Investigations on structure and biosynthesis of ribosomal peptide natural products involved in a tripartite mutualistic interaction

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Maria Guadalupe Soto Zarazua

aus

Guanajuato, México

Bonn, 2023

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Gabriele M. König

2. Gutachter: PD Dr. Christiane Dahl

Tag der Promotion: 12. Dezember 2023

Erscheinungsjahr: 2024

Abstract

Ribosomally synthesized and post-translationally modified peptides (RiPPs) constitute a chemically diverse group of biologically active molecules. Several RiPPs play outstanding roles in the social behavior and physiology of microorganisms. Their biosynthesis is a complex and energetically costly process that is strictly orchestrated in the bacterial cell. The timing and level of expression of bacterial RiPP biosynthetic gene clusters (BGCs) are typically controlled by multiple regulatory pathways that respond to environmental changes, microbial interactions, stress, or master switches during specific growth phases.

Streptomyces griseus S4-7 was isolated from the rhizosphere of strawberry plants as the representative strain producing uncharacterized metabolites with inhibitory effect against the phytopathogenic fungus *Fusarium oxysporum*. The S4-7 strain was further found to be engaged in a mutualistic interaction with plants and pollinator bees where bacteria provide protection from pathogens and the honeybees serve as vectors for effective dispersal.

The present thesis provides insights into the structures and biosynthesis of ribosomal peptide natural products implicated in the protective effect of *S. griseus* S4-7. These metabolites are encoded in two BGCs. The *cpr* BGC encodes an uncharacterized thiopeptide predicted to be the largest member of this subfamily of RiPPs ever reported. The second BGC, here termed mrs_{S47}, which stands for mixed RiPP system, encodes five precursor peptides varying in length and sequence. Since both cryptic BGCs were found to be silent under standard cultivation conditions in the laboratory, several strategies including elicitation and overexpression of pleiotropic and cluster-situated regulators were applied in the natural producer S4-7. As none of these efforts led to the production of the target metabolites, the *cpr* and mrs_{S47} BGCs were independently investigated.

The *cpr* BGC was cloned and refactored for heterologous expression in *Streptomyces*. Constitutive and inducible promoters as well as three different *Streptomyces* species were tested for production of the thiopeptide. An $[M+2H]^{2+}$ ion at *m/z* 1070.08 was ultimately obtained as a *cpr* BGC–related metabolite that still needs to be confirmed as the mature thiopeptide.

On the other hand, the mrs_{S47} BGC was studied through an *in vitro*/*in vivo* reconstitution approach. The five precursor peptides (MrsA1-A5) and the six biosynthetic enzymes (MrsM, MrsKC, MrsP, MrsJ, MrsY, and MrsR) were heterologously expressed in *E. coli*, purified, and subsequently used for reconstitution of the main biosynthetic steps. It was experimentally shown that the five peptides encoded in the mrs_{S47} BGC are independently modified and belong to different subclasses of RiPPs. MrsA1 and MrsA2 were confirmed to be modified by the lanthipeptide synthetase MrsM and the oxidoreductase MrsJ giving linear peptides containing dehydroamino acids and, in the case of MrsA2, D-alanine and D-amino butyric acid residues. The class III lanthipeptide synthetase MrsKC installed a labionin structure in MrsA3 and MrsA4. The S8 serine protease MrsP was proven to be responsible for the maturation of three of the five peptides (MrsA3-A5). Finally, MrsA5 was found to be modified by the radical SAM protein and the iron-containing redox enzyme (MrsY and MrsR, respectively) and is proposed to be the effector of a standalone peptide-processing biological conflict system. Further bioinformatics analyses revealed that the *mrs* BGC of *S. griseus* S4-7 is a member of a widespread family of mixed RiPP systems from Actinobacteria. These *mrs* BGCs share a highly conserved operon likely encoding a peptide-processing biological conflict system, and at least one additional modified peptide. The mrs_{S47} is thus the first member of this newly discovered family of mixed RiPP BGCs to be experimentally characterized.

Table of content

1. Introduction

1.1 Natural Products

Nature has evolved an extraordinary array of biosynthetic machineries that install diverse chemistries into natural products (NPs). Also known as secondary or specialized metabolites, NPs are produced by almost all types of living organisms such as plants, animals and microorganisms, where they mediate defense mechanisms and interactions with their ecosystem (Pham *et al*., 2019). Secondary metabolites are the result of multiple rounds of natural selection for biosynthetic pathways that produce molecules interacting with specific, biologically relevant targets. This advantage has historically situated NPs as the primary source of lead compounds for drug discovery and development (Newman & Cragg, 2020). For over seven decades, bacteria provided a vast number of chemical entities that were developed into drugs; as a prominent example, most of the antibiotics approved for medical use are products of microorganisms or were inspired from their metabolites (Dias *et al*., 2012). Bacterial natural products have also been applied in veterinary, agriculture and food industry (Abdel-Razek *et al.*, 2020). Classical microbial NP discovery workflows consisted of cultivation of bacterial strains and biological screening of crude extracts to identify bioactive hits that were further fractionated following a laborious bioactivity-guided isolation process (Wagenaar, 2008). However, efforts on finding new bioactives applying these classical approaches faced a problem of consistent re-discovery of already characterized molecules (Atanasov *et al*., 2021). In the early 2000s, the growing availability of genome sequences disclosed the potential of microorganisms to produce complex structures through previously unexplored biosynthetic pathways. Thus, advances in genomics, bioinformatics, metabolic engineering, and synthetic biology ushered in a new paradigm in natural product research (Kalkreuter *et al*., 2020).
Nowadays, the integration of sophisticated data mining, genetics, and analytical tools is empowering the discovery of natural products derived not only from unconventional sources, but also from organisms that were thought to have already been exhaustively explored. For example, predictions of biosynthetic gene clusters (BGCs) and genome mining of Actinobacterial genomes uncovered the immense existence of previously unrecognized biosynthetic potential of this phylum (Genilloud, 2018).

1.2 Actinobacteria: the inexhaustible source of natural products

Actinobacteria are well known as a non-exhaustive source of NPs. Most of the classic antibiotics used in the clinic today were originally isolated from actinomycetes. Between the 1950s and 1970s, the period known as the golden era of antibiotics, the most important classes of antimicrobials (macrolides, tetracyclines, glycopeptides, β-lactams, and aminoglycosides) were discovered (Katz, 2016). Their industrial production and development into commercial drugs allowed the treatment of bacterial infections saving millions of lives. Nowadays however, the extensive and sometimes improper use of antibiotics is leading to a threatening rise of resistance, making evident not only the necessity to rationalize their use but also to discover novel molecules with alternative mechanisms of action that fight resistant pathogens (De Simeis, 2021).
In this concern, advances in next-generation sequencing and genome mining are speeding up the holistic exploration of *Streptomyces* isolated from unique habitats (Maiti *et al*., 2020; Menegatti *et al*., 2020). More than 600 *Streptomyces* genome sequences are currently publicly available on the NCBI database (NCBI, 2023), and it is estimated that less than 10% of their genetic potential is being explored (Nett *et al*., 2009). However, it has been generally recognized that most of the genetically predicted pathways are silent under standard fermentation conditions. The activation of silent or low-expressing BGCs is thus still one of the biggest bottlenecks in bacterial NPs research (Kalkreuter *et al*., 2020; van Wezel, 2011).
Recent discoveries of novel antibiotics from Streptomycetes include two hexadepsipeptides, meliponamycin A and meliponamycin B from *Streptomyces* sp. ICBG1318 isolated from *Melipona*

scutellaris nurse bees. These cyclic peptides exhibited strong activity against the human pathogens *Staphylococcus aureus* and *Leishmania infantum* (Menegatti *et al*., 2020). Another example is picolinamycin produced from *Streptomyces* sp. SM01 isolated from Indian soils. This novel antibiotic tailors a central picolinamide moiety and three substructures: substructure A contains a pyridine ring along with a benzene ring, substructure B has a pyrimidine ring and a thiazole ring, and substructure C contains a pyridine ring and a pyrimidine moiety attached with a methane group (Supplementary Figure S1). Picolinamycin was shown to be effective against multi-drug resistant pathogens (Maiti *et al*., 2020). Using the OSMAC (One Strain Many Compounds) strategy, *Streptomyces peucetius* DM07 was found to produce a γ-pyrone macrolide. This novel NP was named peucemycin and proved antibacterial activity and a suppressive effect on the viability of various cancer cell lines (Pham *et al*., 2021).
Over the years, *Streptomyces* have provided a vast number of antibiotics and current studies still highlight this genus as a treasure trove for the discovery of novel natural products.

1.2.1 *Streptomyces*: the convergence of morphological development and secondary metabolism

The genus *Streptomyces* belongs to the family Streptomycetaceae, in the order of Actinomycetales. This large group of multicellular Gram-positive bacteria have a DNA with a GC content of 69-78% and a cellular differentiation that resembles the fungal life cycle (Jones, 2018). The life cycle of *Streptomyces* encompasses three well-defined developmental stages: vegetative hyphae, aerial hyphae and spores (Figure 1). Transition between these morphologically differentiated growth phases is controlled by multiple regulators and is largely related to secondary metabolites production (Jones, 2018).

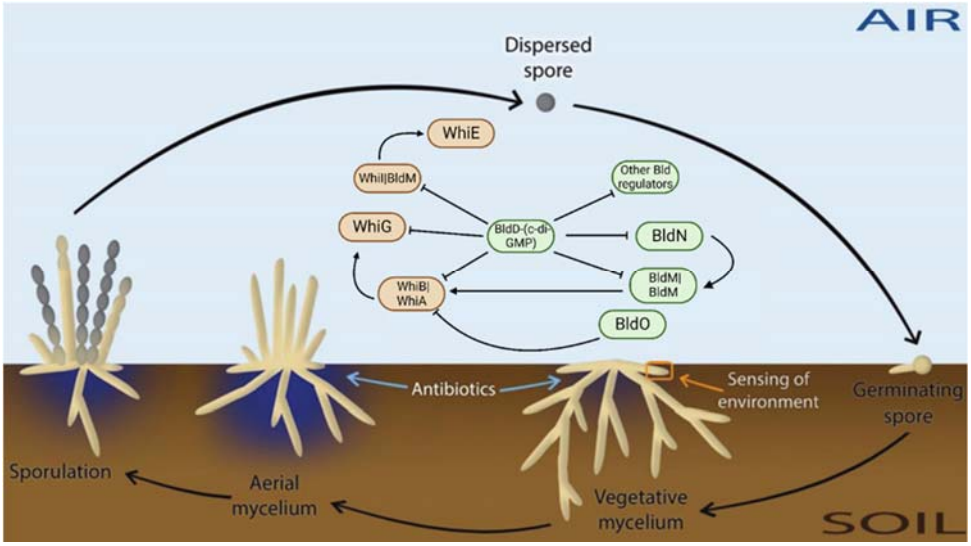


Figure 1. *Streptomyces* life cycle. The traditional *Streptomyces* developmental cycle comprises three stages: vegetative mycelium, aerial mycelium and sporulation. Transition between these morphological stages is regulated by *bld* and *whi* genes. Secondary metabolite production is linked to morphological differentiation and responds to environmental stimuli. (Taken and adapted from Urem *et al*., 2016 and Jones & Elliot, 2018).

Streptomyces are commonly found in soils living as saprophytes (Jones *et al*., 2017). Their life cycle begins with the germination of a single dormant spore. The spore grows out at the hyphal tip to form a dense network of branching hyphae called vegetative mycelium. These filaments anchor the *Streptomyces* colony to its growth substrate through a process mediated by the polarisome, which is a multiprotein involved in recruiting peptidoglycan biosynthesis machinery to the hyphal tips. Under nutrient depletion, a complex developmental program is initiated, whereby the vegetative mycelium is

used as substrate for the second differentiation stage. Genes required for the transition from vegetative to aerial form are known as *bld* genes. The term "bld" refers to the bald phenotype derived from the failure to produce the fluffy white aerial hyphae (Merrick, 1976). Mutants producing aerial hyphae, but no spores are called *whi* mutants given their white phenotype caused by the lack of the gray pigment (Chater, 1972).

The *bld* gene family controls the aerial hyphae stage, which is primarily responsible for regulating aerial mycelia formation. Binding of BldD to c-di-GMP represses *bld* genes such as those encoding BldM, the sigma factor BldN and some *whi* genes including *whiA*, *whiB*, *whiG* and *whiI*. A BldM homodimer represses the sporulation regulator WhiB, and when BldD-mediated repression is relieved, other Blds activate the genes required for development into aerial hyphae. Then, the aerial hyphae rise to the air, where the regulator BldO represses WhiB. Finally, when this repression is released, WhiB along with WhiA and WhiI with BldM activate the genes responsible for sporulation thus initiating a new cycle. The *whi* gene family controls the cessation of aerial hyphae division and guarantees sporogenic cell division at the appropriate time (Xu *et al*., 2022).

Streptomycete morphological development and the onset of secondary metabolite production under natural conditions is linked to complex regulatory mechanisms that are dependent on the environment. Consequently, attempts to obtain secondary metabolites from *Streptomyces* under laboratory conditions have not been entirely successful. When growth in solid sporulating cultures, *Streptomyces* spores germinate and rapidly develop to early compartmentalized hyphae into the agar; then a programmed cell death triggers the differentiation of the multinucleated antibiotic-producing hyphae. This process is followed by formation of the aerial hyphae and a second programmed cell death stage that culminates with the formation of unigenomic spores (Yagüe *et al*., 2016). Although most *Streptomyces* do not sporulate in liquid culture, it has been stated that physiological differentiation is comparable in liquid and solid cultures; the only important difference is that, in liquid cultivation, aerial mycelium formation and sporulation remain blocked (Yagüe *et al*., 2014). Metabolite production may vary in solid and liquid cultures and depending on media composition (Čihák *et al*., 2017; Crüsemann *et al*., 2017).

1.2.2 Regulatory cascades governing *Streptomyces* secondary metabolism

In *Streptomyces*, secondary metabolism is strictly controlled by pyramidal transcriptional regulatory cascades (Figure 2). The top of the pyramid includes general mechanisms common to all species while lower levels imply the regulation of pathways producing a particular secondary metabolite in a specific strain. This complex pyramid-like regulatory architecture can thus be split into four stages, although they are not necessarily independent and may rather constitute a network.

1.2.2.1 Hormone–like signaling molecules

The first regulatory level is the hormone-mediated onset of metabolite biosynthesis. The term "*Streptomyces* hormones" refers to low-molecular-weight compounds that autoregulate secondary metabolism. The best-known hormones are butenolides and γ-butyrolactones (GBLs) which initiate differentiation and secondary metabolism in around 84% of actinomycetes. A-factor (2-isocapryloyl-3R-hydroxymethyl-γ-butyrolactone) was the first signaling molecule discovered and is the representative GBL. Its function on streptomycin production from *S. griseus* has been extensively studied (Kong *et al*., 2019). When A-factor is absent, the receptor ArpA binds to the *adpA* (A-factor-dependent protein) promoter, causing the repression of *adpA*. However, once A-factor reaches a critical concentration (at nanomolar level), the *adpA*/ArpA complex dissociates leading to *adpA* expression (van der Heul *et al*., 2018). AdpA controls more than 500 genes in *S. griseus* thus being considered a

regulator of great importance in the coordination of chemical and morphological differentiation (Higo *et al*., 2012).

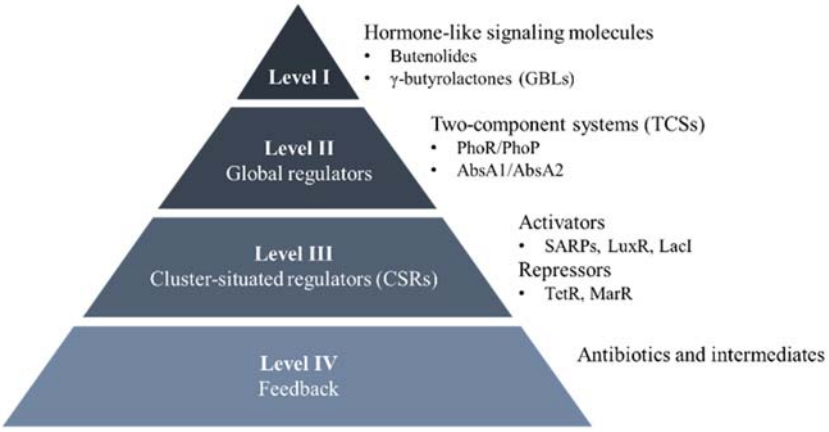


Figure 2. Morphological differentiation and secondary metabolism in *Streptomyces* are controlled by a hierarchical regulatory cascade. At the highest level, hormone-like signaling molecules mediate the transition from vegetative to aerial mycelium and the onset of antibiotic production. Several pleiotropic regulators (level II) such as the two-component systems respond to environmental stimuli and trigger downstream regulatory cascades (at the third level). Antibiotic production can be mediated by feedback (level IV) as a mechanism of self-protection (Xia *et al*., 2020).

1.2.2.2 Global regulators

Global or pleiotropic regulators impact multiple pathways in both primary and secondary metabolism. This regulatory level responds to a variety of chemical and physiological changes such as nutrient limitation, heat shock and pH shift (Xia *et al*., 2020).
The most abundant pleiotropic regulators in *Streptomyces* are the two-component systems (TCSs) like the PhoR-PhoP system. PhoR is a membrane-bound histidine kinase of 426 amino acids which senses specific environmental stimuli while PhoP is the cognate regulator. Under phosphate limitation, PhoR auto-phosphorylates and transfers its phosphate group to PhoP. The phosphorylated PhoP binds to specific sequences in a promoter region known as PHO boxes and ultimately activates gene expression. A PHO box is formed by two 11-nucleotide direct repeat units composed of seven well conserved and four less conserved nucleotides in each (Blanco *et al*., 2002). Another extensively studied TCS from *S. coelicolor* is the AbsA1/AbsA2 system, which is responsible for the activation of calcium-dependent antibiotic transcription (Mingyar *et al*., 2021).

1.2.2.3 Cluster-situated regulators

Cluster-situated regulators (CSRs) represent the ultimate factor determining whether a BGC is expressed or remains silent. The third level focuses on those regulatory proteins whose genes are encoded within a BGC. CSRs directly address the biosynthesis of individual metabolites by responding to pleiotropic regulators and act as a master switch for the expression of biosynthetic genes within the cognate gene cluster. CSRs are classified into families according to their sequence or structure similarity and can work as positive or negative regulators (Mingyar *et al*., 2021). Most of the positive regulators (also called activators) operate either by stabilizing the initial polymerase-promoter complex or by accelerating the transition to the open complex (unwinding of DNA strands). In contrast, negative regulators/repressors bind to promoter regions preventing transcriptional initiation. Characterization of several CSRs in actinobacteria has allowed their organization into superfamilies based on their structure and function. The five families relevant for this thesis are briefly described.

a) *Streptomyces* antibiotic regulatory proteins (SARPs)

SARP stands for *Streptomyces* antibiotic regulatory protein. Members of this family of paralogous proteins show a high specificity for antibiotic production and generally function as activators. SARPs contain a *N*-terminal winged HTH (helix–turn–helix) motif, which is crucial for binding to a conserved DNA sequence within the major groove of the target DNA. The binding motif of SARPs consists of heptameric repeats with 4 bp spacers located stringently 8 bp upstream of the -10 promoter region. This SARP binding sequence overlaps with the -35 region of the target promoter, which is a binding region for most repressors, but not activators. Consequently, SARP-driven transcriptional activation has been proposed to occur via a novel mechanism where the *C*-terminal bacterial activation domain (BTAD) of the SARP activates the transcription of the target genes by recruiting the RNA polymerase (RNAP) to the respective promoter. "Small" SARP-type activators only contain the HTH DNA binding and the BTAD domain, whereas "large" SARPs carry additional domains at the *C*-terminus. These regions include a domain of unknown function belonging to the P-loop NTPase family, and one or more copies of a tetratricopeptide repeat (TPR) motif. The ATPase activity is essential for the isomerization of the closed complex between the SARP and the RNAP to a transcriptionally competent open complex (Tanaka *et al*., 2007).

b) LuxR family

The LuxR protein from *Vibrio fischeri* is a transcriptional activator for the quorum sensing (QS) of luminescence and is a prototype of this large family of regulators. LuxR regulators are proteins of less than 250 amino acids with a *C*-terminal DNA-binding HTH and an *N*-terminal domain that responds to chemical signaling molecules such as γ-butyrolactones. The *N*-terminal region senses and binds its specific QS signal whereas the *C*-terminus binds to the target DNA leading to gene expression (Shadel, 1991). The binding motif of LuxR-like regulators (GGNNNNCCC) was obtained by comparison of more than ten genes confirmed to be activated by LuxR proteins (Ordóñez-Robles, 2016). Although most of the LuxR-type regulators involved in antibiotic production in *Streptomyces* seem to be activators, some studies suggested a possible role as negative regulators (Mingyar *et al*., 2021).

c) TetR proteins

TetR proteins represent another large and important family of one-component signal transduction systems found in the biosynthetic gene clusters of antibiotics and other metabolites in *Streptomyces*. All TetR regulators consist of an *N*-terminal DNA binding domain and a larger *C*-terminal domain which interacts with one or more ligands and alters the ability of the regulator to bind DNA. These proteins are almost exclusively α-helical and function as dimers binding to repeated-palindromic DNA operator sequences (Cuthbertson, 2013).
TetR proteins primarily regulate the expression of efflux pumps required for antibiotic export but may also regulate the expression of other biosynthetic genes. Although they generally work as repressors, it has been hypothesized that there is a correlation between the location of the DNA-binding domain and the repressor or activator activities. According to this model, repressors contain an *N*-terminal DNA-binding domain while activators have the DNA-binding domain at the *C*-terminus (Pérez-Rueda, 2001).

d) Multiple antibiotic resistance regulator (MarR)

The MarR family of transcriptional regulators encompasses more than 19,000 members widespread in bacteria and is the fourth most abundant class of regulators in *Streptomyces*. MarRs control a variety of cellular processes such as multidrug resistance, pathogenicity, stress response and metabolic pathways.

They typically act as homodimeric transcriptional repressors and only in some cases as activators. MarR proteins have a characteristic wHTH DNA-binding domain and bind to palindromic sequences within the target promoter region. The DNA-binding affinity of MarR is reduced by conformational changes as result of small-molecule ligands stimulus, leading to transcriptional regulation blockage (Guo *et al*., 2018). Some studies have shown that cluster-situated MarRs in *Streptomyces* potentially work as activators by binding to 37 bp consensus sequences harboring two highly conserved motifs GAAAT(A/G)(T/C)ATCG and CTTAT(A/G)TA(A/G)GCT (Zhu *et al*., 2013).

e) LacI family

Most LacI transcription factors sense sugar effectors and regulate carbohydrate utilization genes; in the absence of the inducing compounds, they directly interact with DNA as repressors. These regulators work as either activators or repressors by binding a specific effector ligand that either decreases or increases DNA-binding affinity.
The *N*-terminal DNA-binding domains of LacI regulators are similar to the helix-turn-helix domains of other transcriptional factors, whereas the ligand-binding domain is homologous to the periplasmic proteins of ABC-transporters. Such domains can act as a sensor binding small molecule ligands that the DNA-binding domain responds to, recognizing sugars, sugar phosphates, sugar acids and purines. LacI-binding sites are located within the area between 140 and 30 bp upstream of the start codon and in 1%, the sites are within the coding region. These binding sequences are normally palindromes with a conserved GC pair in the center (Ravcheev *et al*., 2014).

1.2.2.4 Feedback regulation

Antibiotics and their biosynthetic intermediates can serve as feedback substances for the processing enzymes to modulate their own expression (Kong *et al*., 2019). One example of this regulatory mechanism is jadomycin B (JdB) from *S. venezuelae*. JarR1 is the main cluster-situated activator controlling the pathway in the presence of traces of JdB. However, when high concentrations of JdB are sensed, JadR1 dissociates from its target promoter, and the JdB-JadR1 interaction directly controls jadomycin production in a dose-dependent manner. Another example is the red-pigmented antibiotic undecylprodigiosin (also known as Red) produced by *S. coelicolor* and *S. lividans*. During Red biosynthesis, RedZ positively regulates transcription of *redD*, the main cluster-situated activator of Red biosynthesis. When Red is found in high concentrations, it interacts with RedZ thus modulating its activity and as consequence the repression of Red production (Wang *et al*., 2009).
Taken together, the complex hierarchical regulation of secondary metabolism in *Streptomyces* still constitutes the major bottleneck to produce novel bioactive compounds under laboratory conditions. Manipulation of global regulators as well as engineering of target regulatory genes in silent BGCs are viable ways to activate biosynthetic pathways in native producers.

1.2.3 Strategies for activation of metabolic pathways in *Streptomyces*

Metabolic pathway silencing can be tackled by manipulating the regulatory network of the producing organism. Strategies targeting high levels of regulation (I-II) are straightforward since no genome engineering is required, yet the rate of success is unpredictable. Conversely, approaching lower levels does require genetic manipulation, but it increases the likelihood of success since a specific pathway is being forced to awake passing over the natural molecule/signal controlling the cascade initiation. Different approaches have been developed for pathway activation and to enhance the production of specialized metabolites. These can be “unselective”, when they cause a global change in secondary metabolite output or “selective”, when a specific BGC is manipulated (Craney, 2013).

