 74, 1539-1555.
[14] Lorenzen K & Anke T (1998) *Basidiomycetes* as a source for new bioactive natural products. *Curr. Org. Chem.*, 2, 329-364.
[15] Fleming A (1929) On the antibacterial action of a *Penicillium*, with special reference to their use in the isolation of *B. influenzae Br. J. Exp. Pathol.*, 10, 226–236.
[16] Raju TN (1999) The Nobel chronicles. 1945: Sir Alexander Fleming (1881-1955); Sir Ernst Boris Chain (1906-79); and Baron Howard Walter Florey (1898-1968). *Lancet*, 353, 936.
[17] Dubos RJ & Hotchkiss RD (1941) The production of bactericidal substances by aerobic sporulating *Bacilli*. *J. Exp. Med.*, 73, 629-640.
[18] Dubos RJ (1939) Studies on a bactericidal agent extracted from a soil *Bacillus*: II. protective effect of the bacteriocidal agent against experimental *Pneumococcus* infections in mice. *J. Exp. Med.*, 70, 11-17.
[19] Dubos RJ (1939) Studies on a bactericidal agent extracted from a soil *Bacillus*: I. preparation of the agent. Its activity in vitro. *J. Exp. Med.*, 70, 1-10.
[20] Bérdy J (2005) Bioactive microbial metabolites. *J. Antibiot.*, 58, 1-26.
[21] Newman DJ & Cragg GM (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.*, 75, 311-335.
[22] Newman DJ & Cragg GM (2007) Natural products as sources of new drugs over the last 25 years. *J. Nat. Prod.*, 70, 461-477.
[23] Katz E & Demain AL (1977) Peptide antibiotics of *Bacillus* - chemistry, biogenesis, and possible functions. *Bacteriol. Rev.*, 41, 449-474.
[24] Stein T (2005) *Bacillus subtilis* antibiotics: Structures, syntheses and specific functions. *Mol. Microbiol.*, 56, 845-857.
[25] Reichenbach H, Gerth K, Irschik H, & Kunze B (1988) Myxobacteria: A source of new antibiotics. *Trends Biotechnol.*, 6, 115-121.
[26] Reichenbach H & Höfle G (1999) Myxobacteria as producers of secondary metabolites. In *Drug Discovery from Nature* (Grabley S & Thiericke R, eds), pp. 149-179. Springer, Berlin.
[27] Wenzel SC & Müller R (2009) The biosynthetic potential of myxobacteria and their impact on drug discovery. *Curr Opin Drug Discov Devel*, 12, 220-230.
[28] Dias DA, Urban S, & Roessner U (2012) A historical overview of natural products in drug discovery. *Metabolites.*, 2, 303-336.
[29] Gerwick WH & Moore BS (2012) Lessons from the past and charting the future of marine natural products drug discovery and chemical biology. *Chem. Biol.*, 19, 85-98.
[30] Bewley CA & Faulkner DJ (1998) Lithistid sponges: Star performers or hosts to the stars. *Angew. Chem. Int. Ed. Engl.*, 37, 2162-2178.
[31] Martins A, Vieira H, Gaspar H, & Santos S (2014) Marketed marine natural products in the pharmaceutical and cosmeceutical industries: Tips for success. *Mar. Drugs*, 12, 1066-1101.
[32] Hentschel U, Hopke J, Horn M, Friedrich AB, Wagner M, Hacker J, & Moore BS (2002) Molecular evidence for a uniform microbial community in sponges from different oceans. *Appl. Environ. Microbiol.*, 68, 4431-4440.
[33] Hentschel U, Usher KM, & Taylor MW (2006) Marine sponges as microbial fermenters. *FEMS Microbiol. Ecol.*, 55, 167-177.
[34] Simister RL, Deines P, Botté ES, Webster NS, & Taylor MW (2012) Sponge-specific clusters revisited: A comprehensive phylogeny of sponge-associated microorganisms. *Environ. Microbiol.*, 14, 517-524.
[35] Fenical W & Jensen PR (2006) Developing a new resource for drug discovery: Marine actinomycete bacteria. *Nat. Chem. Biol.*, 2, 666-673.
[36] Schaberle TF, Goralski E, Neu E, Erol O, Holzl G, Dormann P, Bierbaum G, & Konig GM (2010) Marine myxobacteria as a source of antibiotics - comparison of physiology, polyketide-type genes and antibiotic production of three new isolates of *Enhygromyxa salina*. *Mar. Drugs*, 8, 2466-2479.
[37] Baltz RH (2008) Renaissance in antibacterial discovery from actinomycetes. *Curr. Opin. Pharm.*, 8, 557-563.
[38] Shimkets L, Dworkin M, & Reichenbach H (2006) The Myxobacteria. In *The Prokaryotes* (Dworkin M, ed), pp. 31-115. Springer, Berlin.
[39] Errington J (2003) Regulation of endospore formation in *Bacillus subtilis*. *Nat. Rev. Microbiol.*, 1, 117-126.
[40] Garcia RO, Krug D, & Müller R (2009) Discovering natural products from myxobacteria with emphasis on rare producer strains in combination with improved analytical methods. *Methods Enzymol.,* 458, 59-91.
[41] Zusman DR, Scott AE, Yang Z, & Kirby JR (2007) Chemosensory pathways, motility and development in *Myxococcus xanthus*. *Nat. Rev. Microbiol.*, 5, 862-872.

[42] Goldman BS, Nierman WC, Kaiser D, Slater SC, Durkin AS, 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 (2006) Evolution of sensory complexity recorded in a myxobacterial genome. *P. Natl. Acad. Sci. USA*, 103, 15200-15205.

[43] Bode HB & Müller R (2006) Analysis of myxobacterial secondary metabolism goes molecular. *J. Ind. Microbiol. Biotechnol.*, 33, 577-588.

[44] 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, Munoz-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 FJ, Weissman KJ, Welch RD, Wenzel SC, Whitworth DE, Wilhelm S, Wittmann C, Blöcker H, Pühler A, & Müller R (2007) Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat. Biotechnol.*, 25, 1281-1289.

[45] Han K, Li ZF, Peng R, Zhu LP, Zhou T, Wang LG, Li SG, Zhang XB, Hu W, Wu ZH, Qin N, & Li YZ (2013) Extraordinary expansion of a *Sorangium cellulosum* genome from an alkaline milieu. *Sci. Rep.*, 3, 1-7.

[46] Weissman KJ & Müller R (2010) Myxobacterial secondary metabolites: Bioactivities and modes-of-action. *Nat. Prod. Rep.*, 27, 1276-1295.

[47] Gerth K, Pradella S, Perlova O, Beyer S, & Müller R (2003) Myxobacteria: Proficient producers of novel natural products with various biological activities - past and future biotechnological aspects with the focus on the genus *Sorangium*. *J. Biotechnol.*, 106, 233-253.

[48] Gerth K, Bedorf N, Höfle G, Irschik H, & Reichenbach H (1996) Epothilons A and B: Antifungal and cytotoxic compounds from *Sorangium cellulosum* (Myxobacteria). Production, physico-chemical and biological properties. *J. Antibiot.*, 49, 560-563.

[49] Ringel SM, Greenough RC, Romm S, Conor D, Gutt AL, Blair B, Kanter G, & von Strandtmann M (1977) Ambruticin (W7783), a new antifungal antibiotic. *J. Antibiot.*, 30, 371-375.

[50] Irschik H, Augustiniak H, Gerth K, Höfle G, & Reichenbach H (1995) The ripostatins, novel inhibitors of eubacterial RNA polymerase isolated from myxobacteria. *J. Antibiot.*, 48, 787-792.

[51] Gerth K, Bedorf N, Irschik H, Höfle G, & Reichenbach H (1994) The soraphens: A family of novel antifungal compounds from *Sorangium cellulosum* (Myxobacteria). I. Soraphen A1 alpha: Fermentation, isolation, biological properties. *J. Antibiot.*, 47, 23-31.

[52] Jansen R, Washausen P, Kunze B, Reichenbach H, & Hofle G (1999) Antibiotics from gliding bacteria, LXXXIII - The crocacins, novel antifungal and cytotoxic antibiotics from *Chondromyces crocatus* and *Chondromyces pediculatus* (Myxobacteria): Isolation and structure elucidation. *Eur. J. Org. Chem.*, 1085-1089.

[53] Steinmetz H, Irschik H, Kunze B, Reichenbach H, Höfle G, & Jansen R (2007) Thuggacins, macrolide antibiotics active against *Mycobacterium tuberculosis*: Isolation from myxobacteria, structure elucidation, conformation analysis and biosynthesis. *Chemistry*, 13, 5822-5832.

[54] Trowitzsch W, Wray V, Gerth K, & Höfle G (1982) Structure of Myxovirescin A, a new macrocyclic antibiotic from gliding bacteria. *J. Chem. Soc., Chem. Commun.*, 23, 1340-1342.

[55] Irschik H, Trowitzsch-Kienast W, Gerth K, Höfle G, & Reichenbach H (1988) Saframycin Mx1, a new natural saframycin isolated from a myxobacterium. *J. Antibiot.*, 41, 993-998.

[56] Altmann KH, Gaugaz FZ, & Schiess R (2011) Diversity through semisynthesis: The chemistry and biological activity of semisynthetic epothilone derivatives. *Mol. Divers.*, 15, 383-399.

[57] Singh N & Abraham J (2014) Ribosomally synthesized peptides from natural sources. *J. Antibiot.*, 67, 277-289.

[58] Revermann O (2012) Novel secondary metabolites from myxobacteria and their biosynthetic machinery. Ph.D. thesis, Universität des Saarlandes, Saarbrücken.

[59] Hertweck C (2009) The biosynthetic logic of polyketide diversity. *Angew. Chem Int. Ed Engl.*, 48, 4688-4716.

[60] Staunton J & Weissman KJ (2001) Polyketide biosynthesis: A millennium review. *Nat. Prod. Rep.*, 18, 380-416.

[61] Hopwood DA (1997) Genetic contributions to understanding polyketide synthases. *Chem Rev*, 97, 2465-2497.

[62] Rawlings BJ (2001) Type I polyketide biosynthesis in bacteria (Part A-erythromycin biosynthesis). *Nat. Prod. Rep.*, 18, 190-227.

[63] Moss SJ, Martin CJ, & Wilkinson B (2004) Loss of co-linearity by modular polyketide synthases: A mechanism for the evolution of chemical diversity. *Nat. Prod. Rep.*, 21, 575-593.

[64] Campbell CD & Vederas JC (2010) Biosynthesis of lovastatin and related metabolites formed by fungal iterative PKS enzymes. *Biopolymers*, 93, 755-763.

[65] Izumikawa M, Cheng Q, & Moore BS (2006) Priming type II polyketide synthases via a type II nonribosomal peptide synthetase mechanism. *J. Am. Chem. Soc.*, 128, 1428-1429.

[66] Perlova O, Gerth K, Hans A, Kaiser O, & Müller R (2006) Identification and analysis of the chivosazol biosynthetic gene cluster from the myxobacterial model strain *Sorangium cellulosum* So ce56. *J. Biotechnol.*, 121, 174-191.

[67] Ligon J, Hill S, Beck J, Zirkle R, Molnar I, Zawodny J, Money S, & Schupp T (2002) Characterization of the biosynthetic gene cluster for the antifungal polyketide soraphen A from *Sorangium cellulosum* So ce26. *Gene*, 285, 257-267.

[68] Hoffmann T, Müller S, Nadmid S, Garcia R, & Müller R (2013) Microsclerodermins from terrestrial myxobacteria: An intriguing biosynthesis likely connected to a sponge symbiont. *J. Am. Chem. Soc.*, 45, 16904-16911.

[69] Müller S, Rachid S, Hoffmann T, Surup F, Volz C, Zaburannyi N, & Müller R (2014) Biosynthesis of crocacin involves an unusual hydrolytic release domain showing similarity to condensation domains. *Chem. Biol.*, 21, 855-865.

[70] Chan YA, Podevels AM, Kevany BM, & Thomas MG (2009) Biosynthesis of polyketide synthase extender units. *Nat. Prod. Rep.*, 26, 90-114.

[71] Caffrey P (2003) Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. *ChemBioChem*, 4, 654-657.

[72] Du L & Lou L (2010) PKS and NRPS release mechanisms. *Nat. Prod. Rep.*, 27, 255-278.

[73] Jenke-Kodama H, Sandmann A, Müller R, & Dittmann E (2005) Evolutionary implications of bacterial polyketide synthases. *Mol. Biol. Evol.*, 22, 2027-2039.

[74] Jenke-Kodama H, Müller R, & Dittmann E (2008) Evolutionary mechanisms underlying secondary metabolite diversity. In *Natural Compounds as Drugs Volume I* (Petersen F & Amstutz R, eds), pp. 119-140. Birkhäuser, Basel.

[75] Du L & Shen B (2001) Biosynthesis of hybrid peptide-polyketide natural products. *Curr. Opin. Drug. Discov. Devel.*, 4, 215-228.

[76] Lipmann F, Gevers W, Kleinkauf H, & Roskoski R, Jr. (1971) Polypeptide synthesis on protein templates: The enzymatic synthesis of gramicidin S and tyrocidine. *Adv. Enzymol. Relat. Areas Mol. Biol.*, 35, 1-34.
[77] Sieber SA & Marahiel MA (2005) Molecular mechanisms underlying nonribosomal peptide synthesis: Approaches to new antibiotics. *Chem. Rev.*, 105, 715-738.
[78] Tanovic A, Samel SA, Essen LO, & Marahiel MA (2008) Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science,* 321, 659-663.
[79] Weber T & Marahiel MA (2001) Exploring the domain structure of modular nonribosomal peptide synthetases. *Structure*, 9, R3-R9.
[80] Strieker M, Tanovic A, & Marahiel MA (2010) Nonribosomal peptide synthetases: Structures and dynamics. *Curr. Opin. Struct. Biol.*, 20, 234-240.
[81] Walsh CT, O'Brien RV, & Khosla C (2013) Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angew. Chem. Int. Ed Engl.*, 52, 7098-7124.
[82] O'Donoghue P & Luthey-Schulten Z (2003) On the evolution of structure in aminoacyl-tRNA synthetases. *Microbiol. Mol. Biol. Rev.*, 67, 550-573.
[83] Stachelhaus T, Mootz HD, & Marahiel MA (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.*, 6, 493-505.
[84] Linne U & Marahiel MA (2004) Reactions catalyzed by mature and recombinant nonribosomal peptide synthetases. *Methods Enzymol.*, 388, 293-315.
[85] Belshaw PJ, Walsh CT, & Stachelhaus T (1999) Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science*, 284, 486-489.
[86] Fischbach MA & Walsh CT (2006) Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic, machinery, and mechanisms. *Chem. Rev.*, 106, 3468-3496.
[87] Stachelhaus T & Walsh CT (2000) Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. *Biochemistry*, 39, 5775-5787.
[88] Hoffmann K, Schneider-Scherzer E, Kleinkauf H, & Zocher R (1994) Purification and characterization of eucaryotic alanine racemase acting as key enzyme in cyclosporin biosynthesis. *J. Biol. Chem.*, 269, 12710-12714.
[89] Billich A, Zocher R, Kleinkauf H, Braun DG, Lavanchy D, & Hochkeppel HK (1987) Monoclonal antibodies to the multienzyme enniatin synthetase. Production and use in structural studies. *Biol. Chem. Hoppe Seyler*, 368, 521-529.
[90] Haese A, Pieper R, von OT, & Zocher R (1994) Bacterial expression of catalytically active fragments of the multifunctional enzyme enniatin synthetase. *J. Mol. Biol.*, 243, 116-122.
[91] Konz D, Klens A, Schorgendorfer K, & Marahiel MA (1997) The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC 10716: Molecular characterization of three multi-modular peptide synthetases. *Chem. Biol.*, 4, 927-937.
[92] Julien B, Shah S, Ziermann R, Goldman R, Katz L, & Khosla C (2000) Isolation and characterization of the epothilone biosynthetic gene cluster from *Sorangium cellulosum*. *Gene*, 249, 153-160.
[93] Silakowski B, Schairer HU, Ehret H, Kunze B, Weinig S, Nordsiek G, Brandt P, Blöcker H, Höfle G, Beyer S, & Müller R (1999) New lessons for combinatorial biosynthesis from myxobacteria: The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. *J. Biol. Chem.*, 274, 37391-37399.
[94] Sandmann A, Sasse F, & Müller R (2004) Identification and analysis of the core biosynthetic machinery of tubulysin, a potent cytotoxin with potential anticancer activity. *Chem. Biol.*, 11, 1071-1079.
[95] Shen B, Du L, Sanchez C, Chen MT, & Edwards DJ (1999) Bleomycin biosynthesis in *Streptomyces verticillus* ATCC15003: A model of hybrid peptide and polyketide biosynthesis. *Bioorganic Chem*, 27, 155-191.
[96] Shen B, Du LC, Sanchez C, Edwards DJ, Chen M, & Murrell JM (2002) Cloning and characterization of the bleomycin biosynthetic gene cluster from *Streptomyces verticillus* ATCC15003. *J. Nat. Prod.*, 65, 422-431.
[97] Paitan Y, Alon G, Orr E, Ron EZ, & Rosenberg E (1999) The first gene in the biosynthesis of the polyketide antibiotic TA of *Myxococcus xanthus* codes for a unique PKS module coupled to a peptide synthetase. *Journal of Molecular Biology*, 286, 465-474.
[98] Chang Z, Sitachitta N, Rossi JV, Roberts MA, Flatt P, Jia J, Sherman DH, & Gerwick WH (2004) Biosynthetic pathway and gene cluster analysis of Curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. *J. Nat. Prod.*, 67, 1356-1367.
[99] Motamedi H & Shafiee A (1998) The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506. *Eur. J. Biochem.*, 256, 528-534.
[100] Schwecke T, Aparicio JF, Molnar I, König A, Khaw LE, Haydock SF, Oliynyke M, Caffrey P, Cortes J, Lester JB, Bohm GA, Staunton J, & Leadley PF (1995) The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc. Natl. Acad. Sci. USA*, 92, 7839-7843.
[101] Giessen TW & Marahiel MA (2012) Ribosome-independent biosynthesis of biologically active peptides: Application of synthetic biology to generate structural diversity. *FEBS Lett.*, 586, 2065-2075.
[102] Baltz RH (2014) Combinatorial biosynthesis of cyclic lipopeptide antibiotics: A model for synthetic biology to accelerate the evolution of secondary metabolite biosynthetic pathways. *ACS Synth. Biol.*, 3, 748-758.
[103] Boettger D, Bergmann H, Kuehn B, Shelest E, & Hertweck C (2012) Evolutionary imprint of catalytic domains in fungal PKS-NRPS hybrids. *ChemBioChem*, 13, 2363-2373.
[104] Keating TA, Ehmann DE, Kohli RM, Marshall CG, Trauger JW, & Walsh CT (2001) Chain termination steps in nonribosomal peptide synthetase assembly lines: Directed acyl-S-enzyme breakdown in antibiotic and siderophore biosynthesis. *ChemBioChem*, 2, 99-107.
[105] Tsai SC, Lu HX, Cane DE, Khosla C, & Stroud RM (2002) Insights into channel architecture and substrate specificity from crystal structures of two macrocycle-forming thioesterases of modular polyketide synthases. *Biochemistry*, 41, 12598-12606.
[106] Butcher RA, Schroeder FC, Fischbach MA, Straight PD, Kolter R, Walsh CT, & Clardy J (2007) The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA*, 104, 1506-1509.
[107] Byford MF, Baldwin JE, Shiau CY, & Schofield CJ (1997) The mechanism of ACV synthetase. *Chem. Rev.*, 97, 2631-2650.

[108] Buntin K, Irschik H, Weissman KJ, Luxenburger E, Blöcker H, & Müller R (2010) Biosynthesis of thuggacins in myxobacteria: Comparative cluster analysis reveals basis for natural product structural diversity. *Chem. Biol.*, 17, 342-356.
[109] Robbel L, Hoyer KM, & Marahiel MA (2009) TioS T-TE - a prototypical thioesterase responsible for cyclodimerization of the quinoline- and quinoxaline-type class of chromodepsipeptides. *FEBS J.*, 276, 1641-1653.
[110] Gavalda S, Bardou F, Laval F, Bon C, Malaga W, Chalut C, Guilhot C, Mourey L, Daffe M, & Quemard A (2014) The polyketide synthase Pks13 catalyzes a novel mechanism of lipid transfer in mycobacteria. *Chem. Biol.*, 21, 1660-1669.
[111] Ehmann DE, Gehring AM, & Walsh CT (1999) Lysine biosynthesis in *Saccharomyces cerevisiae*: Mechanism of (alpha)-aminoadipate reductase (Lys2) involves posttranslational phosphopantetheinylation by Lys5. *Biochemistry*, 38, 6171-6177.
[112] Gaitatzis N, Kunze B, & Müller R (2001) In vitro reconstitution of the myxochelin biosynthetic machinery of *Stigmatella aurantiaca* Sg a15: Biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases. *Proc. Natl. Acad. Sci. USA*, 98, 11136-11141.
[113] Li Y, Weissman KJ, & Müller R (2008) Myxochelin biosynthesis: Direct evidence for two- and four-electron reduction of a carrier protein-bound thioester. *J. Am. Chem. Soc.*, 130, 7554-7555.
[114] Bailey AM, Cox RJ, Harley K, Lazarus CM, Simpson TJ, & Skellam E (2007) Characterisation of 3-methylorcinaldehyde synthase (MOS) in *Acremonium strictum*: First observation of a reductive release mechanism during polyketide biosynthesis. *Chem. Commun.*, 2007, 4053-4055.
[115] Silakowski B, Nordsiek G, Kunze B, Blöcker H, & Müller R (2001) Novel features in a combined polyketide synthase/non-ribosomal peptide synthetase: The myxalamid biosynthetic gene cluster of the myxobacterium *Stigmatella aurantiaca* Sga15. *Chem. Biol.*, 8, 59-69.
[116] Edwards DJ & Gerwick WH (2004) Lyngbyatoxin biosynthesis: Sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase. *J. Am. Chem. Soc.*, 126, 11432-11433.
[117] Gao X, Haynes SW, Ames BD, Wang P, Vien LP, Walsh CT, & Tang Y (2012) Cyclization of fungal nonribosomal peptides by a terminal condensation-like domain. *Nat. Chem. Biol.*, 8, 823-830.
[118] Ishida K, Christiansen G, Yoshida WY, Kurmayer R, Welker M, Valls N, Bonjoch J, Hertweck C, Borner T, Hemscheidt T, & Dittmann E (2007) Biosynthesis and structure of aeruginoside 126A and 126B, cyanobacterial peptide glycosides bearing a 2-carboxy-6-hydroxyoctahydroindole moiety. *Chem. Biol.*, 14, 565-576.
[119] Maruyama C, Toyoda J, Kato Y, Izumikawa M, Takagi M, Shin-ya K, Katano H, Utagawa T, & Hamano Y (2012) A stand-alone adenylation domain forms amide bonds in streptothricin biosynthesis. *Nat. Chem. Biol.*, 8, 791-797.
[120] Zaleta-Rivera K, Xu C, Yu F, Butchko RA, Proctor RH, Hidalgo-Lara ME, Raza A, Dussault PH, & Du L (2006) A bidomain nonribosomal peptide synthetase encoded by FUM14 catalyzes the formation of tricarballylic esters in the biosynthesis of fumonisins. *Biochemistry*, 45, 2561-2569.
[121] Lin S, Van Lanen SG, & Shen B (2009) A free-standing condensation enzyme catalyzing ester bond formation in C-1027 biosynthesis. *Proc. Natl. Acad. Sci. USA*, 106, 4183-4188.
[122] Ozcengiz G & Demain AL (2013) Recent advances in the biosynthesis of penicillins, cephalosporins and clavams and its regulation. *Biotechnol. Adv.*, 31, 287-311.
[123] Rodriguez-Saiz M, Diez B, & Barredo JL (2005) Why did the Fleming strain fail in penicillin industry? *Fungal Genet. Biol.*, 42, 464-470.
[124] Rodriguez-Saiz M, Barredo JL, Moreno MA, Fernandez-Canon JM, Penalva MA, & Diez B (2001) Reduced function of a phenylacetate-oxidizing cytochrome P450 caused strong genetic improvement in early phylogeny of penicillin-producing strains. *J. Bacteriol.*, 183, 5465-5471.
[125] Rokem JS, Lantz AE, & Nielsen J (2007) Systems biology of antibiotic production by microorganisms. *Nat. Prod. Rep.*, 24, 1262-1287.
[126] Baltz RH (2014) Spontaneous and induced mutations to rifampicin, streptomycin and spectinomycin resistances in actinomycetes: Mutagenic mechanisms and applications for strain improvement. *J. Antibiot.*, 67, 619-624.
[127] Rowlands RT (1984) Industrial strain improvement - Mutagenesis and random screening procedures. *Enzyme Microb. Technol.*, 6, 3-10.
[128] Parekh S (2004) Strain improvement. In *The Desk Encyclopedia of Microbiology* (Schaechter M, ed.) pp. 960–973. Elsevier Academic, San Diego
[129] Pickens LB, Tang Y, & Chooi YH (2011) Metabolic engineering for the production of natural products. *Annu. Rev. Chem. Biomol. Eng.*, 2, 211-236.
[130] Ryu YG, Butler MJ, Chater KF, & Lee KJ (2006) Engineering of primary carbohydrate metabolism for increased production of actinorhodin in *Streptomyces coelicolor*. *Appl. Environ. Microbiol.*, 72, 7132-7139.
[131] Reeves AR, Brikun IA, Cernota WH, Leach BI, Gonzalez MC, & Weber JM (2007) Engineering of the methylmalonyl-CoA metabolite node of *Saccharopolyspora erythraea* for increased erythromycin production. *Metab Eng*, 9, 293-303.
[132] Han SJ, Park SW, Kim BW, & Sim SJ (2008) Selective production of epothilone B by heterologous expression of propionyl-CoA synthetase in *Sorangium cellulosum*. *J. Microbiol. Biotechnol.*, 18, 135-137.
[133] Thykaer J, Nielsen J, Wohlleben W, Weber T, Gutknecht M, Lantz AE, & Stegmann E (2010) Increased glycopeptide production after overexpression of shikimate pathway genes being part of the balhimycin biosynthetic gene cluster. *Metab. Eng.*, 12, 455-461.
[134] Knaggs AR (1999) The biosynthesis of shikimate metabolites. *Nat. Prod. Rep.*, 16, 525-560.
[135] Yanai K, Murakami T, & Bibb M (2006) Amplification of the entire kanamycin biosynthetic gene cluster during empirical strain improvement of *Streptomyces kanamyceticus Proc. Natl. Acad. Sci. USA*, 103, 12951.
[136] Smith DJ, Bull JH, Edwards J, & Turner G (1989) Amplification of the isopenicillin-N synthetase gene in a strain of *Penicillium chrysogenum* producing high-levels of penicillin. *Mol. Gen. Genet.*, 216, 492-497.
[137] Barredo JL, Diez B, Alvarez E, & Martin JF (1989) Large amplification of a 35-Kb DNA fragment carrying 2 penicillin biosynthetic genes in high penicillin producing strains of *Penicillium chrysogenum*. *Curr. Genet.*, 16, 453-459.
[138] Viehrig K, Surup F, Harmrolfs K, Jansen R, Kunze B, & Müller R (2013) Concerted action of P450 plus helper protein to form the amino-hydroxy-piperidone moiety of the potent protease inhibitor crocapeptin. *J. Am. Chem. Soc.*, 135, 16885-16894.
[139] Chen Y, Smanski MJ, & Shen B (2010) Improvement of secondary metabolite production in *Streptomyces* by manipulating pathway regulation. *Appl. Microbiol. Biotechnol.*, 86, 19-25.

[140] Rachid S, Gerth K, Kochems I, & Müller R (2007) Deciphering regulatory mechanisms for secondary metabolite production in the myxobacterium *Sorangium cellulosum* So ce56. *Mol. Microbiol.*, 63, 1783-1796.
[141] Rachid S, Gerth K, & Müller R (2008) NtcA-A negative regulator of secondary metabolite biosynthesis in *Sorangium cellulosum*. *J. Biotechnol.*, 140, 135-142.
[142] Malla S, Niraula NP, Liou K, & Sohng JK (2010) Self-resistance mechanism in *Streptomyces peucetius*: Overexpression of *drrA*, *drrB* and *drrC* for doxorubicin enhancement. *Microbiol. Res.*, 165, 259-267.
[143] Huo L, Rachid S, Stadler M, Wenzel SC, & Müller R (2012) Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. *Chem. Biol.*, 19, 1278-1287.

Chapter I

Biosynthesis of crocacin involves an unusual hydrolytic release domain showing similarity to condensation domains

Stefan Müller, Shwan Rachid, Thomas Hoffmann, Frank Surup, Carsten Volz, Nestor Zaburannyi and Rolf Müller (2014)

Chemistry & Biology, Volume 21, Issue 7, page 855-865

DOI: http://dx.doi.org/10.1016/j.chembiol.2014.05.012

Summary

The crocacins are potent antifungal and cytotoxic natural compounds from myxobacteria of the genus *Chondromyces*. While total synthesis approaches have been reported, the molecular and biochemical basis guiding the formation of the linear crocacin scaffold remained unknown. Alongside with the identification and functional analysis of the crocacin biosynthetic gene cluster from *Chondromyces crocatus* Cm c5, we here present the identification and biochemical characterization of an unusual chain termination domain homologous to condensation domains responsible for hydrolytic release of the product from the assembly line. In particular, gene inactivation studies and *in vitro* experiments using the heterologously produced domain CroK-C2 confirm this surprising role giving rise to the linear carboxylic acid. Additionally, we determined the kinetic parameters of CroK-C2 by monitoring hydrolytic cleavage of the substrate mimic *N*-acetylcysteaminyl-crocacin B using an innovative high-performance liquid chromatography mass spectrometry based assay.

Introduction

Many pharmaceutically relevant natural products are produced by terrestrial microorganisms (Chin et al., 2006) including numerous compounds isolated from actinobacteria (Lucas et al., 2012) and myxobacteria (Wenzel and Müller, 2009) still holding great promise for future drug candidates (Müller and Wink, 2014). One example are the crocacins which were isolated from *Chondromyces crocatus* Cm c5 and found to exhibit strong antifungal and cytotoxic activities (Jansen et al., 1999). Although chemical syntheses have been described, it is still difficult to produce crocacins and their analogues via the synthetic route because of the complexity of the whole molecule. One alternative path forward is biosynthetic engineering of the producer strain which we have been able to set the stage for in recent efforts dealing with other bioactive compounds found in the multiproducer *C. crocatus* Cm c5. During this work genetic tools were developed and the biosynthesis of chondramides (Rachid et al., 2006), ajudazols (Buntin et al., 2008), chondrochlorens (Rachid et al., 2009), thuggacins (Buntin et al., 2010) and crocapeptins (Viehrig et al., 2013) was studied revealing a number of exceptional biosynthetic mechanisms and demonstrating possibilities for biosynthetic engineering in this species.

As basis for biosynthesis, mostly large multifunctional enzyme complexes such as polyketide synthases (PKS) and nonribosomal peptide synthetases (NRPS) account for product formation via successive assembly of these secondary metabolites from small building blocks (Hertweck, 2009; Sieber and Marahiel, 2005). A minimal type I PKS module consists of a β-ketoacyl synthase (KS) an acyl transferase (AT) plus an acyl carrier protein (ACP) collectively

expanding the polyketide chain via Claisen condensation with an extender unit, a process similar to fatty acid biosynthesis (Hopwood, 1997). Similarly, an NRPS module is composed of different domains performing distinct functions. In detail, an adenylation (A) domain activates a specific amino acid in an ATP dependent manner and subsequently transfers it to the 4'-phosphopantetheine (Ppant) arm of a peptidyl carrier protein (PCP). Finally, a condensation (C) domain catalyzes amide bond formation between two adjacent PCP-bound amino acids to yield the dipeptide. Once the growing intermediate reaches its full length it must be released from the last carrier protein. This reaction is commonly catalyzed by a thioesterase (TE) domain. The TE domain mediated product release is the most widespread mechanism found in natural product biosynthesis. Usually, a TE is located at the end of a PKS or NRPS assembly line and contains a conserved catalytic triad (Ser-His-Asp) required for product release (Tsai et al., 2002). Particularly, the assembled compound is transferred from the last carrier protein to the active site Ser of the TE, thus forming the respective ester bound to the TE domain. Depending on the characteristics of the TE, this intermediate either undergoes hydrolysis or macrocyclization. In the first case the ester is hydrolyzed by the nucleophilic attack of water resulting in a linear carboxylate as reported for numerous natural products, e.g. the δ -(L-α-aminoadipyl)-L-cysteinyl-D-valine tripeptide (Byford et al., 1997) and bacillaene (Butcher et al., 2007; Moldenhauer et al., 2007). Contrary to this, macrocyclization occurs when the respective TE-ester undergoes an intramolecular transesterification with a suitable hydroxyl, amino or thiol group as reported for the thuggacins (Buntin et al., 2010), nystatin (Brautaset et al., 2000) and thiocoraline (Robbel et al., 2009) and many others.

However, Nature has evolved several alternative chain termination strategies involving terminal reductase (Red) domains and terminal C domain variants leading to altered products (Du et al., 2010). Chain termination catalyzed by Red domains demonstrates an additional way to release a product from the Ppant moiety of its cognate carrier protein. In general, these domains are capable of catalyzing two and four electron reduction of the respective carrier protein bound thioester, thereby releasing the former ester as aldehyde or alcohol. The first biochemically characterized example thereof was described by the Walsh group in the context of lysine biosynthesis (Ehmann et al., 1999). Primal observation of a likewise mechanism in natural product biosynthesis was reported for myxochelin formation (Gaitatzis et al., 2001). Additional genetic and biochemical investigations proved that the PCP bound intermediate is reduced to the aldehyde state which is either trapped in a reductive aminotransferase reaction giving rise to the terminal amine in myxochelin B or further reduced to the alcohol as found in myxochelin A (Gaitatzis et al., 2001; Li et al., 2008).

As additional alternative, variants of C domains from NRPS systems are as well capable of performing versatile chain termination reactions. The most common route is typically found in fungi, where terminal C domains catalyze macrocyclization of the respective product through nucleophilic attack of the PCP-bound thioester by an internal amine (Keating et al., 2001; Gao et al., 2012). Accordingly, the biosynthesis of the fungal metabolites cyclosporin (Weber et al., 1994), enniatin (Pieper et al., 1995), HC-toxin (Scott-Craig et al., 1992) and PF1022A (Weckwerth et al., 2000) is believed to follow such a mechanism. Noteworthy, this chain termination strategy is not only restricted to the formation of an intramolecular amide, but can also occur between the carrier protein bound carboxyl group of the assembled compound and a primary amine of an acceptor substrate as found in several biosynthetic pathways e.g. rapamycin (Schwecke et al., 1995), FK506 (Motamedi et al., 1998), aeruginosin (Ishida et al., 2007) and streptothricin (Maruyama et al., 2012). Additionally, esterification displays a second and rather uncommon way of a C domain mediated chain release. The first C domain variant that was shown to catalyze esterification in vitro is that of the PCP-C didomain protein Fum14p involved in fumonisin biosynthesis (Zaleta-Rivera et al., 2006). Particularly, this survey suggests that the C domain of Fump14p mediates chain release of a tricarboxylic PCP-bound intermediate while generating its fumonisin esters. Another instance of a C domain catalyzed chain release via transesterification was reported for the free standing condensation enzyme SgcC5 taking part in the biosynthesis of C-1027 (Lin et al., 2009). In this study the authors experimentally demonstrated that SgcC5 transfers the PCP-bound (S)-3-chloro-5-hydroxy-β-tyrosyl intermediate onto the enediyne core mimic (R)-1-phenyl-1,2-ethanediol while forming the respective ester. Furthermore, they stated that this enzyme still retained reduced canonical amide-forming activity of typical NRPS C domains, thereby providing an evolutionary link between amide- and ester-forming condensation enzymes.

On the basis of current knowledge, it becomes increasingly evident that the catalytic activity of C domain variants of NRPS machineries expands behind the classical formation of an amide bond: Alongside with the identification and functional analysis of the crocacin biosynthetic gene cluster from *C. crocatus* Cm c5, we herein demonstrate a C domain variant exhibiting hydrolytic activity, which holds responsible for the formation of the free carboxylic acid moiety of crocacin B. Thus, this study adds a third function to this enzyme class and additionally provides interesting insights for genetic engineering approaches to generate novel natural products.

Results and Discussion

Identification of the crocacin biosynthetic gene cluster

The crocacins are linear natural products most likely derived from a NRPS/PKS hybrid biosynthetic gene cluster. A retrobiosynthetic analysis of the structure suggested two NRPS modules being specific for the incorporation of glycine. Based on this assumption we used the antiSMASH pipeline (Medema et al., 2011) to screen a draft genome of *C. crocatus* Cm c5 which we achieved previously by performing several sequencing attempts for the presence of such a biosynthetic gene cluster. In fact, this search revealed a distinct DNA segment on the chromosome matching the required criteria and a detailed *in silico* analysis strongly suggested this region as the crocacin biosynthetic gene cluster (Figure1). To unambiguously assign this DNA segment and the encoded functions to crocacin biosynthesis the suicide vector pSUP_croI_KO was introduced into the genome of *C. crocatus* Cm c5 (Figure S2). Integration of this plasmid into the genome via single cross over should lead to the inactivation of *croI*, a hybrid PKS-NRPS gene which should be essential for crocacin biosynthesis and furthermore might affect the expression of the subsequent genes due to polar effects. As expected, analysis of the resulting mutants Cm c5*::*pSUP_croI_KO by high-performance liquid chromatography mass spectrometry showed abolishment of crocacin production and thereby confirmed the participation of this genomic region in crocacin biosynthesis (Figure 2). This result, taken together with additional experiments described below clearly confirmed our assumption that this DNA segment indeed bears the encoded functions responsible for crocacin biosynthesis. The respective DNA fragment was therefore referred to as the *cro* locus.

Sequence analysis of the crocacin biosynthetic gene cluster and the contiguous regions

The *cro* locus has a size of about 57 kb with a GC content of 67.3 %. It contains four PKS (*croA-C* and *croJ*), one NRPS (*croK*) and one hybrid PKS-NRPS (*croI*) gene (Figure 1A; Table 1). The β-branching unit consists of five genes (*croD-H*) and is flanked by *croC* and *croI*. Interestingly, the assembly line stops with a C domain (CroK-C2) but no TE is found. Moreover, no TE domain seems to be encoded in the whole gene cluster as could be observed in an in-depth *in silico* analysis of the whole region (see below).

At this point we focused on PKS-NRPS related domains regarding the conservation of characteristic motifs and key amino acid residues. First, the substrate specificity of all acyltransferase (AT) domains was evaluated as described earlier (Yadav et al., 2003). Each AT domain contains the consensus motif GXSXG bearing the catalytic serine residue in its center

and can therefore be considered as an active domain. With respect to their substrate specificity, CroA-AT and CroB-AT are predicted to load methylmalonyl-coenzyme A (CoA), while the remaining ones use malonyl-CoA. This is in agreement with the observed structure of the crocacins. Furthermore, the gene cluster contains three dehydratase (DH) domains, which all should be active as judged by the presence of the common consensus sequence of the respective active site (Keatinge-Clay, 2008). The analysis of all five ketoreduction domains (KRs) confirmed the presence of the exact Rossman fold motif GXGXXG for CroA-KR, CroI-KR and CroJ-KR. The other two KRs deviate from this motif at the second conserved position. In particular, CroB-KR contains a serine and CroC-KR features an aspartate. Since in both cases the deviating amino acid is followed by a glycine this adjustment should still provide enough space for NADH binding and therefore these domains are considered as functional. The substrate specificity of the two adenylation (A) domains CroI-A and CroK-A was predicted by analyzing their primary sequences with the NRPSpredictor2 tool (Röttig et al., 2011). Following this analysis, both A domains should have specificity for the amino acid glycine. Three methyltransferases are encoded in the *cro* locus. CroA-MT and CroB-MT are part of the respective modules 1 and 2 while the third one is encoded by the single standing gene *croL*. All domains can be classified as *O*-MTs when compared to the described conserved motifs (Ansari et al., 2008). A rather unusual feature of the *cro* cluster is the beta branching cassette encoded by *croD-H*. Commonly, such systems consist of a hydroxymethylglutaryl-CoA synthase homologue (HMG), two enoyl-CoA hydratase homologues, and in each case a free standing KS, AT and ACP domain (Calderone, 2008). In the present case the KS and AT are located on the bifunctional protein CroD. Additionally, this KS does not feature the typical Cys to Ser mutation in the active site that hinders it to take part in chain elongation but still allows decarboxylation of the extender unit. These differences might be due to the fact that CroD is a relic of a former PKS complex now taking part in β-branching. Apart from that no further deviations affecting the beta branching cassette seem to be present.

An investigation of the surrounding region of the *cro* cluster revealed two open reading frames *(orfs)* upstream of *croA*. A BLAST analysis suggests that *orf*1 and *orf*2 encode putative proteins comprising an ABC-1 domain. These domains are often found in protein kinases involved in signal transduction and regulation of cellular processes (Leonard et al., 1998), supporting the idea that these proteins might be involved in regulation of crocacin biosynthesis. A similar adjustment of genes was already reported for the rhizopodin gene cluster but no putative function in the biosynthesis could be assigned (Pistorius and Müller, 2012). To assign or to exclude a function of *orf1* and *orf2* in crocacin biosynthesis, we aimed to disrupt both by integrating derivatives of pSUP_Hyg harboring a homology fragment of the respective loci into

the genome of *C. crocatus* Cm c5. Unfortunately, we could only obtain mutants altered in *orf2*. These mutants showed no significant changes in crocacin production when compared to the wild type strain indicating that the gene product is not involved in crocacin production under the applied conditions (Figure 2).

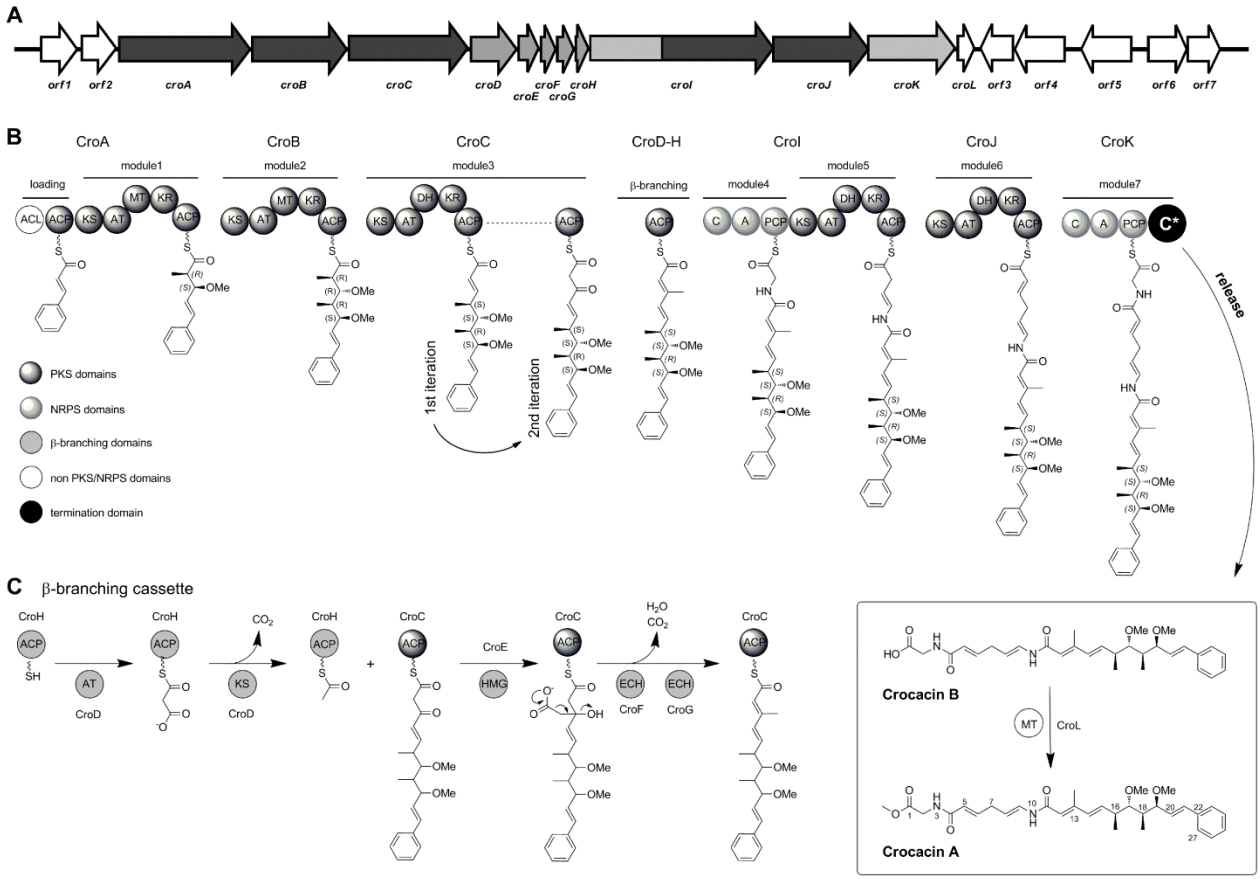


Figure 1: Gene cluster and proposed pathway for biosynthesis of crocacin in *C. crocatus* Cm c5. (A) Genetic organization of the crocacin gene cluster. (B) Proposed biosynthetic pathway for crocacin production in *C. crocatus* Cm c5. (C) Proposed process of β-branching at C13. ECH: enoyl-CoA hydratase homologue.

The downstream region of *croL* comprises five putative genes (*orf3*-*orf7*). Two of them, *orf3* and *orf4* lie on the opposite strand and encode a putative cytochrome P450 (Orf3) and a putative long-chain fatty acid-CoA ligase (Orf4). To clarify their function, we disrupted both genes in separate attempts. Resulting mutants showed no significant changes in crocacin production when compared to the wild type (Figure 2). Therefore, these genes are not considered to be associated with crocacin biosynthesis. Gene *orf5* encodes a putative di-heme-cytochrome-c peroxidase. Targeted inactivation of this gene resulted in a slight decrease of crocacin

production. However, this might be regarded as a secondary effect. The available sequence closes with the overlapping genes *orf6* and *orf7* encoding a putative two-component system consisting of a signal transduction histidine kinase and a LytR family transcriptional regulator. Such systems are usually regulating transcription of target genes upon certain stimuli (West et al., 2001) and may regulate crocacin formation, too. To investigate the role of this system we integrated plasmid pSUP_orf6_KO into the genome of *C. crocatus* Cm c5 thereby impeding expression of the histidine kinase.

Table 1: Putative functions and domain organization within the crocacin gene cluster and the contiguous region

Gene/ protein	Length (bp/aa)	Modules	Domains (coordinates in protein sequence (aa))
croA/CroA	7713/2570	loading	CoA-Lig$_L$ (32-477), ACP$_L$ (606-672)
		module 1	KS (691-1064), AT (1222-1504), O-MT (1620-1876), KR (2160-2354), ACP (2456-2520)
croB/CroB	5700/1899	module 2	KS (36-412), AT (568-885), O-MT (969-1224) KR (1509-1688), ACP (1795-1857)
croC/CroC	5637/1878	module 3	KS (59-401), AT (555-871), DH (925-1096), KR (1251-1358), ACP (1771-1828)
croD/CroD	2793/930	β-branching	KS (39-408), AT (566-886)
croE/CroE	1254/417	β-branching	hydroxymethylglutaryl-coenzyme A synthase homologue
croF/CroF	798/265	β-branching	enoyl-CoA hydratase/isomerase homologue
croG/CroG	789/262	β-branching	enoyl-CoA hydratase/isomerase homologue
croH/CroH	201/66	β-branching	ACP (6-57)
croI/CroI	8892/2963	module 4	C (28-321), A (521-918), PCP (1007-1070)
		module 5	KS (1108-1483), AT (1638-1955), DH (2004-2182), KR (2557-2735), ACP (2984-2913)
croJ/CroJ	5685/1894	module 6	KS (39-414), AT (568-885), DH (935-1102), KR (1475-1664), ACP (1768-1831)
croK/CroK	4611/1536	module 7	C1 (33-328), A (526-923), PCP (1012-1071), C2 (1109-1404)
croL/CroL	723/240	MT	MT (41-144)

Orf	Length (bp/aa)	Accession number of closest homologue	Similarity/identity [%]	Putative function/homologue
1	1827/608	ZP_06890144	54/38	ABC-1 domain protein
2	1698/565	ZP_06890144	51/28	ABC-1 domain protein
3	1248/415	ZP_21238480	69/50	cytochrome P450
4	2037/679	ZP_21238481	66/51	long-chain-fatty-acid-CoA ligase
5	1722/573	ZP_11029676	85/76	di-heme cytochrome-c peroxidase
6	1095/364	ZP_11029675	89/81	two-component histidine kinase
7	783/260	ZP_11029674	93/85	regulatory protein

Again, no significant effect on crocacin production was observed. Nevertheless, these first inactivation experiments allowed us to determine the boundaries of the *cro* cluster. The results of the *in-depth* sequence analysis served as a basis for all further experiments which led to the presentation of a detailed model for crocacin biosynthesis in *C. crocatus* Cm c5 as shown in Figure 1.

Proposed model of crocacin biosynthesis

Based on the structure of crocacin, biosynthesis should most likely start with the activation and tethering of *trans*-cinnamic acid to the ACP-domain of the loading module (Figure 1B). To confirm this, we supplemented growing cultures of *C. crocatus* Cm c5 with isotope-labeled d7-*trans*-cinnamic acid and identified the incorporation of this starter unit by HPLC-MS analysis of the methanolic extract. A mass increase of 7 Da indicates incorporation of the full molecule. (Figure S1). The use of *trans*-cinnamoyl-CoA as starter unit in natural product biosynthesis has already been reported for the stilbenes produced by *Photorhabdus luminescens* TTO1 (Joyce et al., 2008). Detailed studies have shown that in *P. luminescens* TTO1 *trans*-cinnamic acid is derived through non-oxidative deamination of L-phenylalanine by the action of StlA (NP_929491, phenylalanine ammonia lyase [PAL], Enzyme Commission [EC] No. 4.3.1.5.) (Williams et al., 2005). In order to shed light on the origin of the starter molecule in *C. crocatus* Cm c5, we fed ring labeled $^{13}C_6$- L-phenylalanine and in fact observed a mass increase of 6 Da for the crocacins (Figure S1). This suggests a similar origin of *trans*-cinnamic acid in *C. crocatus* Cm c5. To further study this, we analyzed the genome of *C. crocatus* Cm c5 regarding the presence of a PAL encoding gene by employing the amino acid sequence of StlA in a tBLASTn search and found a candidate gene (*pal*^{cmc5}). In this context it has to be mentioned that PALs share very high sequence similarity to histidine ammonia lyases (HAL, EC 4.3.1.3.) but they can be distinguished on the basis of some key amino acid residues (Xiang et al., 2002; Calabrese et al., 2004). Accordingly, Pal^{cmc5} could be classified as HAL and consequently, targeted inactivation of this gene had no effect on crocacin production (Figure 2). Hence, the origin of the *trans*-cinnamic acid starter unit remains elusive, although its incorporation could be verified.

Chain elongation is mediated by PKS modules 1 and 2 using methylmalonyl-CoA as extender unit. Their modular architecture of type KS-AT-MT-KR-ACP perfectly corresponds to the β-methoxy moieties present in crocacin. The origin of the methoxy groups was verified by feeding experiments using isotope-labeled d3-methionine. Mass shifts of +3, +6, and +9 Da are in accordance with the activity of *S*-adenosylmethionine (SAM)-dependent methyl transferases (CroA-MT, CroB-MT and CroL (see below)) responsible for three distinct methylations (Figure S1). Assuming that crocacin biosynthesis follows the “co-linearity rule”, one would expect two

distinct PKS modules after CroB introducing two malonyl-CoA-extender units. However and strikingly, only one PKS module (CroC) with the domain architecture KS-AT-DH-KR-ACP is located behind CroB. To explain this unusual finding we hypothesize that CroC to acts in a programmed iterative manner likewise to PKS modules involved in stigmatellin (Gaitatzis et al., 2002) and aureothin (He et al., 2003) biosynthesis. The way this iteration is likely to occur is extraordinary as the first iteration cycle would use the full set of catalytic domains of module 3 to form an α,β-unsaturated product while the second iteration would be non-reductive. Based on the structure of crocacin it seems likely that the nascent polyketide chain directly undergoes β-branching at C13 after the second iteration step, notably without any β-carbon processing. A plausible explanation for this is that CroE catalyzes the aldol addition of the acetyl group from CroH to the β-keto functionality of C13 much faster than the motif would undergo reduction through CroC-KR and CroC-DH. This hypothesis is supported by the fact that we did not encounter any crocacin derivatives, which would result from a less strict process, e.g. iteration without β-branching or non-iterative use without β-branching. Moreover, the methyl group at C13 is not SAM-derived as reasoned by the above mentioned feeding experiment. In particular, the methyl-branch is likely to be incorporated as follows (Figure 1C); CroD-AT loads malonyl-CoA onto the free standing ACP (CroH). Subsequently, CroD-KS decarboxylates the malonyl-thioester and CroE catalyzes the C-C bond formation between the acetyl moiety and the β-carbon C13 of the enzyme-bound template. Finally, the enoyl-CoA hydratase homologues CroF and CroG process the intermediate by dehydration and decarboxylation to yield the β-methyl moiety. To support this model we disrupted *croE* and *croF*. The respective mutant strains were analyzed by HPLC-MS and showed loss of crocacin production, which suggests the protein's participation in crocacin biosynthesis (Figure 2). Such β-methyl moiety introducing biochemistry on growing PKS templates has been reported for the biosynthesis of other natural products, e.g. mupirocin (El-Sayed et al., 2003), curacin (Chang et al., 2004), myxovirescin (Simunovic et al., 2006), bryostatin (Sudek et al., 2007) and most recently myxopyronin (Sucipto et al., 2013). Besides that, *in vitro* investigations of the β-branching machineries from *pksX* and myxovirescin pathways proved the mechanism proposed earlier (Calderone et al., 2006; Calderone et al., 2007).

After β-branching, the biosynthetic intermediate is passed through module 4 (C-A-PCP) and thereby extended with glycine. Afterwards, two PKS elongation steps with malonyl-CoA are performed by module 5 and module 6, both resulting in an unsaturated polyketide chain. In case of module 5, an unusual position of the double bond is observed, which could be attributed to the conjugational system which is extended for such a constitution. Although the respective DH domain does not obviously differ from common DH domains it may either catalyze an

isomerization (Leesong et al., 1996) or a γ-β elimination as suggested for rhizoxin (Kusebauch et al., 2010).

The assembly of crocacin ends with module 7 comprising the domain order C-A-PCP-C. The first C domain CroK-C1 catalyzes the condensation of a second glycine and the resulting crocacin intermediate remains bound to PCP7. Based on the structure of crocacin B one would expect the presence and subsequent action of a TE domain, leading to the formation of the free carboxylic acid upon hydrolytic release. Surprisingly, the very last domain CroK-C2 of this megasynthetase seems to be another C domain, as could be classified by a homology analysis. *In silico* analysis of the *cro* locus showed no indications of an additional TE encoding gene, pointing towards an unusual release function of CroK-C2. Hereupon, we introduced plasmid pSUP_croK-C2_KO into the genome of *C. crocatus* Cm c5 resulting in an inactivation of *croK-C2* and consequently in mutants which might be able to produce PCP7-bound crocacin. Assuming that a distinct TE domain encoded somewhere else in the genome would be capable of releasing PCP-bound crocacin, this particular gene disruption would result in an unaltered production of crocacins. Interestingly, all clones analyzed by HPLC-MS did not produce crocacins or shunt products thereof (Figure 2). This indicates the involvement of CroK-C2 in the release of the final product although we cannot exclude that the mutagenesis itself results in inactive CroK. To gain further insights into the actual release mechanism we attempted the heterologous production of CroK-C2 alongside with *in vitro* experiments as described below.

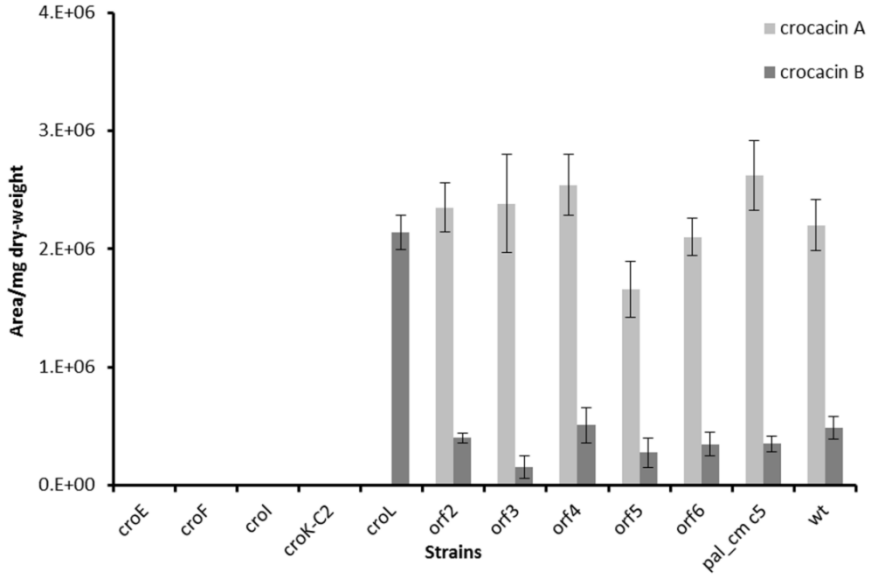


Figure 2: Production of crocacin A and crocacin B of various mutants and the wild type strain. Mutants are altered in the indicated gene. All LC-MS derived area values are normalized to the cell dry weight of each sample. The depicted values are mean values from three independent mutants. Error bars show calculated standard deviations. wt: wild-type.

Crocacin A, the main derivative in *C. crocatus* Cm c5, is most likely derived from crocacin B by a tailoring *O*-methyltransferase acting on the terminal carboxyl group. As described above, the *cro* locus encodes the putative *O*-methyltransferase CroL which could be responsible for the formation of crocacin A. The targeted inactivation of this gene resulted in the abolishment of crocacin A production while elevated levels of crocacin B were detectable in extracts when compared to the wild type (Figure 2). Together with our d3-methionine feeding experiments (Figure S1) these findings confirmed the assumption of CroL being the SAM-dependent tailoring enzyme responsible for *O*-methylation of crocacin B. This is yet another example of a post assembly line modification of a natural compound whereby the product range of the core biosynthetic machinery gets broadened.

Heterologous production and *in vitro* assay of CroK-C2

Gene inactivation experiments revealed that mutants being altered in the last segment of *croK* failed to produce crocacin (see above). This finding, together with the lack of an encoded TE domain in the cluster suggested an extraordinary mode of action of the terminal C domain CroK-C2 in the release mechanism of crocacin B from the assembly line. Because of these interesting facts, we set out to investigate the protein's function in crocacin biosynthesis. First, we applied an *in silico* approach using the NaPDoS tool (Ziemert et al., 2012) to classify this protein (Figure S3 A). According to this evaluation CroK-C2 clusters in a small clade named "modified AA" together with three C domains from the bleomycin and microcystin pathways whereas the remaining domains CroI-C and CroK-C1 group with the classical LCL domains catalyzing condensation of L-amino acids without epimerisation. Even though the group of "modified AA" has not been experimentally characterized, their proposed function is to modify incorporated amino acids, e.g. the dehydration of serine to dehydroalanine (Du et al., 2000; Tillett et al., 2000). A similar function for CroK-C2 is not expected; nevertheless this phylogenetic analysis emphasizes its distinctiveness among common C domains, which suggested a catalytic activity that differs from the standard ones. In order to determine how and to what extent CroK-C2 varies from classical C domains, a primary sequence alignment of all C domains of crocacin biosynthesis was performed (Figure S3 B) and analyzed for the presence of conserved motifs (Marahiel et al., 1997). While CroK-C2 shows 36.8% pairwise similarity to CroI-C and 37.5% to CroK-C1, respectively, the classical C-domains CroI-C and CroK-C1 are 65.8% similar amongst each other. Furthermore, CroK-C2 perfectly features the HHXXXDG consensus motif and only moderately deviates from the consensus motifs C2 and C4. Contrary to this, motifs C1 and C5 seem to be not present at all in CroK-C2.

To investigate the actual function of CroK-C2 in crocacin biosynthesis, we heterologously produced the protein in *E. coli* BL21 (DE3) pLysS and purified it to homogeneity as N-terminal His_6-tagged form (Figure 3A). To test if CroK-C2 is able to release PCP7-bound crocacin, an *N*-acetylcysteaminoyl-crocacin B (crocacin-SNAC) mimicking PCP7-bound crocacin was synthesized and used as a substrate-mimic to set up an enzymatic *in vitro* assay using native and denatured CroK-C2. Indeed, hydrolysis of the crocacin-SNAC ester was only detectable upon incubation with native CroK-C2 (Figure S4 A). This initial test confirmed the proposed role of this terminal unusual C-domain in the release of crocacin B. As a next step we aimed to evaluate the kinetic parameters of the CroK-C2-mediated hydrolysis. As there is no cofactor consumption which could be photometrically monitored, we set up a MS-based method to record the reaction. The high salt concentration of the assay obviously hampered a mass spectrometric approach based on direct infusion. Hence, a fast HPLC-MS method was established, which is capable of separating crocacin-SNAC and its hydrolysis product in less than 1 minute. In particular, consecutive injections of the assay mixture enabled us to track the time course in steps of 0.9 minutes (Figure 3 C). We were not able to reach a plateau phase in the Michaelis-Menten graph, owing to the poor water solubility of the substrate upon which we could not use substrate concentrations higher than 200 µM. We found that on the one hand DMSO has a strong influence on the enzyme activity (Figure S4 C) and on the other hand, a certain concentration of DMSO is mandatory to dissolve the substrate. For these two reasons, we maintained DMSO concentration at 5 % for all assays as the most reasonable assay condition. After all, the formation of crocacin B with variable concentrations of crocacin-SNAC exhibited Michaelis-Menten kinetics with the parameters as follows: K_m= 140.3 ± 26.1 µM, k_{cat}= 749 s^{-1}, and k_{cat}/K_m= 5.34 s^{-1} μM^{-1} (Figure 3 D). On the basis of these parameters we conclude that CroK-C2 efficiently releases crocacin B from the assembly line.

Having confirmed the exceptional activity of CroK-C2, we intended to assess whether the hydrolytic cleavage of crocacin-SNAC is unique to this particular enzyme, or if any C-domain could catalyze this reaction. To evaluate this question we choose the myxochelin megasynthetase MxcG (C-A-PCP-Red) to be assayed with crocacin-SNAC because of two reasons. First, its functionality has already been shown *in vitro* (Gaitatzis et al., 2001; Li et al., 2008) and second, this NRPS ends with a Red domain instead of a TE, thus an undesired hydrolysis of the substrate can most likely be neglected, especially as even NADPH can be omitted in the assay. Following this idea, MxcG was heterologously produced and purified to homogeneity (Figure 3 A). MxcG was then used in an *in vitro* assay under the same conditions as applied in previous experiments. However, we did not monitor the reaction online but rather stopped the reaction at different time points. We observed neither an increase of the product

(crocacin B), nor a significant decrease of the substrate (crocacin-SNAC) and thereby concluded that MxcG, contrary to CroK-C2, is not able to act hydrolytically on crocacin-SNAC (Figure 3 B).

Taken together, these results depict the specialty of CroK-C2 with respect to other C domains and prove its remarkable role in the release of a PCP-bound intermediate by hydrolysis. Obviously, it cannot be ruled out that other hydrolytic C-domains present in currently uncharacterized assembly lines exist. For that reason, we aimed to screen the GenBank nucleotide database for the presence of NRPS and NRPS/PKS hybrid clusters of bacterial origin which do not encode an apparent release function like a TE or terminal reductase domain (TD) (Figure S5).

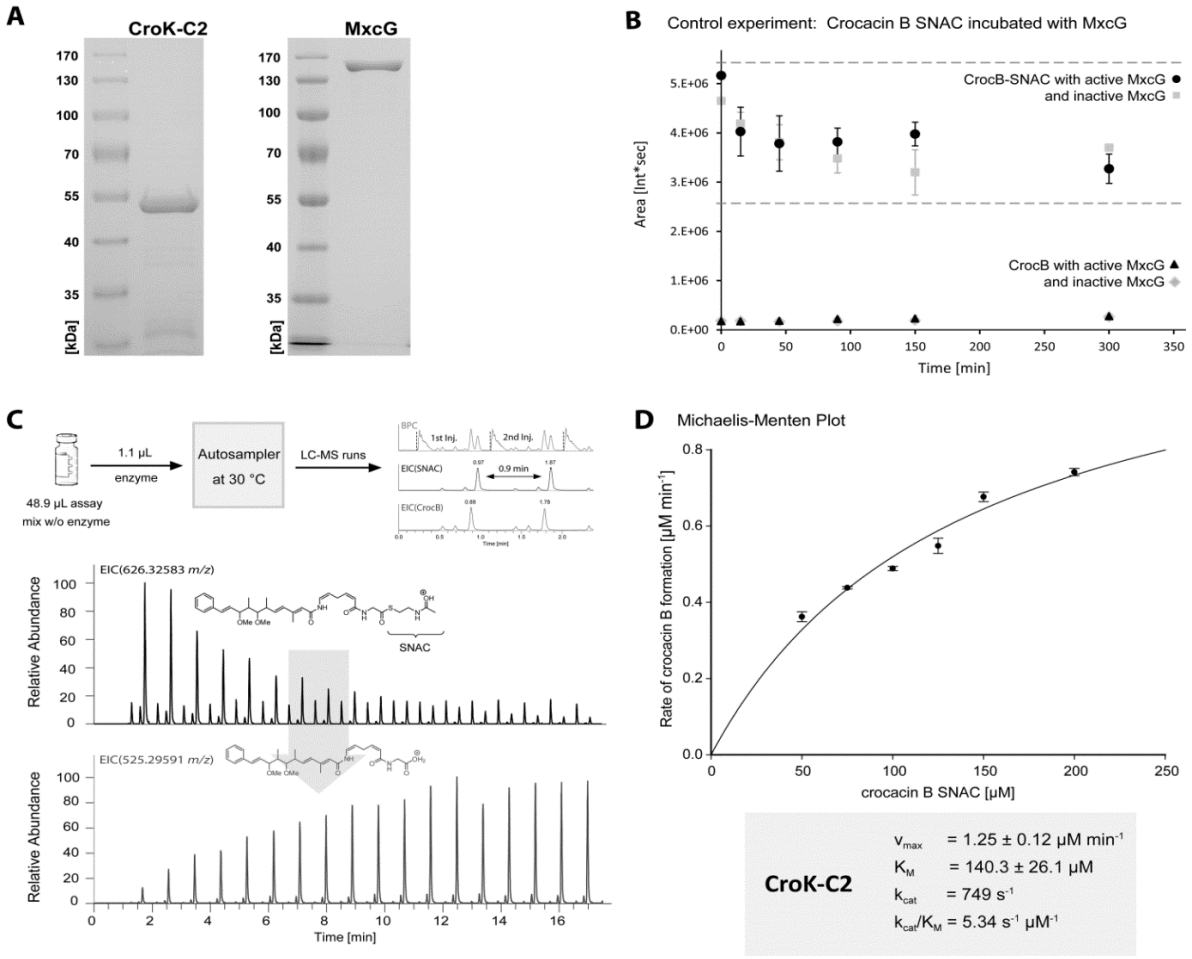


Figure 3: In Vitro Activity Assays of CroK-C2 and MxcG with Respect to Crocacin B SNAC Hydrolysis. (A) SDS-PAGE analysis of heterologously produced and enriched CroK-C2 (52.4 kDa) and MxcG (156.8 kDa). (B) Crocacin-SNAC hydrolysis experiment with active and inactive MxcG shows no effect of MxcG on this substrate. Error bars show calculated standard deviations. (C) Fast HPLC-MS assay used to monitor the formation of crocacin B as a measure for CroK-C2 activity. The reaction mixture was consecutively probed every 0.9 minutes. Different substrate concentrations were measured with this technique to determine initial reaction rates (D) Michaelis-Menten Plot and respective parameters of CroK-C2 as calculated on the basis of crocacin B formation in the assay. Error bars show calculated standard deviations.

Initially, starting from 2390 sequences, 1632 were sorted out because no biosynthetic function has been assigned for them. Secondly, the remaining 758 sequences were analyzed for NRPS signatures with 241 sequences matching these criteria. Out of these sequences 175 contained a TE and 16 a TD. Accordingly, the remaining 50 sequences featured NRPS-related genes while no TE or TD domain was detectable. A detailed *in silico* analysis (Table S1) of these sequences using Pfam (Punta et al., 2012) revealed that 15 additional sequences contained a TE domain that was either not found by the antiSMASH software (due to their poor homology to other TEs) or which was part of a split cluster where the other part contained the TE. Hence, 35 sequences were not harboring any classical termination domains. A number of 6 sequences thereof have been defined as partial sequence in the database which is why they have been excluded from the analysis. From the remaining 29 sequences a number of 16 encode at least partially characterized secondary metabolite pathways in which chain termination should involve a C domain catalyzing amide or ester bond formation. Finally, the 13 residual ones have either not been linked to secondary metabolites, or they do not appear to encode terminal C domains except one polyketide-2 cluster (AB070942) from *Streptomyces avermitilis* (Omura et al., 2001) and the crocacin gene cluster (FN547928) from *C. crocatus* Cm c5 of which an early version has been previously submitted to GenBank. The core cluster from *S. avermitilis* most likely consists of two distinct loci BAB69218.1 and BAB69217.1. The first encodes the loading module while the second bears the information of the loading modules ACP followed by two PKS modules (KS-AT-DH-KR-ACP) and a terminal C domain as the sole NRPS feature. Noteworthy, a Pfam search revealed the presence of a putative haloacid dehalogenase-like hydrolase (E-value 0.32) after the C domain on BAB69217.1 and a putative abhydrolase (E-value 9.7e-13) encoded by the following locus BAB69216.1. This suggests that compound release might rather be performed by one of the putative hydrolases than by the C domain but the actual scenario remains elusive.

According to this *in silico* evaluation and to the best of our knowledge CroK-C2 and its hydrolytic chain terminating activity is currently unique to bacterial NRPS machineries.

Significance

Myxobacteria are well known producers of structurally diverse bioactive natural products. Their assembly is performed by large multienzyme PKS and NRPS complexes and often comprises unprecedented biosynthetic features. Therefore, the analysis and characterization of yet undescribed secondary metabolite gene clusters hold great biotechnological potential, especially when new catalytic functions within these machineries are detected. Knowledge gained could help to engineer related biosynthetic routes towards the formation of altered

products. Our analysis of the crocacin biosynthetic gene cluster in *C. crocatus* Cm c5 indeed revealed such a catalytic activity of a condensation-type domain being part of a modular NRPS. Alongside with a programmed iteration module and an uncommon β-branching cassette, the cluster features a highly remarkable domain order in the last module. With a total absence of typical release domains, the observed hydrolytic release of crocacins was initially not understood. This question was addressed by characterizing the terminal C domain CroK-C2 by reconstitution of its activity *in vitro*, which proved that CroK-C2 is responsible for release of the PCP-bound crocacin. This particular finding can now help to engineer NRPS natural product pathways toward a programmed template release, thus opening the possibility to create new derivatives with terminal carboxylic acid moieties but further studies are required to elucidate the actual molecular basis that triggers a former C domain to catalyse hydrolysis.

Experimental Procedures

Strains and culture conditions

See Supplemental Experimental Procedures

Gene disruption experiments in *C. crocatus* Cm c5

See Supplemental Experimental Procedures

Construction of pSUP_Hyg derivatives used in this study

See Supplemental Experimental Procedures

Isolation and Purification of Crocacin B

Utilized instruments, the strain maintenance and 70 l scale cultivation of *C. crocatus* Cm c5 as well as the obtainment of the crude extract were described previously (Viehrig et al., 2013). As described for the isolation of crocapeptins A_1-A_3, the crude extract was subjected to solvent partition using 85% aqueous methanol, which was extracted twice with heptane. Subsequently, the aqueous methanol was adjusted to 70% methanol and extracted twice with dichloromethane. The dichloromethane fraction contained the main amount of crocacin B, which was fractionated by RP medium-pressure liquid chromatography [column 480 × 30 mm, ODS/AQ C18 (Kronlab), gradient 30−100% methanol in 60 min, flow 30 ml/min, UV peak detection at 210 nm]. Fractions containing crocacin B were combined and further purified by preparative RP HPLC (column 250 x 21 mm, Nucleodur C18 5 µm (Macherey-Nagel), gradient 45-100% methanol with 0.2% acetic acid in 25 min) provided 10.7 mg of crocacin B.

Synthesis of *N*-acetylcysteaminyl-crocacin B

The synthetic strategy used in the generation *N*-acetylcysteaminyl-crocacin B (crocacin-SNAC) is based on the method described by (Carroll et al., 2002). A solution of crocacin B (6.3 mg, 12 µmol) in 1 ml dichloromethane was cooled down to 0 °C. 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide (EDC, 4.6 mg, 24 µmol) and 4-dimethylaminopyridine (DMAP, ~50 µg, cat.) were added. After 2 h stirring at 25 °C, the solution was cooled down to 0 °C and *N*-acetylcysteamine (7.2 mg, 60 µmol, 5 eq.) was added. Subsequently, the mixture was stirred for 16 h at 25 °C and quenched with 2 ml water. After separation of the layers, dichloromethane was removed *in vacuo* and the resulting material was analyzed by HPLC-UV-MS. Subsequently, crocacin-SNAC (1.8 mg, 2.8 µmol, 24 %) was isolated from the crude product by preparative HPLC (column 250 x 21 mm, Nucleodur C18 5 µm (Macherey-Nagel), gradient 65-100% methanol in 25 min).

Heterologous Production and Purification of CroK-C2 and MxcG

CroK-C2 was produced as N-terminal His-tagged fusion protein using the pET28b(+) plasmid (Novagen). For construction of the expression plasmid pET28b_croK-C2 the respective part of *croK* was amplified by PCR using the oligonucleotides CroK-C2-NdeI (CCG CGC CCA TAT GCC TCT CGA CAA GGT GGC CCT G) and CroK-C2-XhoI (ACG CGC TCG AGT CAG GCT GCT GGA GAG GTC G) and integrated into the prepared vector. Production of the His_6 fusion protein was achieved by using *E. coli* BL21 DE3 carrying the pLysS plasmid. Initial cultivation was carried out in a 2 L shaking flask containing 300 mL Luria broth medium supplemented with 25 µg/mL chloramphenicol and 50 µg/mL kanamycin sulfate at 37°C and 200 rpm. When the culture reached an optical density at 600 nm of 0.5-0.7 gene expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.25 mM and the temperature was shifted to 16°C. After further cultivation for approximately 20 h the cells were harvested by centrifugation washed twice with PBS-buffer and were resuspended in ice cold 25 mL HisTrap A buffer (25 mM HEPES, 500 mM NaCl, 25 mM imidazole and 10% v/v glycerol at pH:8.0) containing Roche complete protease inhibitors. Cells were disrupted by using a French press three times at 900 PSI and the debris was removed by centrifugation at 20000xg for 30 minutes at 4°C. Afterwards the soluble fraction was filtered through a 0.45 µm syringe filter before incubation with Ni^{2+}-NTA (Qiagen) for 1 h at 4°C. The protein was eluted with a stepwise gradient of imidazole using different mixtures of HisTrap A and HisTrap B (25 mM HEPES, 500 mM NaCl, 500 mM imidazole and 10% v/v glycerol at pH:8.0). All fractions were subjected to SDS-PAGE and the ones which contained the target

protein (52.4 kDa) were pooled and concentrated. As a second purification step size exclusion chromatography was applied. Therefore an Äkta pure M2 system equipped with a Superdex 200 increase 10/300 GL column was used in HisTrap buffer without imidazole at a flow rate of 0.75 ml/min. Fractions of interest were pooled and analyzed by SDS-PAGE and finally subjected to MADLI-TOF analysis to prove their identity.

MxcG was produced and purified according to the previous described protocol (Gaitatzis et al., 2001; Li et al., 2008). The final fraction was concentrated and subjected to size exclusion chromatography to remove residual dithiothreitol present in the sample. Therefore an Äkta pure M2 system equipped with a Superdex 200 increase 10/300 GL column at a flow rate of 0.75 ml/min in elution buffer (25 mM Tris HCl (pH: 8.0), 50 mM NaCl, 0.1 mM EDTA) was used. The amount of protein in the respective samples was determined using Bio Rad Protein Assay dye reagent and obtained intensities were referred to a calibration line.

In vitro assay for monitoring the hydrolysis of *N*-acetylcysteamine ester of crocacin B by CroK-C2 and MxcG

Test samples were premixed without protein and preheated to 30 °C. The enzyme reaction was started by the addition of active protein or in case of the negative control experiments by addition of heat inactivated protein (5 min at 98 °C). Activity testing of CroK-C2 was carried out by a fast LC-MS method in order to measure the time course of the reaction. Reaction mixtures of 50 µL were prepared in conical HPLC vials containing 100 mM Tris HCl (pH 7.6), 50 mM NaCl, 0.1 µM protein, and *N*-acetylcysteaminyl-crocacin B (crocacin-SNAC) in concentrations ranging from 50 - 200 µM. The DMSO content was fixed to 5 % (v/v) for all assay mixtures. For measuring a time course, the autosampler temperature was set to 30 °C to allow reasonable enzyme activity. Upon starting the reaction by adding enzyme, the same sample was probed consecutively by drawing 1 µL of the mixture every 0.9 min. This sample was injected to the HPLC by an overlapping injection protocol. Separation is achieved by an isocratic elution at 58 % B on a Waters BEH C18, 50 x 2.1 mm, 1.7 µm dp column using (A) H2O + 0.1 % formic acid to (B) acetonitrile + 0.1 % FA at a flow rate of 800 µL/min and 45 °C. Peaks for crocacin B ($[M+H]^+$ = 525.2959 *m/z*) and crocacin SNAC ($[M+H]^+$ = 626.3258 *m/z*) are monitored in SIM mode (450-700 m/z) at a resolution of 7500 and 200 ms fill time using a Thermo Scientific Orbitrap. The concentration of crocacin B was determined by an area-based calibration curve which was measured using the same method. Rates of product formation were calculated from the initial linear portion of the obtained time course and fitted to the Michaelis-Menten equation by using the enzyme kinetics tool of Sigma Plot 11.0 (Systat Software, Inc., San Jose, USA).

Activity of MxcG was monitored by quenching enzyme assays at different time points. Reaction mixtures of 70 µL were incubated in HPLC vials at 30 °C. Each reaction mixture contained 25 mM Tris HCl (pH 8.0), 50 mM NaCl, 0.1 mM EDTA, 5 µM of protein, 20 µM crocacin-SNAC and 2% DMSO. All reactions were started simultaneously by the addition of enzyme and quenched at different time points by the addition of 70 µL methanol. After centrifugation, 2 µL of the reaction mixture was used for LC-MS analysis using the standard method.

N-acetylcysteaminyl-crocacin B

See Supplemental Experimental Procedures

LC-MS standard method

See Supplemental Experimental Procedures

In silico analysis of NRPS-encoding gene clusters with no apparent termination domain

See Supplemental Experimental Procedures

Accession Number

The sequence of crocacin biosynthetic gene cluster from *C. crocatus* Cm c5 has been deposited in GenBank under accession number: KJ868728.

Supplemental Information

Supplemental information includes Supplemental Experimental Procedures, five figures, and two tables.

Acknowledgement

Research in R.M.'s laboratory was funded by the Bundesministerium für Bildung and Forschung and the Deutsche Forschungsgemeinschaft.

Reference List

Ansari, M.Z., Sharma, J., Gokhale, R.S., and Mohanty, D. (2008). *In silico* analysis of methyltransferase domains involved in biosynthesis of secondary metabolites. BMC Bioinformatics. 9, 454

Brautaset, T., Sekurova, O.N., Sletta, H., Ellingsen, T.E., Strm, A.R., Valla, S., and Zotchev, S.B. (2000). Biosynthesis of the polyene antifungal antibiotic nystatin in *Streptomyces noursei* ATCC 11455: Analysis of the gene cluster and deduction of the biosynthetic pathway. Chemistry & Biology. 6, 395-403.

Buntin, K., Irschik, H., Weissman, K.J., Luxenburger, E., Blöcker, H., and Müller, R. (2010). Biosynthesis of thuggacins in myxobacteria: comparative cluster analysis reveals basis for natural product structural diversity. Chemistry & Biology. 4, 342-356.

Buntin, K., Rachid, S., Scharfe, M., Blöcker, H., Weissman, K.J., and Müller, R. (2008). Production of the antifungal isochromanone ajudazols A and B in *Chondromyces crocatus* Cm c5: Biosynthetic machinery and cytochrome P450 modifications. Angewandte Chemie International Edition. 24, 4595-4599.

Butcher, R.A., Schroeder, F.C., Fischbach, M.A., Straight, P.D., Kolter, R., Walsh, C.T., and Clardy, J. (2007). The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis*. Proceedings of the National Academy of Sciences of the United States of America. 5, 1506-1509.

Byford, M.F., Baldwin, J.E., Shiau, C.Y., and Schofield, C.J. (1997). The mechanism of ACV synthetase. Chemical Reviews. 7, 2631-2650.

Calabrese, J.C., Jordan, D.B., Boodhoo, A., Sariaslani, S., and Vannelli, T. (2004). Crystal structure of phenylalanine ammonia lyase: Multiple helix dipoles implicated in catalysis. Biochemistry. 36, 11403-11416.

Calderone, C.T. (2008). Isoprenoid-like alkylations in polyketide biosynthesis. Natural Product Reports. 5, 845-853.

Calderone, C.T., Iwig, D.F., Dorrestein, P.C., Kelleher, N.L., and Walsh, C.T. (2007). Incorporation of non-methyl branches by isoprenoid-like logic: Multiple beta-alkylation events in the biosynthesis of myxovirescin A1. Chemistry & Biology. 7, 835-846.

Calderone, C.T., Kowtoniuk, W.E., Kelleher, N.L., Walsh, C.T., and Dorrestein, P.C. (2006). Convergence of isoprene and polyketide biosynthetic machinery: Isoprenyl-S-carrier proteins in the *pksX* pathway of *Bacillus subtilis*. Proceedings of the National Academy of Sciences of the United States of America. 24, 8977-8982.

Carroll, B.J., Moss, S.J., Bai, L.Q., Kato, Y., Toelzer, S., Yu, T.W., and Floss, H.G. (2002). Identification of a set of genes involved in the formation of the substrate for the incorporation of the unusual "glycolate" chain extension unit in ansamitocin biosynthesis. Journal of the American Chemical Society. 16, 4176-4177.

Chang, Z., Sitachitta, N., Rossi, J.V., Roberts, M.A., Flatt, P., Jia, J., Sherman, D.H., and Gerwick, W.H. (2004). Biosynthetic pathway and gene cluster analysis of Curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. Journal of Natural Products. 8, 1356-1367.

Chin, Y.W., Balunas, M.J., Chai, H.B., and Kinghorn, A.D. (2006). Drug discovery from natural sources. Aaps Journal. 2, E239-E253.

Du, L., and Lou, L. (2010). PKS and NRPS release mechanisms. Natural Product Reports. 2, 255-278.

Du, L., Sanchez, C., Chen, M.T., Edwards, D.J., and Shen, B. (2000). The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. Chemistry & Biology. 8, 623-642.

Ehmann, D.E., Gehring, A.M., and Walsh, C.T. (1999). Lysine biosynthesis in *Saccharomyces cerevisiae*: Mechanism of (alpha)-aminoadipate reductase (Lys2) involves posttranslational phosphopantetheinylation by Lys5. Biochemistry. 19, 6171-6177.

El-Sayed, A.K., Hothersall, J., Cooper, S.M., Stephens, E., Simpson, T.J., and Thomas, C.M. (2003). Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586. Chemistry & Biology. 5, 419-430.

Gaitatzis, N., Kunze, B., and Müller, R. (2001). In vitro reconstitution of the myxochelin biosynthetic machinery of *Stigmatella aurantiaca* Sg a15: Biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases. Proceedings of the National Academy of Sciences of the United States of America. 20, 11136-11141.

Gaitatzis, N., Silakowski, B., Kunze, B., Nordsiek, G., Blöcker, H., Höfle, G., and Müller, R. (2002). The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. Journal of Biological Chemistry. 15, 13082-13090.

Gao, X., Haynes, S.W., Ames, B.D., Wang, P., Vien, L.P., Walsh, C.T., and Tang, Y. (2012). Cyclization of fungal nonribosomal peptides by a terminal condensation-like domain. Nature Chemical Biology. 10, 823-830.

He, J., and Hertweck, C. (2003). Iteration as programmed event during polyketide assembly; Molecular analysis of the aureothin biosynthesis gene cluster. Chemistry & Biology. 12, 1225-1232.

Hertweck, C. (2009). The biosynthetic logic of polyketide diversity. Angewandte Chemie International Edition. 26, 4688-4716.

Hopwood, D.A. (1997). Genetic contributions to understanding polyketide synthases. Chemical Reviews. 7, 2465-2497.

Ishida, K., Christiansen, G., Yoshida, W.Y., Kurmayer, R., Welker, M., Valls, N., Bonjoch, J., Hertweck, C., Borner, T., Hemscheidt, T., and Dittmann, E. (2007). Biosynthesis and structure of aeruginoside 126A and 126B, cyanobacterial peptide glycosides bearing a 2-carboxy-6-hydroxyoctahydroindole moiety. Chemistry & Biology. 5, 565-576.

Jansen, R., Washausen, P., Kunze, B., Reichenbach, H., and Höfle, G. (1999). Antibiotics from gliding bacteria, LXXXIII - The crocacins, novel antifungal and cytotoxic antibiotics from *Chondromyces crocatus* and *Chondromyces pediculatus* (Myxobacteria): Isolation and structure elucidation. European Journal of Organic Chemistry. 5, 1085-1089.

Joyce, S.A., Brachmann, A.O., Glazer, I., Lango, L., Schwär, G., Clarke, D.J., and Bode, H.B. (2008). Bacterial biosynthesis of a multipotent stilbene. Angewandte Chemie International Edition. 10, 1942-1945.

Keating, T.A., Ehmann, D.E., Kohli, R.M., Marshall, C.G., Trauger, J.W., and Walsh, C.T. (2001). Chain termination steps in nonribosomal peptide synthetase assembly lines: Directed acyl-S-enzyme breakdown in antibiotic and siderophore biosynthesis. ChemBioChem. 2, 99-107.

Keatinge-Clay, A. (2008). Crystal structure of the erythromycin polyketide synthase dehydratase. Journal of Molecular Biology. 4, 941-953.

Kusebauch, B., Busch, B., Scherlach, K., Roth, M., and Hertweck, C. (2010). Functionally distinct modules operate two consecutive alpha,beta--beta,gamma double-bond shifts in the rhizoxin polyketide assembly line. Angewandte Chemie International Edition. 49, 1460–1464.

Leesong, M., Henderson, B.S., Gillig, J.R., Schwab, J.M., and Smith, J.L. (1996). Structure of a dehydratase–isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: Two catalytic activities in one active site. Structure 4, 253–264.

Leonard, C.J., Aravind, L., and Koonin, E.V. (1998). Novel families of putative protein kinases in bacteria and archaea: Evolution of the "eukaryotic" protein kinase superfamily. Genome Research. 10, 1038-1047.

Li, Y., Weissman, K.J., and Müller, R. (2008). Myxochelin biosynthesis: Direct evidence for two- and four-electron reduction of a carrier protein-bound thioester. Journal of the American Chemical Society. 24, 7554-7555.

Lin, S., Van Lanen, S.G., and Shen, B. (2009). A free-standing condensation enzyme catalyzing ester bond formation in C-1027 biosynthesis. Proceedings of the National Academy of Sciences of the United States of America. 11, 4183-4188

Lucas, X., Senger, C., Erxleben, A., Gruning, B.A., Doring, K., Mosch, J., Flemming, S., and Gunther, S. (2013). StreptomeDB: A resource for natural compounds isolated from *Streptomyces* species. Nucleic Acids Research. D1, D1130-D1136.

Marahiel, M.A., Stachelhaus, T., and Mootz, H.D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis. Chemical Reviews. 7, 2651-2674.

Maruyama, C., Toyoda, J., Kato, Y., Izumikawa, M., Takagi, M., Shin-ya, K., Katano, H., Utagawa, T., and Hamano, Y. (2012). A stand-alone adenylation domain forms amide bonds in streptothricin biosynthesis. Nature Chemical Biology. 9, 791-797.

Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M.A., Weber, T., Takano, E., and Breitling, R. (2011). antiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Research. 39, W339-W346.

Moldenhauer, J., Chen, X.H., Borriss, R., and Piel, J. (2007). Biosynthesis of the antibiotic bacillaene, the product of a giant polyketide synthase complex of the trans-AT family. Angewandte Chemie International Edition. 43, 8195-8197.

Motamedi, H., and Shafiee, A. (1998). The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506. European Journal of Biochemistry. 3, 528-534.

Müller, R., and Wink, J. (2014). Future potential for anti-infectives from bacteria: How to exploit biodiversity and genomic potential. International Journal of Medical Microbiology. 1, 3-13.

Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y., and Hattori, M. (2001). Genome sequence of an industrial microorganism *Streptomyces avermitilis*: Deducing the ability of producing secondary metabolites. Proceedings of the National Academy of Sciences of the United States of America. 21, 12215-12220.

Pieper, R., Haese, A., Schroder, W., and Zocher, R. (1995). Arrangement of catalytic sites in the multifunctional enzyme enniatin synthetase. European Journal of Biochemistry. 1, 119-126.

Pistorius, D., and Müller, R. (2012). Discovery of the rhizopodin biosynthetic gene cluster in *Stigmatella aurantiaca* Sg a15 by genome mining. ChemBioChem. 3, 416-426.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L.L., Eddy, S.R., Bateman, A., and Finn, R.D. (2012). The Pfam protein families database. Nucleic Acids Research. D1, D290-D301.

Rachid, S., Krug, D., Kochems, I., Kunze, B., Scharfe, M., Blöcker, H., Zabriski, M., and Müller, R. (2006). Molecular and biochemical studies of chondramide formation - highly cytotoxic natural products from *Chondromyces crocatus* Cm c5. Chemistry & Biology. 6, 667-681.

Rachid, S., Scharfe, M., Blöcker, H., Weissman, K.J., and Müller, R. (2009). Unusual chemistry in the biosynthesis of the antibiotic chondrochlorens. Chemistry & Biology. 1, 70-81.

Robbel, L., Hoyer, K.M., and Marahiel, M.A. (2009). TioS T-TE - a prototypical thioesterase responsible for cyclodimerization of the quinoline- and quinoxaline-type class of chromodepsipeptides. FEBS Journal. 6, 1641-1653.

Röttig, M., Medema, M.H., Blin, K., Weber, T., Rausch, C., and Kohlbacher, O. (2011). NRPSpredictor2-a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Research. 39, W362-W367.

Schwecke, T., Aparicio, J.F., Molnar, I., König, A., Khaw, L.E., Haydock, S.F., Oliynyke, M., Caffrey, P., Cortes, J., Lester, J.B., Bohm, G.A., Staunton, J., and Leadley, P.F. (1995). The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. Proceedings of the National Academy of Sciences of the United States of America. 17, 7839-7843.

Scott-Craig, J.S., Panaccione, D.G., Pocard, J.A., and Walton, J.D. (1992). The cyclic peptide synthetase catalyzing HC-toxin production in the filamentous fungus *Cochliobolus carbonum* is encoded by a 15.7-kilobase open reading frame. Journal of Biological Chemistry. *36*, 26044-26049.

Sieber, S.A., and Marahiel, M.A. (2005). Molecular mechanisms underlying nonribosomal peptide synthesis: Approaches to new antibiotics. Chemical Reviews. 2, 715-738.

Simunovic, V., Zapp, J., Rachid, S., Krug, D., Meiser, P., and Müller, R. (2006). Myxovirescin biosynthesis is directed by hybrid polyketide synthases/nonribosomal peptide synthetase, 3-hydroxy-3-methylglutaryl CoA synthases and trans-acting acyltransferases. ChemBioChem. 8, 1206-1220.

Sucipto, H., Wenzel, S.C., and Müller, R. (2013). Exploring chemical diversity of α-pyrone antibiotics: Molecular basis of myxopyronin biosynthesis. ChemBioChem. 13, 1581-1589.

Sudek, S., Lopanik, N.B., Waggoner, L.E., Hildebrand, M., Anderson, C., Liu, H.B., Patel, A., Sherman, D.H., and Haygood, M.G. (2007). Identification of the putative bryostatin polyketide synthase gene cluster from "*Candidatus endobugula sertula*", the uncultivated microbial symbiont of the marine bryozoan *Bugula neritina*. Journal of Natural Products. 1, 67-74.

Tillett, D., Dittmann, E., Erhard, M., von Döhren, H., Börner, T., and Neilan, B.A. (2000). Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: An integrated peptide-polyketide synthetase system. Chemistry & Biology. 10, 753-764.

Tsai, S.C., Lu, H.X., Cane, D.E., Khosla, C., and Stroud, R.M. (2002). Insights into channel architecture and substrate specificity from crystal structures of two macrocycle-forming thioesterases of modular polyketide synthases. Biochemistry. 42, 12598-12606.

Viehrig, K., Surup, F., Harmrolfs, K., Jansen, R., Kunze, B., and Müller, R. (2013). Concerted action of P450 plus helper protein to form the amino-hydroxy-piperidone moiety of the potent protease inhibitor crocapeptin. Journal of the American Chemical Society. 45, 16885-16894.

Weber, G., Schorgendorfer, K., Schneiderscherzer, E., and Leitner, E. (1994). The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame. Current Genetics. 2, 120-125.

Weckwerth, W., Miyamoto, K., Iinuma, K., Krause, M., Glinski, M., Storm, T., Bonse, G., Kleinkauf, H., and Zocher, R. (2000). Biosynthesis of PF1022A and related cyclooctadepsipeptides. Journal of Biological Chemistry. 23, 17909-17915.

Wenzel, S.C., and Müller, R. (2009). The biosynthetic potential of myxobacteria and their impact on drug discovery. Curr. Opin. Drug Discov. Devel. 2, 220-230.

West, A.H., and Stock, A.M. (2001). Histidine kinases and response regulator proteins in two-component signaling systems. Trends in Biochemical Sciences. 6, 369-376.

Williams, J.S., Thomas, M., and Clarke, D.J. (2005). The gene *stlA* encodes a phenylalanine ammonia-lyase that is involved in the production of a stilbene antibiotic in *Photorhabdus luminescens* TT01. Microbiology. 8, 2543-2550.

Xiang, L., and Moore, B.S. (2002). Inactivation, complementation, and heterologous expression of *encP*, a novel bacterial phenylalanine ammonia-lyase gene. Journal of Biological Chemistry. 36, 32505-32509.

Yadav, G., Gokhale, R.S., and Mohanty, D. (2003). Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. Journal of Molecular Biology. 2, 335-363.

Zaleta-Rivera, K., Xu, C., Yu, F., Butchko, R.A., Proctor, R.H., Hidalgo-Lara, M.E., Raza, A., Dussault, P.H., and Du, L. (2006). A bidomain nonribosomal peptide synthetase encoded by FUM14 catalyzes the formation of tricarballylic esters in the biosynthesis of fumonisins. Biochemistry. 8, 2561-2569.

Ziemert, N., Podell, S., Penn, K., Badger, J.H., Allen, E., and Jensen, P.R. (2012). The natural product domain seeker NaPDoS: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. PLoS ONE. 3, e34064.

Supplemental Information

Inventory of Supplemental Information

Supplemental Data

- **Figure S1**, related to Figure 1: Isotopic peak pattern of crocacin A ([M+H] +calc.= 539.31155 *m/z*) reveals incorporation of labeled precursors.
- **Figure S2**, related to Figure 2: Genetic verification of mutants generated in this study.
- **Figure S3**, related to Figure 3: *In silico* analysis of CroK-C2.
- **Figure S4**, related to Figure 3: Results from LC-MS assay for monitoring crocacin SNAC hydrolysis catalyzed by CroK-C2.
- **Figure S5**, related to Figure 3: Workflow for the identification of NRPS-related gene clusters not encoding apparent release domains.
- **DNA-sequence of *pal*$^{cm\ c5}$**
- **Table S1**, related to Figure 3: *In silico* analysis of NRPS gene clusters with no obvious release function.
- **Table S2**, related to experimental procedures: Oligonucleotides used for mutagenesis studies.

Supplemental Experimental Procedures

Strains and culture conditions

Gene disruption experiments in *C. crocatus* Cm c5

Construction of pSUP_Hyg derivatives used in this study

N-acetylcysteaminyl-crocacin B

LC-MS standard method

In silico analysis of NRPS-encoding gene clusters with no apparent termination domain

Supplemental Reference List

Supplemental Data

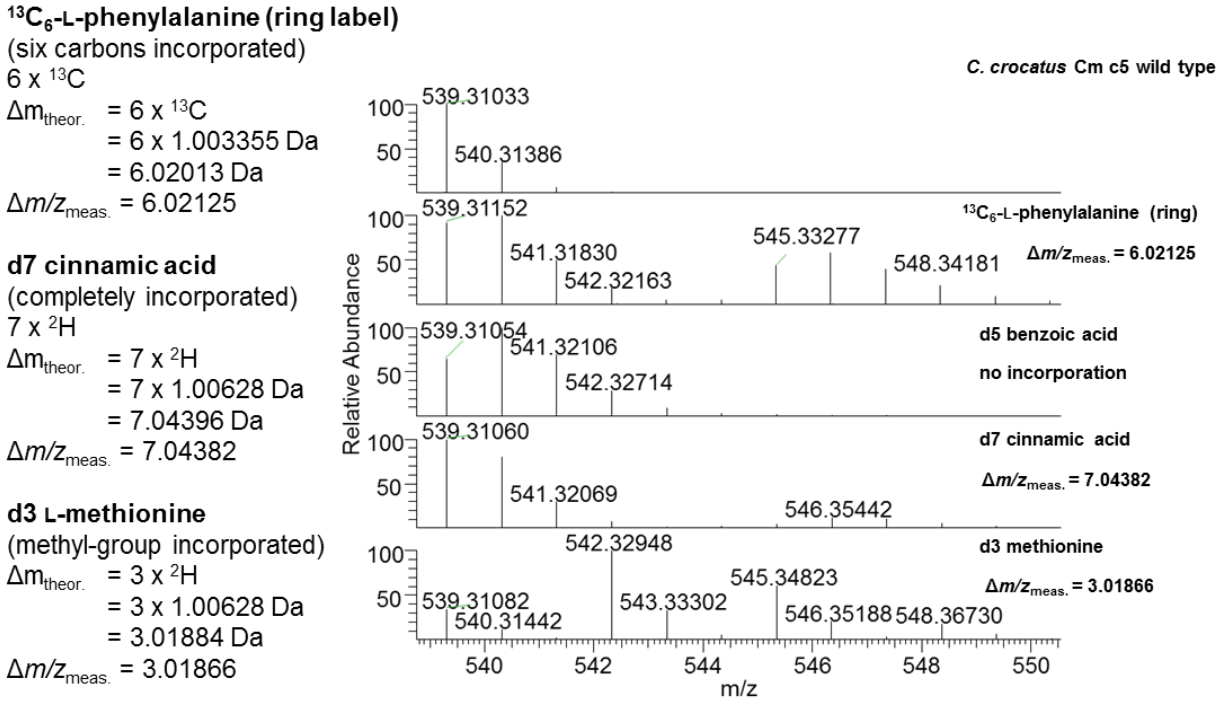


Figure S1, related to Figure 1: Isotopic peak pattern of crocacin A ($[M+H]^+_{calc.}$= 539.31155 *m/z*) reveals incorporation of labeled precursors. The observed mass shifts fit to the heavy isotopes that were incorporated. Ring labelled L-phenylalanine is incorporated into crocacin A which can be concluded due to a mass shift of 6.02125 *m/z*. At the same time, labelled benzoic acid is not incorporated. Feeding experiments with cinnamic acid resulted in a mass shift of 7.04382 *m/z* which proves the complete incorporation of this molecule. Noteworthy, the isotope intensity distribution is sometimes shifted to the higher isotopes. This is likely owing to a partial scrambling of the precursors by primary metabolism and/or exchange of acidic deuterium atoms. The labeled amino acid L-methionine was fed in order to determine the origin of the three methyl groups present in crocacin A. With detected mass shifts of 3.01866 *m/z*, 6.03741 *m/z* and 9.05648 *m/z* this experiment clearly shows that all methyl groups are derived from S-adenosyl-methionine.

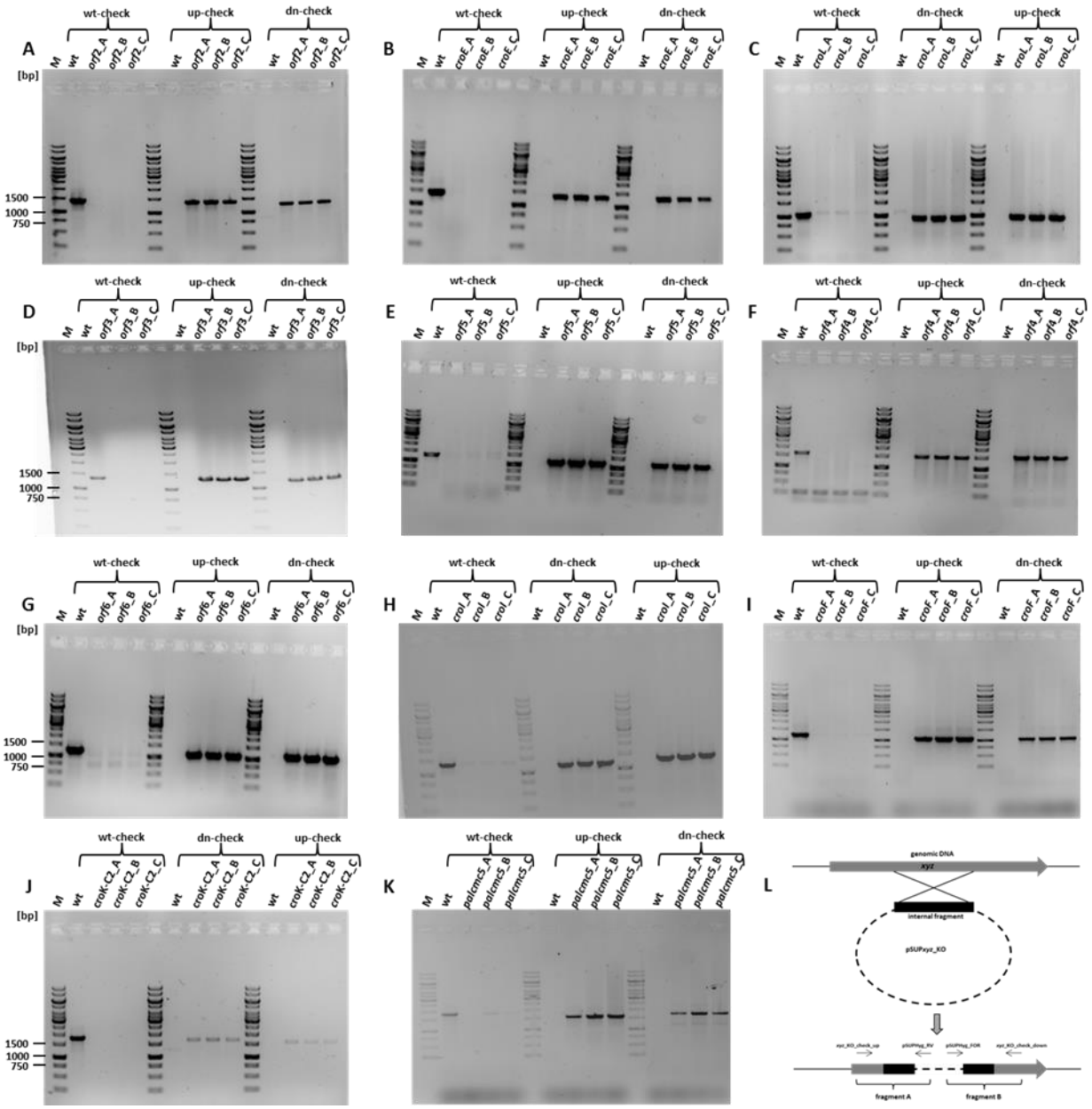
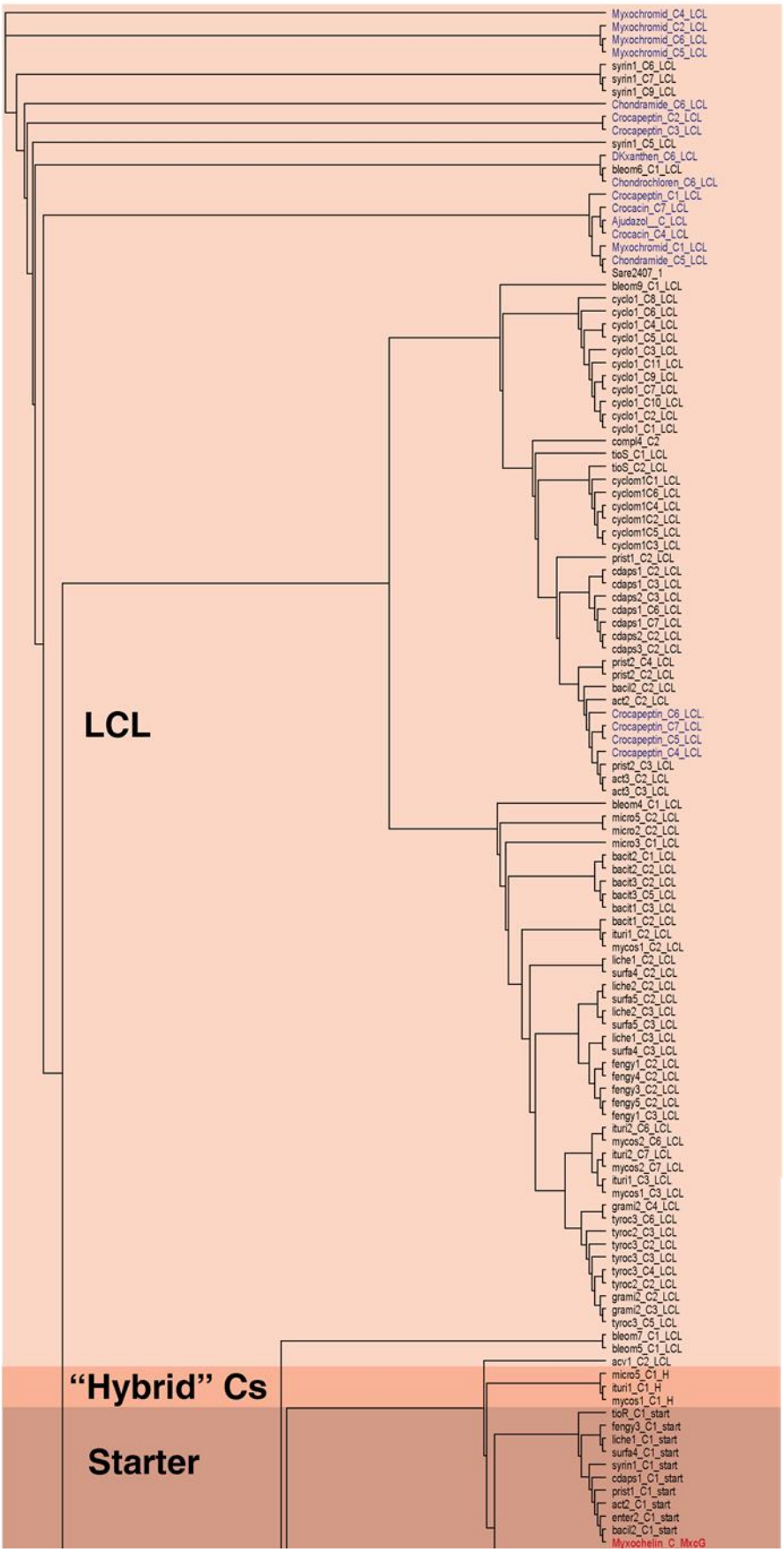
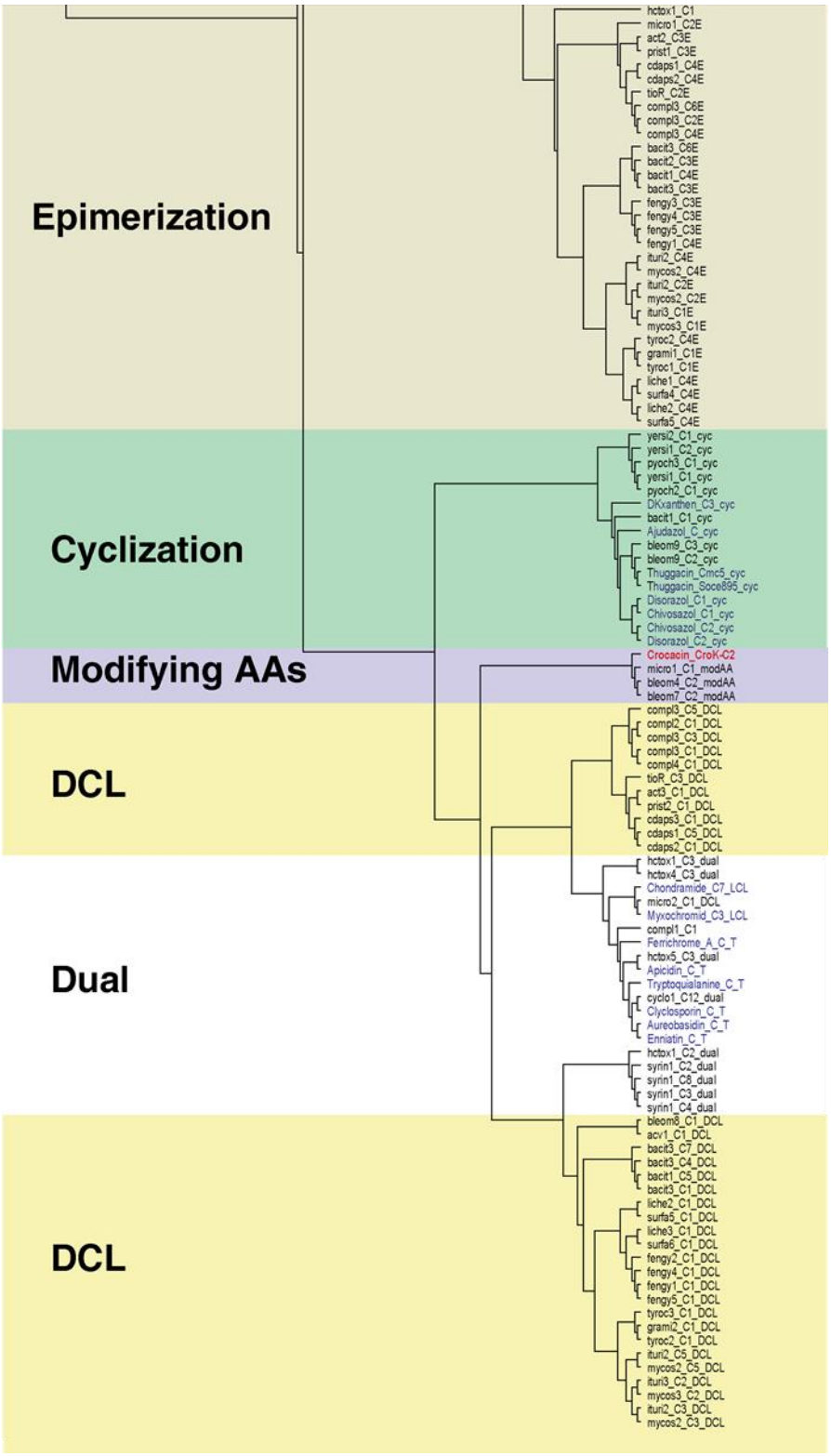


Figure S2, related to Figure 2: Genetic verification of mutants generated in this study. (A): verification of Cm c5::pSUP_orf2_KO strains wt-check (orf2_KO_check_up + orf2_KO_check_dn): 1400 bp; up-check (orf2_KO_check_up + pSUPHyg_FOR): 1259 bp; dn-check: (orf2_KO_check_dn + pSUPHyg_RV): 1239 bp. (B): verification of Cm c5::pSUP_croE_KO strains wt-check (croE_KO_check_up + croE_KO_check_dn): 1426 bp; up-check (croE_KO_check_up + pSUPHyg_FOR): 1268 bp; dn-check (croE_KO_check_dn + pSUPHyg_RV): 1287 bp. (C): verification of Cm c5::pSUP_croL_KO strains wt-check (croL_KO_check_up + croL_KO_check_dn): 1024 bp; dn-check (croL_KO_check_dn + pSUPHyg_RV): 878 bp; up-check (croL_KO_check_up + pSUPHyg_FOR): 851 bp. (D): verification of Cm c5::pSUP_orf3_KO strains wt-check (orf3_KO_check_up + orf3_KO_check_dn): 1363 bp; up-check (orf3_KO_check_up + pSUPHyg_FOR): 1219 bp; dn-check (orf3_KO_check_dn + pSUPHyg_RV): 1189 bp. (E): verification of Cm c5::pSUP_orf5_KO strains wt-check (orf5_KO_check_up + orf5_KO_check_dn): 1276 bp; up-check (orf5_KO_check_up + pSUPHyg_RV): 1122 bp; dn-check (orf5_KO_check_dn + pSUPHyg_FOR): 1128 bp. (F): verification of Cm c5::pSUP_orf4_KO strains wt-check (orf4_KO_check_up + orf4_KO_check_dn): 1613 bp; up-check (orf4_KO_check_up + pSUPHyg_RV): 1457 bp; dn-check (orf4_KO_check_dn + pSUPHyg_FOR): 1468 bp. (G): verification of Cm c5::pSUP_orf6_KO strains wt-check (orf6_KO_check_up + orf6_KO_check_dn): 1194 bp; up-check

(orf6_KO_check_up + pSUPHyg_RV): 1024 bp; dn-check (orf6_KO_check_dn + pSUPHyg_FOR): 1041 bp. (H): verification of Cm c5::pSUP_croI_KO strains wt-check (croI_KO_check_up + croI_KO_check_dn): 1323 bp; dn-check (croI_KO_check_dn + pSUPHyg_RV): 1151 bp; up-check (croI_KO_check_up + pSUPHyg_FOR): 1181 bp. (I): verification of Cm c5::pSUP_croF_KO strains wt-check (croF_KO_check_up + croF_KO_check_dn): 1198 bp; up-check (croF_KO_check_up + pSUPHyg_FOR): 1000 bp; dn-check (croF_KO_check_dn + pSUPHyg_RV): 1014 bp. (J): verification of Cm c5::pSUP_croK-C2_KO strains wt-check (croK-C2_KO_check_up + croK-C2_KO_check_dn): 1771 bp; dn-check (croK-C2_KO_check_dn + pSUPHyg_FOR): 1589 bp; dn-check (croK-C2_KO_check_up + pSUPHyg_RV): 1552 bp. (K): verification of Cm c5::pSUP_ palcmc5_KO strains wt-check (palcmc5_KO_check_up + palcmc5_KO_check_dn): 1794 bp; up-check (palcmc5_KO_check_up + pSUPHyg_FOR): 1794 bp; dn-check (palcmc5_KO_check_dn + pSUPHyg_RV): 1679 bp. (L): Integration of pSUPHyg derivatives into the chromosome of *C. crocatus Cm c5* by single cross-over. Obtained mutants were analyzed using two sets of oligonucleotides binding at the plasmid backbone as well as on the genomic DNA.

A


Myxochromid_C4_LCL
Myxochromid_C2_LCL
Myxochromid_C6_LCL
Myxochromid_C5_LCL
syrin1_C6_LCL
syrin1_C7_LCL
syrin1_C9_LCL
Chondramide_C6_LCL
Crocapeptin_C2_LCL
Crocapeptin_C3_LCL
syrin1_C5_LCL
DKxanthen_C6_LCL
bleom6_C1_LCL
Chondrochloren_C6_LCL
Crocapeptin_C1_LCL
Crocacin_C7_LCL
Ajudazol__C_LCL
Crocacin_C4_LCL
Myxochromid_C1_LCL
Chondramide_C5_LCL
Sare2407_1
bleom9_C1_LCL
cyclo1_C8_LCL
cyclo1_C6_LCL
cyclo1_C4_LCL
cyclo1_C5_LCL
cyclo1_C3_LCL
cyclo1_C11_LCL
cyclo1_C9_LCL
cyclo1_C7_LCL
cyclo1_C10_LCL
cyclo1_C2_LCL
cyclo1_C1_LCL
compl4_C2
tioS_C1_LCL
tioS_C2_LCL
cyclom1C1_LCL
cyclom1C6_LCL
cyclom1C4_LCL
cyclom1C2_LCL
cyclom1C5_LCL
cyclom1C3_LCL
prist1_C2_LCL
cdaps1_C2_LCL
cdaps1_C3_LCL
cdaps2_C3_LCL
cdaps1_C6_LCL
cdaps1_C7_LCL
cdaps2_C2_LCL
cdaps3_C2_LCL
prist2_C4_LCL
prist2_C2_LCL
bacil2_C2_LCL
act2_C2_LCL
Crocapeptin_C6_LCL
Crocapeptin_C7_LCL
Crocapeptin_C5_LCL
Crocapeptin_C4_LCL
prist2_C3_LCL
act3_C2_LCL
act3_C3_LCL
bleom4_C1_LCL
micro5_C2_LCL
micro2_C2_LCL
micro3_C1_LCL
bacit2_C1_LCL
bacit2_C2_LCL
bacit3_C2_LCL
bacit3_C5_LCL
bacit1_C3_LCL
bacit1_C2_LCL
ituri1_C2_LCL
mycos1_C2_LCL
liche1_C2_LCL
surfa4_C2_LCL
liche2_C2_LCL
surfa5_C2_LCL
liche2_C3_LCL
surfa5_C3_LCL
liche1_C3_LCL
surfa4_C3_LCL
fengy1_C2_LCL
fengy4_C2_LCL
fengy3_C2_LCL
fengy5_C2_LCL
fengy1_C3_LCL
ituri2_C6_LCL
mycos2_C6_LCL
ituri2_C7_LCL
mycos2_C7_LCL
ituri1_C3_LCL
mycos1_C3_LCL
grami2_C4_LCL
tyroc3_C6_LCL
tyroc2_C3_LCL
tyroc3_C2_LCL
tyroc3_C3_LCL
tyroc3_C4_LCL
tyroc2_C2_LCL
grami2_C2_LCL
grami2_C3_LCL
tyroc3_C5_LCL
bleom7_C1_LCL
bleom5_C1_LCL
acv1_C2_LCL
micro5_C1_H
ituri1_C1_H
mycos1_C1_H
tioR_C1_start
fengy3_C1_start
liche1_C1_start
surfa4_C1_start
syrin1_C1_start
cdaps1_C1_start
prist1_C1_start
act2_C1_start
enter2_C1_start
bacil2_C1_start
Myxochelin_C_MxcG
LCL
"Hybrid" Cs
Starter


Epimerization
Cyclization
Modifying AAs
DCL
Dual
DCL
hctox1_C1
micro1_C2E
act2_C3E
prist1_C3E
cdaps1_C4E
cdaps2_C4E
tioR_C2E
compl3_C6E
compl3_C2E
compl3_C4E
bacit3_C6E
bacit2_C3E
bacit1_C4E
bacit3_C3E
fengy3_C3E
fengy4_C3E
fengy5_C3E
fengy1_C4E
ituri2_C4E
mycos2_C4E
ituri2_C2E
mycos2_C2E
ituri3_C1E
mycos3_C1E
tyroc2_C4E
grami1_C1E
tyroc1_C1E
liche1_C4E
surfa4_C4E
liche2_C4E
surfa5_C4E
yersi2_C1_cyc
yersi1_C2_cyc
pyoch3_C1_cyc
yersi1_C1_cyc
pyoch2_C1_cyc
DKxanthen_C3_cyc
bacit1_C1_cyc
Ajudazol_C_cyc
bleom9_C3_cyc
bleom9_C2_cyc
Thuggacin_Cmc5_cyc
Thuggacin_Soce895_cyc
Disorazol_C1_cyc
Chivosazol_C1_cyc
Chivosazol_C2_cyc
Disorazol_C2_cyc
Crocacin_CroK-C2
micro1_C1_modAA
bleom4_C2_modAA
bleom7_C2_modAA
compl3_C5_DCL
compl2_C1_DCL
compl3_C3_DCL
compl3_C1_DCL
compl4_C1_DCL
tioR_C3_DCL
act3_C1_DCL
prist2_C1_DCL
cdaps3_C1_DCL
cdaps1_C5_DCL
cdaps2_C1_DCL
hctox1_C3_dual
hctox4_C3_dual
Chondramide_C7_LCL
micro2_C1_DCL
Myxochromid_C3_LCL
compl1_C1
Ferrichrome_A_C_T
hctox5_C3_dual
Apicidin_C_T
Tryptoquialanine_C_T
cyclo1_C12_dual
Cyclosporin_C_T
Aureobasidin_C_T
Enniatin_C_T
hctox1_C2_dual
syrin1_C2_dual
syrin1_C8_dual
syrin1_C3_dual
syrin1_C4_dual
bleom8_C1_DCL
acv1_C1_DCL
bacit3_C7_DCL
bacit3_C4_DCL
bacit1_C5_DCL
bacit3_C1_DCL
liche2_C1_DCL
surfa5_C1_DCL
liche3_C1_DCL
surfa6_C1_DCL
fengy2_C1_DCL
fengy4_C1_DCL
fengy1_C1_DCL
fengy5_C1_DCL
tyroc3_C1_DCL
grami2_C1_DCL
tyroc2_C1_DCL
ituri2_C5_DCL
mycos2_C5_DCL
ituri3_C2_DCL
mycos3_C2_DCL
ituri2_C3_DCL
mycos2_C3_DCL

B

```
   motif                        C1                                             C2
CroK-C2_hydrolytic      AYPLTAMQEVMLDEHAKDAGRRGIYHLQYALHLRDEGFSLPALEAALATLVRRNPALRTV 60
CroK-C1_LCL             PQPTSHGQQALWYLYRLDP-ESPAYNVAFAFRVRSP-IDAPAMRGALLALTERHPVLRTT 58
CroI_LCL                -HPLSYGQRALWFLHQLSP-QSHAYNTLFAVRIRST-LDLAKLAEAFRLMVQRHPTLRST 57
                          * :  *..:   :  .. .   *:  :*.::*.  :. . :  *:  :..*:*.**:.

   motif
CroK-C2_hydrolytic      ILEIPGAPRLQGVLRTAPPSLEIEDLRALTQEERQRVSEARLQADRGRPFETGPSRLLSR 120
CroK-C1_LCL             YLQRDGEA-LRTIHARLDPGFDELDASTWAEAE----LAERVQAARCRPFDLERG-PILR 112
CroI_LCL                YCLEDGVP-RQIIHPSLELPLVEEDAASLDEKQ----LAPRLLERAHRPFDLVRG-PVAR 111
                             * .  : :       :   *  :  : :       *:     ***:   .  : *

   motif                                      C3                                 C4
CroK-C2_hydrolytic      FRILLHTDHTFEIHLSCHHAILDGWSSVELFHELLALYGAHKRGEEIVETPEPHGYRDFV 180
CroK-C1_LCL             GVLFHRAPDDHVLLLVAHHIAVDGWSLWILIDELKALYPAVRAGHPPELPLLSVQYGDFV 172
CroI_LCL                AHVFTRGPEHHVLLVAAHHITCDFFSGASLLEELFAAYSALRCGELPTLPTPALQYTDFV 171
                          :: :  . . : : .**   * :*   *:.** * * * : *.    .  .  * ***

CroK-C2_hydrolytic      ALER---EARDAQPSRAFWLDHLRGSRR----AEPAPATSVESWDRRVRTVHVEAPLLAR 233
CroK-C1_LCL             RWQRDLLGGPEGERLWAYWQQRLSGKIPNLDLPLDLPRPPVQTFRGASHPFTIRREPTQR 232
CroI_LCL                ARQEELLAS-DGEELFAYWQTQLATAPVSLDLPLDRPRPPIPTGRGASVSFHIGSELTRR 230
                          :.    . :.:   *:*  :*         .   * ..: :      .. :      *

   motif                                                                         C5
CroK-C2_hydrolytic      LHELRRRWSVSLRAILLGTYLHALQKIGG-EAVPVGVVSSGRSARLVDPLRTRGLLWNLT 292
CroK-C1_LCL             IVQLSKAHGATLFAVLLAAYQTLLSRHSGERDILVGTPMAGRSRQEFSPVVGDFINMVVL 292
CroI_LCL                ISTLTRSQGATLNMALVAMFQVLLHRYTRESDIVVGSPMAGWSRSEFKGVVGDFINMVVL 290
                        :  * :  ..:*   *:. :   * :      : **   :* *   .. :    :   :

   motif
CroK-C2_hydrolytic      PFHC 296
CroK-C1_LCL             RADL 296
CroI_LCL                RSDL 294
                          .
```

Figure S3, related to Figure 3: *In silico* analysis of CroK-C2. (A): Phylogenetic analysis of CroK-C2 and MxcG (both highlighted in red) using NaPDoS (Ziemert et al., 2012). Domains highlighted in blue are not included in the NaPDoS database but added by the authors. LCL domains catalyze a peptide bond between two L-amino acids; DCL domains catalyze a peptide bond between a D- and a L-amino acid; Starter Cs acylate the first amino acid with a β-hydroxy-carboxylic acid, Cyclization domains catalyze both peptide bond formation and the subsequent cyclization of cysteine, serine, or threonine; Epimerization domains switch the chirality of the last amino acid in the growing peptide; Dual domains catalyze both epimerization and condensation. (B): Primary sequence alignment of all C domains taking part in crocacin biosynthesis using ClustalW2 (Larkin et al., 2007). The presence of conserved motifs C1-C5 (C1: SXAQXR(LM)(WY)XL; C2: RHEXLRTXF; C3: MHHXISDG(WV)S; C4: YXD(FY)AVW; C5: IVGXFVNT(QL)(CA)XR) which are highlighted in grey has been analyzed according to Marahiel et al., 1997.

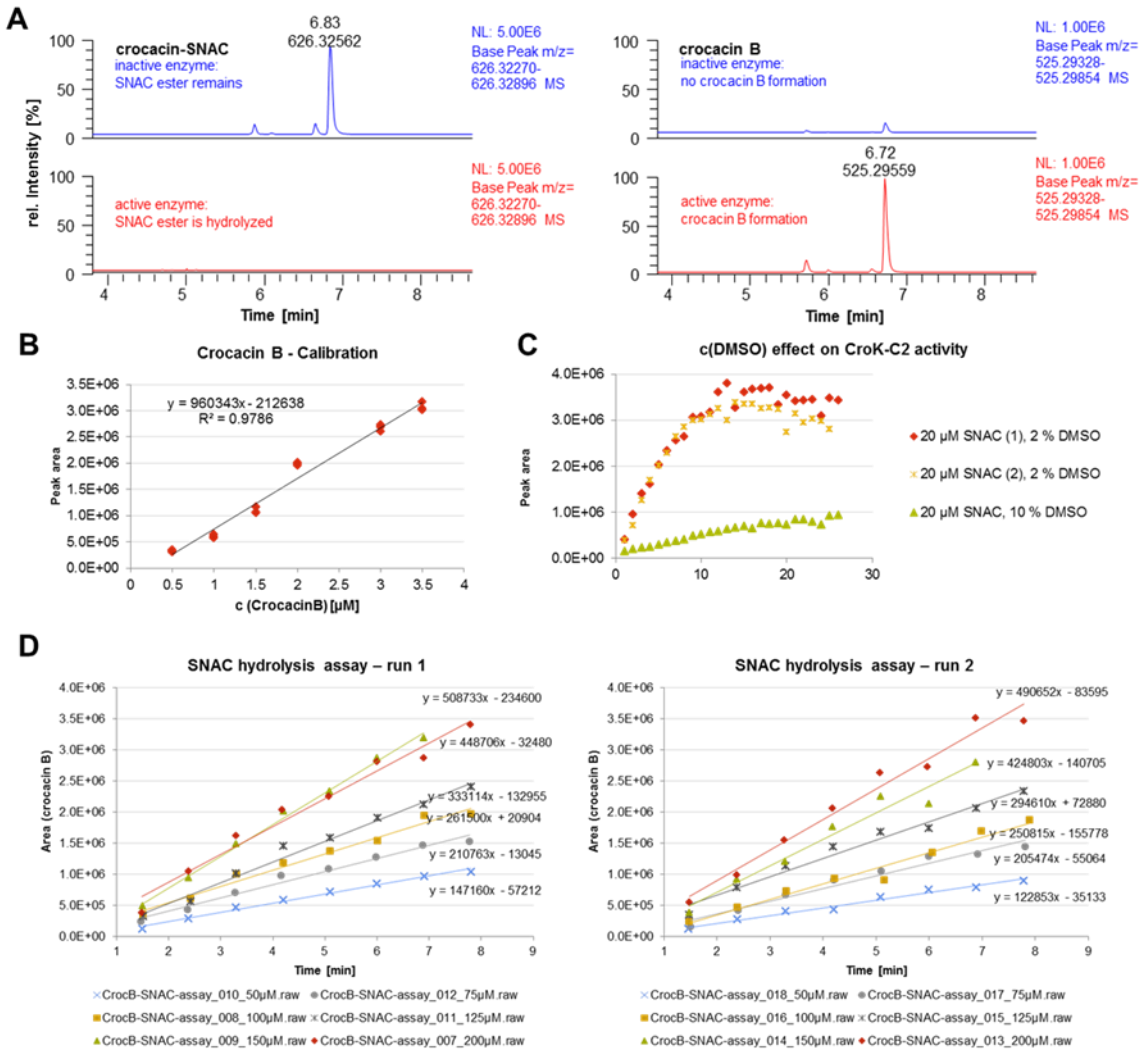


Figure S4, related to Figure 3: Results from LC-MS assay for monitoring crocacin SNAC hydrolysis catalyzed by CroK-C2. (A): HPLC-MS analysis of crocacin-SNAC hydrolysis assays with active (red trace) and inactive (blue trace) CroK-C2. Both assays were prepared from the same master mix containing 100 mM Tris HCl pH: 7.6, 50 mM NaCl, 2% DMSO and 10 µM crocacin-SNAC. The reactions were initialized by addition of active or heat inactivated CroK-C2. After 30 minutes of incubation at 30°C the assays were quenched with methanol and subjected to HPLC-MS analysis. The extracted ion chromatogram (EIC) for crocacin-SNAC ($[M+H]^+_{calc.}$= 626.32583 *m/z*) showed substrate consumption only when active CroK-C2 was added to the reaction mixture. In line with these results the EIC for crocacin B ($[M+H]^+_{calc.}$= 525.29495 *m/z*), showed only significant amounts of crocacin B in the corresponding assay. (B): Crocacin B calibration curve measured with identical conditions as used for the assay. (C): Effect of high DMSO concentration on enzyme activity is clearly shown by comparing the reaction rates for reactions with 2 % and 10 % DMSO. The reproducibility of the method is proven by the replicate measurement. (D): Results of the assays with different substrate concentrations. Only the linear part of the reaction rate curve is shown. Values were linearly fitted and corresponding slopes used to calculate the Michaelis-Menten parameters.

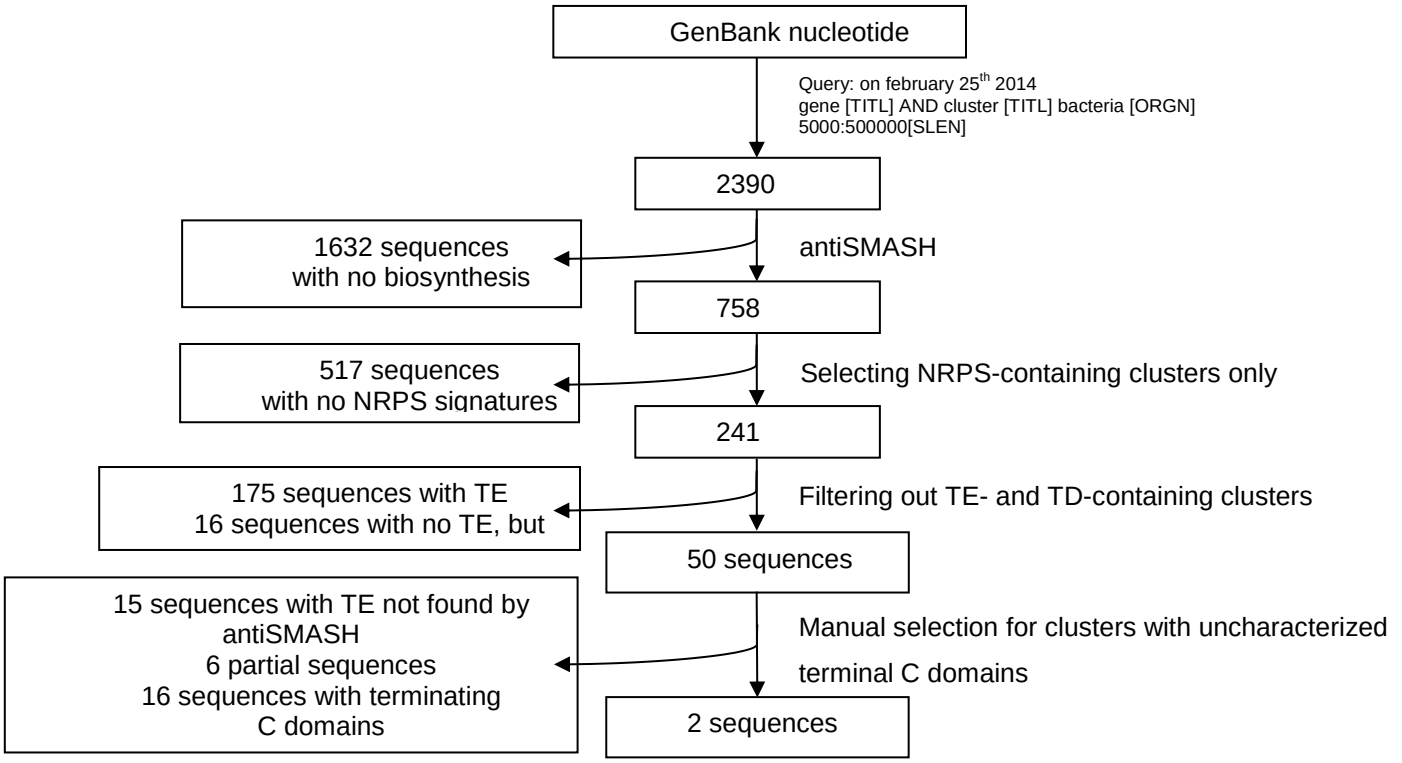


Figure S5, related to Figure 3: Workflow for the identification of NRPS-related gene clusters not encoding apparent release domains. TE: thioesterase domain, TD: terminal reductase domain.

DNA-sequence of *pal*$^{cm\ c5}$: Below, the respective DNA sequence of *pal*$^{cm\ c5}$ (underlined) and 250 bp of the surrounding regions are given:

5':TCCCCTGATCACAGCCGTCCCGGATCTCATCCGGGGCGATGTGCGCTTCCTCGCTGGAGGCGCGCCACCTCGTCCTGGGTG
GCAACGGGGTACGAACTTGCGGGGCTCTGGGCTAGATAGCCTCCCCGTGTCGTCTCCCAACACCCTCGCCCCCGTCCTCCTCG
GCCGTCCCCTCTCCCTTCTCGATCTCGAAGACGTCGCGCGGCGGCGTCGACCAGTCTCCGTGTGTCCCGAGGCTCGCGCCCG
AACGATGGCTTCGCGGCGAGCCATCGACGACATCCTGGAGGCCGGGGATGGTGCCCCCGCCGTTTACGGGGTGAACACTGGC
TTCGGCGCGCTGGCCGAGACGCGCATCGCCGCGAGCGACATCGCCGTCCTCCAGCGCAACCTGGTGCGCAGCCACGCTTCG
GGGGTGGGCCCGGACCTGGGGGATGCCGACGTCCGCGCGATGATCCTTCTGCGGGCGCAGGTCATCGCGCTCGGCTACTCC
GGCGTCCGGCCTCAGGTGCTCGATGCCCTCGTCGGACTGCTGGAGCGGGGAGTCTGCCCACGCATCCCTGCGCAAGGCTCCG
TGGGCGCTTCGGGCGATCTCGCGCCACTCGCGCACCTGGCGCTCACCCTGATCGGGGAGGGCGAGGCGAGGCACGAGGGGA
CCCTTCTTCAGGCGTCGGAGGCGCTCGCGCGCGCGGGGCTCACGCCCGTCGAACTGGTGGCGAAGGAGGGGCTCGCGCTGA
TCAACGGGACCCAGTACATGACGGCGCTCGGGGCGCTGGCCGTGAGTGAAGCGGCCGATCTGTGCGCGCTGGCCGACGTCG
CCGGGGCGATGAGCCTGGAGGCGCTGATGGGCTCGCGTCGCCCCTTCGACGAGCGGCTGATGCACGTGCGCCCCCATCCGG
GGCAGGCCACCGTGGCCCGTAACCTGCGCGCACTGCTCACGGAGAGCGAGATCATGACGTCCCACGCCGGCTGCGCCCGGG
TGCAGGACGCTTACTCGCTCCGGTGCATGCCGCAGGTCCACGGGGCTTCGCGCGACGCGCTCTCCTGGGCGGCCGAGGTGCT
CACCCGCGAGGTGAACAGCGTGACGGACAACCCCACGGTGTTCCTGACGTCAGACGCGCCCGAGCTGCTCTCCGGAGGGAAT
TTCCACGGGCAACCCGTGGCCCTCGCGCTCGATCTCGCCGCCATTGCGGTGGCCGAGCTGGCCAACATCAGCGAGCGCCGCG
TCGAGCAGCTCGTGAACCCGTCGCTCTCGTGCGGGTTGCCTCCGTTCCTCGCGCCGCAGAGCGGGCTGAACTCCGGGTTCAT
GATCGCTCAGGTCGCCAGCGCGGCGCTGGTGAGCGAGAACAAGGTGCTCTGTCATCCGGCGTCCGTGGACTCGATCCCGACC
AGCGCGAACCGCGAGGATCACGTGAGCATGGGCAGCATCTCCGCGCGCAAGCTCTCCCAGGTGGTCGAGAACGTCCGCGCGT
CGATCGCCATCGAACTCCTCGGCGCCGCGCAAGGCATCGATCTGCGCCGCCCCTTGCGCCCGAGCGCCGCCGTGGCAGCAG
CCCACGCCACCGTGCGACGGGTGGTGCCGAAGCTCGAGTCGGACAGGCCGCTCTACCAGGACATCGCCACCGTCACGGCGCT

CATCCGCAGTGGAGAGCTGCTGCGTGAGGTGGAAGCGGTGACGGGTTCGCTGCGGTGAGCGAGCCGCTCCCTCCACCCACTC
CCGCCCCCGGAGAAGCCTCGGAAGGCACCTTGCCGCTGGATGCGGTGCGCGATCTTCTGGGGCTGGTCCGCGCCGTCTATGC
GGCCGCTCGCGCTCGAGGGGCTGCCAGAGTGGAGCTGACGCGGGTCGCGCGGGTGGGTCAGAGGCTGAGCCAGGTGCTCG
AACTCTCGCGCTCACCGCAGCGAGGAAGTGCTGCCGCGGCAGAGGCCTGGGCTCGCCTCGACGA-3'

Table S1, related to Figure 3: *In silico* analysis of NRPS gene clusters with no obvious release function.

Accession number	Organism	Definition	Reference	Classification; Function
AY192157	*Pantoea agglomerans*	andrimid biosynthetic gene cluster, complete sequence	unpublished; Biosynthetic Gene Cluster for the Acetyl-CoA Carboxylase Inhibitor Andrimid. J. Am. Chem. Soc. Aug 23, 2006; 128(33): 10660–10661	TE encoded; AdmM 5.5e-10
AY212249	*Anabaena* sp. 90	hypothetical protein genes, complete cds; microcystin biosynthesis gene cluster, complete sequence; putative transposase (atn1) gene, complete cds	Genes coding for hepatotoxic heptapeptides (microcystins) in the cyanobacterium *Anabaena* strain 90 Appl. Environ. Microbiol. 70 (2), 686-692 (2004)	TE encoded; McyC 4.1e-19
AY652953	*Lyngbya majuscula*	curacin biosynthesis gene cluster,complete sequence	Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacteria Lyngbya majuscula J. Nat. Prod. 67 (8), 1356-1367 (2004)	TE encoded; CurM 4.3e-19
AY735112	*Streptomyces hygroscopicus*	mannopeptimycin biosynthetic genecluster, complete sequence	Biosynthetic pathway for mannopeptimycins, lipoglycopeptide antibiotics active against drug-resistant gram-positive pathogens Antimicrob. Agents Chemother. 50 (6), 2167-2177 (2006)	TE-like beta-lactamase enzyme encoded; MmpK 1.5e-57
DD494859	*Streptomyces virginiae*	virginiamycin M biosynthesis gene, their gene cluster and utilization thereof.	"Virginiamycin M biosynthesis gene, their gene cluster and utilization thereof Patent: JP 2007061045-A 11 15-MAR-2007; OSAKA UNIVERSITY; Characterization of biosynthetic gene cluster for the production of virginiamycin M, a streptogramin type A antibiotic, in *Streptomyces virginiae.* Gene Volume 393, Issues 1–2, 15 May 2007, Pages 31–42	TE encoded; VirJ as reasoned from the publication
GQ859495	*Streptomyces subtropicus* strain CGMCC IA A-182	coelichelin biosynthesis gene cluster, complete sequence	unpublished; Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining Nature Chemical Biology 1, 265 - 269 (2005)	esterase CchJ perfomrs release as reasoned from the publication
GU479979	*Pseudomonas fluorescens* strain BCCM_ID9359	kalimantacin/batumin biosynthesis gene cluster, complete sequence	Isolation and purification of a new kalimantacin/batumin-related polyketide antibiotic and elucidation of its biosynthesis gene cluster Chem. Biol. 17 (2), 149-159 (2010)	TE encoded; Bat3 0.00045
HQ287563	*Streptomyces* sp. DSM 5940	apsamycin biosynthetic gene cluster, complete sequence	Identification of a napsamycin biosynthesis gene cluster by genome mining ChemBioChem 2011 12(3):477-87	TE encoded; NpsP5 3e-13
AJ536156	*Anabaena* sp. 90	microcystin synthesis gene cluster, complete sequence	Characterization of the microcystin synthetase gene cluster from *Anabaena* 90 unpublished	TE encoded; McyC 4.1e-19
AJ586576	*Xanthomonas albilineans*	gene cluster XALB1 for albicidin pathotoxin production, complete sequence	Albicidin pathotoxin produced by Xanthomonas albilineans is encoded by three large PKS and NRPS genes present in a gene cluster also containing several putative modifying, regulatory, and resistance genes Mol. Plant Microbe Interact. 17 (4), 414-427 (2004)	TE encoded; AlbIX 1.3e-19
HM047288	*Pseudomonas* sp. 2663	hypothetical proteins, OmpA family protein, ImcF-related family protein, hypothetical protein, and integrase genes, complete cds; FR901464 gene cluster, complete sequence; andhistone family protein DNA-binding protein and integrase genes, complete cds.	Cloning and elucidation of the FR901464 gene cluster revealing a complex acyltransferase-less polyketide synthase using glycerate asstarter units J. Am. Chem. Soc. 133 (8), 2452-2462 (2011)	TE-like alpha/beta abhydrolase encoded; Fr9I: 3.3e-29
JN982332	*Pseudomonas putida* strain RW10S2	luciferase-like monooxygenase gene,complete cds; WLIP production gene cluster 1, complete sequence; and LysE superfamily transporter protein and major facilitatorsuperfamily transporter protein genes, complete cds.	Genetic and Functional Characterization of Cyclic Lipopeptide White-Line-Inducing Principle (WLIP) Production by Rice Rhizosphere Isolate *Pseudomonas putida* RW10S2 Appl. Environ. Microbiol. 78 (14), 4826-4834 (2012)	split cluster: tandem TE encoded in 2nd part of the cluster (JN982333) *wlpC*

KC297505	*Pseudomonas putida* strain BW11M1	hypothetical protein gene, partial cds; glycosyltransferase, UDP-N-acetylglucosamine 2-epimerase, hypothetical protein, and transcriptional regulator genes, complete cds; xantholysin production first gene cluster, complete sequence; and MarR family protein, O-methyltransferase, and glyoxalase family protein genes, complete cds.	The antimicrobial compound xantholysin defines a new group of pseudomonas cyclic lipopeptides PLoS ONE 8 (5), E62946 (2013)	split cluster: tandem TE encoded in 2nd part of the cluster (KC297506): *xtlC*
EF397502	*Salinispora tropica* strain CNB-476	salinosporamide A biosynthetic gene cluster, partial sequence	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(30):12295-12300	TE encoded; SalF: 9.7e-35
EU199080	*Pseudomonas fluorescens* strain SS101 clone 1	massetolide A biosynthesis gene cluster, partial sequence	Massetolide A biosynthesis in *Pseudomonas fluorescens* J. Bacteriol. 190 (8), 2777-2789 (2008)	split cluster: tandem TE encoded in 2nd part of the cluster (EU199081): *massC*
KF158418	*Streptomyces netropsis* strain CGMCC 4.1650	pyrrolamids biosynthesis gene cluster, partial sequence	unpublished	due to partial sequence no valid statement possible
KF386868	*Streptomyces* sp. MMG1662	phosphonate biosynthetic gene cluster, partial sequence	Diversity and abundance of phosphonate biosynthetic genes in nature Proc. Natl. Acad. Sci. U.S.A. 110 (51), 20759-20764 (2013)	due to partial sequence no valid statement possible
KF386871	*Streptomyces* sp. MMG1522	phosphonate biosynthetic gene cluster, partial sequence.	Diversity and abundance of phosphonate biosynthetic genes in nature Proc. Natl. Acad. Sci. U.S.A. 110 (51), 20759-20764 (2013)	due to partial sequence no valid statement possible
KF386880	*Streptomyces* sp. WM4235	phosphonate biosynthetic gene cluster, partial sequence	Diversity and abundance of phosphonate biosynthetic genes in nature Proc. Natl. Acad. Sci. U.S.A. 110 (51), 20759-20764 (2013)	due to partial sequence no valid statement possible
AF512431	*Saccharothrix mutabilis* subsp. Capreolus	nonribosomal peptide synthetase gene cluster, partial sequence	unpublished	due to partial sequence no valid statement possible
SEG_AY225601S	*Streptomyces vinaceus* strain ATCC 11861	viomycin biosynthesis gene cluster, 5' and 3' fragments	Identification and cloning of genes encoding viomycin biosynthesis from *Streptomyces vinaceus* and evidence for involvement of a rare oxygenase Gene 312, 215-224 (2003)	due to partial sequence no valid statement possible
KF551872	uncultured bacterium clone AR1455	rebeccamycin-like tryptophan dimer gene cluster, partial sequence	Discovery and synthetic refactoring of tryptophan dimer gene clusters from the environment J. Am. Chem. Soc. 135 (47), 17906-17912 (2013)	C domain part of RebD-like chromopyrrolic acid synthase encoded by *asmbl_10*--> believed to catalyze condensation of two tryptophan
AB083344	*Agrobacterium tumefaciens* MAFF301001	*Agrobacterium tumefaciens* ferric iron uptake system gene cluster (agbF, agbC, agbE, agbB, agbA, agbD), complete cds.	Gene cluster for ferric iron uptake in *Agrobacterium tumefaciens* MAFF301001 Genes Genet. Syst. 77 (3), 137-146 (2002)	agbF encodes heterocyclization C domain (BLAST shows EntF homology)--> not included in publication--> function unknown
AM990467	*Planktothrix rubescens* NIVA-CYA 98	NRPS-like 2 gene cluster, strain NIVA-CYA 98	A genome-wide analysis of nonribosomal peptide synthetase gene clusters and their peptides in a *Planktothrix rubescens* strain BMC Genomics 10, 396 (2009)	quote from publication: "NRPS-like gene cluster 2, containing only one NRPS encoding module, could not be linked to oligopeptides"
HQ257512	*Streptomyces* sp. NRRL 30471	muraymycin biosynthetic gene cluster, complete sequence; and flanking sequence.	Identification of the gene cluster involved in muraymycin biosynthesis from *Streptomyces* sp. NRRL 30471 Mol Biosyst 7 (3), 920-927 (2011)	chain ist proposed termination is proposed to be governed by Mur30; a beta-lactamase with homology to CapW catalyzing amide-ester exchange
JN252488	*Streptomyces* sp. ATCC 700974	albomycin biosynthetic gene cluster, complete sequence	Biosynthesis of albomycin delta(2) provides a template for assembling siderophore and aminoacyl-tRNA synthetase inhibitor conjugates ACS Chem. Biol. 7 (9), 1565-1575 (2012)	last and only NRPS AbmQ (A-T-C-T-C-T) assembels tripeptide--> proposed release either by AbmR or AbmC
AF034152	*Mycobacterium smegmatis*	exochelin gene cluster, ExiT (exiT) and FxbB (fxbB) genes, complete cds; and FxbC (fxbC) gene, partial cds	Exochelin genes in *Mycobacterium smegmatis*: identification of an ABC transporter and two non-ribosomal peptide synthetase genes Mol. Microbiol. 29 (2), 629-639 (1998)	release mechanism unclear; no additional C domain found
KF594335	*Streptomyces regensis* strain WC-3844	phosphonate biosynthetic gene cluster, complete sequence	Cyanohydrin Phosphonate Natural Product from *Streptomyces regensis*. J. Nat. Prod. 77 (2), 243–249 (2014)	release mechanism unclear; one single standing C domain encoded by *orf5*; most likely not involved in phosphonate biosynthesis-->phosphonates like phosphinotricin tripeptide contain two type II TEs and one of them is belived to be involved in the release and the other for proofreading
AB476988	*Streptomyces griseus* SANK60196	A-500359s biosynthetic gene cluster	Identification of the biosynthetic gene cluster of A-500359s in *Streptomyces griseus* SANK60196 J. Antibiot. 62 (6), 325-332 (2009)	release mechanism unclear; no terminal C domain present
AB538860	*Streptomyces* sp. SANK 62799	A-503083 biosynthetic gene cluster	unpublished	release mechanism unclear; no terminal C domain present

KF170332	*Streptomyces ansochromogenes* clone	nrps4 metabolite biosynthetic gene cluster, complete sequence	Assembly and features of secondary metabolite biosynthetic gene clusters in *Streptomyces ansochromogenes* Sci China Life Sci 56 (7), 609-618 (2013)	free standing C domain encoded by ctg1_orf00013 most likely belongs to a split NRPS module with ctg1_orf00012 (A- PCP)
GQ370384	*Pseudomonas* sp. J465	Pseudomonas sp. J465 CysB gene, partial cds; UspA domain-containing protein, 5' nucleotidase, 2-dehydropantoate 2-reducatse, hypothetical protein, PilZ, phospho-2-dehydro-3-deoxyheptonate aldolase, and MFS-family transporter genes, complete cds; ddd gene cluster, complete sequence; NHL repeat-containing protein, benzoate transporter, lysine exporter protein, MarR, GntR, glutathione S-transferase, UvrD/REP helicase, isochorismate synthase, enterobactin synthase subunit E, and isochorismatase genes, complete cds; and enterobactin synthetase component F gene, partial cds.	Identification of genes for dimethyl sulfide production in bacteria in the gut of Atlantic Herring (*Clupea harengus*) ISME J 4 (1), 144-146 (2010)	genes for dimethyl sulfide production--> ACY02015.1 encodes C-A module with homology to enterobactin synthetase component F
EF552687	*Streptomyces albus* strain JA3453	oxazolomycin biosynthetic gene cluster, complete sequence.	Oxazolomycin Biosynthesis in Streptomyces albus JA3453 Featuring an "Acyltransferase-less" Type I Polyketide Synthase That Incorporates Two Distinct Extender Units J. biol. Chem. 26: 20097–20108 (2010)	C domain of last module OzmL (C-A-MT-PCP-C) most likely catalyzes cyclization of the Ser side chain to form a β-lactone ring thereby terminating biosynthesis
AB684619	*Streptomyces rochei*	streptothricin biosynthetic gene cluster, strain: NBRC 12908	A stand-alone adenylation domain forms amide bonds in streptothricin Nat. Chem. Biol. 8 (9), 791-797 (2012)	ORF 18 catalyzes the condensation of L-β-lysine oligopeptides covalently bound to ORF 18 with a freely diffusible intermediate to release the streptothrisamine products
AB684620	*Streptomyces lavendulae* subsp. lavendula	streptothricin biosynthetic gene cluster, strain: NBRC 12789	A stand-alone adenylation domain forms amide bonds in streptothricin Nat. Chem. Biol. 8 (9), 791-797 (2012)	ORF 18 catalyzes the condensation of L-β-lysine oligopeptides covalently bound to ORF 18 with a freely diffusible intermediate to release the streptothrisamine products
KC935381	*Streptomyces* sp. TP-A0356	streptothricin biosynthetic gene cluster, partial sequence	Mining of a streptothricin gene cluster from *Streptomyces* sp. TP-A0356 genome via heterologous expression Sci China Life Sci 56 (7), 619-627 (2013)	see streptothricin above
AM071396	*Planktothrix agardhii* NIVA-CYA 126	aeruginoside biosynthesis gene cluster	Biosynthesis and structure of aeruginoside 126A and 126B cyanobacterial peptide glycosides bearing a 2-carboxy-6-hydroxyoctahydroindole moiety Chem. Biol. 14 (5), 565-576 (2007)	second C domain of AerG (C-A-PCP-C-PCP) is proposed to release the assembeled peptide chain via condensation with free an agmatine molecule
FJ609416	*Microcystis aeruginosa* NIES-98	aeruginosin synthetase gene cluster, complete sequence.	Plasticity and evolution of aeruginosin biosynthesis in cyanobacteria. Appl. Environ. Microbiol. 75 (7), 2017-2026 (2009)	C domain of AerG2 (C-PCP) is proposed to release the assembeled peptide chain via condensation with free an agmatine molecule
AY048670	*Streptomyces globisporus*	enediyne antitumor antibiotic C-1027 biosynthetic gene cluster, complete sequence.	A free-standing condensation enzyme catalyzing ester bond formation in C-1027 biosynthesis Proc. Natl. Acad. Sci. U.S.A.106 (11):4183-8 (2009)	free standing C domain Sgc5 catalyzes transersterification thereby releasing the compound
DL080157	*Streptomyces* sp	Biosynthetic Gene Cluster for the Production of a Complex Polyketide.	Biosynthetic Gene Cluster for the Production of a Complex Polyketide Patent: JP 2008501342-A 1 24-JAN-2008; Wyeth	MerP (C-A-PCP-C) is proposed to be responsible for addition of the pipecolate unit to the polyketide chain thereby releasing the compound with subsequent formation of the macrocycle
DQ351275	*Streptomyces* sp. NRRL 30748	meridamycin biosynthetic gene cluster and flanking regions.	Isolation and characterization of meridamycin biosynthetic gene cluster from Streptomyces sp. NRRL 30748. Gene 377, 109-118 (2006)	MerP (C-A-PCP-C) is proposed to be responsible for addition of the pipecolate unit to the polyketide chain thereby releasing the compound with subsequent formation of the macrocycle
DQ885223	*Streptomyces violaceusniger*	meridamycin gene cluster, complete sequence.	Organization of the biosynthetic gene cluster in Streptomyces sp. DSM 4137 for the novel neuroprotectant polyketide meridamycin. Microbiology 152 (PT 12), 3507-3515 (2006)	MerP (C-A-PCP-C) is proposed to be responsible for addition of the pipecolate unit to the polyketide chain thereby releasing the compound with subsequent formation of the macrocycle

EF032505	*Streptoalloteichus hindustanus* strain ATCC 31158	tallysomycin biosynthetic gene cluster, complete sequence	The tallysomycin biosynthetic gene cluster from *Streptoalloteichus hindustanus* E465-94 ATCC 31158 unveiling new insights into the biosynthesis of the bleomycin family of antitumor antibiotics. Mol Biosyst 3 (1), 60-74 (2007)	single standing C domain TlmII most likely catalyzes coupling reaction between the NRPS-bound full length BLM peptide intermediate and the terminal amines
EF484930	*Pseudomonas fluorescens* strain WCS374	siderophore pseudomonine biosynthesis and transport gene cluster, partial sequence.	Three Siderophores from One Bacterial Enzymatic Assembly Line J. Am. Chem. Soc.131 (14), 5056–5057 (2009)	the C domain of PmsG can use synthetic N-OH-histamine as a nucleophilic substrate and release acinetobactin *in vitro*--> mostlikely same mechanism *in vivo*
VCU52150	*Vibrio cholerae*	vibriobactin gene cluster, complete sequence	Cloning of a Vibrio cholerae vibriobactin gene cluster: identification of genes required for early steps in siderophore biosynthesis;J. Bacteriol. 179 (22), 7055-7062 (1997); Vibriobactin biosynthesis in Vibrio cholerae:VibH is an amide synthase homologous to nonribosomal peptide synthetasecondensation domains. Biochemistry. (50):15513-21 (2000)	VibH as a novel amide synthase couples 2,3-DHB thioesterified to holo-VibB to a primary amine of norspermidinewhile regenerating holo VibB
X86780	*Streptomyces hygroscopicus*	gene cluster for polyketide immunosuppressant rapamycin	The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin Proc. Natl. Acad. Sci. U.S.A. 92 (17), 7839-7843 (1995)	RapP (C-A-PCP-C) is proposed to be responsible for addition of the pipecolate unit to the polyketide chain thereby releasing the compound from last ACP of Rap3 with subsequent formation of the macrocycle
AM231613	*Mycobacterium chelonae* type strain CIP104535T.	*Mycobacterium chelonae* GPL gene cluster, genes involved in the biosynthesis of GPIs	Genomics of glycopeptidolipid biosynthesis in *Mycobacterium abscessus* and *M. chelonae* BMC Genomics 8, 114 (2007)	single standing C domain PapA3 is proposed to catalyze the transfer of the Pks-bound fatty acid to the tripeptide
AM231619	*Mycobacterium abscessus* type strain CIP104536T	*Mycobacterium abscessus* GPL gene cluster, genes involved in the biosynthesis of GPL	Genomics of glycopeptidolipid biosynthesis in *Mycobacterium abscessus* and *M. chelonae* BMC Genomics 8, 114 (2007)	single standing C domain PapA3 is proposed to catalyze the transfer of the Pks-bound fatty acid to the tripeptide
AB070942	*Streptomyces avermitilis*	polyketide-2 biosynthetic gene cluster	Genome sequence of an industrial microorganism *Streptomyces avermitilis*:deducing the ability of producing secondary metabolites Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)	the modular and uncharacterized PKS BAB69217.1 ceases with a C domain
FN547928	*Chondromyces crocatus* Cm c5	crocacin biosynthetic gene cluster	unpublished	CroK (C-A-PCP-C) ends with C domain with hydrolytic function

In order to detect NRPS gene clusters that do not encode an obvious release function an *in silico* analysis was employed (see Figure S5). The resulting 50 clusters were further manually examined regarding the presence of uncharacterized terminal C domains using the Pfam database (Punta et al., 2012). 15 sequences contained a TE domain although they were not found by antiSMASH (highlighted in blue), 6 sequences were only partially available (highlighted in green) and 11 sequences were not encoding a terminal C domain or have been not linked to secondary metabolism (highlighted in purple or brown). 16 of the 18 remaining sequences encode C domains which are proposed to take part in chain termination (highlighted in red). After all, one sequence from *Streptomyces avermitilis* and one from *C. crocatus* Cm c5 remained (not highlighted).

Table S2, related to experimental procedures: Oligonucleotides used for mutagenesis studies

oligonucleotide	sequence (5'→3')	purpose
pSUPHyg_for	ATG TAG CAC CTG AAG TCA GCC	verification
pSUPHyg_rev	ACG CAT ATA GCG CTA GCA GC	verification
orf2_KO_for	CAC GTA GTT CAC CTT GAT G	inactivation
orf2_KO_rev	GAT TGA TGC CAT TCT GCT	inactivation
orf2_KO_check_up	CGT TCG AGG GAT ACA TCT CG	verification
orf2_KO_check_down	CCT GGC AGA GCA TCT GG	verification
croE_KO_for	GTG GTG AAT CGA GAA GGT A	inactivation
croE_KO_rev	GTA AGA GAA GAG GGA AAT GC	inactivation
croE_KO_check_up	ATC CTC CCG CTG ACG TCC	verification
croE_KO_check_down	AGC AAC CGT CTT CCA GCG	verification
croF_KO_for	CTC GAC TGA GCC GCG CTG	inactivation
croF_KO_rev	CGA GCT TCT CGA ACC AGG	inactivation
croF_KO_check_up	GAA CGA GCA GCT TCG CAG C	verification
croF_KO_check_down	ATC ATC CAT CCG GAT CTC G	verification
croI_KO_for	GCG GAT AGC GTT CCT CTT C	inactivation
croI_KO_rev	TGT GAG AGC CAT ACG GTC GG	inactivation
croI_KO_check_up	AAG CCC TGC GAG AAC TCG	verification
croI_KO_check_down	GCT CAC GGG TCC TTC CC	verification
croK-C2_KO_for	TCC CTG CTC CAG AGC CTG CC	inactivation
croK-C2_KO_rev	ACA GGT CGT CCA CGG CTT CC	inactivation
croK-C2_KO_check_up	TCA CCG AGC TCT GGC AGC	verification
croK-C2_KO_check_down	AAG ACG CGC ACC AGC TCC	verification
croL_KO_for	GTA CTG CCG ATG TTT ATG A	inactivation
croL_KO_rev	GTG TGC AGG TAG AAC TTG	inactivation
croL_KO_check_up	GTT GCA GAA GAT CCT GAG G	verification
croL_KO_check_down	TCG ACG CGA GGA CTC AGG	verification
orf3_KO_for	GGT CAG ATG GCT GAT GTT C	inactivation
orf3_KO_rev	GAA GAT CAC GTC GCT CTC	inactivation
orf3_KO_check_up	GCT GTC GTC CGA GTT CC	verification
orf3_KO_check_down	AGA TCC TGC AGA CGC GC	verification
orf4_KO_for	CTC TTG TCT GTT CGA GAA TG	inactivation
orf4_KO_rev	GAC GAT GAG ATC CTT CAG AC	inactivation
orf4_KO_check_up	ATG AGC GCA ACG GTC CC	verification
orf4_KO_check_down	TCG TGC TCC TCG GCG ACC	verification
orf5_KO_for	CAT GAT CGA GCA TTC ATC	inactivation
orf5_KO_rev	CTT GAG AGC AGG AAA CAG	inactivation
orf5_KO_check_up	GCT CGC AGT CCG CAA GG	verification
orf5_KO_check_down	TTC TTC GTG CGA GAA CCG	verification
orf6_KO_for	AGC ATC TGA ACG CTC TTC	inactivation
orf6_KO_rev	GGA TGT CCC AAT CGA AG	inactivation
orf6_KO_check_up	ATT GGC TGT CGC ATC TGG	verification
orf6_KO_check_down	CCA CAC GGA GAC CGA CG	verification
palcmc5_KO_for_*Eco*RV	AAG GAT ATC ATC GAC GAC ATC CTG TGA GG	inactivation

oligonucleotide	sequence (5'→3')	purpose
palcmc5_KO_rev_*Hind*III	TCA CAA GCT TCA CCA CCC GTC GCA CGG	inactivation
palcmc5_KO_check_up	ACG AAC TTG CGG GGC TCT GG	verification
palcmc5_KO_check_down	GCG AGA GTT CGA GCA CCT GG	verification

Supplemental Experimental Procedures

Strains and culture conditions

C. crocatus Cm c5 was cultivated in Pol03 media (0.3 % probion, 0.3% soluble starch, 0.05 % $CaCl_2 \cdot 2\ H_2O$, 0.2 % $MgSO_4 \cdot 7\ H_2O$, 50 mM HEPES pH: 7.2) at 30°C and 160 rpm. Feeding experiments with labeled precursors were performed in 25 ml Pol03 medium in a 100 ml flask using 1 % (v/v) XAD-16. Labeled amino acids (5 mg) were dissolved in 250 mM HCl (400 µl) and sterile filtered. The labeled precursor stock was added in four portions to the culture. 100 µl were added on the first and third day followed by another 100 µl on the fifth and sixth day of cultivation, respectively. Labeled benzoic acid and *trans*-cinammic acid were dissolved in 100 µl DMSO and added in 25 µl portions to the respective culture on the same days as for the amino acids. The cultures were harvested after 7 days and extracted with 2 x 30 ml of methanol. The organic solvent was removed and the residuals dissolved in 1 ml of methanol prior to LC-MS measurement using the standard method (see below). For mutants, Pol03 medium was additionally supplemented with 100 µg/ml hygromycin B. For metabolic analysis of respective strains, 500 µl of cells clumps were used to inoculate 50 mL Pol03 medium in 300 ml flasks and cultivation was carried out for at least 5 days while using three independent clones for every targeted inactivation. Afterwards the cells were separated from the medium, shock frozen in liquid nitrogen and lyophilized. Methanolic extracts were prepared by extracting the dried cells with two times of 30 ml methanol, followed by evaporation of the solvent. Finally the extract was dissolved in 1 ml of methanol. The *Escherichia coli* strain DH10B used for cloning and propagation of plasmids was cultivated in Luria Bertani (LB) medium at 37°C supplemented with 70 µg/ml hygromycin B, when necessary.

Gene disruption experiments in *C. crocatus* Cm c5

Genetic modification of strain *C. crocatus* Cm c5 was accomplished according to a previously described protocol (Rachid et al., 2006). Therefore and internal fragment of the target gene was amplified from genomic DNA of *C. crocatus* Cm c5 by PCR and either subcloned into pCR TOPO II vector or directly integrated into the prepared suicide vector pSUP_Hyg. After

verification of the respective plasmids by sequencing they were introduced into *E. coli* strain ET12567 additionally carrying the helper plasmid pUB307. Selection was done using 25 µg/ml kanamycin sulfate 25 µg/ml chloramphenicol and 70 µg/ml hygromycin B. The transfer of the respective pSUP_Hyg derivatives into *C. crocatus* Cm c5 was achieved by biparentale mating with the mentioned *E. coli* strain ET12567 pUB307. Selection for *C. crocatus* Cm c5 clones was carried out on Pol03 agar plates containing 100 µg/ml hygromycin B and 60 µg/ml spectinomycin. Putative mutants were transferred onto new selection plates and further cultivated in liquid medium both containing 100 µg/ml hygromycin B. Genotypic verification of respective clones was done by PCR as shown in Figure S2 L.

Construction of pSUP_Hyg derivatives used in this study

For generation of all inactivation constructs except pSUP_palcmc5_KO, the homology fragments (for the oligonucleotides used see Table S2) have been subcloned cloned into pCR2.1TOPO (Invitrogen), subsequently recovered by restriction digestion (*Eco*RV and *Hind*III) and inserted into the prepared vector pSUPHyg. In case of pSUP_palcmc5_KO the PCR fragment was directly cloned into pSUPHyg by using *Eco*RV and *Hind*III.

N-acetylcysteaminyl-crocacin B

$[\alpha]_D^{25}$ = + 79.8 (*c* = 0.2 in MeOH), UV (MeOH): λ_{max} (lg ε) = 202 nm (4.62), 254 (4.50), 282 (4.30), 291 (4.27); ^{1}H NMR (600 MHz, CD_3OD): similar to crocacin B (Jansen et al., 1999), but δ = 1.92 (s, 3H, 2''-H), 3.01 (t, 6.6 Hz, 2H, 1'-H), 3.31 (t, 6.6 Hz, 2H, 2'-H); ^{13}C NMR (150 MHz, CD_3OD): similar to crocacin B (Jansen et al., 1999), but δ = 199.1 (s, C-1), 173.5 (s, C-1''), 39.9 (t, C-2'), 29.0 (t, C-1'), 22.5 (q, C-2''); HRESIMS: *m/z* 624.3102 $[M - H]^-$ (calcd for $C_{34}H_{46}N_3O_6S$, 624.3113), 670.3157 $[M + HCO_3]^-$ (calcd for $C_{35}H_{48}N_3O_9S$, 670.3168)

LC-MS standard method

Samples are measured on a Dionex Ultimate 3000 RSLC system using a Waters BEH C18, 50 x 2.1 mm, 1.7 µm d_p. Separation was achieved by a linear gradient with (A) H_2O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 600 µL/min and 45 °C. The gradient was initiated by a 0.33 min isocratic step at 5 % B, followed by an increase to 95 % B in 9 min to end up with a 1 min flush step at 95 % B before reequilibration under the initial conditions. Coupling the HPLC to the MS was supported by an Advion Triversa Nanomate nano-ESI system attached to a Thermo

Scientific Orbitrap. Mass spectra were acquired in centroid mode ranging from 200 – 2000 m/z at a resolution of R = 30000.

In silico analysis of NRPS-encoding gene clusters with no apparent termination domain

Gene clusters of bacterial origin with function in secondary metabolite biosynthesis were screened for the presence of thioesterase domains as follows (Figure S5). First, GenBank nucleotide database was queried on 25.02.2014 with "gene [TITL] AND cluster [TITL] bacteria [ORGN] 5000:500000 [SLEN]" search string, resulting in 2390 sequences matching these criteria. Sequences were downloaded and processed through antiSMASH 2.0 tool (Blin et al., 2013) with default parameters, yielding 758 clusters. On these, several levels of filtering were applied to select for gene clusters that contain NRPS signatures, yet do not possess thioesterase (TE) and terminal reduction (TD) domains. The remaining 50 sequences have been analyzed manually using Pfam database (Punta et al., 2012).

Supplemental Reference List

Blin, K., Medema, M.H., Kazempour, D., Fischbach, M.A., Breitling, R., Takano, E., and Weber, T. (2013). antiSMASH 2.0 - A versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Research. 41, W204-W212.

Jansen, R., Washausen, P., Kunze, B., Reichenbach, H., and Hofle, G. (1999). Antibiotics from gliding bacteria, LXXXIII - The crocacins, novel antifungal and cytotoxic antibiotics from *Chondromyces crocatus* and *Chondromyces pediculatus* (Myxobacteria): Isolation and structure elucidation. European Journal of Organic Chemistry. 5, 1085-1089.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2007). Clustal W and clustal X version 2.0. Bioinformatics. 21, 2947-2948.

Marahiel, M.A., Stachelhaus, T., and Mootz, H.D. (1997). Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. Chemical Reviews. 7, 2651-2674.