1.2.3.1 Elicitation

Elicitation is an unselective strategy for bacterial pathway activation. Challenging bacteria with external stimuli (elicitors) has been long recognized as an effective way to biologically activate metabolism and to enable the identification of physiological responses. It has been proposed that elicitors are initially sensed by binding to specific receptors; this alters the level of intracellular signaling effectors such as Ca^{2+} and reactive oxygen species, leading to direct activation of BGCs or causing the expression of regulatory genes.

The most straightforward strategy to stimulate bacterial metabolism is by using physical elicitors. Moderate to drastic changes in environmental factors like temperature and pH can be sufficient to trigger the biosynthesis of compounds. The production of validomycin in *Streptomyces lavendulae* ACR-DA1, for example, varied considerably depending on the cultivation temperature, being 10 ºC the condition yielding the highest production (Sharma *et al.*, 2017). Prodiginine biosynthesis from *Streptomyces coelicolor* M511 increased 2.5-fold when the pH of the culture medium was decreased from 7.2 to 4.0 (Mo, 2013). Addition of strong bases to the fermentation medium increased the validamycin A production from *S. hygroscopicus* (Jiang *et al*., 2018). Although temperature and pH are parameters easily altered in a laboratory setting, the outcome is unpredictable.

A more rational approach to trigger antibiotic production in *Streptomyces* is co-cultivation-based elicitation. This implies a culture containing more than one organism from different species, genera or even kingdoms, where the overall environment is more complex and more representative of the natural habitat, possibly resulting in activation of metabolite production by physical contact. Several *Streptomyces* strains have shown improved antibiotic production when growth in co-culture. The anti-staphylococcal agent holomycin, for example, is only produced in co-culture of *Streptomyces clavuligerus* and *Staphylococcus aureus* (Charusanti *et al.*, 2012). Co-cultivation of 44 soil-derived *Streptomyces* strains with mycolic acid-containing bacteria induced changes in the metabolic profile of 35-40 of the strains tested; overall this derived in the production of 49 putative new compounds that were not detected in monocultures (Wang, *et al*., 2021).

Another alternative to potentiate yields of silent compounds consists of supplementation of cultures with small molecules. A variety of chemical elicitors have been successfully applied in *Streptomyces*, but only in some cases the mechanisms have been elucidated. The addition of rare earth elements like scandium, yttrium or lanthanum are powerful stimuli that activate antibiotic synthesis, as it has been proposed they target RNA polymerase and ribosomes. The use of metals like nickel, cobalt, zinc, chromium and manganese to induce stress-responses has also been reported to activate metabolite production. For example, metal-stress induced the production of two aromatic polyketides, stremycin A and B in *Streptomyces pratensis* (Akhter *et al*., 2018).

1.2.3.2 Selective activation of metabolic pathways

As stated before, selective activation of metabolite biosynthesis implies the manipulation of lower levels in the regulatory cascade. Overexpression (OE) of positive CSRs has been successfully applied for a variety of metabolites in actinobacteria. One example is the 2-fold increase of toyocamycin production by *S. disatotochromogenes* 1628 through OE of the LuxR regulator ToyA under the control of the SPL57 promoter (Zhou, 2020).

In a similar way, inactivation/deletion of negative regulators can enhance metabolite titers in heterologous expression systems and can activate silent pathways when combined with OE of positive regulators in natural producers. Disruption of the TetR regulator CalR3 resulted in a 30-fold and 171-fold increase of calcinycin and cezomycin production respectively (Zhou, 2020). Deletion of negative regulators from the TetR family increased avermectin production in *S. avermitillis*, calcimycin production in *S. chartreusis* and pristinamycin in S. *pristinaespiralis* (Xia *et al*., 2020).

Other strategies include overexpression of transporters and tuning feedback. Many BGC-linked transporters belonging to the ATP-binding cassette (ABC) superfamily and major facilitator superfamily (MFS) are important for metabolite export. Transporters are crucial for secretion of antibiotics to reduce their intracellular accumulation, which may lead to self-toxicity. Overexpression of the ABC-transporter AvtAB in *S. avermitilis* increased the production of avermectin B1a 2-fold (Qiu *et al*., 2011). The biosynthesis of the ribosomally produced peptides planosporicin and microbisporicin was found to be modulated by multiple ABC transporters localized in their respective gene clusters (Xia *et al*., 2020). Similarly, the overexpression of the putative efflux pump BotT encoded in the bottromycin BGC increased production about 20-fold in a heterologous host (Huo *et al*., 2012).
Antibiotic-producing bacteria may be sensitive towards their own compounds during fermentation; bacteria deal with this problem by either self-resistance or not producing the metabolites until they have passed through their complete growth phase. However, when microorganisms are cultivated under defined conditions, the active metabolite(s) can be produced by developing cells, making necessary the implementation of strategies to circumvent toxic effects. In this regard, adsorbent resins like amberlite XAD-16, diaion HP-20, and silica gel have been effectively used for up taking metabolites reducing feedback repression in laboratory cultures (Lee *et al*., 2003). An illustrative example is the 1.25-fold increase of pristinamycin production by *S. pristinaespiralis* by adding 12% (w/v) of JD-1 resin into the culture after 20 hours of inoculation (Jia *et al*., 2006).

1.2.4 Other regulatory systems directing natural products biosynthesis

Beyond all regulatory mechanisms described above, metabolite production is influenced to a great extent by external stimuli. In their natural environment, bacteria encounter many other microbes in their surroundings. In some cases, this contact leads to "conflicts" if the microbes compete for nutrients or if a pathogen attacks (Kaur *et al*., 2020). To defend themselves, bacteria have developed "immune systems" to detect pathogens and to trigger the production and subsequent use of antibiotics and toxins as defense mechanisms (Werren, 2011). Although it is not clear how bacteria evolved these systems, recent investigations provided insight into new immune systems in diverse multicellular bacteria (Kaur *et al*., 2020). In general, these systems are constituted by multiple molecular components and require a large investment of energy to be triggered. Such an activation barrier prevents bacteria from using weapons unless they are in real danger (Kaur *et al*., 2020). Multiple systems have been bioinformatically found to be widespread in actinobacteria, cyanobacteria, planctomycetes, and chloroflexi. One example of conflict systems is the chaperone/co-chaperone-based systems. These complexes are composed by a MoxR-like AAA+ ATPase (ATPases associated with various cellular activities) and its co-chaperone, which is a peptide-binding von Willebrand factor A (vWA) domain (Wong & Houry, 2012). These two proteins are linked with at least one other component, thus forming multiplexes known as ternary systems (Figure 3).

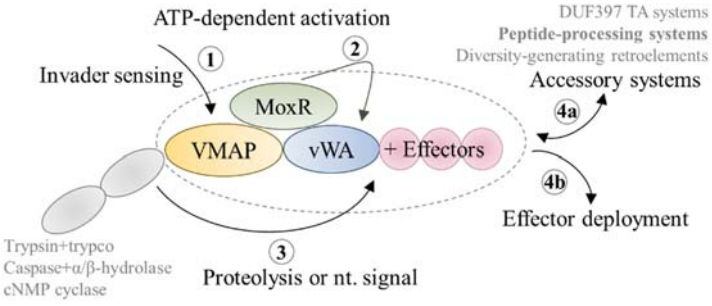


Figure 3: Schematic representation of the plausible mechanisms of action of a classic ternary biological conflict system. The ternary core is composed of a VMAP (vWA-MoxR associated protein), a MoxR AAA+ ATPase, and

a von Willebrand factor A. Invader sensing (1), activation (2), proteolysis (3) and effector deployment (4) are indicated. (Taken and modified from Kaur *et al.*, 2020).

About 20% of the actinobacterial ternary systems are linked with a peptide-processing system. The core of these systems is a gene encoding a zincin-like metallopeptidase and another small peptide (annotated as FxSxx-COOH small protein in GenBank). The peptide is cleaved by the former peptidase to release a small *C*-terminal peptide. A distinctive radical SAM protein is encoded in the proximity of these systems and is predicted to catalyze modifications on the peptide. It is thought that the modified peptide acts as a peptide toxin or signal that amplifies the action of the core system in a subset of the ternary system (Kaur *et al*., 2020). It is also possible these peptide-processing systems exist independently. To date, none of these bioinformatically described systems has been investigated experimentally.

1.3 Ribosomally synthesized and post-translationally modified peptides

Ribosomally synthesized and post-translationally modified peptides (RiPPs) constitute one of the largest, but still primarily underexplored NP families with remarkable structural and functional diversity. These metabolites take advantage of the ribosomal translation machinery to produce linear peptides that are subsequently modified by dedicated biosynthetic enzymes. Although RiPPs are limited to proteinogenic amino acids, the plethora of post-translational modifications (PTMs) that processing enzymes can install provides RiPPs with a high diversity of molecular scaffolds and structural modifications (Arnison *et al*., 2013).
In the last decade, the research field has seen great progress in understanding the mechanisms by which nature assembles these metabolites, in generating new tools for pathway engineering, and in determining potential application of RiPPs. Due to the dramatic increase of novel structures reported during the last years, in 2019 a group of RiPP experts collected all the new discoveries in a RiPP community-wide review. According to that publication, more than 41 classes of RiPPs have been reported and it is expected this number to expand further during the coming years (Montalbán-López *et al*., 2021). This comprehensive report also provided the basis for classifying new RiPPs based on the structural features which rely on types of modification, and where each class represents a unique subset of biosynthetic logic. Even though some fungi, plants and marine organisms have been reported to produce RiPPs, they are essentially produced by bacteria (Li *et al*., 2021).

1.3.1 The purpose of RiPPs in nature

To understand why bacteria produce certain metabolites and to comprehend the regulatory mechanisms controlling the activation of RiPP biosynthesis, it is necessary to consider that bacteria live in complex polymicrobial communities where interactions between individuals shape the composition of populations and their function (Hibbing *et al*., 2010).
Among the different roles of RiPPs in nature, their participation in microbial competition is the best documented. Direct or interference competition means the attack of one individual to another by the secretion of harmful molecules (Raffatellu, 2018). An example of RiPP-interference competition is the production of thuricin CD by *B. thuringiensis* DPC 6431 obtained from a human fecal isolate (Rea *et al.*, 2010). This two-component sactibiotic synergistically kills the pathogen *Clostridium difficile*, which is the causative agent of nosocomial infections. Interestingly, thuricins have a moderate impact on other genera that contribute to microbiota health such as *Lactobacillus casei* and *Bacillus lactis*. Thuricins are an example of how RiPP-producing bacteria can selectively outcompete a pathogen without perturbing their natural environment (Rea *et al.*, 2010). Another example of RiPPs involved in competition is the thiopeptide GE7468 produced by the symbiont of the ant *Trachymyrmex*

septentrionalis. In this case, ants use the thiopeptide to antagonize their competitors, which are mainly other ant-associated bacteria (Chang *et al.*, 2020).

RiPPs also participate in communication. Bacteria communicate with each other using the quorum sensing (QS) strategy, which involves the production, diffusion, and perception of extracellular signaling molecules that subsequently regulate gene expression at a community level. In Gram-positive bacteria, ribosomal peptide NPs commonly mediate QS mechanisms through three different mechanisms (Thoendel, 2010). First, the RiPP itself can be a QS signaling molecule that regulates other physiological processes; second, the RiPP can work as autoinducing peptide (AIP) controlling its own production and last, RiPP biosynthesis is strongly regulated by QS based on other signals (Thoendel, 2010).

Other ecological roles of RiPPs involve microbial physiology, where they can act as promoters of biofilm formation and as enzyme cofactors in morphological development. The thiopeptides thiocillin and thiostrepton, for instance, stimulate biofilm formation in *Bacillus subtilis* and *Pseudomonas aeruginosa,* respectively (Bleich, *et al*., 2015; Ranieri *et al*., 2019). The lanthipeptides SapB from *S. coelicolor* and SapT from *Streptomyces tendae* were shown to be essential for the development of aerial hyphae from vegetative cells (Kodani *et al*., 2004; Kodani *et al.*, 2005). Classic examples of RiPPs working as cofactors are pyrroloquinoline quinone (PQQ) and mycofactocin. Both small ribosomal peptides serve as redox cofactors of dehydrogenases such as alcohol and glucose dehydrogenase, allowing the producer to use these substances as carbon source (Shen *et al*., 2012).

All in all, the least understood role of RiPPs in nature is their mediation of plant-microbe interactions, and their implication in plant protection against pests. RiPPs are produced by living organisms for certain reasons; for science however, they represent an arsenal of novel chemical entities with promising application in environmental, medical, veterinary and food industrial sectors.

1.3.2 Biosynthetic steps common to all RiPPs

RiPP discovery has particularly benefited from genome mining since all the elements required for biosynthesis are encoded within the same DNA stretch, forming BGCs that can be easily accessed by using synthetic engineering. This colocalization of genes encoding the substrate peptide and biosynthetic proteins also facilitates prediction of the entire biosynthetic pathway. A classic RiPP gene cluster (Figure 4) typically encodes one or multiple precursor peptide(s), a variable number of tailoring enzymes (also known as modifying enzymes), CSRs, transporters and occasionally a protease for leader removal (Hudson & Mitchell, 2018).

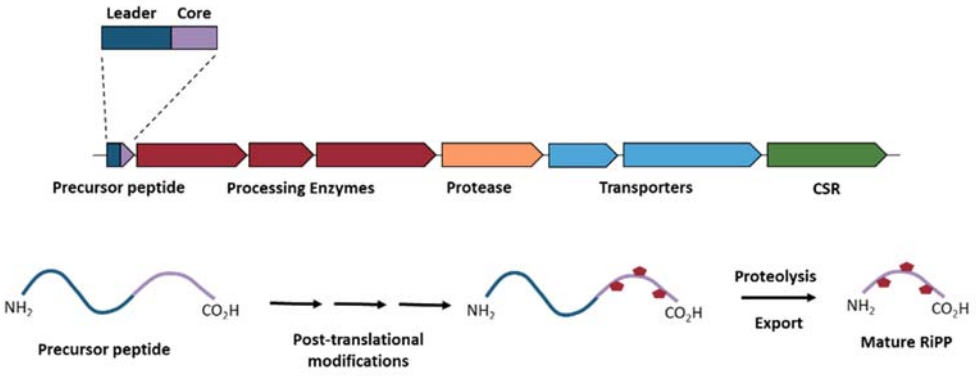


Figure 4. RiPP biosynthetic gene cluster (BCG) and RiPP biosynthesis. **Top:** a generic RiPP biosynthetic gene cluster consisting of a precursor peptide, modifying/processing enzymes, cluster-situated regulators (CSRs), and proteolysis/export enzymes. RiPP precursor features leader and core regions. **Bottom:** after translation, the core peptide is processed by modifying enzymes with a final leader removal and export of a mature RiPP. (Taken and modified from Tan *et al.*, 2019).

Biosynthesis steps common to all RiPPs involve the translation of a precursor peptide normally 20 to 110 residues in length. The *N*-terminal region, known as leader peptide (LP), provides binding affinity for many of the processing enzymes and ensures that PTMs are carried out in the correct order. The LP can serve as an allosteric effector for activation of biosynthetic enzymes and to keep the maturing peptide inactive. The *C*-terminal region is known as core peptide (CP) and is the region where the post-translational modifications take place (Figure 4). Some RiPP precursors also contain a small amino acid chain attached to the *C*-terminus which is known as follower peptide. As defined in the standardized nomenclature for RiPP precursors, residues in the core peptide are indicated with a positive number starting from the junction with the leader sequence, whereas the *C*-terminal residue of the LP is designated as -1 and numbers decrease towards the *N*-terminus (Montalbán-López *et al*., 2021).
Posttranslational modifications can generally be organized into primary and secondary. RiPPs classification is defined by primary modifications which are common to all members of a specific subfamily. Primary PTMs mainly include dehydration, crosslinks between amino acids, azole formation and macrocyclizations (Zhang, 2014). Secondary modifications (or tailoring reactions) are compound-specific and characterize an individual RiPP. Secondary PTMs allow additional molecular interaction between the mature RiPP and its target, leading to a higher affinity and selectivity. Common secondary PTMs are epimerization, methylation, phosphorylation, hydroxylation, halogenation, glycosylation, and additional modifications of *N*- and *C*-termini and other rearrangements (Montalbán-López *et al*., 2021).
RiPP biosynthesis usually ends with the leader removal and export of the mature peptide out of the cell (Figure 4). A deeper description of selected RiPP families relevant for this thesis is provided in sections 1.3.3- 1.3.5.

1.3.3 Thiopeptides

With more than one hundred members, thiopeptides represent one of the largest and most structurally complex subfamilies of RiPPs. Thiopeptides are characterized by a nitrogen-containing six-membered heterocycle attached to the peptide backbone. This central ring serves as scaffold to at least one macrocycle and a tail. Besides the central heterocycle, the characteristic PTMs in thiopeptides include dehydroamino acids and azoles such as thiazoles, oxazoles and thiazolines (Hudson & Mitchell, 2018).
As for all RiPPs, thiopeptide biosynthesis (exemplified by the thiomuracin biosynthesis shown in Figure 5a) initiates with the regular ribosomal translation of a precursor peptide. The assembly process can be divided into pre- and post-macrocyclization. The first stage involves primary PTMs common to all known thiopeptides while the second stage is pathway (compound) specific. Current understanding concerning the order of reactions during the pre-macrocyclization comes from *in vitro* and *in vivo* studies of well-known thiopeptides like thiomuracin, thiostrepton, lactazole and GE2270A (Vinogradov & Suga, 2020).
The cascade of PTMs begins with azole formation. Thiazoles and oxazoles are installed from selected Cys, Thr and Ser residues in a two-step mechanism which involves three biosynthetic enzymes. In a first step, a YcaO enzyme uses ATP to activate a backbone amide and takes the side chain of the adjacent amino acid as nucleophile to perform a heterocyclization yielding azoline. This reaction is leader dependent and requires an additional enzyme to bring the precursor peptide and the YcaO in proximity. This partner is known as Ocin-ThiF-like protein and contains a RiPP recognition element (RRE) which is a conserved domain recognized by processing enzymes to engage their substrates. The second step is the conversion of azolines into azoles. This is an FMN-dependent oxidation performed by a dedicated dehydrogenase. Of note, oxazoles are less common than thiazoles (Vinogradov & Suga, 2020).
It has been proposed that dehydration of amino acids follows azole formation. Serine and threonine residues are converted to dehydroalanine (Dha) and dehydrobutyrine (Dhb) respectively. The two-step

process consists of Ser/Thr side chain activation through glutamylation by a $tRNA_{Glu}$-dependent glutamylation enzyme, and the subsequent glutamate elimination by means of a retro-Michael reaction catalyzed by a glutamate elimination enzyme (Bothwell *et al*., 2019).

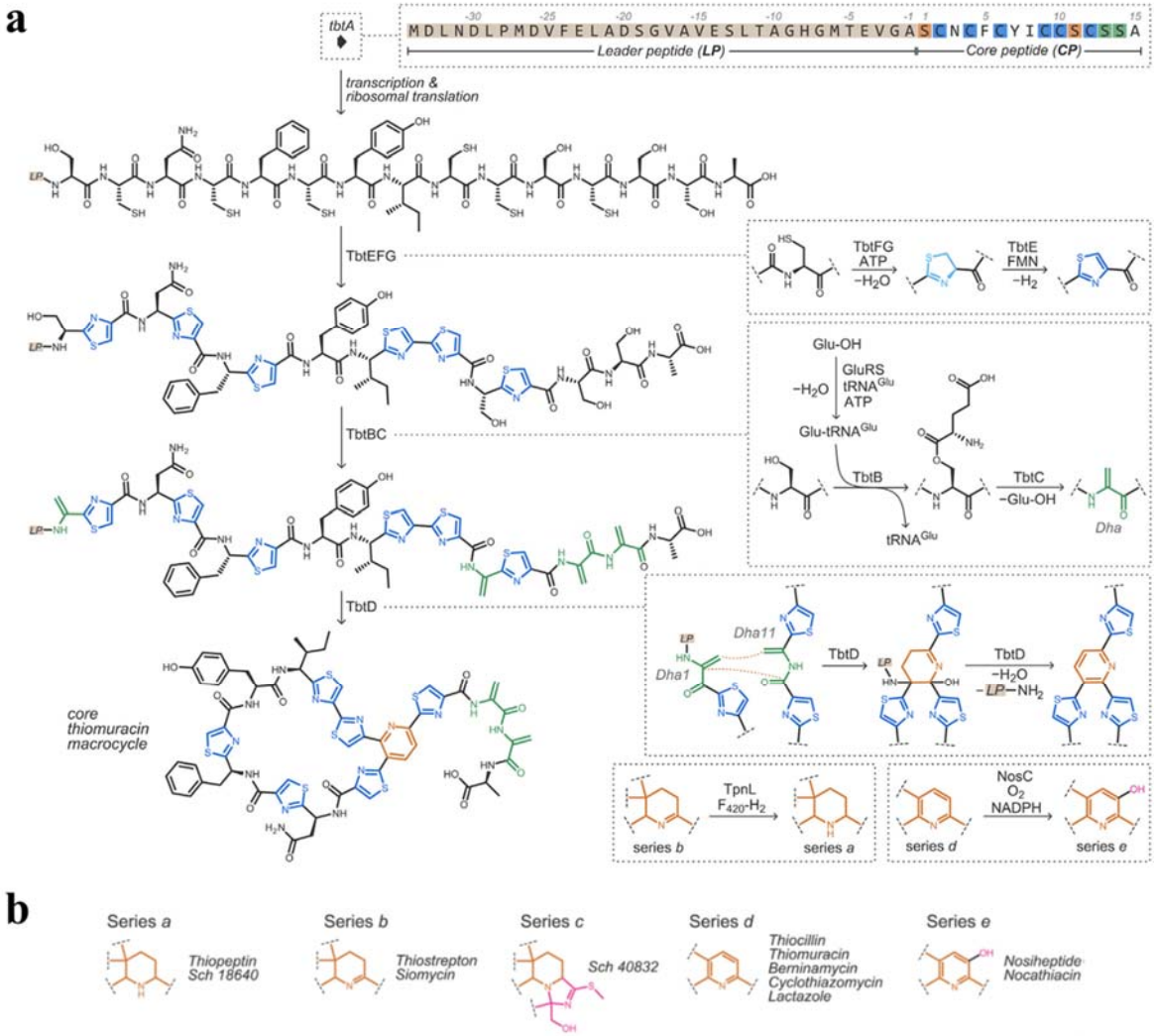


Figure 5. Thiopeptide biosynthesis. **(a)** Assembly process exemplified by the thiomuracin (tbt) biosynthesis. Post-translational modifications (PTMs) catalyzed by individual enzymes are shown on the right panel. Precursor peptide (TbtA), dehydratase glutamylation domain (TbtB); dehydratase elimination domain (TbtC), [2+4]-cycloaddition/macrocyclization (TbtD), FMN-dependent dehydrogenase (TbtE), Ocin-Thif-like (TbtF), and YcaO cyclodehydratase (TbtG). **(b)** Classification into series of thiopeptides according to the oxidation state of the central heterocycle and representative examples. (Taken and modified from Vinogradov & Suga, 2020).

Once oxazoles and dehydroamino acids have been installed, macrocyclization takes place. This reaction is performed by a single enzyme that uses two Dha side chains to catalyze a formal aza-Diels-Alder reaction giving a macrocycle with a dehydropiperidine fixed inside of it. For some thiopeptides, the cycloaddition enzyme consecutively eliminates a molecule of water along with the leader peptide to aromatize the heterocycle; for others however, the mechanism remains still unknown as the heterocycle adopts different oxidation states (Arnison *et al*., 2013). Thiopeptides are classified into five series (a – e) based on the oxidation state and substitution pattern of their central six-membered ring (Figure 5b). Thiopeptides from series a, b and c have a piperidine, a dehydropiperidine and an imidazopiperidine as the core nitrogen heterocycle, respectively. Members of series d contain a trisubstituted pyridine while

series e presents an additional hydroxyl at position 5. To date, the majority of thiopeptides discovered belong to series d and series e. This is in accordance with bioinformatic studies where 97% of over 500 thiopeptide BGCs were predicted to encode series d and e thiopeptides (Schwalen, *et al.*, 2018).
The post-macrocyclization stage in thiopeptide biosynthesis involves a variety of modifications that characterize individual members. Thiostrepton, for instance, tailors a second macrocycle containing a quinalidic acid attached. This unique decoration requires the participation of seven biosynthetic enzymes encoded in the thiostrepton BGC, but after several years of research, the whole mechanism remains still unclear (Duan *et al*., 2012). In a similar manner, five processing enzymes are required for the incorporation of 3-methylindolic acid and the formation of the side ring in nosiheptide (Yu *et al*., 2009).
Other secondary PTMs include methylation and hydroxylation catalyzed by radical SAM (*S*-adenosyl-methionine) proteins and cytochrome P450 (CYP) enzymes, respectively. In additional cases, thiopeptides can be modified at the *C*-terminus, either, by removal of the final residue, or by amide formation (Vinogradov & Suga, 2020).
Besides the classification into series, thiopeptides are organized according to the size of the primary macrocycle. Structures known to date feature 26-, 29-, 32- and 35- membered rings, and a relationship between the ring size and biological activity has been proposed. Thiopeptides possess diverse bioactivities such as antibacterial, antiviral, antiparasitic and immunosuppressive. Studies have proven that thiopeptides act as antibacterials by interfering with protein synthesis. Thiopeptides containing 26- and 35-membered rings such as thiostrepton, nosiheptide, thiocillin and berninamycins inhibit protein synthesis by binding to the large subunit of the bacterial ribosome, targeting the interface of S23 rRNA and L11 ribosomal protein. On the other hand, 29-membered thiopeptides like thiomuracin use a different mechanism inhibiting the elongation factor Tu (EF-Tu) (Montalbán-López *et al*., 2021).