Chapter II

Microsclerodermins from terrestrial myxobacteria: An intriguing biosynthesis likely connected to a sponge symbiont

Thomas Hoffmann, **Stefan Müller**, Suvd Nadmid, Ronald Garcia and Rolf Müller (2013)

Journal of the American Chemical Society, 135 (45), page 16904–16911

DOI: http://dx.doi.org/10.1021/ja4054509

Abstract

The microsclerodermins are unusual peptide natural products exhibiting potent antifungal activity reported from marine sponges of the genera *Microscleroderma* and *Theonella*. We here describe a variety of microbial producers of microsclerodermins and pedeins among myxobacteria along with the isolation of several new derivatives. A retro-biosynthetic approach led to the identification of microsclerodermin biosynthetic gene clusters in genomes of *Sorangium* and *Jahnella* species, allowing for the first time insights into the intriguing hybrid PKS/NRPS machinery required for microsclerodermin formation. This study reveals the biosynthesis of a "marine natural product" in a terrestrial myxobacterium where even the identical structure is available from both sources. Thus, the newly identified terrestrial producers provide access to additional chemical diversity; moreover, they are clearly more amenable to production optimization and genetic modification than the original source from the marine habitat. As sponge metagenome data strongly suggest the presence of associated myxobacteria, our findings underpin the recent notion that many previously described "sponge metabolites" might in fact originate from such microbial symbionts.

Introduction

Natural products have a longstanding tradition as leads for the development of new medicines.[1] In addition to well-established and extensively investigated plant, fungal, and bacterial producers of secondary metabolites, newer screening campaigns increasingly include organisms from less studied taxa and previously underexploited habitats such as terrestrial myxobacteria and marine sponges.[2–5] Their potential as sources of novel chemical scaffolds has been clearly demonstrated and despite the impressive structural diversity originating from these organisms, the overall picture has emerged that structural types obtained from phylogenetically distant producers usually show little overlap.[6] However, as an exception to this general notion the production of several strikingly similar compounds by unrelated species has also been reported. Some of these findings are parallel discoveries of initially sponge-derived metabolite classes from microbial sources, leading to the assumption that the respective natural products might in fact be produced by bacterial sponge symbionts.[7–9] Support for this theory comes from the identification of filamentous bacteria growing within intercellular space inside the sponge.[8,10] However, studies which unambiguously prove the production of a "sponge metabolite" by a symbiotic bacterium are exceedingly rare.[10,11] The same holds true for marine natural products of other host organisms.[12–14] This shortcoming may be attributed to difficulties with isolation and cultivation of symbiotic microbes under laboratory conditions. Notably, the ability to

independently cultivate the "real" producer of a specific secondary metabolite holds great promise, not only for sustained production but also for improving yields using both biotechnological and genetic engineering approaches. These opportunities present an invaluable advantage when further investigating a compound of interest, as the marine organism itself usually faces critical supply limitations and is poorly amenable to genetic manipulation. Moreover, access to a microbial producer facilitates the identification of biosynthetic genes underlying the formation of the metabolite of interest - a crucial prerequisite for understanding the biosynthetic machinery and a promising step toward transferring these genes into a suitable heterologous expression host.[15]

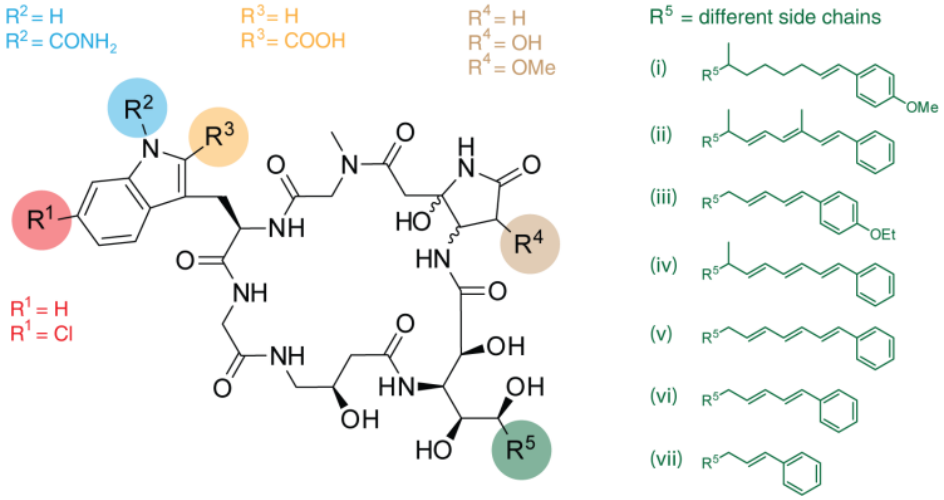


Figure 1: The microsclerodermin scaffold with an overview of the different residues identified so far. Groups R1-R4 are related to the presence of tailoring enzymes during biosynthesis whereas the side chain R5 is derived by the PKS part of the biosynthetic machinery. The pyrrolidone is reported to have R,R- or S,S-configuration, respectively.

Looking at those natural products from marine sources having apparent microbial counterparts, several cases exist where structures of myxobacterial secondary metabolites are indeed strikingly similar to previously discovered sponge-derived compounds. For example, the cyclodepsipeptide jaspamide (jasplakinolide) isolated from the marine sponge *Jaspis.*[16,17] is closely related to the structure of chondramides produced by the myxobacterium *Chondromyces crocatus* Cm c5, suggesting that the biosynthetic pathways responsible for production of these molecules should be largely similar.[18] The same holds true for renieramycin and saframycin MX1, isolated from a *Reniera* sponge and a myxobacterium of the genus *Myxococcus*.[19,20] Moreover, the macrolides salicylihalamide and apicularen were isolated from a *Halicona* sp. sponge and a *Chondromyces* species, respectively.[21,22] Very recently, bengamides were described from a *Jaspis* sponge and a cultured myxobacterium.[23,24] Adding to the list of

"biosynthetic look-alikes", the structure of pedein from the terrestrial myxobacterium *Chondromyces pediculatus* Cm p3 closely resembles that of microsclerodermin,[25] which was isolated in 1994 from *Microscleroderma* sp., a lithistid sponge harvested in New Caledonia.[26] Upon their finding of pedeins in myxobacteria, Kunze *et al.* suggested that the origin of microsclerodermins could be a bacterial sponge symbiont closely related to myxobacteria.[25] Indeed, pedein and microsclerodermin are highly similar, and both exhibit potent antifungal activity. To date several new derivatives belonging to the microsclerodermin class of peptides have been identified from various *Microscleroderma* species as well as from a *Theonella* sponge.[27–30] Nevertheless, the biosynthetic machinery behind this natural product remains so far elusive.

Table 1: Overview of Different Microsclerodermins and Pedeins and Their Origin

derivative	R^1	R^2	R^3	R^4	R^5	pyrrolidone config	sum formula	$(M+H)^+$ [m/z]	[*a*]	[*b*]	[*c*]	[*d*]	[*e*]	ref.
A	H	H	COOH	OH	i	S, S	$C_{47}H_{62}N_8O_{16}$	995.4357	•					26,29
B	H	H	COOH	H	i	S, S	$C_{47}H_{62}N_8O_{15}$	979.4407	•					26,29
C	Cl	$CONH_2$	H	H	vii	R, R	$C_{41}H_{50}N_9O_{13}Cl$	912.3289	•	•				27
D	Cl	H	H	H	vii	R, R	$C_{40}H_{49}N_8O_{12}Cl$	869.3231	•	•	•	•		27, this study
E	H	H	COOH	H	iii	R, R	$C_{45}H_{54}N_8O_{14}$	931.3832	•					27
F + G [f]	H	H	H	H	iv	R, R	$C_{45}H_{56}N_8O_{12}$	901.4090	•					28
H + I [f]	H	H	H	H	ii	R, R	$C_{46}H_{58}N_8O_{12}$	915.4247	•					28
J	H	H	H	H	i	S, S	$C_{46}H_{60}N_8O_{12}$	917.4403	•					29
K	H	H	H	OH	i	S, S	$C_{46}H_{60}N_8O_{13}$	933.4353	•					29
L	Cl	H	H	OMe	vii	R, R	$C_{41}H_{51}N_8O_{13}Cl$	899.3337			•	•		this study
M	H	H	H	H	v	R, R	$C_{44}H_{54}N_8O_{12}$	887.3934					•	this study
Pedein A [g]	Cl	H	H	OMe	vi	R, R	$C_{43}H_{53}N_8O_{13}Cl$	925.3493			•	•		25, this study
Pedein B [g]	H	H	H	OMe	vi	R, R	$C_{43}H_{54}N_8O_{13}$	891.3883			•	•		25, this study

[a] *Microscleroderma* sp. (3 species) [b] *Theonella* sp. (1 species) [c] *Chondromyces* sp. (2 species) [d] *Jahnella* sp. (2 species) [e] *Sorangium* sp. (11 species) [f] The tryptophan side chain is reduced to an α-β-unsaturated amino acid. [g] Based on their same biosynthetic origin, we implicitly include pedeins when referring to the microsclerodermin family in this study.

In this study we present several terrestrial myxobacteria as alternative producers of microsclerodermins and pedeins. Our data show that *Jahnella* and *Chondromyces* species can produce the identical derivate also known from a *Microscleroderma* species. In addition, they produce new derivatives not previously reported from other sources. Access to genomic sequences for two myxobacterial producers allowed us to establish for the first time a biosynthetic model for microsclerodermin formation and also provided us with an opportunity to probe the molecular basis responsible for the structural diversity observed from microsclerodermins. Moreover, it was shown that the myxobacterial pedeins[25] originate from the same biosynthetic machinery as the microsclerodermins; hence, they belong to the same

compound family. Taken together with recent metagenomic studies providing evidence that myxobacterial taxa may even exist as sponge symbionts,[31] our results underpin the assumption that a myxobacterium is the real biosynthetic source of the "marine" natural product microsclerodermin.

Experimental Section

Bacterial Strains and Culture Conditions

Sorangium cellulosum So ce38 was cultivated in H-medium (2 g/L soybean flour, 2 g/L glucose, 8 g/L starch, 2 g/L yeast extract, 1 g/L $CaCl_2 \cdot 2H_2O$, 1 g/L $MgSO_4 \cdot 7H_2O$, 8 mg/L Fe-EDTA, 50 mM HEPES, adjusted to pH 7.4 with 10 N KOH). Mutants of *S. cellulosum* So ce38 were cultivated in H-medium supplemented with hygromycin B (100 µg/mL) and 1 % (w/v) adsorber resin (XAD-16, Rohm & Haas) at 180 rpm and 30 °C. *Jahnella* sp. MSr9139 was cultivated in buffered yeast broth medium VY/2 (5 g/L baker`s yeast, 1 g/L $CaCl_2 \cdot 2H_2O$, 5 mM HEPES pH 7.0 with 10 N KOH) at 180 rpm and 30 °C.[32] The *Escherichia coli* strains DH10B and ET12567 harboring the plasmids pUB307 and pSUP*mscH*_KO for conjugation purposes were cultivated in Luria-Bertani (LB) medium at 37°C. Transformation of strains was performed according to the standard methods described elsewhere.[33] Antibiotics were added with the following final concentrations: chloramphenicol (25 µg/mL), kanamycin sulfate (25 µg/mL) and hygromycin B (100 µg/mL).

Disruption of the *msc*H *L*ocus in So ce38

Gene disruption in So ce38 using biparental mating was carried out according to a previously established protocol.[34] For construction of the plasmid pSUP*mscH*_KO a homologous fragment with the size of 2472 bp was amplified from genomic DNA using the oligonucleotides *mscH*_KO_for (GAT CCA GCG CTG GTT CCT CG) and *mscH*_KO_rev (ACG AGG CTG TCG AAG AGC G) and cloned into pCR-TOPO II-vector, resulting in the plasmid pTOPO_*mscH*_KO. The genomic segment was subsequently recovered from this plasmid using the restriction enzymes *Hind*III and *Eco*RV and further integrated into the prepared vector pSUPHyg.

Isolation of Microsclerodermin M from So ce38

The production medium for So ce38 was P38X medium (2 g/L peptone, 2 g/L glucose, 8 g/L starch, 4 g/L probion, 1 g/L $CaCl_2 \cdot 2H_2O$, 1 g/L $MgSO_4 \cdot 7H_2O$, 8 mg/L Fe-EDTA, 50 mM HEPES, adjusted to pH 7.5 with 10 N KOH). A 100 L fermenter with 2 % (w/v) XAD-16 adsorber resin

(Rohm & Haas) was harvested after 14 days of fermentation. The cells were removed from the XAD before extraction with 3 x 3 L of methanol followed by 1 x 3 L of acetone. The combined fractions yielded 47.2 g dry weight of crude extract. Five grams of this extract was suspended in cold water, the suspension was centrifuged immediately, and the remaining pellet was dissolved in DMSO/MeOH (1:1, v/v) to give a product-enriched solution which was subjected to preparative HPLC using a Waters Autopurifier System equipped with a Waters XBridge C18, 150 x 19 mm, 5 µm d_p column operated at room temperature. The gradient started at 30 % B, increased to 50 % B in 2 min and to 51 % B in another 2 min before increasing to 95 % B in 4 min for column flushing. The combined fractions of interest were lyophilized, dissolved in DMSO/MeOH (1:1, v/v), and forwarded to a semipreparative Dionex HPLC system (P680 pump, TCC100 thermostat, and PDA100 detector) equipped with a Phenomenex Fusion C18, 250 x 4.6 mm, 4 µm d_p column. Separation was achieved by a linear gradient using (A) H_2O and (B) ACN at a flow rate of 5 mL/min and 30 °C. The gradient started at 10 % B and increased to 30 % B in 3 min, followed by an increase to 38 % B in 15 min (0.9 % B/column volume). UV data were acquired at 316 nm. A maximum of 100 µL of the sample was manually injected before fraction collection, yielding 8.1 mg of microsclerodermin M. Microsclerodermin M: white amorphous solid, $[\alpha]_D^{20}$ - 55.7 ° (c 0.10, DMSO/MeOH 8:2).

Isolation of Microsclerodermins from MSr9139

The strain MSr9139 was cultivated in 3 x 1 L shaking flasks containing 500 mL of buffered VY/2 medium for 30 days. The medium was changed every 24 h by pipetting out the liquid broth. The cell pellet was harvested by centrifugation and lyophilized overnight, followed by extraction with 3 x 300 mL of methanol. The combined fractions yielded an orange-brown crude extract which was further partitioned between hexane and MeOH/H_2O 7:3 (v/v) to yield 170 mg of crude extract out of the aqueous phase. Subsequently, the extract was purified by semipreparative HPLC using an Agilent 1260 Infinity system (G1311C quaternary pump, G1330B thermostat, G1315D DAD detector and G1328C manual injector) equipped with a Phenomenex Jupiter Proteo, 250 x 10 mm, 4 µm d_p column. Separation was achieved by a linear gradient using (A) H_2O and (B) ACN at a flow rate of 2.5 mL/min and 22 °C. The gradient started at 20 % B and increased to 50 % B in 35 min (5.7 % B/column volume). UV data were acquired at 280 nm. A maximum of 100 µL of the sample was manually injected before fraction collection yielding 0.7 mg of microsclerodermin D, 0.45 mg of microsclerodermin L, and 0.85 mg of pedein A. Microsclerodermin L: white amorphous solid, $[\alpha]_D^{20}$ - 77.7 ° (c 0.12, MeOH).

LC-MS data acquisition

All measurements were performed on a Dionex Ultimate 3000 RSLC system using a BEH C18, 100 x 2.1 mm, 1.7 µm d_p column (Waters, Germany). Separation of 1 µL sample was achieved by a linear gradient from (A) H2O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 600 µL/min and 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 min to end up with a 2 min step at 95 % B before reequilibration under the initial conditions. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 75 µL/min before entering the maXis 4G hr-ToF mass spectrometer (Bruker Daltonics, Germany) using the Apollo II ESI source. Mass spectra were acquired in centroid mode ranging from 150 – 2500 m/z at 2 Hz scan rate.

16S rRNA Gene and Phylogenetic Analysis

Extraction of the 16S rRNA gene was performed in representative microsclerodermin producing strains of *Sorangium, Jahnella* and *Chondromyces*. Sequences of other myxobacterial strains used in the analysis were obtained from GenBank. The 16S rRNA gene was amplified using a set of universal primers, and phylogenetic analysis was performed as described in a previous study, but using the MUSCLE alignment algorithm and Neighbor-Joining tree method (JC69) as implemented in the Geneious Pro program version 5.6.5.[35]

Genome Data

The *msc* gene cluster sequence was deposited in the GenBank with the accession number KF657738 for *S. cellulosum* So ce38 and accession no KF657739 for *Jahnella* sp. MSr9139.

Results and Discussion

Production of Microsclerodermins by Terrestrial Myxobacteria

In the course of our screening for bioactive natural products from myxobacteria, we observed antifungal activity in extracts from strain MSr9139, a newly isolated *Jahnella* species. Subsequent HPLC-based purification led to several fractions showing antifungal activity which contained compounds featuring an isotopic pattern typical for chlorination in MS analysis. Two compounds from these fractions could be assigned by their exact mass, fragmentation pattern, and retention time as pedein A (925.3493 *m/z*, $[M+H]^+$) and pedein B (891.3883 *m/z*, $[M+H]^+$), antifungal metabolites known from the myxobacterium *Chondromyces pediculatus* Cm p3.[25] Full structure elucidation was carried out for a compound with 869.3231 *m/z* obtained from another

bioactive fraction, and the data unambiguously revealed this candidate as the known marine natural product microsclerodermin D (Table 1 and Figure S1). In addition to this, analysis of the MSr9139 extract led to the isolation and structure elucidation of the new derivative microsclerodermin L, differing from microsclerodermin D by an additional methoxy group, which is also reported for the pedein structure (Table 1 and Figure S3). Notably, the microsclerodermins are the first family of compounds found in the unexplored genus *Jahnella*, a member of the notable secondary metabolite producer myxobacterial family *Polyangiaceae*.[36]

Almost simultaneously, extracts from the myxobacterial strain *Sorangium cellulosum* So ce38 underwent biological profiling and HPLC fractionation, highlighting antifungal activity in the same chromatographic region as previously found from the MSr9139 extract. HPLC purification could narrow down the putatively active compounds to a candidate with 887.3934 *m/z*, and subsequent NMR analysis identified a peptide featuring the pyrrolidone moiety also known from microsclerodermins. NMR data revealed the presence of a new non-chlorinated derivative, microsclerodermin M (Table 1 and Figures 2, S4). It shares the typical cyclic core structure with other microsclerodermins but features an unbranched side chain with three double bonds in conjugation to a phenyl moiety. Like the known microsclerodermins, the newly identified derivatives show potent activity against *Candida albicans* (microsclerodermin M, MIC 0.16 µg/mL; microsclerodermin L, MIC 18 µg/mL, microsclerodermin D, MIC 6.8 µg/mL). The stereochemistry of the isolated microsclerodermins was identified by acetonide formation and chemical degradation experiments followed by advanced Marfey analysis (see supporting information). It is identical to that reported for the microsclerodermins C – I and pedeins.[25,27,28]

Having discovered that myxobacteria from three different genera *Jahnella*, *Sorangium*, and *Chondromyces* are able to produce microsclerodermin congeners including even the exact same structure as previously described from two species of lithistid sponges (microsclerodermin D) was surprising for two reasons: examples of myxobacteria producing an identical scaffold also known from a phylogenetically distant organism are to date exceedingly rare (even when counting among the bacterial kingdom), and according to previous studies the secondary metabolite profiles from strains belonging to different myxobacterial genera usually exhibit little overlap.[6] In order to shed light on the occurrence of microsclerodermins within the myxobacteria, we conducted a search across high-resolution LC-MS data sets measured from almost 800 extracts, thus covering a sufficiently representative sample including most known myxobacterial taxa (Figure S20). On the basis of the evaluation of exact masses, isotope patterns and retention times we could identify a panel of 15 strains from the suborder *Sorangineae* (no single producer was found within the *Cystobacterineae*) as producers of microsclerodermins. Interestingly, our comprehensive LC-MS survey of myxobacterial secondary metabolomes

revealed that producers of microsclerodermins form two mutually exclusive groups: one group comprises 11 *Sorangium* species producing solely the new microsclerodermin M, while the second group includes two strains of *Jahnella* sp. and two *Chondromyces* sp. that produce a variety of different derivatives: i.e. the "marine" microsclerodermin D in addition to the new microsclerodermin L and pedeins A/B (Figures S21 and S22). The various microsclerodermins differ in side chain, tryptophan modification and oxidation state at the pyrrolidone ring, whereas the peptidic core structure is always identical (Figure 1).

The fact that all microsclerodermins, irrespective of their origin, exhibit an identical macrocycle and in most cases even the same stereochemistry supports the idea of a shared biosynthetic origin or even a shared evolutionary ancestor. Moreover, the finding of a group of terrestrial myxobacteria producing exactly the same compound as found in lithistid sponges[27] (microsclerodermin D) fuels speculation about the actual biosynthetic origin of marine microsclerodermins. We herein propose that the marine microsclerodermins actually originate from a myxobacterium phylogenetically related to the *Sorangineae* suborder – possibly a yet uncultured species of the *Chondromyces*, *Jahnella,* or *Sorangium* taxa living in symbiosis with the sponges from which microsclerodermins were previously isolated.

In coincidence with this hypothesis, phylogenetic studies of sponge metagenomes recently identified δ-proteobacteria in the sponge holobiont.[31] Indeed, the phylogenetic tree presented in the work of Simister *et al*. lists a clade containing nine myxobacterial species of terrestrial origin, including *Sorangium cellulosum* and *Chondromyces pediculatus*.[31] These data underpin our assumption that an evolutionary link exists between microsclerodermin biosynthesis in terrestrial and marine producers. Notably, 13 out of 15 producers identified by our LC-MS metabolome survey belong to the genera *Sorangium* or *Chondromyces*. In addition, the new *Jahnella* sp. MSr9139 was isolated from a soil sample collected from the same Philippine island where the sponge *Microscleroderma* was initially found.[27] This intriguing finding suggests that myxobacteria are possibly flushed to the ocean and adapted to an association with a sponge. The diversity and density of microbial flora present in sponges appears to be a good niche for a predator and proteo-bacteriolytic myxobacterium;[7] thus it is expected that in the future myxobacteria will be isolated from this underexplored source.

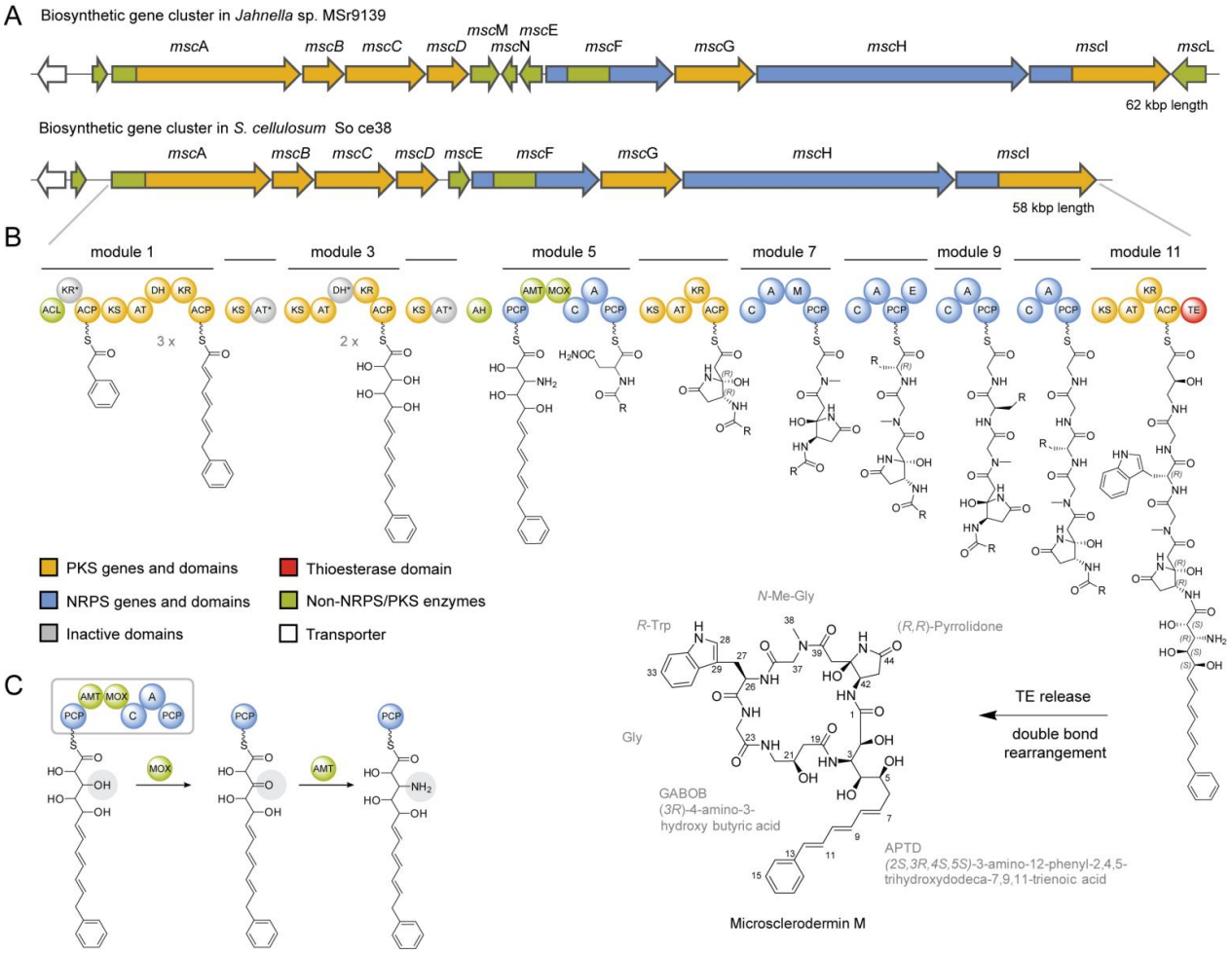


Figure 2: (A) Organization of the *msc* biosynthetic gene cluster in *Jahnella* sp. MSr9139 compared to *Sorangium cellulosum* So ce38. **(B)** Proposed biosynthetic route to microsclerodermin formation in So ce38. **(C)** Postulated biosynthetic steps leading to the amino group that is involved in macrolactam formation. A, adenylation domain; AMT, aminotransferase; ACP, acyl-carrier-protein domain; AT, acyltransferase; C, condensation domain; CoA-Lig, coenzyme A Ligase; DH, dehydratase; E, epimerase; KR, ketoreductase; KS, ketosynthase; MT, methyltransferase; MOX, monooxygenase; PCP, peptidyl-carrier-protein domain.

Access to myxobacterial producers holds a remarkable benefit, as these bacteria may be cultivated in large-scale fermentations, thereby allowing efficient production of the compounds of interest. Microsclerodermin M is produced at 12 mg/L in *S. cellulosum* So ce38 without optimization of growth conditions or genetic modification of the strain. Moreover, it allows us to investigate their biosynthesis, which has not been elucidated from any marine source to date. Thus, we set out to mine genome sequences of the new terrestrial producers for the presence of putative microsclerodermin biosynthetic pathways, using a retrobiosynthetic analysis as the starting point. The genome sequence of the strain *Sorangium cellulosum* So ce38, producer of the new microsclerodermin M, was already available from a previous study.[37] The newly isolated *Jahnella* sp. MSr9139 was selected for additional genome sequencing, as it produces the new

microsclerodermin L in addition to known pedeins A and B and the "marine" microsclerodermin D.

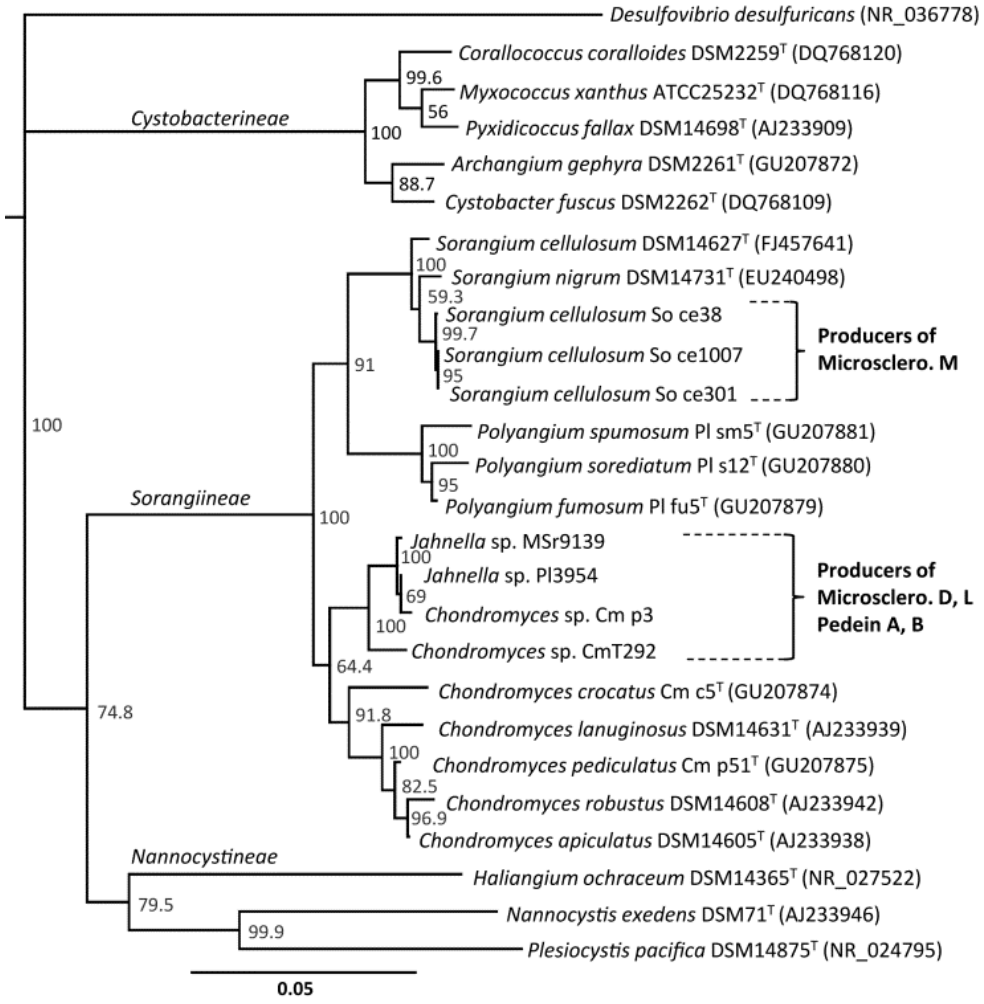


Figure 3: Neighbor-joining tree of myxobacteria inferred from 16S rRNA gene sequences showing the clades of microsclerodermin producing strains in suborder *Sorangiineae*. The numbers at branch points indicate percentage bootstrap support based on 1000 resamplings. GenBank accession numbers are indicated in parentheses. Bar = 0.05 substitutions per nucleotide position.

Microsclerodermin Biosynthetic Machinery

All microsclerodermins share the same cyclic peptide core but feature different lipophilic side chains and modifications of amino acid residues. On the basis of retrobiosynthetic considerations, the biosynthetic machinery for microsclerodermin formation was expected to consist of a multimodular PKS/NRPS system accompanied by a set of enzymes involved in side chain biosynthesis and post assembly line modification. The core PKS/NRPS modules should be conserved between producers, while enzymes involved in side chain biosynthesis and additional tailoring enzymes – responsible for modifications such as halogenation or oxidation of the

pyrrolidone ring – should occur differentially between the two producer groups, as a consequence of evolutionary diversification of the microsclerodermin pathway.

Using *S. cellulosum* So ce38 and *Jahnella* sp. MSr9139 as representative strains from both microsclerodermin producer groups, we sought to identify microsclerodermin biosynthetic pathways in their genomes and subsequently elaborate on the molecular basis for the observed structural variations. The genome sequences of both strains were searched *in silico* for secondary metabolite gene clusters using the antiSMASH analysis pipeline.[38] The assignment of a matching candidate cluster to microsclerodermin biosynthesis was verified in So ce38 via targeted gene disruption by single crossover integration using biparental conjugation (Figure S24). Sequence comparison on the protein and nucleotide level revealed high similarity between gene clusters from both strains (Table 2), and comparison of operons permitted the tentative assignment of cluster boundaries. The microsclerodermin cluster spans a region of 58 kbp (74.7 % GC) in So ce38 and 62 kbp (72.4 % GC) in MSr9139, respectively. In both strains, genes encoding a major facilitator superfamily transporter (*mscK*) followed by a type II thioesterase (*mscJ*) are located upstream to *msc*A. The core biosynthetic assembly line covers five NRPS modules and three PKS modules encoded on genes *mscA* to *mscI*. An additional halogenase is encoded by *mscL* near the downstream boundary of the cluster in MSr9139 (Figure 2 A).

Microsclerodermin biosynthesis is initiated at the side chain to build up a phenyl group in conjugation to a double bond. An activated starter unit such as benzoyl-CoA or *trans*-cinnamoyl-CoA is usually recruited by the enzyme in such a case. However, a retrobiosynthetic proposal tells us that the observed double bond order of the side chain is likely different during biosynthesis (Figure 2, B). The biosynthetic logic requests the incorporation of C2 units, which is only possible if the double bonds are rearranged (Figure S23). A rearrangement of double bonds has already been reported for other natural products like bacillaen, rhizoxin, corallopyronin and ansamitocin where isomerization is likely catalyzed by a dehydratase domain.[39–42] The reason for isomerization of the microsclerodermin side chain remains elusive; however, it is supported by an energetic benefit of a conjugated π-system. On the basis of this hypothesis, the only suitable starter unit is phenylacetyl-CoA, which has already been reported for other natural product biosynthesis.[43] The incorporation of a phenylacetate starter unit was indeed verified by feeding experiments using isotope-labeled precursors. Feeding ring-labeled $^{13}C_6$-L-phenylalanine resulted in a mass increase of 6 Da, whereas the fully labeled $^{15}N,^{13}C_9$-L-phenylalanine led to a mass shift of 8 Da, indicating the incorporation of two side chain carbons (Figure S27). Feeding d5-benzoic acid or d7-*trans*-cinnamic acid resulted in no mass increase (Figure S25). We can conclude that the α and β carbon atoms of phenylalanine -

but not the carboxyl carbon - is incorporated into microsclerodermin. Elongation of the phenylacetate unit is catalyzed by modules MscA and MscC using three times malonate and two times 3-hydroxymalonate as extender units in an iterative manner. Modules 2 (MscB) and 4 (MscD) do only exhibit a combination of a functional KS domain attached to an inactive AT domain as identified by consensus sequence analysis, likely a relic of a former PKS complex.

Table 1: Proteins involved in microsclerodermin biosynthesis as identified in two myxobacterial strains.

	Sorangium cellulosum So ce38		*Jahnella* sp. MSr9139		
protein	length [aa]	domains and position in sequence	length [aa]	domains and position in sequence	identity [%]
MscA	3275	CoA-Lig (264-701), KR*(1034-1197), ACP (1348-1411), KS (1441-1828), AT (1976-2286), DH (2342-2505), KR' (2870-3047), ACP' (3149-3214)	3535	CoA-Lig (215-649), KR*(1032-1126), MT (1229-1509), ACP (1613-1676), KS (1712-2147), AT (2244-2555), DH (2610-2774), KR*' (3132-3309), ACP' (3411-3476)	65.4
MscB	870	KS (27-451), AT* (548-772)	878	KS (39-464), AT* (561-792)	73.4
MscC	1551	KS (36-460), AT (557-859), DH* (956-1076), KR (1158-1336), ACP (1436-1505)	1549	KS (39-464), AT (561-865), DH* (956-1076), KR (1157-1335), ACP (1436-1499)	75.0
MscD	900	KS (36-461), AT* (565-678)	848	KS (36-461), AT* (564-675)	72.0
MscE	446	Putative amidohydrolase	386	Putative amidohydrolase	82.9
MscF	2273	PCP (27-98), AMT (329-660), MOX (828-1127), C (1185-1529), A (1673-2082), PCP' (2169-2237)	2189	PCP (4-75), AMT (280-614), MOX (758-1057), C (1105-1397), A (1594-2000), PCP' (2088-2149)	76.5
MscG	1548	KS (14-439), AT (534-828), KR (1156-1330), ACP (1441-1509)	1511	KS (14-439), AT (531-850), KR (1136-1312), ACP (1404-1472)	72.1
MscH	4141	C (76-377), A (564-966), MT (1037-1256), PCP (1469-1531), C' (1554-1850), A' (2037-2426), PCP' (2515-2574), E (2591-2905), C'' (3074-3374), A'' (3558-3967), PCP'' (4054-4121)	4106	C (48-346), A (534-936), MT (1007-1225), PCP (1442-1505), C' (1526-1827), A' (2013-2405), PCP' (2492-2551), E (2568-2872), C'' (3043-3343), A'' (3527-3932), PCP'' (4019-4083)	78.8
MscI	2904	C (48-346), A (533-936), PCP (1023-1087), KS (1111-1535), AT (1638-1936), KR (2266-2465), ACP (2549-2612), TE (2634-2888)	2945	C (77-375), A (563-965), PCP (1053-1116), KS (1150-1573), AT (1676-1970), KR (2309-2509), ACP (2597-2660), TE (2683-2945)	77.8
MscJ	257	thioesterase type II	263	thioesterase type II	43.5
MscK	415	major facilitator superfamily (MFS) transporter	450	major facilitator superfamily (MFS) transporter	24.5
MscL	-	-	535	Tryptophan halogenase	-
MscM	-	-	438	Fe(II)/α-ketoglutarate dependent oxygenase	-
MscN	-	-	277	methyltransferase	-

A, adenylation domain; AMT, aminotransferase; ACP, acyl-carrier-protein domain; AT, acyltransferase; C, condensation domain; CoA-Lig, coenzyme A Ligase; DH, dehydratase; E, epimerase; KR, ketoreductase; KS, ketosynthase; MT, methyltransferase; MOX, monooxygenase; PCP, peptidyl-carrier-protein domain. * inactive domain

The PKS-derived unit is forwarded to the first PCP domain of module MscF. This module harbors two additional domains of rather uncommon type showing high homology to the amino transferase family (AMT) and to the monooxygenase family, both located downstream to the PCP domain. A biosynthetic proposal to account for this domain order is based on oxidation of the β-hydroxyl group of the bound intermediate to the respective β-keto functionality followed by

conversion to a β-amino moiety that undergoes macrocyclization (Figure 2, C). The use of an aminotransferase is known from other natural product biosynthetic pathways, however, not in combination with the initial oxidation step.[44]

Thereafter, biosynthesis continues with a set of NRPS- and PKS-based reaction cycles. Analysis of the A domain specificities *in silico* is consistent with the amino acids incorporated.[45] We propose the uncommon pyrrolidone moiety is built up by asparagine cyclization. To the best of our knowledge, such an asparagine-derived pyrrolidone system is only found in microsclerodermins and koshikamides, a natural product that was isolated from a *Theonella* species.[46] Indeed, the A domain of MscF is specific for asparagine activation and we did a feeding experiment with fully labeled $^{15}N_2$-$^{13}C_4$-L-asparagine to prove this biosynthetic step. We observed a mass increase of 6 Da according to the incorporation of all carbon and nitrogen atoms of asparagine into the compound (Figures S26, S28). On the basis of this result, a plausible biosynthetic hypothesis requires the nucleophilic attack of the asparagine side chain to the backbone carbonyl atom. A suitable mechanism is known from the intein-mediated peptide cleavage, where intein initiates an intra-molecular asparagine cyclization, notwithstanding the poor reactivity of the side chain's amide.[47] In microsclerodermin biosynthesis, this reaction likely is accompanied by an inversion of the stereochemistry at the α-carbon of the former (*S)*-asparagine. The relative configuration of the pyrrolidone ring was identified by NOE correlations, whereas the absolute (*R,R*)-configuration is derived from degradation experiments (see supporting information). The stereochemistry at this position is thereby identical with that of microsclerodermins C – I. The protein MscE is most likely responsible for the cyclization step, as it is found in both microsclerodermin clusters and shows similarity to the amidohydrolase class, a fairly promiscuous enzyme family able to act on a variety of substrates. However, the exact mechanism involved in this biosynthetic step remains elusive at present.

The forthcoming NRPS modules correspond to the observed structure of microsclerodermin in terms of domain order and predicted substrate specificity (Figure 2, B). For both new microsclerodermins an *R*-configured tryptophan was identified by means of the advanced Marfey method, which is in agreement with the epimerization domain found in module 8. The (3*R*)-configuration of the γ-amino butyric acid (GABA) subunit was identified by the same technique (see Figures S13, S14). Both stereogenic centers have the same configuration as identified in all microsclerodermins so far.

Genetic Basis for the Structural Diversity of Microsclerodermins

The derivatives found in *Jahnella* sp. MSr9139 feature side chains with either one or two double bonds while the side chain in So ce38 comprises strictly three double bonds**.** As the number of PKS modules encoded in the gene cluster does not match the number of required elongation cycles, an iterative function of the type I PKS subunits MscA and MscC as described for the stigmatellin megasynthase may explain this finding.[48] The KS domains of each module are highly identical for both strains and do not comprise any of the postulated sequence-based identifiers of iterative KS domains.[49] Nevertheless, MscB is grouping with iterative KS domains in a phylogenetic analysis (Figure S11). Comparing the entire module MscA of both clusters revealed the insertion of an additional methyl transferase-like domain into the first part of the protein in MSr9139. This domain is likely inactive on the basis of *in silico* analysis as judged by the presence of a corrupted SAM-binding motif (Table S13).[50] Currently, we cannot rule out the possibility that the presence of this additional methyl transferase may influence the iteration process within MscA. In addition to this difference, there is no obvious reason why biosynthesis in So ce38 results in a triene whereas MSr9139 is less strict in iteration. As another hint for a shared origin of the biosynthetic cluster, some of the sponge-derived derivatives exhibit a methyl-branched side chain which could be attributed to this methyl transferase being active in some of the marine producers. Eventually, the variable side chains of the microsclerodermin family are in agreement with the alternating PKS functionality. Halogenation of tryptophan as well as oxidation of the pyrrolidone ring is catalyzed by tailoring enzymes. The halogenase MscL is located downstream to the cluster in MSr9139 and is responsible for chlorination of the tryptophan. It shows 32 % identity on a protein level to a tryptophan halogenase from a *Streptomyces* species (PDB entry 2WET_A). There is no analogue of MscL found in So ce38, which is in agreement with the absence of chlorinated products in this strain. Supplementing KBr or NaBr to the MSr9139 cultivation broth led to the production of brominated microsclerodermins on the basis of LC-MS analysis. Another difference is the inter-region between the main PKS and NRPS parts. In MSr9139 two additional proteins are found in this region. MscN is a member of the SAM-dependent methyl transferase family, and MscM shows homology to Fe(II)/α-ketoglutarate-dependent dioxygenases. On the basis of the structures produced by MSr9139, we conclude that MscM and MscN are responsible for oxidation and methylation of the pyrrolidone ring, respectively. Modifications at the tryptophan as known from some marine-derived microsclerodermins were not observed in this study. Such modifications are attributed to promiscuous acting enzymes that could be related to the producer strain or even to enzymes related to some sponge symbiont.

Conclusion

The discovery of microsclerodermins/pedeins from several myxobacteria represents one of the few findings of identical compounds from marine organisms and terrestrial bacteria reported to date. This study thus strengthens the notion that certain natural products, which have been isolated from marine sources such as sponges or other invertebrates, actually originate from associated microbes. Notably, the identification of microsclerodermin-producing myxobacteria provides meaningful hints for future attempts to isolate the symbiotic microbe. This knowledge is considered particularly helpful because isolation success in many cases critically depends on methods well adapted to the requirements of the genus targeted for isolation, especially when aimed at the rather challenging isolation of slow-growing myxobacteria. Availability of an alternative microbial producer as a sustainable source is an advantage for realizing the potential of a marine natural product for therapeutic applications. Moreover, the myxobacterial producers come along with additional chemical diversity and are amenable to genetic manipulation, as demonstrated in this study. Finally, the identification of two slightly different microsclerodermin biosynthetic gene clusters from two myxobacteria allowed us to establish a conclusive model for microsclerodermin biosynthesis and provided insights into the molecular basis for structural diversity within this compound family. A detailed understanding of microsclerodermin biosynthesis is an important prerequisite for any future efforts toward engineering the pathway for yield improvement or for the production of new derivatives: whether in the native producer, by heterologous expression, or by using synthetic biology approaches.

Author Contributions

All authors have given approval to the final version of the manuscript.

Acknowledgement

Research in R.M.'s laboratory was funded by the Bundes-ministerium für Bildung and Forschung and the Deutsche Forschungsgemeinschaft. We thank Daniel Krug for advice and discussion, as well as for editing the manuscript. We also thank Jennifer Herrmann for performing the bioactivity assays, Lena Etzbach and Alberto Plaza for help with the NMR data, and Manuel Klos for ozonolysis reactions.

Refrences

(1) Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* 2012, *75*, 311–35.
(2) Wenzel, S. C.; Müller, R. *Mol. Biosyst.* 2009, *5*, 567–74.
(3) Weissman, K. J.; Müller, R. *Bioorg. Med. Chem.* 2009, *17*, 2121–36.
(4) Lane, A. L.; Moore, B. S. *Nat. Prod. Rep.* 2011, *28*, 411–28.

(5) Winder, P. L.; Pomponi, S. A.; Wright, A. E. *Mar. Drugs* 2011, *9*, 2643–82.
(6) Wink, J.; Müller, R. *Int. J. Med. Microbiol.* 2013, in press.
(7) Hentschel, U.; Piel, J.; Degnan, S. M.; Taylor, M. W. *Nat. Rev. Microbiol.* 2012, *10*, 641–54.
(8) Taylor, M. W.; Radax, R.; Steger, D.; Wagner, M. *Microbiol. Mol. Biol. Rev.* 2007, *71*, 295–347.
(9) Bewley, C. A.; Faulkner, D. J. *Angew. Chem. Int. Ed. Engl.* 1998, *37*, 2162–2178.
(10) Schmidt, E. W.; Obraztsova, A. Y.; Davidson, S. K.; Faulkner, D. J.; Haygood, M. G. *Mar. Biol.* 2000, *136*, 969–977.
(11) Unson, M. D.; Faulkner, D. J. *Experientia* 1993, *49*, 349–353.
(12) Elshahawi, S. I.; Trindade-Silva, A. E.; Hanora, A.; Han, A. W.; Flores, M. S.; Vizzoni, V.; Schrago, C. G.; Soares, C. A.; Concepcion, G. P.; Distel, D. L.; Schmidt, E. W.; Haygood, M. G. *Proc. Natl. Acad. Sci. U.S.A.* 2013, *110*, E295–304.
(13) Lin, Z.; Torres, J. P.; Ammon, M. A.; Marett, L.; Teichert, R. W.; Reilly, C. A.; Kwan, J. C.; Hughen, R. W.; Flores, M.; Tianero, M. D.; Peraud, O.; Cox, J. E.; Light, A. R.; Villaraza, A. J. L.; Haygood, M. G.; Concepcion, G. P.; Olivera, B. M.; Schmidt, E. W. *Chem. Biol.* 2013, *20*, 73–81.
(14) Kwan, J. C.; Donia, M. S.; Han, A. W.; Hirose, E.; Haygood, M. G.; Schmidt, E. W. *Proc. Natl. Acad. Sci. U.S.A.* 2012, *109*, 20655–60.
(15) Wenzel, S. C.; Müller, R. In *Comprehensive Natural Products II, Vol. 2: Natural Products Structural Diversity-II Secondary Metabolites: Sources, Structures and Chemical Biology*; Moore, B. S.; Crews, P., Eds.; Elsevier, 2010; pp 189-222.
(16) Crews, P.; Manes, L. V; Boehler, M. *Tetrahedron Lett.* 1986, *27*, 2797–2800.
(17) Zabriskie, T. M.; Klocke, J. A.; Ireland, C. M.; Marcus, A. H.; Molinski, T. F.; Faulkner, D. J.; Xu, C.; Clardy, J. *J. Am. Chem. Soc.* 1986, *108*, 3123–3124.
(18) Kunze, B.; Jansen, R.; Sasse, F.; Höfle, G.; Reichenbach, H. *J. Antibiot.* 1995, *48*, 1262–6.
(19) Irschik, H.; Trowitzsch-Kienast, W.; Gerth, K.; Höfle, G.; Reichenbach, H. *J. Antibiot.* 1988, *41*, 993–8.
(20) Frincke, J. M.; Faulkner, D. J. *J. Am. Chem. Soc.* 1982, *104*, 265–269.
(21) Erickson, K. L.; Beutler, J. A.; Cardellina, J. H.; Boyd, M. R. *J. Org. Chem.* 1997, *62*, 8188–8192.
(22) Kunze, B.; Jansen, R.; Sasse, F.; Höfle, G.; Reichenbach, H. *J. Antibiot.* 1998, *51*, 1075–80.
(23) Hoffmann, H.; Haag-Richter, S.; Kurz, M.; Tietgen, H. Bengamide derivatives, method for the production thereof and use thereof for the treatment of cancer. WO2005044803 A1, 2005.
(24) Johnson, T. A.; Sohn, J.; Vaske, Y. M.; White, K. N.; Cohen, T. L.; Vervoort, H. C.; Tenney, K.; Valeriote, F. A.; Bjeldanes, L. F.; Crews, P. *Bioorg. Med. Chem.* 2012, *20*, 4348–55.
(25) Kunze, B.; Böhlendorf, B.; Reichenbach, H.; Höfle, G. *J. Antibiot.* 2008, *61*, 18–26.
(26) Bewley, C. A.; Debitus, C.; Faulkner, D. J. *J. Am. Chem. Soc.* 1994, *116*, 7631–7636.
(27) Schmidt, E. W.; John Faulkner, D. *Tetrahedron* 1998, *54*, 3043–3056.
(28) Qureshi, A.; Colin, P. L.; Faulkner, D. J. *Tetrahedron* 2000, *56*, 3679–3685.
(29) Zhang, X.; Jacob, M. R.; Ranga Rao, R.; Wang, Y. H.; Agarwal, A. K.; Newman, D. J.; Khan, I. A.; Clark, A. M.; Li, X. C. *Res. Rep. Med. Chem.* 2012, *2012*, 7–14.
(30) Zhang, X.; Jacob, M. R.; Ranga Rao, R.; Wang, Y. H.; Agarwal, A. K.; Newman, D. J.; Khan, I. A.; Clark, A. M.; Li, X. C. *Res. Rep. Med. Chem.* 2013, 9.
(31) Simister, R. L.; Deines, P.; Botté, E. S.; Webster, N. S.; Taylor, M. W. *Environ. Microbiol.* 2012, *14*, 517–24.
(32) Garcia, R. O.; Krug, D.; Müller, R. *Meth. Enzymol.* 2009, *458*, 59–91.
(33) Sambrook, J.; Russell, D. W. *Molecular cloning: A laboratory manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2001.
(34) Kopp, M.; Irschik, H.; Gross, F.; Perlova, O.; Sandmann, A.; Gerth, K.; Müller, R. *J. Biotechnol.* 2004, *107*, 29–40.
(35) Garcia, R.; Gerth, K.; Stadler, M.; Dogma, I. J.; Müller, R. *Mol. Phylogenet. Evol.* 2010, *57*, 878–87.
(36) Garcia, R.; Müller, R. In *The Prokaryotes: Deltaproteobacteria and Epsilonproteobacteria*; Rosenberg, E.; DeLong, E. F.; Lory, S.; Stackebrandt, E.; Thompson, F., Eds.; Springer: Heidelberg, 2014, in press.
(37) Jahns, C.; Hoffmann, T.; Müller, S.; Gerth, K.; Washausen, P.; Höfle, G.; Reichenbach, H.; Kalesse, M.; Müller, R. *Angew. Chem. Int. Ed. Engl.* 2012, *51*, 5239–43.
(38) Medema, M. H.; Blin, K.; Cimermancic, P.; de Jager, V.; Zakrzewski, P.; Fischbach, M. A.; Weber, T.; Takano, E.; Breitling, R. *Nucleic Acids Res.* 2011, *39*, W339–46.
(39) Butcher, R. A.; Schroeder, F. C.; Fischbach, M. A.; Straight, P. D.; Kolter, R.; Walsh, C. T.; Clardy, J. *Proc. Natl. Acad. Sci. U.S.A.* 2007, *104*, 1506–9.
(40) Kusebauch, B.; Busch, B.; Scherlach, K.; Roth, M.; Hertweck, C. *Angew. Chem. Int. Ed. Engl.* 2010, *49*, 1460–4.
(41) Erol, O.; Schäberle, T. F.; Schmitz, A.; Rachid, S.; Gurgui, C.; El Omari, M.; Lohr, F.; Kehraus, S.; Piel, J.; Müller, R.; König, G. M. *ChemBioChem* 2010, *11*, 1253–65.
(42) Taft, F.; Brünjes, M.; Knobloch, T.; Floss, H. G.; Kirschning, A. *J. Am. Chem. Soc.* 2009, *131*, 3812–3.
(43) Magarvey, N. A.; Beck, Z. Q.; Golakoti, T.; Ding, Y.; Huber, U.; Hemscheidt, T. K.; Abelson, D.; Moore, R. E.; Sherman, D. H. *ACS Chem. Biol.* 2006, *1*, 766–79.
(44) Tillett, D.; Dittmann, E.; Erhard, M.; von Döhren, H.; Börner, T.; Neilan, B. A. *Chem. Biol.* 2000, *7*, 753–64.
(45) Röttig, M.; Medema, M. H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. *Nucleic Acids Res.* 2011, *39*, W362–7.
(46) Plaza, A.; Bifulco, G.; Masullo, M.; Lloyd, J. R.; Keffer, J. L.; Colin, P. L.; Hooper, J. N. A.; Bell, L. J.; Bewley, C. A. *J. Org. Chem.* 2010, *75*, 4344–55.
(47) Mujika, J. I.; Lopez, X.; Mulholland, A. J. *Org. Biomol. Chem.* 2012, *10*, 1207–18.
(48) Gaitatzis, N.; Silakowski, B.; Kunze, B.; Nordsiek, G.; Blöcker, H.; Höfle, G.; Müller, R. *J. Biol. Chem.* 2002, *277*, 13082–90.
(49) Yadav, G.; Gokhale, R. S.; Mohanty, D. *PLoS Comput. Biol.* 2009, *5*, e1000351.
(50) Kozbial, P. Z.; Mushegian, A. R. *BMC Struct. Biol.* 2005, *5*, 19.

Supporting Information

Structure elucidation

NMR data of Microsclerodermin D

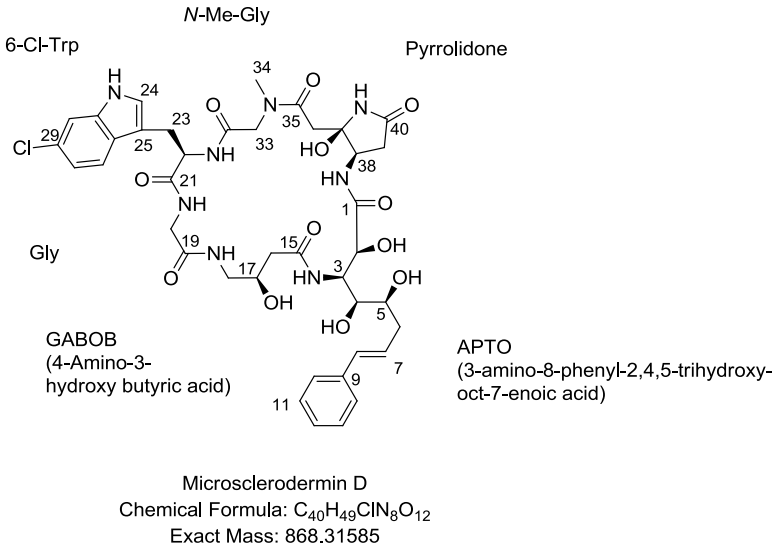


Figure S1: Structure and atom numbering of microsclerodermin D. Compound isolated from marine sponges of the genera *Microscleroderma* sp. and *Theonella* sp. and the myxobacterium *Jahnella* sp. MSr9139.

Table S1: NMR chemical shifts of microsclerodermin D.

Amino acid	Assignment	δ_C [a]	δ_H (mult., *J* in Hz) [b]	HMBC	ROESY
APTO	1	172.3			
	2	69.2	4.39 (d, 3.2)		
	3	53	4.13 (d, 11.3)		
	4	69.5	3.32 (m)		
	5	68.8	3.57 (m)		
	6	36.3	2.35 (m)		
	7	128.0	6.24 (m)		
	8	130.5	6.40 (d, 15.8)		
	9	137.2			
	10, 14	125.4	7.35 (d, 7.3)		
	11, 13	128.2	7.29 (t, 7.6)		
	12	126.5	7.19 (t, 7.3)		
	OH-2		6.58 (brs)		
	OH-4		4.17 (m)		
	OH-5		4.36 (d, 4.5)		
	NH-3		7.47 (d, 6.0)		
GABOB	15	172.3			
	16	40.7	2.43 (d, 14.0)		
			2.15 (d, 14.0)		
	17	66.8	3.72 (m)		
	18	44.7	3.39 (m)		
			2.61 (m)		
	OH-17		4.88 (d, 4.8)		

	NH-18		7.50 (m)
Gly	19	168.5	
	20	42.3	3.75 (d, 7.0)
			3.34 (d, 7.0)
	NH-20		8.54 (t, 6.0)
6-Cl-Trp	21	171.5	
	22	55.1	4.17 (m)
	23	25.8	3.10 (dd, 5.7, 14.7)
			2.98 (dd, 5.7, 14.7)
	24	124.6	7.26 (d, 1.8)
	25	109.5	
	26	125.6	
	27	119.4	7.52 (d, 8.3)
	28	118.4	7.00 (dd, 1.5, 8.3)
	29	c	
	30	110.7	7.37 (s)
	31	136.1	
	NH-22		8.64 (d, 4.3)
	NH-24		11.04 (brs)
N-Me-Gly	32	169.9	
	33	49.5	4.08 (d, 16.3)
			3.84 (d, 16.3)
	34	36.2	2.93 (s)
Pyrrolidone	35	170.2	
	36	38.6	2.84 (d, 17.1)
			2.69 (d, 17.1)
	37	85.2	
	38	50.3	4.47 (m)
	39	35.0	2.27 (m)
	40	172.3	
	NH-37		7.96 (brs)
	NH-38		7.56 (m)
	OH-37		c

[a] Recorded at 175 MHz, referenced to residual solvent DMSO-d_6 at 39.51 ppm.

[b] Recorded at 700 MHz, referenced to residual solvent DMSO-d_6 at 2.50 ppm.

[c] Not observed. [ov] overlapping signals.

NMR data of Dehydromicrosclerodermin D - acetonide

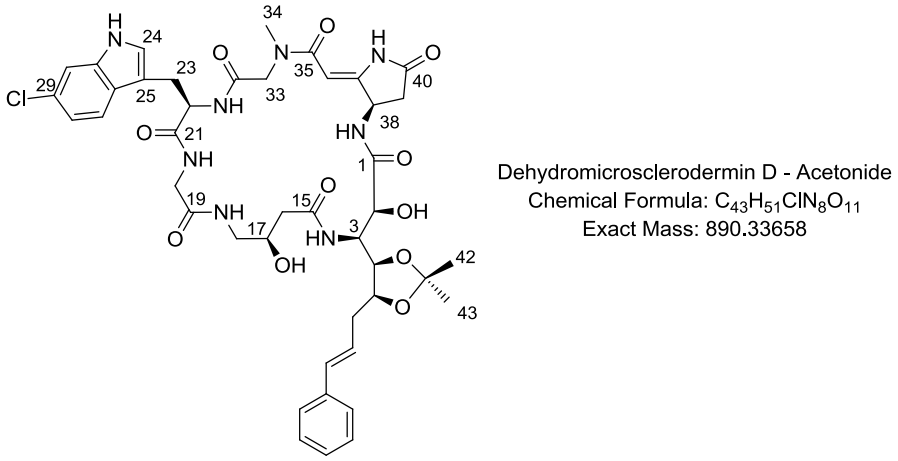


Figure S2: Structure and atom numbering of the acetonide of dehydromicrosclerodermin D. Derivative is based on microsclerodermin D as isolated from the myxobacterium *Jahnella* sp. MSr9139.

Table S2: NMR chemical shifts of dehydromicrosclerodermin D - acetonide.

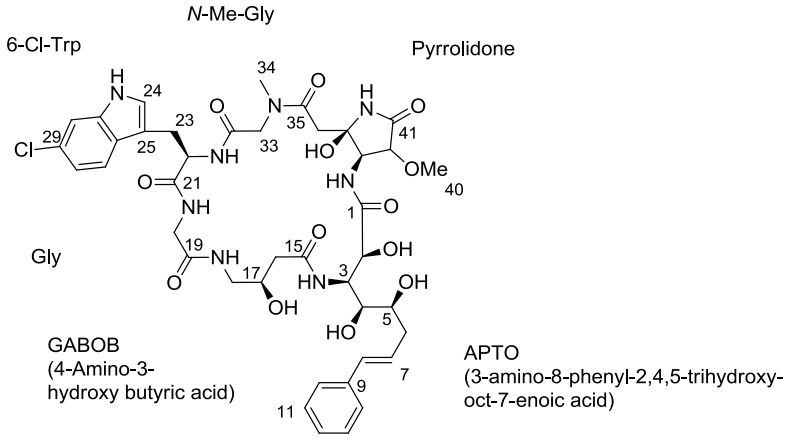
Amino acid	Assignment	δ_C [a]	δ_H (mult., *J* in Hz) [b]
APTO	1	172.3	
	2	70.1	4.19 (d, 4.2)
	3	55.4	4.11 (d, 10.0)
	4	77.3	3.74 (ov)
	5	78.7	4.02 (m)
	6	37.1	2.32 (m) 2.55 (m)
	7	126.5	6.24 (m)
	8	131.3	6.43 (d, 16.0)
	9	c	
	10, 14	125.4	7.38 (m)
	11, 13	128.2	7.30 (t, 7.6)
	12	126.5	7.20 (t, 7.2)
	OH-2		6.23 (ov)
	NH-3		7.09 (ov)
GABOB	15	c	
	16	41.1	2.22 (d, 14.8) 2.01 (m)
	17	65.6	3.89 (m)
	18	44.7	3.23 (m) 2.82 (m)
	OH-17		5.10 (d, 4.5)
	NH-18		7.35 (t, 6.5)
Gly	19	c	
	20	42.6	3.58 (m))
	NH-20		8.38 (t, 6.5)
6-Cl-Trp	21	c	
	22	55.2	4.11 (m)
	23	25.4	3.19 (ov) 3.01 (dd, 5.1, 10.0)
	24	124.6	7.26 (d, 1.8)
	25	109.5	
	26	125.6	
	27	119.4	7.52 (d, 8.3)
	28	118.4	7.00 (dd, 1.5, 8.3)
	29	c	
	30	110.7	7.37 (s)
	31	136.1	
	NH-22		8.77 (d, 5.0)
	NH-24		11.04 (brs)
N-Me-Gly	32	c	
	33	50.5	3.44 (d, 15.7) 4.53 (d, 15.7)
	34	36.8	3.06 (s)
Pyrrolidone	35	c	
	36	88.0	5.25 (s)
	37	c	
	38	45.4	5.21 (m)
	39	33.7	2.44 (m) 2.74 (m)
	40	c	
	NH-37		c
	NH-38		8.36 (d, 8.8)

[a] Recorded at 175 MHz, referenced to residual solvent DMSO-d_6 at 39.51 ppm.

Recorded at 700 MHz, referenced to residual solvent DMSO-d_6 at 2.50 ppm.

[c] Not observed. [ov] overlapping signals.

NMR data of Microsclerodermin L



Microsclerodermin L
Chemical Formula: $C_{41}H_{51}ClN_8O_{13}$
Exact Mass: 898.32641 Da

Figure S3: Structure and atom numbering of microsclerodermin L. Compound isolated from the myxobacterium *Jahnella* sp. MSr9139.

Table S3: NMR chemical shifts of microsclerodermin L.

Amino acid	Assignment	δ_C [a]	δ_H (mult., *J* in Hz) [b]	HMBC	ROESY
APTO	1	172.9			
	2	69.2	4.43 (brs)	1	
	3	53	4.19 (d, 11.3)	4	4
	4	69.7	3.32 (m)		
	5	68.7	3.59 (q, 6.5)		
	6	36.6	2.38 (m)	4, 5, 7, 8	
			2.33 (m)		
	7	128	6.26 (m)	6, 9	
	8	130.5	6.40 (d, 15.8)	6, 9, 10, 14	6
	9	137.2			
	10, 14	125.6	7.36 (d, 7.0)	8, 12	7, 8
	11, 13	128.3	7.29 (t, 7.6)	9, 10, 14	
	12	126	7.19 (t, 7.4)	10, 11, 13, 14	
	OH-2		6.57 (brs)		
	OH-4		4.85 (d, 4.9)		
	OH-5		4.30 (brs)		
	NH-3		7.42 (brs)		16
GABOB	15	172.1			
	16	40.6	2.43 (d, 14.0)	15	
			2.12 (d, 14.0)		
	17	66.6	3.77 (m)	16, 18	OH-17
	18	44.8	3.33 (m)	17	
			2.65 (m)		
	OH-17		4.85 (brs)		
	NH-18		7.43 (brs)		20
Gly	19	168.5			
	20	42.5	3.72 (d, 7.0)	19, 21	
			3.36 (d, 7.0)		

	NH-20		8.51 (t, 6.0)		22
6-Cl-Trp	21	171.6			
	22	55.2	4.17 (m)		NH-20
	23	25.8	3.10 (dd, 5.7, 15.0)	21, 22, 24, 25, 26	
			2.98 (dd, 5.7, 15.0)		
	24	124.7	7.26 (d, 2.0)	25, 31, 32	
	25	109.3			
	26	125.5			
	27	119.4	7.52 (d, 8.5)	31	
	28	118.5	7.00 (dd, 1.8, 8.5)	30, 26	
	29	c			
	30	110.7	7.37 (s)	27	
	31	136.3			
	NH-22		8.68 (d, 4.3)		22, 23, 24, 33
	NH-24		11.03 (d, 2.0)	24, 25, 31, 26	24, 30
N-Me-Gly	32	172.2			
	33	49.6	4.28 (d, 16.4)	32, 34	
			3.69 (d, 16.4)		
	34	35.9	2.92 (s)	33, 35	
Pyrrolidone	35	170.4			
	36	38.6	2.83 (d, 16.9)	35, 37	38
			2.61 (d, 16.9)		
	37	82.3			
	38	55.8	4.44 (d, 9.0)	39	NH-38
	39	79.0	4.04 (d, 8.8)	40, 41	NH-38
	40	56.6	3.39 (s)	39	
	41	171.0			
	NH-37		8.18 (brs)	37, 38, 39	
	NH-38		7.61 (d, 9.6)		2, 39
	OH-37		c		

[a] Recorded at 175 MHz, referenced to residual solvent DMSO-d_6 at 39.51 ppm.

Recorded at 700 MHz, referenced to residual solvent DMSO-d_6 at 2.50 ppm.

[c] Not observed. [ov] overlapping signals.

NMR data of Microsclerodermin M

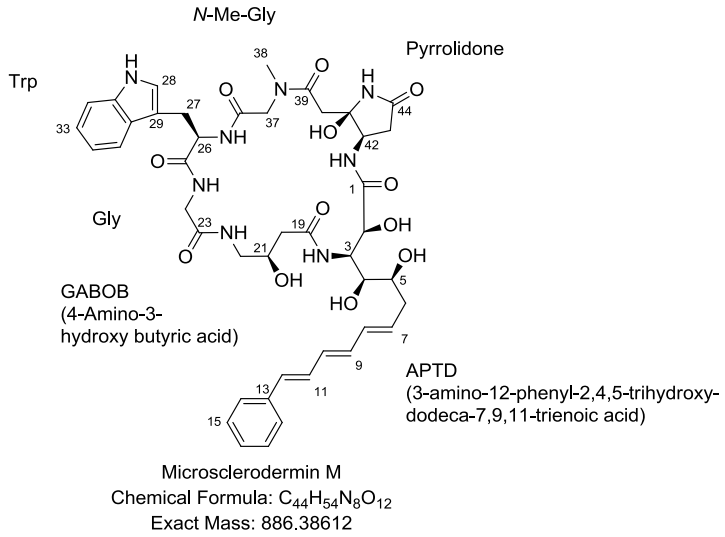


Figure S4: Structure and atom numbering of microsclerodermin M. Compound isolated from the myxobacterium *Sorangium cellulosum* So ce38.

Table S4: NMR chemical shifts of microsclerodermin M.

Amino acid	Assignment	δ_C [a]	δ_H (mult., *J* in Hz) [b]	HMBC	ROESY
APTD	1	172.5			
	2	69.3	4.38 (d, 5.4)	3	3, OH-2
	3	53.0	4.11 (ov)	4, 5, 19	OH-5, NH-3
	4	69.8	3.27 (ov)	2, 3	3, 7, OH-2, OH-3
	5	68.8	3.52 (ov)	4, 6	7, 8
	6	36.4	2.28 (ov)	4, 5, 7, 8	4, 5, 7, 8
	7	133.0	5.74 (dt, 15.0, 7.4)	6, 9	4, 5, 6, 9
	8	131.7	6.17 (dd, 10.0, 15.1)	9, 10	6, 11
	9	133.7	6.34 (ov)	7, 8, 10, 11	7, 11
	10	130.7	6.33 (ov)	8, 11, 12	12
	11	129.2	6.92 (dd, 9.6, 15.6)	9, 10, 13	9, 18
	12	131.3	6.56 (d, 15.6)	10, 14, 18, 13	10, 14
	13	137.0			
	14, 18	125.7	7.44 (d, 7.6)	12, 16	11, 12
	15, 17	128.4	7.31 (d, 8.3)	14, 18	14
	16	127.1	7.21 (d, 7.1)	14, 15, 17, 18	14, 18
	OH-2		6.25 (d, 5.6)	2, 3	2, 4, NH-42
	OH-4		4.34 (d, 4.7)	5, 6	3, 5
	OH-5		4.52 (d, 9.0)	3, 4, 5	3, 4
	NH-3		7.43 (d, 12.6)	2, 3	4, 5, 20, OH-21
GABOB	19	172.6			
	20	40.7	2.44 (d, 14.0)	19, 21, 22	
			2.15 (ov)		NH-3
	21	66.9	3.73 (ov)	21, 23	20, 22, OH-21, NH-22
	22	44.8	3.37 (ov)		21, OH-21
			2.64 (ov)		
	NH-22		7.49 (d, 6.7)		20, 21, 22
	OH-21		4.86 (d, 5.0)	20, 21, 22	20, 21, 22, NH-3
Gly	23	168.7			
	24	42.4	3.76 (d, 7.6)	23, 25	NH-24
			3.36 (d, 5.7)		
	NH-24		8.56 (t, 5.7)		24, 26, NH-22
Trp	25	171.7			
	26	55.2	4.19 (m)	25, 27, 29, 36	27, 30, NH-24
	27	26.2	3.11 (dd, 5.6, 14.6)	25, 26, 29	26, 28, 30, NH-26
			2.98 (dd, 5.6, 14.6)		
	28	123.5	7.21 (d, 6.6)	29, 34	26, 27, NH-26, NH-28
	29	109.6			
	30	126.5			
	31	118.0	7.52 (d, 7.7)	29, 32	26, 27, 32
	32	118.1	6.98 (t, 7.3)	33, 34	31, 34
	33	120.8	7.06 (d, 7.3)	30, 31, 35	31, 34
	34	111.1	7.32 (d, 7.9)	30, 32, 35	32
	35	135.9			
	NH-28		10.87 (brs)	28, 29, 35, 30	28, 34
	NH-26		8.65 (d, 4.2)	26, 27	26, 27, 28, 37
N-Me-Gly	36	169.9			
	37	49.6	4.10 (d, 15.4)	39, 36	38, NH-26
			3.82 (d, 15.4)		
	38	36.2	2.94 (s)	39	37
Pyrrolidone	39	170.4			
	40	38.7	2.85 (d, 17.0)	39, 41, 42	42, OH-41, NH-41
			2.70 (d, 17.0)		
	41	85.5			
	42	50.4	4.47 (ov)	40, 42, 43	40

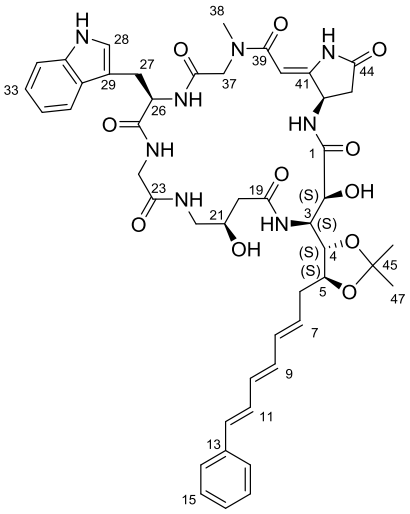
43	34.9	2.28 (ov)	41, 44	42, OH-41, NH-42
44	172.9			
OH-41		6.07 (s)	40, 42	40, 43, NH-41, NH-42
NH-41		7.97 (s)	41, 42, 43	40, 42, OH-41
NH-42		7.54 (ov)	42	42, 43, OH-41

[a] Recorded at 175 MHz, referenced to residual solvent DMSO-d_6 at 39.51 ppm.

Recorded at 700 MHz, referenced to residual solvent DMSO-d_6 at 2.50 ppm.

[c] Not observed. [ov] overlapping signals.

NMR data of Dehydromicrosclerodermin M - acetonide



Dehydromicrosclerodermin M
Acetonide
Chemical Formula: $C_{47}H_{56}N_8O_{11}$
Exact Mass: 908.40685

Figure S5: Structure and atom numbering of the acetonide of dehydromicrosclerodermin M. Derivative is based on microsclerodermin M as isolated from the myxobacterium *Sorangium cellulosum* So ce38.

Table S5: NMR chemical shifts of microsclerodermin M - acetonide.

Amino acid	Assignment	δ_C [a]	δ_H (mult., *J* in Hz) [b]
APTD	1	172.5	
	2	69.9	4.17 (ov)
	3	53.1	4.28 (m)
	4	77.0	3.69 (dd, 7.0, 9.9)
	5	78.6	3.96 (m)
	6	36.7	2.47, 2.24 (ov)
	7	133.0	5.75 (m)
	8	131.7	6.17 (dd, 9.7, 15.0)
	9	133.7	6.34 (ov)
	10	130.7	6.33 (ov)
	11	129.2	6.92 (dd, 9.6, 15.6)
	12	131.3	6.56 (d, 15.6)
	13	137.0	
	14, 18	125.7	7.44 (d, 7.6)
	15, 17	128.4	7.31 (ov)
	16	127.1	7.21 (t, 7.3)
	45	*c*	
	46	26.6	1.35 (s)
	47	26.6	1.35 (s)

	OH-2		6.25
	NH-3		7.43
GABOB	19	172.6	
	20	40.3	2.22 (ov)
			2.00 (dd, 7.7, 10.8)
	21	65.2	3.87 (m)
	22	44.8	3.23 (ov)
			2.82 (m)
	NH-22		7.37
	OH-21		4.86
Gly	23	168.7	
	24	41.9	3.59 (m)
	NH-24		8.40
Trp	25	171.7	
	26	54.8	4.14 (m)
	27	25.1	3.21 (dd, 3.3, 10.3)
			3.00 (dd, 3.3, 10.3)
	28	123.5	7.21 (d, 5.3)
	29	109.6	
	30	126.5	
	31	118.0	7.52 (d, 7.7)
	32	118.1	6.98 (t, 7.3)
	33	120.8	7.06 (t, 7.3)
	34	111.1	7.32 (d, 7.9)
	35	135.9	
	NH-28		10.87 (brs)
	NH-26		8.65
N-Me-Gly	36	169.9	
	37	49.7	4.54 (d, 11.5)
			3.43 (ov)
	38	36.4	3.05 (s)
Pyrrolidone	39	170.4	
	40	87.3	5.23 (ov)
	41	*c*	
	42	50.4	3.69 (m)
	43	33.1	2.45, 2.75 (ov)
	44	172.9	
	NH-41		8.35
	NH-42		*c*

[a] Recorded at 175 MHz, referenced to residual solvent DMSO-d_6 at 39.51 ppm. Recorded at 700 MHz, referenced to residual solvent DMSO-d_6 at 2.50 ppm. [c] Not observed. [ov] overlapping signals.

Analytical data for microsclerodermins and pedeins

Microsclerodermin D

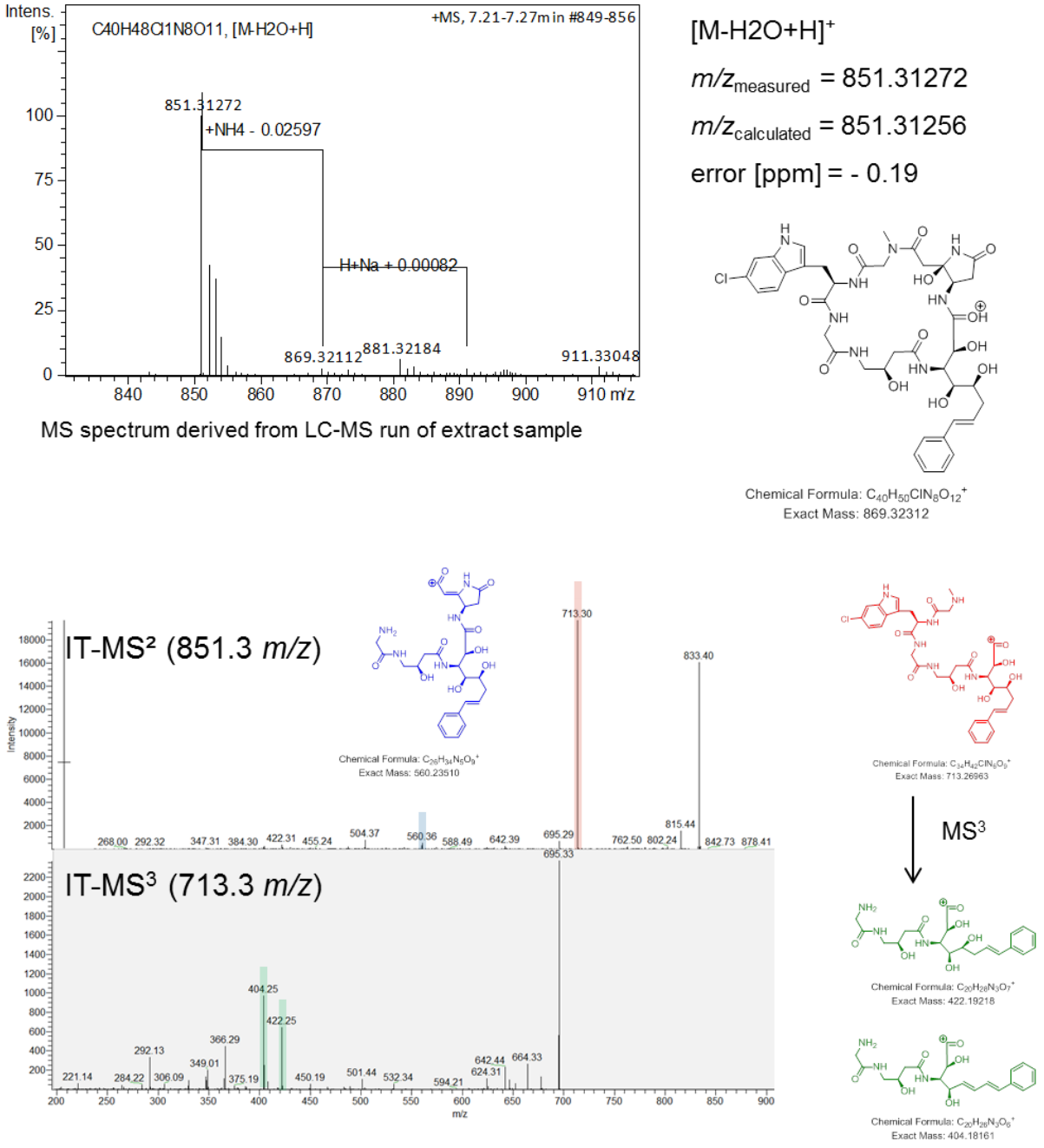
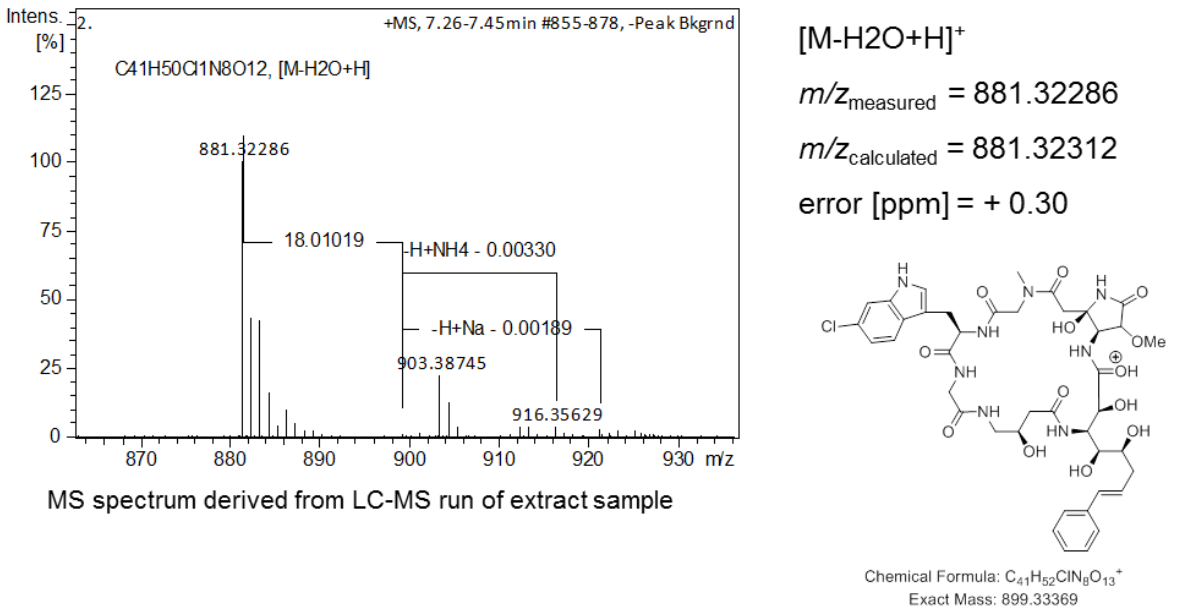


Figure S6: Accurate *m/z* measurement and CID fragmentation pattern of microsclerodermin D.

Microsclerodermin L



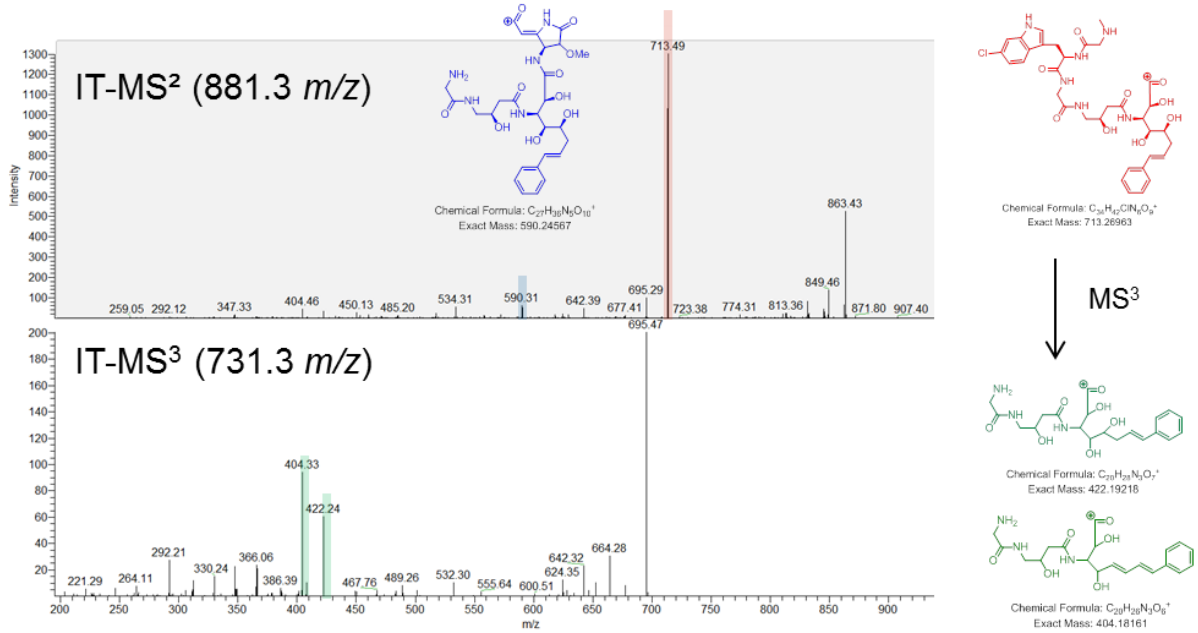


Figure S7: Accurate *m/z* measurement and CID fragmentation pattern of microsclerodermin L.

Microsclerodermin M

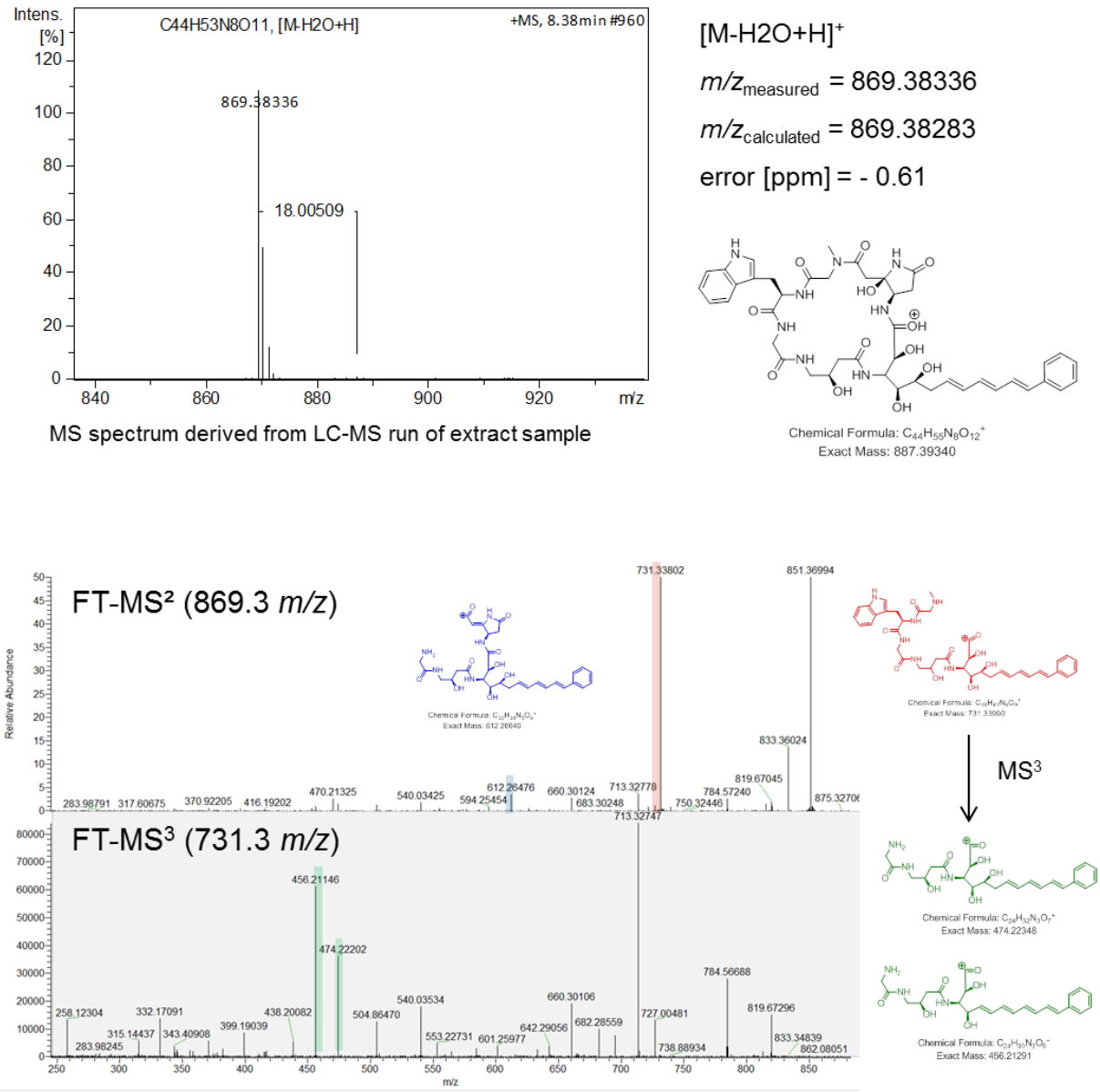


Figure S8: Accurate m/z measurement and CID fragmentation pattern of microsclerodermin M.

Pedein A

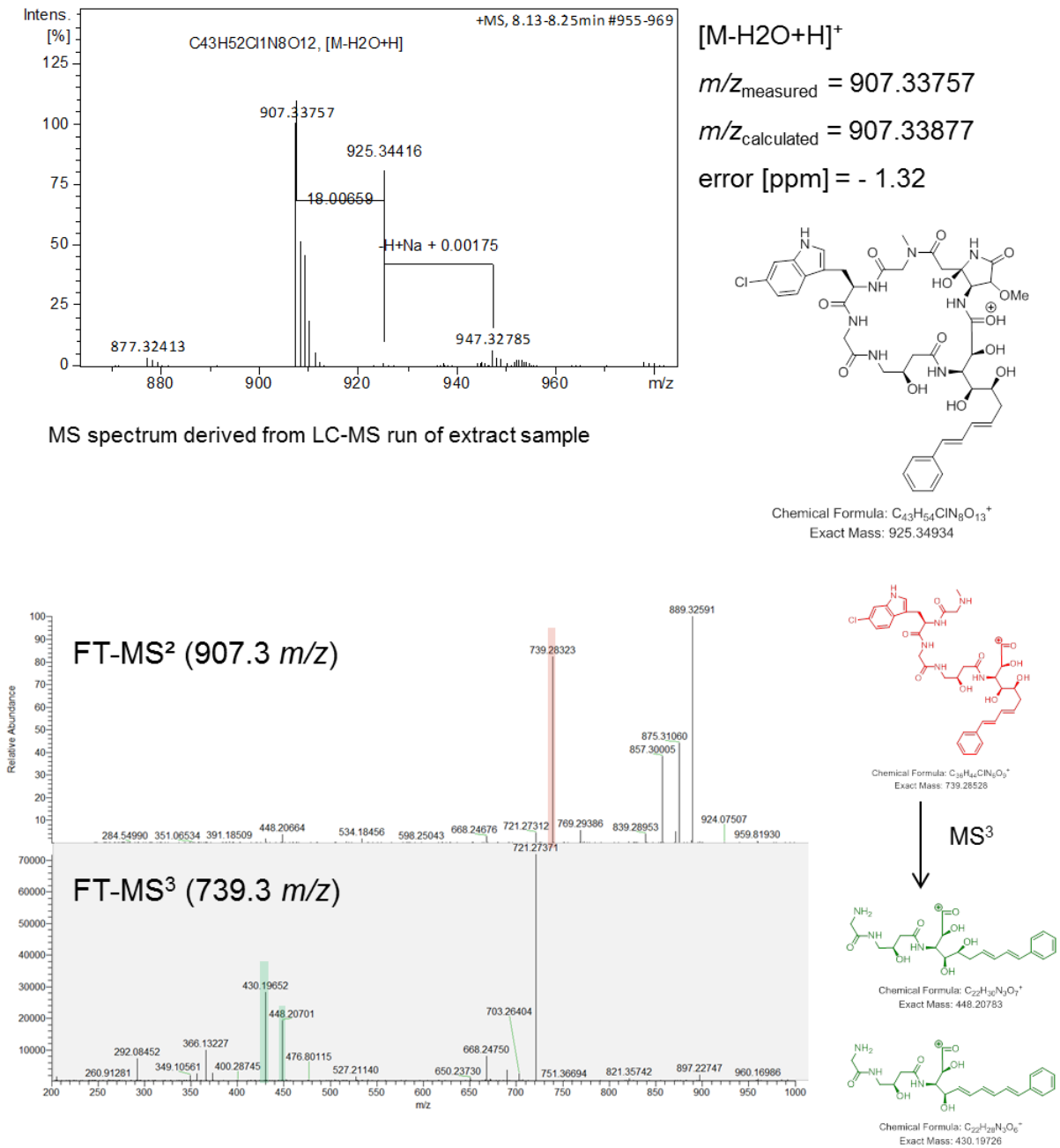
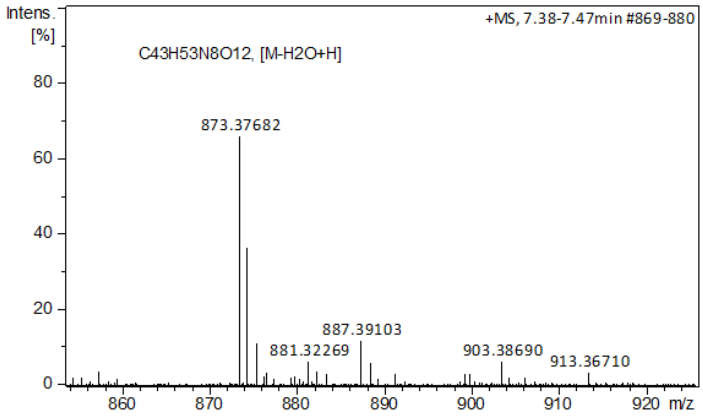


Figure S9: Accurate *m/z* measurement and CID fragmentation pattern of pedein A.

Pedein B



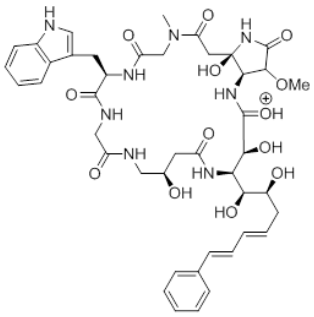
MS spectrum derived from LC-MS run of extract sample, thus coeluting with other derivatives of microsclerodermin.

$[M\text{-}H2O+H]^+$

$m/z_{measured}$ = 873.37682

$m/z_{calculated}$ = 873.37775

error [ppm] = 1.07



Chemical Formula: $C_{43}H_{55}N_8O_{13}^+$
Exact Mass: 891.38831

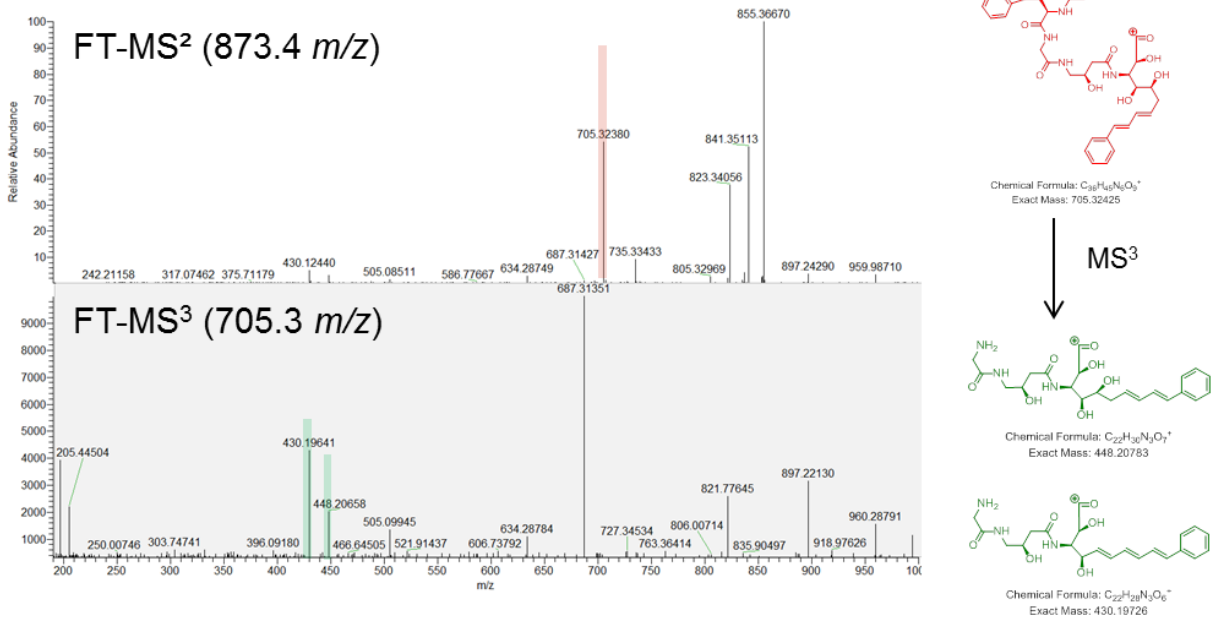


Figure S10: Accurate *m/z* measurement and CID fragmentation pattern of pedein B.

Analysis of the *msc* gene clusters in So ce38 and MSr9139

The regions of interest as well as 20 kb flanking regions were searched for open reading frames (ORFs) using Glimmer 3.0 and subsequently subjected to an automatic annotation by the antiSMASH software tool[1,2]. The *msc*-locus in So ce38 (Table S6) spans 58,049 bp and covers nine modules within the genes *mscA* to *mscI* with a GC content of 74.7 % (GenBank accession no. KF657738). The 5'-end of the gene cluster covers a major facilitator superfamily transporter (MFS-type, *mscK*) and a type II-thioesterase (*mscJ*) located approximately 2.5 kb upstream to *msc*A. In the case of *Jahnella* sp. MSr9139 the cluster has a size of 62,268 bp and is thereby approximately 4.2 kb larger. We find the same genes as in So ce38 starting from *mscA* to *mscI* (Table S6, GenBank accession no. KF657739). There are three additional auxiliary genes named *mscL*, *mscM* and *mscN*. The first one encodes a halogenase specific for tryptophan found at the end of the cluster. This is consistent with the observation that a chlorinated tryptophan residue is found for the microsclerodermins isolated from *Jahnella* specimen. The genes *mscM* and *mscN* are located within the cluster (between *mscD* and *mscE*). They are encoding two proteins with significant similarity to an α-ketoglutarate dependent dioxygenase and a methyltransferase, respectively. The pairwise similarity of both clusters is 74.2 % on DNA-level. Both *msc* clusters have a comparable GC content of around 72 %.

Table S6: Genes involved in microsclerodermin biosynthesis as annotated in the strains *Sorangium cellulosum* So ce38 and *Jahnella* sp. MSr9139

	Sorangium cellulosum So ce38				*Jahnella* sp. MSr9139				
Gene	Start	End	Length [bp]	Function[1]	Start	End	Length [bp]	Function[1]	pairwise identity
mscA	3859	13686	9828	CoA-Lig-KR*-ACP-KS-AT-DH-KR-ACP	4124	14731	10608	CoA-Lig-KR* -MT-ACP-KS-AT-DH-KR-ACP	74.7
mscB	13709	16321	2613	KS-AT*	14718	17354	2637	KS-AT*	79.7
mscC	16327	20982	4656	KS-AT-DH-KR-ACP	17341	21990	4650	KS-AT-DH-KR-ACP	81.5
mscD	20995	23697	2703	KS-AT*	22000	24546	2547	KS-AT*	78.7
mscE	23727	25067	1341	Putative amidohydrolase	27907	26747	1161	Putative amidohydrolase	83.9
mscF	25335	32156	6822	PCP-ATT-MOX-C-A-PCP	28287	34856	6570	PCP-AAT-MOX-C-A-PCP	82.4
mscG	32158	36804	4647	KS-AT-KR-ACP	34858	39393	4536	KS-AT-KR-ACP	80.9
mscH	36844	49269	12426	C-A-M-PCP-C-A-PCP-E-C-A-PCP	39472	51792	12321	C-A-M-PCP-C-A-PCP-E-C-A-PCP	84.5
mscI	49344	58049	8706	C-A-PCP-KS-AT-KR-ACP-TE	51789	60668	8880	C-A-PCP-KS-At-KR-ACP-TE	83.9
mscJ	1473	2246	774	Thioesterase typeII	3006	3794	789	Thioesterase typeII	63
mscK	1248	1	1248	MFS-transporter	1350	1	1350	MFS-transporter	55.9
mscL	-	-	-	-	62268	60661	1608	Trp-halogenase	-
mscM	-	-	-	-	24583	25899	1317	Fe(II)/α-ketoglutarate dependent oxygenase	-
mscN	-	-	-	-	25881	26714	834	Methyltransferase	-

[1] A: adenylation domain, AAT: aminotransferase, ACP: acyl-carrier-protein domain AT: acyltransferase domain, C: condensation domain, CoA-Lig: CoA-Ligase, DH: dehydratation domain, E: epimerase, KR: ketoreductase domain, KS: ketosynthase domain, MT: methyltransferase domain, MOX: monooxygenase, PCP: peptidyl-carrier-protein domain, TE: thioesterase domain, *: inactive domain

Table S7: Proteins involved in microsclerodermin biosynthesis in the strains *Sorangium cellulosum* So ce38 and *Jahnella* sp. MSr9139.

	Sorangium cellulosum So ce38		*Jahnella* sp. MSr9139		
Protein	Length [aa]	Domains and position in sequence[a]	Length [aa]	Domains and position in sequence[a]	Identity [%]
MscA	3275	CoA-Lig (264-701), KR*(1034-1197), ACP (1348-1411), KS (1441-1828), AT (1976-2286), DH (2342-2505), KR' (2870-3047), ACP' (3149-3214)	3535	CoA-Lig (215-649), KR*(1032-1126), MT (1229-1509), ACP (1613-1676), KS (1712-2147), AT (2244-2555), DH (2610-2774), KR*' (3132-3309), ACP' (3411-3476)	65.4
MscB	870	KS (27-451), AT* (548-772)	878	KS (39-464), AT* (561-792)	73.4
MscC	1551	KS (36-460), AT (557-859), DH* (956-1076), KR (1158-1336), ACP (1436-1505)	1549	KS (39-464), AT (561-865), DH* (956-1076), KR (1157-1335), ACP (1436-1499)	75.0
MscD	900	KS (36-461), AT* (565-678)	848	KS (36-461), AT* (564-675)	72.0
MscE	446	Putative amidohydrolase	386	Putative amidohydrolase	82.9
MscF	2273	PCP (27-98), AMT (329-660), MOX (828-1127), C (1185-1529), A (1673-2082), PCP' (2169-2237)	2189	PCP (4-75), AMT (280-614), MOX (758-1057), C (1105-1397), A (1594-2000), PCP' (2088-2149)	76.5
MscG	1548	KS (14-439), AT (534-828), KR (1156-1330), ACP (1441-1509)	1511	KS (14-439), AT (531-850), KR (1136-1312), ACP (1404-1472)	72.1
MscH	4141	C (76-377), A (564-966), MT (1037-1256), PCP (1469-1531), C' (1554-1850), A' (2037-2426), PCP' (2515-2574), E (2591-2905), C'' (3074-3374), A'' (3558-3967), PCP'' (4054-4121)	4106	C (48-346), A (534-936), MT (1007-1225), PCP (1442-1505), C' (1526-1827), A' (2013-2405), PCP' (2492-2551), E (2568-2872), C'' (3043-3343), A'' (3527-3932), PCP'' (4019-4083)	78.8
MscI	2904	C (48-346), A (533-936), PCP (1023-1087), KS (1111-1535), AT (1638-1936), KR (2266-2465), ACP (2549-2612), TE (2634-2888)	2945	C (77-375), A (563-965), PCP (1053-1116), KS (1150-1573), AT (1676-1970), KR (2309-2509), ACP (2597-2660), TE (2683-2945)	77.8
MscJ	257	Thioesterase type II	263	Thioesterase type II	43.5
MscK	415	Major Facilitator Superfamily (MFS) transporter	450	Major Facilitator Superfamily (MFS) transporter	24.5
MscL	-	-	535	Trytophan halogenase	-
MscM	-	-	438	Fe(II)/α-ketoglutarate dependent oxygenase	-
MscN	-	-	277	Methyltransferase	-

[a] A: adenylation domain, AMT: aminotransferase, ACP: acyl-carrier-protein domain AT: acyltransferase domain, C: condensation domain, CoA-Lig: coenzyme A Ligase, DH: dehydratation domain, E: epimerase, KR: ketoreductase domain, KS: ketosynthase domain, MT: methyltransferase domain, MOX: monooxygenase, PCP: peptidyl-carrier-protein domain, TE: thioesterase domain, *: inactive domain

AT domains

Substrate specificity of acyl transferase (AT) domains was determined using conserved motif analysis by aligning the AT domains from both clusters to a reference AT from *E.coli* FAS, 1MLA (PDB 1MLA, UniProtKB P0AAI9) (Table S8).[3] This data shows that most of the listed AT-domains contain the GxSxG consensus motif with the catalytic residue in the center, except MscB in which the catalytic serine residue is replaced by a glycine. The MscD-AT domains are considered inactive as they are heavily truncated and do not align to the canonical motifs of AT domains. Active site analysis was used to predict which extender unit is recruited by a distinct AT domain. The main discriminant between using a methylmalonyl-CoA (mm-CoA) or malonyl-CoA (m-CoA) is the amino acid position 200 in 1MLA, where Ser200 supports mm-CoA and Phe200 supports m-CoA. The *in silico* prediction for microsclerodermin production is in all cases m-CoA and thereby in agreement with the chemical structure observed. We were not able to predict the incorporation of a hydroxymalonyl-CoA substrate for MscC-AT.

Table S8: Active site analysis of the acyl transferase domains (AT) from *S. cellulosum* So ce38 and *Jahnella* sp. MSr9139 according to Yadav et al.[3]

Domain	observed	predicted	11	63	90	91	92	93	94	117	200	201	231	250	255	15	58	59	60	61	62	70	72	197	198	199
1MLA	M	M	Q	Q	G	H	S	L	G	R	S	H	N	Q	V	T	K	T	W	Q	T	S	A	S	V	P
MscA-AT$_{So\ ce38}$	M	M	Q	Q	G	H	S	V	G	R	F	H	N	H	V	Q	R	T	E	Y	T	E	A	S	H	A
MscA-AT$_{Jahnella\ sp.}$	M	M	Q	Q	G	H	S	V	G	R	F	H	N	H	V	Q	R	T	E	Y	T	E	A	S	H	A
MscB-AT$_{So\ ce38}$	inactive	-	E	Q	G	Q	G	A	G	R	-	-	S	D	D	T	Q	T	A	F	T	Q	A	-	-	-
MscB-AT$_{Jahnella\ sp.}$	inactive	-	Q	Q	G	L	G	V	G	R	-	-	S	Q	D	L	Q	A	A	F	A	E	A	-	-	-
MscC-AT$_{So\ ce38}$	Hydroxy malonate	M	Q	Q	G	H	S	I	G	R	F	H	N	H	V	Q	Q	T	A	F	T	E	A	S	H	A
MscC-AT$_{Jahnella\ sp.}$	Hydroxy malonate	M	Q	Q	G	H	S	I	G	R	F	H	N	H	V	F	Q	T	A	F	A	E	A	S	H	A
MscG-AT$_{So\ ce38}$	M	M	Q	Q	G	H	S	I	G	R	F	H	N	H	V	H	D	T	A	I	A	E	A	S	H	A
MscG-AT$_{Jahnella\ sp}$	M	M	Q	Q	G	H	S	V	G	R	F	H	N	H	V	H	D	T	A	L	A	G	A	S	H	A
MscI-AT$_{So\ ce38}$	M	M	Q	Q	G	H	S	V	G	R	F	H	N	H	V	Y	Q	T	R	L	T	E	A	S	H	A
MscI-AT$_{Jahnella\ sp}$	M	M	Q	Q	G	H	S	V	G	R	F	H	N	H	V	H	Q	T	R	L	A	E	A	S	H	A

KS domains

The ketosynthase (KS) domains of both *msc* biosynthetic gene clusters were extracted and analyzed using the NaPDoS web tool.[4] The result is depicted as a phylogenetic tree in Figure S11 and allows easy annotation of the KS subtypes found in the *msc* cluster.

The postulated iterative function of MscB is supported by grouping of these domains at the interface between modular (ochre) and iterative (green) domains. The remaining, functional domains MscA, MscG, and MscI are correctly grouped in the PKS/NRPS hybrid section. MscA is characterized as a hybrid domain because of its linkage to the first module with its uncommon starter unit.

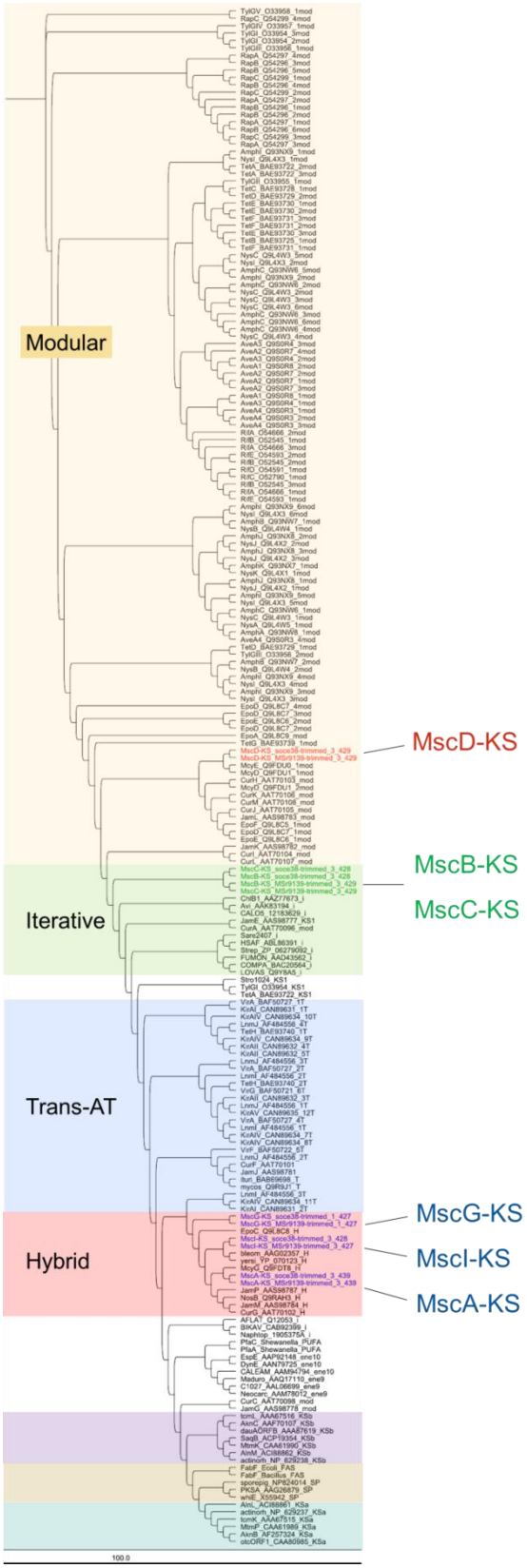


Figure S11: Phylogenetic tree highlighting different KS domain types and the KS domains found in the *msc* biosynthetic gene cluster in So ce38 and MSr9139.

DH domains

Both *msc* clusters contain two dehydration domains present in the proteins MscA and MscC. Based on the proposed biosynthesis of microsclerodermin the DH-domain in MscA must be active. However, it does not exhibit the common consensus-motif $H(X_3)G(X_4)P$ of DH-domains as known from other modular PKS-systems (Table S9). Here, the central glycine is replaced by a valine. This substitution is obviously not affecting the enzyme function. The DH-domains from MscC are considered to be inactive as they do not contain a conserved motif at all. The domains are truncated with a protein length of around 100 amino acids.

Table S9: Active site analysis of the dehydratase domains (DH) of the msc biosynthetic gene cluster.

Consensus	H	x	x	x	G	x	x	x	x	P
MscA-DHSo ce38	H	A	E	G	V	L	V	A	R	P
MscA-DHJahnella sp	H	V	E	G	V	V	V	A	R	P

KR domains

The *msc* cluster from both myxobacteria encodes five ketoreductase domains. The KR domains may be classified as A- and B-type by analyzing the amino acid sequence of two regions within the amino acid sequence.[5] However, it was shown that a reliable prediction of the stereogenic center that emerges upon KR reduction of a keto-function is governed by additional, unknown interactions.[6] Nevertheless, we attempted to classify the domains in this work. The occurrence of a LDD motif is a good indicator for a B-type KR domain. This is sometimes supported by the presence of a Pro144 and an Asp148 according to the numbering based on the DEBS2 KR domain (GenBank #X62569.1). In case of A-type KR domains a highly conserved Trp141 is present. Moreover, a correct co-factor binding motif is mandatory for the enzyme function. Based on this, MscA-KR1 of both strains is predicted to be inactive as it does not exhibit the co-factor binding motif GxGxxGxxxA (Table S10). Additionally, the MscA-KR1 domain of MSr9139 is heavily truncated. The remaining KR domains feature the correct binding motif.

Table S10: Analysis of the co-factor binding motif of the ketoreductase domains (KR) within the *msc* biosynthetic gene cluster.

Domain	G	x	G	x	x	G	x	x	x	A	motif
MscA-KR1 $_{So\ ce38}$	D	L	G	G	V	S	L	Q	L	L	inactive
MscA-KR1 $_{MSr9139}$	A	L	G	S	Y	E	-	-	-	-	inactive
MscA-KR2 $_{So\ ce38}$	G	L	G	A	L	G	R	R	V	A	ok
MscA-KR2 $_{MSr9139}$	G	L	G	A	L	G	R	R	V	A	ok

MscC-KR$_{So\ ce38}$	G	T	G	A	L	G	L	H	I	A	ok
MscC-KR$_{MSr9139}$	G	I	G	A	L	G	L	H	T	A	ok
MscG-KR$_{So\ ce38}$	G	L	G	G	A	G	L	G	I	A	ok
MscG-KR$_{MSr9139}$	G	L	G	G	V	G	L	A	I	A	ok
MscI-KR$_{So\ ce38}$	G	L	G	R	I	G	L	C	L	A	ok
MscI-KR$_{MSr9139}$	G	L	G	R	I	G	L	C	L	A	ok

The two amino acid regions for KR domain classification are shown in Table S11. MscA-KR2 and MscC-KR1 can be classified as B-type and A-type, respectively. The LDD-motif together with an Asp148 is found in MscA-KR2 whereas the highly conserved Trp141 specifies MscC-KR. One stereogenic center originating from the A-type MscC-KR is still present in the final product and should exhibit *S*-configuration. This is in agreement with the configuration of the last hydroxyl group reported for the microsclerodermin side chain. The KR domains of MscG do only feature a LDD-like motif but no additional identifier in the second region. It remains unclear whether this domain is active at all. MscI-KR produces an *R*-configured hydroxyl moiety as identified by advanced Marfey analysis of microsclerodermin (GABA subunit). However, there is no common consensus that allows a classification. Moreover, the KR domains exceed the common KR domain length of approximately 180 amino acids by additional 30 amino acids.

Table S11: Analysis of the ketoreductase domains (KR) within the *msc* biosynthetic gene cluster

	region 88-103																region 134-149														KR type
MscA-KR2 $_{So\ ce38}$	H	L	A	G	A	L	D	D	G	V	L	L	Q	Q	S	W	F	S	S	I	A	S	V	L	G	S	A	A	Q	G	N	Y	B-type
MscA-KR2 $_{MSr9139}$	H	L	A	G	S	L	D	D	G	V	L	V	Q	Q	S	W	F	S	S	I	A	S	I	L	G	S	A	G	Q	G	N	Y	B-type
MscC-KR$_{So\ ce38}$	H	A	A	G	V	A	G	A	R	D	L	S	A	L	D	A	T	S	S	I	A	S	L	W	G	S	R	G	Q	A	H	Y	A-type
MscC-KR$_{MSr9139}$	H	A	A	G	V	G	G	Y	C	E	L	S	R	L	D	A	Y	S	S	I	A	S	L	W	G	S	R	G	Q	A	H	Y	A-type
MscG-KR$_{So\ ce38}$	Y	A	A	G	V	D	D	P	A	L	I	G	-	G	I	G	S	S	S	L	S	A	I	L	G	G	R	G	L	G	A	Y	B-type
MscG-KR$_{MSr9139}$	Y	A	A	A	I	D	D	A	G	L	L	G	-	A	V	G	G	S	S	L	S	A	I	L	G	G	R	G	L	A	A	Y	B-type
MscI-KR$_{So\ ce38}$	H	A	A	G	A	R	G	D	G	T	F	M	V	P	L	A	L	S	S	T	S	A	I	L	G	G	L	G	L	G	P	Y	unknown
MscI-KR$_{MSr9139}$	H	A	A	G	A	R	G	D	G	T	F	M	T	S	L	A	L	S	S	T	S	A	I	L	G	G	L	G	L	G	P	Y	unknown

A domains

Three genes of microsclerodermin biosynthesis (*mscF*, *mscH* and *mscI*) harbor at least one NRPS module. In order to determine the A domain specificity of each module the respective amino acid sequences were analyzed using the NRPSpredictor2 tool.[7] The results are listed in Table S12. The predicted specificity of each A-domain is consistent with the observed amino acid incorporated into the microsclerodermin scaffold.

Table S12: Prediction of A-domain specificity for each A-domain present in the *msc* biosynthetic gene cluster using the NRPSpredictor2 software-tool.[7]

	amino-acid specificity	
domain	So ce38	MSr9139
MscF-A	asn	asn
MscH-A1	gly	gly
MscH-A2	trp	trp
MscH-A3	gly	gly
MscI-A	gly	gly

C domains

All C domains of the *msc* biosynthetic gene cluster were analyzed using the NaPDoS web tool.[4] Especially the hybrid module MscF is of interest as the incorporated *S*-asparagine is (a) cyclized to form and pyrrolidone moiety and (b) converted to the *R*-asparagine. The mechanism behind these reactions is not understood yet. Analysis of the respective C domain, MscF-C1, characterizes the domain as a common PKS/NRPS hybrid domain (Figure S12). Hence, the observed cyclization as well as inversion of asparagine is not done by the C domain. It is more likely related to the amidohydrolase MscE which is found in both clusters. Its detailed function remains elusive at the current stage of research.

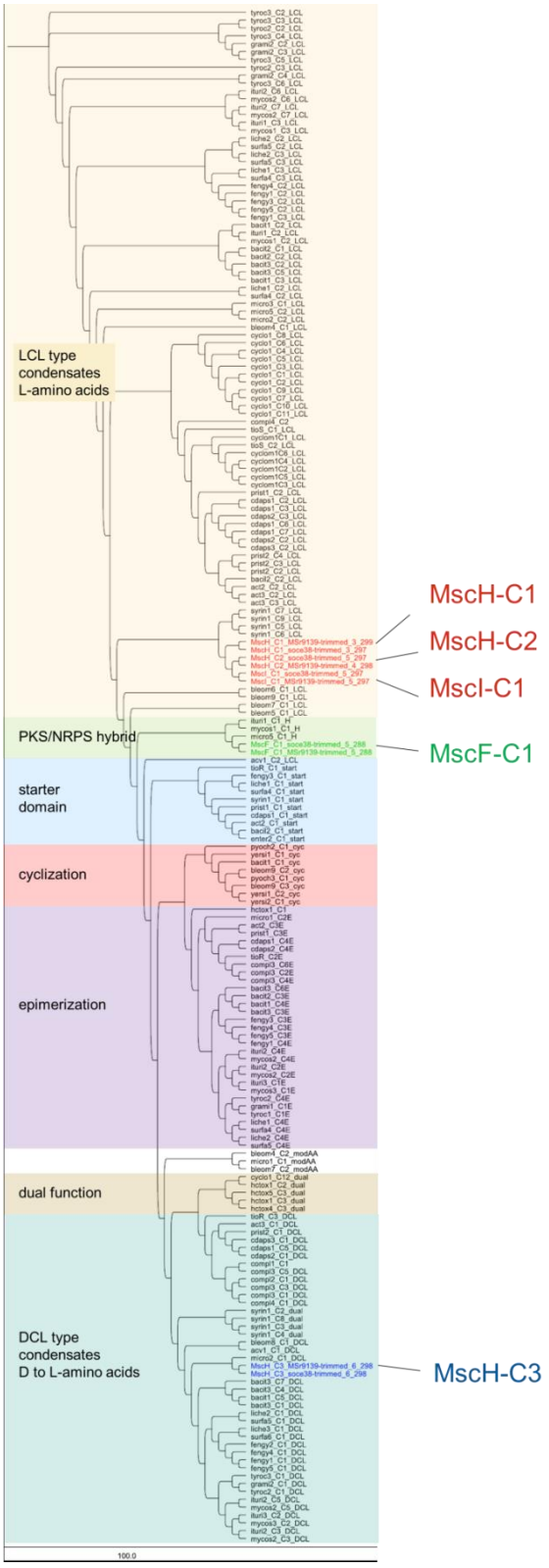


Figure S12: Phylogenetic tree highlighting different C domain types and the C-domains found in the *msc* biosynthetic gene cluster in So ce38 and MSr9139.

MT-domains

The *N*-methyltransferase present in module 1 of MscH is found in both clusters and is responsible for the *N*-methylation of the glycine right next to the pyrrolidone ring. Both of them harbor the common consensus-motif GxGxG at the *N*-terminal end which is responsible for SAM Co-factor binding (Table S13).[8] These *N*-methyltransferases belong to the UbiE/COQ5 family. In addition, the *msc*-cluster of *Jahnella* sp. encodes a second methyltransferase domain which is located in MscA. This methyl transferase has homology to the PRMT5 family. However, the MT domain is likely inactive as the SAM binding motif GxGxG is changed to GxGxE. The exchange of the small glycine to a charged glutamic acid may prevent cofactor binding.

Table S13: Active site analysis of the methyltransferase domains (MT) within the *msc* biosynthetic gene cluster.

active consensus	G	x	G	X	G
MscA-MT$_{Jahnella\ sp.}$	G	T	G	K	E
MscH-MT$_{So\ ce38}$	G	C	G	T	G
MscH-MT$_{Jahnella\ sp.}$	G	C	G	T	G

Determination of configuration

The conformation of all stereogenic centers of the isolated microsclerodermins is identical to that of the microsclerodermins C – I and confirms the assumed stereochemistry of the pedeins.[9–11] In detail, the scaffold is based on an *R*-tryptophan, a (*3R*)-4-NH_2-3-OH-butyric acid moiety, a side chain with (*2S,3R,4S,5S*)-configuration and an (*R,R*)-pyrrolidone.

Absolute configuration of 3-aminobutyric acid and tryptophan

Microsclerodermin has several stereogenic centers. We identified the configuration of the γ-aminobutyric acid (GABA) moiety and tryptophan by acid hydrolysis followed by advanced Marfey analysis using FDLA reagent. (3*S*)-GABA was ordered from Sigma and used as a reference together with *S*-tryptophan and *S*-6-chloro-tryptophan. Derivatization of the γ-amino functionality with L-FDLA and D-FDLA results in two products with reproducible retention time differences (3 seconds using 10 technical replicates, see Figure S13) which allows annotation of the stereochemistry. In conclusion, the new microsclerodermins L and M have an (3*R)*-configured carbon atom in the GABA subunit which is in agreement with the other microsclerodermins reported.

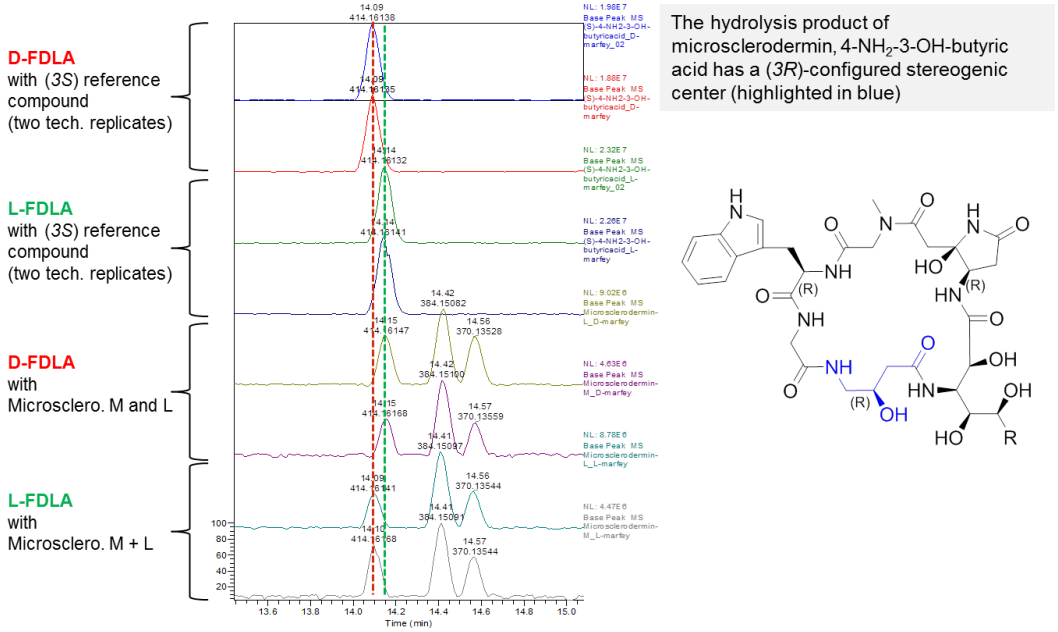


Figure S13: Overlay of LC-MS chromatograms highlighting the retention time difference of the GABA subunit when derivatized with Marfey's reagent (FDLA). The graph features the (3S)-GABA reference compound as well as microsclerodermin M and L.

The tryptophan in microsclerodermin M, D and L is *R*-configured. This applies to all microsclerodermins known so far and is in agreement with the epimerization domain found within the respective module of the biosynthetic assembly lines in So ce38 and MSr9139. Microsclerodermin M was compared to *S*-tryptophan whereas microsclerodermin D and L were compared to a *S*-6-Cl-tryptophan reference (Figure S14).

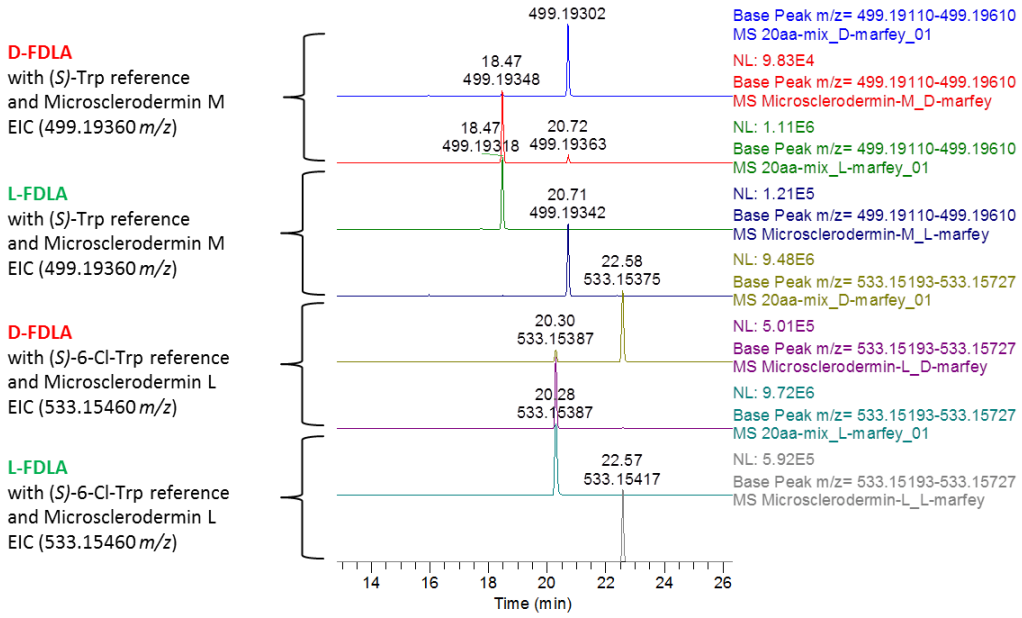


Figure S14: Overlay of the LC-MS chromatograms for tryptophan characterization using advanced Marfey's reagent (FDLA). The stereogenic center of tryptophan is (R)-configured in microsclerodermin M and L.

Configuration of the pyrrolidone moiety

The pyrrolidone moiety of the microsclerodermins is derived from an asparagine as indicated by feeding labeled *S*-asparagine. Incorporation of labeled *S*-asparagine suggested that the pyrrolidone ring keeps this configuration. However, this turned out to be wrong. Microsclerodermins M, D and L underwent the same chemical degradation steps as reported for the microsclerodermins A-I.[9,10,12] After dehydration at the pyrrolidone moiety the respective dehydromicrosclerodermins were subjected to ozonolysis. The reaction products were hydrolyzed and derivatized with Marfey's reagent. Marfey derivatization and HPLC-MS measurements revealed an *R*-configured pyrrolidone moiety in the microsclerodermin derivatives M, L, and D. Although this chemical degradation protocol is used for the microsclerodermins A – I in literature, the method is not that elegant in terms of the result as ozonolysis, oxidative workup and hydrolysis of *R*-tryptophan will result in *R*-asparate as well. At the same time this procedure will generate aspartate out of the pyrrolidone moiety. By doing a short hydrolysis step of not more than 60 minutes we observe a notable amount of *R*-asparagine in the sample which can only originate from the pyrrolidone moiety. Based on these results we were able to identify *R*-tryptophan and *R*-pyrrolidone in microsclerodermin M and D. We verified the results by applying the above mentioned protocol to pure *R*-tryptophan as well as to the synthesized, enantiomeric pure *S*-dehydropyrrolidone fragment (98 % ee). *R*-tryptophan is converted to *R*-aspartate while the *S*-dehydropyrrolidone fragment is converted to asparagine

and aspartate with a *S*:*R*-ratio of approximately 4:1. This partial inversion clarifies the origin of the *R*- and *S*-asparagine mixture which is observed in all samples of microsclerodermin, there with the opposite *S*:*R*-ratio of 1:4 (Figures S15).

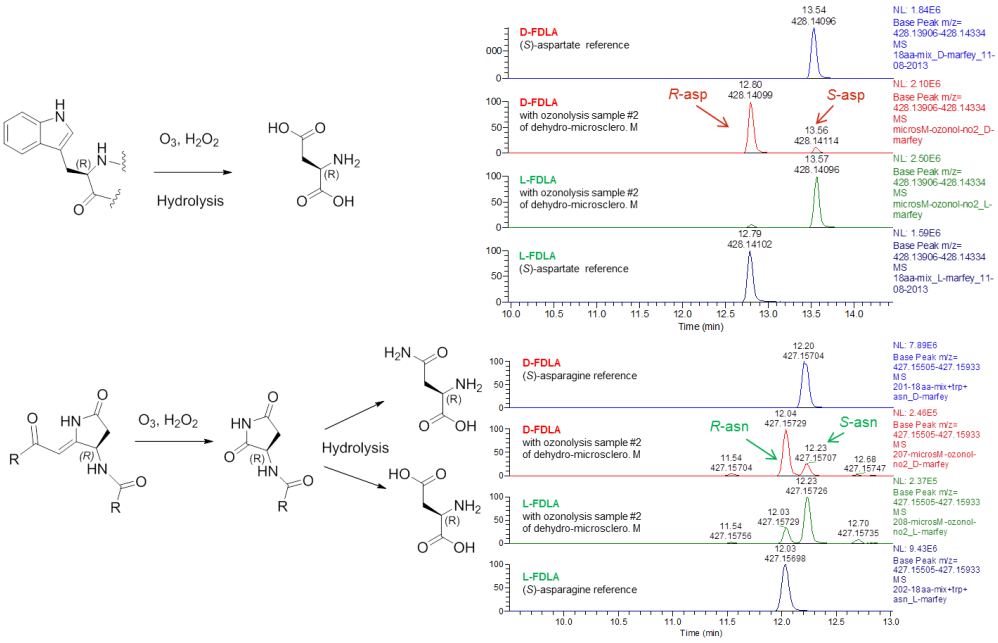


Figure S15: Marfey analysis result of ozonolysis product of microsclerodermin M. The chromatograms show the peaks for asparagine and aspartate, both as an overlay display with the respective S-configured reference compound.

Microsclerodermin D and L are both isolated from strain MSr9139. In case of microsclerodermin L, the pyrrolidone moiety has an additional methoxy group resulting in a 3-methoxy-aspartate after the above mentioned reactions whereas microsclerodermin D has the same pyrrolidone moiety as the M derivative (So ce38). Unfortunately we were not able to get suitable enantiomeric pure reference compounds for 3-methoxy-aspartate. However, based on the rule of thumb that an L-Marfey derivative of an L-configured amino acid elutes earlier than that of an D-Marfey derivative[13] we can conclude that it is the other way round in our sample and we do have an D-configured amino acid, (*2R*)-3-methoxy-aspartate, present in microsclerodermin L (Figure S16). This is supported by the detection of *R*-asparagine in the microsclerodermin D derived sample which is isolated from the same strain and thereby underlying the same biosynthesis. The stereogenic center at the methoxy group was not assigned by this method.

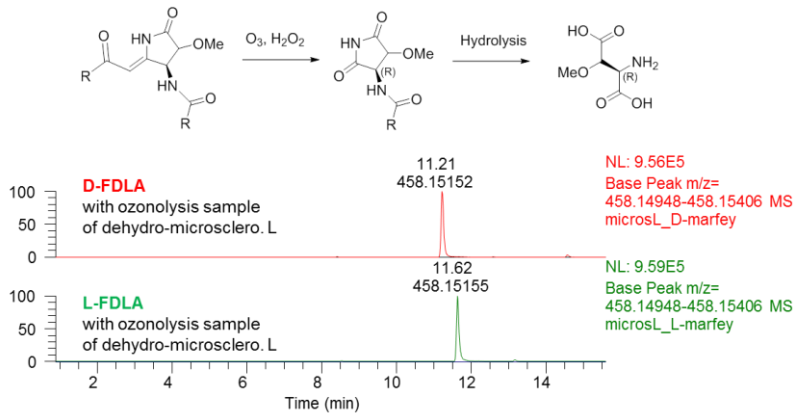


Figure S16: Marfey analysis result of ozonolysis product of microsclerodermin L. The chromatograms show the peaks for 3-MeO-aspartate. As the D-marfey derivative (upper part) elutes before the L-marfey derivative (lower part) the 3-MeO-aspartate has a 2R-configured stereogenic center. There is no reference compound available.

Configuration of the side chain

The side chain's vicinal hydroxyl groups OH-4 and OH-5 were converted to an acetonide in order to identify the relative stereochemistry of this part of the molecule. The acetonide formation was done for microsclerodermin M from *S. cellulosum* So ce38 and microsclerodermin L from *Jahnella* sp. MSr9139. Acetonides were purified using the same HPLC method as for the natural product isolation. Purity was checked by HPLC-MS analysis prior to NMR analysis. Respective NOE couplings are detected using selective irradiation experiments for the protons attached to the acetonide moiety. The NOE couplings are identical to those determined for the microsclerodermins known to literature. Based on this result we conclude to have the same relative stereochemistry as reported for microsclerodermins A – I which have a *2S,3R,4S,5S*-configured side chain. We assigned the absolute stereochemistry of the side chain by diol cleavage at OH-4 and OH-5. The product is oxidized to the respective carboxylic acid and hydrolyzed using 6 N HCl. After Marfey derivatization we observed a peak for L-threo-ß-OH-aspartate (*2S,3S* configuration) which would be in agreement with the *2S,3R,4S,5S*-configured side chain.

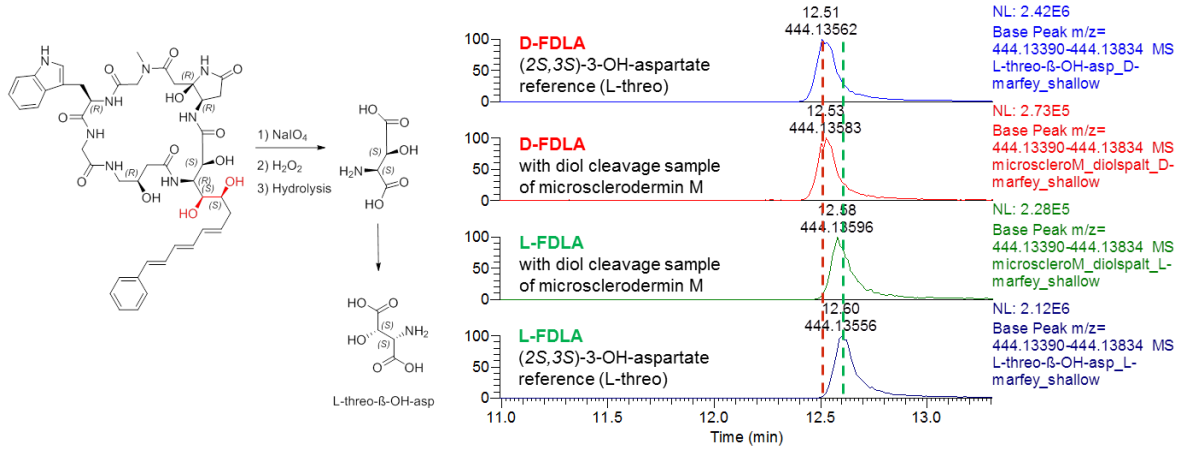


Figure S17: Marfey analysis of diol cleavage product (Microsclerodermin M).Peak identified by reference compound L-threo-ß-OH-aspartate.

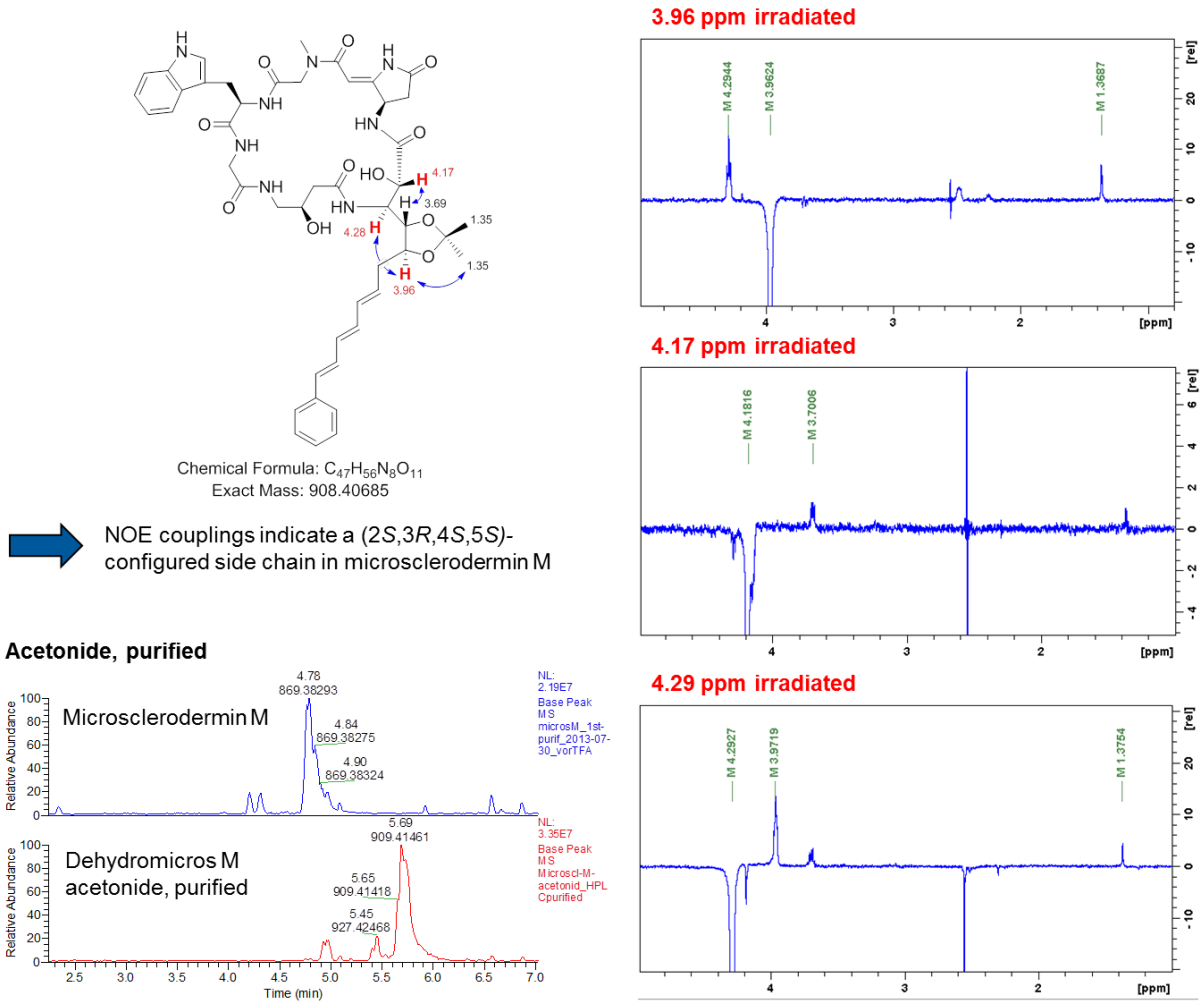
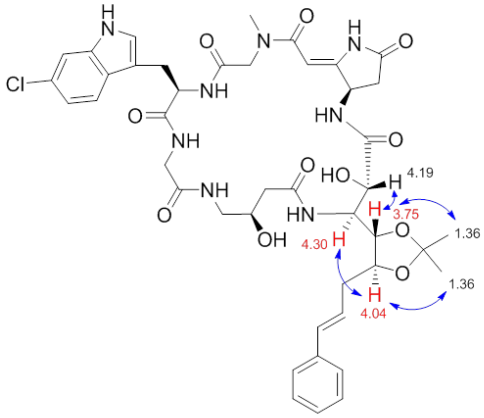


Figure S18: Acetonide purification and selective irradiation experiments of dehydromicrosclerodermin M.



Dehydromicrosclerodermin D acetonide
Chemical Formula: $C_{43}H_{51}ClN_8O_{11}$
Exact Mass: 890.33658

NOE couplings indicate a (2*S*,3*R*,4*S*,5*S*)-configured side chain in microsclerodermin D

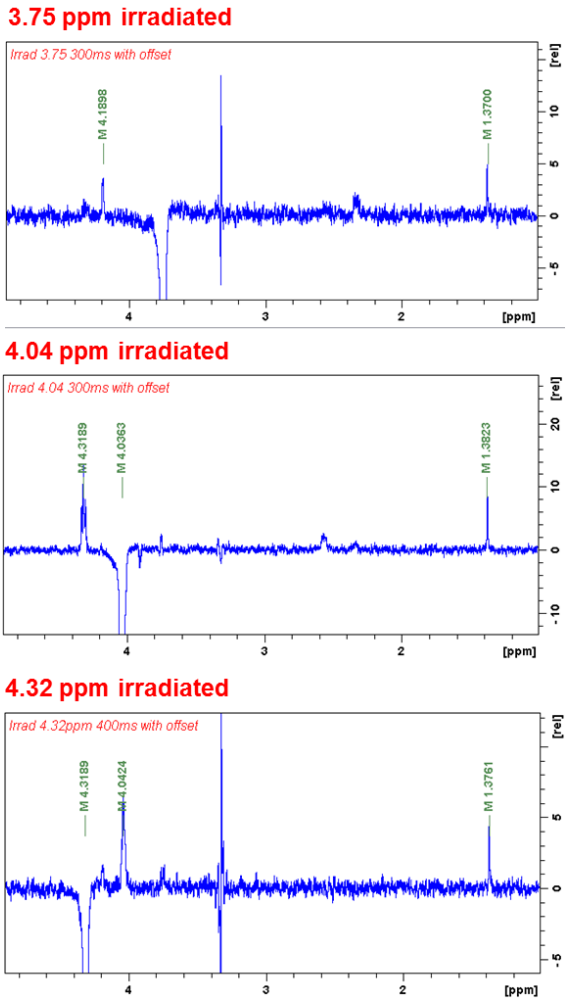


Figure S19: Acetonide purification and selective irradiation experiments of dehydromicrosclerodermin D.

Experimental section

Dehydration of microsclerodermin

A solution of microsclerodermin (2 mg) in 100 µl DMSO was diluted with 1 % aqueous TFA (400 µl). The mixture was kept at 40 °C for 15 minutes and dried afterwards using a GeneVac Evaporator (GeneVac Ltd, Suffolk, UK). Quantitative conversion to the respective dehydromicrosclerodermin was verified by dissolving the products in methanol and measuring them by LC-MS analysis.

Ozonolysis of dehydromicrosclerodermin

Dehydromicrosclerodermin (100 µg) was dissolved in 400 µL dry methanol. A stream of ozone was bubbled through the cooled solution (-78 °C) for 20 minutes. The excess reagent was removed by a stream of nitrogen before warming the solution to 20 °C. Remaining solvent was

then removed under reduced pressure and the dry ozonide dissolved in 1 mL formic acid/hydrogen peroxide (2:1, v/v). The reaction mixture was heated to 80 °C for 30 min before drying the mixture under reduced pressure. The sample was then used for hydrolysis and Marfey derivatization.

Diol cleavage of microsclerodermin

Microsclerodermin (500 µg) was dissolved in 50 µL DMSO and added to 300 µL of sodium periodate solution in water (0.5 mg/mL), adjusted to pH 4.0 with acetic acid. The sample was stirred at room temperature for 16 hours and dried afterwards. The residue was dissolved in hydrogen peroxide (1 mL) and formic acid (0.5 mL) and heated to 70 °C for 20 min. After removing the solvent under reduced pressure the sample was ready for hydrolysis and Marfey derivatization.

Acetonide of dehydromicrosclerodermin

Under an atmosphere of dry nitrogen 200 µL of 2,2-dimethoxy propane is added to a stirred solution of dehydromicrosclerodermin (2 mg) in 150 µL dry DMF. The reaction is initiated by adding catalytic amounts of pyridinium p-toluene sulfonic acid and stirred for 16 hours at room temperature. The reaction is quenched by adding 20 µl pyridine and dried using a stream of nitrogen. The residues are diluted in 100 µL H_2O/DMSO (1:1, v/v) and the respective acetonide purified using the same HPLC method as for the initial purification of the microsclerodermins in this work.

Marfey derivatization protocol

Marfey derivatization was performed with NMR, ozonolysis and diol-cleavage samples of the compounds. The protocol is as follows:

- Put at least 50 µg of sample into a 1.4 mL glass vial
- Add 100 µL of 6 N HCl. Fill the vial with nitrogen and close it. Keep it at 110 °C for 45 minutes. Open the vial to let dry at 110 °C for another 15 minutes. Do not exceed 60 min as tryptophan easily decomposes.
- Dissolve the residues in 110 µL H_2O. Prepare two 1.5 mL PP tubes and add 50 µL of the aqueous solution in each.
- Add 20 µL of 1 N $NaHCO_3$ in each tube (pH adjusted to approx. 9)
- Add 20 µL of 1 % (w/v) Marfey's reagent in acetone: D-FDLA or L-FDLA, respectively.
- Keep for 1 hour at 40 °C, 700 rpm.
- Add 10 µL of 2 N HCl to stop reaction and 300 µL of ACN to end up with 400 µL total volume.
- Centrifuge the sample and measure the supernatant by HPLC-MS

HPLC-MS analysis of Marfey samples

All measurements were performed on a Dionex Ultimate 3000 RSLC system using a Waters BEH C18, 100 x 2.1 mm, 1.7 µm d_p column by injection of 1 µl sample. Separation was achieved by a gradient using (A) H_2O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 550 µl/min and 45 °C. The gradient was as follows: starting at 5 % B to increase to 10 % B in 1 min, from 1 to 15 min increase to 35 % B, from 15 to 22 min increase to 50 % B, from 22 to 25 min increase to 80 % B. After a 1 min hold at 80 % B the system was reequilibrated with initial conditions for 5 minutes. UV data was acquired at 340 ± 8 nm and MS detection was performed simultaneously. Coupling the HPLC to the MS was supported by an Advion Triversa Nanomate nano-ESI system attached to a Thermo Fisher Orbitrap. LC flow is split to 500 nL/min before entering the ion source. Mass spectra were acquired in centroid mode ranging from 150 – 1000 m/z at a resolution of R = 30000.

HPLC-MS based screening

We searched our in-house database for producers of microsclerodermins. This screening included LC-MS data from 791 myxobacterial extracts covering a broad diversity of myxobacterial genera (Figure S20). Eleven different *Sorangium* sp. were identified as producers of microsclerodermin M. Microsclerodermin M is a rather instable compound that readily forms isobaric isomers. This effect is most likely related to the unsaturated side chain and causes several peaks that nearly coelute under standard LC-MS screening conditions. Thus, extracted ion chromatograms (EICs) for the $[M\text{-}H_2O+H]^+$ signal at 869.38283 *m/z* (blue trace, Figure S21) show up with a broadened peak including multiple isomers. In addition, the $[M+H]^+$ signal at 887.39340 *m/z* is detected albeit with lower abundance. We observe an additional, later eluting peak in the EICs with a $[M+H]^+$ signal at 869.38283 *m/z*, most likely caused by a loss of water at the pyrrolidone ring upon extract preparation and storage.

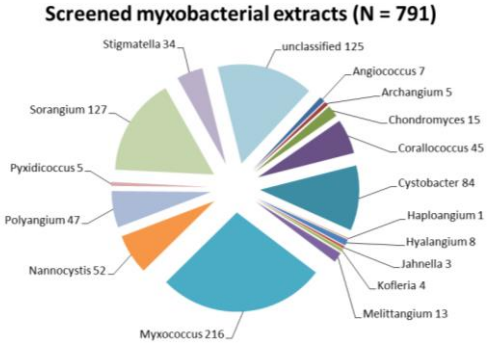


Figure S20: Myxobacterial genera as covered in our LC-MS based screening data set.

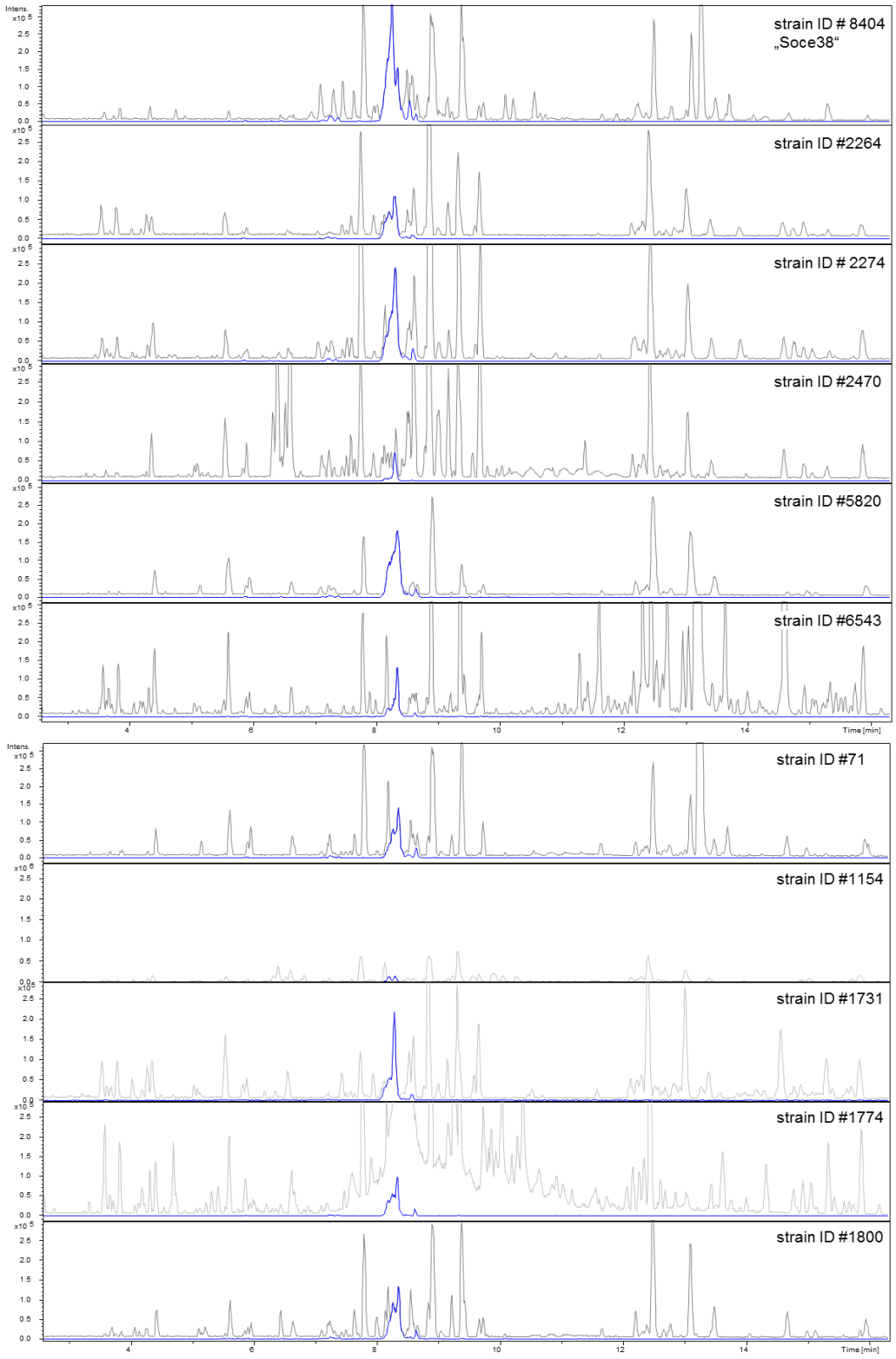


Figure S21: List of microsclerodermin M producers as identified by searching our in-house database. All strains are of the genus *Sorangium*. Microsclerodermin M and its isobaric isomers are highlighted by the EIC of 869.38283 m/z ($[M\text{-}H_2O+H]^+$ ion, blue trace). Abundance of derivatives is related to the different producer strain. Irregular peak shape is due to isobaric, almost coeluting derivatives occurring during storage of the extracts. Identification of microsclerodermin M is based on accurate *m/z*, isotope pattern fit, retention time and MS^2 pattern compared to the reference compound.

Another group of myxobacteria is producing the known microsclerodermin D as well as pedein A and B. In addition, one new derivative (Microsclerodermin L) is identified in this work (Figure S22). The strain *Chondromyces pediculatus* Cm p3 is also part of this group.[11]

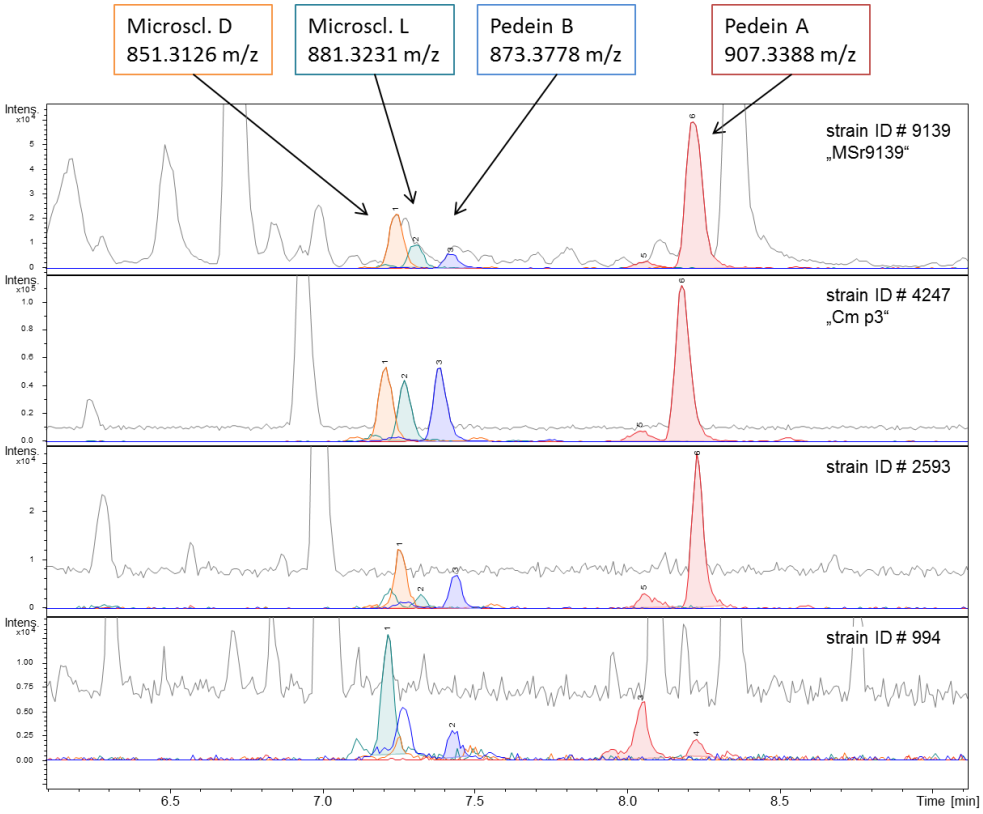


Figure S22: The known derivatives of microsclerodermin in four producer strains are highlighted by colored extracted ion chromatograms (EICs) overlaying the base peak chromatogram (BPC, grey). The producers are of the genus *Jahnella* and *Chondromyces*. Abundance of derivatives is related to the different producer strain. Identification of microsclerodermin derivatives is based on accurate *m/z*, isotope pattern fit, retention time and MS^2 pattern compared to the reference compounds.

LC-MS method for screening

All measurements were performed on a Dionex Ultimate 3000 RSLC system using a BEH C18, 100 x 2.1 mm, 1.7 µm dp column (Waters, Germany). Separation of 1 µl sample was achieved by a linear gradient from (A) H2O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 600 µl/min and 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 min to end up with a 2 min step at 95 % B before reequilibration with initial conditions. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 75 µl/min before entering the maXis 4G hr-ToF mass spectrometer (Bruker Daltonics, Germany) using the Apollo ESI source. Mass spectra were acquired in centroid mode ranging from 150 – 2500 m/z at 2 Hz scan rate.

Double Bond rearrangement

A retro-biosynthetic approach of the microsclerodermin side chain will initially result in a putative incorporation of a very uncommon C3 extender unit. This finding is owing to the double bond order which usually indicates the C2 extender units in between the double bonds. However, in microsclerodermin this does not hold true as the double bonds can rearrange in order to form a π-system conjugated to the phenyl ring. Taking this into account, the retro-biosynthetic proposal is based on a phenylacetate derived starter unit followed by C2 units, which is in agreement with conventional biosynthetic logic and the results from feeding experiments with labeled precursors. It remains elusive if this rearrangement is the result of sample extraction and workup or already takes place during/after biosynthesis *in vivo*.

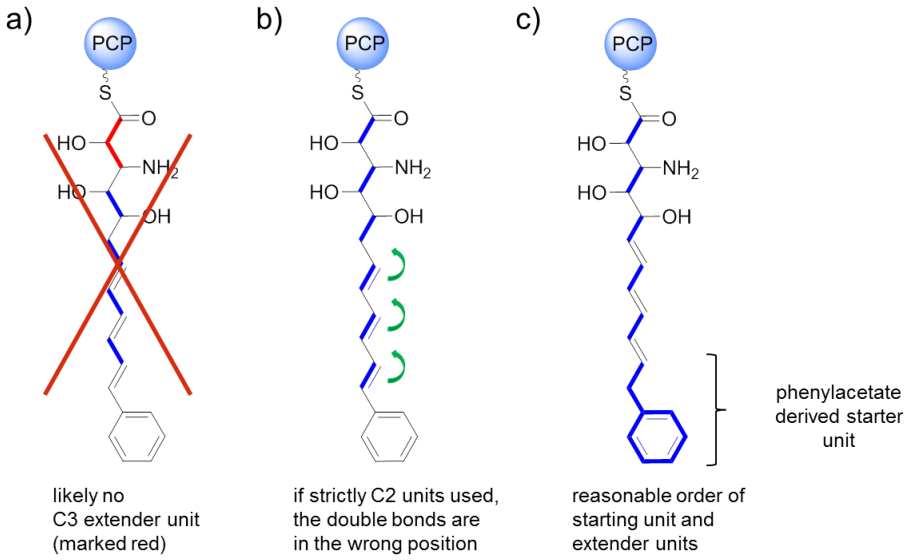


Figure S23: Retro-biosynthetic proposal for the microsclerodermin side chain. a) C3 extender unit is likely not incorporated during biosynthesis. b) A double bond rearrangement leads to C2 extender units. c) A reasonable biosynthetic proposal is starting with a phenylacetate starter followed by several C2 extender units.

Targeted inactivation of the *msc* locus in So ce38

Genetic modification of the So ce38 strain was accomplished according to a previously described protocol.[14] Putative single cross-over transformants were grown in H-Medium (0.2 % soybean meal, 0.2 % glucose, 0.8% starch, 0.2% yeast-extract, 0.1 % $CaCl_2 \cdot 2\ H_2O$, 0.1 % $MgSO_4 \cdot 7H_2O$, 8 mg/L FeEDTA, 50 mM HEPES, adjusted to pH 7.4 using 10N KOH) supplemented with hygromycin (100 µg/mL) and 1 % adsorber resin (XAD-16, Rohm & Haas) for at least 5 days at 180 rpm and 30 °C. Methanolic extracts of cells and the XAD resin were analyzed by LC-MS and compared with WT extracts. For isolation of chromosomal DNA the

mutant strain was grown in M-medium containing 100 µg/mL hygromycin. The integration of the plasmid at the right position in the genome was verified by PCR as shown in supporting Figure S24, using mutant and wild-type DNA as templates.

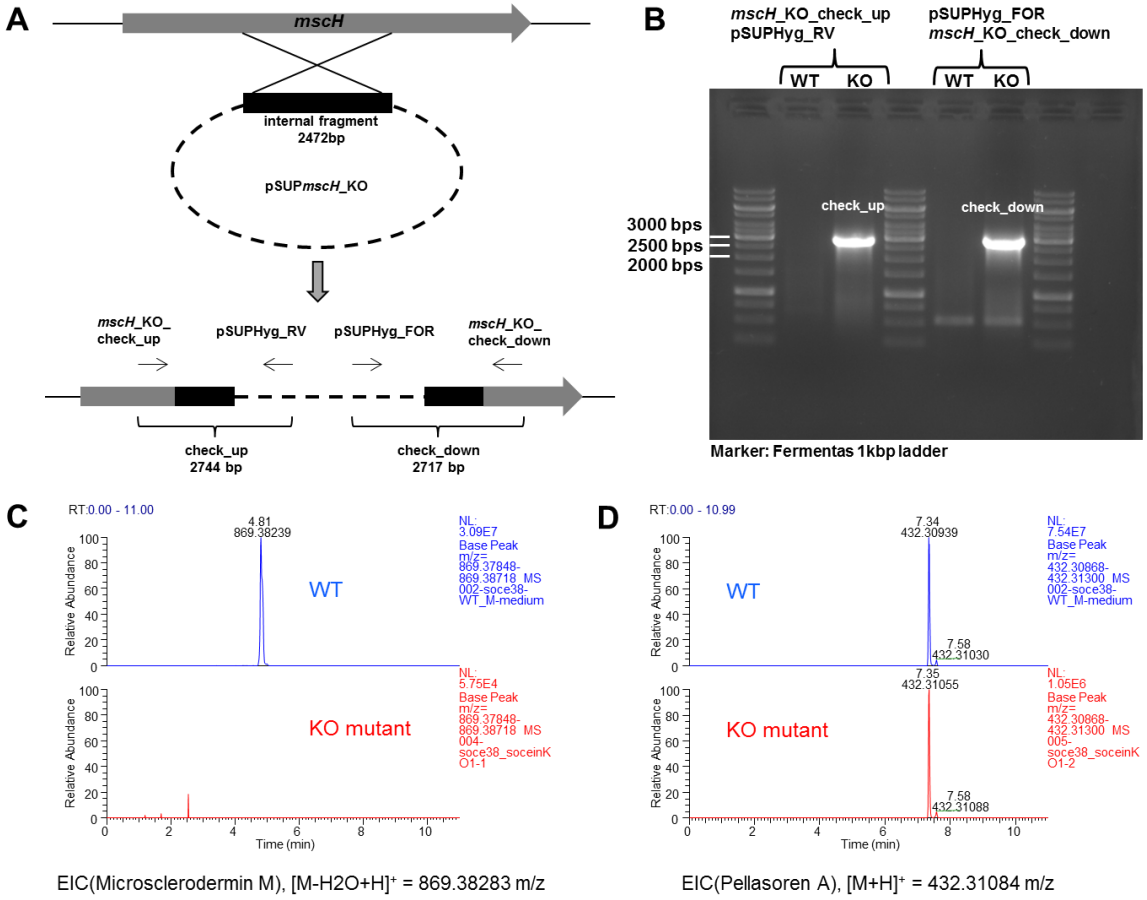


Figure S24: (A) Site of homologous recombination in So ce38. (B) Verification of KO-mutant So ce38::pSUP*mscH*_KO using PCR resulting in two amplicons "check_up" and "check_down". (C) Abolishment of microsclerodermin production as detected by comparative LC-MS analysis of wildtype and mutant strains. An extracted ion chromatogram for the most abundant $[M\text{-}H_2O+H]^+$ signal is shown. (D) Regular secondary metabolite production of the mutant strain was verified by identification of a known secondary metabolite in both wild-type and mutant strain.

Table S14: Primers used for amplification and verification of the microsclerodermin knock-out mutant in So ce38.

oligonucleotide	sequence
*mscH*_KO_for	GAT CCA GCG CTG GTT CCT CG
*mscH*_KO_rev	ACG AGG CTG TCG AAG AGC G
pSUPHyg_fwd	ATG TAG CAC CTG AAG TCA GCC
pSUPHyg_rev	ACG CAT ATA GCG CTA GCA GC
*mscH*_KO_check_up	ACA ACT TCT TCG CGC TCG G
*mscH*_KO_check_down	TCG TCG TAC GAG AGC CGG

Feeding experiments using labeled precursors

Feeding in *S. cellulosum* So ce38

$^{15}N_2$-$^{13}C_4$-L-asparagine
(completely incorporated)
$\Delta m_{theor.}$ = 2 x ^{15}N and 4 x ^{13}C
= (2 x 0.997035 Da) + (4 x 1.003355 Da)
= 6.00749 Da
$\Delta m/z_{meas.}$ = 6.0075

^{15}N-$^{13}C_9$-L-phenylalanine:
(eight carbons incorparated)
8 x ^{13}C
$\Delta m_{theor.}$ = 8 x ^{13}C
= 8 x 1.003355 Da
= 8.02684 Da
$\Delta m/z_{meas.}$ = 8.0267

$^{13}C_6$-L-phenylalanine (ring label)
(six carbons incorporated)
6 x ^{13}C
$\Delta m_{theor.}$ = 6 x ^{13}C
= 6 x 1.003355 Da
= 6.02013 Da
$\Delta m/z_{meas.}$ = 6.02023

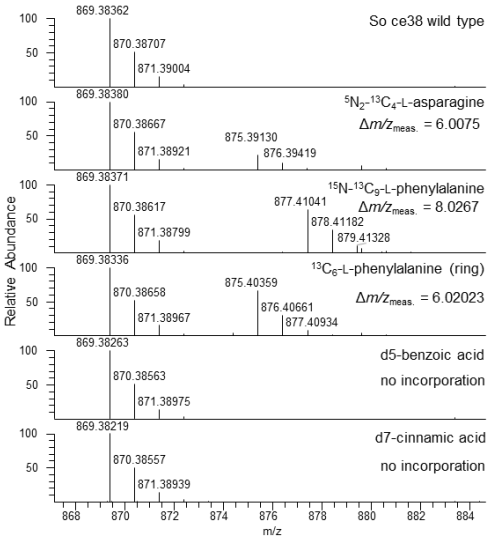


Figure S25: Isotopic peak pattern of the highly abundant [M-H2O+H]+ signal at 869.38283 *m/z* proves incorporation of labeled precursors into microsclerodermin M (So ce38). The observed mass shifts do perfectly fit to the heavy isotopes that were incorporated.

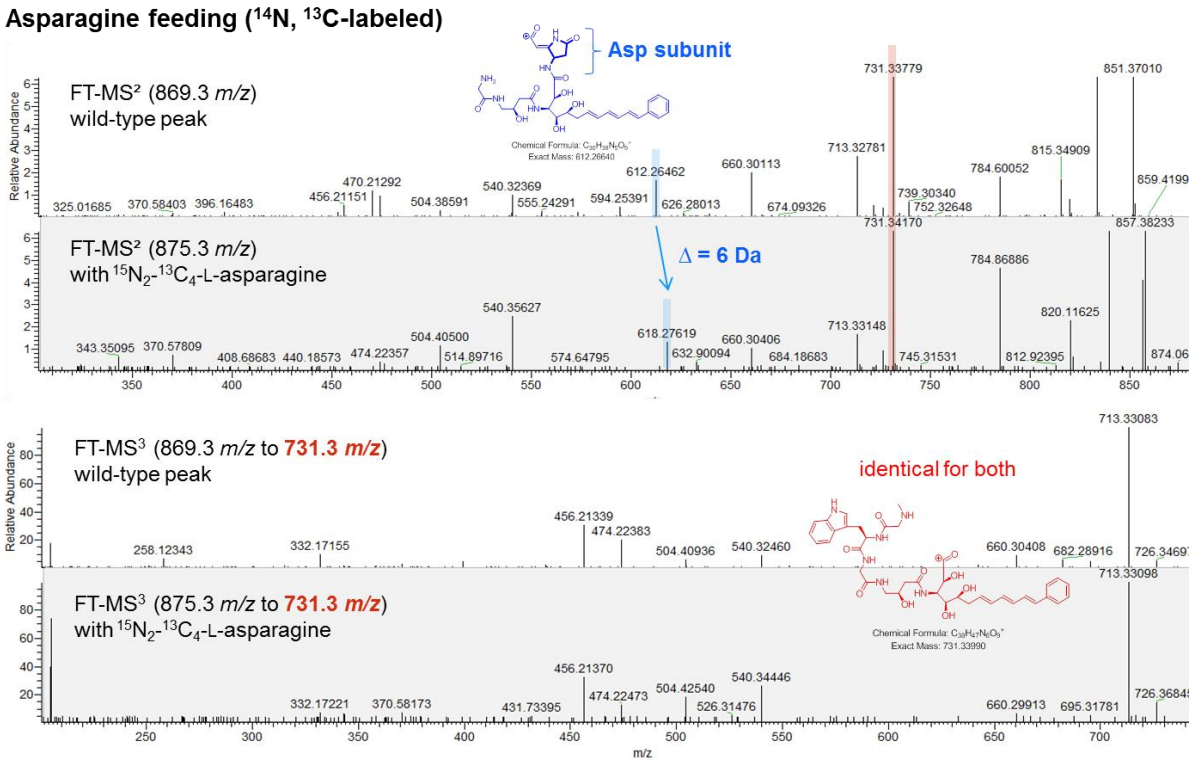


Figure S26: Incorporation of labeled L-asparagine subunit characterized by MS² and MS³ spectra acquisition. Data shows the feeding experiment in *S. cellulosum* So ce38. The most abundant fragment with 731.3 m/z is identical for the wild type peak and the enriched derivative (see MS³ data).