1.3.4 Lanthipeptides

Lanthipeptides, the largest sub-family of RiPPs, are polycyclic structures distinguished by the presence of sulfur-to-β–carbon thioether crosslinks and multiple dehydroamino acids (Montalbán-López *et al*., 2021). Lanthipeptides are of particular importance due to their ability to kill Gram-positive bacteria, especially those exhibiting drug resistance-like methicillin-resistant *Staphylococcus aureus*, *Streptococcus pneumoniae* and vancomycin-resistant *Enterococci* and *Clostridium difficile* (Karbalaei-Heidari, 2020). When lanthipeptides are confirmed to have antibiotic activity, they are named lantibiotics. Most lanthipeptides exert their antibacterial activity by binding to lipid II which is an essential intermediate in peptidoglycan biosynthesis; this results in blockage of cell wall formation and disruption of the membrane integrity due to pore formation. Alternatively, lanthipeptides with broader spectrum act by recognizing and combining phosphatidylethanolamine, an important component of membrane structures and viral envelopes (Li *et al*., 2021).
Nisin was the first-identified lantibiotic reported in 1928 and is currently applied as a preservative food additive in more than 80 countries. This lantibiotic has been used for over 60 years without the appearance of significant bacterial resistance. In addition, recent studies have pointed out a broader potential application of lanthipeptides as antiviral, immunomodulators and anti-allosteric molecules (van Staden *et al*., 2021). Other lantibiotics like duramycin, NAI-107, mutacin, and NVB-302 have been tested in the clinic and demonstrated potent antimicrobial activity *in vivo* (Zhao & Kuipers, 2021).
Several studies have reported the existence of two-component lanthipeptide systems that exert their antimicrobial activity against Gram-positive bacteria by using synergistic mechanisms of action. Cytolysin, lacticin 3147 and haloduracin are the best studied two-component lanthipeptides (Walker *et al*., 2020). In these systems, individual peptides have little or no activity while equimolar concentrations of the two peptides exhibit significantly higher activity. The first peptide, termed as the α-peptide,

contains the lipid II-binding motif, and although it interacts with lipid II, it is not able to inhibit cell wall biosynthesis on its own. However, binding of the α-peptide to the lipid II causes the recruitment of the β-peptide which aids in pore formation. Two-component lanthipeptide systems are commonly found in Firmicutes and only two such systems, roseocin and birimositide, have been reported from Actinobacteria (Walker *et al*., 2020).

1.3.4.1 Lanthipeptide biosynthesis and classification

Lanthipeptide biosynthesis comprises four main steps: (i) translation of precursor peptide; (ii) formation of thioether rings which involves the dehydration of Ser and Thr residues to form dehydroalanine (Dha) and dehydrobutyrine (Dhb), respectively, and a subsequent Michael-type addition of a Cys residue onto the dehydroamino acid to form lanthionine and methyllanthionine, respectively; (iii) secondary posttranslational modifications, and (iv) removal of the leader peptide (Li *et al*., 2021).
According to the enzymes catalyzing dehydration and cyclization, lanthipeptides are divided into five classes (Figure 6a). Class I are synthesized by two enzymes, a discrete aminoacyl-tRNA dependent dehydratase LanB, and a cyclase LanC (Figure 6b). Class II lanthipeptides are formed by a single enzyme (generically called LanM) that has a dual function for dehydration and cyclization. LanM homologues contain an *N*-terminal ATP-dependent dehydratase domain that bears no homology to LanB, and a *C*-terminal cyclase domain. LanM synthetases phosphorylate Ser and Thr side chains on their substrate by using ATP (Figure 5b); this reaction is followed by β-elimination of the phosphorylated residue and subsequent cyclization (Repka *et al*., 2017). Class III and IV lanthipeptides are synthesized by trifunctional enzymes LanKCs and LanLs, respectively. These enzymes contain three domains: a central kinase which activates the hydroxyl group of Ser/Thr residues through NTP-dependent phosphorylation, an *N*-terminal lyase domain responsible for phosphate elimination to yield Dha/Dhb, and a *C*-terminal cyclase domain that catalyzes the nucleophilic attack by the cysteine thiol of the unsaturated double bound. Of note, the cyclase domain of LanMs and LanLs contains a conserved zinc-binding motif (Cys-Cys-His/Cys) while that of LanKCs lacks these conserved residues (Krawczyk, *et al*., 2012; Hegemann & van der Donk, 2018).
A distinctive characteristic of LanKC synthetases is their ability to form bicyclic structures known as labionin and methyllabionin. During (methyl)labionin synthesis, the enolate intermediate obtained from the thioether bridge formation undergoes a second nucleophilic attack of another Dha moiety yielding a carbocyclic ring (Figure 6b). Requirement for LanKCs to install labionin structures is the presence of a conserved $S_{X2}S_{X2\text{-}5}C$ motif in the substrate core peptide (Meindl *et al*., 2010).
Class V is a recently discovered group of lanthipeptides (Xu *et al*., 2020). In contrast to classes I to IV, the biosynthesis of class V lanthipeptides involves three independent enzymes responsible for phosphorylation, dehydration and cyclization. Class V synthetases show remarkable differences to other modifying enzymes and recent bioinformatics analyses suggested a high diversity in their structures and catalytic mechanisms (Xu *et al*., 2020). This implies a richer structural and biosynthetic diversity; thus, in the near future, a more complex classification for lanthipeptides will likely be required.
Secondary PTMs identified in lanthipeptides include halogenation, acylation, hydroxylation, installment of *C*-terminal (2-aminovinyl)-cysteine, lysoalanine, D-amino acids, among others (Repka *et al*., 2017). Epilancin 15x for instance, is a class I lanthipeptide that contains an unusual *N*-terminal lactate group. During Epilancin 15x biosynthesis, proteolytic removal of the leader peptide by the subtilisin-like serine peptidase ElxP exposes an *N*-terminal Dha which hydrolyzes giving a pyruvyl group. Then, the short chain dehydrogenase ElxO reduces the ketone of the pyruvyl group forming an *N*-terminal lactyl moiety (Ortega *et al*., 2014). A second example is the class III lanthipeptide NAI-112, which contains a 6-deoxyhexose moiety *N*-linked to a central tryptophan side chain. This PTM is installed by a glycosyltransferase found within the NAI-112 gene cluster (Iorio *et al*., 2014). Another

example is the biosynthesis of lipolanthines that involves two additional PTMs: a guanidine fatty acid attached to the *N*-terminal of the core peptide, and an avionin structure originated by decarboxylation of the labionin at the *C*-terminus. Avionin formation requires the action of a LanKC and a FAD-dependent cysteine decarboxylase (Wiebach *et al*., 2020).

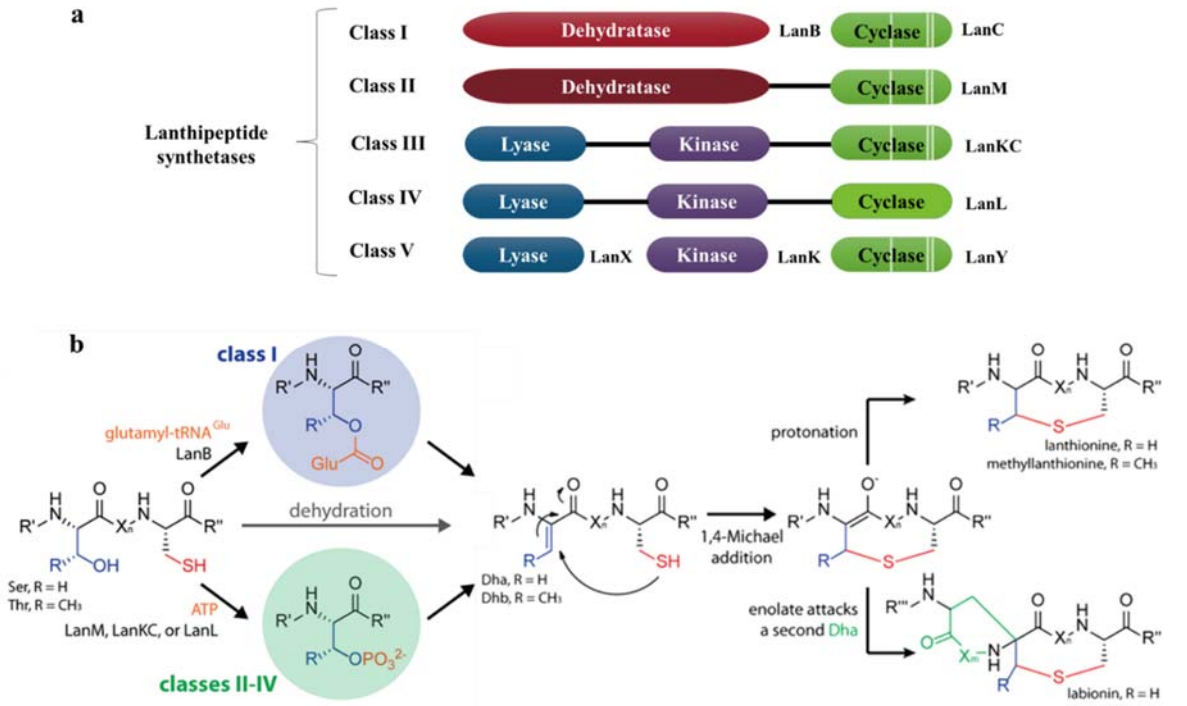


Figure 6. Classification of lanthipeptide synthetases and primary post-translational modifications in lanthipeptide biosynthesis. **(a)** Five classes of lanthipeptide modifying enzymes install dehydroamino acids (Dha/Dhb) and thioether crosslinks. "Lan" is the generic term used to name lanthipeptide synthetases. In the cyclase domain, lines indicate the zinc binding motif. The zinc ion is thought to activate the Cys thiols of the precursor peptide for nucleophilic attack on the dehydroamino acids. **(b)** Main steps in lanthipeptide biosynthesis. Peptide processing starts with the dehydration of Ser/Thr residues to yield Dha and Dhb followed by cyclization through a 1,4-Michael addition. (Taken and modified from Hegemann *et al*., 2019).

One of the most valuable secondary PTM in lanthipeptide biosynthesis is the incorporation of D-amino acids. This modification imparts a range of favorable properties like reduced susceptibility to proteolysis, higher bioactivity, induction of turns, and self-assembly into higher order structures (Yang & van der Donk, 2015). D-amino acids are introduced into lanthipeptides by hydrogenation of Dha/Dhb, resulting in D-Ala and D-amino butyric acid residues. The enzymes carrying out this reaction are encoded in the cognate BGCs and fall into four dehydrogenase families generically termed $LanJ_{(A\text{-}D)}$ (Figure 7). Members of the $LanJ_A$ family are zinc-dependent dehydrogenases. The protein $LtnJ_A$ involved in lacticin 3147 biosynthesis was the first functionally identified member of this group of reductases (Cotter *et al*., 2005). Other $LanJ_A$ reductases are PenN and SacJ for pediocin A and staphylococcin C55 biosynthesis, respectively (Cotter *et al*., 2005). $LanJ_B$ proteins belong to the flavin oxidoreductase superfamily (Yang & van der Donk, 2015). Identified members of this subgroup are CrnJ in carnolysin, and LasN in lactocin S biosynthesis (Lohans *et al*., 2014). The third group of reductases is represented by $BrtJ_C$ from *Streptomyces rimosus* (Walker *et al*., 2020). This protein is a luciferase-like monooxygenase that converts two Dha residues into D-Ala to yield the two-component lanthipeptide birimositide (Walker *et al*., 2020). All $LanJ_{A\text{-}C}$ proteins form D-amino acids by reducing Dha/Dhb formed by LanM lanthipeptide synthetases. Conversely, the last subclass (here termed $LanJ_D$)

was found in a class V lanthipeptide BGC. LxmJ is a F_{420} H_2-dependent reductase that exclusively converts Dha into D-Ala to form lexapeptide (Xu *et al*., 2020).

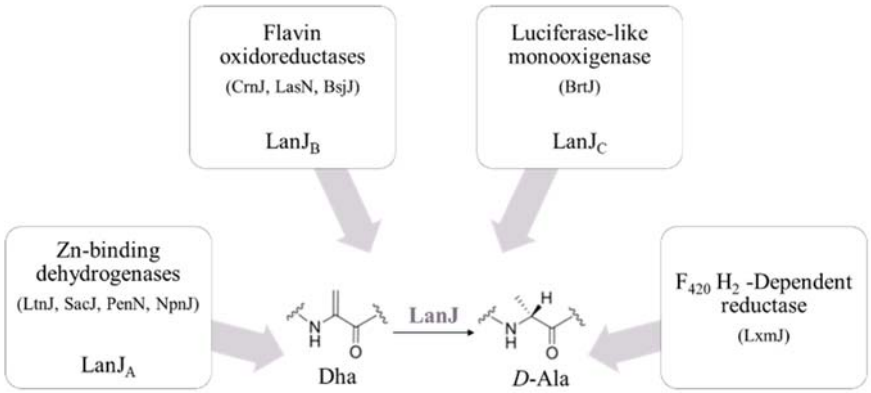


Figure 7. Classification of reductases (generically named LanJ) involved in lanthipeptide biosynthesis. Through stereoselective reactions, LanJs convert Dha and Dhb into D-Ala and D-amino butyric acid, respectively (Yang & van der Donk, 2015; Walker *et al.*, 2020; Xu *et al*., 2020).

Maturation of all five classes of lanthipeptides involves a protease-mediated cleavage between leader and core peptide (an overview of proteases is depicted in Figure 8). In many cases, the leader peptide protects the producing organism by keeping lantibiotics inactive inside the host until secretion (Chen *et al.*, 2019). Leader removal of class I lanthipeptides is typically executed by S8 subtilisin-like serine proteases termed LanPs. Classical examples of this group of enzymes are NisP and ElxP in the biosynthesis of nisin and epilancin 15x, respectively (Ortega *et al*., 2014). Some proteases for class I lanthipeptide maturation are not encoded in the cognate BGCs. Correlational analyses and integration of publicly available transcriptomic data revealed unclustered lanthipeptide protease-encoding genes (Xue *et al*., 2022). The paenilan BGC from *Paenibacillus polymyxa* ATCC 842, for instance, is the representative of a group of 44 class I lanthipeptide BGCs that do not harbor any protease-encoding gene (Xue *et al*., 2022). The S8 protease Pll-prot686, encoded elsewhere in the genome of *P. polymyxa* ATCC 842, was experimentally confirmed as the protease responsible for leader removal in paenilan biosynthesis (Xue *et al*., 2022).

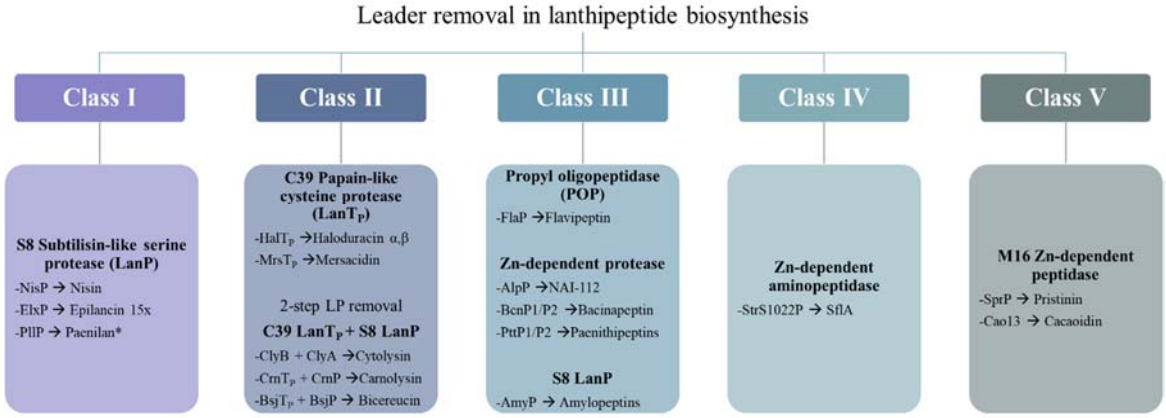


Figure 8. Leader removal in lanthipeptide biosynthesis. Depending on the lanthipeptide class, proteases belonging to different families are responsible for peptide maturation (Chen *et al.*, 2019).

For most class II lanthipeptides, leader cleavage takes place after a highly conserved Gly-Gly motif at the *C*-terminus of the LP. This step is mediated by bifunctional enzymes that contain an *N*-terminal papain-like cysteine protease domain (C39 protease) fused to an ABC transporter. These proteins are termed $LanT_P$ and are responsible for peptide maturation and export of mature products (Håvarstein *et al*., 1995). Certain class II lanthipeptides like lichenicidin, cytolysin, and carnolysin utilize a combination of $LanT_P$ and LanP proteases to complete peptide maturation. In these cases, a $LanT_P$ trims off the modified peptide during export and then an extracellular LanP protease cleaves the remaining residues to give the mature lanthipeptide (Barbosa *et al*., 2023). Three different protease systems have been described for leader removal in class III lanthipeptide biosynthesis. The propyl oligopeptidase (POP) FlaP was proved to be responsible for leader removal in flavipeptin biosynthesis by *Kribbella flavida*; FlaP is specific to the modified peptide and can discriminate between *N*- and *C*-terminal rings (Völler *et al*., 2013). The second system consists of Zn-dependent bifunctional proteases with both endo- and aminopeptidase activities. The model of this group is AplP for NAI-112 biosynthesis. AlpP works first as endopeptidase by removing a section of the LP after a highly conserved EL-Q motif, and subsequently eliminates the remaining residues through its aminopeptidase activity (Chen *et al.*, 2019). The third system was recently reported and consists of an S8 LanP protease that recognizes an EVE motif in the leader peptide for binding. The only example of LanPs involved in class III lanthipeptide maturation is AmyP in the biosynthesis of a set of seven peptides called amylopeptins (Zhang *et al*., 2022). Proteases for leader cleavage in class IV and class V lanthipeptides have been also explored, albeit to a lesser extent. For the class IV lanthipeptide SflA, a Zn-dependent aminopeptidase StrS1022P was reported from *Streptomyces* NRRL S-1022 (Ren *et al*., 2020). Finally, two members of the M16 family of peptidases SprP and Cao15 have been described for leader removal of the class V lanthipeptides pristinin and cacaoidin (Kloosterman *et al*., 2020).

During the last two decades, the number of predicted and identified lanthipeptides has tremendously increased. In 2015, Zhang and colleagues from the group of Prof. van der Donk at the University of Illinois reported a comprehensive analysis of lanthipeptide-like BGCs from *Actinobacteria*. This study suggested that lanthipeptide biosynthetic pathways are much more diverse than ever anticipated. In this large-scale study, Zhang *et al*. organized 1,163 lanthipeptide-like clusters into 100 different gene cluster families (GCFs) based on sequence and organizational similarities in entire clusters rather than specific enzymes (Zhang *et al*., 2015).
Two-component lantibiotics had been confined to *Firmicutes*; however, in this study, novel multicomponent lanthipeptide systems were found to be abundant in *Actinobacteria*. Four GCFs showed the most interesting examples (Figure 9). The GCF30 has 12 members, these clusters encode two LanMs and two precursor peptides containing the Gly-Gly/Ala motif and multiple Cys and Ser/Thr residues. With only 6 members, the GCF46 seemed to represent class III two-component systems. These clusters encode two LanA precursors containing a $S_{X2}S_{X2\text{-}5}C$ motif, and two LanKC enzymes sharing about 40% identity. Another interesting group was the GCF140 (29 members) encoding three precursor peptides, a single LanKC and a set of hypothetical enzymes.
Nevertheless, the most intriguing was the GCF146 (30 members) that was the second largest lanthipeptide-like cluster family. Members of this group encode a LanM and a LanKC lanthipeptide synthetase and two precursor peptides. The first precursor contains a strictly conserved $CSx_{1\text{-}2}Sx_{2\text{-}5}$ motif and is predicted as the LanKC substrate. A radical SAM protein is highly conserved in these clusters and was proposed to catalyze an unprecedented PTM in lanthipeptide maturation. The second precursor was suggested to be the putative LanM substrate, but it does not contain Cys residues in the core region. Its amino acid sequence resembles that of the NpnAs which are a set of six linear peptides with very conserved leader sequences but highly diverse core peptides (Zhang *et al*., 2014).

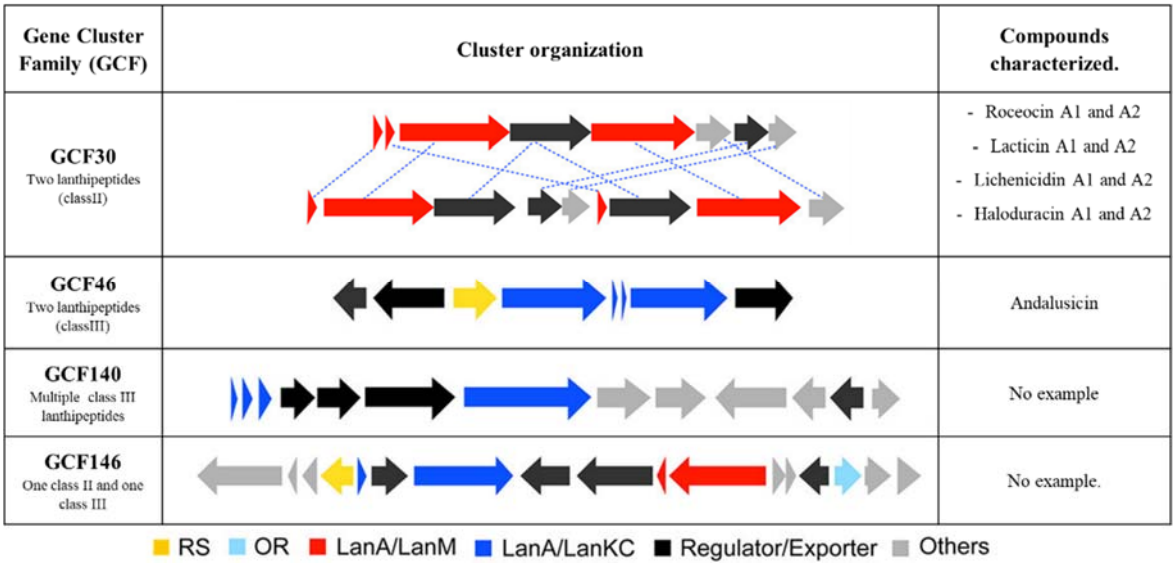


Figure 9. Organization of lanthipeptide-like gene clusters into Gene Cluster Families (GCFs) according to sequence and organizational similarities in the entire cluster. RS: radical SAM protein. OR: oxidoreductase. LanA: lanthipeptide precursor peptide. LanM: class II lanthipeptide synthetase. LanKC: class III lanthipeptide synthetase. Examples of characterized members of each family are indicated in the right column. (Adapted from Zhang *et al*., 2015).

An additional enzyme encoded in the GCF146 was hypothesized to further modify the LanM substrate. This enzyme is an oxidoreductase containing a NAD-binding Rossmann fold that may form D-amino acids by stereoselective reduction of dehydroamino acids formed by the LanM dehydratase (Zhang *et al*., 2015). Despite members of the GCF146 seem to represent a hybrid class II-class III lanthipeptide family of great importance for many Actinobacteria, to date, neither their ecological role nor the final products of this pathway have been experimentally characterized.

1.3.5 Radical SAM enzyme-catalyzed modifications in RiPP biosynthesis

In recent years, post-translational modifications mediated by radical *S*-adenosy-*L*-methionine-dependent (rSAM) enzymes have drawn special attention. In RiPP biosynthesis, reported crosslinks involve an aliphatic residue of alanine, proline, leucine or isoleucine, which are rarely observed since these residues are notoriously difficult to activate chemically or enzymatically. Due to their catalytic versatility, rSAM proteins have the potential to perform these chemical transformations (Lewis *et al*., 2021).

Members of the radical SAM superfamily of enzymes are generally identified on the basis of a highly conserved Cx_3Cx_2C motif, whose thiolates coordinate the three iron corners of a catalytic iron-sulfur (Fe-S) metallocluster. rSAM proteins achieve their complex chemistry by using a [4Fe-4S] cluster to reductively cleave SAM forming a 5′-deoxyadenosyl radical (5′-dAdo·) (Figure 10). This highly reactive radical abstracts a hydrogen from inert sp^3 carbons in the precursor peptide. Activated carbons can subsequently fuse to thiols, hydroxyls or aromatic rings, resulting in the formation of carbon-carbon bonds, carbon-sulfur bonds, carbon-oxygen bonds, epimerization, and methylation, among other unique transformations in the RiPP core peptide (Lewis *et al.*, 2021).