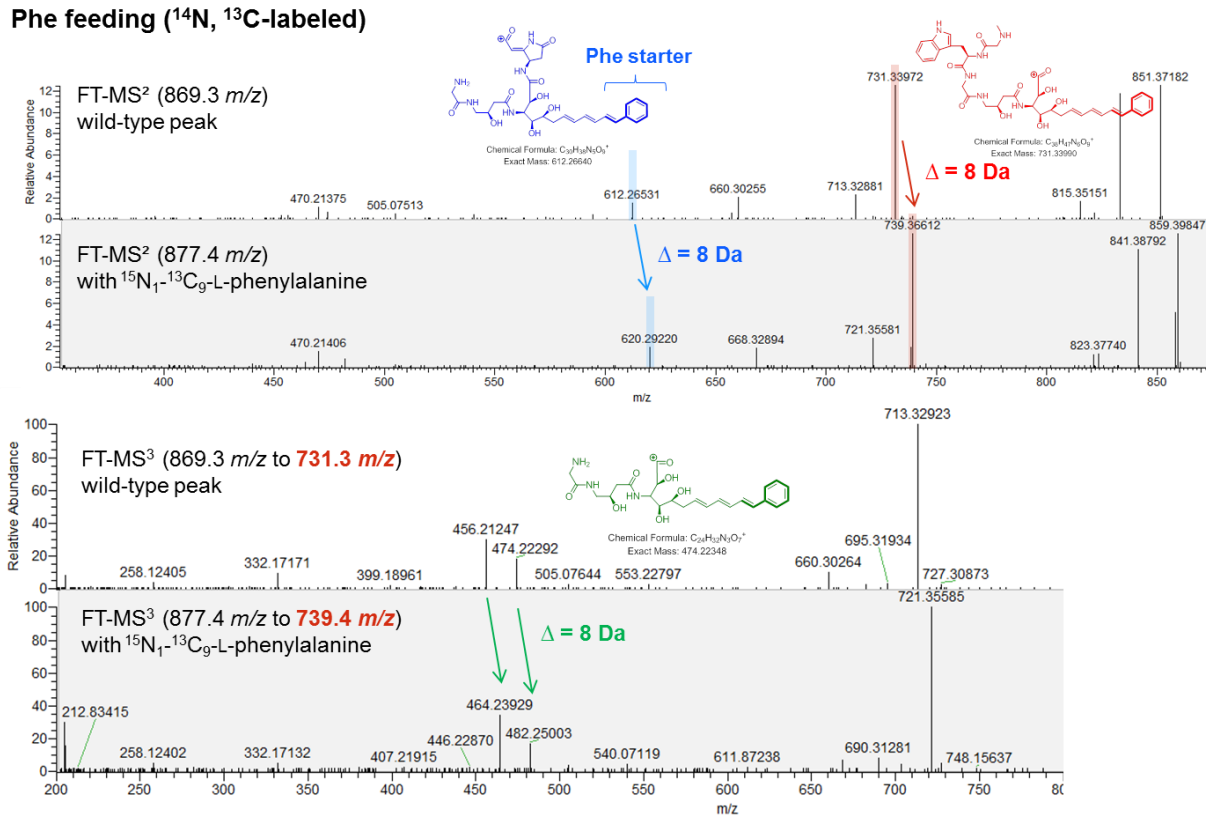


Figure S27: Incorporation of labeled phenylalanine subunit characterized by MS² and MS³ spectra acquisition. Data shows the feeding experiment in *S.cellulosum* So ce38. The mass difference of 8 Da is present in all the highlighted fragments indicating that phenylalanine is incorporated into the side chain.

Feeding in So ce38: Cultivation of So ce38 was performed in 25 mL H medium using 1 % (w/v) XAD-16. Labeled amino acids (5 mg) were dissolved in 250 mM HCl (320 µL) and sterile filtered. The labeled precursor stock (0.16 mg/mL) was added in three portions to the culture. 80 µL were added on first and third day followed by the remaining 160 µL on the fifth day of cultivation. Labeled benzoic acid and cinammic acid were dissolved in 80 µL DMSO and added in 2 x 20 µL and 40 µl portions to the respective culture. The cultures were harvested after 9 days and extracted with 2 x 25 mL methanol. The organic solvent was removed and the residuals dissolved in 500 µL methanol.

Feeding in *Jahnella* sp. MSr9139

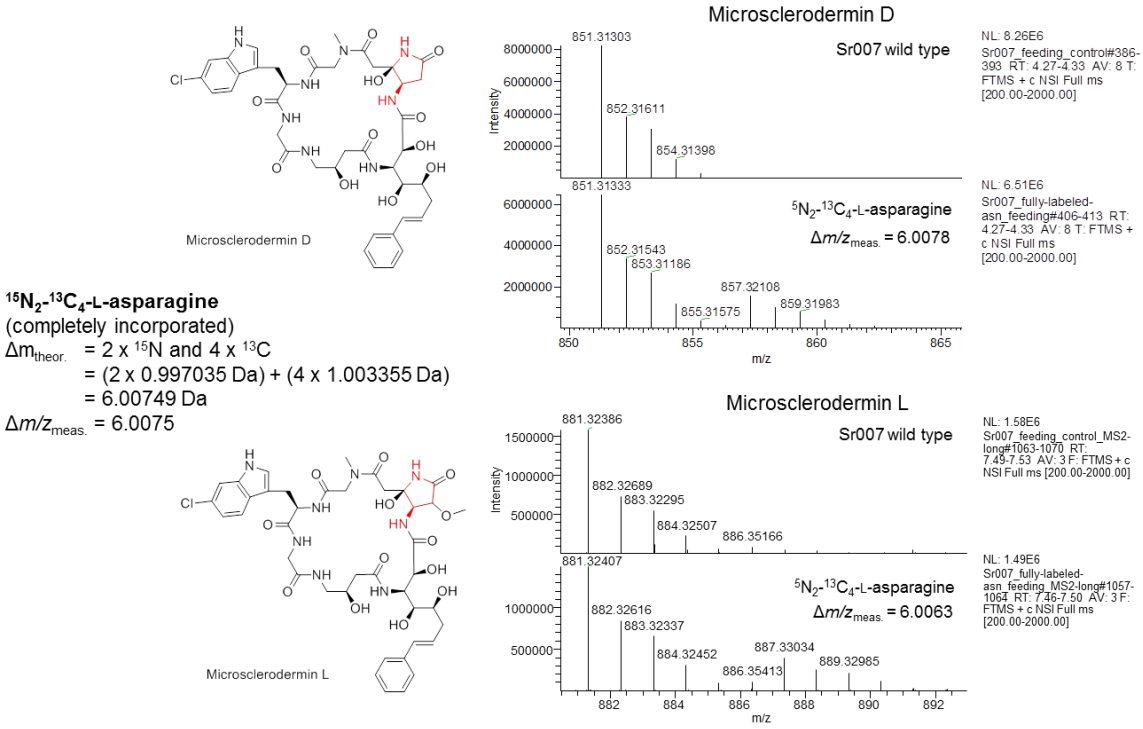


Figure S28: Isotopic peak pattern of the highly abundant $[M\text{-}H_2O+H]^+$ signals for microsclerodermin D (851.31303 *m/z*) and L (881.32312 *m/z*) prove incorporation of labeled precursor in *Jahnella* sp. MSr9139. The observed mass shifts do perfectly fit to the heavy isotopes that were incorporated.

LC-MS analysis of feeding experiments: Measurements were performed on a Dionex Ultimate 3000 RSLC system using a Waters BEH C18, 50 x 2.1 mm, 1.7 µm d_p column by injection of two µl methanolic sample. Separation was achieved by a linear gradient with (A) H_2O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 600 µL/min and 45 °C. The gradient was initiated by a 0.3 min isocratic step at 5 % B, followed by an increase to 95 % B in 9 min to end up with a 1 min flush step at 95 % B before reequilibration with initial conditions. UV and MS detection were performed simultaneously. Coupling the HPLC to the MS was supported by an Advion Triversa Nanomate nano-ESI system attached to a Thermo Fisher Orbitrap. Mass spectra were acquired in centroid mode ranging from 200 – 2000 *m/z* at a resolution of R = 30000. The measurements were repeated to obtain high resolution CID data in order to support the experiment with annotation of the obtained fragments

.

Bioactivity Testing

Bacterial Cultures

All microorganisms were handled under standard conditions recommended by the depositor. Overnight cultures of microorganisms were prepared in EBS medium (0.5 % peptone casein, 0.5 % proteose peptone, 0.1 % peptone meat, 0.1 % yeast extract; pH 7.0) or TSB medium (1.7 % peptone casein, 0.3 % peptone soymeal, 0.25 % glucose, 0.5 % NaCl, 0.25 % K_2HPO_4; pH 7.3). The latter medium was used for *E. faecalis* and *S. pneumonia* cultures. Yeast and fungi were grown in Myc medium (1% phytone peptone, 1 % glucose, 50 mM HEPES, pH 7.0).

Microbial Susceptibility Assay (MIC)

Overnight cultures of microorganisms were diluted to OD_{600} 0.01 (bacteria) or 0.05 (yeast/fungi) in the respective growth medium. Serial dilutions of compounds were prepared as duplicates in sterile 96-well plates. The cell suspension was added and microorganisms were grown overnight on a microplate shaker (750 rpm, 30 °C or 37 °C). Growth inhibition was assessed by measuring the OD_{600} on a plate reader. MIC_{50} values were determined as average relative to respective control samples by sigmoidal curve fitting.

References

(1) Delcher, A. L.; Bratke, K. A.; Powers, E. C.; Salzberg, S. L. *Bioinformatics* 2007, *23*, 673–9.
(2) Medema, M. H.; Blin, K.; Cimermancic, P.; de Jager, V.; Zakrzewski, P.; Fischbach, M. A.; Weber, T.; Takano, E.; Breitling, R. *Nucleic Acids Res.* 2011, *39*, W339–46.
(3) Yadav, G.; Gokhale, R. S.; Mohanty, D. *Journal of Molecular Biology* 2003, *328*, 335–363.
(4) Ziemert, N.; Podell, S.; Penn, K.; Badger, J. H.; Allen, E.; Jensen, P. R. *PLoS ONE* 2012, *7*, e34064.
(5) Caffrey, P. *ChemBioChem* 2003, *4*, 654–7.
(6) Kwan, D. H.; Tosin, M.; Schläger, N.; Schulz, F.; Leadlay, P. F. *Org. Biomol. Chem.* 2011, *9*, 2053–6.
(7) Röttig, M.; Medema, M. H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. *Nucleic Acids Res.* 2011, *39*, W362–7.
(8) Kozbial, P. Z.; Mushegian, A. R. *BMC Struct. Biol.* 2005, *5*, 19.
(9) Schmidt, E. W.; John Faulkner, D. *Tetrahedron* 1998, *54*, 3043–3056.
(10) Qureshi, A.; Colin, P. L.; Faulkner, D. J. *Tetrahedron* 2000, *56*, 3679–3685.
(11) Kunze, B.; Böhlendorf, B.; Reichenbach, H.; Höfle, G. *J. Antibiot.* 2008, *61*, 18–26.
(12) Bewley, C. A.; Debitus, C.; Faulkner, D. J. *J. Am. Chem. Soc.* 1994, *116*, 7631–7636.
(13) Fujii, K.; Ikai, Y.; Oka, H.; Suzuki, M.; Harada, K. *Analytical Chemistry* 1997, *69*, 5146–5151.
(14) Kopp, M.; Irschik, H.; Gross, F.; Perlova, O.; Sandmann, A.; Gerth, K.; Müller, R. *J. Biotechnol.* 2004, *107, 29-40.*

NMR spectra collection

The NMR spectra collection is not included in this thesis but availible online at DOI: http://dx.doi.org/10.1021/ja4054509

Chapter III

Studies towards enhanced production of ambruticin and derivatives thereof via genetic engineering

Introduction

The 2012 Revision of the World Population Prospects conducted by the Department of Economic and Social Affairs of the United Nations estimates that the number of people worldwide will reach 8.1 billion in the year 2025 and further rise up to 9.6 billion in 2050 [1]. Accordingly, the demand for aliments is increasing and this is a huge challenge, especially in times of climate change, which agriculture is facing. In order to successfully tackle this problem, farmers are constantly urged to deliver consistent quality and to improve the yields of their products while restricting crop failure to a minimum. Besides sufficient soil irrigation together with manuring, the protection of the crops against weeds, pests and diseases is a key factor for successful industrial farming. The significance of crop protection products in agriculture is reflected by its market volume which reached a total of € 7,683 million at the manufacturer level in Europe (EU 27 and EFTA nations) in 2011 [2]. Herbicides are the most applied pesticides closely followed by fungicides and the comparatively rarely used insecticides. In the year 2012 over 7,700 tons of herbicides, 6,100 tons of fungicides and 440 tons of insecticides were applied to crops of any kind in the UK (Department for Environment, Food & Rural Affairs, UK; Pesticide Usage Statistics).

The majority of fungicides used in our days are acting on vital functions of the fungus. The imidazoles and pyrimidines for example are inhibiting sterol biosynthesis at the stage of C_{14}-demethylation [3] whereas the strobilurins, likewise to the myxothiazols, are blocking the electron transport within the cytochrome bc_1 complex by binding to the Q_o-site [4-6]. Contrary to these “direct-targets”, molecules like the phenylpyrroles fenpiclonil and fludioxonil, both deduced from the natural product pyrrolnitrin [7], interfere with signal-transduction cascades of the fungus. These compounds target the high-osmolarity glycerol (HOG1)-pathway in filamentous fungi by interfering with a specific group III histidine kinase as shown for *Alternaria brassicicola* [8], *Pyricularia oryzae* [9], and *Neurospora crassa* [10]. Furthermore, the heterologous expression of respective kinase genes in the naturally phenylpyrrol resistant yeast *Saccharomyces cerevisiae* confers sensitivity against such fungicides [11;12]. In particular, the activation of the HOG1-pathway results in the intracellular accumulation of high concentrations of glycerol, triacylglycerides and free fatty acids what leads to cell death in the absence of high external osmolarity [13;14].

The myxobacterial secondary metabolite ambruticin and its derivatives exhibit strong antifungal activity against various pathogenic fungi e.g. *Histoplasma capsulatum*, *Coccidioides immitis* and *Blastomyces dermatitidis* [15] by exerting a similar effect to the phenylpyrroles on the HOG1-pathway [11;13]. Ambruticin was first isolated in 1977 from an ethyl acetate extract of

Polyangium cellulosum var. *firlvum* ATCC 25532 and it is the first antibiotic from myxobacteria for which a chemical structure could be presented [15]. Besides structure elucidation and bioactivity testings, efforts to reveal the biosynthetic basis for production of the ambruticins have been undertaken [16]. In the course of these studies the ambruticin gene cluster could be identified and a detailed biosynthetic model was presented. Accordingly, the ambruticins are build up by a PKS assembly line comprising 9 modules and employing several unusual post-PKS modification steps to generate at least 6 known derivatives (F, S, VS1, VS3, VS4 and VS5). Besides its biosynthesis, nothing is known so far about the regulatory mechanisms directing compound formation.

Interestingly, the ambruticins share high similarity with the jerangolids (A and B) which also exhibit fungicidal activity and were isolated as well from a myxobacterium belonging to the genus *Sorangium* [17]. In particular, the eastern part of both molecules is identical, suggesting a related biosynthetic origin and this section to be crucial for biological activity. Indeed, the first 5 modules of both assembly lines are composed of almost identical domains and the genomic regions flanking the structural genes for compound formation look strikingly similar [16]. Notably, both the ambruticin and the jerangolid cluster encode a protein termed Amb7 and Jer7, respectively. These enzymes show very high homology to particular fungal group III histidine kinases involved in the HOG-1 pathway, which are assumed to be the target of ambruticin and other fungicides [18].

In contrast to ambruticin, which has never been made commercially available, fludioxonil was first marketed in France in the year 1993 as a cereal seed dressing [19] and it is still an ingredient in various fungicides e.g. CELEST® (Syngenta Agro GmbH), Arena C (Bayer CropScience AG), and Mozart® TR (BASF SE). This might be caused by the limited availability of ambruticin rather than by its limited activity. In a *S. cerevisiae* sensitivity model ambruticin displayed an even better minimal inhibitory concentration (MIC) than fludioxonil or vinclozolin [11]. Additionally, no significant adverse effects in animals have been reported to date [20]. As a consequence, ambruticin might be an attractive candidate for the development of a commercial fungicide. Although successful chemical synthesis procedures for ambruticin and derivatives have been reported [21-23] it is still difficult to produce these molecules via the synthetic route with commercially viable yields because of their complexity. One possibility to overcome this issue is a classical strain improvement of the natural producer towards higher production yields either by random mutagenesis (UV or X-ray irradiation and mutagenic chemicals) or by directed genetic manipulation. Even though genetic manipulation of myxobacteria is often challenging, this approach allows rational alterations to be made in the genome of the producer (if the DNA sequence is available) and makes excessive screening of randomly generated mutant libraries

dispensable. A promising strategy to boost metabolite yields is to enhance the expression of the biosynthetic genes by exchanging the native promoters against promoters exhibiting a strong and constitutive activity [24]. Using this strategy, the formation of the myxobacterial metabolites thuggacin and crocapeptin could be significantly increased [25;26]. As an alternative, the elucidation and subsequent modification of regulatory mechanisms controlling compound production holds great promise for strain optimization. Unfortunately and in contrast to many *Streptomycetes* [27], regulators of myxobacterial secondary metabolite pathways are frequently not encoded within or in close vicinity of the corresponding gene clusters. The method of so-called “promoter-fishing” can help to circumvent this issue and to identify the respective regulatory proteins. Here, biotinylated PCR products bearing the promoter regions of the structural genes of interest are used as a probe similar to a fishing hook. Probes and a cell-free lysate of the producer strain are coincubated where potential regulatory proteins should bind specifically. Proteins bound to these probes are subsequently recovered employing streptavidin coupled magnetic beads which can be easily isolated using a simple magnet. The bound proteins can finally be identified by MALDI-TOF-analysis. Using this method, two regulators (ChiR and NtcA) specifically interacting with a chivosazol gene cluster promoter region in *Sorangium cellulosum* So ce56 could be identified. ChiR turned out to be a direct positive regulator, consequently the overexpression of *chiR* in a meroploid mutant lead to a 5-fold overproduction of chivosazol [28] whereas NtcA was shown to regulate metabolite formation in a negative way [29].

In the context of this work it was intended to optimize the production-yield of an ambruticin synthesizing strain to obtain amounts of this compound allowing for the first time initial steps of commercial development. This should enable bioactivity-testings and should furthermore set the stage for an industrial production strain. In particular, our in-house database (Myxobase) [30] should be used to allow identification of the best ambruticin producers from our strain collection. Based on these findings the most suitable strain with regard to a stable production, susceptibility to antibiotics and growth in suspension should be selected, for which a genetic manipulation system had to be developed. Eventually, potential elements controlling ambruticin formation should be identified to allow rational strain optimization via genetic engineering.

Experimental procedures

Strains, Oligonucleotides and Plasmids

The following tables summarize the used or constructed strains, the used/generated plasmids and the oligonucleotides which were employed in this study.

Table 1: Myxobacterial strains used and generated in this study

Strain	Characteristics
Sorangium cellulosum strain #8405	wild type strain; obtained from the laboratory collection of the Department of Microbial Drugs, Helmholtz Centre for Infection Research, Braunschweig
Sorangium cellulosum strain #8405::pMycoMAR_hyg	*Sorangium cellulosum* strain #8405 with randomly integrated transposable element of pMycoMAR_hyg (*hpt*$^+$) →generated while establishing electroporation protocol
Sorangium cellulosum strain #8405::pSM_lacZ_hyg_ntcA_up+dn1	*Sorangium cellulosum* strain #8405 with pSM_lacZ_hyg_ntcA_up+dn1 specifically integrated into the genome via the down homology region. (*hpt*$^+$, *lacZ*$^+$) →purpose: in-frame deletion of *ntcA* gene
Sorangium cellulosum strain #8405::pSM_lacZ_hyg_sacB_ntcA_up+dn1 clones 1 and 2	*Sorangium cellulosum* strain #8405 with pSM_lacZ_hyg_sacB_ntcA_up+dn1 specifically integrated into the genome via the up homology region. (*hpt*$^+$, *lacZ*$^+$, *sacB*$^+$) →purpose: in-frame deletion of *ntcA* gene
Sorangium cellulosum strain #8405::pSM_lacZ_hyg_sacB_ntcA_up+dn1 clones 3 and 6	*Sorangium cellulosum* strain #8405 with pSM_lacZ_hyg_sacB_ntcA_up+dn1 specifically integrated into the genome via the dn homology region. (*hpt*$^+$, *lacZ*$^+$, *sacB*$^+$) →purpose: in-frame deletion of *ntcA* gene
Sorangium cellulosum strain #8405 ΔntcA::lacZ,hpt clones 3,5 and 8	*Sorangium cellulosum* strain #8405 with replacement of *ntcA* against 7924 bp *Hind*III/*Nsi*I fragment of pSM_lacZ_hyg _ntcA_up+dn2 (*hpt*$^+$, *lacZ*$^+$) → 918bp deletion of *ntcA*
Sorangium cellulosum strain #8405 Δamb01::lacZ,hpt clones L1_1 and L1_3	*Sorangium cellulosum* strain #8405 Δamb01::lacZ,hpt with 7620 bp *Hind*III/*Nsi*I fragment of pSM_lacZ_hyg_amb01_KO_up1+dn1 integrated into the genome (exact place not defined) (*hpt*$^+$, *lacZ*$^+$) →incorrect Δ*amb01* mutants
Sorangium cellulosum strain #8405::pSM_lacZ_hyg_T7A1_amb01 clones 1,2,3,9 and10	*Sorangium cellulosum* strain #8405 with pSM_lacZ_hyg_T7A1_amb01 specifically integrated into the genome (*hpt*$^+$, *lacZ*$^+$) →promoter exchange in front of *amb01*
Sorangium cellulosum strain #8405::pMycoMAR_tet clones K1-12	*Sorangium cellulosum* strain #8405 with randomly integrated transposable element of pMycoMAR_tet (*tetA*$^+$) →generated while establishing *tetA* as novel selection marker for this strain

Strain	Characteristics
Sorangium cellulosum strain #8405 Δamb07::gfp,tetA clones L7_7, L7_8, L7_9, L7_b and L7_d	*Sorangium cellulosum* strain #8405 with replacement of *amb07* against 6151 bp *Xba*I/*Nsi*I fragment of pSM_gfp_tet_amb07_KO_up1+dn$_{neu}$ (*tetA*$^+$, *gfp*$^+$) →only L7_9 harbors correct 6206 bp deletion of *amb07*
Sorangium cellulosum strain #8405 Δamb04::lacZ,hpt clones L4_1-6	*Sorangium cellulosum* strain #8405 Δamb04::lacZ,hpt with 7827 bp *Hind*III/*Nsi*I fragment of pSM_lacZ_hyg_amb04_KO_up+dn1 integrated into the genome (*hpt*$^+$, *lacZ*$^+$) →only the strains L4_2,3,5 and 6 harbor the correct 1308 bp deletion of *amb04*
Sorangium cellulosum strain #8405::pSM_lacZ_hyg_T7A1_amb03-04 clones 1-8, 10,12,13 and 16	*Sorangium cellulosum* strain #8405 with pSM_lacZ_hyg_T7A1_amb03-04 specifically integrated into the genome (*hpt*$^+$, *lacZ*$^+$) →promoter exchange in front of *amb03-04* operon
Sorangium cellulosum strain #8405 Δamb04::lacZ,hpt::pSM_gfp_tet_partlylacZ_T7A1_amb04 clones 1,2,4,5 and 6	*Sorangium cellulosum* strain #8405 strain L4_3 with pSM_gfp_tet_partlylacZ_T7A1_amb04 specifically integrated via *lacZ* into the genome (*hpt*$^+$, *tetA*$^+$, *gfp*$^+$ and *lacZ*$^-$) → attempt to complement Δ*amb04* phenotype
Sorangium cellulosum strain #8405 Δamb04::lacZ,hpt::pSM_gfp_tet_partlylacZ_aphII_amb04 clones 1,2 and 3	*Sorangium cellulosum* strain #8405 strain L4_3 with pSM_gfp_tet_partlylacZ_aphII_amb04 specifically integrated via *lacZ* into the genome (*hpt*$^+$, *tetA*$^+$, *gfp*$^+$ and *lacZ*$^-$) → attempt to complement Δ*amb04* phenotype
Sorangium cellulosum strain #8405::pSM_lacZ_hyg_T7A1_ambA clones 2 and 3	*Sorangium cellulosum* strain #8405 with pSM_lacZ_hyg_T7A1_ambA specifically integrated into the genome (*hpt*$^+$, *lacZ*$^+$) →promoter exchange in front of *ambA*

Table 2: Oligonucleotides used in this study

Name	Sequence 5'→3'	Target genes/purpose
ntcA_up_HindIII/for	TGT AAG CTT GCC GTA GAT CGG GAT CGG	amplification of 1741 bp up region of *ntcA* for integration into pSM_lacZ_hyg; amplification of Southern Blot probe in ntcA_KO experiments
ntcA_up-XbaI/rev	TAT TCT AGA CTG CAC GGG GGA GCC AGT TC	amplification of 1741 bp up region of *ntcA* for integration into pSM_lacZ_hyg; amplification of Southern Blot probe in ntcA_KO experiments
ntcA_dn1_XbaI/for	ATA TCT AGA ATC TGA GCT CGC TCG CCC G	amplification of 1730 bp dn1 region of *ntcA* for integration into pSM_lacZ_hyg_ntcA_up
ntcA_dn1-NotI/rev	TAT GCG GCC GCC GCA AAA GCA GGG CGT GGC	amplification of 1730 bp dn1 region of *ntcA* for integration into pSM_lacZ_hyg_ntcA_up
ntcA_dn2-BamHI/for	ATA GGA TCC ATC TGA GCT CGC TCG CCC G	amplification of 1626 bp dn2 region of *ntcA* for integration into pSM_lacZ_hyg_ntcA_up
ntcA_dn2-NsiI/rev	TTT ATG CAT CCG CTG GAG GCG GAG AAG G	amplification of 1626 bp dn2 region of *ntcA* for integration into pSM_lacZ_hyg_ntcA_up

Name	Sequence 5'→3'	Target genes/purpose
ntcA_wt_for	AAA GAG GCG CAT CCA ACG TCC	oligo binding 53 bp upstream of *ntcA;* for verification of mutants
ntcA_wt_rev	CGA AGC AGC GCT TGG AGC AGG	oligo binding 152 bp downstream of *ntcA;* for verification of mutants
lev_ScaI_for	CCG GAG TAC TAT CAA AAA GAG TAT TGA CTT AAA GTC TAA CC	amplification of *sacB* from psaccat for integration into pSM_lacZ_hyg_ntcA_up+dn1
lev_ScaI_rev	TAG CAG TAC TTC ATT TGT TAA CTG TTA ATT GTC CTT GTT CAA GG	amplification of *sacB* from psaccat for integration into pSM_lacZ_hyg_ntcA_up+dn1
amb01_KO_dn_for_XmaJI	ATC CCT AGG GGT CGA TGG ATA GCA CGG CG	amplification of 1506 bp dn1 fragment of region of *amb01* for integration into pSM_lacZ_hyg_amb01_KO_ up1; amplification of Southern Blot probe in amb01_KO experiments
amb01_KO_dn_rev_NsiI_1	CTT ATG CAT ATC TGG CAC CAG CAT GGC	amplification of 1506 bp dn1 fragment of region of *amb01* for integration into pSM_lacZ_hyg_amb01_KO_ up1; amplification of Southern Blot probe in amb01_KO experiments
amb01_KO_up_rev_XbaI	GAT TCT AGA CCT TCC TCG TCC TGA GCG C	amplification of 1585 bp up1 fragment of region of *amb01* for integration into pSM_lacZ_hyg
amb01_KO_up_for_HindIII_1	GAA AAG CTT CGA TCT TCA CGA GCT GCG	amplification of 1585 bp up1 fragment of region of *amb01* for integration into pSM_lacZ_hyg
T7A1_amb01_rev_XmaJI	GCC TAG GAT CAA AAA GAG TAT TGA CTT AAA GTC TAA CCT ATA GGA TAC TTA CAG CCA TCG AGA GGT GTA CAA GGA AAC AGC TAT GCG CCT GCA GTC GGT C	amplification of 1056 bp fragment of T7A1_amb01 for integration into pSM_lacZ_hyg
T7A1_amb01_for_NsiI	TGA TGC ATT CAG GAC GAG GAA GGT CAG	amplification of 1056 bp fragment of T7A1_amb01 for integration into pSM_lacZ_hyg
T7A1_amb01_check_dn	TAC GGA CCG AGC AGG ACG AGC	verification of *Sorangium cellulosum* strain #8405::pSM_lacZ_hyg_ T7A1_amb01
T7A1_amb01_check_up	CAGCAGATGCGGACGACGAAG	verification of *Sorangium cellulosum* strain #8405::pSM_lacZ_hyg_ T7A1_amb01
amb07_KO_up_rev_NotI_1	ATG CGG CCG CCC GGA GGG GCG CGC GGC C	amplification of 1768 bp up1 fragment region of *amb07* for integration into pSM_gfp_tet
amb07_KO_up_for_XbaI	ATA TCT AGA CGC AGC ACC ACG CCG AGC C	amplification of 1768 bp up1 fragment region of *amb07* for integration into pSM_gfp_tet

Name	Sequence 5'→3'	Target genes/purpose
amb07_KOdn_rev_NsiI_neu	TGC ATG CAT TGT CTC GAG CGT GTA GC	amplification of 1757 bp dn_{neu} fragment region of for integration into pSM_gfp_tet_amb07_KO_ up1; amplification of Southern Blot probe in amb07_KO experiments
amb07_KO_dn_for_XmaJI	ATC CTA GGG CTA ACG CGC GCG GCG GG	amplification of 1757 bp dn_{neu} fragment region of *amb07* for integration into pSM_gfp_tet_amb07_KO_ up1; amplification of Southern Blot probe in *amb07*_KO experiments
T7A1_amb07_rev_XmaJI	AAC CTA GGA TCA AAA AGA GTA TTG ACT TAA AGT CTA ACC TAT AGG ATA CTT ACA GCC ATC GAG AGG TGT ACA AGG AAA CAG CTG TTC TTT AGT GTG GCT CAC CTC	amplification of 1848 bp fragment of T7A1_amb07 for integration into pSM_lacZ_hyg
T7A1_amb07_for_NsiI_neu	ACA TGC ATA CCT TCG ACA CGT CGC GG	amplification of 1848 bp fragment of T7A1_amb07 for integration into pSM_lacZ_hyg
amb04_KO_up_for_HindIII	CGT AAG CTT CGA TCT CTT GAA GCT CGC C	amplification of 1659 bp up fragment region of *amb04* for integration into pSM_lacZ_hyg*;* amplification of Southern Blot probe in amb04_KO and complementation experiments
amb04_KO_up_rev_XbaI	GAC TCT AGA TCA TGG CTT CTC CTG CGG C	amplification of 1659 bp up fragment region of *amb04* for integration into pSM_lacZ_hyg*;* amplification of Southern Blot probe in amb04_KO and complementation experiments
amb04_KO_dn1_rev_NsiI_1	GAT ATG CAT CAA CGA CAT GGC GGA GTC G	amplification of 1653 bp dn1 fragment region of *amb04* for integration into pSM_lacZ_hyg_amb04_KO_ up*;* amplification of Southern Blot probe in amb04_KO complementation experiments
amb04_KO_dn_for_XmaJI_1	AGC CTA GGA TCG AGC GAC CCT GTA CCG	amplification of 1653 bp dn1 fragment region of *amb04* for integration into pSM_lacZ_hyg_amb04_KO_ up*;* amplification of Southern Blot probe in amb04_KO complementation experiments
T7A1_amb04_for_XmaJI	GAC CTA GGA TCA AAA AGA GTA TTG ACT TAA AGT CTA ACC TAT AGG ATA CTT ACA GCC ATC GAG AGG TGT ACA AGG AAA CAG CTA TGA GCG GTC GCG TCC TG	amplification of 1480 bp fragment of T7A1_amb04 for integration into pSM_gfp_tet_partlylacZ
T7A1_amb04_rev_NsiI	GCA TGC ATT CAC GGC TTC GGG TCG CTC	amplification of 1480 bp fragment of T7A1_*amb04* for integration into pSM_gfp_tet_partlylacZ

Name	Sequence 5'→3'	Target genes/purpose
T7A1_amb03-04_for_XmaJI	GAC CTA GGA TCA AAA AGA GTA TTG ACT TAA AGT CTA ACC TAT AGG ATA CTT ACA GCC ATC GAG AGG TGT ACA AGG AAA CAG CTA CGA CGC CGG GCA ATG AAG	amplification of 1745 bp fragment of T7A1_amb03-04 for integration into pSM_lacZ_ hyg
T7A1_amb03-04_rev_NsiI	AGA TGC ATT CTC CAG GTC GCG GTT CG	amplification of 1745 bp fragment of T7A1_amb03-04 for integration into pSM_lacZ_ hyg
T7A1_amb03-4_check_up_for	GAC GTT GAG CTA CGC CGG	verification of *Sorangium cellulosum* strain #8405::pSM_lacZ_ hyg _T7A1_amb03-04
T7A1_amb03-4_check_dn_rv	TTC TGG AGC TCG CGG ACG	verification of *Sorangium cellulosum* strain #8405::pSM_lacZ_ hyg _T7A1_amb03-04
pBlues_FOR2	CAC CGA TCC GGA GGA ACT GG	verification oligonucleotide binding at backbone of pSM_lacZ_ hyg
pBlues_RV2	GAG CAG ACA AGC CCG TCA GG	verification oligonucleotide binding at backbone of pSM_lacZ_ hyg and pSM_gfp_tet; sequencing
pBlues_FOR	CTC ACT CAT TAG GCA CCC CAG G	verification oligonucleotide binding at backbone of pSM_lacZ_hyg and pSM_gfp_tet; sequencing
pBlues_RV	CGG CCA GTG AAT CCG TAA TCA TGG	verification oligonucleotide binding at backbone of pSM_lacZ_hyg and pSM_gfp_tet; sequencing
southern_blot_Tet_for	GGA GGC AGA CAA GGT ATA GG	ampification of *tetA* Southern Blot probe (1059 bp)
southern_blot_Tet_rev	GCC TCT TGC GGG ATA TCG TC	ampification of *tetA* Southern Blot probe (1059 bp)
southern-blot-hyg-fwd	GAA CTG CGC CAG TTC CTC C	ampification of *hpt* Southern Blot probe (1021 bp)
southern-blot-hyg-rev	GAT CTG ATG GCG CAG GGG	ampification of *hpt* Southern Blot probe (1021 bp)
pblueHyg_For_PmeI	AAT GTT TAA ACA TGA CAC AAG AAT CCC TGT TAC TTC TCG	amplification of 1021 bp fragment of *hpt* gene for exchange of *cat* against *hpt* in pBluecat
pblueHyg_RV_AanI:	ATT AAT TAT AAT CAG GCG CCG GGG GCG G	amplification of 1021 bp fragment of *hpt* gene for exchange of *cat* against *hpt* in pBluecat
NotI-T7A1lacZRBS-megfp	TGC GGC CGC ATC AAA AAG AGT ATT GAC TTA AAG TCT AAC CTA TAG GAT ACT TAC AGC CAT CGA GAG GTG TAC AAG GAA ACA GCT ATG GTG AGC AAG GGC G	amplification of 816 bp fragment of *megfp* for exchange of *lacZ* in pSM_lacZ_hyg against *megfp*
megfp_rev_NdeI	GTA AGG CAT ATG TCA CTT GTA CAG CTC GTC C	amplification of 816 bp fragment of *megfp* for exchange of *lacZ* in pSM_lacZ_hyg against *megfp*

Name	Sequence 5'→3'	Target genes/purpose
Red/ET_tet_for	GGC GCC GGG GGC GGT GTC CGG CGG CCC CCA GAG GAA CTG CGC CAG TTG TTT AAA CGC GAG GTG CCG CCG GCT TC	amplification of *tetA* gene flanked by *Pme*I restriction sites and a region homologous to pMycoMAR_hyg (1434 bp)
Red/ET_tet_rev	TGG CGC AGG GGA TCA AGA TCT GAT CAA GAG ACA GGA TGA GGA TCG TTT GTT TAA ACT CAG CCC CAT ACG ATA TAA GTT GTA ATT CTC ATG	amplification of *tetA* gene flanked by *Pme*I restriction sites and a region homologous to pMycoMAR_hyg (1434 bp)
partly_lacZ_rev_NotI	ATG CGG CCG CTG CCA GGC GCT GAT GTG	amplification of an internal part of *lacZ* (1907 bp) for integration into pSM_gfp_tet
partly_lacZ_for_XbaI	GAT CTA GAA ACT GGC AGA TGC ACG G	amplification of an internal part of *lacZ* (1907 bp) for integration into pSM_gfp_tet
aphII_for_XmaJI	TGC CTA GGC ACG CTG CCG CAA GCA C	amplification of 374 bp *aphII* promoter with 19 bp homology arm to *amb04*
aphII_rev_overlap_amb04	GAC CGC TCA TGG CTT CTC CAA ACG ATC CTC ATC CTG TCT C	amplification of 374 bp *aphII* promoter with 19 bp homology arm to *amb04*
overlap_aphII_amb04_for	GAG ACA GGA TGA GGA TCG TTT GGA GAA GCC ATG AGC GGT C	amplification of 1425 bp fragment of *amb04* including native RBS with 21 bp overlap to the *aphII* promoter
amb04_rev_NsiI	TCG ATG CAT TCA CGG CTT CGG GTC G	amplification of 1425 bp fragment of *amb04* including native RBS with 21 bp overlap to the *aphII* promoter
T7A1_ambA_for_XbaI	ATC TCT AGA ATC AAA AAG AGT ATT GAC TTA AAG TCT AAC CTA TAG GAT ACT TAC AGC CAT CGA GAG GTG TAC AAG GAA ACA GCT ATG GGG CTA CGG GAT TCA GAT GC	amplification of 1581 bp fragment of T7A1_*ambA* for integration into pSM_lacZ_hyg
T7A1_ambA_rev_NotI	AGT GCG GCC GCG TGC TCC GTG AGC CAC	amplification of 1581 bp fragment of T7A1_*ambA* for integration into pSM_lacZ_hyg
T7A1_ambA_check_up_for	CTG TGT CAA GGT GCG CCG	verification of *Sorangium cellulosum* strain #8405::pSM_lacZ_hyg_T7A1_ambA
T7A1_ambA_check_dn_rev	GAA GAC CAC CTT GCC CCG C	verification of *Sorangium cellulosum* strain #8405::pSM_lacZ_hyg_T7A1_ambA
Ambruticin_for	GCC GGT CAT CGT CAT CAC C	amplification of 1038 bp fragment of *amb04* for verification of strain #8405 strain L4_3::pSM_gfp_tet_partlylacZ_T7A1_amb04 and L4_3::pSM_gfp_tet_partlylacZ_aphII_amb04

Name	Sequence 5'→3'	Target genes/purpose
Ambruticin_rev	AGG ACG CGC AGG ATG TAG C	amplification of 1038 bp fragment of *amb04* for verification of strain #8405 strain L4_3::pSM_gfp_tet_partlylacZ _T7A1_amb04 and L4_3::pSM_gfp_tet_partlylacZ _aphII_amb04
16s_Soce10_for	CGT AAA GCG CAT GTA GGC GG	amplification of 196 bp part of 16 s gene for q-real-time PCR
16s_Soce10_rev	TTC GCA TCT CAG CGT CAG TC.	amplification of 196 bp part of 16 s gene for q-real-time PCR
RT_amb04_for	TGG ACG ACT CCC ACC GTT AC	amplification of 198 bp part of *amb04* gene for q-real-time PCR
RT_amb04_rev	AGT TCA CCG CGA CGA ACG	amplification of 198 bp part of *amb04* gene for q-real-time PCR
RT_amb04a_for	TCT CGT CCA GAC CGG AGC AG	amplification of 214 bp part of *amb04a* gene for q-real-time PCR
RT_amb04a_rev	TCC TGG GGA GAG AGG CTG TC	amplification of 214 bp part of *amb04a* gene for q-real-time PCR
amb04_for_BamHI	GAG GGA TCC ATG AGC GGT CGC GTC CTG	amplification of *amb04* (1386 bp) for cloning it into pSUMO3_ck4
amb04_rev_KpnI	CCG GTA CCT CAC GGC TTC GGG TC	amplification of *amb04* (1386 bp) for cloning it into pSUMO3_ck4
T7prom	TAA TAC GAC TCA CTA TAG G	sequencing of pSUMO3_ck4_amb04
T7term	GCT AGT TAT TGC TCA GCG G	sequencing of pSUMO3_ck4_amb04
Seq_Tet_out_down 2	TTA TGA TTC TTC TCG CTT CC	verification oligonucleotide binding at backbone of pSM_gfp_tet; sequencing
EMSA_sec_HEX2	HEX-GCT CTC GGT GCC CTT GTG	amplification of HEX-labeled DNA fragments for EMSA
EMSA_sec_HEX3	HEX-GGA ATG GGC CGG GTA CTG	amplification of HEX-labeled DNA fragments for EMSA
amb3hex3F	GGA ATG GGC CGG GTA CTG CCG AAG AGA CGT TGA GCT ACG C	amplification of 355 bp *amb3* EMSA fragment
amb3hex2R	GCT CTC GGT GCC CTT GTG ATG GCG TCT CTG CGG ATC TG	amplification of 355 bp *amb3* EMSA fragment
amb4ahex3F	GGA ATG GGC CGG GTA CTG CGA CCC GAA GCC GTG AC	amplification of 253 bp *amb4* EMSA fragment
amb4ahex2R	GCT CTC GGT GCC CTT GTG CCG TGG TGT GGC TCA TGT TC	amplification of 253 bp *amb4* EMSA fragment
ambAhex3F	GGA ATG GGC CGG GTA CTG ATC GGA TGC AAC GTG ATC GAG	amplification of 708 bp *ambA* EMSA fragment

Name	Sequence 5'→3'	Target genes/purpose
ambAhex2R	GCT CTC GGT GCC CTT GTG AAG CGA CAT CCC ATT CCC AC	amplification of 708 bp *ambA* EMSA fragment
ambGhex3F	GGA ATG GGC CGG GTA CTG TCT TCC GGT CAG TGC TCA ATC C	amplification of 299 bp *ambG* EMSA fragment
ambGhex2R	GCT CTC GGT GCC CTT GTG CTT GTT CCT TCC CTG CTC ATG C	amplification of 299 bp *ambG* EMSA fragment
ambNQhex3F	GGA ATG GGC CGG GTA CTG ATG AGC GGC TCC TCG TTT GC	amplification of 253 bp *ambNQ* EMSA fragment
ambNQhex2R	GCT CTC GGT GCC CTT GTG CCT GTC AAG CCC AGG TGG TG	amplification of 253 bp *ambNQ* EMSA fragment
ambShex3F	GGA ATG GGC CGG GTA CTG ACC GAG GCT GAC GTG GAC TG	amplification of 281 bp *ambS* EMSA fragment
ambShex2R	GCT CTC GGT GCC CTT GTG CAT GAG CTG GCT CAG CTC GAC	amplification of 281 bp *ambS* EMSA fragment
amb7hex3F	GGA ATG GGC CGG GTA CTG CAT GTT GGA CCC TCT CCA GG	amplification of 306 bp *amb7* EMSA fragment
amb7hex2R	GCT CTC GGT GCC CTT GTG TAA TCA GAC GGA GCT GCG AGA G	amplification of 306 bp *amb7* EMSA fragment
amb5/6hex3F2	GGA ATG GGC CGG GTA CTG ACC ACG ATG GCC ACG CTC	amplification of 333 bp *amb5/6* EMSA fragment
amb5/6hex2R2	GCT CTC GGT GCC CTT GTG CGT CCA GAG GAC CGA GCT G	amplification of 333 bp *amb5/6* EMSA fragment
ambBhex3F	GGA ATG GGC CGG GTA CTG GGA ATG TTT TCT TCG AGC AAG GAC TC	amplification of 380 bp *ambB* EMSA fragment
ambBhex2R	GCT CTC GGT GCC CTT GTG AAA CCG GCA CGC CAT CG	amplification of 380 bp *ambB* EMSA fragment
ambHhex3F	GGA ATG GGC CGG GTA CTG CGG CTC TCG ACG TCG AGA G	amplification of 273 bp *ambH* EMSA fragment
ambHhex2R	GCT CTC GGT GCC CTT GTG ACG CCG ATG ATC GCG ATC	amplification of 273 bp *ambH* EMSA fragment
ambRhex3F	GGA ATG GGC CGG GTA CTG CGG TGT CCT GTT CCG CCT C	amplification of 333 bp *ambR* EMSA fragment
ambRhex2R	GCT CTC GGT GCC CTT GTG CAA CAA ACC TCC CCT GAA GTT CG	amplification of 333 bp *ambR* EMSA fragment
5'-P-3'-Cy5	TGG CGC GGA TGA-Cy5	Cy5-labeling of small EMSA fragments
amb3-1	GCG CCG CGC GCC ACC TGC GCT ACG ACG CCG GGC AAT GAA GCT CGC GCG CA	generation of 50 bp EMSA fragment with 15 bp overlap to amb3-2 fragment
amb3-1r	TCA TCC GCG CCA TGC GCG CGA GCT TCA TTG CCC GGC GTC GTA GCG CAG GTG GCG CGC GGC GC	generation of 50 bp EMSA fragment with 15 bp overlap to amb3-2 fragment; 12 bp complementary to 5'-P-3'-Cy5 for labeling

Name	Sequence 5'→3'	Target genes/purpose
amb3-2	GAC GGC GTC GGG CGG GCG GCT GGA CGC GCG CAC CCG CGC CGC GCG CCA CC	generation of 50 bp EMSA fragment with 15 bp overlap to amb3-3 and amb3-1 fragment
amb3-2r	TCA TCC GCG CCA GGT GGC GCG CGG CGC GGG TGC GCG CGT CCA GCC GCC CGC CCG ACG CCG TC	generation of 50 bp EMSA fragment with 15 bp overlap to amb3-3 and amb3-1 fragment; 12 bp complementary to 5'-P-3'-Cy5 for labeling
amb3-3	GCG GGT GCC GCA CAA CGC GGC ATG TCG CAT TTT GCG ACG GCG TCG GGC GG	generation of 50 bp EMSA fragment with 15 bp overlap to amb3-4 and amb3-2 fragment
amb3-3r	TCA TCC GCG CCA CCG CCC GAC GCC GTC GCA AAA TGC GAC ATG CCG CGT TGT GCG GCA CCC GC	generation of 50 bp EMSA fragment with 15 bp overlap to amb3-4 and amb3-2 fragment; 12 bp complementary to 5'-P-3'-Cy5 for labeling
amb3-4	CCG GGC GAG CGC GGC GCC GCG GCA GGA CGG CCG ACG CGG GTG CCG CAC AA	generation of 50 bp EMSA fragment with 15 bp overlap to amb3-5 and amb3-3 fragment
amb3-4r	TCA TCC GCG CCA TTG TGC GGC ACC CGC GTC GGC CGT CCT GCC GCG GCG CCG CGC TCG CCC GG	generation of 50 bp EMSA fragment with 15 bp overlap to amb3-5 and amb3-3 fragment; 12 bp complementary to 5'-P-3'-Cy5 for labeling
amb3-5	AGC TCG CTG CGC GTT CGC TTC CCC GCA GCC TGA AGC CGG GCG AGC GCG GC	generation of 50 bp EMSA fragment with 15 bp overlap to amb3-4 fragment
amb3-5r	TCA TCC GCG CCA GCC GCG CTC GCC CGG CTT CAG GCT GCG GGG AAG CGA ACG CGC AGC GAG CT	generation of 50 bp EMSA fragment with 15 bp overlap to amb3-4 fragment; 12 bp complementary to 5'-P-3'-Cy5 for labeling
amb4a-1	TGG CAT GAG AGC TGC CTG AAA GCC GGG CGC GAA CAT GAG CCA CAC CAC GG	generation of 50 bp EMSA fragment with 15 bp overlap to amb4a-2 fragment
amb4a-1r	TCA TCC GCG CCA CCG TGG TGT GGC TCA TG TTC GCG CCC GGC TTT CAG GCA GCT CTC ATG CCA	generation of 50 bp EMSA fragment with 15 bp overlap to amb4a-2 fragment; 12 bp complementary to 5'-P-3'-Cy5 for labeling
amb4a-2	CGC GGG CGC GCG CGG GCG CCC GCG ACT CTG CCT CGT GGC ATG AGA GCT GC	generation of 50 bp EMSA fragment with 15 bp overlap to amb4a-3 and amb4a-1 fragment
amb4a-2r	TCA TCC GCG CCA GCA GCT CTC ATG CCA CGA GGC AGA GTC GCG GGC GCC CGC GCG CGC CCG CG	generation of 50 bp EMSA fragment with 15 bp overlap to amb4a-3 and amb4a-1 fragment; 12 bp complementary to 5'-P-3'-Cy5 for labeling
amb4a-3	GAC ACC CGC GCC GCC GCG CGA TCG GCA GCG CCG CTC GCG GGC GCG CGC GG	generation of 50 bp EMSA fragment with 15 bp overlap to amb4a-4 and amb4a-2 fragment
amb4a-3r	TCA TCC GCG CCA CCG CGC GCG CCC GCG AGC GGC GCT GCC GAT CGC GCG GCG GCG CGG GTG TC	generation of 50 bp EMSA fragment with 15 bp overlap to amb4a-4 and amb4a-2 fragment; 12 bp complementary to 5'-P-3'-Cy5 for labeling

Name	Sequence 5'→3'	Target genes/purpose
amb4a-4	GGC GAC TCC GCT CGT CCC TCG CTG TTG CAG AAC GCG ACA CCC GCG CCG CC	generation of 50 bp EMSA fragment with 15 bp overlap to amb4a-5 and amb4a-3 fragment
amb4a-4r	TCA TCC GCG CCA GGC GGC GCG GGT GTC GCG TTC TGC AAC AGC GAG GGA CGA GCG GAG TCG CC	generation of 50 bp EMSA fragment with 15 bp overlap to amb4a-5 and amb4a-3 fragment; 12 bp complementary to 5'-P-3'-Cy5 for labeling
amb4a-5	TCG CGC CGG AGG TGA TGC CCG GAG AGC CTC GCG GCG GCG ACT CCG CTC GT	generation of 50 bp EMSA fragment with 15 bp overlap to amb4a-4 fragment
amb4a-5r	TCA TCC GCG CCA ACG AGC GGA GTC GCC GCC GCG AGG CTC TCC GGG CAT CAC CTC CGG CGC GA	generation of 50 bp EMSA fragment with 15 bp overlap to amb4a-4 fragment; 12 bp complementary to 5'-P-3'-Cy5 for labeling
amb33-1	GCG GGA ATT AAA CAA CGC GGC ATG TCG CAT TTT GCG ACG GCG TCG GGC GG	generation of 50 bp EMSA fragment of amb3-3 with altered repeat1
amb33-1R	TCA TCC GCG CCA CCG CCC GAC GCC GTC GCA AAA TGC GAC ATG CCG CGT TGT TTA ATT CCC GC	generation of 50 bp EMSA fragment of amb3-3 with altered repeat1; 12 bp complementary to 5'-P-3'-Cy5 for labeling
amb33-2	GCG GGT GCC GCA CAA CAA TTA ATG TCG CAT TTT GCG ACG GCG TCG GGC GG	generation of 50 bp EMSA fragment of amb3-3 with altered repeat1
amb33-2R	TCA TCC GCG CCA CCG CCC GAC GCC GTC GCA AAA TGC GAC ATT AAT TGT TGT GCG GCA CCC GC	generation of 50 bp EMSA fragment of amb3-3 with altered repeat1; 12 bp complementary to 5'-P-3'-Cy5 for labeling
amb33-3	GCG GGA ATT AAA CAA CAA TTA ATG TCG CAT TTT GCG ACG GCG TCG GGC GG	generation of 50 bp EMSA fragment of amb3-3 with altered repeat1
amb33-3R	TCA TCC GCG CCA CCG CCC GAC GCC GTC GCA AAA TGC GAC ATT AAT TGT TGT TTA ATT CCC GC	generation of 50 bp EMSA fragment of amb3-3 with altered repeat1; 12 bp complementary to 5'-P-3'-Cy5 for labeling
amb33-4	GCG GGT GCC GCA CAA CGC GGC AAA TTA AAT TTT GCG ACG GCG TCG GGC GG	generation of 50 bp EMSA fragment of amb3-3 with altered repeat2
amb33-4R	TCA TCC GCG CCA CCG CCC GAC GCC GTC GCA AAA TTT AAT TTG CCG CGT TGT GCG GCA CCC GC	generation of 50 bp EMSA fragment of amb3-3 with altered repeat2; 12 bp complementary to 5'-P-3'-Cy5 for labeling
amb33-5	GCG GGT GCC GCA CAA CGC GGC ATG TCG CAT TTT AAT TAA GCG TCG GGC GG	generation of 50 bp EMSA fragment of amb3-3 with altered repeat2
amb33-5R	TCA TCC GCG CCA CCG CCC GAC GCT TAA TTA AAA TGC GAC ATG CCG CGT TGT GCG GCA CCC GC	generation of 50 bp EMSA fragment of amb3-3 with altered repeat2; 12 bp complementary to 5'-P-3'-Cy5 for labeling
amb33-6	GCG GGT GCC GCA CAA CGC GGC AAA TTA AAT TTT AAT TAA GCG TCG GGC GG	generation of 50 bp EMSA fragment of amb3-3 with altered repeat2

Name	Sequence 5'→3'	Target genes/purpose
amb33-6R	TCA TCC GCG CCA CCG CCC GAC GCT TAA TTA AAA TTT AAT TTG CCG CGT TGT GCG GCA CCC GC	generation of 50 bp EMSA fragment of amb3-3 with altered repeat2; 12 bp complementary to 5'-P-3'-Cy5 for labeling
amb33-7	GCG GGA ATT AAA CAA CAA TTA ATG TCG CCA AGC GCG ACG GCG TCG GGC GG	generation of 50 bp EMSA fragment of amb3-3 with altered repeat1 and altered spacer sequence of repeat2
amb33-7R	TCA TCC GCG CCA CCG CCC GAC GCC GTC GCG CTT GGC GAC ATT AAT TGT TGT TTA ATT CCC GC	generation of 50 bp EMSA fragment of amb3-3 with altered repeat1 and altered spacer sequence of repeat2; 12 bp complementary to 5'-P-3'-Cy5 for labeling
amb44-1	GGC GAC TCC GCT CGT CCC TCG CAA TTA AAG AAC AAT TAA CCC GCG CCG CC	generation of 50 bp EMSA fragment of amb4-4 with altered repeat
amb44-1R	TCA TCC GCG CCA GGC GGC GCG GGT TAA TTG TTC TTT AAT TGC GAG GGA CGA GCG GAG TCG CC	generation of 50 bp EMSA fragment of amb4-4 with altered repeat; 12 bp complementary to 5'-P-3'-Cy5 for labeling
amb44-2	GGC GAC TCC GCT CGT CCC TCG CTG TTG CGA CCA GCG ACA CCC GCG CCG CC	generation of 50 bp EMSA fragment of amb4-4 with altered spacer sequence
amb44-2R	TCA TCC GCG CCA GGC GGC GCG GGT GTC GCT GGT CGC AAC AGC GAG GGA CGA GCG GAG TCG CC	generation of 50 bp EMSA fragment of amb4-4 with altered spacer sequence; 12 bp complementary to 5'-P-3'-Cy5 for labeling
ambB_forbtn	[Btn]-ATC GGA TGC AAC GTG ATC GAG	amplification of 676 bp fragment in front of *ambA* for DNA-protein pull-down experiments with 5' biotin label
ambB_rev2	GGG CAA GCG ACA TCC CAT TC	amplification of 676 bp fragment in front of *ambA* for DNA-protein pull-down and EMSA experiments
ambB_for_hex	GCT CTC GGT GCC CTT GTG ATC GGA TGC AAC GTG ATC GAG	amplification of 676 bp EMSA fragment in front of *ambA*
MXAN_3950_pro_rev	[Btn]-GCG GCC CAG GCA TTG CAA CT	amplification of P_{TaA} promoter fragment as negative control in DNA-protein pull-down experiments with 5' biotin label
MXAN_3951_pro_for	AAT GCG CCG GCT TCC GTT CA	amplification of P_{TaA} promoter fragment as negative control in DNA-protein pull-down experiments with 5' biotin label

Table 3: Plasmids used and generated in this study

Name	Genotype	Restriction sites for cloning purposes	Reference
pBluecat	*bla*, *cat*, *lacZ* under control of T7A1 promoter	progenitor of pSM_lacZ_hyg	Dr. Carsten Volz

Name	Genotype	Restriction sites for cloning purposes	Reference
pSaccat	*bla*, *cat*, *sacB* under control of T7A1 promoter	used to amplify T7A1_sacB	Dr. Carsten Volz
pMycoMAR_hyg	*aacC1*, *hpt*, *rep* (R6K), *rep* (p15a), omega-ter, P_{aphII} in front of hpt, *tnp*	progenitor of pMycoMAR_tet	[31]
pACYC184	*cat*, *tetA*, *rep* (ColE1)	used to amplify *tetA* for Red/ET	New England Biolabs
pPm1-mEGFP-rppA	*npt*, *xylS*, xylose inducible P_m promoter, *rep* (V), *megfp* fused to *rppA*	used to amplify *gfp*	Dr. Katja Gemperlein
pMycoMAR_tet	derivative of pMycoMAR_Hyg with *hpt* exchanged against *tetA*	cloned by using Red/ET technology	This study
pSM_lacZ_hyg	pBluecat derivative with *cat* exchanged against *hpt* from pMycoMAR_ hyg	*Pme*I/*Aan*I	This study
pSM_gfp_hyg	pSM_lacZ_hyg derivative with *lacZ* exchanged against *gfp*	*Not*I/*Nde*I	This study
pSM_gfp_tet	pSM_gfp_hyg derivative with *hpt* exchanged against *tetA*	*Pme*I/*Aan*I	This study
pSM_lacZ_hyg_ntcA_up	pSM_lacZ_hyg with up region of *ntcA*	*Hind*III/ *Xba*I	This study
pSM_lacZ_hyg_ntcA_up+ dn1	pSM_lacZ_hyg_ntcA_up with dn1 fragment of *ntcA*-region	*Xba*I/*Not*I	This study
pSM_lacZ_hyg_sacB_ ntcA_up+dn1	pSM_lacZ_hyg_ntcA_up+dn1 with T7A1_sacB fragment of pSaccat	*Sca*I	This study
pSM_lacZ_hyg_ntcA_up+ dn2	pSM_lacZ_hyg_ntcA_up with dn2 fragment of *ntcA* region	*Bam*HI/*Nsi*I	This study
pSM_lacZ_hyg_amb01_ KO_up1	pSM_lacZ_hyg with up1 fragment of *amb01*-region	*Hind*III/ *Xba*I	This study
pSM_lacZ_hyg_amb01_ KO_up1+dn1	pSM_lacZ_hyg_amb01_KO_up1 with dn1 fragment of *amb01*-region	*Xma*JI/*Nsi*I	This study
pSM_lacZ_hyg_T7A1_ amb01	pSM_lacZ_hyg with T7A1_amb01 fragment	*Xma*JI/*Nsi*I	This study
pSM_gfp_tet_amb07_KO _up1	pSM_gfp_tet with up1 fragment of *amb07* region	*Xba*I/*Not*I	This study
pSM_gfp_tet_amb07_KO _up1+dn_{neu}	pSM_gfp_*tet*_amb07_KO_up1 with dn_{neu} fragment of *amb07* region	*Xma*JI/*Nsi*I	This study
pSM_lacZ_hyg_T7A1_ amb07	pSM_lacZ_hyg with T7A1 amb07 fragment	*Xma*JI/*Nsi*I	This study
pSM_lacZ_hyg_amb04_ KO_up	pSM_lacZ_hyg with up fragment of *amb04* region	*Hind*III/ *Xba*I	This study
pSM_lacZ_hyg_amb04_ KO_up+dn1	pSM_lacZ_hyg_amb04_KO_up dn1 fragment of *amb04* region	*Xma*JI/*Nsi*I	This study
pSM_gfp_tet_partlylacZ	pSM_gfp_tet with 1889 bp part of *lacZ*	*Xba*I/*Not*I	This study
pSM_gfp_tet_partlylacZ_ T7A1_amb04	pSM_gfp_tet_partlylacZ with T7A1_amb04 fragment	*Xma*JI/*Nsi*I	This study
pSM_gfp_tet_partlylacZ_ aphII_amb04	pSM_gfp_tet_partlylacZ with aphII_amb04 fragment	*Xma*JI/*Nsi*I	This study
pSM_lacZ_hyg_T7A1_ amb03-04	pSM_lacZ_hyg with T7A1_amb03-04 fragment	*Xma*JI/*Nsi*I	This study
pSM_lacZ_hyg_T7A1_ ambA	pSM_lacZ_hyg with T7A1_ambA fragment	*Xba*I/*Not*I	This study
pSUMO3_ck4	*npt*, *lacI*, His_6-SUMO (Genbank: KJ633124.1)	-	Dr. Carsten Kegler
pSUMO3_ck4_amb04	pSUMO3_ck4 with *amb04* fragment	*Bam*HI/*Kpn*I	This study

Culture conditions

Sorangium cellulosum strain #8405 was obtained from the laboratory collection of the Department of Microbial Drugs, Helmholtz Centre for Infection Research, Braunschweig, Germany. For routine propagation the strain was cultivated in liquid M-medium (1.0% soybean peptone, 1.0 % maltose, 0.1% $CaCl_2 \cdot 2\ H_2O$, 0.1% $MgSO_4 \cdot 7H_2O$, 8 mg/l NaFe-EDTA and 1.19% HEPES pH 7.4). HS-medium (0.15% casitone, 0.1% KNO_3, 0.1% $MgSO_4 \cdot 7\ H_2O$, 8 mg/l Fe-EDTA and 0.2% HEPES pH 7.2, supplemented with 0.00625% K_2HPO_4, 0.4% glucose and 0.0075% $CaCl_2 \cdot 2\ H_2O$ after autoclaving) was used during electroporation experiments and for initial cultivation of mutants. If necessary, antibiotics were added in the following concentrations: oxytetracycline 5-10 µg/ml; hygromycin B 100-150 µg/ml. For metabolic analysis the respective strains were almost exclusively cultivated in A-medium (0.4% glycerol, 0.8% soluble starch, 0.4% soybean meal, 0.2% yeast extract, 0.1% $CaCl_2 \cdot 2\ H_2O$, 0.1% $MgSO_4 \cdot 7\ H_2O$, 8 mg/l NaFe-EDTA and 2.83% HEPES, pH 7.4). For initial evaluation of the best production medium the strain was additionally grown in H-medium (0.2 % soybean meal, 0.2 % glucose, 0.8% starch, 0.2% yeast-extract, 0.1 % $CaCl_2 \cdot 2\ H_2O$, 0.1 % $MgSO_4 \cdot 7H_2O$, 8 mg/l NaFe-EDTA, 1.19% HEPES, pH 7.4) and P-medium (0.2% peptone Marcor S, 0.8% soluble starch, 0.4% probion, 0.2% yeast extract, 0.1% $CaCl_2 \cdot 2\ H_2O$, 0.1 % $MgSO_4 \cdot 7H_2O$, 8 mg/l NaFe-EDTA and 2.83% HEPES, pH 7.4). Solid growth media of any kind contained 1.5% agar. All cultivations were performed at 30°C. Liquid cultures were placed on a rotary shaker set to 200 rpm. For long term preservation, the respective strains were grown on agar plates (with appropriate antibiotics) until they formed a lawn. Thereafter, pieces of agar were cut out and stored in cryotubes at – 80°C.

Escherichia coli DH10B [32] as well as *E. coli* XL1-blue (Stratagene) were used for all conventional cloning procedures and for plasmid propagation. When non-methylated plasmid DNA was required, the *E. coli* strain SCS110 (Stratagene) was employed. Strain *E. coli* GB05-red [33] was used in Red/ET recombineering experiments. Heterologous production of His_6-SUMO-Amb04 was carried out using the *E. coli* strain Rosetta (Novagen). In all cases, the cells were cultivated either in LB-medium (1% tryptone, 0.5% yeast extract and 0.5% NaCl) or on LB-agar plates (supplemented with 1.5% agar) at 30 or 37°C and shaken at 200 rpm in case of liquid cultures. When necessary, the growth medium was supplemented with hygromycin B (70-100 µg/ml), oxytetracycline (10 µg/ml) or kanamycin (50 µg/ml).

Metabolic analysis of *S. cellulosum* strain #8405 and respective mutants

For metabolic analysis of the wild-type and the mutant strains generated in this study, cells were always cultivated in 300 ml flasks containing 50 ml of A-medium (0.4% glycerol, 0.8%

soluble starch, 0.4% soybean meal, 0.2% yeast extract, 0.1% $CaCl_2 \cdot 2\ H_2O$, 0.1% $MgSO_4 \cdot 7\ H_2O$, 8 mg/l NaFe-EDTA and 2.83% HEPES, pH 7.4) for 7 days at 30°C and 200 rpm. Before the cultures were harvested in a falcon tube using a Beckmann Avanti J-25 centrifuge equipped with a JA-12 rotor at 9.500 rpm and RT for 30 min, 1 ml of every culture was hold back for subsequent determination of the sample's protein content (see below). After centrifugation the supernatant was transferred into a new falcon tube, 2% (v/v) XAD-16 resin was added and the mixture was incubated on a tube rotator at 40 rpm for 30 min. Next, the supernatant was discarded and the XAD-16 resin as well as the cell pellet were separately extracted with 40 ml methanol. After complete evaporation of the solvent the respective extracts were dissolved in 1 ml of methanol and analyzed by LC-MS. Therefore the samples are measured on a Dionex Ultimate 3000 RSLC system using a Waters BEH C18, 50 x 2.1 mm, 1.7 µm d_p. Separation was achieved by a linear gradient with (A) H_2O + 0.1% FA to (B) ACN + 0.1% FA at a flow rate of 600 µl/min and 45°C. The gradient was initiated by a 0.33 min isocratic step at 5% B, followed by an increase to 95% B in 9 min to end up with a 1 min flush step at 95% B before reequilibration under the initial conditions. Coupling the HPLC to the MS was supported by an Advion Triversa Nanomate nano-ESI system attached to a Thermo Scientific Orbitrap. Mass spectra were acquired in centroid mode ranging from 200 – 2000 m/z at a resolution of R = 30000. For data evaluation the Thermo Xcalibur tool (Thermo Scientific) was used. Finally, obtained areas for the respective ambruticin derivatives were correlated to the total protein content of the sample which was determined using a modified version of the previously described method by Bradford, M.M. [34]. Therefore, the cells were disrupted by sonification (2x 15 sec, 70% amplitude, small probe; vibra-cell™, Zinsser Analytik GmbH). The resulting lysate was used in a microtiter plate based Bradford assay.

Routine molecular biological techniques

All common molecular biological techniques, such as preparation of genomic or plasmid DNA, directed cloning of DNA fragments and separation of nucleic acids were performed as described elsewhere [35]. The amplification of DNA fragments from either genomic or plasmid DNA was carried out according to a previously described method [36] by using either Taq or Phusion DNA-polymerase. The standard reaction mixtures (25 µl) for both polymerase chain reactions (PCR) are shown in Table 4. Usually the following standard conditions were applied: initialization: 2 min at 96°C, followed by 30 repetitive cycles of denaturation (1 min at 95°C), oligonucleotide annealing (20-30 sec at variable temperature (60-72 °C)) and elongation (variable time (1 min/kb for Taq; 15 sec/kb for Phusion) at 72°C). Before storage at 4°C a final elongation step of 5 min at 72°C should ensure the completion of the amplification process.

Table 4: Composition of a standard PCR mixture using Taq or Phusion polymerase

Component	Taq	Phusion
oligonucleotides [10µM] (for/rev)	1.0 µl	1.25 µl
dNTP's [1.25 µM each]	4.0 µl	4.0 µl
reaction buffer [10x/5x]	2.5 µl	5.0 µl
$MgCl_2$ [25 mM]	2.0 µl	-
glycerol [50%]	4.0 µl	-
DMSO [100%]	-	1.25 µl
template [100 ng]	1.0 µl	0.5 µl
DNA-polymerase	0.1 µl	0.1 µl
H_2O	9.4 µl	11.65 µl

All enzymes and buffers used for the contemplated applications were obtained from Thermo Scientific or New England Biolabs.

Construction of inactivation and overexpression plasmids used in this study

All oligonucleotides used for specific amplification of DNA fragments are given in Table 2. The plasmids generated for inactivations or promoter exchanges are descendants of pBluecat (Dr. Carsten Volz, unpublished). Based on that vector, pSM_lacZ_hyg was constructed by exchanging the chloramphenicol resistance gene (*cat*) against the hygromycin resistance gene (*hpt*). Therefore the *hpt* gene was amplified in a PCR reaction using the oligonucleotides pblueHyg_For_PmeI and pblueHyg_RV_AanI with pMycoMAR_hyg as template. The resulting 1021 bp DNA fragment was cleaved with *Pme*I/*Aan*I and subsequently cloned into the prepared pBluecat vector. Plasmid pSM_gfp_tet was generated in two consecutive steps using the pSM_lacZ_hyg vector. First, *lacZ* was exchanged against *megfp* by integrating the 816 bp DNA fragment which was amplified from pPm1-mEGFP-rppA (Dr. Katja Gemperlein, unpublished) with the oligonucleotides NotI-T7A1lacZRBS-megfp and megfp_rev_NdeI into pSM_lacZ_hyg using the indicated restriction enzymes thereby generating pSM_gfp_hyg. In the second step, the 1332 bp *Pme*I fragment from pMycoMAR_tet containing *tetA* was introduced into pSM_gfp_hyg which was cleaved beforehand with *Pme*I/*Aan*I.

For construction of single cross-over integration plasmids, the respective homology fragments were PCR amplified using the relevant oligonucleotides depicted in Table 2 and subsequently cloned into pSM_lacZ_hyg by using the indicated restriction sites.

In order to obtain linear DNA fragments suitable for gene replacements in strain #8405, the respective up and down regions flanking the target gene were amplified and consecutively integrated into one of the plasmids (pSM_lacZ_hyg or pSM_gfp_tet). Finally, the resulting plasmid was digested with the external restriction enzymes used for initial cloning of the flanking

regions, separated by agarose gel electrophoresis and the required fragment was recovered from the gel.

For genetic complementation the vector pSM_gfp_tet_partlylacZ was constructed. This vector allows the reintroduction of a gene deleted previously by insertion of a pSM_lacZ_hyg derivative into the region of interest. Integration of the respective pSM_gfp_tet_partlylacZ derivative is achieved by single cross over recombination via the lacZ allele of the integrated pSM_lacZ_hyg derivative (see Figure 25). To construct pSM_gfp_tet_partlylacZ an internal part of *lacZ* (1907 bp) was amplified from pSM_lacZ_hyg (using the oligonucleotides partly_lacZ_for_XbaI and partly_lacZ_rev_NotI) and cloned into pSM_gfp_tet by employing the indicated restriction sites. Based on this vector two complementation plasmids (pSM_gfp_tet_partlylacZ_T7A1_amb04 and pSM_gfp_tet_partlylacZ_aphII_amb04) have been constructed. The first was generated by cloning a 1480 bp fragment (amplified with T7A1_amb04_for_XmaJI and T7A1_amb04_rev_NsiI) containing the T7A1 promoter fused to *amb04* into pSM_gfp_tet_partlylacZ. The second plasmid contains *amb04* under the control of the *aphII* promoter. Therefore the *aphII* promoter and the *amb04* gene were amplified in two separate PCR reactions ((aphII_for_XmaJI and aphII_rev_overlap_amb04; 374 bp aphII_overlap_*amb04*) and (overlap_aphII_amb04_for and amb04_rev_NsiI; 1480 bp overlap_aphII_*amb04*) with each amplicon comprising a sufficient overhang for merging both fragments in a third PCR reaction (aphII_for_XmaJI and amb04_rev_NsiI; 1759 bp aphII_amb04)). Finally, the complete PCR fragment was cloned into pSM_gfp_tet_partlylacZ by the use of *Xma*JI/*Nsi*I restriction sites, yielding pSM_gfp_tet_partlylacZ_aphII_amb04.

Construction of pMycoMAR_tet using the Red/ET technology

A single colony of *E. coli* strain GB05-red [33] carrying pMycoMAR_hyg was used to inoculate 5 ml of LB-medium containing 70 µg/ml hygromycin B and cultivated overnight at 37°C. On the next day, 30 µl of this culture were transferred to 1.4 ml LB-medium containing 70 µg/ml hygromycin B and cultivated at 30°C and 1000 rpm until an optical density (OD_{600}) of 0.2 was reached. Thereafter, the medium was supplemented with L-arabinose to a final concentration of 0.15% and incubated at 37°C and 1000 rpm for approximately one hour. At an OD_{600} of 0.35-0.4 the cells were harvested at 11000 rpm (30 sec at 4°C) in a bench top centrifuge and washed twice with 1 ml ice-chilled water. For transformation with 0.3 µg of the 1434 bp *tetA* PCR-product (amplified with Red/ET_tet_for and Red/ET_tet_rev (Table 2) from pACYC184) the cells were resuspended in 50 µl distilled water, the DNA-fragment was added and transformed via electroporation at 1350 V, 10 µF, and 600 Ω into the cells. For phenotypic expression, the cells were quickly resuspended in 600 µl of LB-medium and incubated for one hour at 37°C and

900 rpm. Thereafter, the cells were plated on solid LB-media containing 10 µg/ml tetracycline and incubated overnight at 37°C. Obtained clones were cultivated in liquid LB-medium and have been further analyzed.

Sequencing of DNA constructs

To prove the integrity of plasmids generated in this study, they were sequenced at Seq-IT GmbH & Co. KG, Kaiserslautern or LGC-Genomics GmbH, Berlin. All derivatives of plasmid pSM_lacZ_hyg were sequenced using one or more of the following oligonucleotides: pBlues_For, pBlues_For2, pBlues_RV and pBlues_RV2. In case of pSM_gfp_tet derivatives, the oligonucleotides pBlues_For, pBlues_RV, pBlues_RV2 and Seq_Tet_out_down 2 were applied. Important regions of the plasmid pSUMO3_ck4 were sequenced with the oligonucleotides T7prom and T7term.

Transformation of *S. cellulosum* strain #8405 via electroporation

A stationary preculture of strain #8405 grown in HS-medium was used to inoculate 50 ml of HS-medium in a 300 ml flask with volumina ranging from 100-500 µl. Subsequent cultivation is carried out at 30°C and 180 rpm until the culture reaches an OD_{600} at 0.8-1.0. For the preparation of electrocompetent cells, 2-3 ml of the culture were harvested in 2 ml Eppendorf tubes in a benchtop centrifuge at 15.000 rpm, RT and 10 min. These cells were washed three times with 10% glycerol, resuspended in 95 µl of 10% glycerol, and 1-3 µg of plasmid or linear DNA (dissolved in 5 µl H_2O) was added. Afterwards, the cell suspension was transferred to an electroporation cuvette (1mm). Electroporation was carried out with a Bio-Rad Gene Pulser using exponential decay pulses at: 1200 V 200 Ω and 25 µF. Thereafter, the cuvette was flushed with 500 µl HS-medium and the cell-suspension was transferred into a large test-tube containing additional 3 ml of HS-medium. After subsequent cultivation for approximately 20 h @ 30°C and 180 rpm, the cells were harvested and resuspended in 300 µl of HS-medium. The cells were spread onto HS-agar containing the appropriate antibiotic(s) (hygromycin B 150 µg/ml; oxytetracycline 10 µg/ml) and, if necessary, 80 µg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal). After incubation for approximately 10-14 days, the first colonies could be observed. For further analysis, the respective clones were transferred onto new selective agar plates. If these clones were still able to grow, each one was used to inoculate a small volume of HS-medium (5 ml medium in a 25 ml flask) containing the appropriate antibiotics (hygromycin B (100 µg/ml) and/or oxytetracycline (10 µg/ml)), and cultivation was carried out as described above.

Usually, 16 transformation attempts have been sufficient to obtain an adequate number of single cross-over mutants. The construction of gene replacement mutants using linear DNA was in general less effective.

Extraction of genomic DNA from Myxobacteria

50 ml of a culture grown in HS-medium (0.15% casitone, 0.1% KNO_3, 0.1% $MgSO_4 \cdot 7\ H_2O$, 8 mg/l Fe-EDTA and 0.2% HEPES pH 7.2, supplemented with 0.00625% K_2HPO_4, 0.4% glucose and 0.0075% $CaCl_2 \cdot 2\ H_2O$ after autoclaving) for at least 5 days at 30°C and 200 rpm were harvested by centrifugation (10 min at 7000xg and RT). Thereafter, the cell pellet was resuspended in 4 ml SET-buffer (25 mM Tris/HCl pH 8.1, 75 mM NaCl and 25 mM EDTA). Subsequently, 200 µl proteinase K (20 mg/ml in H_2O), 400 µl SDS (10% w/v) and 100 µl RNase A (10 mg/ml in H_2O) were added. After inverting the tube several times, it was incubated at 55°C for at least 2 h. Thereafter 1.6 ml of 5 M NaCl and 4 ml phenol:chloroform:isoamyl alcohol (25:24:1) were added and the mixture was incubated for 30 min on a tube rotator at 10 rpm. After centrifugation (10 min at 7000xg and 4°C) the upper phase was transferred to a new tube. The extraction procedure was repeated additional two times. After the last centrifugation step, the upper phase was transferred to a new tube and one volume of isopropanol was added. Subsequently, the tube was gently inverted several times until “cloudy-like” DNA appeared which was collected by wrapping it around the tip of a glass Pasteur pipette. Before the DNA was dissolved in sterile H_2O, it was washed by immersing the tip of the pipette in a tube filled with 70% ice-cold ethanol. Finally, concentration and purity of the sample was determined using a NanoDrop 2000 UV-Vis spectrophotometer (Thermo Scientific) and stored at 4°C.

Verification of generated mutants by PCR

Mutants which have been created by single-cross over homologous recombination were analyzed by PCR as described elsewhere [37]. Therefore, three distinct reactions were conducted in which wild type- as well as mutant-DNA was applied. To detect the loss of the wt-allel in mutants, one set of oligonucleotides binding up- and downstream of the homologous region (usually named xxx_check_up and xxx_check_dn) was employed in reaction 1. The other two reactions were carried out in each case using a pair of oligonucleotides with one being specific for the vector backbone and the other for a region outside of the homologous area.

Verification of generated mutants by Southern Blot analysis

For Southern Blot analysis of genetically altered strains generated in this study, they were grown in HS-medium (0.15% casitone, 0.1% KNO_3, 0.1% $MgSO_4 \cdot 7\ H_2O$, 8 mg/l Fe-EDTA and

0.2% HEPES pH 7.2, supplemented with 0.00625% K_2HPO_4, 0.4% glucose and 0.0075% $CaCl_2 \cdot 2\ H_2O$ after autoclaving) for 5 days at 30°C and 200 rpm. Next, the cells were harvested by centrifugation (10 min at 7000xg and RT), and genomic DNA was extracted. 5-10 µg DNA were cleaved overnight with appropriate restriction endonucleases. After electrophoretic separation (25 V for approx. 20 h) of the respective samples using a TAE-buffered (40 mM Tris, 20 mM acetic acid and 1 mM EDTA) agarose gel (1%), the gel was treated with 250 mM HCl for 15 min. Subsequently, the gel was washed with water followed by two times incubation for 15 min in denaturation solution (0.5 M NaOH and 1.5 M NaCl). Before the gel was placed for 30 min in 20x SSC buffer (3 M NaCl and 300 mM trisodium citrate, pH 7.0) it was washed with water and treated twice with neutralizing solution (1 M Tris/HCl pH 7.5 and 1.5 M NaCl) for 15 min. Afterwards, the DNA was transferred onto a positively charged nylon membrane (Roche Applied Science) by the method of capillary blotting. Then, the membrane was treated for 2 min with 0.4 M NaOH and 0.25 M Tris/HCl pH 7.5, respectively. Finally, it was placed for 2 h in an oven set to 80°C.

For non-radioactive visualization of specific DNA fragments bound to the membrane the DIG-system from Roche Applied Science was used. For the purpose of prehybridization, the membrane was incubated with 25 ml DIG Easy Hyb solution (Roche Applied Science) for 1 h at 42°C in a glass tube under rotation in a hybridization oven. Thereafter the solution was exchanged against 25 ml of DIG Easy Hyb containing at least 7.5 ng/ml of a specific DNA probe and rotative incubation at 42°C was carried out until the next day. All probes used in this study were about 1.0-1.7 kb in size and were generated using the PCR DIG Labeling Mix (Roche Applied Science), which allows direct labeling of amplification products with digoxigenin-dUTP in a PCR with Taq or Phusion DNA-polymerase. The respective oligonucleotides used therefore are given in Table 2. If the double stranded DNA probes were used for the first time, they were denatured at 95°C for 10 min and cooled down on ice before they were added to the DIG Easy Hyb solution. In case they were reused, the whole DIG Easy Hyb solution containing the probe was incubated at 70°C for 10 min. On the following day the membrane was washed twice with 25 ml of 2x washing solution (2x SSC, 0.1% (v/v) SDS) for 20 min at RT succeeded by a second washing step with two times 25 ml of 0.5x washing solution (0.5x SSC, 0.1% (v/v) SDS) at 58 °C in a hybridization oven for 20 min. Next, 25 ml of DIG2-buffer (1% Blocking Reagent (Roche Applied Science) in DIG1-buffer (150 mM NaCl, 100 mM maleic acid, pH 7.5)) were added to the membrane and incubation was carried out at RT for 1 h. Subsequently, the membrane was incubated under rotation with 25 ml DIG2-buffer, containing 3 µl of Anti-digoxigenin-AP Fab fragments (Roche Applied Science), for 30 min at RT. Hereupon, the membrane was washed twice for 15 min at RT under rotation with DIG1-buffer comprising 0.3% (v/v) Tween®20. Before

the membrane was incubated for 2 min in DIG3-buffer (100 mM NaCl, 100 mM Tris/HCl pH 9.5) it was rinsed several times with DIG1-buffer without Tween®20. Thereafter the membrane was transferred into a clear plastic bag, covered with CDP-*Star*, ready-to-use (Roche Applied Science) and chemiluminescence was detected by the use of a gel documentation system (Fusion FX, Vilber Lourmat). For removal of the probe, the membrane was alternately incubated (2x 20 min each) with water and stripping solution (0.2 M NaOH, 0.1% (v/v) SDS), air-dried and stored at RT.

Extraction of RNA and quantitative real time PCR analysis

Cells were grown in 300 ml flasks containing 100 ml A-medium (0.4% glycerol, 0.8% soluble starch, 0.4% soybean meal, 0.2% yeast extract, 0.1% $CaCl_2$•2 H_2O, 0.1% $MgSO_4$•7 H_2O, 8 mg/l NaFe-EDTA and 2.83% HEPES, pH 7.4). Every 24 h 10 ml of the culture were harvested by centrifugation (10 min 7000xg and RT), shock frozen in liquid nitrogen, and stored at -80°C. For RNA extraction the cell pellet was resuspended in 4 ml preheated (60°C) NAE saturated phenol/chloroform (phenol/chlorofom 6:1, 5 mM sodium acetate, 1 mM EDTA, pH: 5.0). Before 4 ml NAES-buffer (50 mM sodium acetate, 10 mM EDTA and 1% (v/v) SDS, pH: 5.0) were added to the suspension, it was incubated for 5 min at 60°C in a water bath. After a second incubation for 5 min at 60°C the sample was cooled down on ice and subsequently centrifuged for 10 min at 7000xg and 4°C. Thereafter, the upper phase was transferred into a new falcon tube, 4 ml of phenol/chloroform (6:1) were added, and the tube was inverted several times before it was centrifuged again. This procedure was repeated two additional times and afterwards the upper phase was gently mixed with 400 µl of 3 M sodium acetate in 4 ml isopropanol. After incubation overnight at -20°C the tube was centrifuged, the supernatant was discarded and the pellet was washed with 3 ml 70% ethanol. Finally, the dried pellet was dissolved in 200 µl diethylpyrocarbonate (DEPC) treated H_2O. The last steps of RNA preparation were performed using the RNeasy Midi Kit (Qiagen).

In order to generate complementary-DNA (c-DNA), 5 µl of the RNA solution were mixed with 1 µl dNTP's (10 mM each), 1 µl random hexamer oligonucleotides (100 ng/ml), ad. 14.5 µl DEPC H_2O. This mixture was incubated for 5 min at 65°C, cooled down on ice, and added to 1 µl RevertAid Reverse Transcriptase (Thermo Scientific), 0.5 µl RiboLock RNase Inhibitor (Thermo Scientific), and 4 µl of 5x reverse transcriptase reaction buffer in a PCR tube. Conditions for the reverse transcription PCR (RT-PCR) were: 30 min at 60°C followed by 10 min at 85°C.

For quantitative real time PCR (q-PCR), a single reaction contained 5 µl 2x KAPA SYBR Fast reaction mix (Qiagen), 0.5 µl of the respective oligonucleotides 3 µl DEPC H_2O and 1 µl

cDNA or RNA (in concentration used in RT-PCR). Oligonucleotides for amplification of specific gene-sections are given in Table 2. Conditions for q-PCR in a Peq-Star 96Q cycler (Peqlab) were: 3 min at 95°C followed by 40 cycles of 3 sec 95°C, 20 sec 62 °C and 1 sec 72°C. Calculations were performed out using the peq-star 96Q software (Peqlab).

Heterologous production and purification of Amb04

Amb04 was heterologously produced as N-terminal His_6-SUMO-tagged protein in *E. coli* strain Rosetta. Therefore, the *amb04* gene was amplified in a PCR reaction with Phusion polymerase using the oligonucleotide pair amb04_for_BamHI/amb04_rev_KpnI (Table 2) and genomic DNA as template. The resulting 1403 bp PCR fragment was digested with *Bam*HI/*Kpn*I and subsequently cloned into the prepared pSUMO3_ck4 vector. After verification of the DNA sequence, the correct plasmid pSUMO3_ck4_amb04 was transformed into *E. coli* strain Rosetta. For protein production, 600 ml LB-medium (1% tryptone, 0.5% yeast extract and 0.5% NaCl) in a 2 l shaking flask were inoculated with an overnight preculture of *E. coli* strain Rosetta harboring pSUMO3_ck4_amb04 to a final OD_{600} of 0.1. Afterwards the cells were cultivated at 37°C and 180 rpm until they reached an OD_{600} of about 0.7-0.8. At this point the culture was cooled down in an ice bath for 15 min and gene expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM. Subsequent cultivation was carried out at 16 °C and 180 rpm for approximately 20 h. Thereafter the cells were harvested using a Beckmann Avanti J-25 centrifuge equipped with a JLA10-500 rotor at 6000 rpm and 4°C for 10 min and washed twice with PBS buffer (1.06 mM KH_2PO_4, 155.1 mM NaCl and 2.97 mM $Na_2HPO_4 \cdot 7H_2O$). Afterwards the cells were resuspended in 25 ml HisTrap A buffer (25 mM HEPES, 500 mM NaCl, 25 mM imidazole and 10% v/v glycerol, pH 8.0) containing Roche complete protease inhibitors (EDTA-free). For cell disruption the suspension was passed three times through a French Press at 900 PSI. The debris was removed by centrifugation at 20000xg for 30 min at 4°C. Afterwards the soluble fraction was filtered through a 0.45 µm syringe filter and incubated with 3 ml Ni^{2+}-NTA (Qiagen) for 1 h at 4°C. The protein was eluted with a stepwise gradient of imidazole using different mixtures of HisTrap A and HisTrap B buffers (25 mM HEPES, 500 mM NaCl, 500 mM imidazole and 10% v/v glycerol, pH 8.0). All fractions were subjected to SDS-PAGE. Fractions containing the target protein (≈63.0 kDa) were pooled and the HisTrap buffer was exchanged with SUMO2 protease buffer (20 mM Tris/HCl pH: 8.0 and 150 mM NaCl) by using Amicon® Ultra-15 centrifugal filter devices (30,000 NMWL) at 4500xg and 4°C. For doubtless identification of the produced protein 10 µl of the sample with a concentration of 0.8 mg/ml were subjected to LC-MS analysis (see LC-MS analysis of intact proteins). The protein solution was transferred into a 15 ml falcon tube and the SUMO tag was

cut off by the SUMO2 protease (4 U/mg protein) overnight at 4°C. On the next day, 1.5 ml of Ni^{2+}-NTA (Qiagen) in SUMO2 protease buffer were added and further incubation was carried out for 1 h at 4°C. The whole mixture was then applied onto an open plastic column and the flow through which contained Amb04 without SUMO tag was collected. Finally, the protein solution was stored at -80°C containing glycerol in a final concentration of 10%.

LC-MS analysis of intact proteins

All ESI-MS measurements were performed on a Dionex Ultimate 3000 RSLC system using an Aeris Widepore XB-C8, 150 x 2.1 mm, 3.6 µm d_p column (Phenomenex, USA). Separation of 2 µl sample was achieved by a linear gradient using (A) H_2O + 0.1 % FA and (B) ACN + 0.1 % FA at a flow rate of 300 µl/min and 45 °C. The gradient was initiated by a 1.0 min isocratic step at 2 % B, followed by an increase to 75 % B in 8 min to end up with a 3 min step at 75 % B before reequilibration under the initial conditions. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 35 µl/min before entering the maXis 4G hr-ToF mass spectrometer (Bruker Daltonics) using the standard Bruker ESI source. Dry gas temperature was set to 180°C at a flow of 6 l/min. The capillary voltage was held at 4000 V and the nebulizer was set to 1.1 bar. Mass spectra were acquired in positive ionization mode ranging from 600-1800 *m/z* at a 2.5 Hz scan rate. Protein masses were deconvoluted using the Maximum Entropy algorithm as implemented in the DataAnalysis software (DA 4.2 b383).

Electrophoretic mobility shift assay (EMSA)

EMSAs were performed on the basis of a previously described protocol [38] using native polyacrylamide gels of a concentration of 4.0%. A typical preparation contained 0.1 pmol HEX-labeled or 0.2 pmol Cy5-labeled DNA-fragments, 0.5 µg of the DNA mimic poly (deoxyinosinic-deoxycytidylic) acid sodium salt (Sigma-Aldrich) and Amb04 (100 fold molar excess to labeled DNA fragments) in gel-shift buffer (50 mM K-acetate, 8 mM Mg-acetate, 0.5 mM dithiothreitol and 15 mM Tris-acetate pH: 7.3). For *in vitro* phosphorylation of Amb04, the protein was incubated with 30 mM acetyl phosphate for 30 min at 37°C. EMSA assays with cell lysates of strain #8405 were likewise performed, with the exception that 5 or 10 µl of the respective lysates were used instead of heterologously produced protein. Gels were run at 4°C with 12 mA per gel for 120-150 min. Imaging of the labeled DNA fragments was performed by scanning the gel with the Typhoon™ 9410 gel imager (GEhealthcare) using the following settings: 5′-hexachlorofluorescein phosphoramidite (HEX), excitation laser at 532 nm, emission filter 555 nm; Cy5, excitation laser at 633 nm, emission filter 670 nm.

Labeling of DNA fragments for EMSA

DNA fragments of 200 bp and longer were generated by two consecutive PCR reactions employing Phusion polymerase. First, the fragments of interest were amplified using area specific oligonucleotides (Table 2) additionally harboring 5' extensions for subsequent 5′-HEX-labeling with the oligonucleotides EMSA_sec_HEX2 and EMSA_sec_HEX3 (Table 2). Thus, defined HEX-labeled fragments around the promoter and/or the 5' regions of the respective genes were generated for *amb03*, *amb04a*, *amb05/06*, *ambA*, *ambB*, *ambG*, *ambH*, *ambNQ*, *ambR*, *ambS* and *amb07*. The 676 bp DNA fragment in front of *ambA* used in EMSA studies with cell lysates was only labeled once. Accordingly, the second PCR reaction was performed using the oligonucleotides EMSA_sec_HEX2 and ambB_rev2 (Table 2).

DNA fragments of around 50 bp were generated in an oligonucleotide based approach. Therefore, complementary oligonucleotides (20 µM in H_2O; Table 2) were mixed, heated up to 95°C for 1 min and cooled down to RT to allow annealing. One of the oligonucleotides in every mix contained a 5' sequence complementary to a third oligonucleotide harboring a Cy5 label at its 3'-end (5'-P-3'-Cy5). Finally, the annealed oligonucleotide pair (0.25 µM) and the 5'-P-3'-Cy5 oligonucleotide (0.25 µM) were ligated. After heat inactivation of the ligation reaction (10 min at 65°C), the fragments were used in EMSA experiments.

DNA-protein pull-down

In order to identify proteins specifically binding to a promotor region of interest, cell free lysates of strain #8405 were used in DNA-protein pull-down assays. This was performed to separate these unknown proteins from all other proteins in the lysates. At this, a 676 bp DNA probe was used being biotinylated at its 5´-end. As a negative control, a biotinylated DNA probe bearing promotor P_{taA} of the myxovirescin cluster of *Myxococcus xanthus* DK1622 was coincubated in parallel with the lysates. The DNA probes were generated by PCR using one specific oligonucleotide bearing a 5'-biotin residue in each case (Table 2). All experiments were performed like described before by Volz *et al.* [38]. An SDS PAGE analysis was performed to visualize the result of the DNA-protein pull down assays. All visible protein bands were cut out and subjected to MALDI-TOF analysis to identify the respective proteins.

Results and discussion

Identification and initial characterization of ambruticin producer strains

In order to detect good ambruticin producers from our strain collection, we made use of our in-house database, the Myxobase, which allows - amongst numerous other functions - the effective *in silico* screening for producers of known myxobacterial compounds. The underlying data repository is based on a highly standardized and thoroughly monitored sample preparation and LC-MS measuring workflow which applies for all our myxobacterial strains as well as purified compounds. Those LC-MS data sets are subsequently searched against a library of known natural products based on retention time fit, accurate *m/z* and isotope pattern fit. Resulting lists of known compounds are uploaded to the Myxobase and are freely searchable, thus allowing rapid identification of all producers of a distinct compound across all uploaded data sets. At the end of the year 2013 this database contained approximately 2500 secondary metabolome datasets from various myxobacterial genera [30]. However, at the beginning of this project, when the search for ambruticin producers was conducted, the number of available LC-MS data sets was significantly lower with only 757 extracts. The extracts covered myxobacterial strains from various genera as depicted Figure 1.

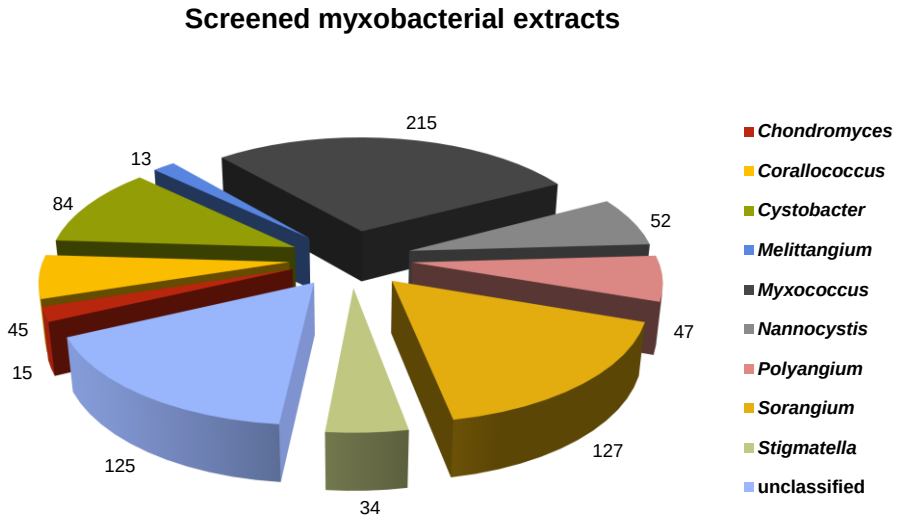
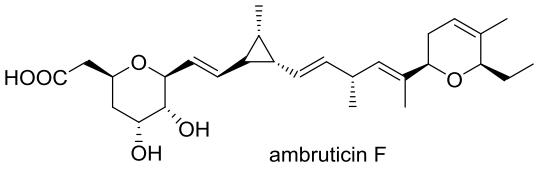
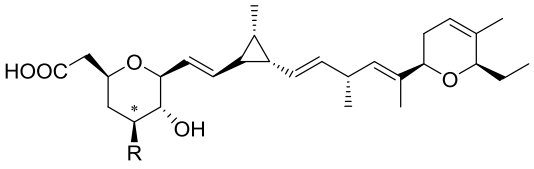


Figure 1: Screened myxobacterial extracts from different genera in terms of ambruticin production (N=757).

As an outcome we could identify nine strains producing ambruticins, which all belong to the genus *Sorangium*. The strains were then further analyzed in detail regarding their production

spectrum in comparison with the derivatives known so far [39]. Accordingly, they can be classified into two major groups; I: producers of post-PKS modified ambruticins (VS-ambruticins) and II: producers of <u>non</u>-post-PKS modified ambruticin (ambruticin F). The major difference between these derivatives is the degree of modification at the pyran ring (marked with ***** Figure 2). While ambruticin F features a hydroxyl group at this position, the different VS-ambruticins possess an amino group with various degrees of methylation (Figure 2). Noticeable, the detailed examination of the high resolution-MS-data revealed, that in most of the cases those strains showing high abundance of ambruticin F in their extracts synthesized comparatively low amounts of VS-ambruticins and vice versa (Figure 3). This becomes especially apparent for the best producers of ambruticin F and the VS-ambruticins, respectively. In particular, strain #7282 shows the best yield of ambruticin F while almost no VS-ambruticins are present in the corresponding extract. Contrary to this, strain #8405 seems to process the PKS product much more efficiently which is why the proportion is strongly shifted to the VS-ambruticins.





ambruticin	R=
S	OH
VS1	$N(CH_3)_3$
VS3	$N(CH_3)_2$
VS3 N-oxide	$N(CH_3)_2OH$
VS4	NCH_3
VS5	NH_2

Figure 2: Different ambruticins produced by *S. cellulosum* strain #8405.

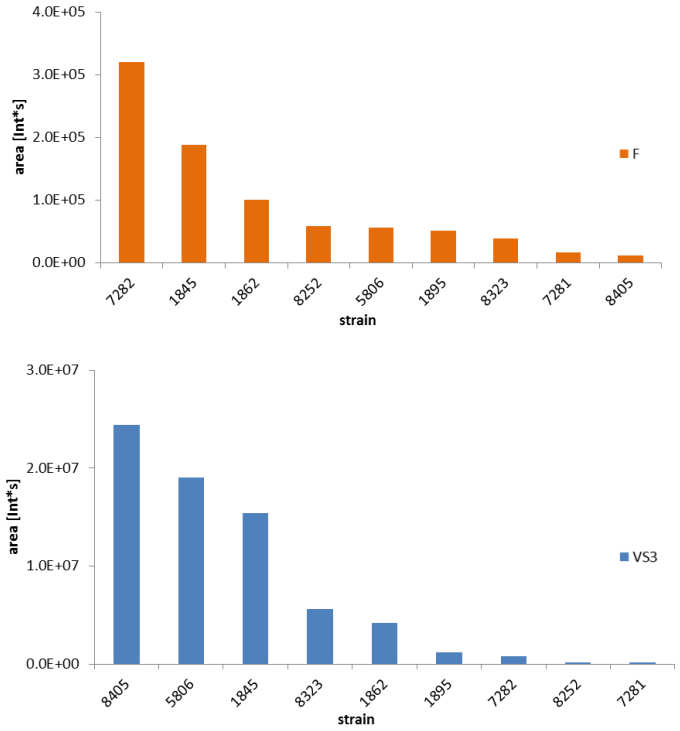


Figure 3: Production properties of all identified ambruticin producing strains with respect to their capability to synthesize ambruticin F and VS3 (typically the most prevalent VS-ambruticin and therefore representative for these derivatives); extracts are prepared according to a standard procedure, which is why the amounts are comparable.

As the focus of this work is on VS-ambruticins and our initial analysis highlighted strain #8405 as the best tangible producer, we decided to concentrate all our efforts on optimizing this strain towards an enhanced compound formation.

Since it is well known that the composition of the culture broth can have a tremendous impact on the production yield of a particular compound, we tried to assess which cultivation medium is most suitable for ambruticin formation by strain #8405. Therefore, strain #8405 was grown in several different media known to facilitate production of secondary metabolites. Subsequently, methanolic extracts of the corresponding cultures were prepared, subjected to LC-MS analysis, and the production of ambruticin F and VS3 was evaluated (Figure 4).

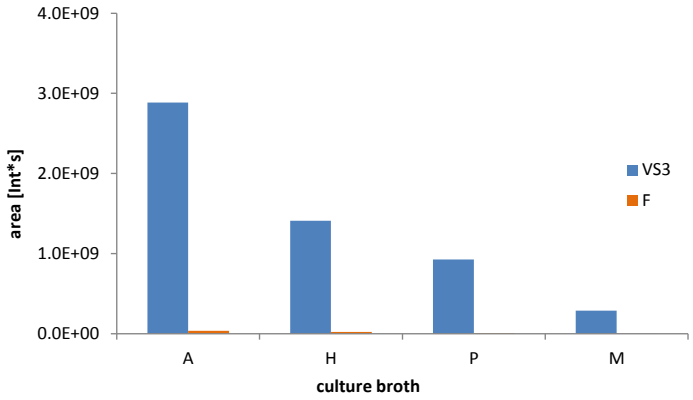


Figure 4: Impact of various culture media for strain #8405 on the production of ambruticin F and VS3.

Based on these results, A-medium could be determined as the most favorable culture broth for production of the ambruticins among all tested media (approximately 10 mg/l). Accordingly, we used this medium in any further experiments dealing with the evaluation of genetically engineered strains.

Besides the already good and stable production properties regarding ambruticins, strain #8405 is able to grow in suspension what further qualified it for this project. This capability should enable the testing of conjugation and electroporation in order to establish a gene transfer system. However, before first attempts to genetically alter strain #8405 could be performed, the resistance spectrum of this strain had to be determined to clarify which antibiotics effectively prevent its growth. Therefore, five different antibiotics (apramycin ≤ 60 µg/ml, kanamycin ≤ 60 µg/ml, spectinomycin ≤ 60 µg/ml, hygromycin B ≤ 200 µg/ml and oxytetracycline ≤ 10 µg/ml) were tested in an agar plate based sensitivity screen. As a result, hygromycin B and oxytetracycline were found to inhibit growth of the organism at concentrations above 150 and 5 µg/ml, respectively. Since the minimal inhibitory concentration (MIC) of an antibiotic can vary between liquid and solid media, additional growth experiments in the presence and absence of the respective antibiotics in liquid M-medium have been performed (Figure 5).

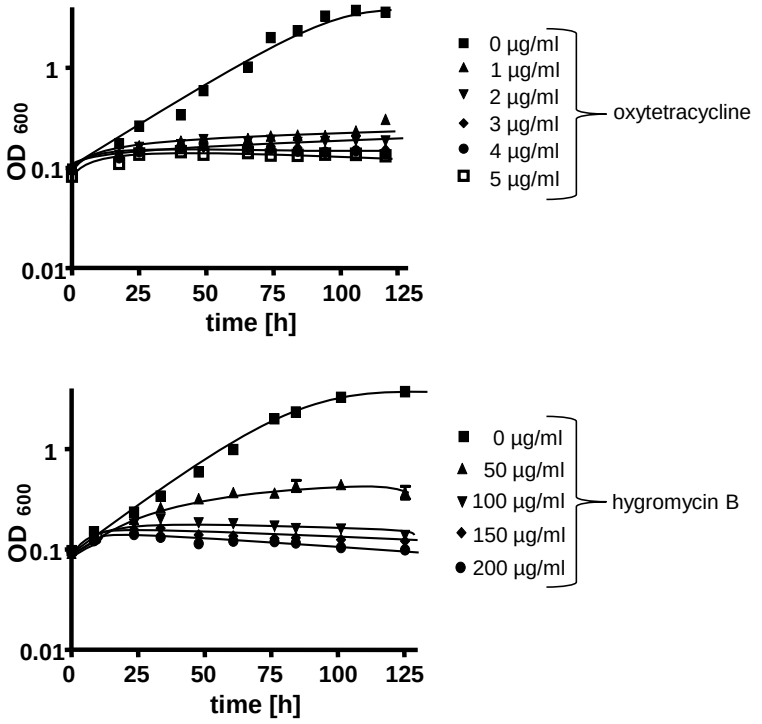


Figure 5: Growth experiments with strain #8405 in M-medium for the MIC determination of oxytetracycline and hygromycin B.

This examination clearly showed that oxytetracxcline already inhibits growth of the organism at concentrations above 1 µg/ml in liquid culture broth whereas solid media needs to be supplemented with at least five times more of the antibiotic. A similar effect is observed for hygromycin B where in liquid medium only two-thirds of the amount which is used in agar plates is required.

In summary, the screening for ambruticin producers available from our strain collection identified nine organisms belonging to the genus *Sorangium*. Strain #7282 exhibits the highest yields for ambruticin F while strain #8405 is the best producer of the VS-ambruticins. Besides its good production properties, planctonic growth of strain #8405 should facilitate strain handling and both, hygromycin B and oxytetracycline susceptibility are prerequisites for the development of genetic tools and protocols.

Development of a gene transfer protocol for strain #8405

To enable strain optimization by genetic engineering, we aimed to establish a gene transfer protocol. To date, there are four different options known how exogenous DNA can be introduced into a bacterial cell: transduction, natural competence, conjugation and electroporation. The process of transduction, first described for *Salmonella typhimurium*, is mediated by phages

which are capable of moving genetic elements from one bacterium to the other [40]. Today this method is often used to shift mutations between various strains of *Bacillus subtilis* [41;42]. Contrary to this, DNA uptake in the context of natural competence is mediated by the cell itself and does not require any donor organism [43;44]. Similar to phage transduction, this phenomenon is wildly-used to transfer exogenous DNA into a variety of bacteria such as *Streptococcus pneumoniae* [45], *Thermus thermophilus* [46], and *Helicobacter pylori* [47]. Conjugation illustrates the third possibility for the exchange of genetic material between two strains and has been discovered in the year 1946 by Lederberg & Tatum [48]. This kind of horizontal gene transfer significantly accounts for the propagation of plasmids harboring various antimicrobial-determinants (R-plasmids) between different pathogenic bacteria [49] what contributes to the occurrence of multi-resistant strains hard to control with conventional methods [50]. In the laboratory this system is employed to introduce circular DNA into for example numerous actinobacteria [51] pseudomonads [52;53], and different myxobacteria [31;54]. The fourth option is electroporation. It can be used for prokaryotic as well as for eukaryotic systems [35;55;56] and represents an additional possibility to manipulate myxobacteria like *M. xanthus*, *Stigmatella aurantiaca*, and *S. cellulosum* [57-59].

In order to establish a genetic manipulation system for strain #8405 electroporation was chosen over conjugation because of two facts: First, DNA transfer by conjugation is more laborious and therefore more time-consuming. Second, the use of a further antibiotic to dispose the donor organisms may cause undesired secondary effects.

For initial determination of adequate electroporation conditions, a transposon based approach was employed. As described earlier, hygromycin B and oxytetracycline were found to efficiently inhibit the growth of strain #8405. Accordingly, the myxobacterial transposons pMycoMAR_hyg [31] and pMycoMAR_tet (this study) were used. Within various attempts applying four different electroporation settings (I: 850 V, 400 Ω, 25 µF; II: 1200 V, 200 Ω, 25 µF; III: 1200 V, 400 Ω, 25 µF and IV: 2000 V, 400 Ω, 25 µF), we obtained clones resistant to the respective antibiotics only under condition I and II with the latter one being more efficient. After further cultivation of these clones, genomic DNA was extracted, digested with appropriate restriction enzymes and subjected to Southern Blot analysis (Figure 6).

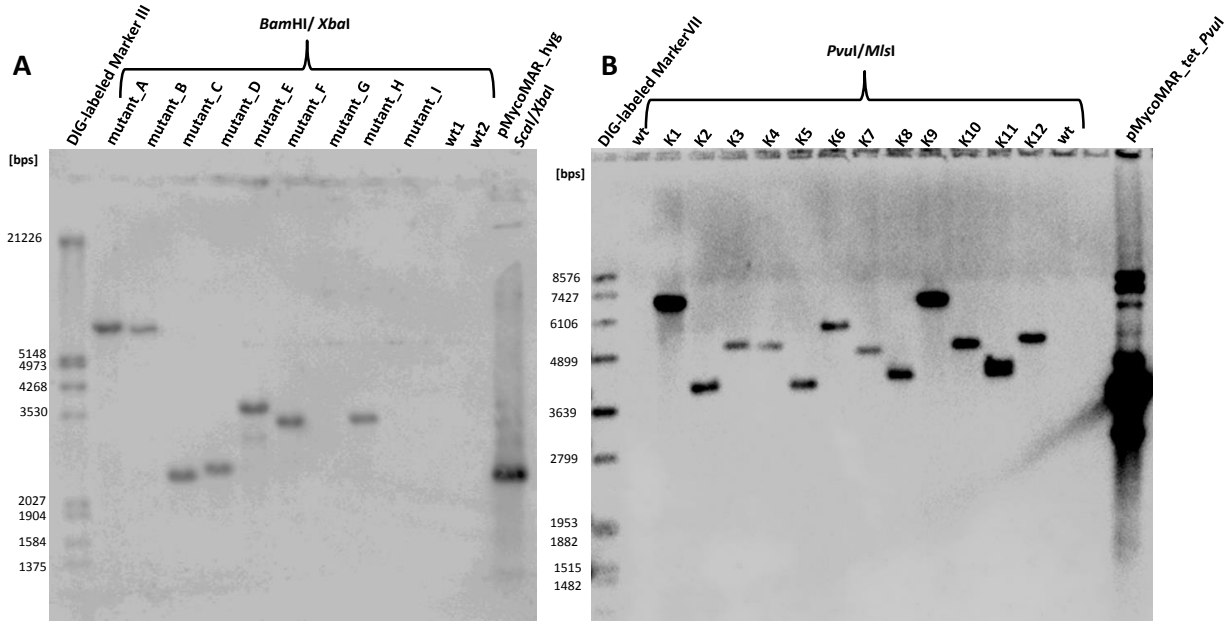


Figure 6: Southern Blot hybridization using genomic DNA of clones resulting from integration of transposable elements of either pMycoMAR_hyg (A) or pMycoMAR_tet (B). **A:** Indicated DNA was digested with *Bam*HI/*Xba*I thereby cutting within the transposable element (3249 bp) and the genome; pMycoMAR_hyg digested with *Sca*I/*Xba*I serves as positive control (2395 bp); probe against *hpt:* hygromycin phosphotransferase. **B:** Indicated DNA (K1-K12: individual mutants) was digested with *Pvu*I/*Mls*I not cutting within the transposable element (3762 bp); pMycoMAR_tet digested with *Sca*I/*Xba*I serves as positive control; probe against *tetA:* tetracycline efflux protein.

As it can be seen from Figure 6A the hygromycin resistance gene (*hpt*) is present in most of the genomes of the analyzed clones thereby proving integration of the transposable element. Only the mutants G and I did not show such a signal which is why these clones were regarded as false positives. In contrast to this, no incorrect mutants (K1-K12) have been obtained using pMycoMAR_tet in combination with oxytetracycline (Figure 6B). The appearance of erroneous clones on hygromycin selection is a well-known problem and might be due to its instability in combination with the detoxification mechanism by hygromycin phosphotransferase (*hpt*). This leads to a continuous decrease of the antibiotic over time what supports the growth of wild-type cells. Moreover, the event of spontaneous resistance displays a second possibility leading to incorrect clones. Nevertheless, hygromycin in combination with the selection marker *hpt* can be used for the construction of genetically altered strains of strain #8405.

In summary, a transposon based gene transfer system using the method of electroporation was established for strain #8405. This represents a promising starting point for the construction of directed mutants via homologous recombination. While the *hpt* gene is used routinely as selection marker for the identification of *Sorangium* mutants the application of *tetA* in

combination with oxy-/tetracycline is only described once in the literature [60]. The availability of two functional selection markers for strain #8405 should permit the creation of double-mutants thereby allowing the examination of a genes function via genetic complementation studies and the analysis of mutants defective in two distinct genes. Other potential selection markers for this strain are the *ble* gene conferring resistance to certain aminoglycosides (bleomycin, phleomycin) or *cat* that detoxifies chloramphenicol. Both were already successfully applied in the genus *Sorangium* [61;62].

Identification of putative transcriptional regulators of the ambruticin cluster

To selectively enhance the production of ambruticins by strain #8405 we aimed to elucidate regulatory mechanisms controlling compound formation. Hence, strain #8405 was subjected to Illumina sequencing and subsequent identification of the ambruticin gene cluster in the draft genome was achieved by employing the DNA sequence of the published *amb* locus from *S. cellulosum* So ce10 (GenBank: DQ897667.1) in a nucleotide BLAST search. A subsequent *in silico* analysis showed that both clusters share >99% similarity and furthermore allowed the identification of putative regulatory elements encoded within the ambruticin cluster (Identification of putative regulatory elements encoded within the ambruticin gene cluster).

As regulatory elements of myxobacterial secondary metabolite pathways are often not encoded within or in close vicinity of the corresponding gene clusters a combination of EMSA and “promotor-fishing” experiments was performed, in order to detect additional putative regulators encoded somewhere else in the genome (Identification of an additional putative transcriptional regulator of the ambruticin gene cluster).

Identification of putative regulatory elements encoded within the ambruticin gene cluster

We aimed to assess the function of genes encoded within the *amb* gene cluster of strain #8405 that could be involved in the regulation of ambruticin biosynthesis. For this purpose the respective DNA sequence was evaluated using the antiSMASH pipeline [63]. In the course of this analysis the previous annotation of all genes by Julien *et al.* [16] in strain So ce10 was confirmed. Moreover, one meaningful open reading frame (669 bp, bold shape in Figure 7) encoding a putative 28 kDa protein featuring a conserved domain (uncharacterized protein family 0158; UPF0158) was additionally detected. Due to its location within the sequence it was termed *amb04a*. As shown in Figure 7 the *amb* gene cluster in strain #8405 comprises 27 genes, of which 9 encode for the core PKS-machinery (black) and an additional 8 for proteins that turn the PKS-product into the respective derivatives (light grey). The remaining 10 cannot directly be linked to ambruticin biosynthesis (dark grey).

Figure 7: Overview of the ambruticin biosynthetic gene cluster in strain #8405. Genes encoding the core PKS machinery are shown in black. Post-PKS tailoring genes: light grey color; genes encoding proteins not directly involved in ambruticin biosynthesis are depicted in dark grey. The additionally detected gene *amb04a* is shown in bold shape.

According to the conserved domains identified in this analysis, three genes namely *amb01* (putative LysR-family transcriptional regulator), *amb04* (putative AtoC like response regulatory protein) and *amb07* (putative hybrid sensor kinase response regulator) might be good candidates for a closer investigation of their possible influence on ambruticin production. In order to shed light on their function, we intended to delete and to overexpress each single gene in separate attempts. A subsequent analysis of the corresponding mutants with respect to their ambruticin production should permit a solid statement whether these loci are involved in the regulation of ambruticin formation or not. Knowledge gained herein will serve as a starting point for a rational optimization of the producer via genetic engineering to obtain a higher productivity.

Identification of an additional putative transcriptional regulator of the ambruticin gene cluster

In order to identify additional putative regulatory elements encoded elsewhere in the genome we followed a similar approach to the one previously described by Rachid *et al.* [28] in which the promoter region upstream of the first structural gene of the chivosazol gene cluster (*chiA*) was used to isolate proteins specifically binding to this DNA-region. Prior to such experiments it should be assessed at which time point during the growth of strain #8405 potential regulatory elements are binding to this region. This was to enhance the probability of the identification of such regulatory proteins. Thus, several cultures were grown in A-medium, whereupon every 24 hours one culture was harvested. This lead to cell pellets spanning a period of 240 hours, from early exponential to the stationary growth phase. Subsequently, cell lysates were prepared, incubated with a DNA fragment containing the presumed promoter region of *ambA* (P_{ambA}) and subjected to EMSA analysis (Figure 8).

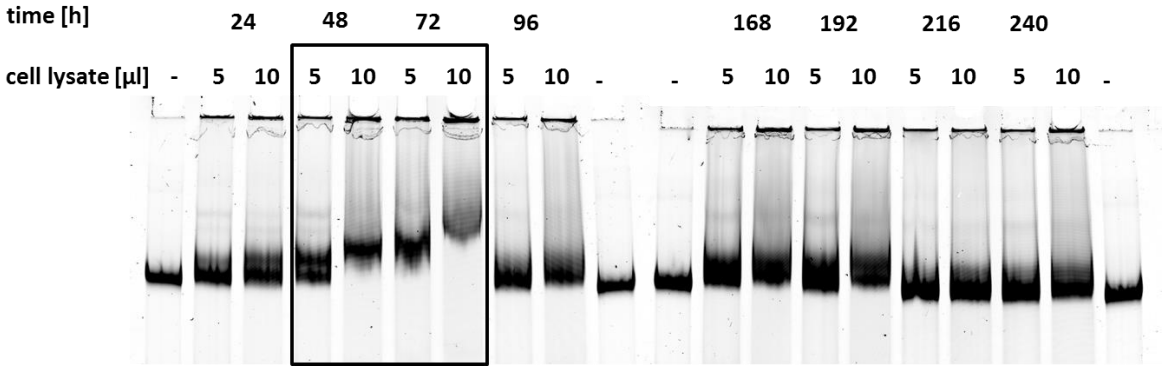


Figure 8: The time dependent occurrence of unknown proteins binding to P_{ambA} during growth of strain #8405. 5 or 10 µl of lysates from cells harvested at the indicated times were incubated in each case with the DNA fragment containing the P_{ambA} region and afterwards separated on a native polyacrylamide gel. Lanes marked with a **–** depict a negative control without lysate. The framed section shows a shift of the DNA fragment upon incubation with lysates from cells grown 48 and 72 hours.

In the context of these experiments we observed a clear shift of the P_{ambA} DNA fragment, which had been previously incubated with cell lysates from cultures harvested after 48 and 72 hours. This indicates binding of one or more proteins, present in the organism at that time, to the DNA fragment. Based on these findings a DNA-protein pull-down experiment with these two lysates was performed. Figure 9 shows the result of the pull-down assay after separation of the samples by SDS-PAGE.

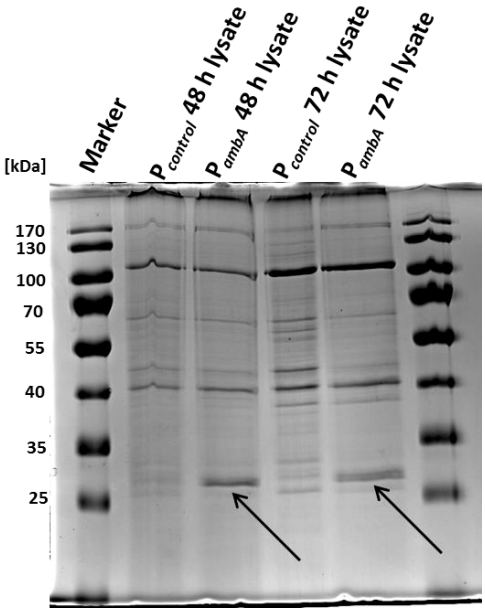


Figure 9: SDS-PAGE of proteins specifically binding to the presumed promoter region of P_{ambA} at 48 and 72 h. Additionally a control experiment using a P_{taA} promoter fragment of *M. xanthus* DK1622 was performed to exclude candidate proteins which result from nonspecific binding (designated $P_{control}$). The arrows highlight proteins that selectively bind to P_{ambA}. These proteins have been cut out of the SDS gel and were identified by MALDI-TOF analysis.

As can be seen in Figure 9, in both cases (48 and 72 h lysate), one protein band is only visible using P_{ambA} as a probe which cannot be seen using the control promoter fragment (marked with an arrow). These bands were cut out of the gel and the proteins were identified by MALDI-TOF analysis. Both proteins were identified as the same NtcA-type regulator. The associated gene was localized in the draft genome of strain #8405. The gene, referred to as *ntcA*, is not encoded in the direct vicinity of the *amb* cluster. Nevertheless, it could be involved in the regulation of ambruticin synthesis. As mentioned above, it is not an unusual observation that regulators of myxobacterial secondary metabolite clusters are encoded elsewhere in the genome. A BLAST analysis employing the protein's amino acid sequence highlighted two conserved domains: an effector molecule binding domain and a helix-turn-helix DNA binding domain. NtcA-type regulators are generally distributed among cyanobacteria [64], where they can act as both, transcriptional activators of genes for nitrate assimilation (e.g. *narB*: nitrate reductase and *nirA*: nitrite reductase), or transcriptional repressors of genes that negatively affect nitrogen incorporation (e.g. *gifA* and *gifB* that encode glutamine synthetase-inactivating factors IF7 and IF17, respectively) [65]. Furthermore, the expression of the *ntcA* gene itself is repressed by high levels of ammonia, what ensures an adequate response to changing environmental conditions. NtcA-type regulators are characterized by the ability to specifically bind to a palindromic sequence signature for which the consensus motif $GTAN_8TAC$ has been defined [66;67].This motif can also be found on the P_{ambA}-bearing DNA fragment used in EMSA and pull-down assays. As mentioned above, an NtcA-type regulator which negatively influences secondary metabolite formation in So ce56 has been previously identified by using an experimental setup, comparable to our experiments [29]. A primary sequence alignment of $NtcA_{\#8405}$ and $NtcA_{So\ ce56}$ revealed more than 95 % sequence identity and moreover the genomic context of those genes in the respective chromosomes is strikingly similar. Conferring these facts to the circumstances of ambruticin formation in strain #8405, the targeted inactivation of *ntcA* might lead to formation of elevated levels of ambruticin.

In summary, an NtcA-like transcriptional regulator, similar to the one previously identified in So ce56, binding to P_{ambA} could be identified by the use of a combinatory approach consisting of EMSA and DNA-protein pull-down investigations. The first experiment allowed the determination of the time point of the appearance of proteins binding to P_{ambA} and the latter one led to the identification of the DNA-binding protein itself. At this stage, genetic experiments needed to be performed in order to examine the potential role of NtcA in ambruticin biosynthesis in strain #8405 (see below).

Inactivation of *ntcA* via in-frame deletion or gene replacement

As pointed out in the previous section, EMSA experiments in combination with a DNA-protein pull-down identified an NtcA-like protein as potential regulator of ambruticin biosynthesis. To clarify this, we aimed first to inactivate the *ntcA* gene. The commonly used method for directed gene inactivations in the genus *Sorangium* is the introduction of a suicide plasmid, harboring a truncated (and mutated) version of the target gene, into the genome of the recipient [68;69]. Genomic integration occurs via single cross-over recombination which usually needs a homology fragment of around 1.0 kb in size to take place efficiently. Eventually, this strategy results in two disrupted copies of the gene, enabling in most cases the analysis of the gene's function. This procedure is the method of choice to rather quickly inactivate genes with a size of more than 1.5 kb e.g. genes encoding full NRPS or PKS modules. Especially, if the size of the target gene is less than 1.0 kb, a directed disruption becomes difficult. On the one hand, the homology size of the fragment must be large enough to permit reasonable recombination efficiencies, and on the other hand it must be small enough to ensure both gene copies to be disrupted.

Since *ntcA* has a size of 933 bp, a gene disruption by single cross over was not taken into account. We therefore attempted to develop an in-frame deletion system which could be used for seamless deletion of this gene. Such a system is already available for the myxobacteria *S. aurantiaca* DW4/3-1 [70], *M. xanthus* [71], and *M. fulvus* Mx f50 [72]. In all cases, the strategy is based on the "pop-in pop-out" method using the *sacB* (levansucrase) gene from *B. subtilis* as counterselection marker [73]. However, the application of *sacB* is often inefficient and the frequent occurrence of false positives clones makes an excessive screening for clones which have undergone excision of the plasmid indispensable. To circumvent this issue, we aimed to establish a reliable method for the visible detection of clones having undergone the second single cross-over likewise to the utilization of *gusA* (beta-D-glucuronidase) in actinomycetes [74]. Hence, strain #8405 was grown on M-agar plates containing either 5-Bromo-4-chloro-3-indolyl β-D-glucuronide (X-Gluc), the substrate of GusA or 5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-Gal), the substrate of LacZ (beta-D-galactosidase) in order to exclude any intrinsic activity of strain #8405 towards these substrates. As a result, no blue precipitate was visible indicating that both systems might be applicable. Finally, *lacZ* from *E. coli* was chosen over *gusA*, because the use of *lacZ* in *M. xanthus* is well established [75;76] and therefore its functional expression in the genus *Sorangium* was considered to be more likely. Consequently, pSM_lacZ_hyg was constructed containing a P_{aphII} driven hygromycin resistance (*hpt*) and *lacZ*

controlled by the T7A1 promoter. This vector served as basis for the construction of numerous plasmids used in this study (see below).

To delete *ntcA* in the genome of strain #8405, plasmid pSM_lacZ_hyg_ntcA_up+dn1 was constructed featuring a truncated version of *ntcA* (6 bp of the respective 5' and 3' ends separated by a 6 bp *Xba*I restriction site) flanked on each side by approx. 1.7 kb of ithe genomic region adjacent to *ntcA*. Consequently, the integration and subsequent excision of this plasmid from the chromosome will either yield in a 921 bp in-frame deletion within *ntcA* or in a revertant to the wt allel. After the transfer of pSM_lacZ_hyg_ntcA_up+dn1 into strain #8405, the gDNA of two of the obtained hygromycin resistant clones was isolated and integration of the introduced DNA was verified by Southern Blot analysis (Figure 10).

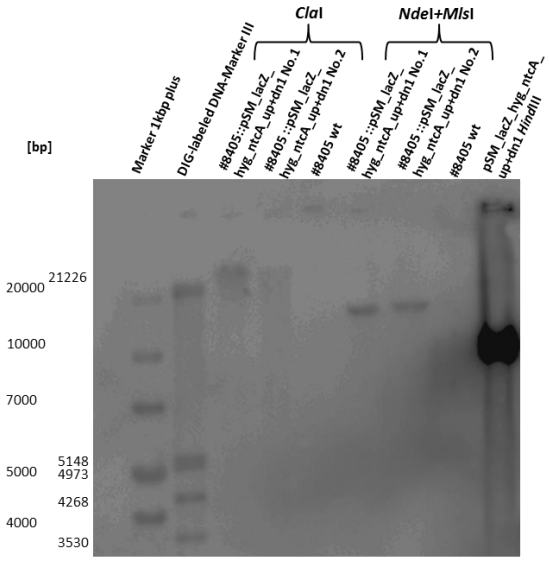


Figure 10: Southern Blot analysis to verify integration of pSM_lacZ_hyg_ntcA_up+dn1. A photograph of the Southern Blot hybridization with a probe against *hpt:* hygromycin phosphotransferase using genomic DNA of the wild type and of clones resulting from integration of pSM_lacZ_hyg_ntcA_up+dn1 into the genome of strain #8405 is shown. All DNAs have been digested with *Cla*I or *Nde*I/*Mls*I, respectively. Expected signal sizes are: genomic integration via up fragment *Cla*I: 17736 bp, *Nde*I/*Mls*I: 15231 bp; genomic integration via down fragment *Cla*I: 16821 bp, *Nde*I/*Mls*I: 14316 bp. pSM_lacZ_hyg_ntcA_up+dn1 cut with *Hind*III as positive control: 10411 bp.

In case of the *Cla*I digested DNAs, the signals appear a little bit too high but this might be caused by the poor electrophoretic separation, whereas the samples cut with *Nde*I/*Mls*I show signals of the expected sizes. Unfortunately, this investigation does not allow a definite statement if the integration of the plasmid into the chromosome of strain #8405 occurred via homologous recombination with the upstream or the downstream fragment flanking *ntcA*. A subsequent experiment using a probe for the up region showed integration via the down region for clone No.1 (Figure 12).

After proving the correct integration of pSM_lacZ_hyg_ntcA_up+dn1 in the genome of strain #8405, it was tested if the recombinant strains would now be able to hydrolyze X-Gal. Thus strain #8405*::*pSM_lacZ_hyg_ntcA_up+dn1 No.1 and No.2 were cultured on apportioned M-agar plates with and without 80 µg/ml X-Gal. Upon incubation for two days at 30°C a blue precipitate was formed on plates containing X-Gal. This can be attributed to the auxiliary acquired β-galactosidase activity of these strains (Figure 11).

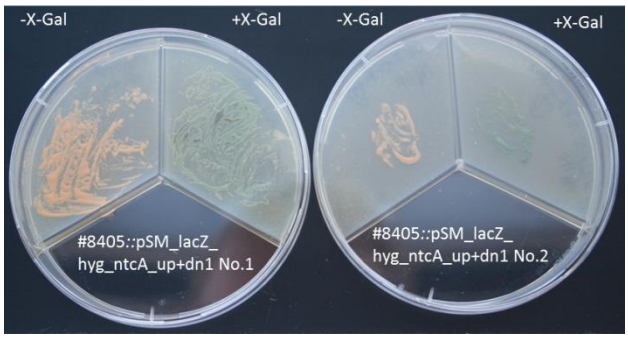


Figure 11: Cultivation of strain #8405*::*pSM_lacZ_hyg_ntcA_up+dn1 No.1 and No.2 on solid M-agar with and without X-Gal. The blue precipitate is caused by hydrolytic activity of LacZ on X-Gal.

According to these results, LacZ activity can be directly correlated to the presence of a pSM_lacZ_hyg derivative in the chromosome of any strain #8405 strain. Furthermore, the *lacZ*-marker should be beneficial to sort out false positive clones after transformation of a pSM_lacZ_hyg based plasmid. In this case blue colonies are the correct clones, orange ones are false positives. For this reason the *lacZ*-marker should also be suitable for the definite identification of clones which have undergone the second recombination event (to yield the desired in-frame deletion or the wild type revertant), which is accompanied by a loss of the ability to hydrolyze X-Gal.

In order to enable the second cross over event to obtain the desired in-frame deletion within *ntcA*, strain #8405*::*pSM_lacZ_hyg_ntcA_up+dn1 No.1 was cultivated in M-medium without the addition of hygromycin B. After subcultivating this strain 24 times in liquid M-medium (approx. 90 days), orange clones were obtained on solid M-medium containing X-Gal. As pointed out earlier, the orange phenotype may indicate two possible genotypes: revertant to wild type or in-frame deletion mutant. For further analysis, selected clones were cultivated, genomic DNA was isolated, and subjected to Southern Blot analysis (Figure 12).

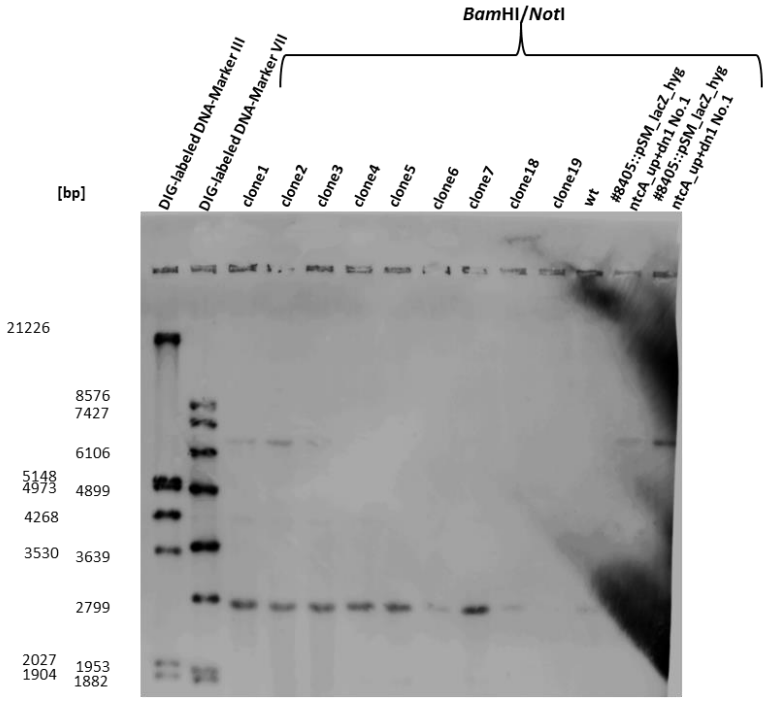


Figure 12: Southern Blot analysis to verify plasmid excision. A photograph of the Southern Blot hybridization of genomic DNA (in each case digested with *Bam*HI/*Not*I) of cured, orange colonies, the wild type, and strain #8405*::*pSM_lacZ_hyg_ntcA_up+dn1 No.1 with a probe against the up region is shown. Expected sizes are as follows: integration via down fragment: 6679+2752 bp (still integrated plasmid); in-frame deletion: 4660 bp; wt-genotype: 2752 bp.

As it can be concluded from Figure 12, genomic DNA of the clones 1-3 led to signals identical to genomic DNA from strain #8405*::*pSM_lacZ_hyg_ntcA_up+dn1 clone 1, suggesting that at least a part of the cells still harbor the integrated plasmid backbone. The residual strains seem to be cured completely because they only show one signal corresponding to the wild type genotype. This experiment clearly shows that double cross over is possible in strain #8405. Unfortunately, only revertant clones and no clones carrying an in-frame deletion of *ntcA* could be obtained, which might be due to the relatively small amount of obtained putative mutants which were analyzed. Another possibility why no in-frame mutants were obtained might be that the deletion of *ntcA* negatively influences the viability of strain #8405 under the applied conditions. In order to exclude this scenario, at least two additional cultivation settings should be tested. However, during the rather long period of cultivation until orange clones were obtained (approx. 90 days), the accumulation of undesired secondary mutations is most likely. Therefore, this system might not be appropriate to reliably study the effect of a single mutation, and even less for the construction of multiple mutants. Since from this data it seemed that the first three clones are still carrying pSM_lacZ_hyg_ntcA_up+dn1 although they do not turn blue on X-Gal media, we hypothesized that these “cured” cultures still contain small amounts of cells which have not excised the plasmid yet. To test the purity of these cultures we did replica plating with all

obtained clones on media containing X-Gal and on media which was moreover supplemented with hygromycin B (Figure 13).

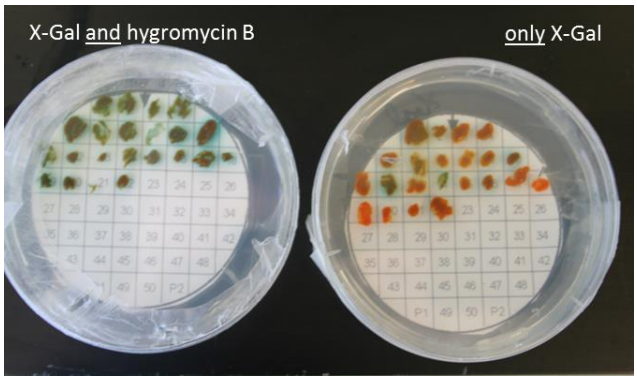


Figure 13: Test for culture homogeneity of cured mutant strains. A photograph of the replica plating of clones analyzed by Southern Blot as well as of additional cured clones which have undergone a second single cross over event is shown.

In contrast to our initial assumption, not only the clones 1-3 are "contaminated" with such cells that still harbor pSM_lacZ_hyg_ntcA_up+dn1 integrated in the genome, but rather all cultures seem to contain a certain amount of such cells. This is indicated by the fact that most of these cultures only turn blue on media additionally containing hygromycin B. To tackle this issue, pSM_lacZ_hyg_ntcA_up+dn1 was expanded by the constitutively expressed *sacB* gene (driven by P_{T7A1} in a similar manner to *lacZ*) thereby creating pSM_lacZ_hyg_sacB_ntcA_up+dn1. Initially we hoped that this combination of a lethal (*sacB*) and a visible (*lacZ*) marker should facilitate effective selection for double cross over clones, but after this plasmid was integrated into the genome of strain #8405 (as verified by Southern Blot analysis, Figure 14) the corresponding mutants and the wild type were still able to grow on media containing sucrose up to 9% (v/v). Thus, *sacB* was considered not to be appropriate for counterselection in the genus *Sorangium*.

In summary, we can state that in principle the "pop-in pop-out" method can be used in *Sorangium* species. Unfortunately, only wild type revertants but no mutants containing an in-frame deletion in *ntcA* have been obtained. This might be caused by the applied culture conditions. The fact that all cultures which have undergone the second single cross over event still contained portions of the parent strain pinpoints the need for a lethal counterselection marker for the genus *Sorangium*. Among the applied markers, only *lacZ* is functional while *sacB* seems to be not. Moreover, *lacZ* might rather be used in the future as indicator for the genomic integration of exogenous DNA into the genome of strain #8405 than for the detection of its

excision. Additionally, LacZ might as well be used for determination of promoter activities or any other related applications.

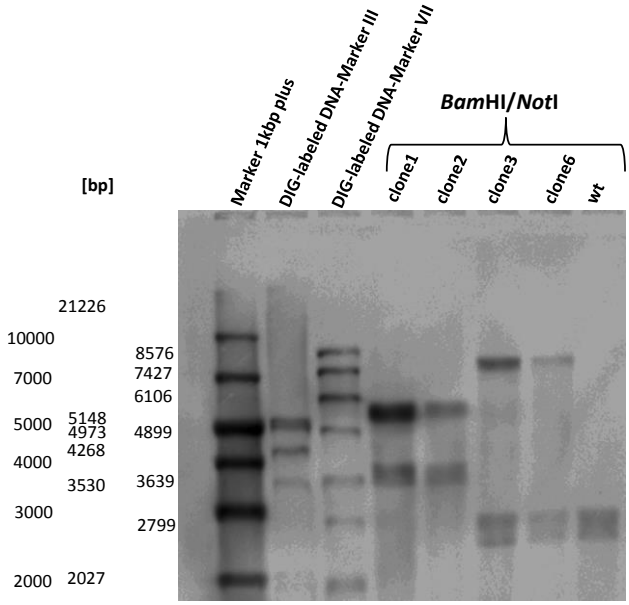


Figure 14: Southern Blot analysis to verify integration of pSM_lacZ_hyg_sacB_ntcA_up+dn1. A photograph of the Southern Blot hybridization with genomic DNA (cut with *Bam*HI/*Not*I) from the wild type and clones obtained after transferring pSM_lacZ_hyg_sacB_ntcA_up+dn1 into strain #8405 is shown. Expected sizes are as follows: integration via up fragment: 3876+5700 bp; via down fragment: 2752+7614 bp; wt: 2752 bp.

In addition to the “pop-in pop-out” gene deletion strategy described above, the methodology of gene replacement by the use of linear DNA depicts an additional and widespread procedure for the deletion of a designated chromosomal region in a variety of bacteria [77-79], plants [80;81], yeast [82], filamentous fungi [83-85], and mammalian cells [86;87]. For this purpose, a linear DNA fragment is assembled which contains a selection marker flanked by segments homologous to the proximate up and downstream region of the target gene. This fragment is then transferred into the respective organism where it integrates via double cross over at a specific site, predefined by the corresponding homologous regions. Hereby, the native DNA sequence is replaced with the exogenous DNA, which is why the method is referred to as gene replacement. Notably, mutants generated this way are stable what makes perpetuation of selective conditions dispensable. To achieve reasonable integration efficiencies, usually a minimum of around 1 kb of homologous DNA to the target site is essential but in some cases it can be significantly reduced by making use of the λ recombination system [83;88-90].

As we failed to generate in-frame deletion mutants using the “pop-in pop-out” method, we wanted to establish the use of linear DNA for gene replacement in the genus *Sorangium* in order

to allow a deletion of *ntcA* to be made. Therefore the, immediate up- and downstream regions flanking the *ntcA* gene were cloned into pSM_lacZ_hyg in a way that they frame *hpt* and *lacZ*. This plasmid was named pSM_lacZ_hyg_ntcA_up+dn2. After restriction with *Hind*III and *Nsi*I, the respective fragment was recovered from an agarose gel and transferred into strain #8405 (Figure 15 A).

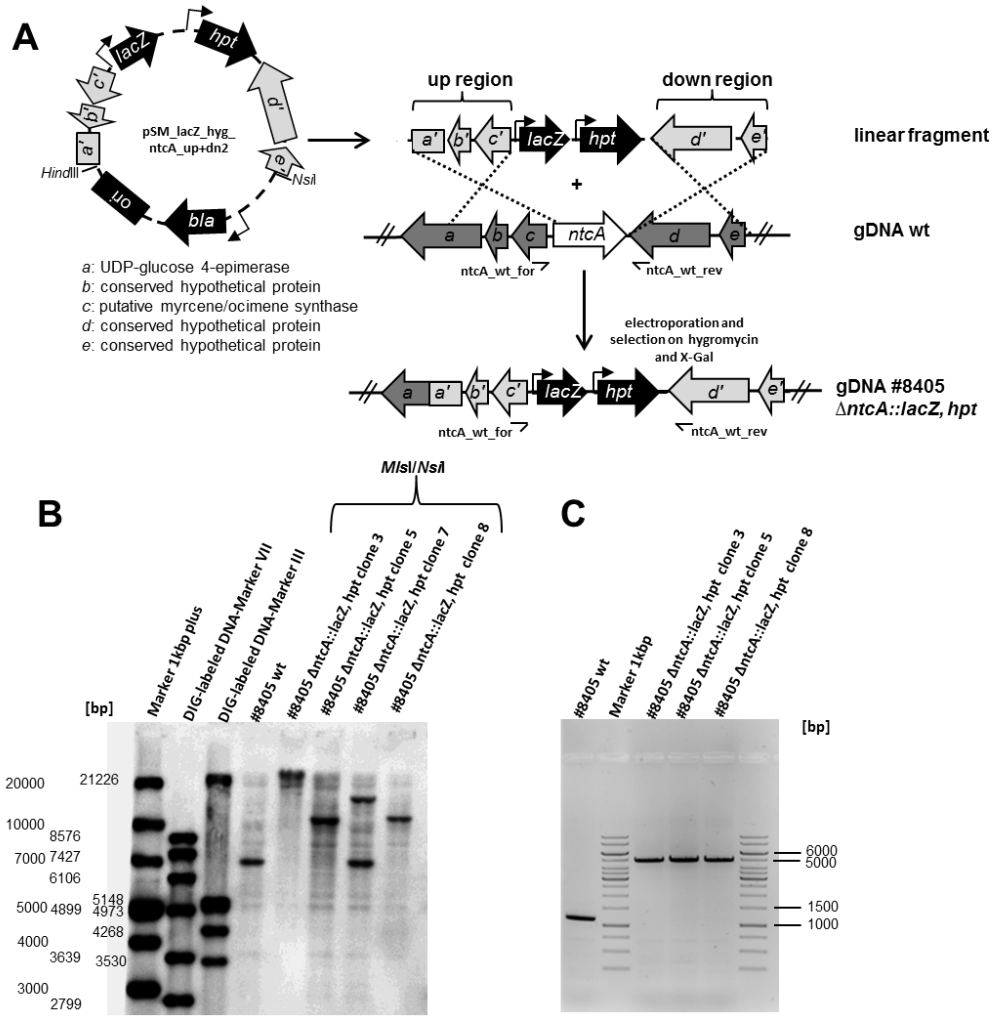


Figure 15: Construction and verification of Δ*ntcA* deletion strains. A: Scheme of the *ntcA* deletion by gene replacement; *bla*: beta-lactamase, *hpt:* hygromycin phosphotransferase, *lacZ:* beta-galactosidase, ori: origin of replication. **B**: Genotypic verification of ΔntcA::lacZ,hpt mutants by Southern Blot analysis using a probe against the up-fragment. Expected sizes are as follows: wt-allel: 6785 bp; replacement of *ntcA*: 10431 bp. **C**: Genotypic verification of strain #8405ΔntcA::lacZ,hpt mutants by PCR using ntcA_wt_for and ntcA_wt_rev as depicted in A. Expected sizes are as follows: wt-allel: 1138 bp, replacement of *ntcA*: 4784 bp.

According to the Southern Blot analysis (Figure 15 B), clone 5 as well as clone 8 have incorporated the linear fragment into their genomes, thereby replacing *ntcA*. In the course of this experiment, a valid verification for clone 3 failed, because its DNA was not properly separated.

However, a subsequent PCR targeting the site of integration proved the deletion of *ntcA* (Figure 15 C). Noteworthy, the examination of clone 7 by Southern Blot led to two signals. This is caused by the integration of a circular molecule of pSM_lacZ_hyg_ntcA_up+dn2 resulting from incomplete digestion of the plasmid. In conclusion, these investigations clearly showed that selective deletions in the chromosome of strain #8405 are possible by applying the method of gene replacement.

In this context, it should be noted that the use of linear DNA for targeted gene replacement depicts a novel and rather uncomplicated way for a reliable analysis of a gene's function in the genus *Sorangium*. This procedure allows the deletion of designated chromosomal regions in *Sorangium* and might replace the "traditional" method of gene disruptions by single cross-over in strains which are amneable to a DNA transfer by electroporation.

Ambruticin production in a *ntcA*⁻ mutant

After confirming the deletion of *ntcA* in the chromosomes of clones 3, 5 and 8, their ambruticin production rates were analyzed and compared to the wild type (Figure 16).

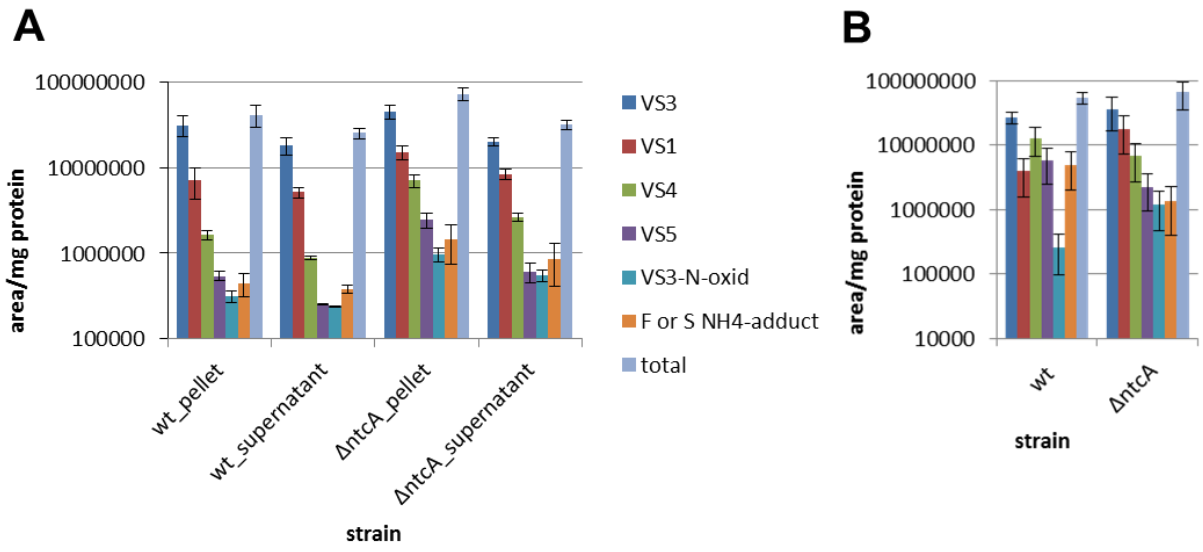


Figure 16: Phenotypic analysis of the wild type and the Δ*ntcA* mutants with respect to their capability of ambruticin production. A: Differential analysis of the cell pellet and the supernatant of wild type and mutant (Δ*ntcA*) cultures. **B:** Earlier analysis of the wild type and the mutants cell pellets. Given values in A and B are means of three independent cultures. Ambruticin titers are presented relative to the protein content of the samples.

In general, the differential analysis of the cell pellets and the culture broth (supernatant) (Figure 16 A) showed significant differences between the wild type and the Δ*ntcA* mutants. While

an evaluation of the ambruticin titers in the supernatant of the wild type and the mutants revealed slightly elevated levels of some derivatives (VS1, VS4, VS5 and VS3-N-oxid) in the mutants, the intracellular ambruticin content of the mutants is significantly increased (about 40%) when compared to the wild type. These results are ambiguous since in an earlier analysis the production rates of the mutants displayed no noteworthy changes (Figure 16 B). However, and in accordance with the high degree of similarity between $NtcA_{\#8405}$ and $NtcA_{So\ ce56}$ the rise in ambruticin production upon the deletion of *ntcA* in strain #8405 was expected. To clearly confirm that mutants defective in *ntcA* are in fact producing higher amounts of ambruticin, when compared to the wild type, additional cultivations need to be performed and analyzed. In addition to the knockout studies described above, a directed overexpression of the *ntcA* gene by exchanging its native promoter against the strong and constitutive T7A1 promoter is ongoing. Unfortunately, this experiment could not be finished in the course of the present work. On the basis of the current evidence, mutants overexpressing *ntcA* are likely to produce either fewer or equal amounts of ambruticin in comparison to the wild type, depending on the degree of repression NtcA exerts.

In summary, a targeted deletion of *ntcA* by gene replacement and the subsequent analysis of mutants suggests that NtcA is a negative regulator of ambruticin biosynthesis in strain #8405, similar to NtcA in So ce56. However, a directed overexpression of this gene is still pending and will give further insight into the regulation of ambruticin biosynthesis in the future.

Amb01: Deletion, overexpression and effects on ambriticin production of a putative LysR-family transcriptional regulator

The *amb01* gene is located on the antisense strand at the 5'-end of the *amb* gene cluster (Figure 7) and encodes a putative LysR-family transcriptional regulator. This type of regulator is ubiquitous among prokaryotes and characterized by an N-terminal helix turn helix DNA binding domain, as well as a C-terminal domain for binding of an inducer substrate [91]. The best characterized protein of this family and also its namesake is LysR from *E. coli*, which acts as transcriptional activator of *lysA* (diaminopimelate decarboxylase) in the presence of diaminopimelate [92;93]. Regulators belonging to this family were originally described to positively regulate divergently encoded genes e.g. AmpR activates *ampC* (beta lactamase) and CatR or ClcR activates the *catBC* or *clcABD* operon, respectively [94;95], but more recent studies suggest that LysR-like proteins may act as global regulators that can stimulate but also repress the transcription of target genes [96;97]. Common to all these regulators is that the

respective inducer substrates (usually an educt, product or intermediate of the regulated pathway) are usually required in order to allow transcriptional activation or repression [98-100].

In the case under consideration this could mean, that Amb01 is either activating or repressing transcription of genes associated with ambruticin production upon binding of an adequate substrate. In the event of repression that substrate might be ambruticin itself, whereupon a targeted deletion or overexpression of *amb01* should result in an enhanced or decreased production. By contrast, if Amb01 activates transcription of the *amb* genes, it might bind a precursor (e.g. malonyl- or methymmalonyl-CoA) and consequently a deletion would result in a decrease and an overexpression possibly in an increase of ambruticin formation. Again, a prerequisite for these scenarios is the presence of the inducer substrate.

To test these hypotheses we aimed to delete and to overexpress *amb01* in separate attempts following phenotypic characterization of corresponding mutants. For a selective deletion by gene replacement plasmid pSM_lacZ_hyg_amb01KO_up1+dn1 was constructed and afterwards linearized with *Hind*III/*Nsi*I. Thereafter, it was transferred into strain #8405 and two resulting clones (L1_3 and L1_5) were examined by Southern Blot analysis (Figure 17). As it can been seen from Figure 17 B, the wt-signals perfectly correspond to expected ones, but the signals obtained with DNA of the mutants do not fit in both cases. Therefore, the mutants are regarded as false positives. Several attempts to generate correct mutants defective in *amb01* were undertaken, but no additional clones have been obtained yet.

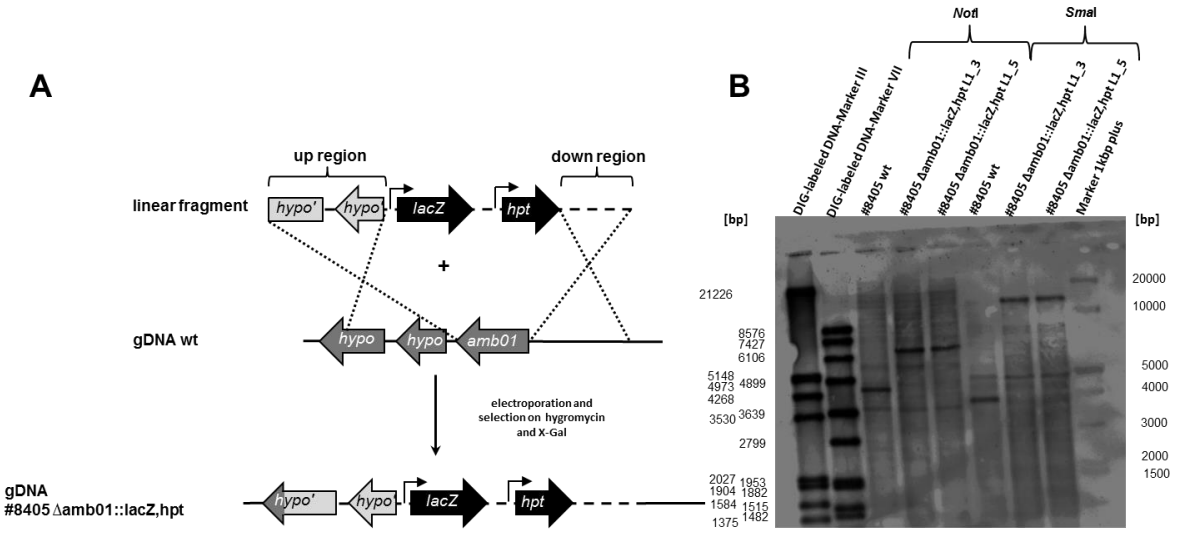


Figure 17: Construction and verification of ∆*amb01* deletion strains. A: Scheme of the gene replacement strategy for deletion of *amb01; hpt:* hygromycin phosphotransferase, *hypo:* hypothetical protein, *lacZ:* beta-galactosidase*.* **B**: Genotypic verification of strain #8405 ∆amb01::lacZ,hpt mutants by Southern Blot analysis using a probe against the down-region. Expected sizes are as follows: *Not*I wt: 4338bp; ∆amb01: 7606bp; *Sma*I wt: 3778bp; ∆amb01: 7370bp.

Besides *amb01* deletion, we also attempted to exchange the promoter in its upstream region against the constitutive T7A1 promoter. Therefore, plasmid pSM_lacZ_hyg_T7A1_amb01 was constructed and transferred into strain #8405. This approach resulted in a meroploid strain with two functional copies of this gene, one under the control of the native promoter and a second one whose expression is driven by T7A1 (Figure 18).

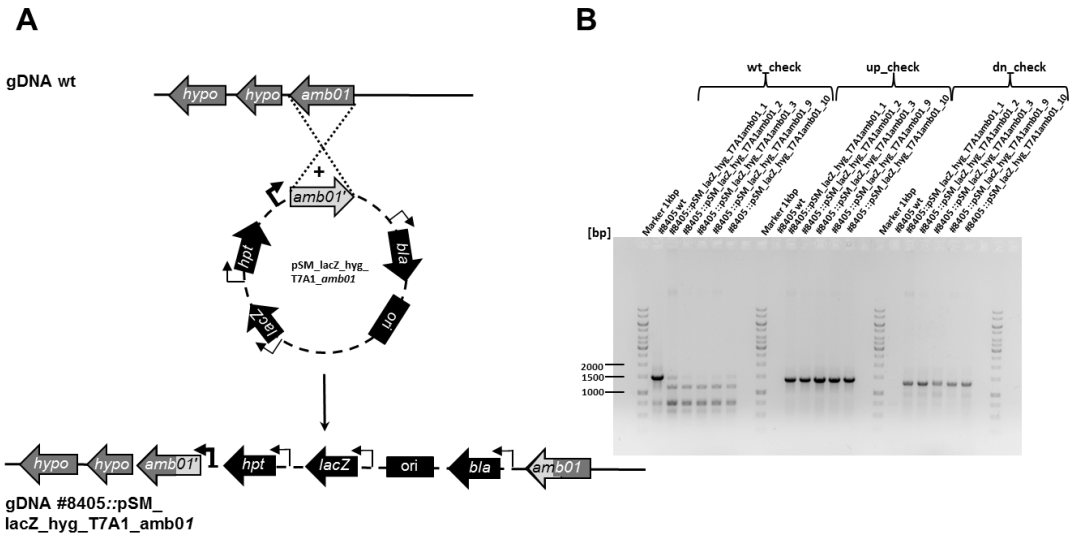


Figure 18: Construction and verification of strains overexpressing *amb01*. A: Schematic depiction describing the promoter exchange strategy for overexpression of *amb01; bla:* beta-lactamase *hpt:* hygromycin phosphotransferase, *lacZ:* beta-galactosidase, ori: origin of replication. **B**: Verification of resulting mutants strain #8405::pSM_lacZ_hyg_T7A1_amb01 Expected sizes are as follows: wt_check: 1465 bp; up_check: 1371 bp; dn_check: 1296 bp).

For phenotypic characterization three clones (1, 9 and 10) were cultivated and analyzed with respect to their ambruticin content (Figure 19).

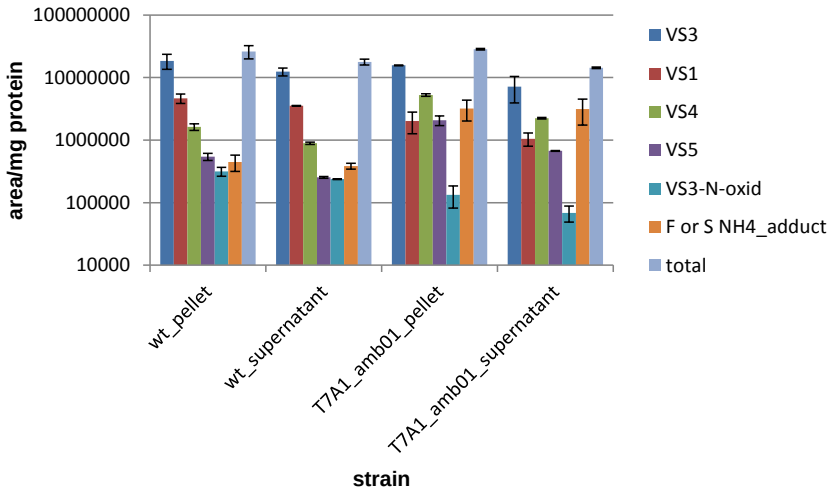


Figure 19: Phenotypic analysis of the T7A1_amb01 mutants and the wild type with respect to their capability of ambruticin production. Given values are means of three cultivations. Ambruticin titers are presented relative to the protein content of the samples.

According to this evaluation, the introduction of a second and constitutively expressed copy of *amb01* has no effect on ambruticin production under the conditions applied. Due to this result it can be concluded, that the expression of this gene is not a limiting factor for ambruticin biosynthesis in the wild type. Unfortunately, no deletion mutants, which will give further insights into the function of *amb01,* were obtained yet. Hence, additional attempts to generate such mutant stains should be undertaken.

Amb07: Deletion, overexpression and effects on ambruticin production of a putative hybrid sensor kinase response regulator

A BLAST analysis using the amino acid sequence of Amb07 as a query classified this protein as putative group III hybrid sensor kinase response regulator. Typically, proteins belonging to this class are characterized by the presence of several N-terminal, so called HAMP domains, a histidine kinase phospho-acceptor domain (HisKA), a histidine kinase domain (HATPase c) and a response regulator receiver domain [101]. In particular, they recognize a stimulus whereupon the ATPase domain transfers a phosphate group to a conserved His residue of HisKA which is eventually forwarded to a specific aspatate of the REC domain [102]. **HAMP** domains are repetitive sequences occurring in a wide range of proteins such as **h**istidine kinases, **a**denylcyclases, **m**ethyl accepting chemotaxis proteins and **p**hosphatases, all associated with signal transduction processes [103]. Bacterial signaling proteins like the osmosensor EnvZ from *E. coli* usually possess one single HAMP domain [104].

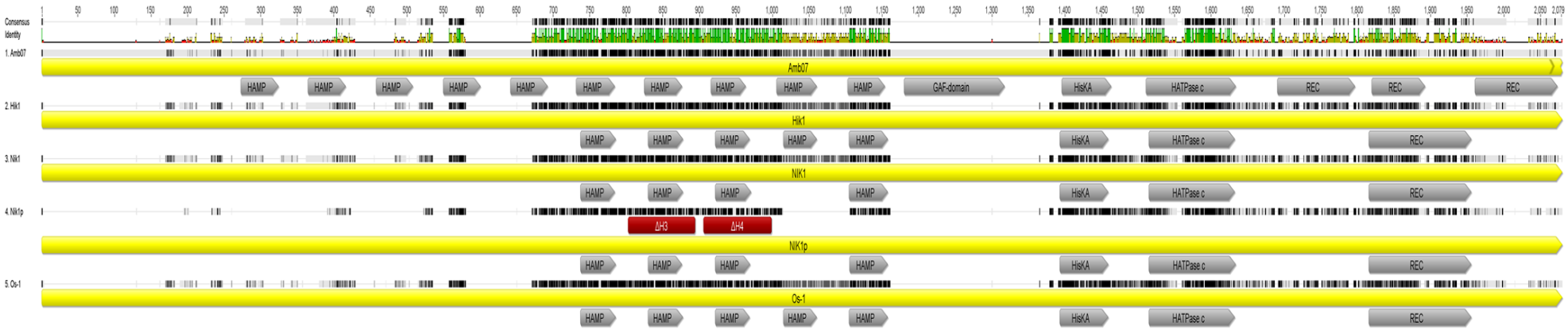


Figure 20: Geneious alignment of the primary sequences of group III hybrid sensor kinase response regulators. Amb07 (ABK32275) from strain #8405, Hik1 (BAB40947) from *Magnaporthe grisea,* Nik1 *(*AAU10319) from *Alternaria brassicicola,* Nik1p *(*AAC72284) from *Candida albicans* and Os-1 (EAA35235) from *Neurospora crassa*. Grey regions highlight the respective domains: HAMP: HAMP domain, HisKA: His Kinase A (phospho-acceptor) domain, HATPase c: Histidine kinase domain, REC: Response regulator receiver domain. Red areas designated ΔH3 and ΔH4 depict those regions of Nik1p which have been deleted in the study published by Buschart *et al.* [12].

Contrary to this, group III sensor kinases involved in osmoregulation of several filamentous fungi (e.g. Hik1, Nik1 Nik1p and Os-1), the designated targets of several fungicides [8-10], contain multiple consecutive HAMP domains [105], an observation which applies as well to Amb07 from strain #8405. Hence, a primary sequence alignment of Amb07 and the respective fungal proteins was performed (Figure 20). This analysis showed significant homology among these sequences, especially in the range of the HAMP, the kinase and partly the receiver domain(s). Remarkably, Amb07 possesses 10 HAMP domains of which the rear ones share 65.3 % sequence identity to those found in fungi, suggesting a perception of a related signal. A study by Buschart *et al.* aimed to address the function of these domains [12]. In detail, they used a yeast sensitivity model expressing truncated versions of Nik1p from *Candida albicans* in which either one (ΔH3 or ΔH4) or two (ΔH3 and ΔH4) of the HAMP domains were precisely deleted (Figure 20). These strains were treated with several fungicides (ambruticin VS3, jerangolid A, fludioxonil and iprodione) and their survival rates were compared to an untreated control.

Although differences in the viability of the cells in dependence of the fungicide which was used were observed, they stated: “Any deletion examined led to a decrease in fungicide susceptibility”. These results are pointing towards an interaction of the fungicides tested with the HAMP domains of Nik1p, suggesting that these regions might be the actual molecular target of those compounds. Consequently, the corresponding HAMP domains of Amb07 might be able to bind ambruticin too, and therefore this protein may function as an ambruticin sensor in strain #8405. A difference between Amb07 and the fungal proteins is the presence of an extra 5 HAMP domains at its N-terminus. This part might be involved in sensing ambruticin as well. In a similar manner to the HAMP domains, Amb07 contains three receiver domains, whereas in the fungal proteins just one of those domains is found. Besides that, Amb07 possesses one domain, which is not found at all in the other sensors: a **GAF**-domain. This conserved domain is named after the proteins in which it occurs: cyclic **G**MP-regulated nucleotide phosphodiesterases, **a**denylyl cyclases and the prokaryotic transcription factor **F**hlA and it is known to bind next to others cyclic nucleotide monophosphates such are cGMP or cAMP what effects enzymatic activity [106;107].

To evaluate a possible function of Amb07 in regulation of ambruticin biosynthesis we aimed to generate deletion mutants as well as strains overexpressing the gene. For selective deletion of *amb07* plasmid pSM_gfp_tet_amb07_KO_up1+dn$_{neu}$ has been constructed. Likewise to the previous gene deletions using linear DNA, the plasmid was digested (*Xba*I/*Nsi*I), the respective linear fragment was gel recovered, and subsequently introduced into the genome of strain #8405 (Figure 21 A). The resulting clones were further grown in selective medium and their genomic DNA was isolated, digested and subjected to Southern Blot analysis (Figure 21 B). As reasoned from genotypic verification of the mutants, only strain L7_9 showed the correct signal pattern. The remaining ones may result from single cross integrations of the intact plasmid or an insertion of the linear fragment at the wrong locus. The poor frequency of correct mutants can most likely be attributed to the fact that a rather large genomic fragment (6206 bp) was deleted leaving only the first 58 bp of *amb07*. For phenotypic characterization this strain was grown in triplicates, methanolic extracts of whole cells and the supernatant were prepared and analyzed by HPLC-MS (Figure 22). This examination showed a 10 fold decrease in overall ambruticin production in comparison to the wild type. Thus, Amb07 should be involved in the regulation of ambruticin formation but up to now its actual function is unclear. Auxiliary evidence that Amb07 plays a crucial role in this system is deduced from the finding that the corresponding gene is found in each *amb*/*jer* gene cluster at a similar position in all genomes of ambruticin and jerangolid producing strains sequenced to date (strain #8405, strain #7282, and strain #1795).

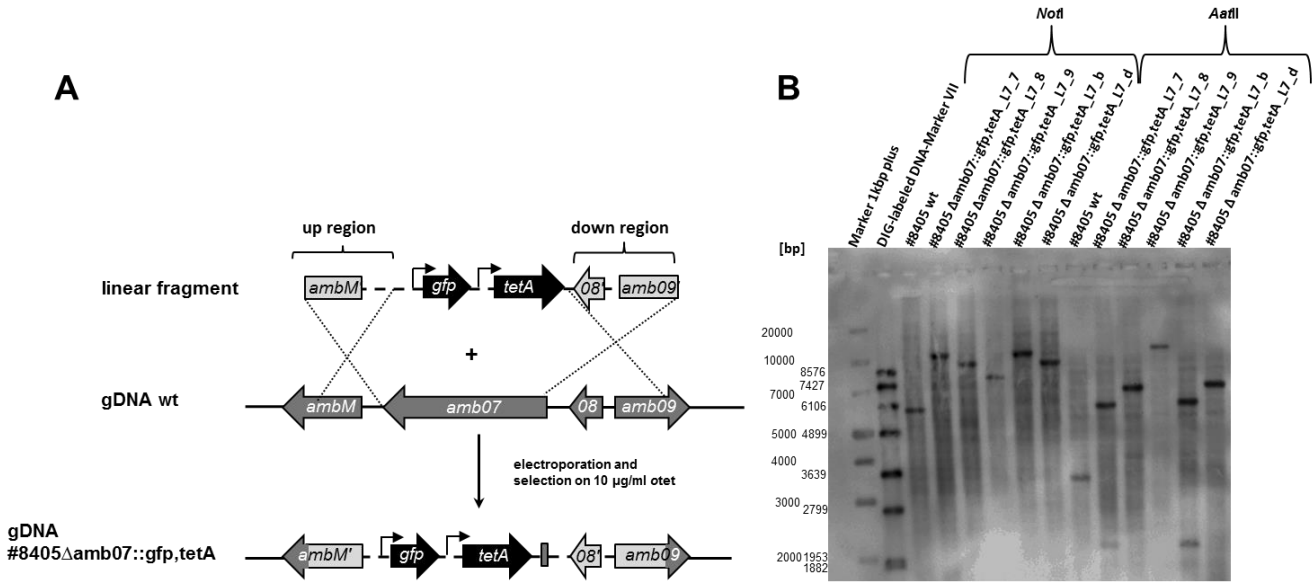


Figure 21: Construction and verification of Δ*amb07* deletion strains. A: Scheme describing the gene replacement strategy for deletion of amb07; *gfp*: green fluorescent protein, *tetA*: tetracycline efflux protein. **B**: Genotypic verification of Δamb07::gfp,tetA mutants by Southern Blot analysis using a probe against the down-region. Expected sizes are as follows: *Not*I wt: 5734 bp; Δ*amb07*: 7742 bp; *Aat*II wt: 3447 bp; Δ*amb07*: 10617 bp.

A plausible explanation for this drop in production cannot be given at this time, because the factor interacting with Amb07 is not known so far, but should be a subject of further investigations in order to get insights into the underlying regulation cascade.

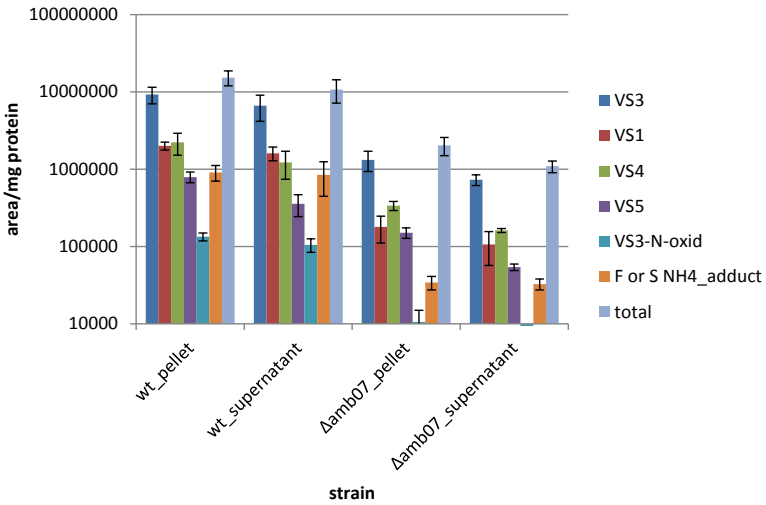


Figure 22: Phenotypic analysis of the Δ*amb07* mutant and the wild type with respect to their capability of ambruticin production. Given values are means of three independent cultivations. Ambruticin titers are presented relative to the protein content of the samples.

Along with deletion experiments, a T7A1 mediated overexpression of *amb07* was attempted to investigate the production properties of such a strain. Based on the observation that a deletion of *amb07* results in a decrease in production, in the easiest case, one would expect enhanced ambruticin productivity in clones overexpressing *amb07*. Although numerous efforts were undertaken to generate this strain, no correct clones have been obtained. This may be caused by the fact that the use of P_{T7A1}, one of the strongest known prokaryotic promoters [108], results in too much transcript which might be harmful for the organism and eventually leads to cell death. In order to circumvent this plausible issue, an inducible gene expression system as described for *M. xanthus* [109] would be highly desirable for strain #8405. A modulated gene expression of *amb07* would allow a more delicate investigation of *amb07*'s function than the all or nothing manner of overexpression and deletion.

As a summary, it should be noted that Amb07, which is encoded in every *amb/jer* gene cluster sequenced so far, shows significant homology to group III hybrid sensor kinase response regulators assumed to be the target of several fungicides. In particular, the high identity of certain HAMP domains of Amb07 with the fungal ones indicates binding of ambruticin. The deletion of *amb07* caused a 10 fold decrease in ambruticin production when compared to the wild type strain. This fact proved the protein's participation in the regulation of compound biosynthesis. On the basis of this finding further experiments should be carried out to identity the interaction partner(s) of Amb07. Contrary to a targeted deletion, an *amb07* overexpression strain could not be obtained, which might be due to deleterious effects upon excessive transcription and/or translation of *amb07*. Here, a conditional gene expression system will be established in the future.