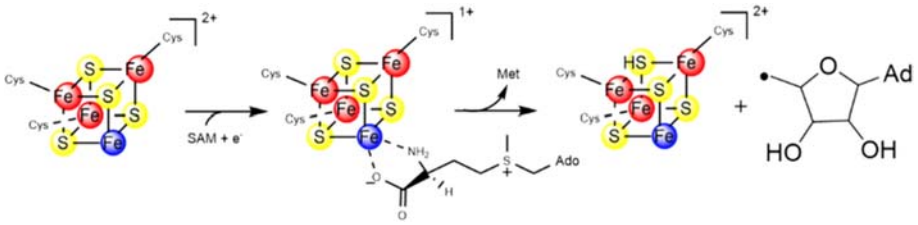


Figure 10. Structure of an [4Fe-4S] cluster found in radical SAM proteins and reductive cleavage of S-adenosyl-L-methionine (SAM). Radical SAM enzymes contain a canonical Cx_3Cx_2C motif that binds the metallocluster; the fourth iron is used to bind the SAM cofactor. When the cluster is reduced, it catalyzes the reductive cleavage of SAM giving the radical 5'-Ado• that subsequently abstracts an H atom from the cognate substrate to initiate catalysis. (Taken from https://www.cbm-lab.fr/en/Pages/Biocat/Thema_02.aspx).

In particular, rSAM/SPASM proteins (SPASM stands for subtilosin A, pyrroloquinoline quinone, anaerobic sulfatase, and mycofactocin) have come into focus as a rich source of unique RiPP maturases. rSAM/SPASMs contain the characteristic *N*-terminal [4Fe-4S] cluster common to all radical SAM enzymes, but the defining feature of this subgroup is the *C*-terminal SPASM domain that binds two auxiliary [4Fe-4S] clusters (aux I and aux II). The function of these additional clusters is not yet fully elucidated, but they are thought to play a role in peptide binding or electron transfer (Ma *et al*., 2021). In previous years, several rSAM/SPASM enzymes found in RiPP BGCs such as PqqE (Barr *et al*., 2016), MftC (Khaliullin *et al*., 2017), StrB (Davis *et al*., 2017), WgkB (Bushin *et al*., 2018), RrrB (Caruso *et al*., 2019) and DarE (Nguyen *et al*., 2022) were proven to catalyze C-C bond formation between an aromatic ring and a sp^3-carbon. Other findings included TqqB, the first enzyme described to install ether bonds giving rise to a new RiPP subfamily called rotapeptides (radical oxygen-to-alpha-carbon-linked). TqqB forms an aliphatic ether crosslink at an inactivated carbon center, linking the oxygen of a threonine side chain to the α-carbon of a Gln residue (Clark *et al*., 2019). SjiB mediates the oxidation of the Ser in the Ω1-X2-S3 motif to produce a cyclophane along with formylglycine (Fgly) and aminomalonate (Ama). SjiB exhibits remarkable promiscuity and was the first report of Ama production in a ribosomal peptide (Ma *et al*., 2021).
Recently developed bioinformatic tools have bolstered the discovery of a plethora of prospective rSAM-catalyzed modifications in RiPP biosynthesis and have unlocked hidden enzymology for peptide cyclization, thus expanding the chemical space of cyclic ribosomal peptide NPs. Examples include the rSAM/SPASM SbtM that displayed a novel mechanism for cyclization of a Cys/SeCys residue in a minimal peptide substrate leading to the formation of thiooxazoles (Lewis *et al.*, 2021). Besides, TvgB catalyzed multiple bond formations between the γ-carbon on the precursor valine giving a poly-cyclopropylglycine-containing peptide (Kostenko *et al.*, 2022). FwwB mediated the formation of multiple Phe- and Trp-derived 3-residue cyclophanes yielding the recently reported triceptides, which are characterized by three-residue cyclophanes (Phan & Morinaka, 2022). He and colleagues bioinformatically identified 32,220 prospective rSAM-catalyzed RiPP BGCs from 161,954 bacterial genomes and prioritized 25 families with new biosynthetic architectures in precursor patterns (He *et al.*, 2022). Among these, three new enzymes were found to catalyze cysteine-glycine (BlaB), histidine-aliphatic side chain (ScaB) and tyrosine/histidine-arginine (VguB) crosslinks (He *et al.*, 2022). Using a rSAM enzyme-transporter co-occurrence approach, Clark and Seyedsayamdost recently presented an Atlas of ~15,500 largely uncharacterized BGCs that organized into 800 families across different microbial phyla. To validate the utility of this massive BGC collection, they examined one family of BGCs encoding a YcaO enzyme that installs a backbone amidine into the precursor peptide, and a radical SAM maturase that forms a C-C bond between the inactivated terminal δ-methyl group of Ile and a Trp side chain (Clark & Seyedsayamdost, 2022).

2. Description of the project

In 2016, Cha *et al*. characterized the root microbiome of strawberry plants from a 15-year continuous monoculture field, which had become suppressive to *Fusarium* wilt disease. Within the isolated microorganisms, *Streptomyces griseus* S4-7 was found to be the representative strain related to *Fusarium* suppression in plants (Cha *et al*., 2016). Of note, *Fusarium spp.* are among the most devastating soilborne pathogens able to cause diseases in many economically important crops worldwide. To date, there are no effective alternatives for protecting plants after they have become infected, and the dispersion and establishment of the pathogen are particularly difficult to control as it produces chlamydospores that persist in soils for years (Burkhardt *et al*., 2019).

Investigations to explore the ability of S4-7 to suppress *Fusarium* wilt depicted a strong positive correlation between *Fusarium* growth inhibition and high expression levels of a gene encoding a class II lanthipeptide synthetase (the *lanM* gene). It was found that a minimum of seven years is required for development of specific soil suppressiveness to *Fusarium* wilt disease. Moreover, *lanM* was proposed as a diagnostic marker to determine whether a soil is conductive or suppressive (Kim *et al*., 2019).

In later studies, Kim *et al*. used a managed strawberry ecosystem to trace the tripartite mutualism between *Streptomyces* SP6C4 (a strain with a genome 99.99% identical to *S. griseus* S4-7), plants and pollinator partners (Kim *et al*., 2019 (2)). When bacteria are established in the rhizosphere, they can move upward internally as endophytes and colonize strawberry flowers and pollen (Figure 11a - b). During pollination, bacteria attach to the bees and are carried from plant to plant (Figure 11c-d). This partnership provides the potential for effective bacterial dispersion and, as the vector in this process, pollinator bees gain protection against common entomopathogens (Figure 11e). Besides, the flowers serve as the hub in the tripartite system and acquire advantages against phytopathogenic bacteria and fungi (Figure 11f). In this comprehensive study, the protective effect of SP6C4 (S4-7) for plants and pollinators was validated by determining the distribution and expression levels of *lanM* (Kim *et al*., 2019 (2)).

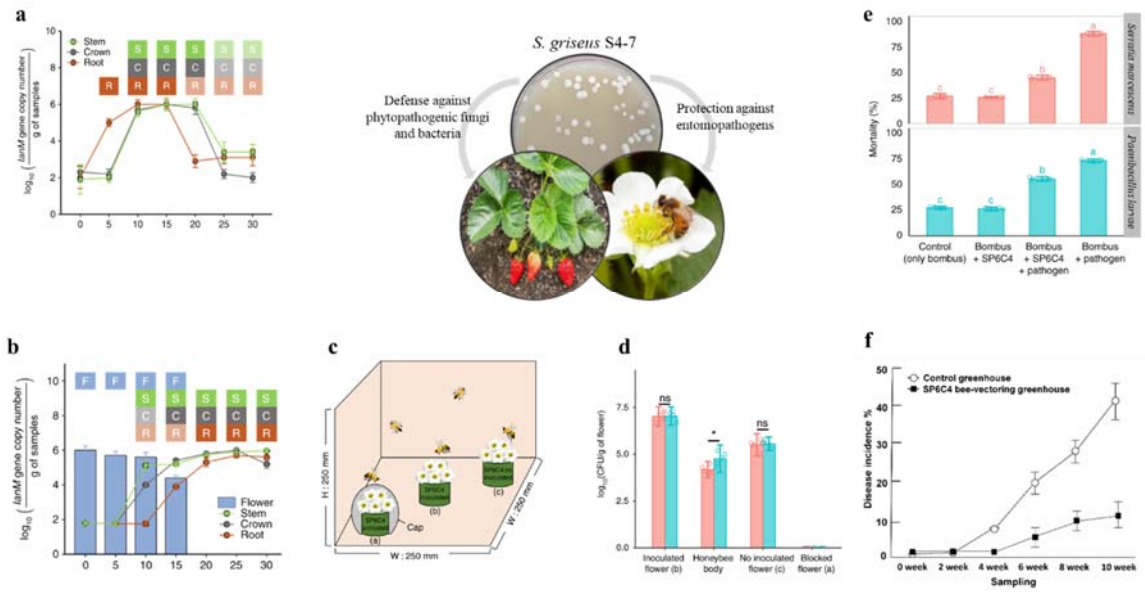


Figure 11. Mutualistic interaction between *Streptomyces* bacteria, strawberry plants and pollinating bees. The genome of *Streptomyces* SP6C4 was found to be 99.99% identical to that of *S. griseus* S4-7. **(a-b)** translocation of *S.* SP6C4 from below- to above-ground and vice versa. In **a**, bacteria were introduced onto the root surface in a suspension with 0.1% methyl cellulose. *Streptomyces* were detected with a *lanM*-specific primer using qPCR SYBR Green TOYOBO master mix. Bars represent ± SE (n = 3, 3 independent experiments). **(b)** *S.* SP6C4 introduced on the flower and detected by *lanM*-specific qPCR. Movement of the strain from above- to below-ground plant parts. Bars represent ± SE (flower: n = 5, 3 independent experiments; root and stem: n = 3, 3 independent experiments; crown: n = 3, 3 independent experiments). **(c)** Illustration of honeybee transport of *S.*

SP6C4 from plant to plant. (a): *S.* SP6C4 untreated and honeybee access blocked by a cap, (b): flower stamens and carpels inoculated with a suspension (100 µL) of *S.* SP6C4 (10^6 cfu/mL), (c): *S.* SP6C4 untreated flower ($n = 5$, 3 independent experiments). **(d)** Population densities of *Streptomyces* on flowers and honeybee bodies. Surface bacteria were detached by sonication for 30 min at 35 kHz. Treatments were analyzed by independent two sample *t*-test, $*P < 0.05$; inoculated flower: $P = 0.9025$, honeybee body: $P = 0.0212$, no inoculated flower: $P = 0.7141$. Bars represent standard deviation. **(e)** Reduction of *Bombus impatiens* mortality by *S.* SP6C4 dispersed from flower to flower. Entomopathogens (105 cfu/mL) were added with or without *S.* SP6C4 (106 cfu/mL) to pollen. Bars with different letters are significantly different according to Duncan's test and ANOVA (*Serratia marcescens*: P = 2e−16; *Paenibacillus larvae*: P = 2e−16). Bars represent standard deviation. **(f)** gray mold disease incidence in plots with *S.* SP6C4 delivered by bee-vectoring and control. Bars represent standard error. (Taken and adapted from Kim *et al*., 2019 (2)).

These studies undoubtedly displayed the ecological relevance of this *Streptomyces* strain. However, its potential as a source of novel NPs was only highlighted after genome sequencing and prediction of BGCs encoding novel metabolites. Bioinformatic analyses revealed the presence of 27 BGCs encoding, amongst others, several RiPPs. Two of these clusters drew attention as no chemical structures were connected, which opened the possibility to discover novel peptide NPs. The first cluster was predicted to encode a thiopeptide with an unusually large and unique precursor peptide sequence. Also, two peptidases with no reports in thiopeptide biosynthesis were found in the cluster. Conprimycin was the name proposed for this novel thiopeptide; in this thesis, the term "*cpr*" was preferred since the structure of the peptide has not yet been elucidated. The second BGC here named mrs_{S47} (mixed RiPP system) was studied as it contains the *lanM* gene, which implied a great relevance for the ecological role of the S4-7 strain.

To confirm the contribution of these RiPP BGCs to the antifungal activity, core biosynthetic genes were knocked out and the individual mutants were tested against *F. oxysporum*. Figure 12 shows an antagonism test where it is observed how the activity was completely lost by the mutants, which confirmed the participation of these predicted metabolites in the global anti-*Fusarium* activity, and also suggested these RiPPs inhibit *F. oxysporum* through synergistic mechanisms.

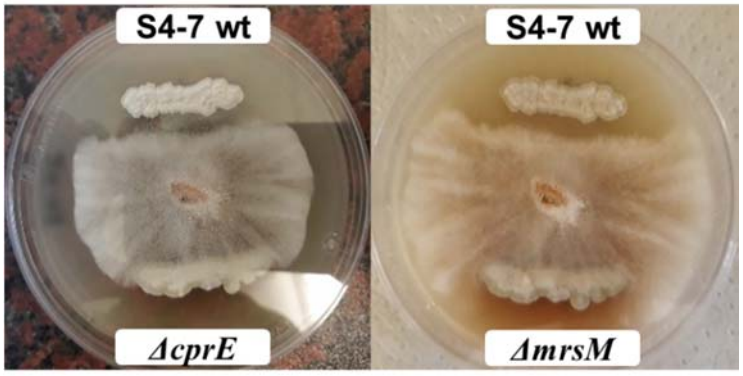


Figure 12. Antagonism of *Fusarium oxysporum* by *S. griseus* S4-7 and deletion mutants (*ΔcprE* and *ΔmrsM*). Bacteria were placed onto a PDK agar at 7 cm each other. Plates were incubated for three days at 30 ºC and then, a piece of agar containing a *F. oxysporum* inoculum was placed in the center of the plate and incubation continued for seven days at room temperature. Deletion mutants were provided by Prof. Dr. Youn-Sig Kwak.

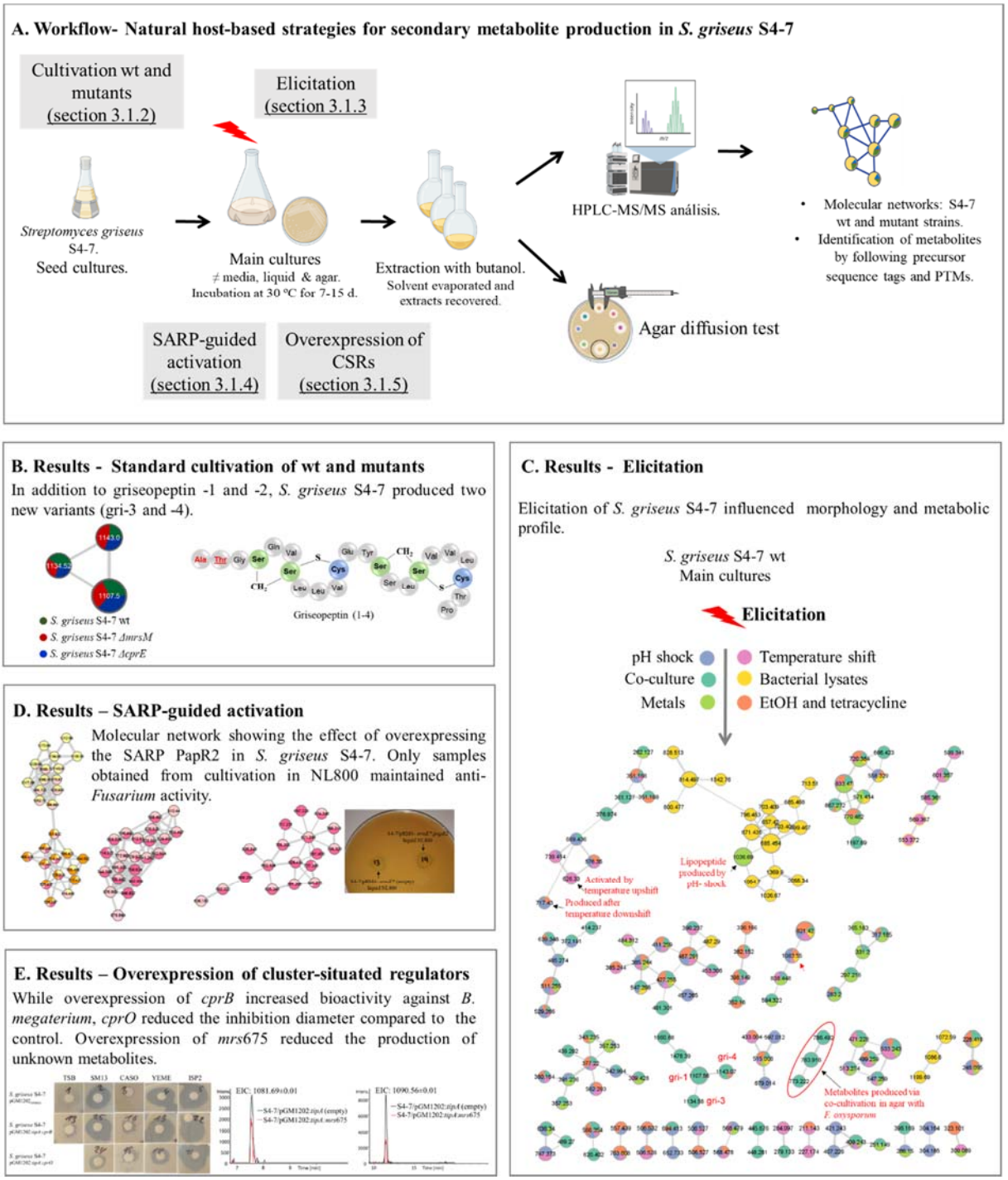
Altogether, this previous research points to the products of the *cpr* and mrs_{S47} BGCs as specialized metabolites with remarkable importance for the ecology of *S. griseus* S4-7. However, neither the structure nor the biosynthesis of these ribosomal peptide NPs had been experimentally characterized. The aim of the present thesis was thus to elucidate the structures of these compounds and to describe their biosyntheses. In this context, several approaches including traditional bacterial cultivation,

cocultivation, elicitation, regulator manipulation, heterologous expression, refactoring, and *in vitro* reconstitution of biosynthetic pathways were tested.
Section 3.1 describes some strategies applied in the wild-type S4-7 strain for production of the target RiPPs whose clusters resembled silent or down-regulated under standard cultivation. Pathway activation was addressed by targeting different levels of regulation, and results were assessed by mass spectrometry and agar diffusion tests. Given the pleiotropic effects caused by the manipulation of regulatory cascades in *Streptomyces*, the production of other secondary metabolites by S4-7 was also covered. In section 3.2, the efforts to heterologously express the *cpr* BGC are summarized. Here, several refactoring steps as well as the effect of promoters and heterologous hosts are discussed. Finally, section 3.3 describes the detailed investigation of the mrs_{S47} BGC by bioinformatics and experimental studies. Combination of bioinformatic tools allowed to establish the set of precursor peptides and biosynthetic enzymes and provided insights into the regulation of the system. Cloning, heterologous expression in *E. coli* and purification of precursor peptides and modifying enzymes for *in vitro* reconstitution of the main biosynthetic steps are described. In this last section, results of genomic context analyses of the mrs reported for the first time the existence of a widely distributed family of mixed RiPP systems in Actinobacteria.

3. Results

3.1 Activation of secondary metabolite production in *Streptomyces griseus* S4-7

Graphical abstract



3.1.1 Genomic context of *Streptomyces griseus* S4-7

The genome of *Streptomyces griseus* S4-7 (NCBI Reference Sequence: NZ_JYBE00000000.2) consists of 7,697,387 bp and has 71.5% G+C content. Analysis by antiSMASH v7.0.0β (analysis updated in January 2023) predicted 27 putative BGCs coding for known and unknown secondary metabolites. Six clusters thereof (4, 11, 14, 15, 16, 19) are predicted to encode RiPPs. Table 1 displays a summary of predicted BGCs indicating the type of metabolite and the most similar known BGC.

Table 1: Putative secondary metabolites predicted by antiSMASH in *S. griseus* S4-7.

Cluster	Type	Most similar known cluster	Type of most similar metabolite	Similarity
1	NRPS, T1PKS	herbimycin A	Polyketide	10%
2	T3PKS	alkylresorcinol	Polyketide	100%
3	Melanin	melanin	Other	100%
4	RiPP-like	streptamidine	RiPP: other	66%
5	T1PKS, NRPS	SGR PTMs	NRP + polyketide	100%
6	Terpene	Hopene	Terpene	69%
7	T1PKS, NRPS	phosphinothricintripeptide	NRP	6%
8	Siderophore	Schizokinen	Other	20%
9	Terpene	-	-	-
10	Ectoine	showdomycin	Other	35%
11	Lanthipeptide class III	AmfS/griseopeptin	RiPP	100%
12	T1PKS, NRPS-like	Enduracidin	NRP	8%
13	Butyrolactone	Kedarcidin	polyketide: iterative type I + enediyne type I	1%
14	Lanthipeptide class IV	labyrinthopeptin A2/ labyrinthopeptin A1/ labyrinthopeptin A3	RiPP	40%
15	Lassopeptide	keywimycin/SRO15-2005	RiPP	100%
16	**LAP, thiopeptide**	**this study**	**-**	**-**
17	NRP-metallophore	albomycin delta2	NRP	100%
18	Siderophore	desferriozamin B	Other	100%
19	**Lanthipeptide class III, lanthipeptide class II**	**this study**	**-**	**-**
20	Ectoine	Ectoine	Other	100%
21	Terpene	steffimycin D	polyketide: type II + saccharide: hybrid/tailoring	19%
22	Terpene	isorenieratene	Terpene	87%
23	T3PKS	Lasalocid	Polyketide	9%
24	NRP-metallophore	griseobactin	NRP	100%
25	NRPS, terpene	clipibicyclene/azabicyclene B/azabicyclene C/azabicyclene D	NRP+ polyketide	16%
26	Butyrolactone	coelimycin P1	polyketide: Modular type I	16%
27	Terpene	isorenietatene	Terpene	100%

Among the RiPP BGCs, cluster 4 matches 66% to the streptamidine BGC from *Streptomyces albidoflavus* J1074. This amidine-containing ribosomally-synthesized peptide was recently characterized as a metabolite widely distributed in nature (Russell *et al*., 2021). Cluster 11 corresponds to the class III lanthipeptide AmfS (also known as griseopeptin); this metabolite is a regulator for the onset of aerial-mycelium formation as is secreted by the AmfAB-mediated transport system to act as an extracellular peptide morphogen (Ueda *et al*., 2002). Structure and biosynthesis of AmfS have

already been characterized (Völler *et al*., 2012). Cluster 14 is 40% similar to that of labyrinthopeptins and is predicted to encode a class IV lanthipeptide with no chemical structure reported. Of note, the precursor peptide contains the sequence EQQ at the *C*-terminus; this small sequence is also contained in one precursor peptide from cluster 19. This observation was especially important to avoid misleading annotations during identification of NPs in the crude extracts. Cluster 15 corresponds to the lassopeptide keywimycin previously characterized from *Streptomyces* sp. NRRL F-5702 (Tietz *et al.*, 2017). Finally, clusters 16 and 19 encode the RiPPs of interest for this thesis. Cluster 16 encodes the cpr thiopeptide while cluster 19 encodes the mixed RiPP system; this last is predicted to yield various modified peptides belonging to different RiPP subclasses. Detailed information about these BGCs will be provided in sections 3.2 and 3.3. Other BGCs predicted by antiSMASH correspond to ectoine, butyrolactones, geosmin, siderophores, five terpenes and several NRPSs. The genomic context of *S. griseus* S4-7 was particularly useful to evaluate the effect of different cultivation conditions and to facilitate genome mining.

In the course of this project, the wild-type *S. griseus* S4-7 was subjected to different cultivation experiments whose results were evaluated by means of LC-MS/MS analyses. Ions were considered "potential candidates" when their *m/z* values indicated masses within a range from 700 to 2500 Da. These values were established based on the precursor peptide sequences, the mass of linear core peptides and the prediction of PTMs according to the biosynthetic enzymes found in the BGCs. Dehydroamino acids (Dha and Dhb), thiazoles and oxazoles were generally expected to be found in the cpr thiopeptide. For the mrs_{S47}, some PTMs like dehydroamino acids and thioether crosslinks were expected. Potential candidates were discarded when no MS/MS spectra were available (mostly due to low signal intensity) or when the MS/MS presented a poor fragmentation pattern that impeded a reliable assignment.

3.1.2 Standard cultivation and identification of metabolites

Streptomyces griseus S4-7 wild-type and two mutants, derived from the deletion of the *cprE* in cluster 16 and *mrsM* in cluster 19, were grown in different media (i. e. PDK and MS) to determine the optimal conditions for production of metabolites (section 6.2.3.1). Cultures were extracted with butanol and crude extracts were analyzed by HPLC-MS/MS (section 6.2.7.1 and 6.2.7.3). MS data were submitted to the Global Natural Product Social Molecular Networking (GNPS) workflow to create molecular networks in order to facilitate the identification of candidates (Wang *et al.*, 2016). To create a molecular network, GNPS first aligns each MS2 spectrum in a dataset to each of the others and assigns a cosine score to each combination to describe their similarity (considering precursor ions, fragment ions, and peak intensities). A cosine score of 1 indicates identical spectra and a cosine score of 0 means no similarity at all. Identical masses (same precursor ion *m/z*) are collapsed into a single node, also named consensus cluster. Structurally related molecules generate comparable MS2 spectra due to commonalities in their gas-phase chemistry and are represented by separate nodes that connect within the network via edges (Aron *et al*., 2020). The sample MS2 spectra are then compared to those in the GNPS library, which contains 13,790 MS2 spectra, to identify known molecules in the samples. Besides, DEREPLICATOR+, which is a database search tool incorporated into GNPS, allows the annotation of known metabolites in MS/MS data using *in silico* fragmentation graphs. Dereplication algorithms search for compound-spectrum matches and score them based on similarities between theoretical spectra derived from metabolites in the chemical library and experimental tandem spectra. Statistically significant matches with the highest score (against a given spectrum) are reported as a putative annotation. The minimum score to consider a metabolite-spectrum match as significant is 12 (Mohimani *et al*., 2018).