Amb04 Deletion, overexpression and effects on ambruticin production of a putative AtoC-like response regulatory protein

As reasoned from a BLAST analysis, *amb04* is encoding a putative AtoC-like response regulatory protein that may be transcribed along with *amb03* (putative sensor kinase), thereby building up an operon. Noteworthy, this set of genes is much conserved in the *jer* gene cluster of strain #1795 but is not found in the *amb* gene cluster of strain #7282. In this context it is important to mention that *amb03* exhibits a 681 bp in frame deletion relative to *jer3*, what should affect the functionality of Amb03 (HAMP and histidine phospho-acceptor domains are missing). However, these genes (*amb03*/*04*) are located approximately 6.0 kb upstream of *ambA* and might be connected to ambruticin production comparable to the sensor kinase AtoS and the transcriptional activator AtoC which are activating the *ato*-operon in *E. coli* upon the presence of acetoacetate [110-113]. In particular, AtoS gets autophosphorylated upon the presence of

acetoacetate. Hereupon, the phosphate group is transferred to a specific aspartate residue of AtoC, which in turn binds to the promotor region of the *ato*-operon to stimulate transcription. A similar scenario might be conceivable for Amb03/04 and transcription of the *amb*-genes. However, since Amb03 is expected to be non-functional, Amb04 may get phosphorylated alternatively, or bind to DNA in its unphosphorylated state. To examine this hypothesis, we aimed to delete *amb04*. Hence, we constructed plasmid pSM_lacZ_amb04_KO_up+dn1 harboring a 1630 bp fragment upstream of *amb04* (ceasing with *amb04*'s ATG) and a 1634 bp fragment downstream of *amb04* (carrying the terminal 74 bp of *amb04* at its 5'-end). The 1312 bp deletion of the respective part of *amb04* was achieved by gene replacement (Figure **23**).

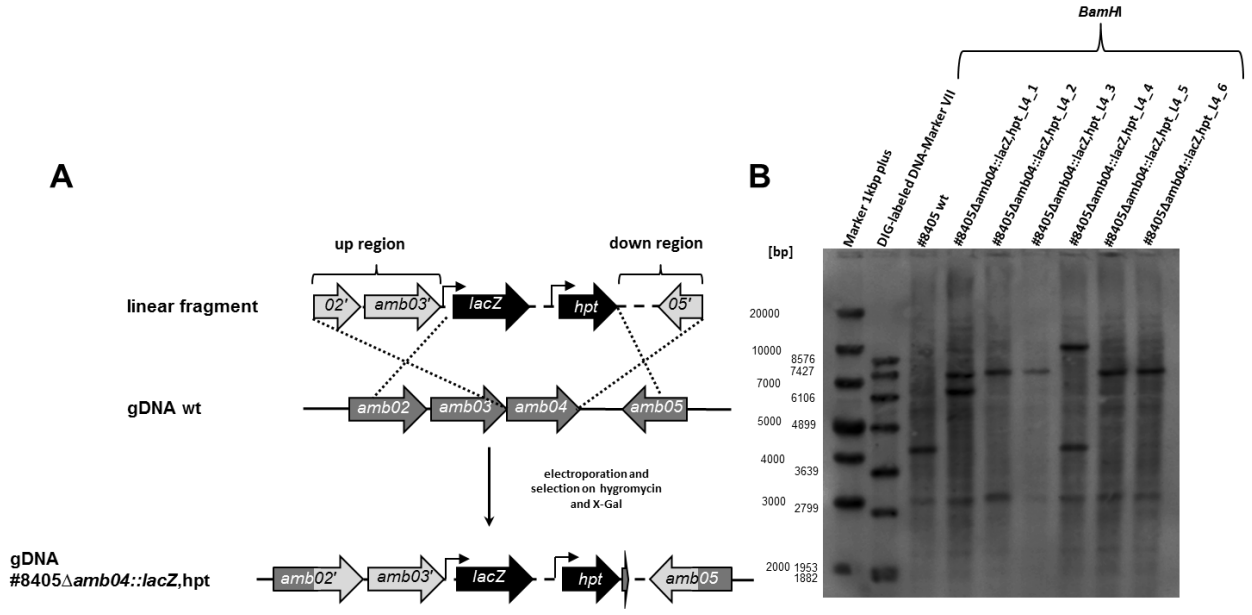


Figure 23: Construction and verification of Δ*amb04* deletion strains. A: Schematic depiction describing gene replacement strategy for deletion of *amb04*; *hpt:* hygromycin phosphotransferase, *lacZ:* beta-galactosidase, ori: origin of replication. **B**: Genotypic verification of Δamb04::lacZ,hpt mutants by Southern Blot analysis using a probe against the up-region. Expected sizes are as follows: wt: 4096 bp; Δ*amb04*: 7351 bp.

As it can be seem from Figure 23 B all samples show a signal at the size of approximately 3.0 kb, regardless if the DNA was prepared from wild type or mutant cultures. This can be attributed to an unspecific binding of the probe. Besides that, the digested DNA from clones L4_2, L4_3, L4_5, L4_6 show only one additional signal at the expected size for successful deletion of *amb04*. The remaining clones L4_1 and L4_4 show a signal pattern which corresponds to the integration of the circular plasmid via the up or dn fragment which results in a duplication of the homology regions. Again, this is due to an incomplete digestion of the KO plasmid. For phenotypic characterization the correct clones and the wild type were cultivated in 300 ml flasks containing 50 ml A-medium for 7 days. Subsequently, methanolic extracts of whole

cells and the supernatant were prepared and analyzed by HPLC-MS (Figure 24). The data showed a significant drop in ambruticin production for the *amb04* deletion mutants when compared to the wild type (two-three orders of magnitude). This suggests that Amb04 is exerting a positive influence on ambruticin biosynthesis; however if this effect is direct (e.g. stimulating transcription of structural genes through binding at promoter region(s) within the *amb* cluster) or indirect (positively affecting other mechanisms associated with ambruticin biosynthesis) remained elusive at this point.

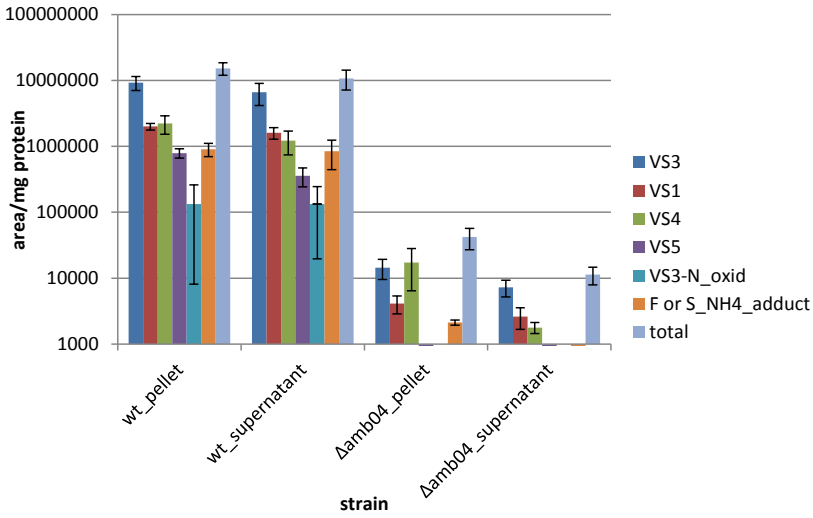


Figure 24: Phenotypic analysis of the Δ*amb04* mutants and the wild type with respect to their capability of ambruticin production. In case of the wt values are means of three independent cultivations while the values for the mutant are means of four independent cultivations. Ambruticin titers are presented relative to the protein content of the samples.

To finally proof the participation of Amb04 in the regulation of ambruticin biosynthesis we attempted to genetically complement the Δ*amb04* mutant by the introduction of a new copy of this gene under the control of the constitutive T7A1 promoter. The possibility of genetic complementation opened up since we found *tetA* to be a second functional resistance gene in strain #8405. Since there are no functional plasmids for myxobacteria available, we integrated the respective complementation plasmid into the genome of strain #8405Δamb04::lacZ,hpt, more precisely L4_3. To alter the genetic background of this strain as little as possible, we decided to integrate the plasmid for complementation into the *lacZ* gene which has been introduced while creating the *amb04* deletion at this locus. This strategy has three advantages in comparison to an integration somewhere else in the genome e.g. by transposition: first, complementation would rather occur in *cis*, than in *trans* (the gene is put back at the location where it has been encoded before); second, no additional alteration within the genome are needed and third, screening for mutants which have integrated the plasmid at the correct

position is easy (screening for orange colonies on X-Gal media, in comparison to screening for blue ones in the deletion procedure carried out before). For construction of the complementation plasmid, an 1889 bp part of *lacZ* was introduced into plasmid pSM_gfp_tet using *Xba*I/*Not*I restriction sites, creating plasmid pSM_gfp_tet_partlylacZ. Into this plasmid, a new T7A1 driven copy of *amb04* was introduced thereby yielding plasmid pSM_gfp_tet_partlylacZ_T7A1_amb04. Subsequently this plasmid was introduced into strain #8405Δamb04::lacZ,hpt thereby generating strain #8405Δamb04::lacZ,hpt::pSM_gfp_tet_T7A1_amb04 (Figure 25 A).

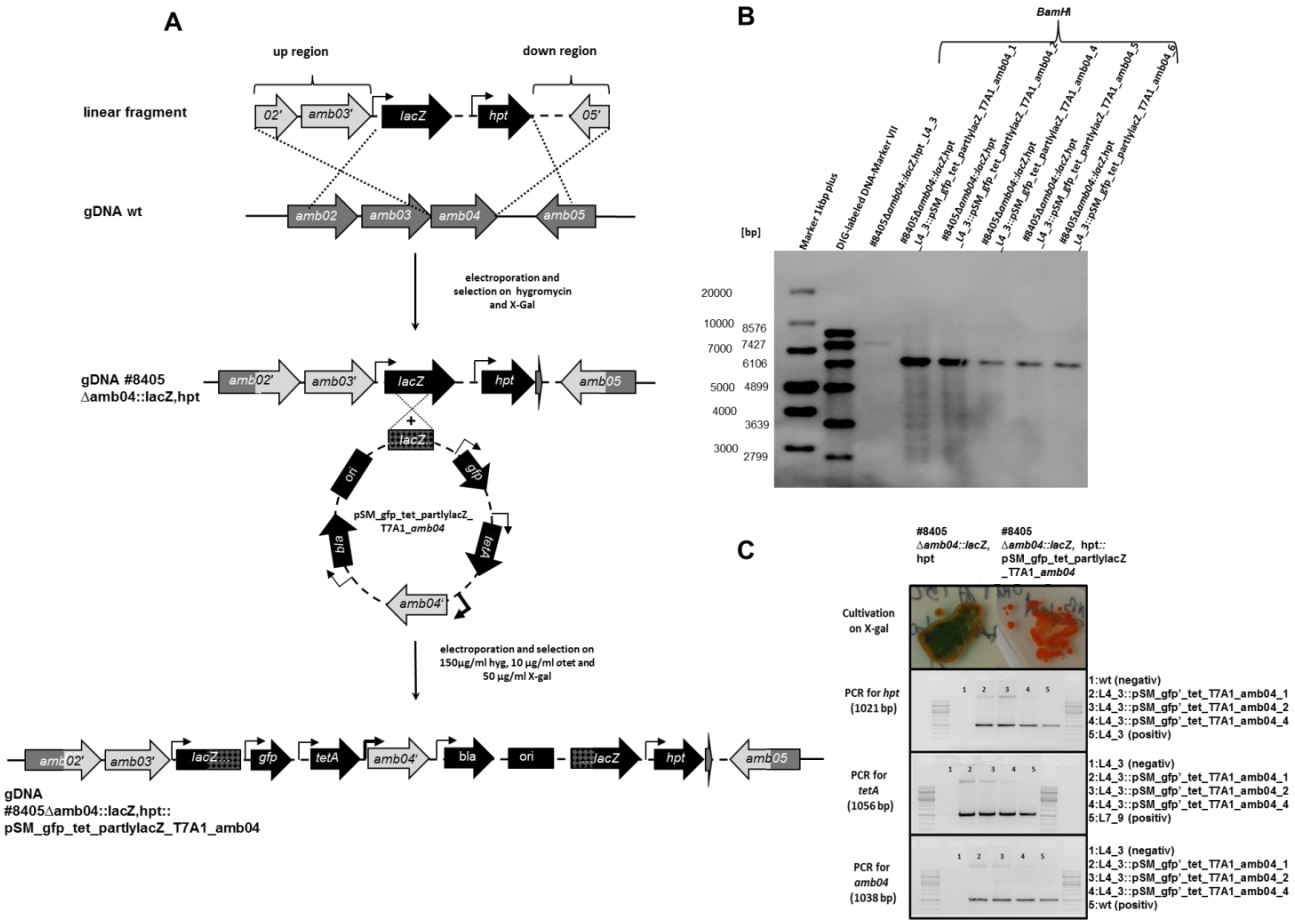


Figure 25: Construction and verification of Δamb04 complementation strains. A: Scheme describing the integration of pSM_gfp_tet_partlylacZ_T7A1_amb04 into *lacZ* of strain #8405Δamb04::lacZ,hpt thereby yielding strain #8405Δamb04::lacZ,hpt::pSM_gfp_tet_partlylacZ_T7A1_amb04; *bla:* beta-lactamase, *gfp*: green fluorescent protein, *hpt*: hygromycin phosphotransferase, *lacZ*: beta-galactosidase, ori: origin of replication, *tetA*: tetracycline efflux protein. **B:** Verification of strain #8405Δamb04::lacZ,hpt::pSM_gfp_tet_partlylacZ_T7A1_amb04 mutants by Southern Blot analysis using a probe against the up-region. Expected sizes are as follows: Δamb04: 7351 bp; Δamb04::pSM_gfp_tet_partlylacZ_T7A1_amb04: 5935 bp. **C:** Cultivation of strain #8405Δamb04::lacZ,hpt (parent strain; *lacZ*$^{+}$) and strain #8405Δamb04::lacZ,hpt::pSM_gfp_tet_partlylacZ_T7A1_amb04 (complementation strain *lacZ*$^{-}$) on HS-media containing X-Gal, and verification PCR's targeting *tetA, hpt* and *amb04*.

Putative mutant strains #8405Δamb04::lacZ,hpt::pSM_gfp_tet_partlylacZ_T7A1_amb04 obtained in this attempt were not able to cleave X-Gal anymore, therefore it was likely that the plasmid integrated at the desired position in the genome (Figure 25 C). To assure this assumption, an additional Southern Blot analysis was performed which proved correct plasmid integration (Figure 25 B). Besides that, PCRs targeting *amb04*, *tetA* and *hpt* were performed to ensure that this strain still carries the initial deletion of *amb04* as well as the newly integrated complementation plasmid (Figure 25 C). Again, the results were fitting the expectations. After genotypic verification, all 5 strains were subjected to phenotypic analysis and compared to the Δ*amb04* deletion strain (Figure 26).

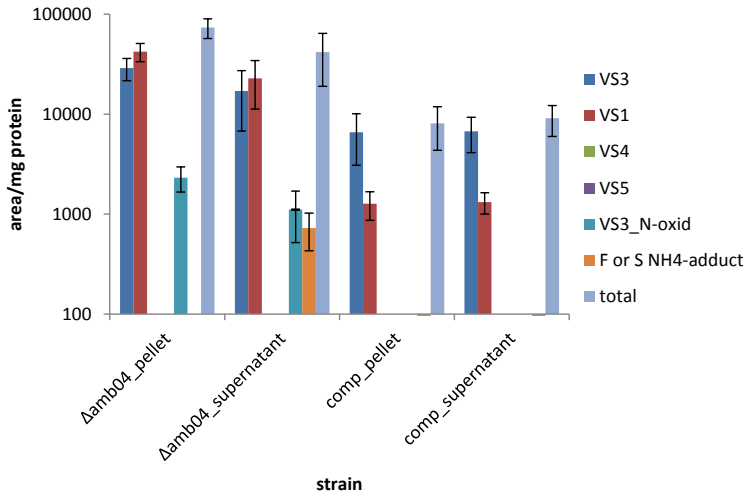


Figure 26: Phenotypic analysis of the Δ*amb04* deletion strain and the respective complementation strains with respect to their capability of ambruticin production. Given values are in case of strain #8405Δamb04::lacZ,hpt_L4_3 (Δ*amb04)* means of three independent cultivations and in case of strain #8405Δamb04::lacZ,hpt::pSM_gfp_tet_partlylacZ_T7A1_amb04 (comp) means of cultivations of five different strains. Ambruticin titers are presented relative to the protein content of the samples.

As it can be seen from this analysis, ambruticin production could not be restored in these strains. This was surprising, because from the knock out experiments of *amb04* it was assumed that Amb04 would be a positive regulator of ambruticin biosynthesis, which is why complementation with T7A1_amb04 should be sufficient. Various explanations could describe why the compound production could not be restored to wild type levels; first, the T7A1_amb04 arrangement does not work and the gene is not transcribed (very unlikely, because this arrangement is used as well for expressing the *lacZ* gene in pSM_lacZ_hyg), second the gene gets transcribed, but the mRNA forms an adverse secondary structure so that the mRNA is not translated, third, Amb04 or its phosphorylated version Amb04-P is present in unfavorable amounts in the cell what eventually might hinder compound production.

To test the first hypothesis, we performed a quantitative real time PCR analysis (q-PCR) to clarify if the gene gets efficiently transcribed. Therefore two independent wild type strains and strain #8405∆amb04::lacZ,hpt::pSMgfp_tet_partlylacZ_T7A1_amb04 1 and 2 (comp 1 and 2) were cultivated in 100 ml A-medium and after 24 h, samples were harvested for RNA isolation. According to this analysis it is clearly evident, that the T7A1 promoter efficiently drives transcription of *amb04* (Figure 27). Conclusively, failure of complementation is not due to the absence of *amb04* transcript in the complementation strains.

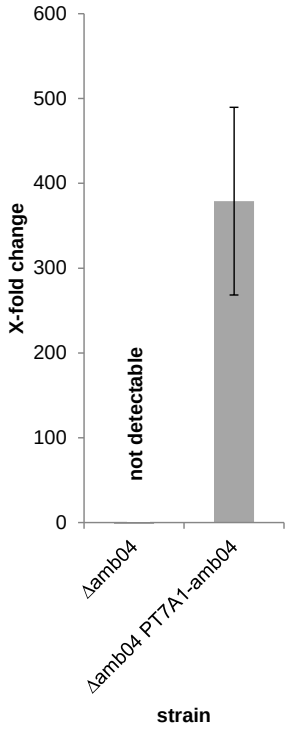


Figure 27: q-PCR analysis to verify transcription of amb04 in the complementation strains. Result of the q-PCR experiment analyzing the amount of amb04 mRNA in strain #8405∆amb04::lacZ,hpt (∆amb04) and strain #8405∆amb04::lacZ,hpt::pSM_gfp_tet_partlylacZ_T7A1_amb04 (∆amb04 PT7A1-amb04) *relative to the wild type is shown.*

Despite an inadequate transcription of *amb04*, it is possible, that its mRNA is forming an unfavorable secondary structure in the 5' region, which hinders translation. Consequently, a second complementation plasmid (pSM_gfp_tet_partlylacZ_aphII_amb04) was generated exhibiting a P_{aphII} driven copy of *amb04* with a promoter arrangement similar to the hygromycin resistance gene from pSM_lacZ_hyg. This plasmid was transferred into strain #8405∆amb04*::*lacZ,hpt L4_3. Three of the resulting clones were verified by PCR and the subsequent phenotypic analysis again showed no differences in ambruticin production when compared to the strain #8405∆amb04*::*lacZ,hpt strain (data not shown). Considering all results

from the complementation experiments, it can be stated, that the reintegration of a functional copy of *amb04* into the genome of strain #8405Δamb04*::*lacZ,hpt was successful, nevertheless ambruticin production could not be restored. Assuming that the vast amount of *amb04* mRNA in the complementation strain results in an excess of Amb04, it is possible that this excess itself is the cause for the lack of production. To test this, again an inducible system would be needed to allow modulated gene expression to encompass wild type levels of Amb04 in a Δ*amb04* background.

In addition to the experiments mentioned above, we also intended to overexpress *amb04* in wild type background, to see if such a strain is able to synthesize elevated levels of ambruticin. Since this gene is part of the putative *amb03-04* operon and these genes are overlapping, we exchanged the native promoter in front of *amb03* against P_{T7A1}, thereby overexpressing both genes. Thus plasmid pSM_lacZ_hyg_T7A1_amb03-04 was constructed and integrated into the genome of the wild type (Figure 28 A). After genotypic verification by PCR (Figure 28 B) six different clones (1, 2, 3, 4, 12 and 16) were cultured and their metabolic profiles were examined by LC-MS (Figure 29). Interestingly, two different phenotypes were observed; the clones 1, 2, 3 and 4 which are combined under the acronym T7A1_amb03-04_A show a significant drop (about 12-14 fold) in total ambruticin production when compared to the wild type. Contrary to this, clones 2 and 4 (T7A1_amb03-04_B) produce ambruticin in an amount similar to the wild type. However, the majority (two-thirds) of the clones analyzed show a decreased production of ambruticins upon the overexpression of *amb03-04.* This indicates once more, that the amount of the putative activator of ambruticin biosynthesis Amb04 does not correlate to the amount of compound built up by the cell. The discrepancy between the results obtained for T7A1_amb03-4_A and B is conspicuous and might be explained by the occurrence of a compensatory mutation(s) allowing the latter clones to produce ambruticin at wild type levels (e.g. by downregulating Amb04 amounts to a “normal” level).

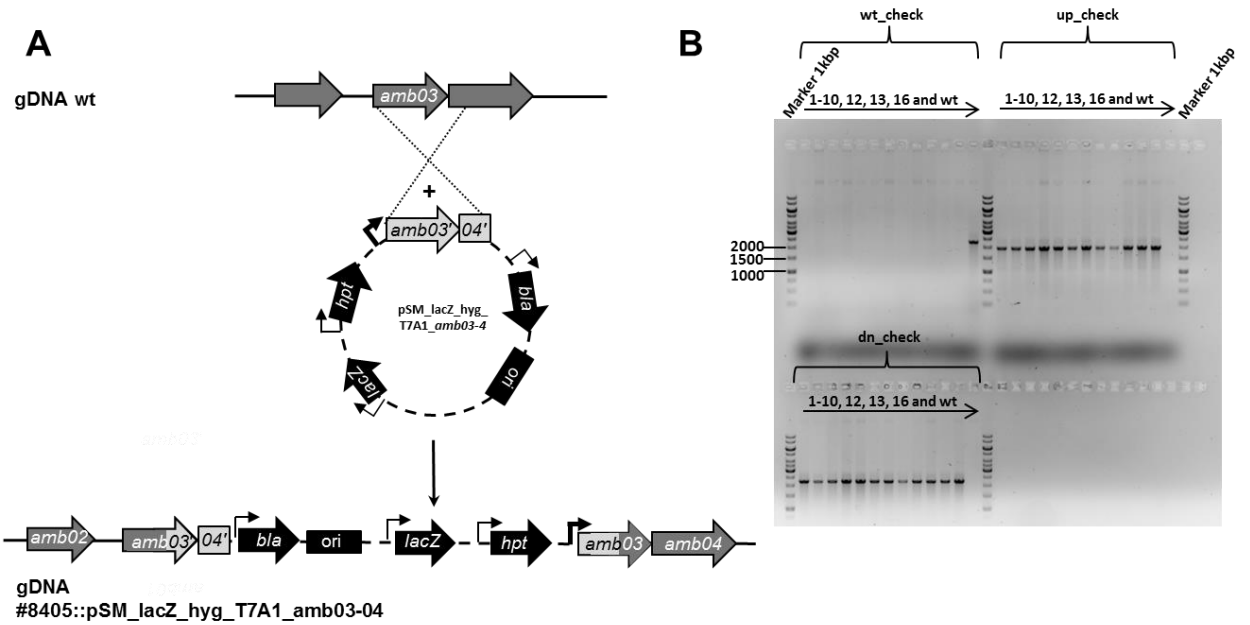


Figure 28: Construction and verification of T7A1_amb03-04 strains. ***A:*** Scheme describing the promoter exchange strategy for overexpression of *amb03-04*; *bla:* beta-lactamase*, hpt*: hygromycin phosphotransferase, *lacZ*: beta-galactosidase, ori: origin of replication. **B:** Verification of resulting mutants of strain #8405::pSM_lacZ_hyg_T7A1_amb03-04 1,2,3,4,5,6,7,8,10,12,13,16 and wt control (from left to right)*;* Expected sizes are as follows: wt_check: 2255 bp; up_check: 1939 bp; dn_check: 2219 bp).

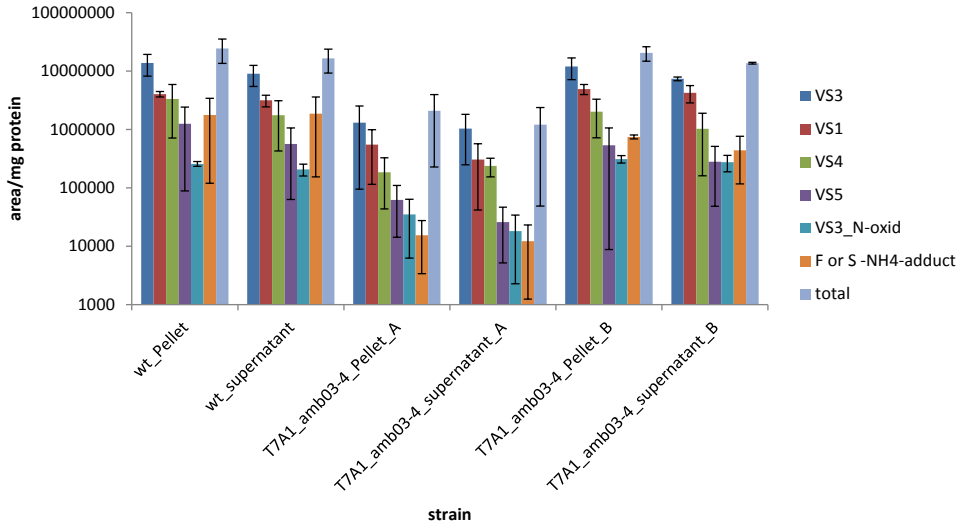


Figure 29: Phenotypic analysis of the T7A1_amb04 overexpressing strains and the wild type with respect to their capability of ambruticin production. Given values are in case of the wild type means of three independent cultivations, in case of strain #8405::pSM_lacZ_hyg_T7A1_amb03-04_A means of four different clones (1,2,3 and 4) and in case of strain #8405::pSM_lacZ_hyg_T7A1_amb03-04_B means of two different clones (12 and 16). Ambruticin titers are presented relative to the protein content of the samples.

Another idea for an explanation could be that the overexpression of *amb03-04* would basically boost ambruticin formation but the cell then downregulates or stops the biosynthesis in order to protect itself. In any case, this might not be brought in line with the fact that the complementation strains show no differences at all in ambruticin production in comparison to the knock out.

As a summary, the deletion of *amb04* results in almost complete abolishment of ambruticin production in the corresponding strains, strongly suggesting that this protein is an activator of ambruticin biosynthesis, although at this point it is not clear if this is a direct (by binding to one or more promoter regions of the *amb* cluster) or indirect regulation (by influencing elements regulating ambruticin formation). The fact that the functional reintegration of *amb04* in the genome of strain #8405Δamb04*::*lacZ,hpt did not restore production is puzzling. In this context it was shown that the mRNA of *amb04* is present in the complementation strains (up to 300 fold more when compared with the wild type). Additionally, the respective part was amplified from the genome of two clones (1, 2) and eventually sequenced, thereby proving its integrity. Together with the fact that complementation of Δ*amb04* with P_{aphII}_amb04 was not possible either, it can be assumed that lack of production is not caused by non-functional Amb04 but rather the presence of too much of it or by counteraction of the cell upon interfering with its regulatory network. In the case of *amb03-04* overexpression, the majority of the examined clones show a 12-14 fold reduction of the intra- and extracellular ambruticin content. This is not as low as for the deletion of *amb04* but significantly lower than wild type production. The reason for that might be yet again counteraction of the cell. Assuming, that Amb04 in its "active" state (e.g. with or without phosphorylation) is not only controlling ambruticin biosynthesis, but also other processes, it is conceivable that the cell reduces this form to a minimum which results in reduced compound levels. However, this might not explain why not a least a "slight" complementation of the Δ*amb04* phenotype could be observed in clones overexpressing this gene. At this point further experiments are needed to shed light onto the actual role of Amb04. Therefore this protein was heterologously produced and subjected to further analysis (see below).

Genes of the ambruticin gene cluster regulated by Amb04

In order to get insights into the fashion Amb04 is regulating ambruticin biosynthesis we aimed to heterologously produce this protein for the purpose of electrophoretic mobility shift assay (EMSA) analysis. Thereby we wanted to assess if Amb04 is directly binding to presumed promoter regions of the *amb* gene cluster or if it indirectly affects compound formation. Hence, Amb04 was heterologously produced as His6-SUMO-tagged protein in *E. coli* strain Rosetta and

enriched using immobilized metal affinity chromatography (IMAC) with Ni^{2+}-NTA beads. Next, the likely His_6-SUMO-Amb04 protein was subjected to LC-MS analysis in order to prove its identity.

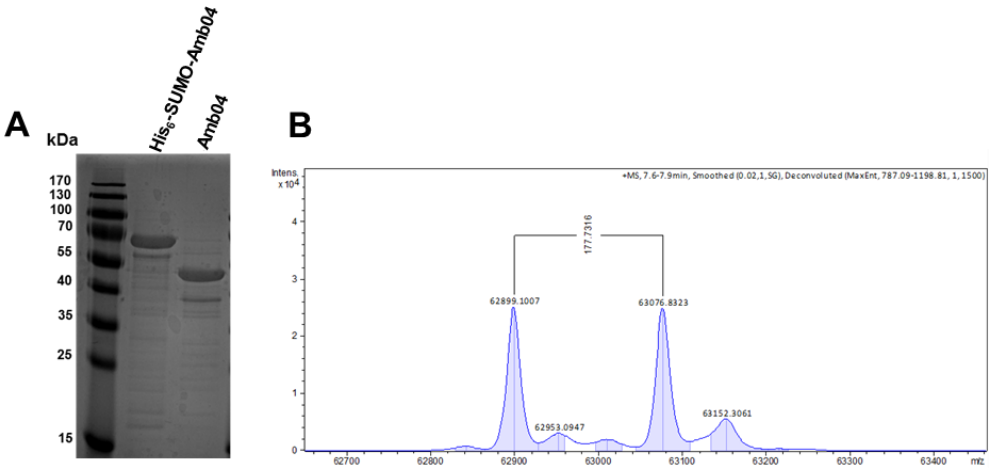


Figure 30: Analysis of heterologously produced Amb04. A: SDS-PAGE of enriched His_6-SUMO-Amb04 (≈63 kDa) and Amb04 (≈51 kDa). **B:** Detail of the deconvoluted LC-MS spectrum of His_6-SUMO-Amb04 analysis. The $\Delta m/z$ of 178 Da results from phosphogluconoylation of the His_6-tag which is common for proteins produced in *E. coli*.

The measured mass of 62899.10 Da fits to the expected one of 62897.51 Da (missing the first methionine) with an error of 2.41 Da, thereby confirming the identity of the protein (Figure 30 B). After incubation with SUMO2 protease and subsequent IMAC the protein was analyzed again by SDS-PAGE (Figure 30 A). After having successfully produced Amb04 it was used in an EMSA analysis with the indicated 5′-hexachlorofluorescein phosphoramidite (HEX)-labeled DNA fragments of presumed promoter regions of the *amb* cluster (Figure 31). Since the binding properties of Amb04 might vary upon phosphorylation, indicated preparations (+P) contained protein which was incubated with 30 mM acetyl phosphate for 30 min at 37°C prior to the experiment.

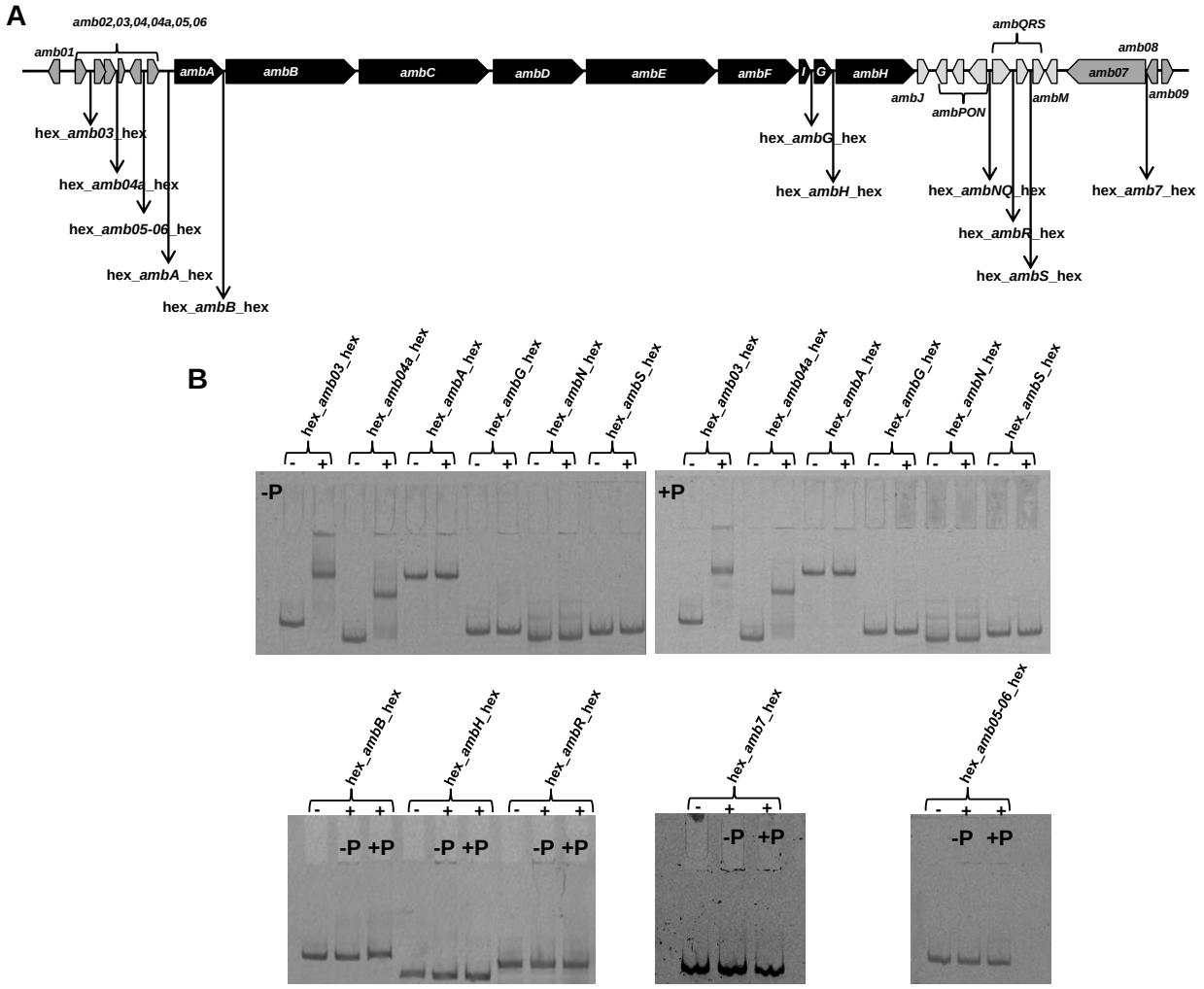


Figure 31: Schematic description and results of EMSA studies with Amb04. A: Illustration of the *amb* gene cluster indicating the fragments used in this analysis. Genes encoding the core PKS machinery are shown in black. Post-PKS tailoring genes have light grey color and genes encoding proteins not directly involved in ambruticin biosynthesis are depicted in dark grey. **B:** Results of EMSA studies with Amb04; **+/**: with or without protein, **+/-P**: with or without incubation of Amb04 with acetyl phosphate prior to EMSA analysis.

From Figure 31 B it can be concluded, that the incubation of Amb04 with acetyl phosphate did not affect the binding to the respective DNA fragments. Consequently, binding of Amb04 is either independent of its phosphorylation status, or *in vitro* phosphorylation was insufficient. Nevertheless, these studies show that unphosphorylated Amb04 is binding to two of the eleven DNA fragments bearing putative promoter regions employed; the presumed promoter of the putative *amb03-04* operon (autoregulation is a common phenomenon among bacterial two component systems [114]) and approximately 2.3 kb downstream to the promoter of *amb04*a. The same pattern was also observed in case for Amb04 preincubated with acetyl phosphate. Based on these results, Amb04 presumably exerts its effect on ambruticin biosynthesis in an indirect manner, either via Amb04a or another member of its regulon.

After binding of Amb04 to the DNA-fragments hex_*amb03*_hex and hex_*amb04a*_hex was observed, we wanted to define the DNA binding motif of Amb04. Thus, each of the respective fragments was divided into smaller pieces of 50 nucleotides with a 15 bp overlap to the adjacent ones (Figure 32 A) and another EMSA analysis was performed (Figure 32 B).

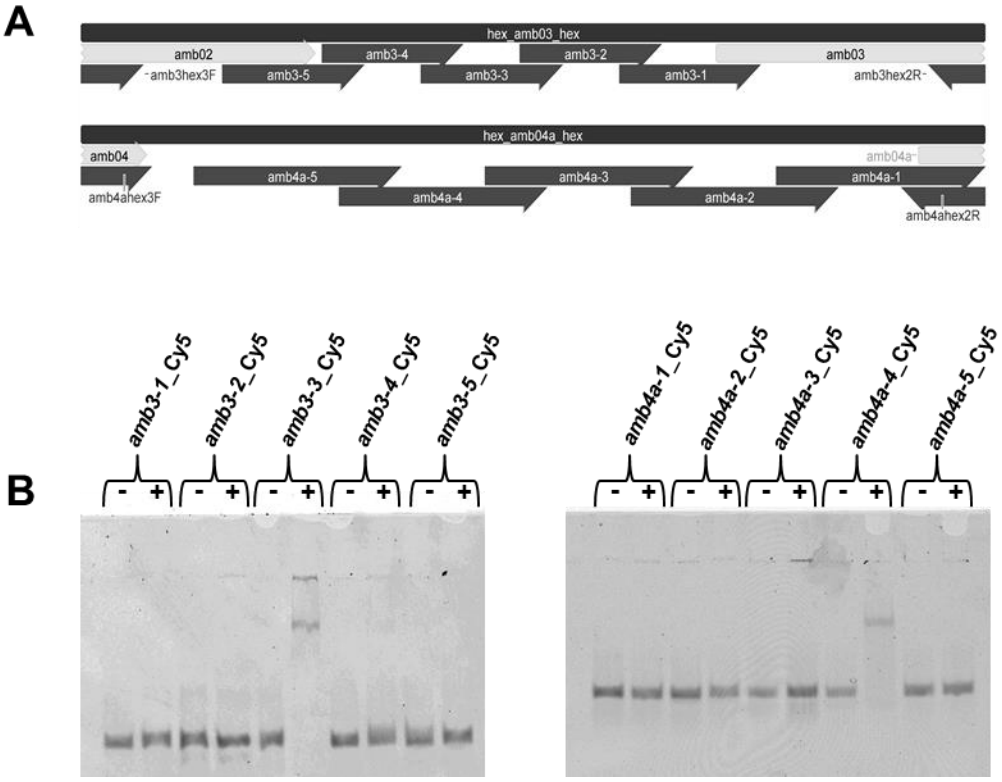


Figure 32: EMSA analysis to narrow down the DNA binding site of Amb04. A: Illustration showing the dissection of hex_amb03_hex and hex_amb04a_hex into smaller fragments which were then used in EMSAs with Amb04. **B:** Results of EMSA studies with Amb04; **+/-**: with or without protein. In this analysis only non-phosphorylated protein was used.

As a result, only the fragments *amb3-3*_Cy5 and *amb4a-4*_Cy5 showed retardation when incubated with Amb04 prior to the EMSA (Figure 32 B). A closer investigation of these segments revealed the presence of three similar motifs, imperfect inverted repeats with a 5 bp spacer sequence, (two on *amb3-3*_Cy5 and one on *amb4a-4*_Cy5) featuring the consensus sequence: TGYYGCABWWYGCGRCR (Figure 33).

Figure 33: Determination of putative consensus sequence for Amb04 binding. MAFFT alignment of different Amb04 binding sites present on *amb3-3_Cy5* and *amb4a-4_Cy5*. Y: pyrimidine (C or T), B: not A (G or C or T), W: weak (A or T), R: purine (A or G), V: not T (A or G or C).

Next, it was attempted to clarify which part(s) of the respective sequences are crucial for binding of Amb04 to the DNA. Therefore several mutated versions of the original fragments *amb3-3*_Cy5 and *amb4a-4*_Cy5 were prepared and used in further EMSAs (Figure 34). As a result, we observed that alterations in repeat 1, as found on fragments *amb33-1*_Cy5- *amb33-3*_Cy5, are causing a reduction of the shift's intensity when compared to the unmodified version (*amb3-3*_Cy5). That clearly proves the assumed interaction of Amb04 with this particular sequence. In a likewise manner repeat 2 was changed and resulted in the effect, that modification of only one portion of the repeat (*amb33-4*_Cy5 and *amb33-5*_Cy5) did not significantly affect binding of Amb04, whereas mutation of both sites did (*amb33-6*_Cy5). Moreover, it was tested if modifications in the spacer region on *amb3-3*_Cy5 would have an impact on binding properties of Amb04. Therefore we used fragment *amb33-7*_Cy5 which harbored a mutated spacer region in addition to a completely altered repeat 2 as found on *amb33-7*_Cy5. In this case, retardation of the respective DNA was not observed anymore. All together these experiments identified repeat 1 and 2 on *amb3-3*_Cy5 as binding sites of Amb04. Additionally, we analyzed the repeat on *amb4a-4*_Cy5 by using fragments altered in both, the repeats as well as the spacer region (*amb4a4-1*_Cy5 and *amb4a4-2*_Cy5) whereupon shifting was completely abolished. Again, these tests confirmed the assumed binding site of Amb04 on *amb4a-4*_Cy5.

A

Fragment	Sequence 5'-3'
*amb3-3*_Cy5	GCGGG**TGCCGC*ACAAC*GCGGCA*TGTCGC**ATTTT*GCGACG**GCGTCGGGCGG
*amb33-1*_Cy5	GCGGGAATTAA***ACAAC***GCGGCA*TGTCGC**ATTTT*GCGACG**GCGTCGGGCGG
*amb33-2*_Cy5	GCGGG**TGCCGC*ACAAC*AATTAA*TGTCGC**ATTTT*GCGACG**GCGTCGGGCGG
*amb33-3*_Cy5	GCGGGAATTAA***ACAAC***AATTAA*TGTCGC**ATTTT*GCGACG**GCGTCGGGCGG
*amb33-4*_Cy5	GCGGG**TGCCGC*ACAAC*GCGGCA**AATTAA*ATTTT*GCGACG**GCGTCGGGCGG
*amb33-5*_Cy5	GCGGG**TGCCGC*ACAAC*GCGGCA*TGTCGC**ATTTT*AATTAAGCGTCGGGCGG
*amb33-6*_Cy5	GCGGG**TGCCGC*ACAAC*GCGGCA**AATTAA*ATTTT*AATTAAGCGTCGGGCGG
*amb33-7*_Cy5	GCGGGAATTAA***ACAAC***AATTAA*TGTCGC**CAAGC**GCGACG**GCGTCGGGCGG
*amb4a-4*_Cy5	GGCGACTCCGCTCGTCCCTCGC**TGTTGC*AGAAC*GCGACA**CCCGCGCCGCC
*amb4a4-1*_Cy5	GGCGACTCCGCTCGTCCCTCGCAATTAA***AGAAC***AATTAACCCGCGCCGCC
*amb4a4-2*_Cy5	GGCGACTCCGCTCGTCCCTCGC**TGTTGC**GACCA**GCGACA**CCCGCGCCGCC

B

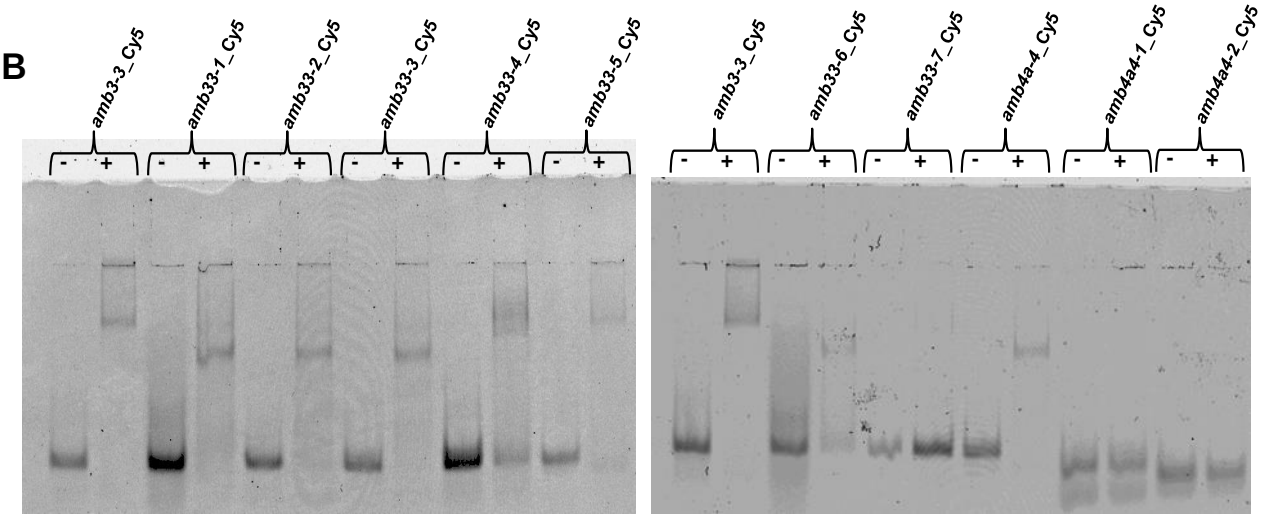


Figure 34: EMSA analysis to determine the DNA binding site of Amb04. A: Fragments used in EMSAs for determination of residues which are crucial for binding of Amb04. Nucleotides in bold text highlight the consensus sequence (imperfect inverted repeats just bold, spacer sequence additionally italic); nucleotides underlined are changed with respect to the native sequence. **B:** Results of EMSA studies with Amb04; **+/-**: with or without protein. In this analysis only non-phosphorylated protein was used.

As a summary it can be stated, that Amb04 was successfully produced as a His_6-SUMO-tagged form, what permitted extensive DNA binding studies. This facilitated detection of genes around the *amb* cluster whose expression is controlled by Amb04. Interestingly this investigation revealed that none of the structural genes required for ambruticin biosynthesis might be directly regulated by Amb04. This was an unexpected finding, since a deletion of this gene resulted in almost complete loss of ambruticin formation. Nevertheless, we found that Amb04 seems to regulate its own expression by binding to two defined sections of the presumed promoter region of the putative *amb03-04* operon. If this results in multiplication of the stimulus or simply in downregulating its own expression upon adequate availability remains elusive. Furthermore, it is unclear if the phosphorylation status of the protein has an impact on its binding properties. Our

in vitro experiments suggest that phosphorylation is of no importance, but the actual scenario *in vivo* is still unknown. Since *amb03*, encoding the sensor kinase, exhibits a huge in frame deletion in the N-terminal part it is most likely that perception of a suitable stimulus and signal transmission to Amb04 might be performed by a different protein, maybe Amb05 (putative sensor His-kinase). Besides that, we could detect an additional binding site in front of *amb04a* which is located downstream of *amb04* and encodes a conserved uncharacterized protein comprising an UPF0158 domain. According to these results, it must be expected that Amb04 is indirectly affecting ambruticin biosynthesis either via Amb04a or another member of its unknown regulon. Fortunately, our EMSA analysis identified three binding sites that allowed determination of a first consensus motif (TGYYGCABWWYGCGRCR) for binding of Amb04. Subsequently, this knowledge was used in a genome wide analysis to detect other genes potentially regulated by Amb04 (see below).

The putative Amb04 regulon: Possible effects on ambruticin biosynthesis

As pointed out above, comprehensive EMSA analysis using Amb04 and several fragments of presumed promoter regions from the *amb* cluster led to the identification of a preliminary consensus motif (TGYYGCABWWYGCGRCR) to which Amb04 is binding *in vitro*. Since it was likely that the first adenine in the spacer part is of great importance, whereas the remaining four nucleotides might be exchangeable, the consensus sequence was adjusted accordingly. Based on this deliberation we employed the sequence TGYYGCA4NGCGRCR in a genome wide analysis in order to search for genes that are preceded by a similar segment (max. distance to start codon 400 bp), so their expression may also be controlled by Amb04. Initially, this search produced 38 hits and after reasonable considerations (intergenic location and correct orientation of the motif) this number was reduced to a total of 10, including *amb04a* and the putative *amb03-04* operon (Table 5). Unfortunately, no valid statement concerning the concrete chromosomal positions of genes other than *amb03-04* and *amb04a* can be made, because the draft genome of strain #8405 has not been fully assembled to date. From this diverse set of genes not all might play a role in regulation of ambruticin production or the biosynthesis itself. Consecutively, each gene and its putative function will be discussed.

Table 5: Genes potentially regulated by the transcriptional regulator Amb04

gene or #	Length (bp/aa)	Accession number of closest and annotated homologue	Similarity/identity [%]	Putative function/homologue
amb04a	669/222	WP_016253989	26/46	uncharacterized protein family 0158
amb03-04	780/259 1386481	ABK32251 ABK32252	100/100 100/100	damaged sensor His-kinase, AtoC response regulator homolog
1	693/230	WP_020740624	96/99	cytochrome c family protein
2	534/178 (partial sequence)	WP_012235106	69/75	IMP dehydrogenase/GMP reductase
3	1398/465	WP_020740117	94/96	major facilitator transporter
4	660/219	WP_028252326	82/88	glutathione S-transferase
5	558/185	EYF07579	78/87	regulatory protein, ArsR-like
6	498/165	WP_012240457	76/85	major facilitator superfamily permease
7	1899/632	EYF03936	59/69	serine/threonine protein kinase
8	783/260	WP_012236875	98/99	thiazole synthase

The gene *amb04a* encodes a small protein with a molecular mass of approx. 25 kDa containing a conserved but uncharacterized domain (UPF0158). According to the Conserved Domain Architecture Retrieval Tool (CDART) [115] accessible via NCBI this domain is almost exclusively found in bacterial proteins from organisms belonging to the phylum of proteobacteria, actinobacteria, firmicutes and members of the genus of *Chlamydia*. The vast majority of these proteins is rather small (up to 300 aa) and contains no further domains, similar to Amb04a. Moreover, UPF0158 domains are rarely found around the N-terminus of proteins that additionally harbor a GCN5-related N-acetyltransferase domain at their C-terminus. Finally, such domains are also present in putative adenylate cyclases from *Frankia* sp. and in some hypothetical proteins of actinobacteria that accessorily feature a pRiA4b ORF-3-like domain, which has not been characterized either. Consequently, any speculation about the possible role of Amb04a is untenable at the moment. In order to get an idea about the function of Amb04a we wanted to assess in which way a deletion or overexpression of *amb04* influences the expression of *amb04a*. Therefore q-PCR experiments were performed using cDNA derived from RNA that was isolated from the wild type, the Δ*amb04* deletion mutant (strain #8405Δamb04*::*lacZ,hpt), and the Δ*amb04* complementation strain (strain #8405Δamb04*::*lacZ,hpt*::*pSMgfp_tet_partlylacZ_T7A1_amb04) (Figure 35).

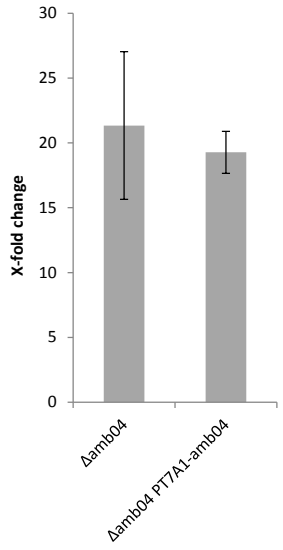


Figure 35: q-PCR analysis determining amounts of *amb04a* mRNA. Results of quantitative real time PCR experiments analyzing amounts of amb04a mRNA in strain #8405Δamb04::lacZ,hpt (Δamb04) and strain #8405Δamb04::lacZ,hpt::pSM_gfp_tet_partlylacZ_T7A1_ amb04 (Δamb04 PT7A1-amb04*)* relative to the wild type.

This analysis showed elevated levels of *amb04a* mRNA in both mutant strains when compared to the wild type. Since both strains are impaired in ambruticin production we interfered that the reduction could be due to increased amounts of Amb04a. If this hypothesis is true, a targeted deletion of *amb04a* should result in higher product yields. This experiment is planned and currently ongoing.

The EMSAs described above have also shown that Amb04 specifically binds to a region around the presumed promoter of its operon, indicating that this regulator controls its own expression. In the context of these experiments the scenario of its autoregulation has already been discussed and therefore it will be referred to this section of the chapter (Genes of the ambruticin gene cluster regulated by Amb04).

In case of the cytochrome c family protein an involvement in the regulation of ambruticin production might be debarred. It builds up a likely operon, together with a putative molybdopterin oxidoreductase and putative Fe-S cluster containing hydrogenase. Such enzymes are rather involved in primary metabolism of adequate substrates than in regulation of secondary metabolite formation.

The putative IMP dehydrogenase/GMP reductase might be as well regulated by Amb04. These enzymes are usually necessary for *de novo* synthesis of GTP (IMP dehydrogenase) [116] or maintaining the intracellular balance of adenine and guanine nucleotides (GMP reductase) [117]. In compliance with their assigned function, a potential role in ambruticin regulation might be ruled out as well.

Transporters belonging to the major facilitator superfamily (MFS) are ubiquitously found in bacteria, archaea, and eukarya. They employ the chemiosmotic gradient of the cell to im- or export a diverse set of molecules such as sugar monomers, oligosaccharides, inositols, drugs, amino acids, and a large variety of ions [118]. Accordingly, it could be conceivable that the deletion of *amb04* results in loss of transporters that import an effector molecule which initiates compound production via an unknown route. Limitations in precursor supply (acetate, propionate *S*-adenosylmethionine) upon lapse of importers are unlikely, but might also be a reason for switching off ambruticin biosynthesis.

Bacterial glutathione S-transferases (GST) play a key role in cellular detoxification of harmful substances like antimicrobials or xenobiotics and protect the cell against oxidative or chemical stresses [119]. Common to all GSTs is the ability to transfer the tripeptide glutathione to electrophilic groups of dedicated molecules thereby "labelling" them. For *Proteus mirabilis* it has been shown, that exposure to antibiotics led to increased levels of a distinct GST (B1-1) and that a null mutant exhibits a higher sensitivity towards the respective drugs [120]. A possible self-resistance mechanism against ambruticin by this GST may be excluded, since extracts prepared from wild type cells of strain #8405 contain huge amounts of the unmodified compound inside the cell. Furthermore, at the moment it is unknown if this GST is positively or negatively regulated by Amb04, which is why no solid statement about the function of this enzyme in the context of ambruticin formation can be made.

The expression of a gene encoding a putative thiazole synthase might be as well regulated by Amb04. This protein is the key enzyme for generating the thiazole moiety of thiamin pyrophosphate [121]. This co factor is used by several enzymes involved in important metabolic pathways such as pyruvate dehydrogenase, α-ketoglutarate dehydrogenase and transketolase [122]. Since ambruticin does not feature a thiazole moiety it is most unlikely that this gene is directly involved in compound formation. It may be upregulated by Amb04 while extensive secondary metabolite biosynthesis in order to ensure sufficient provision of energy by the respective catabolic pathways.

The consensus sequence for binding of Amb04 is as well found upstream of two separate genes encoding a putative ArsR-like regulatory protein and a serine/threonine protein kinase, respectively. In contrast to most of the genes discussed above, these ones could play a concrete role in regulating ambruticin biosynthesis. The ArsR transcriptional repressor from *E. coli* responds to the presence of arsenic, antimony or bismuth salts. This leads to the dissociation from its binding site what causes expression of genes encoding detoxifying enzymes to maintain cellular homeostasis [123;124]. Assuming that the ArsR-like protein of strain #8405 regulates a process which somehow (directly or indirectly) affects ambruticin

biosynthesis, the deletion of *amb04* could eventually result in loss of compound formation. The same might apply for the putative serine/threonine protein kinase. These types of proteins are well known for their ability to control diverse cellular processes. When sensing adequate external stimuli, those enzymes undergo autophosphorylation and transfer the phosphorous group to a serine or threonine of the target protein, whereupon it gets shifted to an active or inactive state [125]. In case the respective kinase from strain #8405 senses the signal needed to initiate ambruticin biosynthesis and hands it over to the effector which then turns on expression of the *amb* genes, the deletion of *amb04* finally causes the drop in ambruticin production.

As a summary it can be stated, that the expression of at least 10 genes on the chromosome of strain #8405 is potentially controlled by the transcriptional regulator Amb04. Besides its own expression this protein could consequently target genes required for primary metabolism and cellular homeostasis as well as regulatory functions. Besides this broad *in silico* investigation, hard experimental data is needed to corroborate this hypothetical regulon, for example by means of a comparative q-PCR analysis of the wild type and the Δ*amb04* mutant strain. Clear differences in the expression profiles of the mentioned genes should give further insights into the actual regulation cascade behind the Δ*amb04* phenotype. At this point, all the proposed scenarios remain purely speculative.

Attempts to increase ambruticin production by overexpression of the biosynthetic genes

Although all efforts undertaken so far indeed shed light onto interesting regulatory mechanisms underlying ambruticin formation, we could not generate a rationally engineered strain with significantly enhanced production properties. Therefore, we tried to generate an overproducing strain the “easy-way”. Several studies have shown that the exchange of the native promoter in front of the biosynthetic genes against a strong and constitutive one can tremendously increase compound yields [25;26]. This leads on one hand to more transcript of the respective genes and on the other hand predominant regulatory mechanisms are bypassed. Of course this strategy can only be successfully applied if either the desired molecule is not toxic to the cell itself or by simultaneously boosting the resistance determinant. As strain #8405 is already producing high amounts of ambruticin without obvious defects in viability, we had littleconcerns that this approach could not be used accordingly. Therefore, plasmid pSM_lacZ_hyg_T7A1_ambA was constructed and transferred into strain #8405 (Figure 36).

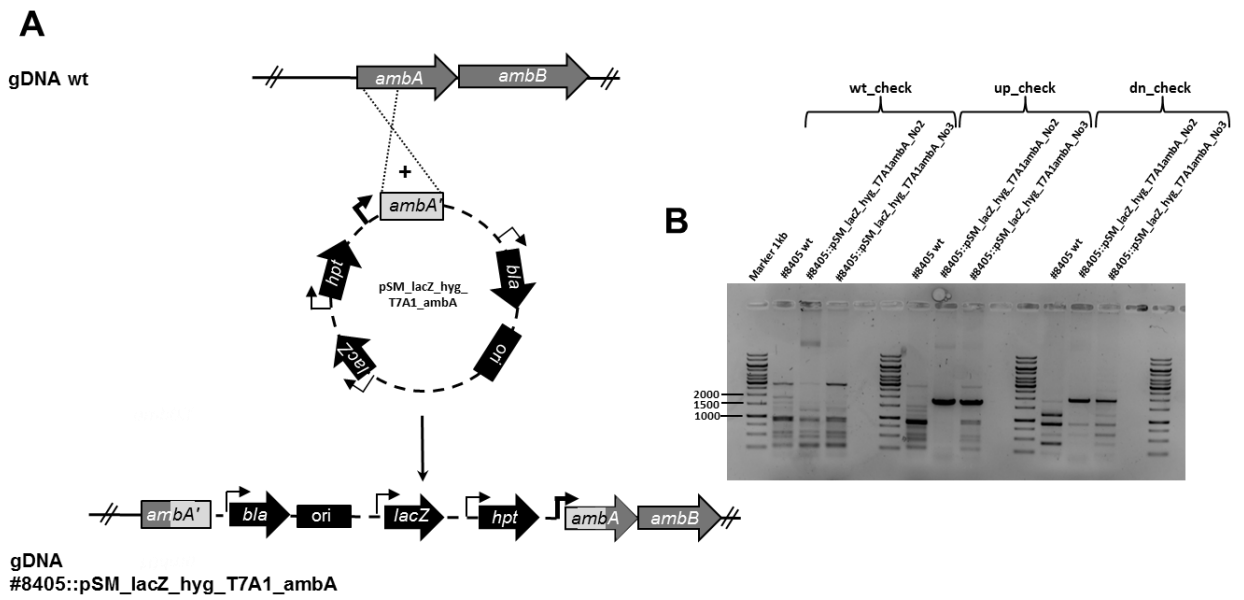


Figure 36: Construction and verification of T7A1_ambA strains. A: Scheme describing the promoter exchange strategy for overexpression of the biosynthetic genes for ambruticin production; *bla*: beta-lactamase, *hpt*: hygromycin phosphotransferase, *lacZ*: beta-galactosidase, ori: origin of replication. **B**: Verification of resulting mutants of strain #8405::pSM_lacZ_hyg_T7A1_ambA; Expected sizes are as follows: wt_check: 1822 bp; up_check: 1816 bp; dn_check: 1875 bp).

Interestingly, only two clones were obtained, although due to the large homology size of the *ambA* fragment (1511 bp) many more were expected. A subsequent phenotypic characterization of those mutants showed impairment of ambruticin production (data not shown).

This result was unexpected and again raises the need for a conditional gene expression system for the genus *Sorangium*. The most obvious explanation for loss of production might be that the P_{T7A1} mediated overexpression boosts the ambruticin biosynthetic machinery far too much whereupon the cell counteracts either by accumulation of compensatory mutations, or by interfering with some regulatory mechanisms. This phenomenon might be often observed but does not find its way into the literature. An inducible system would allow tuning of gene expression to a level which does not cause any antagonizing effects, but eventually leads to enhanced production properties of the strain.

Summary and Conclusion

The long-term aim of this project is the construction of a rationally engineered strain which synthesizes significantly higher amounts of the potent fungicide ambruticin. As starting point the best producer accessible from our strain collection was identified to be *S. cellulosum* strain #8405. Subsequently, an electroporation protocol for the transfer of exogenous DNA into this organism was established, thereby replacing the laborious technique of conjugation described

earlier for this strain. Moreover, this system provided the basis for introducing the procedure of gene replacement to the genus *Sorangium*. This allows the efficient deletion of genes which is superior to the usually applied technique of single cross over gene disruptions. Besides that, the presentation of *tetA* as second selection marker alongside with the *hpt* gene, offered the possibility of genetic complementation studies to be made in this genus. Additionally, it was shown that *lacZ* can reliably be used as reporter gene in *Sorangium* what facilitates fast and easy detection of *lacZ*$^+$ mutants. Furthermore, our efforts to use *lacZ* as counterselection marker for seamless gene deletions showed its functionality, even though this process could be too long-lasting to generate trustable mutants. Attempts to accelerate the detection of cured clones through the application of *sacB* were unsuccessful.

Eventually, the establishment of those tools allowed to study the function of genes potentially involved in the regulation of ambruticin biosynthesis. In particular, four genes of interest were examined: *ntcA* (NtcA-like transcriptional regulator), *amb01* (LysR-family transcriptional regulator), *amb04* (AtoC-like response regulatory protein), and *amb07* (hybrid sensor kinase response regulator). A targeted deletion of *ntcA* had no dramatic effect on ambruticin production in comparison to the wild type. Overexpression of *ntcA* is still pending, which is why no definite statement about the gene's function in ambruticin biosynthesis can be made to date. However, according to significant similarities to NtcA from So ce56, which was shown to be a negative regulator of secondary metabolite pathways (chivosazol and etnangien) it is assumed, that this regulator may also affect ambruticin production in a negative way in strain #8405. In case of *amb01*, only overexpression but no deletion mutants were obtained. The subsequent phenotypic analysis of the respective strains showed no changes in ambruticin production, suggesting that this regulator does not influence ambruticin biosynthesis. Again, to make a valid statement both, the overexpression and the deletion mutants have to be examined.

Contrary to the *amb01* gene discussed above, *amb07* is involved in ambruticin biosynthesis. Not just that it resembles the presumed target of ambruticin in fungi, it is also found in every ambruticin/jerangolid producing strain sequenced to date as a part of the corresponding gene cluster. The similarity to the fungal target indicates a potential role in the perception of ambruticins within the cell. This finding may facilitate the construction of a sensor that can be used for screening of overproducers. However, direct experimental prove to support this hypothesis is missing so far. A targeted deletion of this gene led to a 10 fold drop in productivity, indicating that Amb07 somehow influences ambruticin biosynthesis via a regulation cascade. Unfortunately, up to date no overexpression mutants could be obtained which may give further insights into the real function of this protein.

The targeted deletion of *amb04* caused almost complete abolishment of ambruticin production, thereby demonstrating that Amb04 positively affects compound formation. According to this theory we assumed that a repatriation of *amb04* in the genome of the deletion mutant should restore the ability to produce ambruticins, but a phenotypic analysis of the complementation strains showed no differences in comparison to the parent strain. That effect might be caused by too high levels of Amb04 upon which the cell stops to synthesize ambruticin. An alleviated consequence could be observed as well for strains overexpressing the whole operon. Since these genetic experiments were not providing any evidence about the kind of regulation (direct or indirect), it was aimed to detect the binding sites of Amb04 within the *amb* cluster by EMSA analysis. As a result, no interaction of Amb04 with any presumed promoter regions of the biosynthetic genes was detected, thereby pointing towards an indirect regulatory mechanism. However, Amb04 seems to regulate its own expression and that of a small and so far undiscovered gene (*amb04a*, encoding a protein with an uncharacterized domain UPF0158). Elevated levels of the latter one might cause lack of compound formation and this is now subject of further investigations. Additionally, the EMSA experiments enabled us to define the consensus motif for Amb04 binding. Hence, we could propose a putative Amb04 regulon. This circuit consists of 10 genes/operons of which only two might be involved in regulatory processes. To evaluate these candidate genes comparative q-PCR experiments should be performed, which will give advanced insights into the process. All in all, it must be stated that Amb04 is regulating a process which directly or indirectly affects the biosynthesis of ambruticin, but how this takes place remains elusive.

After all, even if no improved ambruticin producer could be generated to date, numerous beneficial genetic tools originated out of this study which significantly improved *Sorangium* genetics.

Perspective

For initial completion of the functional analysis of genes (potentially) involved in ambruticin biosynthesis, mutants separately overexpressing the *ntcA* and *amb07* gene as well as an *amb01* deletion strain need to be constructed. Eventually, a comparison of the production properties from the deletion and the overexpression mutants will allow to draw conclusions about the actual functions of those genes. Knowledge gained hereby should then be used to improve compound yields. A drawback of this strategy is that an overexpression of dedicated genes cannot be performed using an inducible promoter. This fact may have prevented the construction of a P_{T7A1}_amb07 strain or the complementation of the Δ*amb04* phenotype, because of negative effects on the viability of resulting clones. A similar system is available for *M. xanthus* [109] and

future efforts should be undertaken to transfer it to the genus *Sorangium*. Especially for studying the effect of pleiotropic regulators like Amb04, such a procedure would be extremely valuable. This would permit the analysis of actual changes in the transcriptome that are causing a phenotype, rather than to examine steady-state conditions in which the cell has already adapted to the new circumstances.

Besides an examination of the first set of genes described above, a second compilation (Amb04 regulon), identified through a combinatorial approach of EMSAs and an *in silico* analysis, should be studied. First, candidates potentially involved in the regulation of ambruticin biosynthesis should be validated by q-PCR analysis and subsequently subjected to further studies by knock out and overexpression experiments. A primary examination of the Δ*amb04* strain has shown elevated levels of *amb04a* transcript, which is why this gene is now a target for deletion studies.

In addition to rational strain development, the classical way via UV or chemical mutagenesis may be taken into account as well. Therefore, a sensor system detecting ambruticin titers would be desirable in order to significantly reduce screening efforts. A comparative *in silico* analysis with the presumed targets of ambruticin in fungi suggests that Amb07 might bind ambruticin. Accordingly, the whole protein or its HAMP domains should be heterologously produced in *E. coli* to further investigate a potential binding of ambruticin by Surface Plasmon Resonance spectroscopy experiments.

References

[1] United Nations, Department of Economic and Social Affairs, Population Division (2013) World Population Prospects: The 2012 Revision, *Volume I; Comprehensive Tables ST/ESA/SER.A/336.*

[2] European Crop Protection (2012) Association Annual Report 2011-2012

[3] Berg D, Buchel KH, Plempel M, & Zywietz A (1986) Action mechanisms of cell-division-arresting benzimidazoles and of sterol biosynthesis-inhibiting imidazoles, 1,2,4-triazoles, and pyrimidines. *Mykosen*, 29, 221-229.

[4] Bartlett DW, Clough JM, Godwin JR, Hall AA, Hamer M, & Parr-Dobrzanski B (2002) The strobilurin fungicides. *Pest Manag. Sci.*, 58, 649-662.

[5] Becker WF, von JG, Anke T, & Steglich W (1981) Oudemansin, strobilurin A, strobilurin B and myxothiazol: New inhibitors of the bc_1 segment of the respiratory chain with an E-beta-methoxyacrylate system as common structural element. *FEBS Lett.*, 132, 329-333.

[6] Thierbach G & Reichenbach H (1981) Myxothiazol, a new inhibitor of the cytochrome bc_1 segment of th respiratory chain. *Biochim. Biophys. Acta*, 638, 282-289.

[7] Arima K, Imanaka H, Kousaka M, Fukuda A, & Tamura G (1965) Studies on pyrrolnitrin, a new antibiotic. I. Isolation and properties of pyrrolnitrin. *J. Antibiot. (Tokyo)*, 18, 201-204.

[8] Avenot H, Simoneau P, Iacomi-Vasilescu B, & Bataille-Simoneau N (2005) Characterization of mutations in the two-component histidine kinase gene AbNIK1 from *Alternaria brassicicola* that confer high dicarboximide and phenylpyrrole resistance. *Curr. Genet.*, 47, 234-243.

[9] Motoyama T, Kadokura K, Ohira T, Ichiishi A, Fujimura M, Yamaguchi I, & Kudo T (2005) A two-component histidine kinase of the rice blast fungus is involved in osmotic stress response and fungicide action. *Fungal. Genet. Biol.*, 42, 200-212.

[10] Zhang Y, Lamm R, Pillonel C, Lam S, & Xu JR (2002) Osmoregulation and fungicide resistance: The *Neurospora crassa os-2* gene encodes a HOG1 mitogen-activated protein kinase homologue. *Appl. Environ. Microbiol.*, 68, 532-538.

[11] Vetcher L, Menzella HG, Kudo T, Motoyama T, & Katz L (2007) The antifungal polyketide ambruticin targets the HOG pathway. *Antimicrob. Agents Chemother.*, 51, 3734-3736.

[12] Buschart A, Gremmer K, El-Mowafy M, van den Heuvel J, Mueller PP, & Bilitewski U (2012) A novel functional assay for fungal histidine kinases group III reveals the role of HAMP domains for fungicide sensitivity. *J. Biotechnol.*, 157, 268-277.

[13] Knauth P & Reichenbach H (2000) On the mechanism of action of the myxobacterial fungicide ambruticin. *J. Antibiot.*, 53, 1182-1190.

[14] Hohmann S (2002) Osmotic stress signaling and osmoadaptation in yeasts. *Microbiol. Mol. Biol. Rev.*, 66, 300-372.

[15] Ringel SM, Greenough RC, Romm S, Conor D, Gutt AL, Blair B, Kanter G, & von Strandtmann M (1977) Ambruticin (W7783), a new antifungal antibiotic. *J. Antibiot.*, 30, 371-375.

[16] Julien B, Tian ZQ, Reid R, & Reeves CD (2006) Analysis of the ambruticin and jerangolid gene clusters of *Sorangium cellulosum* reveals unusual mechanisms of polyketide biosynthesis. *Chem. Biol.*, 13, 1277-1286.

[17] Gerth K, Washausen P, Höfle G, Irschik H, & Reichenbach H (1996) The jerangolids: A family of new antifungal compounds from *Sorangium cellulosum* (Myxobacteria). Production, physico-chemical and biological properties of jerangolid A. *J. Antibiot.*, 49, 71-75.

[18] Kojima K, Takano Y, Yoshimi A, Tanaka C, Kikuchi T, & Okuno T (2004) Fungicide activity through activation of a fungal signalling pathway. *Mol. Microbiol.*, 53, 1785-1796.

[19] Hong W (2009) Agricultural products based on fluorinated heterocyclic compounds. In *Fluorinated Heterocyclic Compounds: Synthesis, Chemistry, and Application* (Petrov VA, ed), pp. 397-428. Wiley-Blackwell, Hoboken.

[20] Ringel SM (1990) New antifungal agents for the systemic mycoses. *Mycopathologia*, 109, 75-87.

[21] Liu P & Jacobsen EN (2001) Total synthesis of (+)-ambruticin. *J. Am. Chem. Soc.*, 123, 10772-10773.

[22] Kirkland TA, Colucci J, Geraci LS, Marx MA, Schneider M, Kaelin DE, Jr., & Martin SF (2001) Total synthesis of (+)-ambruticin S. *J. Am. Chem. Soc.*, 123, 12432-12433.

[23] Lee E, Choi SJ, Kim H, Han HO, Kim YK, Min SJ, Son SH, Lim SM, & Jang WS (2002) Total synthesis of ambruticin. *Angew. Chem. Int. Ed Engl.*, 41, 176-178.

[24] Wenzel SC & Müller R (2009) Myxobacteria-'microbial factories' for the production of bioactive secondary metabolites. *Mol. Biosyst.*, 5, 567-574.

[25] Buntin K, Irschik H, Weissman KJ, Luxenburger E, Blöcker H, & Müller R (2010) Biosynthesis of thuggacins in myxobacteria: Comparative cluster analysis reveals basis for natural product structural diversity. *Chem. Biol.*, 17, 342-356.

[26] Viehrig K, Surup F, Harmrolfs K, Jansen R, Kunze B, & Müller R (2013) Concerted action of P450 plus helper protein to form the amino-hydroxy-piperidone moiety of the potent protease inhibitor crocapeptin. *J. Am. Chem. Soc.*, 135, 16885-16894.

[27] Chen Y, Smanski MJ, & Shen B (2010) Improvement of secondary metabolite production in *Streptomyces* by manipulating pathway regulation. *Appl. Microbiol. Biotechnol.*, 86, 19-25.

[28] Rachid S, Gerth K, Kochems I, & Müller R (2007) Deciphering regulatory mechanisms for secondary metabolite production in the myxobacterium *Sorangium cellulosum* So ce56. *Mol. Microbiol.*, 63, 1783-1796.

[29] Rachid S, Gerth K, & Müller R (2008) NtcA-A negative regulator of secondary metabolite biosynthesis in *Sorangium cellulosum*. *J. Biotechnol.*, 140, 135-142.

[30] Krug D & Müller R (2014) Secondary metabolomics: The impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products. *Nat. Prod. Rep.*, 31, 768-783.

[31] Kopp M, Irschik H, Gross F, Perlova O, Sandmann A, Gerth K, & Müller R (2004) 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.*, 107, 29-40.

[32] Grant SG, Jessee J, Bloom FR, & Hanahan D (1990) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc. Natl. Acad. Sci. USA*, 87, 4645-4649.

[33] Fu J, Bian X, Hu S, Wang H, Huang F, Seibert PM, Plaza A, Xia L, Müller R, Stewart AF, & Zhang Y (2012) Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.*, 30, 440-446.

[34] Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principleof protein-dye binding. *Anal. Biochem.*, 72, 248-254.
[35] Sambrook J & Russell DW (2001) *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[36] Mullis K, Falona F, Scharf S, Saiki R, Horn G, & Erlich H (1986) Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.*, 51 Pt 1, 263-273.
[37] Müller S, Rachid S, Hoffmann T, Surup F, Volz C, Zaburannyi N, & Müller R (2014) Biosynthesis of crocacin involves an unusual hydrolytic release domain showing similarity to condensation domains. *Chem. Biol.*, 21, 855-865.
[38] Volz C, Kegler C, & Müller R (2012) Enhancer binding proteins act as hetero-oligomers and link secondary metabolite production to myxococcal development, motility and predation. *Chem. Biol.*, 19, 1447-1459.
[39] Höfle G, Steinmetz H, Gerth K, & Reichenbach H (1991) Ambruticins VS: New members of the antifungal ambruticin family from *Sorangium cellulosum*. *Liebigs Ann.*, 941-945.
[40] Zinder ND & Lederberg J (1952) Genetic exchange in *Salmonella*. *J. Bacteriol.*, 64, 679-699.
[41] Branda SS, Gonzalez-Pastor JE, Ben-Yehuda S, Losick R, & Kolter R (2001) Fruiting body formation by *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA*, 98, 11621-11626.
[42] Chu F, Kearns DB, Branda SS, Kolter R, & Losick R (2006) Targets of the master regulator of biofilm formation in *Bacillus subtilis*. *Mol. Microbiol.*, 59, 1216-1228.
[43] Chen I & Dubnau D (2004) DNA uptake during bacterial transformation. *Nat. Rev. Microbiol.*, 2, 241-249.
[44] Griffith F (1928) The Significance of Pneumococcal Types. *J. Hyg. (Lond)*, 27, 113-159.
[45] Meiers M, Volz C, Eisel J, Maurer P, Henrich B, & Hakenbeck R (2014) Altered lipid composition in *Streptococcus pneumoniae cpoA* mutants. *BMC Microbiol.*, 14, 12.
[46] Salzer R, Joos F, & Averhoff B (2014) Type IV pilus biogenesis, twitching motility, and DNA uptake in *Thermus thermophilus*: Discrete roles of antagonistic ATPases PilF, PilT1, and PilT2. *Appl. Environ. Microbiol.*, 80, 644-652.
[47] Papadakos KS, Sougleri IS, Mentis AF, & Sgouras DN (2013) A mutagenesis method for the addition and deletion of highly repetitive DNA regions: The paradigm of EPIYA motifs in the *cagA* gene of *Helicobacter pylori*. *Helicobacter.*, 18, 229-241.
[48] Lederberg J & Tatum EL (1946) Gene recombination in *Escherichia coli*. *Nature*, 158, 558.
[49] Kruse H & Sorum H (1994) Transfer of multiple drug resistance plasmids between bacteria of diverse origins in natural microenvironments. *Appl. Environ. Microbiol.*, 60, 4015-4021.
[50] Tenover FC (2006) Mechanisms of antimicrobial resistance in bacteria. *Am. J. Med.*, 119, S3-10.
[51] Kieser T, Bibb M, Buttner MJ, Chater KF, & Hopwood DA (2000) *Practical Streptomyces Genetics*. The John Innes Foundation, Norwich, England.
[52] Hill DS, Stein JI, Torkewitz NR, Morse AM, Howell CR, Pachlatko JP, Becker JO, & Ligon JM (1994) Cloning of genes involved in the synthesis of pyrrolnitrin from *Pseudomonas fluorescens* and role of pyrrolnitrin synthesis in biological control of plant disease. *Appl. Environ. Microbiol.*, 60, 78-85.
[53] Wenzel SC, Gross F, Zhang Y, Fu J, Stewart FA, & Müller R (2005) Heterologous expression of a myxobacterial natural products assembly line in pseudomonads via Red/ET recombineering. *Chem. Biol.*, 12, 349-356.
[54] Rachid S, Krug D, Kochems I, Kunze B, Scharfe M, Blöcker H, Zabriski M, & Müller R (2006) Molecular and biochemical studies of chondramide formation - highly cytotoxic natural products from *Chondromyces crocatus* Cm c5. *Chem. Biol.*, 14, 667-681.
[55] Neumann E, Schaefer-Ridder M, Wang Y, & Hofschneider PH (1982) Gene transfer into mouse lyoma cells by electroporation in high electric fields. *EMBO J.*, 1, 841-845.
[56] Dower WJ, Miller JF, & Ragsdale CW (1988) High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.*, 16, 6127-6145.
[57] Kimsey HH & Kaiser D (1991) Targeted disruption of the *Myxococcus xanthus* orotidine 5'-monophosphate decarboxylase gene: Effects on growth and fruiting-body development. *J. Bacteriol.*, 173, 6790-6797.
[58] Stamm I, Leclerque A, & Plaga W (1999) Purification of cold-shock-like proteins from *Stigmatella aurantiaca* - molecular cloning and characterization of the *cspA* gene. *Arch. Microbiol.*, 172, 175-181.
[59] Kopp M, Irschik H, Pradella S, & Müller R (2005) Production of the tubulin destabilizer disorazol in *Sorangium cellulosum*: Biosynthetic machinery and regulatory genes. *ChemBioChem*, 6, 1277-1286.
[60] Tu Y, Chen GP, & Wang YL (2007) Autonomously replicating plasmid transforms *Sorangium cellulosum* So ce90 and induces expression of green fluorescent protein. *J. Biosci. Bioeng.*, 104, 385-390.
[61] Jaoua S, Neff S, & Schupp T (1992) Transfer of mobilizable plasmids to *Sorangium cellulosum* and evidence for their integration into the chromosome. *Plasmid*, 28, 157-165.
[62] Xia ZJ, Wang J, Hu W, Liu H, Gao XZ, Wu ZH, Zhang PY, & Li YZ (2008) Improving conjugation efficacy of *Sorangium cellulosum* by the addition of dual selection antibiotics. *J. Ind. Microbiol. Biotechnol.*, 35, 1157-1163.
[63] Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, & Breitling R (2011) antiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.*, 39, W339-W346.
[64] Frias JE, Merida A, Herrero A, Martin-Nieto J, & Flores E (1993) General distribution of the nitrogen control gene *ntcA* in cyanobacteria. *J. Bacteriol.*, 175, 5710-5713.
[65] Vega-Palas MA, Madueno F, Herrero A, & Flores E (1990) Identification and cloning of a regulatory gene for nitrogen assimilation in the cyanobacterium *Synechococcus sp.* strain PCC 7942. *J. Bacteriol.*, 172, 643-647.
[66] Luque I, Flores E, & Herrero A (1994) Molecular mechanism for the operation of nitrogen control in cyanobacteria. *EMBO J.*, 13, 2862-2869.
[67] Herrero A, Muro-Pastor AM, & Flores E (2001) Nitrogen control in cyanobacteria. *J. Bacteriol.*, 183, 411-425.
[68] Pradella S, Hans A, Sproer C, Reichenbach H, Gerth K, & Beyer S (2002) Characterisation, genome size and genetic manipulation of the myxobacterium *Sorangium cellulosum* So ce56. *Arch. Microbiol.*, 178, 484-492.

[69] Jahns C, Hoffmann T, Müller S, Gerth K, Washausen P, Höfle G, Reichenbach H, Kalesse M, & Müller R (2012) Pellasoren: Structure elucidation, biosynthesis, and total synthesis of a cytotoxic secondary metabolite from *Sorangium cellulosum*. *Angew. Chem. Int. Ed Engl.*, 51, 5239-5243.
[70] Weinig S, Mahmud T, & Müller R (2003) Markerless mutations in the myxothiazol biosynthetic gene cluster: A delicate megasynthetase with a superfluous nonribosomal peptide synthetase domain. *Chem. Biol.*, 10, 953-960.
[71] Wu SS & Kaiser D (1996) Markerless deletions of *pil* genes in *Myxococcus xanthus* generated by counterselection with the *Bacillus subtilis sacB* gene. *J. Bacteriol.*, 178, 5817-5821.
[72] Sucipto H, Wenzel SC, & Müller R (2013) Exploring chemical diversity of α-pyrone antibiotics: Molecular basis of myxopyronin biosynthesis. *ChemBioChem*, 14, 1581-1589.
[73] Reyrat JM, Pelicic V, Gicquel B, & Rappuoli R (1998) Counterselectable markers: Untapped tools for bacterial genetics and pathogenesis. *Infect. Immun.*, 66, 4011-4017.
[74] Myronovskyi M, Welle E, Fedorenko V, & Luzhetskyy A (2011) Beta-glucuronidase as a sensitive and versatile reporter in actinomycetes. *Appl. Environ. Microbiol.*, 77, 5370-5383.
[75] Kuspa A, Kroos L, & Kaiser D (1986) Intercellular signaling is required for developmental gene expression in *Myxococcus xanthus*. *Dev Biol*, 117, 267-276.
[76] Kroos L & Kaiser D (1984) Construction of Tn5 *lac*, a transposon that fuses *lacZ* expression to exogenous promoters, and its introduction into *Myxococcus xanthus*. *Proc. Natl. Acad. Sci. USA*, 81, 5816-5820.
[77] Simon D & Ferretti JJ (1991) Electrotransformation of *Streptococcus pyogenes* with plasmid and linear DNA. *FEMS Microbiol. Lett.*, 66, 219-224.
[78] Dabert P & Smith GR (1997) Gene replacement with linear DNA fragments in wild-type *Escherichia coli*: Enhancement by Chi sites. *Genetics*, 145, 877-889.
[79] Zealey GR, Loosmore SM, Yacoob RK, Cockle SA, Boux LJ, Miller LD, & Klein MH (1990) Gene replacement in *Bordetella pertussis* by transformation with linear DNA. *Biotechnology*, 8, 1025-1029.
[80] Kempin SA, Liljegren SJ, Block LM, Rounsley SD, Yanofsky MF, & Lam E (1997) Targeted disruption in *Arabidopsis*. *Nature*, 389, 802-803.
[81] Shaked H, Melamed-Bessudo C, & Levy AA (2005) High-frequency gene targeting in *Arabidopsis* plants expressing the yeast RAD54 gene. *Proc. Natl. Acad. Sci. USA*, 102, 12265-12269.
[82] Rothstein R (1991) Targeting, disruption, replacement, and allele rescue: Integrative DNA transformation in yeast. *Methods Enzymol.*, 194, 281-301.
[83] Nayak T, Szewczyk E, Oakley CE, Osmani A, Ukil L, Murray SL, Hynes MJ, Osmani SA, & Oakley BR (2006) A versatile and efficient gene-targeting system for *Aspergillus nidulans*. *Genetics*, 172, 1557-1566.
[84] Hensel M, Arst HN, Jr., Aufauvre-Brown A, & Holden DW (1998) The role of the *Aspergillus fumigatus areA* gene in invasive pulmonary aspergillosis. *Mol. Gen. Genet.*, 258, 553-557.
[85] Miller BL, Miller KY, & Timberlake WE (1985) Direct and indirect gene replacements in *Aspergillus nidulans*. *Mol. Cell Biol.*, 5, 1714-1721.
[86] Li J & Baker MD (2000) Mechanisms involved in targeted gene replacement in mammalian cells. *Genetics*, 156, 809-821.
[87] Thomas KR, Folger KR, & Capecchi MR (1986) High frequency targeting of genes to specific sites in the mammalian genome. *Cell*, 44, 419-428.
[88] Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, & Court DL (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, 97, 5978-5983.
[89] Karlinsey JE (2007) lambda-Red genetic engineering in *Salmonella enterica* serovar Typhimurium. *Methods Enzymol.*, 421, 199-209.
[90] Datsenko KA & Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA*, 97, 6640-6645.
[91] Maddocks SE & Oyston PC (2008) Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology*, 154, 3609-3623.
[92] Stragier P, Richaud F, Borne F, & Patte JC (1983) Regulation of diaminopimelate decarboxylase synthesis in Escherichia coli. I. Identification of a *lysR* gene encoding an activator of the *lysA* gene. *J. Mol. Biol.*, 168, 307-320.
[93] Stragier P & Patte JC (1983) Regulation of diaminopimelate decarboxylase synthesis in *Escherichia coli*. III. Nucleotide sequence and regulation of the *lysR* gene. *J. Mol. Biol.*, 168, 333-350.
[94] Lindquist S, Lindberg F, & Normark S (1989) Binding of the *Citrobacter freundii* AmpR regulator to a single DNA site provides both autoregulation and activation of the inducible *ampC* beta-lactamase gene. *J. Bacteriol.*, 171, 3746-3753.
[95] Parsek MR, McFall SM, Shinabarger DL, & Chakrabarty AM (1994) Interaction of two LysR-type regulatory proteins CatR and ClcR with heterologous promoters: functional and evolutionary implications. *Proc. Natl. Acad. Sci. USA*, 91, 12393-12397.
[96] Heroven AK & Dersch P (2006) RovM, a novel LysR-type regulator of the virulence activator gene *rovA*, controls cell invasion, virulence and motility of *Yersinia pseudotuberculosis*. *Mol. Microbiol.*, 62, 1469-1483.
[97] Hernandez-Lucas I, Gallego-Hernandez AL, Encarnacion S, Fernandez-Mora M, Martinez-Batallar AG, Salgado H, Oropeza R, & Calva E (2008) The LysR-type transcriptional regulator LeuO controls expression of several genes in *Salmonella enterica* serovar Typhi. *J. Bacteriol.*, 190, 1658-1670.
[98] Bartowsky E & Normark S (1993) Interactions of wild-type and mutant AmpR of *Citrobacter freundii* with target DNA. *Mol. Microbiol.*, 10, 555-565.
[99] Parsek MR, Shinabarger DL, Rothmel RK, & Chakrabarty AM (1992) Roles of CatR and cis,cis-muconate in activation of the *catBC* operon, which is involved in benzoate degradation in *Pseudomonas putida*. *J. Bacteriol.*, 174, 7798-7806.
[100] Celis RT (1999) Repression and activation of arginine transport genes in *Escherichia coli* K 12 by the ArgP protein. *J. Mol. Biol.*, 294, 1087-1095.
[101] Catlett NL, Yoder OC, & Turgeon BG (2003) Whole-genome analysis of two-component signal transduction genes in fungal pathogens. *Eukaryot. Cell*, 2, 1151-1161.
[102] El-Mowafy M, Bahgat MM, & Bilitewski U (2013) Deletion of the HAMP domains from the histidine kinase CaNik1p of *Candida albicans* or treatment with fungicides activates the MAP kinase Hog1p in *S. cerevisiae* transformants. *BMC Microbiol.*, 13, 209.

[103] Aravind L & Ponting CP (1999) The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. *FEMS Microbiol. Lett.*, 176, 111-116.
[104] Yoshida T, Phadtare S, & Inouye M (2007) Functional and structural characterization of EnvZ, an osmosensing histidine kinase of *E. coli*. *Methods Enzymol.*, 423, 184-202.
[105] Kaur H, Singh S, Rathore YS, Sharma A, Furukawa K, Hohmann S, Ashish, & Mondal AK (2014) Differential role of HAMP-like linkers in regulating the functionality of the group III histidine kinase DhNik1p. *J. Biol. Chem.*, 289, 20245-20258.
[106] Heikaus CC, Pandit J, & Klevit RE (2009) Cyclic nucleotide binding GAF domains from phosphodiesterases: Structural and mechanistic insights. *Structure.*, 17, 1551-1557.
[107] Ho YS, Burden LM, & Hurley JH (2000) Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. *EMBO J.*, 19, 5288-5299.
[108] Deuschle U, Kammerer W, Gentz R, & Bujard H (1986) Promoters of *Escherichia coli*: A hierarchy of in vivo strength indicates alternate structures. *EMBO J.*, 5, 2987-2994.
[109] Iniesta AA, García-Heras F, Abellón-Ruiz J, Gallego-García A, & Elías-Arnanz M (2012) Two systems for conditional gene expression in *Myxococcus xanthus* inducible by isopropyl-ß-D-thiogalactopyranoside or vanillate. *J. Bacteriol.*, 194, 5875-5885.
[110] Jenkins LS & Nunn WD (1987) Regulation of the *ato* operon by the *atoC* gene in *Escherichia coli*. *J. Bacteriol.*, 169, 2096-2102.
[111] Pauli G & Overath P (1972) ato Operon: A highly inducible system for acetoacetate and butyrate degradation in *Escherichia coli*. *Eur. J. Biochem.*, 29, 553-562.
[112] Lioliou EE & Kyriakidis DA (2004) The role of bacterial antizyme: From an inhibitory protein to AtoC transcriptional regulator. *Microb. Cell Fact.*, 3, 8.
[113] Lioliou EE, Mimitou EP, Grigoroudis AI, Panagiotidis CH, Panagiotidis CA, & Kyriakidis DA (2005) Phosphorylation activity of the response regulator of the two-component signal transduction system AtoS-AtoC in *E. coli*. *Biochim. Biophys. Acta*, 1725, 257-268.
[114] Goulian M (2010) Two-component signaling circuit structure and properties. *Curr. Opin. Microbiol.*, 13, 184-189.
[115] Geer LY, Domrachev M, Lipman DJ, & Bryant SH (2002) CDART: Protein homology by domain architecture. *Genome Res.*, 12, 1619-1623.
[116] Collart FR & Huberman E (1988) Cloning and sequence analysis of the human and Chinese hamster inosine-5'-monophosphate dehydrogenase cDNAs. *J. Biol. Chem.*, 263, 15769-15772.
[117] Mehra RK & Drabble WT (1981) Dual control of the gua operon of *Escherichia coli* K12 by adenine and guanine nucleotides. *Journal of General Microbiology*, 123, 27-37.
[118] Pao SS, Paulsen IT, & Saier MH, Jr. (1998) Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.*, 62, 1-34.
[119] Allocati N, Federici L, Masulli M, & Di IC (2009) Glutathione transferases in bacteria. *FEBS J.*, 276, 58-75.
[120] Allocati N, Favaloro B, Masulli M, Alexeyev MF, & Di IC (2003) *Proteus mirabilis* glutathione S-transferase B1-1 is involved in protective mechanisms against oxidative and chemical stresses. *Biochem. J.*, 373, 305-311.
[121] Settembre EC, Dorrestein PC, Zhai H, Chatterjee A, McLafferty FW, Begley TP, & Ealick SE (2004) Thiamin biosynthesis in *Bacillus subtilis*: Structure of the thiazole synthase/sulfur carrier protein complex. *Biochemistry*, 43, 11647-11657.
[122] Bunik VI, Tylicki A, & Lukashev NV (2013) Thiamin diphosphate-dependent enzymes: From enzymology to metabolic regulation, drug design and disease models. *FEBS J.*, 280, 6412-6442.
[123] Wu J & Rosen BP (1991) The ArsR protein is a trans-acting regulatory protein. *Mol. Microbiol.*, 5, 1331-1336.
[124] Silver S, Budd K, Leahy KM, Shaw WV, Hammond D, Novick RP, Willsky GR, Malamy MH, & Rosenberg H (1981) Inducible plasmid-determined resistance to arsenate, arsenite, and antimony (III) in *Escherichia coli* and *Staphylococcus aureus*. *J. Bacteriol.*, 146, 983-996.
[125] Pereira SF, Goss L, & Dworkin J (2011) Eukaryote-like serine/threonine kinases and phosphatases in bacteria. *Microbiol. Mol. Biol. Rev.*, 75, 192-212.

.

Final Discussion

Studies towards the formation of crocacin in *Chondromyces crocatus* Cm c5

The crocacins are natural products isolated from various strains of the genus *Chondromyce*s [1;2]. To date, four different derivatives (crocacin A-D) are known (Figure 1).

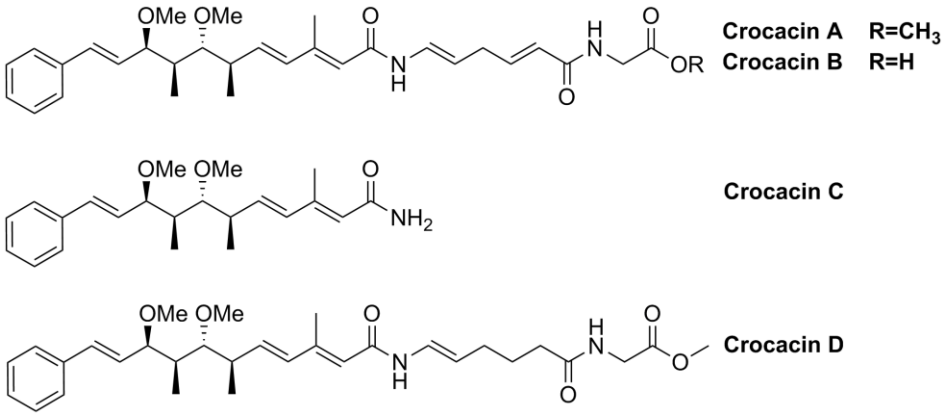


Figure 1: Chemical structures of the known crocacin derivatives. Crocacin A-C were isolated from extracts of different *C. crocatus* strains, whereas crocacin D was purified from an extract of *C. pediculatus* Cm p17.

These linear dipeptides are characterized by the presence of a reactive *N*-acyl enamide functionality also known from the apicularens [3], the lobatamides [4], and salicylihalamide A [5]. While crocacin A is the main derivative produced by *C. crocatus* Cm c5, crocacin B is only present in small amounts. Additionally, the isolation of crocacin C was reported, but as it was stated, this derivative might be formed during the isolation process and thus, represents an artifact [1]. Furthermore, crocacin D was purified from an extract of *C. pediculatus* Cm p17 which is lacking the Δ^5-double bond in comparison to crocacin A.

The crocacins exhibit strong antifungal and cytotoxic properties, whereas no notable antimicrobial effects have been described [2]. However, distinct differences in the activities of the respective derivatives were detected. Crocacin D is exhibiting good properties against both, fungi and mouse fibroblast cells, while crocacin C was less effective. Probably, the loss of the enamide functionality in crocacin C is accompanied with reduction of activity. A similar effect was already reported for a synthetic version of salicylihalamide, where the saturation of the enamide double bond results in a diminished activity [6]. Mode of action studies identified the respiratory chain as potential target of the crocacins, and further experiments revealed that the electron flow within the cytochrome bc_1 complex is inhibited, likewise to myxothiazol [7] and stigmatellin [8].

Besides structure elucidation, activity profiling and mode of action studies, even the total synthesis of all crocacin derivatives has already been accomplished [9-11]. After all, a detailed understanding of how the compound is synthesized by its natural producer was missing. Hence, we set out to discover and characterize the crocacin biosynthetic pathway.

To detect a candidate gene cluster in the genome of a producer, several procedures can be employed. One of the oldest methods relies on a combination of a random transposon mutagenesis, together with bioassay and/or LC-MS screenings of the clones, followed by the recovery of the transposable element from a strain that is impaired to produce the target molecule. Based on available genome data from *Sorangium cellulosum* So ce56, Kopp *et al.* could identify the chivosazol biosynthetic gene cluster in *S. cellulosum* So ce12 in a similar approach [12]. Another option for the assignment of a gene cluster to a certain compound is depending on retrobiosynthetic considerations, which permit the directed analysis of a chromosomal library from the producer (e.g. cosmid or fosmid), and finally the confirmation of the suspected locus by means of knock out experiments. Accordingly, among others, the gene clusters bearing the information for chondramid and chondrochloren production in *C. crocatus* Cm c5 were identified [13;14]. Today, with recent advances in genome sequencing, the most common approach employs bioinformatic tools such as antiSMASH. These programs enable the identification of secondary metabolite loci and help to predict building blocks of natural products based on the specific domain architecture as found for a given loci [15;16]. Though, the underlying pathways are often too complex to allow precise prediction of their products. Nevertheless, since myxobacterial genomes are usually encoding a plurality of different natural product assembly lines, such programs help to uncover potential candidates. Once a gene cluster has been assigned to a certain compound, detailed biosynthetic studies can be performed.

For primary identification of the crocacin gene cluster in the genome of *C. crocatus* Cm c5 Dr. Shwan Rachid screened a chromosomal library of approximately 2,300 clones for the presence of A domains, predicted to activate the amino acid glycine (this building block is used twice during biosynthesis). As additionally anticipated from the chemical structure, crocacin should be derived from a NRPS/PKS hybrid. Consequently, positive cosmids were further probed for PKS characteristics and promising candidates were end sequenced. This resulted in the identification of a cosmid (E:I22) containing the 3' part of the crocacin gene cluster. Based on the 5' end sequence of E:I22, cosmid A:C6 could be identified, which was shown to comprise the remaining part of the cluster. Later, the chromosome of *C. crocatus* Cm c5 has been sequenced and the resulting draft genome was subjected to an antiSMASH analysis. Among the 35 putative biosynthetic gene clusters that were detected, only one locus fulfilled the

requirements of the retrobiosynthetic proposal. As expected, this cluster was identical to the one previously identified by Dr. Shwan Rachid. Hence, identification of the crocacin gene cluster by using the antiSMASH tool was straightforward, although some discrepancies between the predicted and the actual structure were apparent (see below).

As reasoned from feeding experiments with labeled precursors, crocacin biosynthesis is initiated through the activation of *trans*-cinnamic acid catalyzed by the N-terminal CoA-ligase of CroA followed by its transfers to the first ACP. However, the use of *trans*-cinnamic acid as starter unit could not be predicted by the antiSMASH pipeline. This might be due to the fact that CoA-ligases involved in natural product biosynthesis activate a wide range of substrates (e.g. linear or branched fatty acids, benzoate, or phenylacetate). In addition, the number of characterized *trans*-cinnamic acid employing CoA-ligases from bacterial secondary metabolite pathways reported in literature is exceedingly rare (StlB from *Photorhabdus luminescens* [17] and EncH from *Streptomyces maritimus* [18]), which is why no common consensus motifs could be determined yet. Noteworthy, chain initiation is often catalyzed either by an AT-ACP loading module selecting acetate or propionate units, or by a KS^{Q}-AT-ACP module using malonate or methylmalonate [19]. In those cases, prognosis of the starter molecule is usually correct.

In contrast to the starter unit, the prediction of the extender units employed in crocacin biosynthesis was in agreement with the structure of the crocacins. The reason for that is rather simple: The selection of building blocks by A and AT domains has been intensively studied by mutagenesis, sequence analysis, crystal structures, and biochemical experiments [20-24]. Accordingly, if biosynthesis is carried out by a bacterial type I/II PKS, or an NRPS pathway, the chemical structure of the core scaffold can be predicted precisely. Still, if certain PKS modules are used iteratively or uncommon biosynthetic entities are present within the assembly line, prediction fails. In case of the crocacin pathway, an iteratively acting PKS module, alongside with an unusual β-branching cassette and the lack of a termination domain were disturbing a proper structure prediction based on *in silico* analysis tools. Consequently, a detailed model of the biosynthesis was devised as discussed below.

Nowadays, iteratively acting PKS units can be identified by probing the corresponding KS domains [25;26]. In a phylogenetic analysis CroC-KS groups at the interface between iterative and modular KS domains, thereby corroborating our proposal that module 3 is used twice during crocacin assembly (Figure 2).

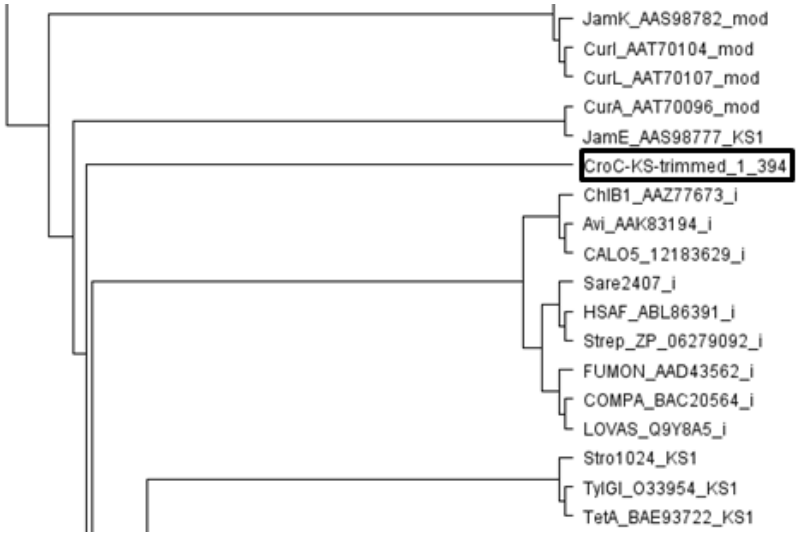


Figure 2: Detail of a phylogenetic analysis of the primary sequence of CroC-KS using the NaPDoS tool [26]. mod: modular KSs, KS1: KSs of multimodular enzymes, including KSs incorporating starter and unusual building blocks, i: iteratively acting KSs.

In contrast to the general assignment of iteratively acting entities, the prediction of the number of extension cycles that are carried out by a distinct module is almost impossible to date. Yadav *et al.* [25] are assuming that the cavity volumes of the substrate binding pocket from different iterative type I KSs, seem to control the repetition count; the larger the cavity, the more cycles are carried out. This particular cavity is determined by a set of key amino acid residues, which may in turn serve as indicators in a primary sequence alignment. Although this finding could help to assess the actual number of extensions, a precise prediction is not possible. A bioinformatic tool that is able to reliably predict the number of extension cycles would be of great biotechnological value in order to alter the number of iterations in type I PKSs by site directed mutagenesis. The more biosynthetic pathways will be characterized, the better the understanding of the underlying mechanisms will be and this may promote the development of such a tool in the future.

The identification of the β-branching cassette was puzzling at a first glance. Prototypes of these systems, as found in the mupirocin [27] or bacillaene [28;29] pathways, comprise six distinct proteins: a free standing ACP, a free standing KS that lacks the conserved cysteine residue required for chain extension, a hydroxymethylglutaryl-CoA synthase homolog (HMG), and finally two enoyl-CoA hydratase/isomerase homologs (ECH) [30]. In case of crocacin, a KS domain, which still features the cysteine residue, and an AT domain are present on the bifunctional protein CroD, while the remaining parts of the β-branching cassette meet the expectations. However, it is known that some β-branching cassettes are deviating from the

standard ones, which is why CroD was suspected to pick and decarboxylate the malonyl-CoA building block needed for C13 methylation. Another possibility could have been that CroD is somehow taking part in chain extension and therefore CroC would not act iteratively. The lack of an ACP domain in CroD made this scenario rather unlikely. Moreover, this proposal would demand a trans-acting KS and AT pair encoded at another locus in the chromosome. Since such a pair could not be detected in the draft genome of *C. crocatus* Cm c5, all evidence led to the assumption that CroD constitutes to the β-branching machinery. To unambiguously prove this proposal, the reaction needs to be reconstituted *in vitro*. In this certain case, knock out experiments by single cross over are not sufficient, as it is likely that expression of the downstream genes is perturbed due to polar effects. This is reasoned form the *croE* (HMG) and *croF* (ECH) inactivations, where either inactivation led to complete abolishment of crocacin production. Contrary to this, in-frame deletions would rather result in the C13 non-methylated version of the crocacins, but it cannot be excluded that this intermediate may not be accepted as a substrate by the subsequent C domain. Unfortunately, due to the current lack of an in-frame deletion system for *C. crocatus* Cm c5, such experiments cannot be performed to date. However, similar β-branch incorporating systems from other secondary metabolite pathways (e.g. bacillaene [28;29;31], bryostatin [32;33], and myxovirescin A (see below)) have already been studied intensively. Particularly, the investigations towards the formation of myxovirescin, which possesses two structurally distinct β-branches at C12 (β-methoxymethyl) and at C16 (β-ethyl), revealed several mechanistic features [34-36]. In detail, through targeted gene deletions in the native producer *Myxococcus xanthus* DK1622, phenotypic characterization of the respective mutants, and structure elucidation of the novel analogue myxovirescin ΔF, several conclusions could be drawn: 1) the two ATs present on TaV load malonyl-CoA and methylmalonyl-CoA onto their cognate ACP domains (TaB and TaE); 2) TaK decarboxylates the malonyl- and methylmalonyl-S-ACP intermediates; 3) the ACP domains, TaB and TaE, which provide the nucleophile, specifically interact with their corresponding HMGs (TaC and TaF); 4) the HMGs are the key-determinants for the identity and the location of the β-branch in the polyketide. Besides these *in vivo* studies, Calderone *et al.* performed a series of *in vitro* experiments with heterologously produced entities from the myxovirescin machinery, which corroborated the conclusions made before [37].

By far the most unusual feature of crocacin assembly is the hydrolytic product release mediated by the terminal C domain CroK-C2. On the basis of our inquiries, we concluded that CroK-C2 catalyzes a novel and currently unique reaction among C domains from NRPS systems. To take advantage of this particular finding the molecular foundation that allows a C domain to act hydrolytically needs to be elucidated. Hence, attempts to crystalize the protein

together with the substrate mimic *N*-acetylcysteaminyl-crocacin B were undertaken. With respect to this, initial trials carried out by Dr. Peer Lukat (Helmholtz Institute for Pharmaceutical Research Saarland (MINS), Saarbrücken, Germany) did not yield in any crystals. It should, however, be noted that only a limited number of crystallization conditions were tested to date. Thus, a systematic approach testing a variety of settings should be undertaken in order to determine the three dimensional structure of CroK-C2. Such efforts seem to be promising, because the protein can be produced in good yields, high purity and most importantly in a functional state. However, based on the *in silico* analysis of this domain it can be speculated how hydrolysis of the assembled product might be accomplished. As outlined in Chapter I, both, the N- and the C-terminus of CroK-C2 deviate from ordinary C domains, while the middle part, including the common HHXXXDG consensus sequence, is more conserved. The second histidine residue of the aforementioned motif is regarded crucial for enzymatic activity. Site-directed mutagenesis in the heterologously produced C domain of module 2 from the tyrocidine pathway (H147V), and in the C domain of module F from the enterobactin synthetase (H138A) abolished activity in condensations assays [38;39]. In particular, it was proposed that the respective histidine residue functions as catalytic base which deprotonates the amine functionality of the acceptor that in turn attacks the carbonyl moiety of the donor, resulting in amide bond formation. Accordingly, CroK-C2 could facilitate hydrolysis of crocacin analogous, with the difference that the acceptor molecule is not a PCP tethered amino acid, but a water molecule. Nevertheless, the fact that CroK-C2 does not group with classical C domains in a NaPDoS analysis (Chapter I) suggest a different mechanism; maybe that of VibH. VibH is a single standing C domain taking part in vibriobactin biosynthesis. Likewise to CroK-C2, VibH catalyzes the chain release of a carrier protein bound intermediate (2,3-hydroxybenzoyl-S-VibB) through condensation with a freely diffusible substrate (norspermidine), thereby regenerating holo-VibB [40]. Thus, these processes are likely comparable. Keating *et al.* solved the crystal structure of VibH and examined important amino acid residues using site-directed mutagenesis [41]. Surprisingly, they revealed that, despite a favorable positioning of the conserved histidine residues at the active site of the protein, specific alterations did not significantly affect the enzyme's activity, indicating that these particular residues are not critical for base catalysis in VibH. Therefore, amide bond formation might be permitted by selective binding of just the deprotonated version of norspermidine, likewise to VanA ligase from vancomycin-resistant *Enterococci*, which largely rejects the protonated form of its substrate D-alanine (no need for base catalysis) [42]. However, considering that the condensation of both molecules is thermodynamically favored over the existing thioester, VibH may achieve linking through binding of both substrates in close proximity in an adequate surrounding. The second possibility may

apply as well for crocacin biosynthesis, but given the high turnover rate determined for CroK-C2, an additional activation of the water molecule is likely.

In summary, all experiments performed within this study allowed the proposal of a detailed model for crocacin biosynthesis in *C. crocatus* Cm c5. As mentioned above, *C. pediculatus* Cm p17 is producing another derivative of the crocacins, crocacin D (Figure 1). In comparison to crocacin A, this derivative is lacking the Δ^5-double bond, which is why the underlying biosynthetic machinery is supposed to be highly similar. The sole exception should be that module 5 from *C. pediculatus* Cm p17 comprises an additional ER domain in order to fully reduce the malonyl-CoA extender unit. From a biosynthetic point of view the presence of the methylester crocacin D requests the existence of the free acid as found in crocacin B. To prove this hypothesis, a methanolic extract of *C. pediculatus* Cm p17 was analyzed by HPLC-MS, particularly with regard to crocacin D and its postulated precursor (Figure 3).

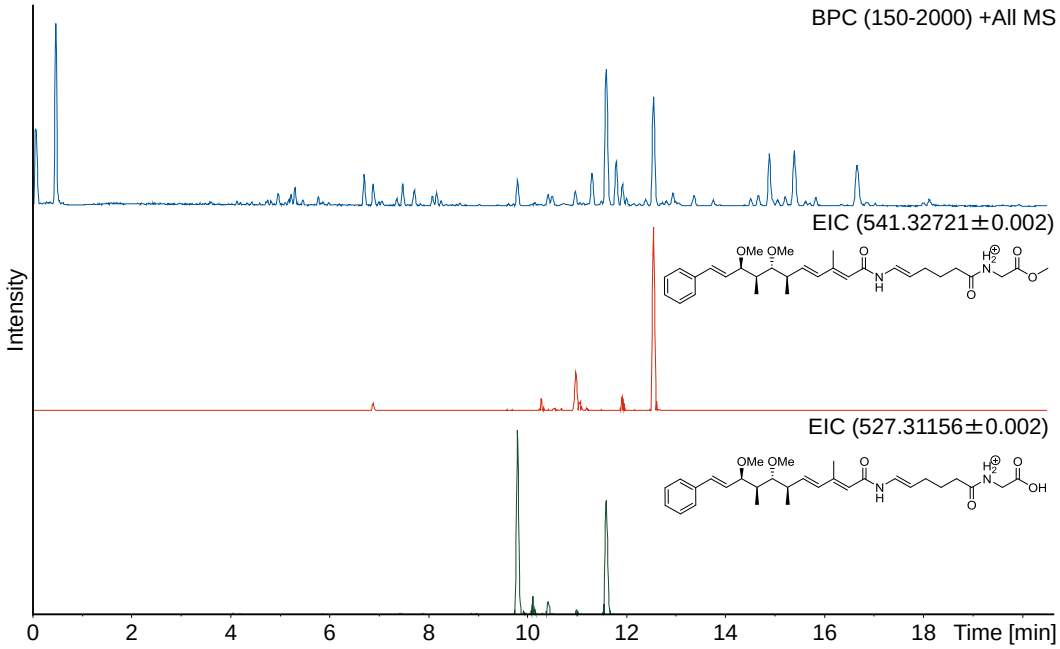


Figure 3: HPLC-MS profile of a methanolic extract from *C. pediculatus* Cm p17. Blue trace: base peak chromatogram, red trace: extracted ion chromatogram of crocacin D (541.32721±0.002 *m/z*), green trace: extracted ion chromatogram of the postulated precursor of crocacin D (527.31156±0.002 *m/z*).

As reasoned from Figure 3, the extract contained high amounts of crocacin D and additional signals observed in the respective extracted ion chromatogram are most likely caused by isomers. Besides that, chromatographic peaks with the expected *m/z* value for the putative precursor of crocacin D were detected as well. Further evidence to corroborate this biosynthetic hypothesis comes from a comparison of the MS2 fragmentation pattern of crocacin D and the two main isomers of the putative precursor (Figure 4).

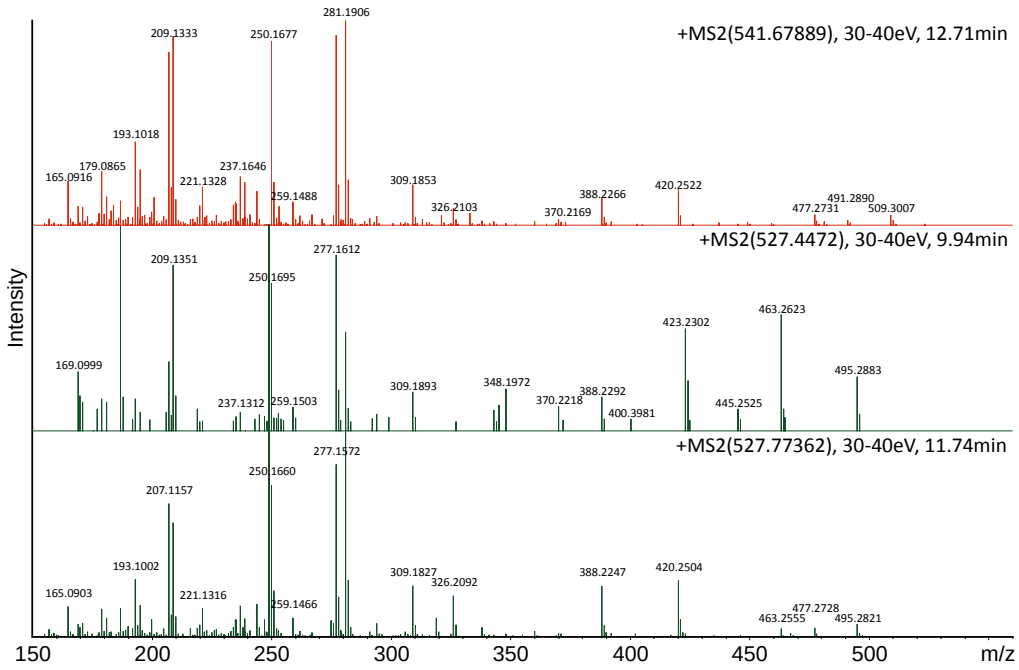


Figure 4: MS2 fragmentation pattern of crocacin D (red trace) and two of the putative precursor isomers eluting at 9.94 and 11.74 min (green traces).

The fragmentation pattern of the putative precursors show significant homology to that of crocacin D. Moreover, in some cases, a shift by 14 *m/z* between the red and the green trace is observed, indicating that the precursor might lack a methyl group, which is in agreement when comparing a free acid and its corresponding methyl ester. Accordingly, and in line with the biosynthetic investigations in *C. crocatus* Cm c5, it is proposed that *C. pediculatus* Cm p17 is producing a yet unidentified crocacin derivative (crocacin E), that serves as precursor for the main product crocacin D. For unambiguous confirmation of this hypothesis, the purified compound needs to be subjected to NMR analysis and the appendant biosynthetic gene cluster of *C. pediculatus* Cm p17 must be examined.

Microsclerodermins from terrestrial myxobacteria

The microsclerodermins are a group of macrocyclic natural products initially identified from extracts of lithistid sponges of the species *Microscleroderma* and *Theonella*. Microsclerodermins, which exhibit potent antifungal and cytotoxic activities, were isolated in reasonable amounts ranging from 0.005 to 0.19 % of biomass dry weight from the respective sponges [43-46]. Access to this natural host is, however, rather complicated. For example, the *Microscleroderma* species which the first microsclerodermins (A-B) were purified from was

collected at a depth of around 300 meters in the Norfolk Rise near New Caledonia. Hence, an easily accessible and sustainable source of the microsclerodermins would be desirable, in order to further evaluate the compounds properties by extended biological and chemical profiling. Noteworthy, pedein A and B, which are derivatives of the microsclerodermins, have been previously isolated from the terrestrial myxobacterium *Chondromyces pediculatus*, thereby providing such a source [47]. In the course of our study, a set of terrestrial bacteria from our strain collection was identified, which, in addition to some of the already known microsclerodermins and pedeins, produce the novel derivatives microsclerodermin L and M. In contrast to the sponges, these microbes are relatively simple to cultivate, and genetic manipulation procedures for at least two of the three producing genera (*Chondromyces* and *Sorangium*) are available. Given the fact, that *Sorangium cellulosum* So ce38 synthesizes microsclerodermin M at a yield of 12 mg/l, it can be expected that systematic strain optimization can result in even higher titers. Accordingly, large scale production of at least some of the microsclerodermins becomes be possible.

One of the most interesting facts revealed in this survey is the identification of terrestrial organisms producing exactly the same compound as previously isolated from a marine sponge. More precisely, the myxobacteria *Jahnella* sp. MSr9139, *Jahnella* sp. PI3954, *Chondromyces* sp. Cmp3, and *Chondromyces* sp. CmT292 all produce microsclerodermin D, which was initially identified in marine sponges from *Microscleroderma* and *Theonella* species. Results from a recent study provide a strong hint that myxobacteria of the genera *Chondromyces* and *Sorangium* are associated with marine sponges [48]. To date, it remains unclear whether the microsclerodermins originate from the sponge or the bacteria. The fact that bacteria produce the same compound as found in sponges points toward a bacterial producer that lives inside the sponge. To address this question, compound localization studies might help on the way to identify the actual producer. Such experiments rely on the fact that sponges dissociate into individual cells if they are transferred into artificial see water lacking magnesium and calcium ions. Further separation of bacterial and sponge cells is achieved either by gradient centrifugation or flow cytometry [49-51]. Using these methods a number of molecules could be attributed to bacterial producers. For example theopalauamide and swinholide A, both isolated from *Theonella swinhoei*, were shown to be localized in distinct populations of bacterial symbionts [52;53]. Besides that, it was demonstrated that halogenated compounds present in the sponge *Dysidea herbacea* are associated with the cyanobacterial symbiont *Oscillatoria spongeliae* [49;54]. Although such investigations provide good evidence for a bacterial source of these compounds, it cannot be ruled out, that other natural products isolated from sponges are accumulated by a sequestration mechanism. Hence, the genetic material encoding the

respective pathways needs to be detected in the bacterial fractions. Since in the course of our study the genetic foundation of microsclerodermin biosynthesis was unambiguously identified, simple PCR or Southern Blot analysis would be sufficient for identification of the gene cluster in sponge samples.

As difficult the identification of the real producers might be, as easy is the answer to the question: “Why are sponges accumulating bioactive natural products like the microsclerodermins”? Sponges are sessile organisms which do not exhibit good mechanical defense mechanisms and since they are under constant predatory pressure, they accumulate chemical protectants [55]. Given that most of these predators are eukaryotic organisms like fish or turtles, it is not surprising that the majority of compounds isolated from sponges exhibit cytotoxic activities [56]. Actually, some of those molecules have already been evaluated in clinical trials [57], among which are the bengamides, produced as well by a terrestrial myxobacterium [58], and microsclerodermin A [59].

The microbiome of a sponge usually comprises a diverse range of symbiotic microorganisms with broad physiological activities that can constitute up to 35 % of its biomass [60]. These bacteria are involved in various dissimilatory as well as assimilatory processes providing essential primary metabolites or energy by means of photosynthesis to the sponge [61]. Conversely, through extensive filterfeeding (1 kg sponge can filter up to 24,000 l of seawater per day) and excretion of ammonia as a metabolic end product, the sponge represents a nutrient-rich environment for the symbionts. Usually, the microbes inhabit the mesophyl that represents the sponge’s “gut”. To prevent their own digestion, some symbiotic bacteria may shield themselves by producing slime-capsules [62]. As we postulate in our study that the microsclerodermins are rather produced by a symbiotic bacterium of the suborder *Sorangineae* than by the sponge itself, the scenario could be as follows: A dedicated myxobacterium is producing the microsclerodermins, thereby supplying the sponge with a potent toxin. Furthermore, the myxobacterium could contribute to proteolytic degradation of complex substrates as well. To avoid being consumed by the sponge, the microbes will produce exo- and lipopolysaccharides, which in fact is typical to myxobacteria. In return, the sponge hosts the bacterium and assures a constant nutrient supply.

Access to the microbial microsclerodermin producers enabled us to identify and to characterize the biosynthetic gene cluster bearing the information for compound production. Particularly, a retrobiosynthetic proposal together with site-directed gene inactivation allowed unambiguous assignment of the candidate cluster in the draft genome of *S. cellulosum* So ce38. Nevertheless, this DNA sequence exhibited gaps which needed to be closed to finalize a valid biosynthetic model for microsclerodermin production. This is a frequent issue when describing

gene clusters from myxobacterial draft genomes. Myxobacterial genomes are rather large (usually more than 10 MB) with a high GC content and repetitive sequences, thus next generation sequencing techniques like Illumina or 454 and assembly of the sequencing reads often fail to produce good results. Accordingly, remaining gaps need to be closed "manually" either by PCR amplification, subcloning, and sequencing of the missing pieces or by preparing, screening, and sequencing of genomic libraries. To close the three gaps that were existent in the microsclerodermin cluster both approaches were pursued (Figure 5). In the end, all gaps were closed, but the overall procedure was very laborious and time consuming. Nevertheless, it was an essential prerequisite for progressing within the microsclerodermin project.

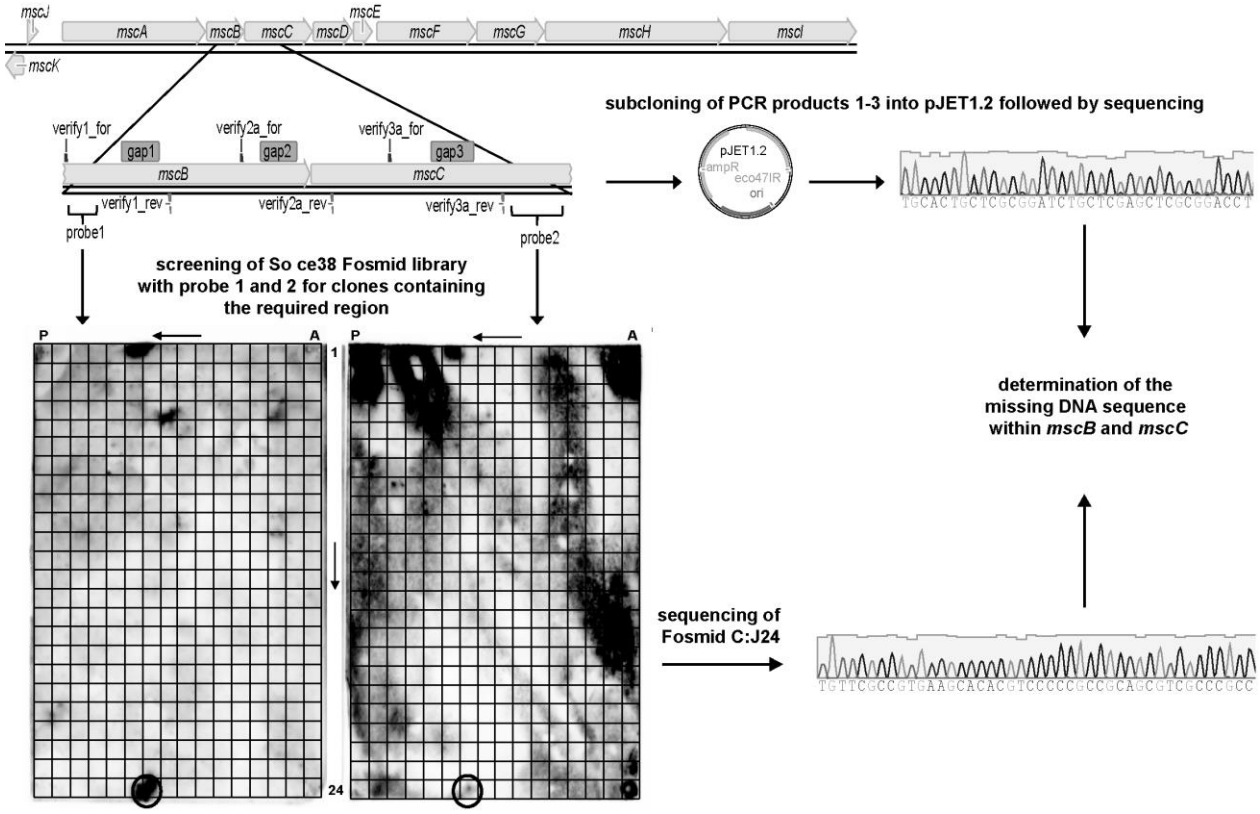


Figure 5: Schematic description of gap closure within the *msc* gene cluster from *S. cellulosum* So ce38 employing conventional cloning and screening of a Fosmid library. The positive hit in the fosmid screening is marked with a circle.

Based on the sequence of the *msc* gene cluster from *S. cellulosum* So ce38, its counterpart in the draft genome of *Jahnella* sp. MSr9139 could be identified. Surprisingly, although being sequenced as well using the Illumina technology, the latter sequence did not contain any missing parts. After having the genetic information of both gene clusters in hand, a comprehensive bioinformatic analysis was performed, which resulted in a hypothesis explaining the differences between the derivatives produced by the respective strains (Figure 6).

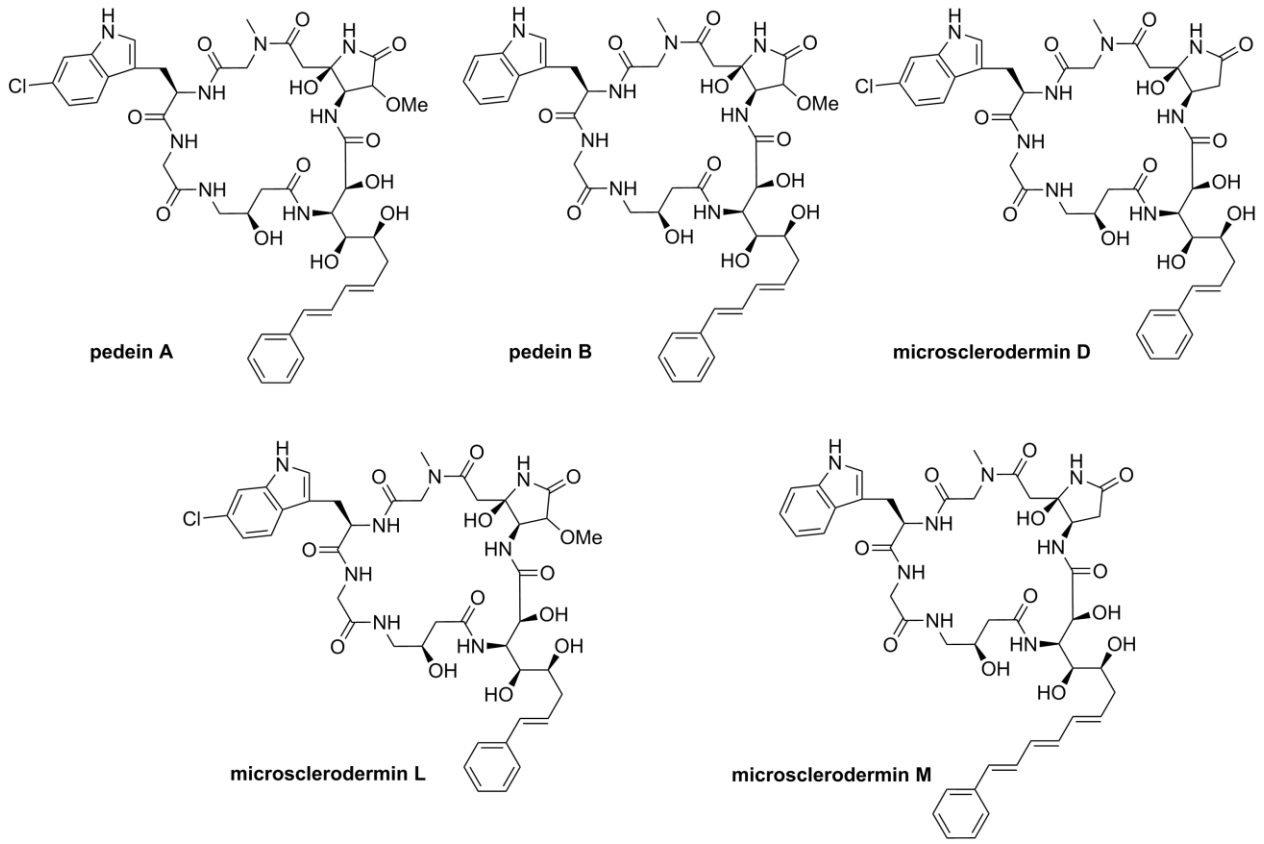


Figure 6: Overview of pedeins and microsclerodermins produced by terrestrial myxobacteria. Pedein A and B as well as microsclerodermin D and L are synthesized by *Jahnella* sp. MSr9139 while, *S. cellulosum* So ce38 exclusively produces microsclerodermin M.

According to this, chlorination of the tryptophan moiety as found in most of the derivatives produced by MSr9139 (pedein A, microsclerodermin D and L), but not by So ce38 (microsclerodermin M), is catalyzed by a dedicated halogenase (MscL) in a tailoring reaction. Furthermore, the methoxy functionality at the pyrrolidone ring, again only present in some derivatives isolated from MSr9139 (pedein A and B, microsclerodermin L), can be linked to the sequential action of MscM (Fe(II)/α-ketoglutarate-dependent dioxygenase) and MscN (SAM-dependent *O*-methyl transferase). As both, the modified and the unmodified pyrrolidone moiety occur, it is assumed that these steps take place after release of the compound from the assembly line. It is worth mentioning in this context, that only the methoxylated, but not the hydroxylated version of the pyrrolidone ring is found. However, even if MscN functions very efficiently, at least traces of the molecule which is only hydroxylated should exist, presumably in amounts below the detection limit for the HPLC-MS setup used in this study. Another interesting aspect of microsclerodermin biosynthesis is the length and assembly of the alkyl side chain. While the derivatives produced by MSr9139 feature side chains that result from incorporation of either one (microsclerodermin D and L) or two (pedein A and B) malonyl-CoA extender units, the

side chain of microsclerodermin M from So ce38 is composed of three such entities. These variations were attributed to differential usage of the iterative type I PKS subunits of MscA. What actually causes these distinctions is not known. As discussed earlier for the crocacins, bioinformatic tools to reliably predict the actual number of iterations are highly desirable, but do not exist to date. The present case can serve as another example to shed light onto this issue. It is largely accepted, that the nature of KS domains from iterative PKS modules influences the number of repetitions. Since a primary sequence alignment of $KS_{MscA\_So\ ce38}$ and $KS_{MscA\_MSr9139}$ shows only very slight differences (83.5 % identity, 91.2 % positives), a comparative structural characterization of both domains might pinpoint crucial factors which determine iteration count. Additionally, the heterologous production of the full length modules encoded by each *mscA* might be taken into account, in order to investigate the nature of this iteration *in vitro*.

Finally, the microsclerodermin study highlights how valuable a broad and well characterized strain collection is. It led to the identification of 13 alternative microsclerodermin producers and furthermore permitted the establishment of a correlation between phylogeny and derivative production. Although in this certain case, we did not depend on any alternative producers to purify and characterize the compounds, other investigations, as exemplified by the myxoprincomide, project strictly rely on them [63]. In the course of this study, a novel secondary metabolite from *Myxococcus xanthus* DK1622 was identified by applying the method of principal-component analysis to preprocessed LC-MS datasets of the wild type and specific mutant strains. As myxoprincomide was present in rather low amounts in the crude extract of *M. xanthus* DK1622, which were not sufficient for isolation and structure elucidation, a collection of 98 wild type *M. xanthus* isolates [64] was screened for another strain with better production properties. This examination revealed *M. xanthus* A2 to produce 15 times more of the compound than the initially identified producer *M. xanthus* DK1622. Upon genetic engineering of *M. xanthus* A2, the yield could be doubled, thereby providing suitable amounts for isolation and structure elucidation of myxoprincomide.

Studies towards enhanced production of ambruticin and derivatives thereof

Ambruticin and its derivatives were initially isolated in 1977 from an ethyl acetate extract of *Polyangium cellulosum* var. *firlvum* ATCC 25532 [65]. They exhibit their potent fungicidal activity by interfering with the osmoadaptation circuit (HOG-1 pathway) of the target organisms [66;67]. In particular, it is believed that ambruticin binds to a specific group III hybrid sensor kinase response regulator, thereby perturbing signal transduction [68]. In terms of bioactivity the VS3 derivative of ambruticin displays even better properties in a *Saccharomyces cerevisiae* sensitivity model than commercially available fungicides (fludioxonil and vinclozolin) [69] and

since no serious adverse effects in animals have been reported [70], ambruticin might be a prospective candidate for the development of a commercial fungicide. To supply sufficient amounts for an evaluation of the compounds potential and beyond, the availability of a strain with suitable production properties is of great value. Accordingly, the overall objective of this project was the identification of the best accessible myxobacterial ambruticin producer and its improvement by means of rational mutagenesis.

An *in silico* survey of the secondary metabolite profiles of 757 different myxobacterial strains from our collection identified nine ambruticin producers. Interestingly, all of them belong to the genus *Sorangium*, what accentuates the general notion that the secondary metabolite spectrum of different myxobacterial genera usually exhibit little overlap [71]. Out of these nine organisms, *Sorangium cellulosum* strain #8405 displays the best production rates for the VS ambruticins. Hence, this strain was chosen for subsequent yield improvement. In order to alter the chromosome of this strain, two mutagenic strategies, based on circular (pop-in/pop-out method) and linear DNA (gene replacement), were applied. Since these tools were already successfully used in other *S. cellulosum* species (strain #8404 and strain #1566), it is assumed that they are functional in a wide range of phylogenetic related organisms.

The ability to construct site directed mutants of strain #8405 allowed us to study elements potentially influencing ambruticin biosynthesis. In particular, two different approaches were employed. On the one hand loci encoding possible regulatory functions located in the direct vicinity of the gene cluster were examined, and on the other hand we sought to determine proteins binding to the presumed promotor region upstream of the *amb* operon that may direct transcription. Recapitulatory, four distinct genes (*ntcA*, *amb01*, *amb04*, and *amb07*) were studied, of which three (*ntcA*, *amb04*, and *amb07*), are definitely linked to ambruticin biosynthesis.

The first gene analyzed encodes an NtcA-like regulatory protein that was found to selectively bind to a region upstream of *ambA*. The role of a related protein from *S. cellulosum* So ce56 has already been investigated [72]. Specifically, this study revealed that high ammonia concentrations (10 mM) are causing a significant drop in chivosazol production in the wild type (5.5 fold less), whereas this inhibitory influence was significantly lower in the *ntcA*$^{-}$ mutants (13.5 fold more chivosazol with respect to the wild type). Accordingly, these data suggest that the repressive effect of ammonia on chivosazol biosynthesis is directly mediated via the transcriptional regulator NtcA. As $NtcA_{\#8405}$ is strikingly similar to $NtcA_{Soce56}$, an analogous function in strain #8405 is likely. In line with this hypothesis, the comparative analysis of the Δ*ntcA* strains and the wild type showed, at least in one case, elevated titers of ambruticin inside the mutants (Chapter III). However, since these cultivations were carried out in a very complex

medium (A medium), the actual impact of ammonia on the formation of ambruticin remains unclear to date. Thus, the production properties of the Δ*ntcA* deletion strains need to be evaluated again by using an experimental setup similar to that utilized previously for cultivation of *S. cellulosum* So ce56 wild type and the *ntcA*⁻ mutants (cultivation in SG medium with and without the addition of ammonium sulfate). Nevertheless, it is well known that the excess of certain nutrients including, ammonium salts and glucose, can suppress the biosynthesis of secondary metabolites in a variety of organisms [73]. Accordingly, transcriptional regulation via NtcA might be a general mechanism in the genus *Sorangium* to govern the biosynthesis of some bioactive metabolites. To support this theory, the genomes of 14 additional *S. cellulosum* strains from our collection were screened for such an NtcA-encoding gene. Interestingly, a similar gene occurs in every genome in exactly the same genetic context (Figure 7).

Figure 7: Genetic organization of the *ntcA* locus as found in *S. cellulosum* strain #8405. A similar arrangement is found as well in the genomes of in the *S. cellulosum* strains So ce56, #1424, #1433, #1566, #1760, #1795, #1941, #5797, #6597, #7237, #7282, #8242, #8404, #8412, #9515. *hypo1*: putative myrcene/ocimene synthase, *ntcA*: NtcA transcriptional regulator, *hypo2*: putative protein showing similarity to the TIGR04562 superfamily, *hypo3*: putative hypothetical protein.

Furthermore, the amino acid sequences of the NtcA homologues from the respective *S. cellulosum* strains were compared to that of $NtcA_{Soce56}$ in a MAFFT alignment (Figure 8).

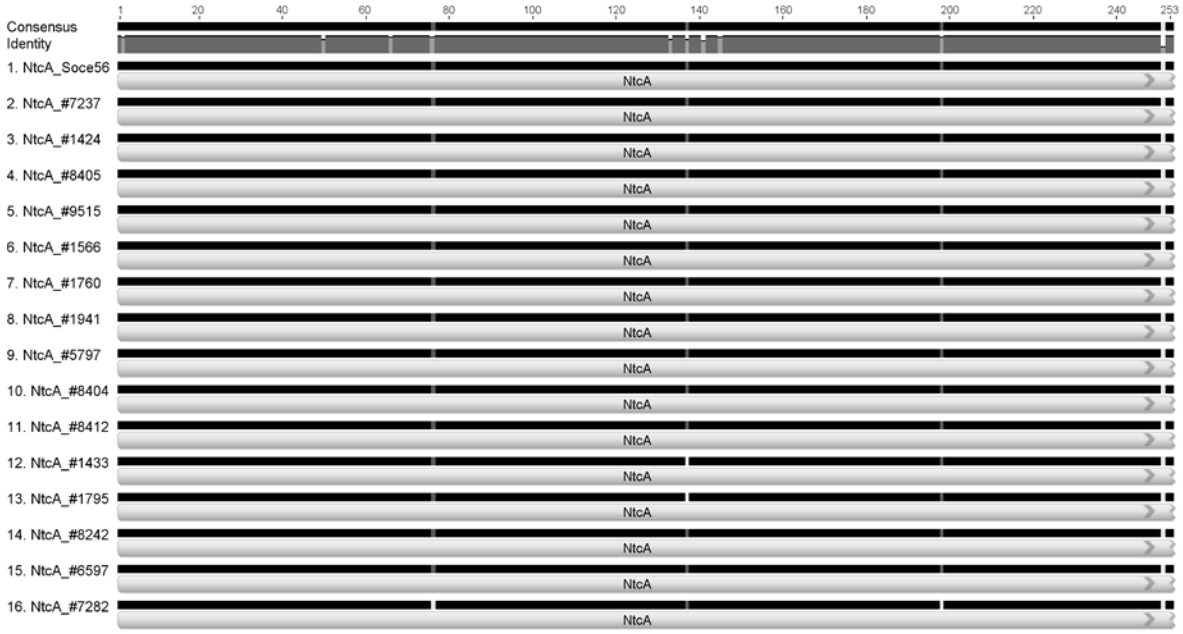


Figure 8: MAFFT alignment of $NtcA_{Soce56}$ and its homologues encoded in the genomes of the indicated *S. cellulosum* strains.

This evaluation revealed extraordinary similarity among all of the NtcA proteins from different *S. cellulosum* strains investigated. Hence, the regulation of secondary metabolite production, as a function of ammonia availability, might in fact be controlled by NtcA in the genus *Sorangium*. Still, against the background that representatives of this genus usually encode a plurality of transcriptional regulators (So ce56 σ^{54}-dependent regulators: 80, So0157-2: 188) [74;75] it is expected that numerous other proteins are directing natural product formation as well.

In conclusion, the discovery that an NtcA-type regulator, proven to influence secondary metabolite biosynthesis, is ubiquitous found in all strains analyzed, suggests that its manipulation may be used as a general instrument to affect natural product formation in the genus *Sorangium*.

The deletion of *amb07* (putative hybrid sensor kinase response regulator) caused a clear phenotype (10 fold drop in ambruticin productivity), proving the gene's involvement in regulation of ambruticin biosynthesis. As outlined in Chapter III, *amb07* is located downstream of every ambruticin/jerangolid gene cluster which has been examined to date. Moreover, Amb07 resembles the presumed target of ambruticin in filamentous fungi, suggesting that it might bind this molecule. Nevertheless, biochemical evidence (e.g. surface plasmon resonance binding studies) to corroborate this hypothesis is missing so far. Accordingly, the full length protein or at least its N-terminal part comprising the HAMP domains should be heterologously produced and subjected to such an analysis. If a specific interaction of ambruticin and the protein can be demonstrated, Amb07 might facilitate the development of a biosensor system that responds in a product dependent fashion. However, as Amb07 will only be able to transmit a signal upon interaction with the effector molecule, a dedicated transcriptional regulator of the underlying signal transduction cascade needs to be identified and characterized. Knowledge gained herein should subsequently allow to selectively control the expression of a reporter gene. If that regulator would stimulate gene expression, a positive reporter such as an antibiotic resistance marker or a fluorescent protein could be chosen. Accordingly, under ideal circumstances the extent of antibiotic-resistance or fluorescence intensity will directly correlate to the amount of ambruticin present inside the cell. Conversely, if the regulator would repress transcription, the use of a negative selection marker should be taken into account. Unfortunately, as shown in our study, the *sacB* gene could not be employed to efficiently inhibit growth of *S. cellulosum* strain #8405 on medium containing sucrose. To circumvent this issue, again, a fluorescent protein could be used, with the difference, that screening should now be directed towards detection of a diminished signal. In both cases, the identification of genetically altered clones with improved production properties can be carried out by screening for, or against the respective reporter activity. Several comparable systems have already been successfully applied, for example, to

obtain biocatalysts from genomic libraries [76;77], for the detection and isolation of proficient producers of amino acids [78;79], and for other purposes as summarized in a recent review by Eggeling *et al.* [80]. However, there is one striking difference between these processes and a putative application of Amb07. While Amb07 is initiating a signal transduction cascade which eventually causes a response; the other biosensors, mostly transcription factors, directly control expression of the reporter upon interaction with the effector. This instance represents a drawback, because the signaling pathway may be influenced as well by other stimuli. Thus, a substantive correlation between ambruticin formation and the activity of the reporter may not be given. Still, as Amb07 is currently the only possible candidate for the development of an ambruticin recognizing biosensor, further studies on this subject should be undertaken.

Based on the current knowledge about Amb07, different working hypothesis are possible explaining its function. One option would be that Amb07 controls the amount of ambruticin synthesized through product recognition and consequent downstream signaling. In particular, if sufficient amounts of ambruticin are present, Amb07 may promote repression of further production efforts, and contrary if inadequate levels are detected, the pathway is upregulated. But as the knock out mutant displays significantly lower yields in comparison to the wild type, this theory may be mistaken. Here, one would assume that the lapse of such a regulatory element would rather result in enhanced ambruticin titers than in a decrement. Alternatively, Amb07 could confer resistance to ambruticin by initiating countermeasures. In this context the phenotype of the mutant matches the expectations (loss of resistance mechanism results in a diminished productivity). In any case, additional experiments are needed in order to reveal the physiological function of Amb07 in *S. cellulosum* strain #8405. Unfortunately, mutants overexpressing *amb07* could not be generated and phenotypically characterized. This might be owing to the lack of an inducible gene expression system for the genus *Sorangium*, as outlined in detail in Chapter III. Nevertheless, the outcome of the experiments conducted so far unambiguously linked Amb07 to ambruticin biosynthesis and furthermore highlighted interesting features of the protein itself, all of potential biotechnological importance.

Among all genes examined within this project, *amb04* encoding a putative AtoC-like response regulatory protein has been studied in most detail. Since the targeted deletion of this gene almost completely abolished ambruticin production, we concluded that the corresponding protein is an activator of ambruticin biosynthesis. According to this observation, we attempted to genetically complement the Δ*amb04* phenotype by introducing a P_{T7A1} driven copy of *amb04* into the genome of the respective strain. Unfortunately, production of the ambruticins could not be restored, even though the new copy was shown to be highly transcribed (up to 400 times more than in the wild type). In fact, these tremendously elevated transcription levels might also be the

reason for failure of complementation. In line with this, the directed P_{T7A1} overexpression of the putative *amb03-04* operon in the wild type led to a decreased production as well. A comprehensive EMSA analysis together with an extensive *in silico* evaluation suggests that Amb04 is regulating the expression of at least 10 genes/operons, notably including another putative transcriptional regulator (ArsR-like protein). This investigation suggests that the effect of Amb04 on compound formation is an indirect one, because the heterologously produced protein is not binding to any DNA fragments containing presumed promotor regions of the *amb* genes. After all, either kind of intervention (overexpression or deletion of *amb04*) might affect the integrity of the complex regulatory network within the cell in a way that eventually the production of ambruticin is perturbed. Hence, the regulatory element directly influencing ambruticin biosynthesis has to be identified. One of the most promising candidates seems to be the *amb04a* gene, encoded immediately downstream of *amb04* and being regulated as well by Amb04 (see Chapter III). In order to evaluate if the putative *amb03-04* operon and the *amb04a* gene are invariably linked to the ambruticin and jerangolid gene clusters, available draft-genome datasets of the other producers *S. cellulosum* strain #7282 (ambruticin), *S. cellulosum* strain #1795 (jerangolid), and *S. cellulosum* strain #1433 (jerangolid) were analyzed accordingly. As an outcome, only *S. cellulosum* strain #7282 does not feature a similar gene arrangement as found in *S. cellulosum* strain #8405. This finding supports our assumption that Amb04a is in fact involved in the regulation of ambruticin and jerangolid biosynthesis, but actual experimental data is needed to corroborate this. Noteworthy, homologues of the *amb03-04* operon encoded together with the *amb04a* gene are found in the draft-genomes of 7 other *S. cellulosum* strains (So ce56, #1566, #1433, #1760, #1795, #6597, #7237, and #9515), however, as it can be concluded from the existing data, not in the direct context of a secondary metabolite gene cluster. This may argue for a more general mechanism, but since in the draft-genomes of other *S. cellulosum* strains (#1941, #5797, #8242, and #8404) such a gene-association is not found at all, this mechanism may not be as prevalent as the nitrogen-dependent regulation via NtcA.

As described above, the identification and subsequent manipulation of regulatory elements is laborious and does not necessarily lead to a reasonable production enhancement on a short timescale. Therefore, other possibilities like deletion of competing secondary metabolite pathways and increased precursor supply should be taken into account as well. This might also be in combination with manipulation of *ntcA*, *amb04* and *amb07*. The first approach may not be that effective in the present case, because strain #8405 is almost exclusively producing ambruticin (small amounts of ajudazol and thuggacin are produced as well). Secondly, the introduction of an additional set of genes encoding for acetyl-CoA carboxylase (ACC) and propionyl-CoA carboxylase (PCC), respectively will most likely help to increase the intracellular

pool of malonyl-CoA (used 5 times during biosynthesis) and methylmalonyl-CoA (used 3 times during biosynthesis), what consequently should enable higher product yields. In particular, both approaches can be combined in a way that the genetic locus of the competing cluster is used as integration site for the introduction of linear pSM_lacZ_hyg and pSM_gfp_tet derivatives each harboring genes encoding for ACC or PCC. This is a promising strategy to generate an ambruticin overproducing strain with a stable genetic background.

References

[1] Jansen R, Washausen P, Kunze B, Reichenbach H, & Höfle G (1999) Antibiotics from gliding bacteria, LXXXIII - The crocacins, novel antifungal and cytotoxic antibiotics from *Chondromyces crocatus* and *Chondromyces pediculatus* (Myxobacteria): Isolation and structure elucidation. *Eur. J. Org. Chem.*, 5, 1085-1089.

[2] Kunze B, Jansen R, Höfle G, & Reichenbach H (1994) Crocacin, a new electron transport inhibitor from *Chondromyces crocatus* (myxobacteria). Production, isolation, physico-chemical and biological properties. *J. Antibiot.*, 47, 881-886.

[3] Jansen R, Kunze B, Reichenbach H, & Höfle G (2000) Antibiotics from gliding bacteria LXXXVI, Apicularen A and B, cytotoxic 10-membered lactones with a novel mechanism of action from *Chondromyces* species (myxobacteria): Isolation, structure elucidation, and biosynthethis. *Eur. J. Org. Chem.*, 6, 913-919.

[4] McKee TC, Galinis DL, Pannell LK, Cardellina JH, Laakso J, Ireland CM, Murray L, Capon RJ, & Boyd MR (1998) The lobatamides, novel cytotoxic macrolides from southwestern pacific tunicates. *J. Org. Chem.*, 63, 7805-7810.

[5] Erickson KL, Beutler JA, Cardellina JH, & Boyd MR (1997) Salicylihalamides A and B, novel cytotoxic macrolides from the marine sponge *Haliclona* sp. *J. Org. Chem.*, 62, 8188-8192.

[6] Wu YS, Seguil OR, & De Brabander JK (2000) Synthesis and initial structure-activity relationships of modified salicylihalamides. *Org. Lett.*, 2, 4241-4244.

[7] Thierbach G & Reichenbach H (1981) Myxothiazol, a new inhibitor of the cytochrome bc_1 segment of th respiratory chain. *Biochim. Biophys. Acta*, 638, 282-289.

[8] Thierbach G, Kunze B, Reichenbach H, & Höfle G (1984) The mode of action of stigmatellin, a new inhibitor of the cytochorme bc_1 segment of the respiratory chain. *Biochim. Biophys. Acta*, 765, 227-235.

[9] Chakraborty TK & Laxman P (2003) Total synthesis of (+)-crocacin A. *Tetrahedron Lett.*, 44, 4989-4992.

[10] Feutrill JT, Lilly MJ, & Rizzacasa MA (2000) Total synthesis of (+)-crocacin C. *Org. Lett.*, 2, 3365-3367.

[11] Feutrill JT, Lilly MJ, & Rizzacasa MA (2002) Total synthesis of (+)-crocacin D. *Org. Lett.*, 4, 525-527.

[12] Kopp M, Irschik H, Gross F, Perlova O, Sandmann A, Gerth K, & Müller R (2004) 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.*, 107, 29-40.

[13] Rachid S, Krug D, Kochems I, Kunze B, Scharfe M, Blöcker H, Zabriski M, & Müller R (2006) Molecular and biochemical studies of chondramide formation - highly cytotoxic natural products from *Chondromyces crocatus* Cm c5. *Chem. Biol.*, 14, 667-681.

[14] Rachid S, Scharfe M, Blöcker H, Weissman KJ, & Müller R (2009) Unusual chemistry in the biosynthesis of the antibiotic chondrochlorens. *Chem. Biol.*, 16, 70-81.

[15] Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, & Weber T (2013) antiSMASH 2.0 - A versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.*, 41, W204-W212.

[16] Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, & Breitling R (2011) antiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.*, 39, W339-W346.

[17] Joyce SA, Brachmann AO, Glazer I, Lango L, Schwär G, Clarke DJ, & Bode HB (2008) Bacterial biosynthesis of a multipotent stilbene. *Angew. Chem. Int. Ed.*, 47, 1942-1945.

[18] Piel J, Hertweck C, Shipley PR, Hunt DM, Newman MS, & Moore BS (2000) Cloning, sequencing and analysis of the enterocin biosynthesis gene cluster from the marine isolate *Streptomyces maritimus*: Evidence for the derailment of an aromatic polyketide synthase. *Chem. Biol.*, 7, 943-955.

[19] Bisang C, Long PF, Cortés J, Westcott J, Crosby J, Matharu AL, Cox RJ, Simpson TJ, Staunton J, & Leadlay PF (1999) A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature*, 401, 502-505.

[20] Yadav G, Gokhale RS, & Mohanty D (2003) Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol.*, 328, 335-363.

[21] Keatinge-Clay AT, Shelat AA, Savage DF, Tsai SC, Miercke LJW, O'Connell JD, Khosla C, & Stroud RM (2003) Catalysis, specificity, and ACP docking site of *Streptomyces coelicolor* malonyl-CoA:ACP transacylase. *Structure*, 11, 147-154.

[22] Challis GL, Ravel J, & Townsend CA (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.*, 7, 211-224.

[23] Stachelhaus T, Mootz HD, & Marahiel MA (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.*, 6, 493-505.

[24] Serre L, Verbree EC, Dauter Z, Stuitje AR, & Derewenda ZS (1995) The *Escherichia coli* malonyl-coa:Acyl carrier protein transacylase at 1.5-angstroms resolution: Crystal structure of a fatty acid synthase component. *J. Biol. Chem.*, 270, 12961.

[25] Yadav G, Gokhale RS, & Mohanty D (2009) Towards prediction of metabolic products of polyketide synthases: An *in silico* analysis. *PLoS Comput. Biol*, 5, e1000351.

[26] Ziemert N, Podell S, Penn K, Badger JH, Allen E, & Jensen PR (2012) The Natural Product Domain Seeker NaPDoS: A Phylogeny Based Bioinformatic Tool to Classify Secondary Metabolite Gene Diversity. *PLoS ONE*, 7, e34064.

[27] El-Sayed AK, Hothersall J, Cooper SM, Stephens E, Simpson TJ, & Thomas CM (2003) Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586. *Chem. Biol.*, 10, 419-430.

[28] Butcher RA, Schroeder FC, Fischbach MA, Straight PD, Kolter R, Walsh CT, & Clardy J (2007) The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA*, 104, 1506-1509.

[29] Moldenhauer J, Chen XH, Borriss R, & Piel J (2007) Biosynthesis of the antibiotic bacillaene, the product of a giant polyketide synthase complex of the trans-AT family. *Angew. Chem. Int. Ed.*, 46, 8195-8197.

[30] Calderone CT (2008) Isoprenoid-like alkylations in polyketide biosynthesis. *Nat. Prod. Rep.*, 25, 845-853.

[31] Calderone CT, Kowtoniuk WE, Kelleher NL, Walsh CT, & Dorrestein PC (2006) Convergence of isoprene and polyketide biosynthetic machinery: Isoprenyl-S-carrier proteins in the *pksX* pathway of *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA*, 103, 8977-8982.

[32] Buchholz TJ, Rath CM, Lopanik NB, Gardner NP, Hakansson K, & Sherman D (2010) Polyketide β-branching in bryostatin biosynthesis: Identification of surrogate acetyl-ACP donors for BryR, an HMG-ACP synthase. *Chem. Biol.*, 17, 1092-1100.

[33] Sudek S, Lopanik NB, Waggoner LE, Hildebrand M, Anderson C, Liu HB, Patel A, Sherman DH, & Haygood MG (2007) Identification of the putative bryostatin polyketide synthase gene cluster from "*Candidatus* endobugula sertula", the uncultivated microbial symbiont of the marine bryozoan *Bugula neritina*. *J. Nat. Prod.*, 70, 67-74.

[34] Simunovic V, Zapp J, Rachid S, Krug D, Meiser P, & Müller R (2006) Myxovirescin biosynthesis is directed by hybrid polyketide synthases/ nonribosomal peptide synthetase, 3-hydroxy-3-methylglutaryl CoA synthases and trans-acting acyltransferases. *ChemBioChem*, 7, 1206-1220.

[35] Simunovic V & Müller R (2007) Mutational analysis of the myxovirescin biosynthetic gene cluster reveals novel insights into the functional elaboration of polyketide backbones. *ChemBioChem*, 8, 1273-1280.

[36] Simunovic V & Müller R (2007) 3-Hydroxy-3-methylglutaryl-CoA-like synthases direct the formation of methyl and ethyl side groups in the biosynthesis of the antibiotic myxovirescin A. *ChemBioChem*, 8, 497-500.

[37] Calderone CT, Iwig DF, Dorrestein PC, Kelleher NL, & Walsh CT (2007) Incorporation of nonmethyl branches by isoprenoid-like logic: Multiple beta-alkylation events in the biosynthesis of myxovirescin A1. *Chem. Biol.*, 14, 835-846.

[38] Stachelhaus T, Mootz HD, Bergendahl V, & Marahiel MA (1998) Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. *J. Biol. Chem.*, 273, 22773-22781.

[39] Roche ED & Walsh CT (2003) Dissection of the EntF condensation domain boundary and active site residues in nonribosomal peptide synthesis. *Biochemistry*, 42, 1334-1344.

[40] Keating TA, Marshall CG, & Walsh CT (2000) Vibriobactin biosynthesis in *Vibrio cholerae*: VibH is an amide synthase homologous to nonribosomal peptide synthetase condensation domains. *Biochemistry*, 39, 15513-15521.

[41] Keating TA, Marshall CG, Walsh CT, & Keating AE (2002) The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.*, 9, 522-526.

[42] Lessard IA, Healy VL, Park IS, & Walsh CT (1999) Determinants for differential effects on D-Ala-D-lactate vs D-Ala-D-Ala formation by the VanA ligase from vancomycin-resistant Enterococci. *Biochemistry*, 38, 14006-14022.

[43] Schmidt EW & Faulkner DJ (1998) Microsclerodermins C-E, antifungal cyclic peptides from the lithistid marine sponges *Theonella* sp. and *Microscleroderma* sp.. *Tetrahedron*, 54, 3043-3056.

[44] Bewley CA, Debitus C, & Faulkner DJ (1994) Microsclerodermins A and B. Antifungal cyclic peptides from the lithistid sponge *Microscleroderma* sp. *J. Am. Chem. Soc.*, 116, 7631-7636.

[45] Qureshi AA, Colin PL, & Faulkner DJ (2000) Microsclerodermins F-I, antitumor and antifungal cyclic peptides from the lithistid sponge *Microscleroderma* sp.. *Tetrahedron*, 56, 3679-3685.

[46] Zhang X, Jacob MR, Ranga Rao R, Wang YH, Agarwal AK, Newman DJ, Khan IA, Clark AM, & Li XC (2012) Antifungal cyclic peptides from the marine sponge *Microscleroderma herdmani*. *Res. Rep. Med. Chem.*, 2, 7-14.

[47] Kunze B, Bohlendorf B, Reichenbach H, & Höfle G (2008) Pedein A and B: Production, isolation, structure elucidation and biological properties of new antifungal cyclopeptides from *Chondromyces pediculatus* (Myxobacteria). *J. Antibiot.*, 61, 18-26.

[48] Simister RL, Deines P, Botté ES, Webster NS, & Taylor MW (2012) Sponge-specific clusters revisited: A comprehensive phylogeny of sponge-associated microorganisms. *Environ. Microbiol.*, 14, 517-524.

[49] Unson MD, Holland ND, & Faulkner DJ (1994) A brominated secondary metabolite synthesized by the cyanobacterial symbiont of a marine sponge and accumulation of the crystalline metabolite in the sponge tissue. *Mar. Biol.*, 119, 1-11.

[50] Garson MJ, Thompson JE, Larsen RM, Battershill CN, Murphy PT, & Bergquist PR (1992) Terpenes in sponge cell membranes cell separation and membrane fractionation studies with the tropical marine sponge *Amphimedon* sp. *Lipids*, 27, 378-388.

[51] Flowers AE, Garson MJ, Webb RI, Dumdei EJ, & Charan RD (1998) Cellular origin of chlorinated diketopiperazines in the dictyoceratid sponge *Dysidea herbacea* (Keller). *Cell Tissue Res.*, 292, 597-607.

[52] Schmidt EW, Bewley CA, & Faulkner DJ (1998) Theopalauamide, a bicyclic glycopeptide from filamentous bacterial symbionts of the lithistid sponge *Theonella swinhoei* from Palau and Mozambique. *J. Org. Chem.*, 63, 1254-1258.

[53] Bewley CA, Holland ND, & Faulkner DJ (1996) Two classes of metabolites from *Theonella swinhoei* are localized in distinct polulations of bacterial symbionts. *Experientia*, 52, 716-722.

[54] Unson MD & Faulkner DJ (1993) Cyanobacterial symbiont biosynthesis of chlorinated metabolites from *Dysidea herbacea* (Porifera). *Experientia*, 49, 349-353.

[55] Paul VJ, Arthur KE, Ritson-Williams R, Ross C, & Sharp K (2007) Chemical defenses: From compounds to communities. *Biol. Bull.*, 213, 226-251.

[56] Mehbub MF, Lei J, Franco C, & Zhang W (2014) Marine sponge derived natural products between 2001 and 2010: Trends and opportunities for discovery of bioactives. *Mar. Drugs*, 12, 4539-4577.

[57] Crews P, Gerwick WH, Schmitz FJ, France D, Bair KW, Wright AE, & Hallock Y (2003) Molecular approaches to discover marine natural product anticancer leads - An update from a drug discovery group collaboration. *Pharm. Biol.*, 41, 39-52.

[58] Johnson TA, Sohn J, Vaske YM, White KN, Cohen TL, Vervoort HC, Tenney K, Valeriote FA, Bjeldanes LF, & Crews P (2012) Myxobacteria versus sponge-derived alkaloids: The bengamide family identified as potent immune modulating agents by scrutiny of LC-MS/ELSD libraries. *Bioorg. Med. Chem*, 20, 4348-4355.

[59] Guzman EA, Maers K, Roberts J, Kemami-Wangun HV, Harmody D, & Wright AE (2015) The marine natural product microsclerodermin A is a novel inhibitor of the nuclear factor kappa B and induces apoptosis in pancreatic cancer cells. *Invest. New Drugs*., 33, 86-94.

[60] Taylor MW, Radax R, Steger D, & Wagner M (2007) Sponge-associated microorganisms: Evolution, ecology, and biotechnological potential. *Microbiol. Mol. Biol. Rev.*, 71, 295-347.

[61] Hentschel U, Piel J, Degnan SM, & Taylor MW (2012) Genomic insights into the marine sponge microbiome. *Nat. Rev. Microbiol.*, 10, 641-654.

[62] Wilkinson CR, Garrone R, & Vacelet J (1984) Marine sponges discriminate between food bacteria and bacterial symbionts - electron microscope autoradiography and insitu evidence. *Proc. Roy. Soc. London B*, 220, 519-528.

[63] Cortina NS, Krug D, Plaza A, Revermann O, & Müller R (2012) Myxoprincomide: A natural product from *Myxococcus xanthus* discovered by comprehensive analysis of the secondary metabolome. *Angew. Chem. Int. Ed Engl.*, 51, 811-816.
[64] Krug D, Zurek G, Revermann O, Vos M, Velicer GJ, & Müller R (2008) Discovering the hidden secondary metabolome of *Myxococcus xanthus*: A study of intraspecific diversity. *Appl. Environ. Microbiol.*, 74, 3058-3068.
[65] Ringel SM, Greenough RC, Romm S, Conor D, Gutt AL, Blair B, Kanter G, & von Strandtmann M (1977) Ambruticin (W7783), a new antifungal antibiotic. *J. Antibiot.*, 30, 371-375.
[66] Knauth P & Reichenbach H (2000) On the mechanism of action of the myxobacterial fungicide ambruticin. *J. Antibiot.*, 53, 1182-1190.
[67] Hohmann S (2002) Osmotic stress signaling and osmoadaptation in yeasts. *Microbiol. Mol. Biol. Rev.*, 66, 300-372.
[68] Buschart A, Gremmer K, El-Mowafy M, van den Heuvel J, Mueller PP, & Bilitewski U (2012) A novel functional assay for fungal histidine kinases group III reveals the role of HAMP domains for fungicide sensitivity. *J. Biotechnol.*, 157, 268-277.
[69] Vetcher L, Menzella HG, Kudo T, Motoyama T, & Katz L (2007) The antifungal polyketide ambruticin targets the HOG pathway. *Antimicrob. Agents Chemother.*, 51, 3734-3736.
[70] Ringel SM (1990) New antifungal agents for the systemic mycoses. *Mycopathologia*, 109, 75-87.
[71] Müller R & Wink J (2014) Future potential for anti-infectives from bacteria - How to exploit biodiversity and genomic potential. *Int. J. Med. Microbiol.*, 304, 3-13.
[72] Rachid S, Gerth K, & Müller R (2008) NtcA-A negative regulator of secondary metabolite biosynthesis in *Sorangium cellulosum*. *J. Biotechnol.*, 140, 135-142.
[73] Demain AL (2006) From natural products discovery to commercialization: a success story. *J. Ind. Microbiol. Biotechnol.*, 33, 486-495.
[74] 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, Munoz-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 FJ, Weissman KJ, Welch RD, Wenzel SC, Whitworth DE, Wilhelm S, Wittmann C, Blöcker H, Pühler A, & Müller R (2007) Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat. Biotechnol.*, 25, 1281-1289.
[75] Han K, Li ZF, Peng R, Zhu LP, Zhou T, Wang LG, Li SG, Zhang XB, Hu W, Wu ZH, Qin N, & Li YZ (2013) Extraordinary expansion of a *Sorangium cellulosum* genome from an alkaline milieu. *Sci. Rep.*, 3, 1-7.
[76] van Sint FS, van Beilen JB, & Witholt B (2006) Selection of biocatalysts for chemical synthesis. *Proc. Natl. Acad. Sci. USA*, 103, 1693-1698.
[77] Uchiyama T & Miyazaki K (2010) Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes. *Appl. Environ. Microbiol.*, 76, 7029-7035.
[78] Binder S, Schendzielorz G, Stabler N, Krumbach K, Hoffmann K, Bott M, & Eggeling L (2012) A high-throughput approach to identify genomic variants of bacterial metabolite producers at the single-cell level. *Genome Biol.*, 13, R40.
[79] Schendzielorz G, Dippong M, Grunberger A, Kohlheyer D, Yoshida A, Binder S, Nishiyama C, Nishiyama M, Bott M, & Eggeling L (2014) Taking control over control: Use of product sensing in single cells to remove flux control at key enzymes in biosynthesis pathways. *ACS Synth. Biol.*, 3, 21-29.
[80] Eggeling L, Bott M, & Marienhagen J (2015) Novel screening methods-biosensors. *Curr. Opin. Biotechnol.*, 35C, 30-36.

Author's effort in the work presented in this thesis

Chapter I

The author, together with Dr. Shwan Rachid, generated and characterized all mutant strains used in this study and elaborated the biosynthetic model. Furthermore, the author performed heterologous production and purification of CroK-C2 and MxcG. Feeding studies as well as determination of the enzyme kinetics were carried out in cooperation with Dr. Thomas Hoffmann, who also measured HPLC samples. Furthermore, the author did the complete *in silico* analysis of the gene cluster, which had been initially identified by Dr. Shwan Rachid. The *in silico* examination of pathways with no obvious release function was performed together with Dr. Nestor Zaburannyi. Eventually, the author wrote the draft of the manuscript, which in turn was edited by the co-authors.

Chapter II

The author performed mutagenesis experiments, closed the gaps in the DNA sequence of the *msc* gene cluster from *Sorangium cellulosum* So ce38 and contributed to the *in silico* analysis of both gene clusters. Moreover, the author edited the manuscript.

Chapter III

The author developed the genetic tools used in this study and performed all experiments except for identification of NtcA and some of the q-PCR analysis, which were carried out by Dr. Carsten Volz. Besides that, Dr. Thomas Hoffmann initially identified all ambruticin producers available from our strain collection and measured HPLC samples.

Danksagung

An erster Stelle möchte ich mich bei Prof. Dr. Rolf Müller bedanken, der mir die Arbeit an diesen spannenden Themen unter exzellenten Bedingungen ermöglicht hat und es durch seine hervorragende Betreuung immer verstand mich zu motivieren.

Prof. Dr. Rolf Hartmann danke ich sehr für die Übernahme des Koreferates sowie für die ergiebigen Diskussionen während der Thesis Committee Meetings.

Ein ganz besonderer Dank gilt Dr. Carsten Volz. Zum einen für seine stete sowie uneingeschränkte Hilfs- und Diskussionsbereitschaft in wissenschaftlichen Belangen und zum anderen für das freundschaftliche Miteinander im Labor aber auch darüber hinaus.

Des Weiteren sei auch allen Kooperationspartnern aus den verschiedenen Projekten für die angenehme und fruchtbare Zusammenarbeit gedankt. Insbesondere möchte ich hier Dr. Thomas Hoffmann hervorheben, von dem ich speziell zu Beginn der Promotion viel gelernt habe und der mir ferner auf dem Gebiet der Analytik eine große Hilfe war.

Allen Mitgliedern der Arbeitsgruppe danke ich hiermit für ihre Hilfsbereitschaft und Kollegialität. Gesondert möchte ich in diesem Zusammenhang Dr. Katja Gemperlein, Dr. Ronald Garcia, Dr. Stephan Hüttel, Dr. Jennifer Herrmann, Dr. Sascha Baumann und Angie Kling nennen. Vielen Dank! Schließlich sei noch Prof. Dr. Peter Gross erwähnt, der mich von Anfang an mit seiner heiteren Art erreichte.

Der größte Dank jedoch gilt meiner Familie und Dr. Thomas Weeber, für die jahrelange und ausnahmslose Unterstützung. Ohne euch wäre das alles nicht möglich gewesen!