Within the hits obtained by DEREPLICATOR+ during analysis of crude extracts from S4-7, an $[M+H]^{1+}$ signal at *m/z* 482.12 was related (with a score 13) to the type I polyketide niizalactam B produced by *Streptomyces* sp. NZ-6 (Hoshino *et al*., 2015). The extracted ion chromatogram (EIC) showed a retention time of 20.4 min (data not shown). Nevertheless, none of the BGCs found in the genome of S4-7 really matched with that of niizalactam B.

On the other hand, the class III lanthipeptide encoded in cluster 11 was undoubtedly detected in all generated extracts (Figure 13a). AmfS also referred to as griseopeptin (gri-1) is produced by *S. griseus* along with its analog gri-2, which lacks the *N*-terminal Thr. The corresponding $[M+2H]^{2+}$ signals at *m/z* 1107.05 and 1056.52 for gri-1 and gri-2 were found in all samples, albeit gri-2 in very low abundance. Interestingly, two ions were networked with gri-1 (Figure 13b). MS/MS spectral analyses confirmed these signals to be new griseopeptin variants. The peak at *m/z* 1143.06 (gri-3) corresponds to a peptide containing one additional Ala at the *N*-terminus while the *m/z* 1134.57 (gri-4) presents, in addition to the Ala, one Dhb at position 2 (Figure 13c). During griseopeptin biosynthesis, the peptide undergoes a stepwise leader processing that is mediated by putative proteases yielding mixtures of differently *N*-terminal-processed class III lanthipeptides. This phenomenon has been observed in other *Streptomyces*-derived lanthipeptides like avermipeptin and erythreapeptin (Völler *et al.*, 2012). Besides, structural characterization of these variants suggested they are mixtures of peptides containing both (me)labionin and (me)lanthionine (Völler *et al*., 2012). This dual-mode cyclase activity of LanKC synthetases has implications for maturation and structure of peptides and should be considered when analyzing other class III lanthipeptides from *Streptomyces*.

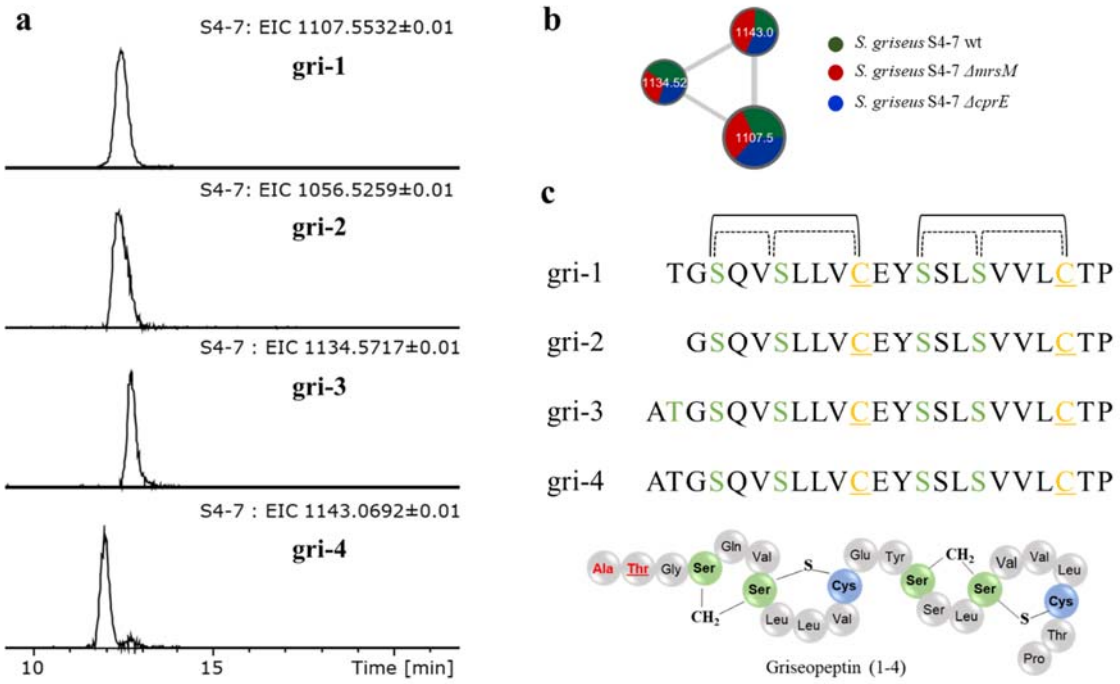


Figure 13. Production of griseopeptins (gri) in *Streptomyces griseus* S4-7. **(a)** Extracted ion chromatograms (EICs) of the four variants detected in extracts from *S. griseus* S4-7. Variants gri-3 and gri-4 are new griseopeptins found in this study. **(b)** Molecular network showing the connectivity between gri-1 and its new derivatives; green: *S. griseus* S4-7; red: Δ*mrsM*; blue: Δ*cprE*. **(c)** General peptide sequence and structure of griseopeptins; Ser/Thr residues that are dehydrated are shown in green; likely labionin crosslinks are indicated with discontinued lines.

Since griseopeptins were produced under all standard conditions, they were used as chemical markers to select the adequate solvent and to establish the optimal extraction conditions. The best peptide yields were found when samples were extracted overnight at room temperature with one volume of butanol.

Regarding the compounds of interest, chromatographic profiles of the wild type (wt), *ΔmrsM* and *ΔcprE* extracts were compared to identify potential candidates. Precursor peptides were used as reference for correlation between fragmentation patterns and amino acid sequence. For the thiopeptide, some masses in the range of 1,500- 2,500 Da, that appeared only in the wt and *ΔmrsM*, were explored. Unfortunately, peptidogenomic (Kersten *et al.*, 2011) connections of MS/MS spectra with the thiopeptide sequence were not possible because of the very low intensity and poor MS2 fragmentation. Due to the limited information about the lanthipeptides and even after comparison of the wt to the *ΔmrsM* extract, it was not possible to clearly identify modified peptides. Absence of potential candidates was in accordance with agar diffusion test results, where no activity against *Fusarium* was obtained. This was to some extent expected considering that these peptides are directly associated to the ecological role of S4-7. High expression levels for the biosynthetic genes have been observed when S4-7 was exposed to *Fusarium oxysporum* in soils (Cha *et al*., 2016; Kim *et al*., 2019), suggesting that pathway activation is mediated by external stimuli.
This evident pathway silencing emphasized the necessity to activate the biosynthesis in the natural producer and to develop heterologous systems for production of metabolites and biosynthesis investigations.

3.1.3 Elicitation of *S. griseus* S4-7 for production of antifungal RiPPs

To unlock the production of anti-*Fusarium* RiPPs in *S. griseus* S4-7, the second level of regulation was initially targeted by using an elicitation-based approach. A total of 12 different elicitors (Section 6.2.3.2) including physical, chemical and biological molecules that were successfully used in the past for cluster activation in *Streptomyces* were applied (Tomm *et al*., 2019; Abdelmohsen *et al*., 2015).
Each elicitor was treated as an independent experiment and after cultivation and elicitation, crude extracts were analyzed by HPLC-MS/MS according to section 6.2.7.3. The overall effect of each elicitor was assessed by comparison of metabolomic profiles. Molecular networks were created to facilitate the analysis and to easily identify potential changes in the S4-7 metabolome (Figure 14).
Physical elicitation was carried out using temperature and pH shock. Temperature upshifts did not alter biomass production nor morphology but a new peak at *m/z* 626.33 $[M+H]^{1+}$ was exclusively observed in this treatment. Unfortunately, the poor MS/MS fragmentation did not allow the analysis of this metabolite and no additional efforts were made as this mass was out of the target range. In contrast, temperature downshift caused an evident increase in cell adherence and some new masses were present in the organic extract. These masses were not observed in the network likely due to their low signal intensity. The corresponding MS2 spectra were manually inspected but none of them corresponded to peptide sequences and the masses were also out of the target range. Temperature-mediated regulation enhances transcription of stress response sigma factors as well as global regulators like PhoR/PhoP and other TCSs. This activation induces synthesis of heat-shock proteins which alter primary metabolism and morphological development (Servant & Mazodier, 2001). Although the identity of the produced compounds could not be established, it can be inferred that they are related to morphological differentiation.
A significant growth inhibition during pH shock did not permit the evaluation of drastic pH shifts; still cultures exposed to moderate changes were evaluated. While no alterations in the metabolic profile at pH 8.5 were observed, a new $[M+2H]^{2+}$ signal at *m/z* 1093.62 and $[M+H]^{1+}$ at *m/z* 717.43 were detected at pH 6 (Figure 14). The doubly charged ion, which was only found after manual inspection of the MS/MS data and not in the molecular network, was promising since the compounds of interest were predicted to have masses near 2,000 Da. Nevertheless, it was not possible to connect this new signal with the target RiPPs due to the lack of MS2 spectra. Interestingly, the $[M+H]^{1+}$ signal at *m/z* 717.43 was also present in the sample where cotton was added to the flask. Using cotton for cultivation has

shown upregulation of metabolites production in species that form biofilm (Tomm *et al*., 2019). According to the MS2 spectrum, this compound did not correspond to a peptide and was therefore ruled out.

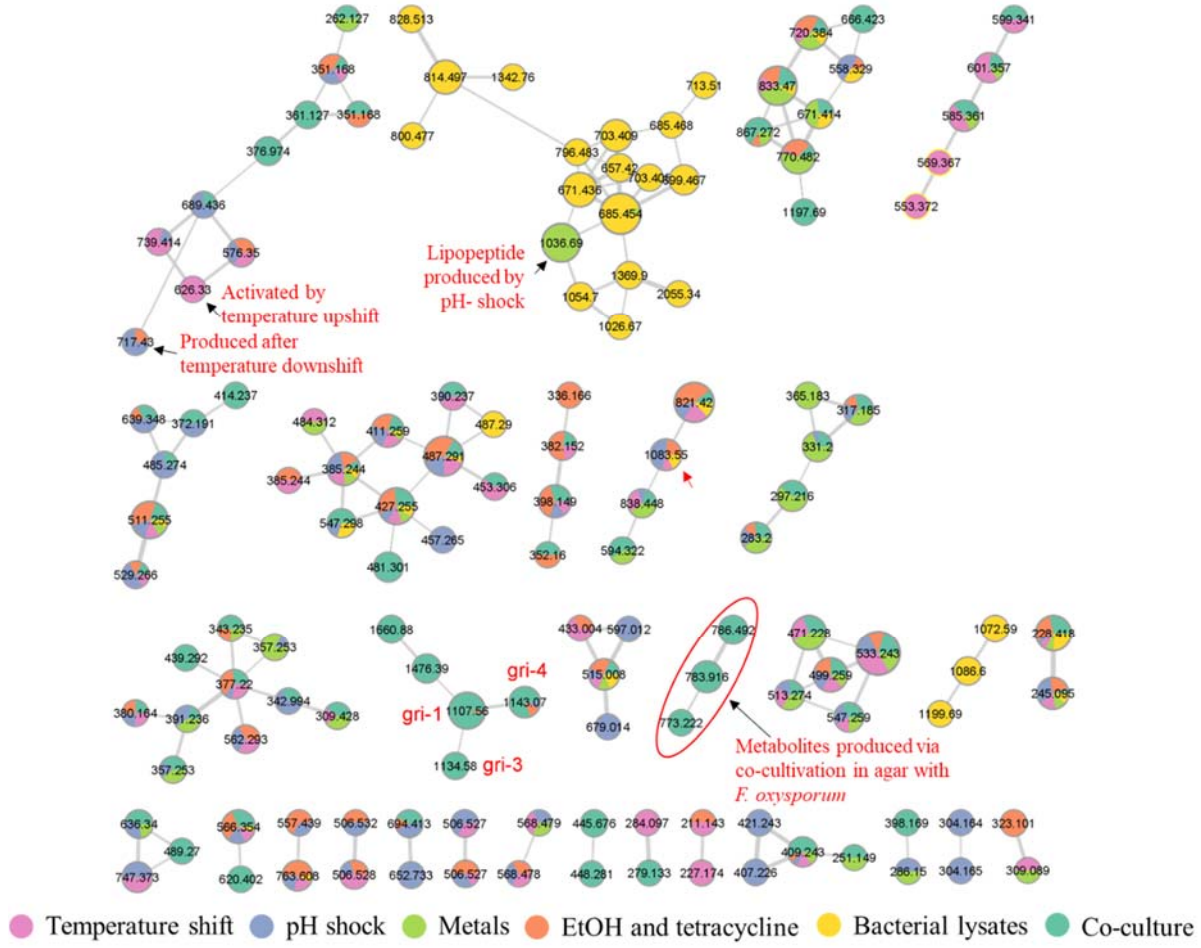


Figure 14. Effect of elicitors on the metabolite production of *S. griseus* S4-7. Twelve elicitors were tested for activation of secondary metabolite production in *S. griseus* S4-7. Elicitors were applied as independent experiments and the MS/MS data generated was grouped into six groups (indicated by a different color) to be submitted to GNPS. Temperature shift: 16 ºC and 37 ºC; pH shock: 5.0 and 8.5; metals: Zn and $LaCl_3$; bacterial lysates: *B. subtillis* and *S. coelicolor*; co-culture: with *F. oxysporum* in liquid and solid media. The molecular network was generated using GNPS according to the parameters described in section 6.2.8.1. Nodes represent the respective *m/z* values while bridges indicate their possible correlations.

Chemical elicitors were also tested for pathway activation. Supplementation with traces of zinc led to the production of a compound with a mass of 1035.68 Da $[M+H]^+$: *m/z* 1036.69. This signal was related by DEREPLICATOR+ to the lipopeptides bacircine-4 (score 20) and esperin (score 19) (Salazar *et al*., 2022). Both are members of the surfactin family and are heptapeptides with a β-hydroxyl fatty acid (C12-C16) connected with the peptide through linear branches (Figure 15). MS/MS fragmentation allowed to confirm this metabolite to be a lipopeptide and to annotate the amino acids sequence; interestingly, the first residue in the peptide from S4-7 seemed to be Asp and not Glu as in the hits. The lipopeptide signal was networked with others only present in the sample containing a lysate of *Bacillus subtilis*. This connection was reasonable since lipopeptides are reported to be produced by *Bacillus*, *Pseudomonas* and *Streptomyces* species (Raaijmakers *et al*., 2010; Zhang *et al*., 2022). The lipopeptide connected to a node at *m/z* 685.45 which was also present in the sample where $LaCl_3$ was added. The

use of rare earths like scandium and lanthanum has been shown to activate a variety of biosynthetic genes in *Streptomyces* (Tanaka *et al*., 2010). Rare earths are found in nature in small amounts, and it is conceivable that bacteria have acquired the ability to respond to low levels of these elements as a mechanism to adapt their physiology to the environment. Other small molecules tested were ethanol and nanomolar concentrations of antibiotics. The use of tetracycline, for example, led to the production of a peak at *m/z* 1083.54 (indicated with a red arrow). This signal was also detected, albeit at a lower intensity, after use of other elicitors like ethanol, lanthanum, temperature shift and pH 6.0. The nature of this compound was not further explored because of its low production and poor MS/MS fragmentation.

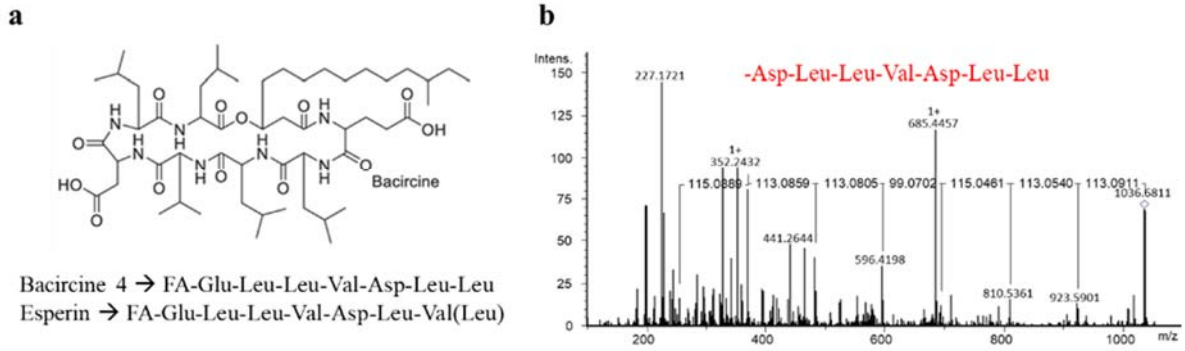


Figure 15. Lipopeptide produced by *S. griseus* S4-7 after elicitation with zinc traces. **(a)** Structure of the lipopeptides bacircine and esperin identified as hits by DERREPLICATOR+. **(b)** MS/MS spectrum of a metabolite with *m/z* 1036.69 detected in the crude extract of *S. griseus* S4-7 after elicitation with zinc traces delineates a peptide sequence (indicated in red).

Given the capacity of *S. griseus* S4-7 to inhibit *Fusarium* growth in plants, it was proposed that co-cultivation of *S. griseus* S4-7 and *Fusarium* could trigger the production of the antifungal peptide NPs. Co-cultivation in liquid and solid media was then tested. After three days of monoculture, the main cultures of S4-7 were inoculated with a piece of agar containing *Fusarium* mycelium and incubation continued for 7 days. Crude extracts were prepared with butanol according to section 6.2.7.1. Co-cultivation in liquid media did not produce any new peak; indeed, base peak chromatograms (BPCs) obtained were almost identical (Supplementary Figure S2 a). Conversely, co-cultivation on agar led to the production of three new peaks with retention times near to 12 minutes (Supplementary Figure S2 b). Although the masses were within the target range, detailed analysis of these signals was not possible due to the absence of MS2 spectra. Production of specialized metabolites to be used as defense weapons depends on biotic interactions in the ecological niche and is thus often difficult to mimic under laboratory conditions (Zong *et al.*, 2022). It therefore cannot be omitted that S4-7 could also need interaction with the plant to initiate the biosynthesis of the antifungal NPs.

Finally, the impact of elicitation on the production of griseopeptins was qualitatively assessed by comparing the area under the curve between the samples. Extracts obtained from monocultures of *S. griseus* S4-7 were used as reference (a value of 1 was assigned). As shown in Table 2, the use of elicitors dramatically impacted the production of griseopeptins. Physical and chemical elicitation completely abolished the formation of gri-2 and gri-3 variants and reduced by about 70% the production of gri-1 and gri-4. Co-cultivation of S4-7 with *Fusarium* in agar lead to a decrease by 10% and 15% in the production of gri-1 and gri-4 respectively. In contrast, co-cultivation in liquid medium maintained the production of all griseopeptins with a 30% increase in gri-3.

Table 2: Effect of elicitation in the production of griseopeptins from *S. griseus* S4-7. The relative production of griseopeptins after elicitation was estimated using the monoculture of the wt as reference (value 1). nd: not detected.

Elicitor	gri-1	gri-2	gri-3	gri-4
Monoculture S4-7	1.0	1.0	1.0	1.0
Cotton	nd	nd	nd	nd
pH-6.0	nd	nd	nd	nd
pH-8.5	0.15	nd	nd	0.16
37 ºC	nd	nd	nd	nd
16 ºC	0.18	nd	nd	0.28
Ethanol	0.33	nd	nd	0.14
Tetracycline	0.10	nd	nd	0.04
S. coelicolor *B. subtilis* lysate	nd	nd	nd	nd
Zinc traces	nd	nd	nd	nd
$LaCl_3$ traces	0.02	nd	nd	nd
Coculture *Fusarium oxysporum* (agar)	0.91	nd	0.30	0.85
Coculture *Fusarium oxysporum* (liquid)	1.03	1.0	1.30	1.09

Although the application of this small number of elicitors failed to activate the production of antifungal RiPPs, further analyses could be performed to identify the metabolites produced after elicitation to gain more insights into the secondary metabolism of this ecologically relevant strain.

3.1.4 Regulator-based semi-targeted activation of metabolites production

The next attempt to awake the BGCs of interest in *S. griseus* S4-7 focused on the activating nature of regulators to initiate the transcription of biosynthetic genes. SARPs are well-known activators of antibiotic biosynthesis in *Streptomyces*. SARP-encoding genes are found in clusters from various classes of metabolites like ribosomal and non-ribosomal peptides, polyketides, and β-lactams. In addition to their role as pathway-specific activators, studies have shown conserved CSRs can activate BGCs in foreign *Streptomyces* strains (Martínez-Burgo *et al.*, 2019). A successful example is the "small" SARP-type regulator PapR2 found in the pristinamycin BGC in *Streptomyces pristinaespiralis*. Overexpression of *papR2* increased the biosynthesis of pristinamycin whereas its deletion completely abolished production (Mast *et al.*, 2015). PapR2 was proven to activate biosynthetic pathways in other bacterial strains. In *Streptomyces lividans* for instance, overexpression of *papR2* led to activation of undecylprodigiosin (Red) production. Electromobility shift assays and transcriptional analysis indicated PapR2 takes over the regulatory function of RedD, which is the main SARP-type positive regulator in Red biosynthesis (Krause *et al*., 2020). Similarly, PapR2 increased the production of the nucleoside antibiotic plicacetin in the poorly studied Indonesian *Streptomyces* sp. SHP22-7. However, no direct interaction of PapR2 with the promoter region was found, and it was not possible to conclude if the enhanced production of plicacetin occurred via direct or indirect pathway activation. Additional studies concluded that the activation of plicacetin BGC resulted from multiple regulatory interactions and that PapR2 was able to overcome them all (Krause *et al*., 2020).

Based on the evidence that PapR2 can mimic the function of SARP-type regulators in several *Streptomyces* strains, it was hypothesized that PapR2 could activate SARP-containing BGCs in *S. griseus* S4-7. Indeed, both BGCs investigated in this thesis (*cpr* BGC and mrs_{S47} BGC) contain SARP-type regulators being potential candidates for PapR2-mediated activation. The thiopeptide BGC has a

unique SARP regulator (*cprB*) encoded upstream the biosynthetic genes. The HTH region in CprB has 62% identity and 70% similarity with the same domain in PapR2. On the other hand, the mrs_{S47} BGC encodes four CSRs including one SARP (*mrs*705) with an HTH region with 35% identity and 51% similarity to PapR2. To test the potential of PapR2 to activate these BGCs in *S. griseus* S4-7, a plasmid carrying *papR2* under the control of the *ermE** promoter was transferred to S4-7 via conjugation (section 6.2.1.4); then, engineered strains (pRM4-empty and pRM4- *ermE**:*papR2*) were grown in four different liquid and solid media (R5, NL800, NL300, and NL19). Crude extracts were analyzed by HPLC-MS/MS (using an Agilent instrument) and mass spectrometry data were organized into four groups to create molecular networks (Figure 16).

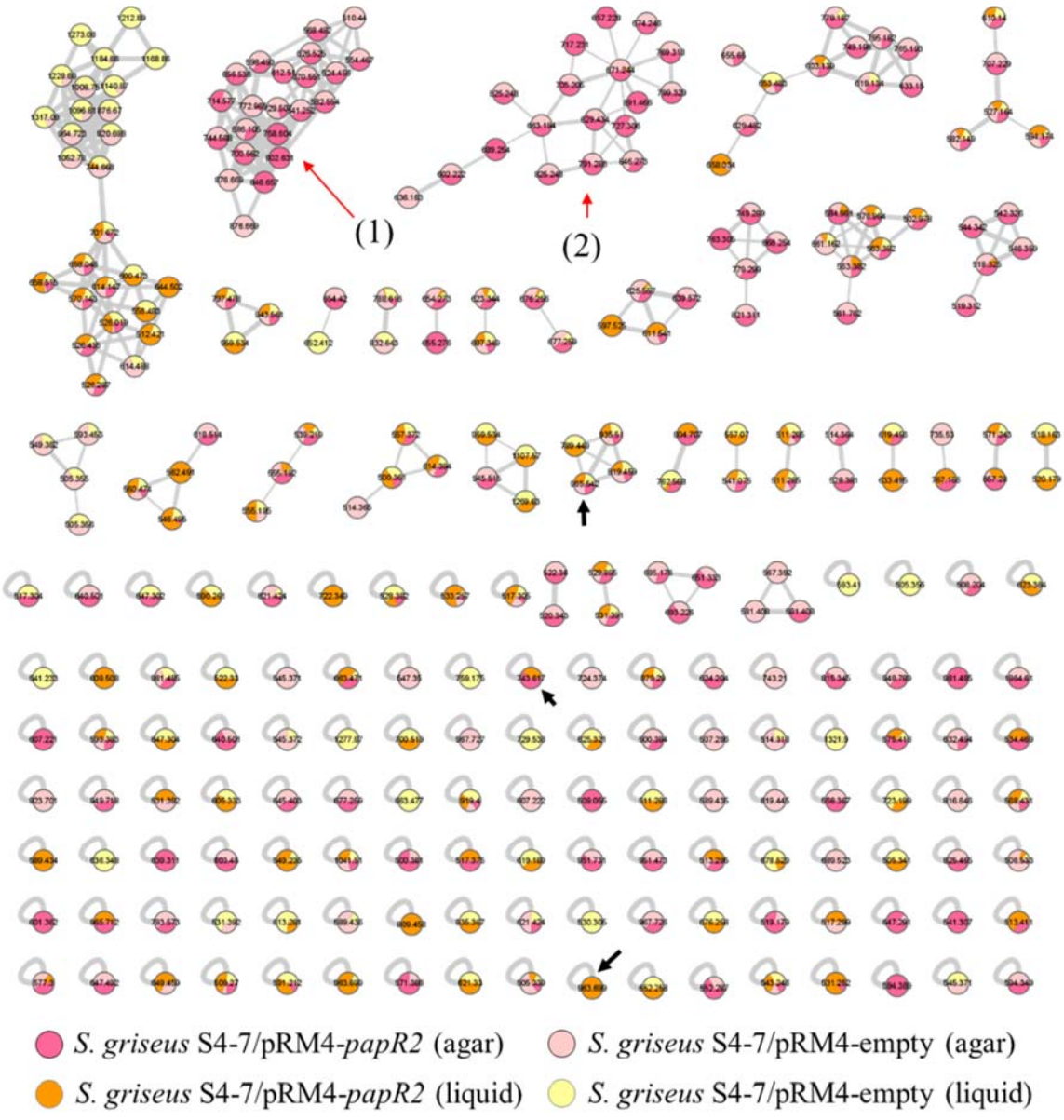


Figure 16. Overexpression of the SARP-like regulator PapR2 in *S. griseus* S4-7. A plasmid carrying the *papR2* gene under control of the constitutive promoter *ermE** was transferred to *S. griseus* S4-7 for overexpression. Four media compositions in liquid and solid cultivation were employed for cultivation. Generated MS/MS data were grouped into four groups (indicated by a different color) to be submitted to GNPS. The molecular network was generated using GNPS according to the parameters described in section 6.2.8.1. Nodes represent the respective *m/z* values while bridges indicate their possible correlations.

Networks clustering several ions in the range from 500 to 900 Da corresponded to metabolites produced through cultivation in agar (indicated with red arrows in Figure 16). One network (1) contained only signals from the cultivation in MS agar that was previously tested for standard cultivation of the wild type (section 3.1.2). The second network (2), however, clustered signals from all media except R5, which is a complex medium routinely used for protoplast regeneration.
On the other hand, ions near 1,000 Da were only observed in the control extracts in both liquid and solid cultures. Such signals were particularly found in the S4-7/pRM4-empty growth in NL300 medium. This network evidenced the effect of media composition in the chemical profile of S4-7 but also called the effect of PapR2 on the metabolism of S4-7 into question.
Promising ions (*m/z* at 965.71, *m/z* at 743.61 and *m/z* at 963.69, indicated with black arrows in Figure 16) were found in the PapR2 overexpression in NL800 medium. Chemical analysis was complemented with bioactivity assessment of the respective samples against *Fusarium oxysporum* (6.2.1.6). Interestingly, only extracts from cultivation in NL800 exhibited antifungal activity (Figure 17). To deeply explore the chemical profile of bioactive samples, the S4-7/pRM4- *ermE*:papR2* grown in NL800 medium was reanalyzed in a different high resolution HPLC-MS/MS instrument (Bruker) to allow comparison of the MS/MS data with that obtained in previous experiments.

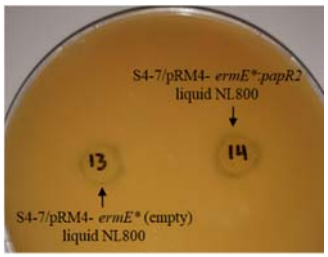


Figure 17. Agar diffusion test of selected extracts against *Fusarium oxysporum*. The different extracts prepared after overexpression of *papR2* in *S. griseus* S4-7 were analyzed (per sample and culture medium) at a concentration of 10 mg/mL. Inhibition was assessed after 7 days of incubation at 25 ºC. The mean inhibition radius was measured for each sample (disk radius: 5 mm, not included in the count). Only samples obtained from cultivation of *S. griseus* S4-7 growth in NL800 medium retained anti-*Fusarium* activity and overexpression of PapR2 slightly increased the inhibition diameter.

This second analysis of the same sample led to significant changes in the obtained data; indeed, the potential ions observed in the first measurement (performed with the Agilent instrument) were not found. Within the new signals detected in this sample were two $[M+H]^{1+}$ ions at *m/z* 969.50 and *m/z* 1039.63; both MS/MS spectra delineated peptide sequences (Supplementary Figure S3). The first ion seemed to contain dehydroamino acids (observed sequence: VDhaL(I)DhbD) while the second showed to be unmodified (AL(I)L(I)L(I)). Nevertheless, none of these sequences corresponded to the precursor peptides encoded in *cpr* BGC or mrs_{S47} BGC. It is possible that these two new signals correspond to peptide NPs connected to other BGCs such as clusters 4, 14 and 15, which also encode RiPPs. Due to the lack of correlation and the very low signal intensities, these compounds were not further investigated.
The variability of MS data obtained from two different instruments also played a role in the assessment of metabolite production after overexpression of PapR2. When extracts were measured in an Agilent instrument, known metabolites such as griseopeptins (gri-2, -3 and -4) were not detected at all. However, when the sample S4-7/pRM4- *ermE*:papR2* growth in NL800 was reanalyzed by using a Bruker instrument, all griseopeptin variants along with potential ions (within the target range) were detected. The sensitivity of an instrument directly relates to its effectiveness on producing gas-phase ions from the liquid matrix, and its capacity for transferring analytes from atmospheric pressure to the low-pressure zone of the system (Pitt, 2019). The variability in the obtained data displays the impact of

ionization and transmission efficiencies of an instrument on the data generated and highlights the importance of these parameters in the interpretation of results.

3.1.5 Pathway activation by overexpression of CSRs

The third level of regulation in secondary metabolism was lastly explored for pathway activation in *S. griseus* S4-7. For this purpose, regulatory genes found in the two biosynthetic gene clusters of interest were targeted.
Since both the *cpr* BGC and mrs_{S47} BGC were silent under standard cultivation and after elicitation, the main purpose of this approach was the induction of production of the mature metabolites for structure elucidation. Therefore, this experiment focused on those regulatory genes predicted to positively influence gene transcription and not on studying the regulatory cascade itself.
Overexpression experiments were preceded by sequence analyses of the target CSRs and their comparison with similar proteins previously characterized. In addition, precursor peptide sequences were used to guide the identification of potential metabolites or intermediates that could indicate cluster activation. The study was complemented with bioactivity testing against *Bacillus megaterium* based on the observation that the wt S4-7 extract showed to be active against this Gram-positive microorganism.

3.1.5.1 Overexpression of CSRs from the *cpr* BGC

Three genes predicted to encode CSRs are found in the thiopeptide BGC (see Figure 22). *cprB* encodes a SARP-type regulator 669 amino acids in length, flanking the biosynthetic genes. Sequence analysis of CprB shows the characteristic HTH and BTAD domains, and one AAA ATPase domain that identifies this protein as a "large" SARP regulator. The other two regulatory genes (*cprO* and *cprR*) are located at the downstream end of the cluster. CprO was predicted to be a member of the LuxR family as it contains a DNA-binding HTH domain at the *C*-terminus, and one *N*-terminal domain of unknown function probably involved in sensing external molecules. A TerR protein was encoded in *cprR*; in addition to the *N*-terminal TetR repressor domain, CprR contains a WHG domain (named after three conserved residues near the *C*-terminus of the domain) which is around 80 amino acids in length and is located at the *C*-terminus of the DNA-binding HTH domain. This part of the protein is predicted to bind unknown ligand(s) allowing a transcriptional regulation response to that molecule (Figure 18).

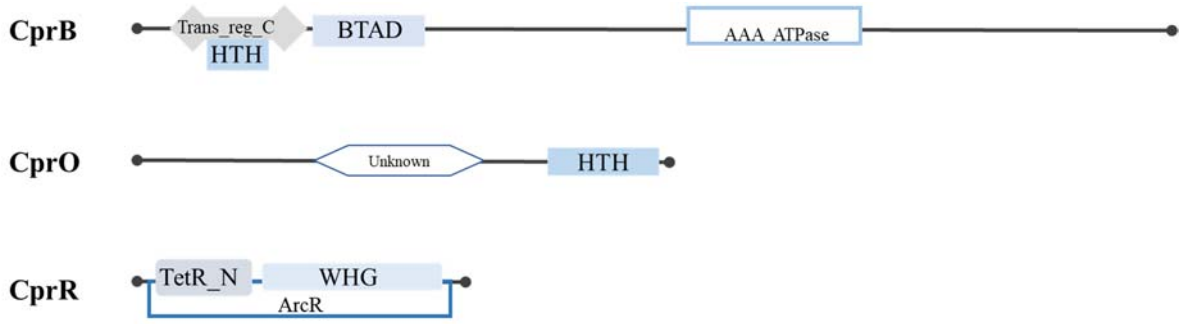


Figure 18. Cluster-situated regulators found in the *cpr* BGC from *S. griseus* S4-7. CprB is a "large" SARP regulator (NCBI protein ID: WP_030803239.1) containing an *N*-terminal HTH domain (aa 40-91), a central BTAD (aa 98-241), and an AAA ATPase (aa 353-408). CprO (NCBI protein ID: Y9K0CJJX013) is a LuxR-like protein containing a central domain of unknown function (aa 129-220) and a *C*-terminal HTH domain (aa 267-324). CprR is a member of the TetR family of repressors. It consists of the characteristic TetR (aa 18-64) domain, a *C*-terminal WHG (aa 82-184), and an ArcD spanning from residue 2 to 184.

Since SARPs are mostly activators and LuxR proteins can function as positive or negative regulators, it was first investigated whether overexpression of *cprB* or *cprO* alone could unlock the biosynthetic pathway. Individual genes were cloned into pSET-*ermE** vector to be transferred to *S. griseus* S4-7 via conjugation (section 6.2.1.4). Exconjugants were confirmed and cultivated for overexpression of the CSRs.

Engineered strains carrying empty vector and regulatory genes showed growth and morphological characteristics identical to the wild-type strain, suggesting CprB and CprO have no role in bacterial differentiation. Results showed that individual overexpression of *cprB* and *cprO* was not sufficient for pathway activation as no new signals were detected when crude extracts were analyzed by HPLC-MS/MS (data not shown).

The replicative vector pGM1202 (Addgene plasmid #69615; http://n2t.net/addgene:69615; RRID:Addgene_69615) harboring the inducible promoter *tipA* was then tested for overexpression of *cprB* and *cprO*. The new constructs (pGM1202*:tipA:cprB* and pGM1202*:tipA:cprO*) were transferred to *S. griseus* S4-7 as described in section 6.2.1.4; validated exconjugants were grown and the *tipA* promoter was induced with thiostrepton. An inducible promoter could help to establish the conditions such as inducer concentration and induction time for optimal transcription initiation. Since culture nutrients can dramatically impact metabolite production in *Streptomyces*, five different media were also used for cultivation of the engineered S4-7 strains. Crude extracts were first tested against *Bacillus megaterium* to prioritize samples with improved activity. Overexpression of *cprB* in *S. griseus* S4-7 grown in TSB, CASO, SM13 and YEME media produced larger zones of inhibition in comparison to the S4-7 harboring the empty vector. Interestingly, a decrease in inhibition zone diameter was generally observed in those samples where *cprO* was overexpressed, except for the ISP-2 sample that presented a considerably larger diameter (Figure 19). These results highlight the importance of testing different media compositions for bacteria cultivation since nutrients in the medium directly impact secondary metabolism and bioactivity.

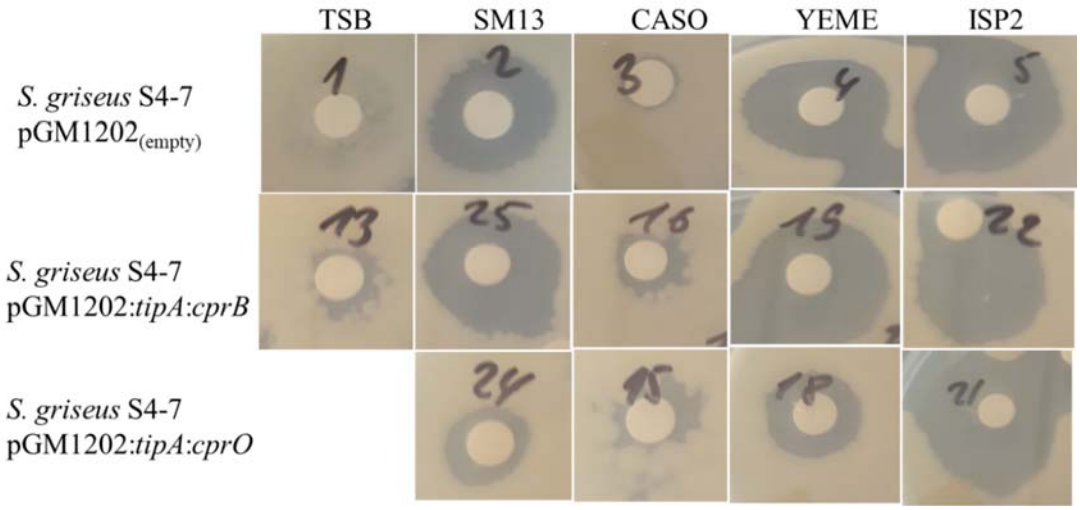


Figure 19. Agar diffusion test against *Bacillus megaterium* of crude extracts obtained from overexpression of CSRs found in the *cpr* BGC. Extracts were analyzed (per sample and culture medium) at a concentration of 10 mg/mL. The inhibition radius was measured for each sample (disk diameter: 5 mm, not included in the count). When the inhibition zone was not constant, an average between the major and minor inhibition radius was calculated.

Samples with the largest inhibition zones were selected for HPLC-MS/MS measurements to identify the associated compounds and to determine if any corresponded to the thiopeptide (Supplementary Figure S4 a). Within the promising species detected, a single charged ion at *m/z* 1827.52 was found in

the S4-7/pGM1202*:tipA:cprB* cultivated in ISP-2 medium. However, the MS/MS spectrum presented a fair fragmentation (3-5 ions per spectrum) being insufficient to certainly delineate a modified peptide. The same signal was also detected in other samples, albeit in considerably lower intensity; this ruled out any possible relation with the searched thiopeptide. Analysis of S4-7/pGM1202*:tipA:cprO* did not show repression of any metabolite that could explain the decreased bioactivity and the chromatographic profile was identical to the control (Supplementary Figure S4 b). The reduction in the activity against *B. megaterium* could have been due to partial degradation of thiostrepton used as inducer of the *tipA* promoter. This assumption was supported by an attenuated thiostrepton signal in such samples. Taken together, these experiments derived in no detection of thiopeptide production.

Biosynthesis of certain thiopeptides is mediated by a main CSR that activates the whole BGC. In *Streptomyces actuosus* for instance, the SARP-type regulator NosP promotes the transcription of biosynthetic genes resulting in nosiheptide biosynthesis. Overexpression of NosP in the native producer enhanced nosiheptide production by approximately 1.2-fold, while gene deletion abrogated biosynthesis. Structural genes in the nosiheptide BGC are organized in two divergent operons and NosP serves as the only cluster-situated activator binding to the intergeneric region of both operons. Besides, NosP senses the biosynthetic process by interactions with peptidyl and small-molecule ligands that result from parts of the leader and core peptides (Li *et al*., 2018). A similar combined mechanism has been observed in other organisms like *Nocardia* sp. where the SARP regulator NocP controls nocathiacin biosynthesis (Li *et al*., 2018). A different mechanism is that of cyclothiazomycin, a thiopeptide produced by *Streptomyces hygroscopicus* 5008. The corresponding BGC contains three regulatory genes that are within and flanking the cluster. The SHJG8833 regulator, member of the LuxR family, is crucial for cyclothiazomycin biosynthesis as it activates seven consecutive structural genes. The other two regulators SHJG8837 (member of the HTH-XRE family) and SHJG8833 (HTH-AraC family) are not critical for cyclothiazomycin production but they might have a role in the regulation of SHJG8833 (Zhang, *et al*., 2014).

Based on amino acid sequence similarity (34-36% identity) of CprB with NosP and NocP, the *cpr* BGC was initially hypothesized to be regulated in a similar manner as the nosiheptide and nocathiacin BGCs. However, the obtained results contradicted this hypothesis and showed that individual overexpression of positive CSRs is not sufficient for activation of cpr production. Biosynthesis of the thiopeptide might be strictly controlled by an interplay between activators, *cprB* and *cprO,* and repressor *cprR*. In this regard, combined approaches will be required to get more insights into the complex hierarchical signaling cascade leading to the cpr production in *S. griseus* S4-7. The most promising strategy in this context would be the overexpression of positive regulators along with deletion of the repressor. Complementary techniques like electrophoretic mobility shift assays and RT-PCR would allow the characterization of the regulatory function of CSRs in the thiopeptide from *S. griseus* S4-7.

Finally, the major constraint in this project relies on the fact that *cpr* BGC is not functional in the native producer. Therefore, the development of a heterologous system for thiopeptide production would enable structure elucidation and comprehensive investigations of the biosynthesis and its regulatory network. Some efforts conducted in this direction are described in section 3.2.

3.1.5.2 Overexpression of CSRs for activation of the mixed RiPP system (mrs_{S47})

The mrs_{S47} cluster (see Figure 28) contains four regulatory genes that were provisionally named *mrs*675, *mrs*705, *mrs*710 and *mrs*725 (Figure 20). According to sequence analyses, Mrs675 and Mrs710 are members of the LuxR family, Mrs705 is a SARP-type regulator and Mrs725 likely belongs to the LacI family.

Mrs675, a protein of 1079 amino acids in length, is unusually large and does not contain the HTH domain. It only presents an adenylate/guanylate cyclase domain with a pair of zinc ribbon domains and

a predicted ATPase domain (AAA_16). The adenylate/guanylate cyclase domain is the catalytic domain of the mononucleotidyl cyclases. Homologous cyclase domains are often coupled with several regulatory modules, granting them the ability to sense a large variety of input signals besides their role as intracellular cAMP/cGMP generators. On the other hand, Mrs710 contains an *N*-terminal P-loop AAA domain and a *C*-terminal HTH DNA-binding domain of LuxR family regulatory proteins which is hypothetically involved in the binding of the promoter sequence upstream of its target gene.
*mrs*705 encodes a protein 642 amino acids in length with a Trans_reg_C (transcriptional regulatory protein, *C* terminal) domain that includes an HTH, and a BTAD. Mrs705 carries a NB-ARC, a signaling motif found in eukaryotes and bacteria which is the regulatory domain that determines if a protein is active or not.
Finally, Mrs725 contains a LacI-type HTH domain at its *N*-terminus and a periplasmic binding protein-like (type III) domain at the *C*-terminus. Mrs725 is a member of LacI/PurR family. PurR is a global transcriptional regulator of *E. coli* that controls the biosynthesis of purines, some steps of pyrimidine biosynthesis, polyamine metabolism, and nitrogen assimilation. In purine biosynthesis, PurR functions as a repressor.

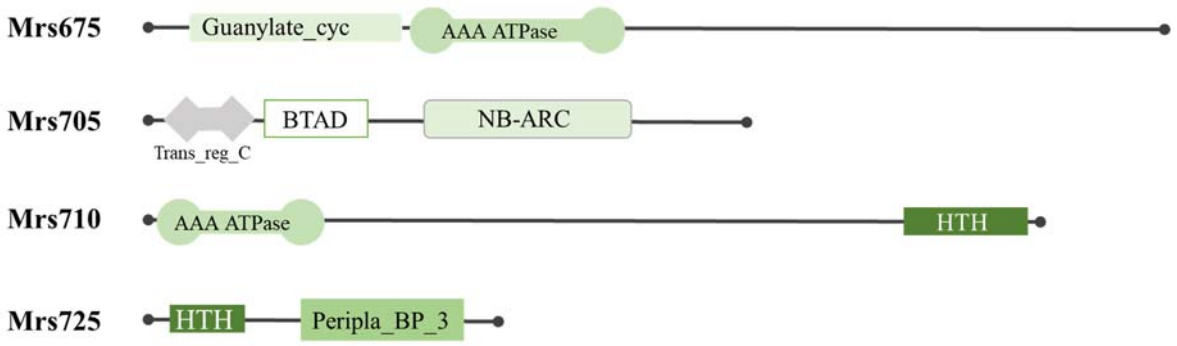


Figure 20. Cluster-situated regulators found in the *mrs*$_{S47}$ BGC from *S. griseus* S4-7. Mrs675 (NBCI Protein ID: WP_073777048.1) contains an adenylate/guanylate cyclase domain (aa 53-230) and an (AAA_16) ATPase domain (aa 248-400). Mrs705 (NBCI Protein ID: WP_044375298.1) contains a Trans_reg_C domain (aa 41-95), a BTAD domain (aa 118-263) and a NB-ARC domain (aa 307-499). Mrs710 (NBCI Protein ID: WP_030813341.1) contains an AAA_16 ATPase domain (aa 8–175) and a *C*-terminal HTH domain of the LuxR type (aa 914-967). Mrs725 (NBCI Protein ID: WP_030813350.1) contains a LacI-type domain (aa 3-49) and a periplasmic binding protein-like III domain (Peripla_BP_3, aa 188-366).

Predictions about the function of regulatory proteins in the *mrs*$_{S47}$ BGC represented a big challenge because the products of the BGC were unknown, and due to the limited information about these RiPP systems. Besides, the existence of four CSRs from different families suggested a complex network controlling the transcription of multiple operons.
First, it was investigated if any of the CSRs operated as the main pathway activator, or whether overexpression of individual proteins could lead to at least partial activation of the system. In a similar manner as for the *cpr* BGC, replicative plasmids carrying individual regulatory genes were transferred to *S. griseus* S4-7 for overexpression in different culture media (6.2.3.4). Samples were extracted with butanol and tested against *Bacillus megaterium* to identify those with increased bioactivity (section 6.2.1.6). Selected samples were analyzed by HPLC-MS/MS and metabolomic profiles were compared to the wild-type carrying the empty vector cultivated under same conditions, including induction of the *tipA* promoter by addition of thiostrepton.
In general, cloning of individual genes into the pGM1202*:tipA* vector was carried out with no major issues and the number of exconjugants in each experiment was comparable to that of the S4-7/pGM1202-empty. Thiostrepton-mediated expression did not alter morphology nor bacterial growth.

Overexpression of *mrs*675 using liquid R5 medium for cultivation yielded several ions in the extracts that had not been previously observed. Specifically, a $[M+H]^{1+}$signal at *m/z* 1081.69 drew attention, as its MS/MS fragmentation pattern showed a peptide sequence (Figure 21a). The signal was found in both *mrs*675 OE and empty vector albeit in lower abundance in the overexpression sample; this might suggest Mrs675 exerts a negative regulation in the production of this peptide. The MS/MS spectra showed a peptide sequence with no modified residues (Supplementary Figure S5). A $[M+2H]^{2+}$ signal at *m/z* 1090.56 was also detected in both *mrs*675 OE and control samples; its signal intensity was again lower in the overexpression extract (Figure 21b). For this doubly charged ion, no additional information could be collected since the MS/MS spectrum showed poor fragmentation. Although these signals could not be clearly related to the mixed RiPP system, they showed the effect of Mrs675 in the production of these metabolites.

On the other hand, crude extracts from S4-7/pGM1202:*tipA*:*mrs*705 cultivated in ISP-2 and CASO media were selected for HPLC-MS/MS analysis since they demonstrated activity against *B. megaterium*, unlike the samples harboring the empty vector. Two promising ions (*m/z* 789.73 and 985.79) were deeply analyzed as they were within the mass range and signal intensity was higher than the controls (Figure 21c-d). However, the MS2 spectra were generally poor in fragmentation, and it was not possible to connect these ions to any of the precursor peptides predicted for the mixed RiPP system.

Since Mrs705 alone was not sufficient to activate the BGC, a new plasmid harboring *mrs*705 and *mrs*710 for simultaneous overexpression of the SARP and the LuxR regulator was created. This simplified the cloning process, as both genes are encoded together, and the experiment would provide insight into any combined regulatory mechanism. In this case, liquid R5 medium was selected for cultivation of the engineered strain. The sample could not be analyzed due to time constraints and sample prioritization.

Finally, overexpression of *mrs*725 in S4-7 cultivated in TSB medium only resulted in a $[M+2H]^{2+}$ signal at *m/z* 1098.54 (Figure 21e). Its signal intensity was significantly higher than in the corresponding control. The evident variability in the production of metabolites observed in these experiments reinforces the necessity of testing different media composition for bacterial cultivation.

Results obtained so far suggested the absence of a main/partial pathway activator and promoted an interplay of regulatory proteins for the activation of the mixed RiPP system. To prove this, the cloning of a plasmid harboring all genes predicted to influence gene transcription was attempted. The new construct would form a polycistronic operon unit carrying all genes under the control of the upstream promoter thus assuring co-transcription and production of all regulatory proteins.

A two-step Gibson assembly was proposed to clone all target genes into the pGM1202 vector (section 6.2.2.6 b). Even though insertion of *mrs*710-*mrs*705 into the already obtained pGM1202:*tipA*:*mrs*725 was successful, some troubles arose afterward. In the last cloning step, *mrs*675 was cloned into the construct previously obtained; the assembly was assessed by PCR using specific primers binding the vector and the new insert, and the reaction product was used as template. Although preliminary screening resulted in the correct amplicon, no colonies were obtained after chemical transformation into *E. coli* α Select Silver. Transformation via electroporation was then attempted as this technique is more efficient when handling large constructs; this resulted however in numerous negative clones. Alternative *E. coli* strains like DH10-β and EPI300 that are suitable for transformation of large constructs, and multiple optimizations like the amount of DNA to be transformed or variations in recovery time were tried in order to obtain colonies carrying the pGM1202:*tipA*:*mrs*725*:mrs*705-710*:mrs*675 construct; but in all cases it resulted in the absence of positive clones. Several reasons could have caused this failure. Studies have shown that relatively large plasmids (6 to 16 kb) resulted toxic in electrotransformation within less than 45 minutes after the pulse (Lesueur *et al*., 2016). In other cases, exogenous DNA results toxic due either to its nucleotide sequence or its expression. Also, the lack of colonies could be attributable to leakage expression during recovery that derived in cell death. The

experimental observations suggested that leaky expression of genes contained in the pGM1202:*tipA*:*mrs*725:*mrs*705-710:*mrs*675 plasmid could take place during recovery despite the absence of inducer. This event appeared to be particularly lethal for constructs harboring the *mrs*675; indeed a reduced transformation efficiency had been observed during transformation of pGM1202:*tipA*:*mrs*675. The inability to obtain the final construct did not allow the subsequent transfer and simultaneous overexpression of all regulatory genes in *S. griseus* S4-7.

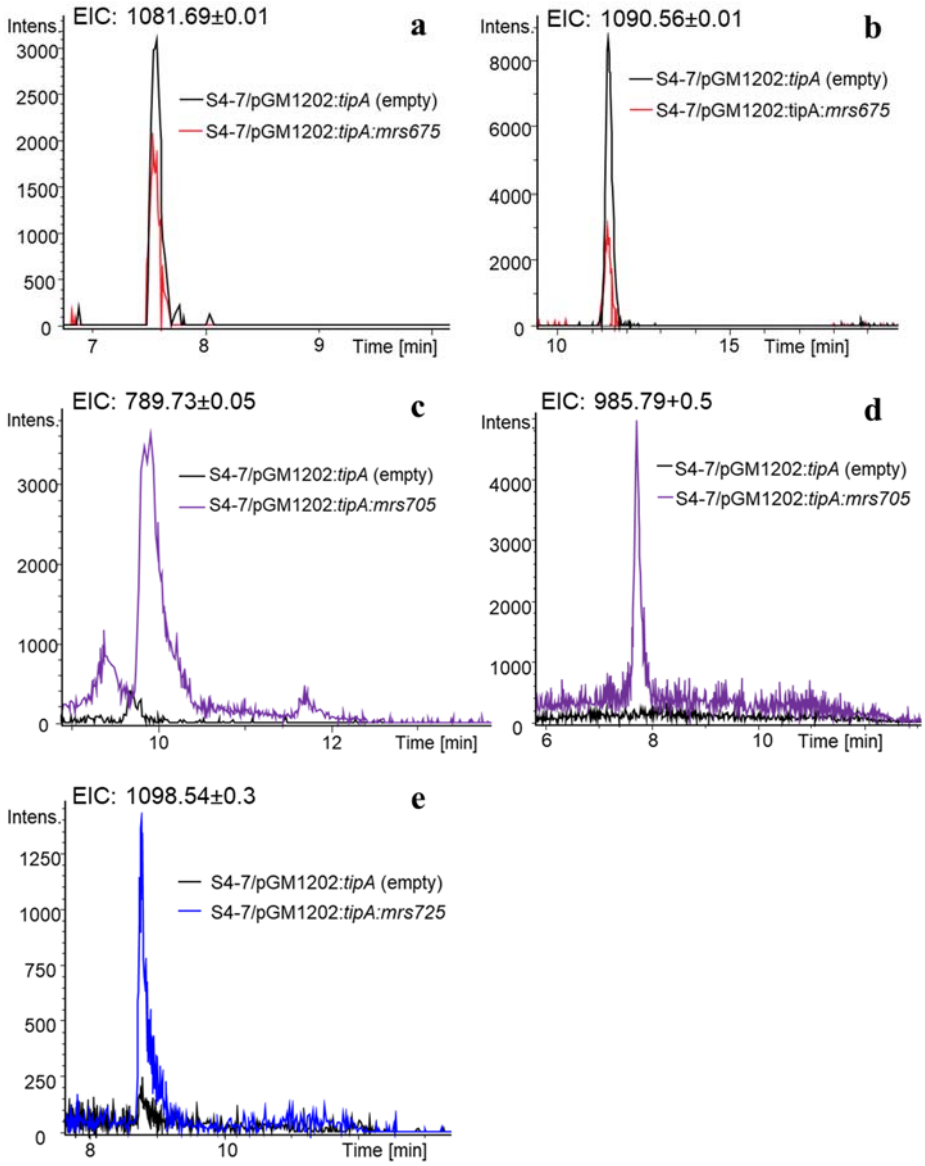


Figure 21. Overexpression of CSRs from the *mrs* BGC in *S. griseus* S4-7. **(a-b)** Extracted ion chromatograms (EICs) of two $[M+H]^+$ ions at *m/z* 1081.69 and *m/z* 1090.56 obtained from overexpression of the *mrs*675 gene. Overexpression of *mrs*675 led to a decrease in the production of these metabolites as depicted by a lower signal intensity. **(c-d)** Extracted ion chromatograms (EICs) of two $[M+2H]^{2+}$ ions at *m/z* 789.73 and *m/z* 985.79 detected after overexpression of the *mrs*705 SARP-like regulator gene. **(e)** Extracted ion chromatograms (EIC) of the $[M+2H]^{2+}$ ions at *m/z* 1098.54 obtained from overexpression of the *mrs*725 regulatory gene.

Through overexpression of individual regulatory genes, some potential ions were detected. These compounds were unfortunately produced in very tiny amounts preventing the generation of reliable MS2 spectra. These experimental issues, along with the scarce information on the PTMs expected in some of the peptides, limited the further characterization of ions. In this context, scaling up the

production of those metabolites without any clear evidence of their correspondence with the target peptides would have been inauspicious, mainly because of the time and resources required to produce significant titers of pure compounds for chemical characterization.
Despite the little results, some deductions can be made, especially regarding *mrs*675. As described above, expression levels of certain metabolites not necessarily connected to the mrs_{S47} BGC underwent a decrease when the *mrs*675 was overexpressed. This observation suggests Mrs675 exerts a negative regulation on genes implicated in the biosynthesis of other metabolites. Mrs675 is thus a clear example of regulatory proteins that interfere with the activation/repression of other clusters (cross-regulation) and reinforces the preference to use the term "cluster-situated regulator" over “pathway-specific regulator”. Also, these results prompted Mrs675 to be a key element in the regulation of mrs_{S47} BGC activation; however, the overexpression experiments conducted in this part of the project were not sufficient to shed light on the actual role of this protein.

3.1.6 Discussion and outlook

In this first part of the project, different strategies were applied to produce the ribosomal peptide NPs encoded in the *cpr* and mrs_{S47} BGCs. In fact, these metabolites must have been produced in co-cultivation with *F. oxysporum* during antagonism tests (Figure 12). However, analysis by LC-MS/MS of extracts derived from these assays did not yield clear candidates. Initial cultivation of the wild-type S4-7 in monoculture confirmed both clusters to be silent under the tested conditions (section 3.1.2). MS/MS data generated from these cultivation experiments were used to create molecular networks to further explore the metabolome of S4-7. Some ions were indeed identified by DEREPLICATOR+ as hits of already known compounds; however, none of them could be connected to BGCs predicted in the S4-7 genome. Dereplication implies the recognition and elimination of active substances already studied to prevent re-isolation of known metabolites (Mohimani *et al*., 2017). Several factors like generalized bond breakage and sequential fragmentation are crucial for successful dereplication (Mohimani *et al*., 2018).
Elicitation was tested for activation of metabolite production in *S. griseus* S4-7; however, none of the tested elicitors could activate the clusters of interest. Understanding the regulatory mechanisms that activate BGCs could be a time-consuming and laborious task. Elicitation is an easy way to trigger metabolite production in bacteria; however, this approach is very unspecific and mostly depends on trial and error. Techniques like high-throughput elicitor screening (HiTES) have been developed for rapid identification of chemical elicitors to increase success rate (Xu *et al*. 2017). In this method, bacteria are cultured on a 96-well plate and exposed to a diverse library of small molecule elicitors. Then, cultures are subjected to different assays such as mass spectrometry and bioactivity testing to identify effective elicitors (Okada & Seyedsayamdost, 2017). Silent BGCs can be targeted by inserting a reporter gene into the BGC of interest and then, the bacterial strain is screened against a library of known potential elicitors (Xu *et al*. 2017). The HiTES approach has been improved by using imaging mass spectrometry to investigate the metabolome of a given microorganism under a range of cultivation conditions. This combined method is particularly interesting for organisms producing bioactive metabolites at very low titers (Xu *et al*., 2019).
The effect of elicitation on the biosynthesis of metabolites whose role is the mediation of morphological differentiation in *S. griseus* S4-7 was displayed by griseopeptins production. Also, differences in the production of griseopeptin variants illustrated the effect of environmental conditions on the activity of biosynthetic enzymes (Table 2).
Another strategy attempted to activate the silent BGCs in *S. griseus* S4-7 consisted of overexpressing the validated SARP-like activator PapR2. Only one out of the eight culture conditions tested for overexpression of PapR2 restored the anti-*Fusarium* activity. Analysis of crude extracts by high

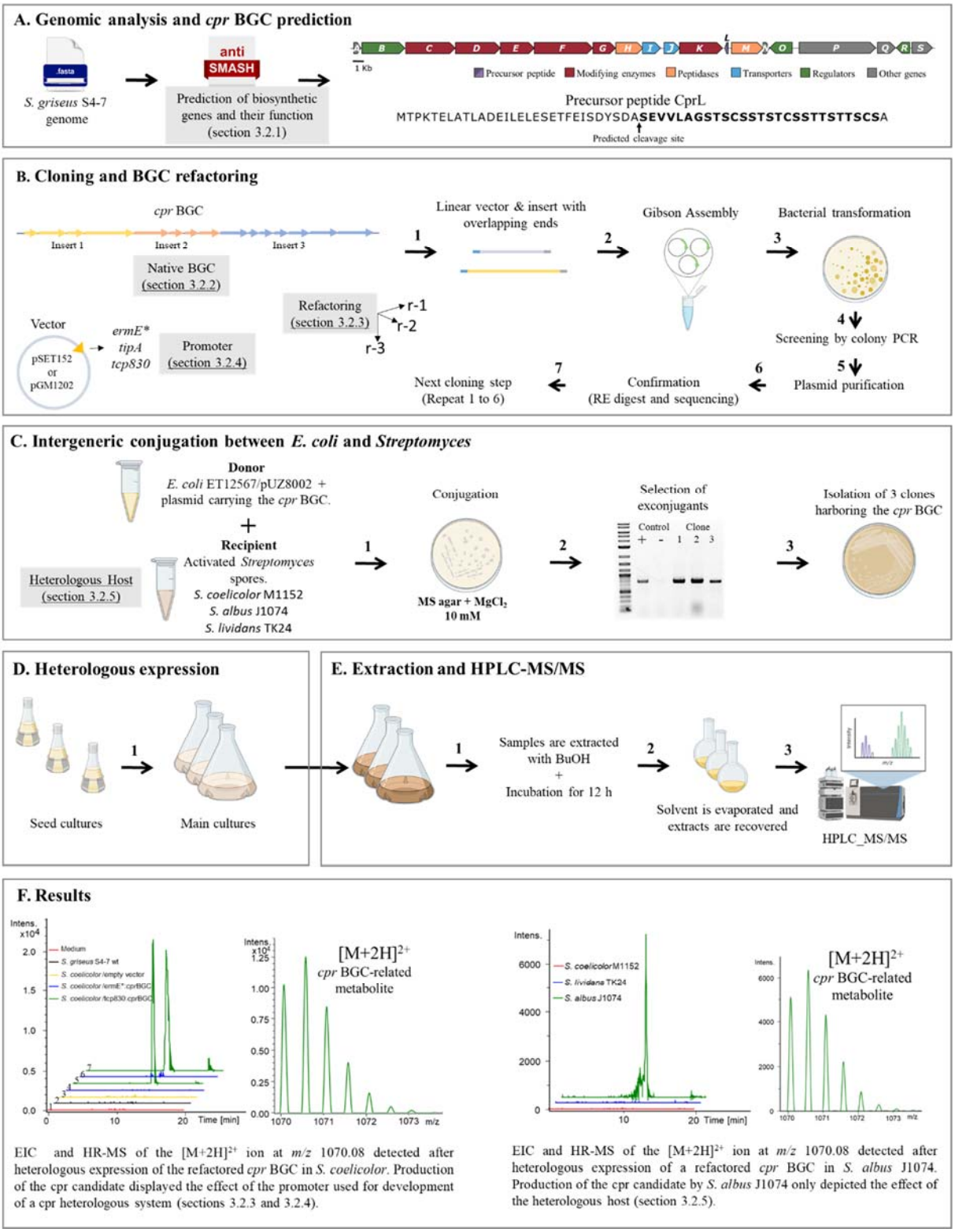
resolution LC-MS/MS detected multiple ions present only in the overexpression sample but not in the negative control. Still, the low signal intensity did not permit an adequate fragmentation of metabolites thus impeding their confirmation as potential cluster-specific metabolites. Low production of antifungals in the overexpression cultures was in accordance with the low-level production in the antagonism test (Section 3.1.4). These results supported the idea that the anti-*Fusarium* activity of *S. griseus* S4-7 results from the concerted action of several metabolites, and that the products of the *cpr* and mrs_{S47} BGCs are essential to maintain the activity, but they seem to be required only in tiny amounts.
The compound-specific approach consisted of overexpression of cluster-situated regulators in *S. griseus* S4-7. Results described in section 3.1.5 demonstrated that activation of the *cpr* and mrs_{S47} BGCs do not rely on a single main activator. These clusters seem to be under control of multiple regulatory effectors and for that reason, activation of metabolite production by individual overexpression of regulatory proteins was not accomplished. Targeting CSRs is a classical strategy that has shown to be effective for production of several *Streptomyces*-derived NPs (Ren *et al*., 2017). However, the success of this strategy is completely dependent on the capacity of the CSR for overcoming the global regulatory cascade governing its activation. When a unique protein regulates a multi-operonic BGC, overexpression of such CSR is sufficient to activate or enhance metabolite production. Nonetheless, other clusters such as the *cpr* BGC may require a combination of activator overexpression and repressor deletion for the NP to be produced. Deciphering the pathway-specific regulatory network for production of cpr and the mixed RiPP system may require the implementation of other technics like gel electrophoresis mobility shift assay to detect protein-DNA complexes. This would guide the connection between coding sequences, native promoters and regulatory effectors.
Other drawbacks worth mentioning are the examination of MS/MS data obtained from crude extracts and the association of potential candidates with the cpr and mrs precursor peptides. LC-MS generally comprises the separation of individual components from a mixture followed by ionization and separation of ions based on their mass-to-charge ratio (*m/z*). Fragmentation of a molecule in a mass spectrometer depends on the stability of the ion relative to that of the fragment (Gallia *et al*., 2013). In the course of this thesis, lack of MS/MS data and poor ion fragmentation were recurrent problems that impeded the analysis of potential candidates. Main causes for unfavorable peptide fragmentation could be the insufficient proton mobility due to charge localization at strongly basic side chains, the presence of post-translational modifications, a bad ionization, and the low concentrations of metabolites (Leitner *et al*., 2007).
The effect of low signal intensity and poor MS fragmentation was exemplified during molecular networking. As shown in figure 13b for example, only three griseopeptin variants (gri-1, -3 and -4) were filtered and networked; however, manual inspection of the MS/MS data confirmed the presence of traces of gri-2. Hence, manual inspection of MS data was fundamental. Major constituents in the crude extracts, like griseopeptins, also masked low-abundant candidates and prevented their fragmentation. Treatment of extracts prior to LC-MS analysis would simplify the matrix and allow a better chromatographic separation; nonetheless, the method used for this purpose should be carefully selected as this could lead to loss of the compounds of interest.
The difficulty of analyzing RiPPs by MS/MS relies on their architectural complexity. Even though precursor peptide sequences can be used as reference, and the types of PTMs could be speculated based on the biosynthetic enzymes found in the respective BGCs, several possible PTM sites on the precursor result in numerous probable structures. In addition, the occurrence of cross-links and other modifications in the core sequences, like oxazole formation or cyclization, that destroy the peptide bond further complicate traditional tandem MS-based structure elucidation (Mohimani *et al*., 2014). Isobaric PTMs could be misidentified because of mass measurement errors that depend on the type of mass spectrometer. Amino acid substitutions can also mimic post-translational modifications when they have the mass of other modified amino acids. Methylation of Asp or Val for instance, results in a mass

increase of 14.01 Da that is isobaric with Glu and Leu/Ile, respectively. Such considerations were of particular importance as cpr and mrs precursor peptides contain those amino acid residues. Selection of suitable instrument settings for generation of reliable spectra could help to circumvent problems related to poor fragmentation patterns thus allowing for a better identification of PTMs in ribosomal peptide NPs.

Overall, these results illustrated several constraints associated with metabolite production regulation and highlighted the necessity to explore alternative approaches for production of the peptide NPs encoded in the *cpr* and *mrs*$_{S47}$ BGCs.

3.2 Cloning, refactoring and heterologous expression of the thiopeptide biosynthetic gene cluster (*cpr* BGC)

Graphical abstract



EIC and HR-MS of the $[M+2H]^{2+}$ ion at *m/z* 1070.08 detected after heterologous expression of the refactored *cpr* BGC in *S. coelicolor*. Production of the cpr candidate displayed the effect of the promoter used for development of a cpr heterologous system (sections 3.2.3 and 3.2.4).

EIC and HR-MS of the $[M+2H]^{2+}$ ion at *m/z* 1070.08 detected after heterologous expression of a refactored *cpr* BGC in *S. albus* J1074. Production of the cpr candidate by *S. albus* J1074 only depicted the effect of the heterologous host (section 3.2.5).

3.2.1 Genomic overview of the *cpr* BGC

The unique thiopeptide BGC found in the genome of *S. griseus* S4-7 is 30.8 kb long and consists of 19 open reading frames (ORFs). Six of them correspond to biosynthetic genes and two additional encode peptidases probably involved in secondary PTMs. A single precursor peptide gene along with three CSRs, two ABC-transporters and five genes of unknown function complete the cluster (Figure 22a). The role of each biosynthetic element was predicted by comparison of protein sequences to thiopeptide BGCs available at the MIBiG (Minimum Information about a Biosynthetic Gene cluster, https://mibig.secondarymetabolites.org/) repository (Terlouw *et al*., 2023). Protein domain architecture was also compared to protein databases by using BlastP.

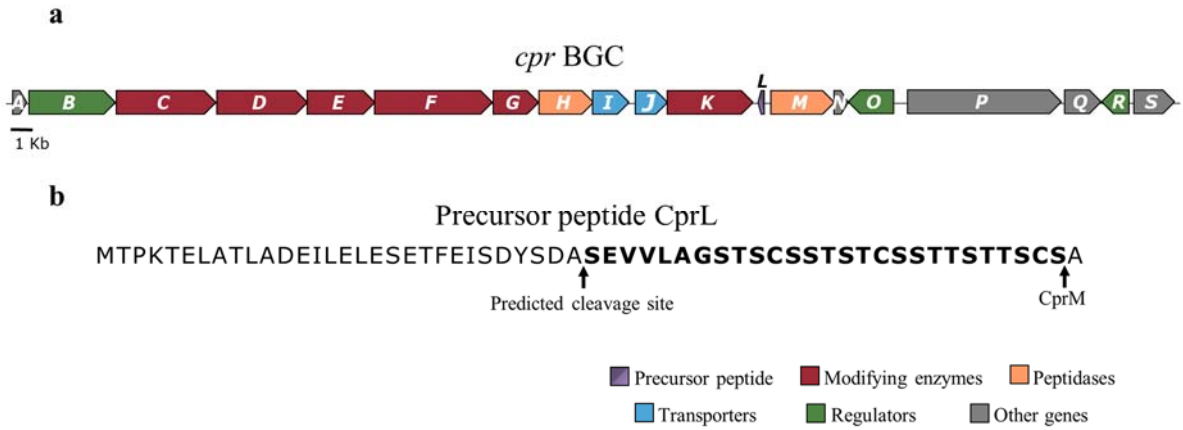


Figure 22. The *cpr* BGC from *S. griseus* S4-7 encodes the largest thiopeptide ever reported. **(a)** The *cpr* BGC contains one precursor peptide (*cprL*), six modifying enzymes, two peptidases, three CSRs, two transporters, and other proteins of unknown function. **(b)** Thiopeptide precursor peptide. Core peptide is highlighted in bold, and the hypothetical cleavage site of CprM is indicated with an arrow.

As explained in section 2, conprimycin (cpr) will be the name assigned to the final product of this BGC, hence biosynthetic genes were named *cpr*. ORFs were provisionally annotated according to the Liu nomenclature (Vinogradov & Suga, 2020) and their hypothetical function was summarized in Table 3. *cprL* encodes the precursor peptide that is 58 amino acids long. The cleavage site predicted using RiPPMiner (Agrawal *et al*., 2017) suggested that 28 amino acids constitute the core structure, which would represent the largest thiopeptide ever discovered (Figure 22b). Also, the abundance of threonines in the CP pointed out the possibility of a structure containing oxazoles. Based on the linear core peptide sequence and the maximum number of PTMs this could undergo, the mass of the mature thiopeptide was expected to be within a range from 1500 to 2500 Da.

Thiopeptide biosynthesis begins with azolines formation and their subsequent conversion into azoles. In contrast to classic thiopeptide BGCs that contain a single YcaO homolog, it was found that *cprD* and *cprK* each encode YcaO enzymes with high homology. Both proteins contain a *N*-terminal cyclodehydratase and a *C*-terminal YcaO domain; this architecture is similar to that of TbtG, TclJ and NosG in thiomuracin, thiocillin and nosiheptide BGCs, respectively. The existence of two proteins with the same apparent function, rather than inconsistent, supports the hypothesis of oxazole formation. The thiopeptide berninamycin A, for instance, contains three oxazoles whose formation involves three independent YcaO proteins that work through a more complex mechanism (Malcolmson *et al*., 2013) (Supplementary Figure S6).

Azoline formation is typically assisted by one Ocin-Thif-like protein (Dunbar *et al.*, 2015). Although *cprC* seems to encode an Ocin-Thif-like protein, it also contains a dehydratase domain which is not

found in other congeners. The FMN reductase required for conversion of azolines into azoles (thiazoles or oxazoles) is encoded in *cprE*.

The next step in thiopeptide biosynthesis is the dehydration of Ser and Thr residues. *cprF* was found to be a dehydratase glutamylation domain whose sequence was 28.0% identical (99.8% coverage) to its homolog in cyclothiazomycin biosynthesis. CprF would transfer glutamic acid to the side chains of substrate Ser and/or Thr; then, a second enzyme CprG would eliminate glutamate yielding Dha and/or Dhb. CprG is 33.0% (95.6% coverage) similar to the dehydratase elimination domain from cyclothiazomycin biosynthesis. The function of CprF and CprG could be confirmed by comparing their domain architecture with that of class I lanthipeptide synthetases as they use the same mechanism. However, it was not possible to predict the number and position of dehydrated amino acids.

Table: 3 Proposed function of open reading frames (ORFs) in the *cpr* biosynthetic gene cluster.

ORF	Protein size (aa)	Proposed function	Hypothetical protein	Protein homolog	Identity/ coverage (%/%)
cprA	82	-	Unknown	-	-
cprB	669	Main pathway activator	SARP-type regulator	NosP	36/38
cprC	763	[4+2]-cycloaddition/macrocyclization and/or azoline formation	Ocin-Thif-like /dehydratase	-	-
cprD	673	Azole formation	YcaO cyclodehydratase	TbtG	36/91
cprE	528	Azole formation	Nitro-FMN reductase	-	-
cprF	918	Dehydroamino acids formation	Dehydratase glutamylation domain	CtlE	28/99
cprG	360	Dehydroamino acids formation	Dehydratase elimination domain	CtlF	33/95
cprH	378	-	Peptidase	-	-
cprI	320	Export	ABC-transporter	Cao18	46/74
cprJ	251	Export	ABC-transporter		
cprK	675	Azole formation	YcaO cyclodehydratase	TpdO	34/93
cprL	58	Thiopeptide producto	Precursor peptide	-	-
cprM	507	*C*-terminal modification	Amidohydrolase	TbtH	52/81
cprN	90	-	Unknown	-	-
cprO	330	Transcriptional regulator	Transcriptional regulator LuxR	SprR	59/100
cprP	1177	Resistance	Unknown	-	-
cprQ	96	-	Unknown	-	-
cprR	195	Transcriptional repressor	Transcriptional regulator TetR	-	-
cprS	289	-	Unknown	-	-

Protein homologs were selected from RiPP BGCs available at the MIBiG database (https://mibig.secondarymetabolites.org/). Nos, nosiheptide; Tbt, thiomuracin; Ctl, cyclothiazomycin; Cao, cacaoidin; Tpd, GE2270 thiopeptide; Spr, pristinin.

Interestingly, when a gene encoding the enzyme responsible for the six-membered ring formation was searched, none of the biosynthetic elements in the *cpr* BGC was clearly related. Nevertheless, the dehydratase domain found in CprC might perform this step as many of the known macrocyclases are indeed single dehydratases. This would of course evoke the question whether CprC participates in both steps (azoline formation and macrocyclization) and what the directionality and the correct order of PTMs would be. It is also possible that the two YcaOs do not require the auxiliary protein to form

azolines, and they rather follow a mechanism similar to that for berninamycin biosynthesis (Malcolmson *et al*., 2013).
Additionally, *cprM* encodes a peptidase 52% similar (81.7% coverage) to TbtH, which is an amidohydrolase responsible for removal of the *C*-terminal Ala in thiomuracin. The presence of *cprM* in the *cpr* BGC suggested the existence of similar secondary PTMs in the thiopeptide.
Two additional genes worth mentioning are *cprH* and *cprP*. The first encodes a protein belonging to the M50 family of peptidases. These are proteins of 700-800 amino acids in length with six *N*-terminal or central protein transmembrane segments (TMS). These integral membrane zinc metalloproteases are frequently encoded in the vicinity of ABC-type transporters in lantibiotic BGCs. Some M50 peptidases are directly related to transport since they harbor ABC-type ATPase domains. These peptidases also participate in the final step in bacteriocin biosynthesis, in which a peptidase and a transporter form a complex in the membrane to remove the leader whereas the mature molecule is being exported. A few head-to-tail cyclized bacteriocin BGCs also encode a M50 peptidase which is located next to genes encoding proteins putatively involved in transport (Major *et al*., 2021). However, there are no reports on M50 peptidases involved in thiopeptide biosynthesis. The presence of *cprH* in the cluster brings additional questions around cpr maturation. On one side, it can be hypothesized that CprC catalyzes the macrocyclization, formation of the heterocycle and as consequence, elimination of the LP. However, the presence of an Ocin-Thif-like in CprC is intriguing, as this domain is essential for leader-dependent PTMs, but not for the [4-2]-macrocyclization. On the other side, the presence of *cprH* suggests the LP is removed during export, which would imply that macrocyclization follows a novel mechanism or does not take place at all.
Based on the knowledge obtained from previous studies about the ecological role of cpr, and since the pathway is silent under laboratory cultivation, it was the question whether the *cpr* BGC would include any resistance gene. Indeed, when the *cpr* BGC was analyzed using ARTS (Antibiotic Resistant Target Seeker; https://arts.ziemertlab.com), *cprP* was annotated as a potential resistance gene. CprP contains a CarD-like TRCF (transcription-repair coupling factor) domain required for activation of light- and starvation-inducible genes, and a DEAD/DEAH box helicase domain related to RNA metabolism like nuclear transcription, pre-mRNA splicing, ribosome biogenesis, translation, etc. Still, no report on these proteins as resistance elements has been published.
Bioinformatic analysis of the *cpr* BGC provided valuable insights into cpr maturation. Nonetheless, prediction of structures and calculation of theoretical masses was still difficult and limited the search for potential candidates in the crude extracts of *S. griseus* S4-7. These limitations emphasized the necessity of developing a heterologous system for expression of the *cpr* BGC.

3.2.2 Cloning and heterologous expression of the native *cpr* BGC

The first attempt to create a heterologous system for cpr production consisted of cloning and expressing the native gene cluster. Due to the size and high GC content (30.8 kb and 74%, respectively), a two-step Gibson assembly strategy was employed (Section 6.2.4.1; Figure 54). The integrative pSET152 vector was used to clone a first insert (*cprABCDE*) to be under the control of the *ermE** promoter. Then, the obtained construct was linearized by restriction digest and used as backbone for the second step where *cprFGHIJKL* was cloned. Despite the challenge that amplification of large DNA fragments represented, both cloning steps were overall successful (Figure 23).

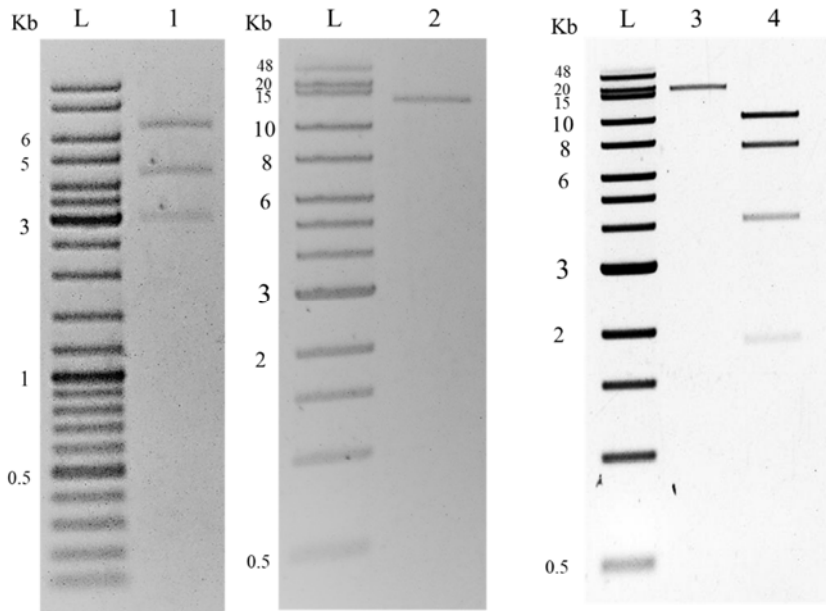


Figure 23. Control digest of cloned constructs to confirm the insertion of the native *cpr* BGC into the pSET-*ermE** vector. The first cloning step introduced the fragment *cprABCDE*. **(1)** pSET:*ermE**:*cprABCDE* double digested with *Eco*RI and *Not*I (expected bands: 6,641, 4,464 and 3,067 bp), and **(2)** linearized with *Pme*I (expected bands: 14 349 bp). The second cloning step added *cprFGHIJKL* to yield the final construct. **(3)** Linearization of pSET:*ermE*:cpr*BGC$_{(native)}$ with *Pme*I (expected band: 23,761 bp) and **(4)** digestion with *Ale*I (expected bands: 10,187, 7,516, 4,168, and 1,890 bp). L: DNA ladder. The size of the linear empty vector is 5,830 bp. In each case 10 µl of the digested plasmid were applied to 0.7% agarose gel.

The final construct (pSET:*ermE**:*cpr*BGC$_{(native)}$) was transferred by conjugation to the heterologous host *Streptomyces coelicolor* M1152. Gene transfer was confirmed by the screening of exconjugants through colony PCR using specific primers that bind at the downstream end of the cluster. It has been observed that exconjugants or transformants containing the same expression construct produce variable levels of metabolite (Gomez-Escribano & Bibb 2012). Hence, for examination of cpr production, two to three independent exconjugants were selected for cultivation. Crude extracts were prepared (section 6.2.7.1) and then analyzed by HPLC-MS/MS. Obtained data were used to create molecular networks to facilitate identification of cluster-specific metabolites. A typical molecular network contained data from the heterologous expression (two independent expression cultures), the negative control (heterologous host carrying the empty vector), and the extracted medium. MS2 spectra of those ions showing masses within the expected range were manually verified.

Results showed that the expression of the native *cpr* BGC did not yield new signals that could relate to cpr (Supplementary Figure S7). As explained in section 3.1, pathway activation in *Streptomyces* is mediated by complex regulatory networks that are not necessarily clustered with the structural genes; thus, pathway activation turns challenging when a BGC is transferred to another organism. It is estimated that only 30% of BGCs from *Streptomyces* will be successfully expressed in another *Streptomyces* host and BGC refactoring must be applied to overcome this problem (Myronovskyi, 2019).

A possible reason for the failure of this first experiment relates to the precursor peptide gene *cprL*. As seen in Figure 22a, all biosynthetic genes are encoded in the sense strand and only *cprL* is found in the antisense; so even if the enzymatic machinery was expressed, *cprL* may not have been transcribed as it was not under direct control of *ermE**. Another reason could have been the absence of *cprM* that was predicted as the peptidase responsible for secondary PTMs, thus being essential for cpr maturation.

3.2.3 Refactoring of the *cpr* BGC

Refactoring implies the genetic manipulation of an entire BGC by decoupling its expression from the native regulatory network (Li *et al*., 2021). "Major refactoring" means the substitution of regulatory elements such as promoters, ribosome binding sites (RBSs) and operators by synthetic elements with a predictable expression behavior. This "rewriting" aims to balance gene expression so that all biosynthetic elements for producing the NP are available and sufficiently activated. In some cases, replacement, deletion or reorganization of elements within the BGC is sufficient to activate the pathway. The first *cpr* BGC refactoring consisted of relocating *cprL* in front of the *ermE** in the sense strand to ensure its transcription, and to include *cprM* in the construct. Since there were no restriction sites downstream of the *ermE** promoter for insertion of *cprL* by traditional cloning, homologous recombination based on the lambda red system was initially attempted. This system derived from the lambda red bacteriophage and can be used to make an assortment of modifications like insertion and deletion of sequences, point mutations, and modification of bacterial artificial chromosomes (BACs) and plasmid DNA (Sharan *et al.*, 2009). The thiopeptide precursor gene (*cprL*) was amplified by PCR using primers with 30 bp of homology to the target site for recombination, which was located a few base pairs downstream the RBS. The obtained amplicon was used as linear donor DNA substrate to be transformed into *E. coli* expressing the lambda red enzymes (carried into the pKD46 helper plasmid and induced with arabinose). These enzymes would then catalyze the homologous recombination of the substrate with the target DNA (pSET:*ermE*:cpr*BGC$_{(native)}$). After recombination, bacteria were grown at 42 ºC to "cure" the helper plasmid that harbored a heat sensitive origin of replication, and recovered bacteria were plated onto selective agar plates. This experiment resulted in several clones that were verified by colony PCR. However, after screening a significant number of colonies, it was not possible to obtain positive clones. The lack of an antibiotic resistance gene for selection via resistance marker, and the low efficiency of the method made it impossible to obtain colonies harboring the refactored construct. In homologous recombination, the typical frequency of recombinants is one positive clone out of 10^4 to 10^5 colonies (Sharan *et al.*, 2009).

An alternative strategy consisted of cloning *cprL* directly into the pSET152 vector backbone (giving pSET:*ermE**:*cprL*) and the subsequent insertion of the native *cpr* BGC (Section 6.2.4.2; Figure 55). This facilitated the creation and verification of the newly assembled construct as the plasmid obtained previously (section 3.2.2) was used as template. The new plasmid also included the gene encoding the peptidase CprM (Figure 24 and 25a). The refactored construct named pSET:*ermE**:*cpr*BGC$_{r\text{-}1(cprL)}$ (r-1) was then transferred to *S. coelicolor* M1152 via conjugation for heterologous expression (as described in section 6.2.1.4).

To facilitate the visualization of potential candidates, the MS data of the generated extracts was used to create a molecular network and masses within 700 and 2500 Da were filtered. Several ions indeed appeared only in the expression cultures (Supplementary Figure S8); however, they were discarded after manual inspection of MS2 spectra as none of them delineated a modified peptide or did not match with the cpr tag sequence. Since this new construct contained all the biosynthetic genes under control of the *ermE**, there was no apparent reason for this experiment to fail. Still, these unsuccessful results suggested either an imbalance in gene transcription or the presence of elements within the cluster causing repression. These elements might be intergeneric regions or even the CSRs (even though *cprR* was not included in the construct).

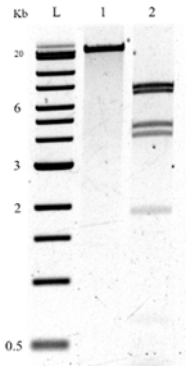


Figure 24. Control digest of refactored *cpr* BGC (r-1). The first refactoring approach consisted of inserting the precursor peptide gene downstream the *ermE** promoter to guarantee its transcription and *cprM* was added to the construct. The refactored construct pSET:*ermE*:cpr*BGC$_{r\text{-}1(cprL)}$ was linearized with *Pme*I (lane **1**; expected band: 25,174 bp) and digested with *Ale*I (lane **2**; expected bands: 7,516, 6,941, 4,482, 4,345, and 1,890 bp). In each case, 10 µl of the digested DNA were applied to the agarose gel. L: Quick-Load 1 kb Extend DNA Ladder (New England Biolabs/Ipswich, USA).

As shown in section 3.1.5, individual overexpression of positive CSRs was not sufficient to activate the pathway in the native producer. This raised the possibility that regulatory proteins negatively impact the production of the thiopeptide in the heterologous host. Hence, a second plasmid lacking regulators and intergeneric regions was constructed. Plasmid pSET:*ermE*:cprL* was taken as initial vector backbone to insert three fragments containing the biosynthetic genes by using a stepwise Gibson assembly strategy (Figure 56 in section 6.2.4.3 summarizes the step-by-step procedure). The new refactored construct (pSET:*ermE*:cpr*BGC$_{r\text{-}2(minimalBGC)}$ (r-2)) was called "minimal gene cluster" as it only contained the structural genes (Figure 25b). This minimal BGC also aimed to assure a balanced expression of all biosynthetic elements. Unfortunately, even after cultivation in different media, the *cpr* BGC$_{r\text{-}2}$ was not activated (Supplementary Figure S9). This recurrent failure suggested that rather than alterations in the BGC architecture, additional elements such as promoter and the heterologous host should be explored to enable production of cpr.

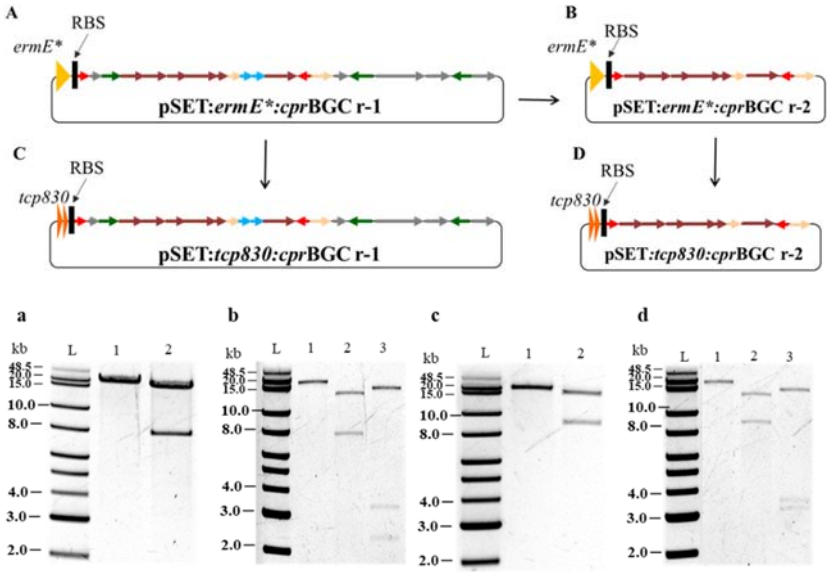


Figure 25. Control digest of refactored *cpr* BGC cloning and promoter exchange. Upper panel: **(A)** representation of the construct encoding the refactored *cpr* BGC (r-1) carrying the precursor peptide *cprL* gene (in red color) downstream the *ermE** constitutive promoter. **(B)** representation of the construct containing the *cpr* minimal cluster (r-2) under control of the *ermE** promoter. The minimal cluster was cloned to contain the precursor peptide

and biosynthetic enzyme genes only. **(C)** *cpr* BGC (r-1) under control of the inducible promoter *tcp*830. **(D)** *cpr* minimal cluster (r-2) carrying the *tcp*830 promoter. Lower panel: plasmid digest to confirm the correct insertion of the *cpr* BGC into pSET152 vector. **(a)** pSET:*ermE*:cpr*BGC$_{r\text{-}1(cprL)}$ linearized with *Pme*I (expected band: 25,174 bp) and digested with *Ale*I (expected bands: 7,516, 6,941, 4,482, 4,345, and 1,890 bp). **(b)** pSET:*ermE**:*cpr*BGC$_{r\text{-}2(\text{minimal BGC})}$ digested with *Nhe*I (expected band: 21,032 bp), *Bgl*II (expected bands: 13,338 and 7,694 bp), and *Pme*I+ *Nhe*I+*Spe*I (expected bands: 15,197, 3,411, and 2,424 bp). **(c)** pSET:*tcp*830*:cpr*BGC$_{r\text{-}1(cprL)}$ digested with *Nhe*I (expected band: 26,503 bp) and *Bgl*II (expected bands: 17,480 and 9,023 bp). **(d)** pSET:*tcp*830:*cpr*BGC$_{r\text{-}2(\text{minimal BGC})}$ digested with *Pme*I (expected band: 22,361 bp), *Bgl*II (expected bands: 13,338 and 9,023 bp) and *Pme*I+*Nhe*I+*Spe*I (expected bands: 15,197, 3,753, and 3,411 bp). L: Quick-Load 1 kb Extend DNA Ladder (New England Biolabs/Ipswich, USA).

3.2.4 Effect of the promoter on the heterologous expression of the *cpr* BGC

A promoter is a DNA sequence located directly upstream of the transcription start site (TSS) and defines the point where transcription of a gene is initiated. Bacterial promoters contain two short sequences located at the -10 and -35 positions from the TSS. While the -10 is essential for transcription initiation, the -35 box controls transcriptional rate (Myronovskyi & Luzhetskyy, 2019).

In *Streptomyces*, identification of native promoters is challenging because a whole intergeneric region spanning from 100 to 1000 base pairs can function as promoter. Therefore, when a BGC is cloned and refactored, promoter integrity is not always assured (Myronovskyi & Luzhetskyy, 2019). During refactoring, replacing native by reliable (smaller than 100 bp) synthetic promoters is one of the most suitable alternatives to guarantee transcription of the whole BGC. Promoters can be constitutive, meaning they are active *in vivo* in all circumstances, or inducible, when they require the presence of an inductor (a molecule, temperature, light, etc.) to initiate transcription.

Initial heterologous expression and refactoring of the *cpr* BGC were conducted using the *ermE** promoter. The *ermE* (originated from the promoter region of the erythromycin resistance gene of *Streptomyces erythraeus*) contains two different promoter regions, *ermEp1* and *ermEp2,* that initiate transcription at 1 and 72 bp from the translational start, respectively. The *ermE** is a stronger variant obtained by trinucleotide TGG deletion within the -35 region of *emrEp1* (Bibb *et al.*, 1985). Even though different variants of *ermE* have successfully activated metabolic pathways in *Streptomyces*, the use of *ermE** for activation of the *cpr* BGC (native, r-1 and r-2) did not lead to thiopeptide production. This strong constitutive promoter could have caused a continuous expression of the *cpr* biosynthetic genes, deriving in toxicity due to lack of peptide maturation. Another explanation for the unsuccessful expression could be that certain constitutive promoters vary under certain conditions or growth stages. It is known, for example, that *ermE* triggers a significantly stronger expression during the stationary phase. However, the expression of the majority of BGCs reaches the maximum level in the late stationary phase.

The evaluation of other promoters was suggested as an alternative for activation of the *cpr* gene cluster. In this context, inducible promoters can be suitable as they enable switching genes on and off to the desired level and specific time point. Inducible promoters contain a regulator, a promoter with an operator site for binding of the activator or repressor, and an inducer. When the inducer is absent in the medium, the repressor binds to the promoter region blocking the access of RNA polymerase thus preventing transcription initiation. Once the inducer is added, it interacts with the repressor leaving the promoter region accessible for the RNA polymerase. For activators, this binding recruits RNA polymerase to the promoter and stimulates the formation of the open RNA complex (Myronovskyi & Luzhetskyy, 2019).

A promoter tested for activation of the *cpr* BGC was *tipA* (Murakami *et al*., 1989). This promoter is derived from *Streptomyces lividans* and its induction with thiostrepton leads to the overexpression of several proteins. TipAL and TipAS are two proteins encoded in the same gene *tipA*. Both proteins bind

to thiostrepton to form a complex that activates transcription of its own gene. TipAL binds the inverted sequence located at -13 to -36 bp upstream of the TSS. Thiostrepton induces transcription of the *tipA* promoter at more than 200-fold basal level, and due to its autogenous transcriptional activation by thiopeptides, *tipA* is widely employed for heterologous expression in *Streptomyces* (Murakami *et al*., 1989).

The pGM1202 (Addgene plasmid # 69615) vector harboring the *tipA* promoter was used to clone and express the *cpr* BGC (r-2) in *S. coelicolor* M1152. Thiostrepton was added to the main cultures 24 hours after inoculation and fermentation continued for 7 days. The concentration of thiostrepton used for induction did not alter bacterial growth nor morphology. LC-MS/MS analysis of crude extracts derived from this expression experiments did not show potential cluster-related metabolites thus indicating no thiopeptide production.

Using inducible promoters to activate the *cpr* BGC could balance the gene expression within the gene cluster. This might be particularly advantageous for producing a peptide that may be toxic for the heterologous host. Another promoter tested was the tetracycline-inducible promoter *tcp*830 (Rodríguez-García *et al*., 2005). This synthetic promoter is based on the *ermEp1* and contains two operators inserted, one between the -10 and -35 and the other between the -10 and +1. The transcriptional activity of *tcp830* is dose-dependent. Inducer concentrations ranging from 1 to 100 ng/mL have been reported to successfully activate expression in *S. coelicolor* M1152, but 100 ng/mL is the concentration where the promoter has shown a full induction (Rodríguez-García *et al*., 2005). Anhydrotetracycline (a non-active derivative of tetracycline) is preferred as an inducer because it can be added at higher concentrations without inhibiting cell growth.

To test *tcp*830 for activation of the *cpr* BGC, the *ermE** was replaced with a cassette containing the *tcp*830 promoter region in both refactored versions (r-1 and r-2) of the *cpr* BGC (Section 6.2.4.4, Figure 57). After insertion and confirmation of the correct promoter assembly, the kanamycin resistance cassette (used for selection of clones harboring the *tcp*830) was removed from the construct to enable its transfer to *S. coelicolor* (Figure 25c and d). Inducer concentrations and time points were tested for expression; it was observed that higher concentrations of tetracycline slowed down cell growth. A concentration of 3 ng/mL was optimal to maintain a biomass production comparable to that observed using the *ermE* promoter. When selected crude extracts obtained from these experiments were analyzed by mass spectrometry, a new $[M+2H]^{2+}$ signal at *m/z* 1070.08 with a retention time of 11.7 minutes was detected. The signal was present in both refactored versions of the cluster after induction of the *tcp830* promoter (Figure 26) but not in the empty vector. The MS/MS spectrum of this potential candidate was studied using the cpr CP sequence to guide the annotation.

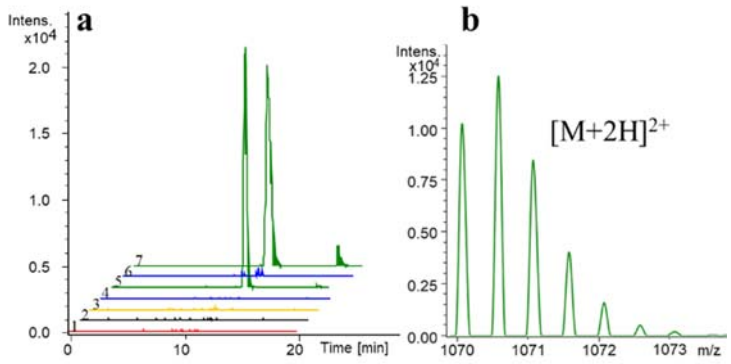


Figure 26. Heterologous expression of the refactored *cpr* BGC in *S. coelicolor* M1152 and effect of the promoter. **(a)** LC/ESI-MS EIC of the $[M+2H]^{2+}$ ion at *m/z* 1070. 08 obtained by heterologous expression of the *cpr* BGC r-1 and r-2 (blue and green, respectively). The S4-7 wild type was growth under the same conditions. Cultivation of all strains was carried out in SM13 medium. 1: crude extract obtained from the SM13 medium. 2: *S. griseus*

S4-7 wt. 3: *S. coelicolor* M1152/pSET-empty used as negative control. 4: *S. coelicolor* M1152/ pSET:*ermE**:*cpr*$BGC_{r\text{-}1(cprL)}$. 5: *S. coelicolor* M1152/ pSET:*tcp*830:*cpr*$BGC_{r\text{-}1(cprL)}$. 6: pSET:*ermE**:*cpr*$BGC_{r\text{-}2(minimal\ BGC)}$. 7: pSET:*tcp*830:*cpr*$BGC_{r\text{-}2(minimal\ BGC)}$. **(b)** HR-MS of the candidate ion observed in the expression sample after induction of the *tcp*830 promoter (sample 5 in panel a).

Thiopeptides are highly modified molecules featuring multiple oxazoles and thiazoles where the peptide bonds are used for cyclization, thus impeding the identification of modified residues by using traditional peptide sequencing. In the case of cpr, the core peptide contains the sequence EVVLAG which is expected to be intact; therefore, this sequence was generally used as tag for identifying potential cpr candidates. Unfortunately, it was not possible to connect the tag sequence to the fragmentation pattern obtained and none of the fragments could be related to hypothetical modified fragments of the peptide based on sequence and expected PTMs.

The *tcp*830 promoter has been used in the past to activate BGCs in *Streptomyces*. For example, a single *tcp*830 was sufficient to transcribe the novobiocin BGC, which is 23.4 kb and contains 20 genes for biosynthesis, regulation and resistance (Dangel *et al*., 2010). Similarly, Flinspach *et al*. increased by 4-fold the production of the thiopeptide GE2270 by inserting a copy of *tcp*830 upstream the entire gene cluster. In the same study, additional copies of *tcp*830 were introduced at different positions within the cluster, but no significant enhancement was observed. When the promoter was placed in front of the structural gene, for instance, accumulation of a GE2270 intermediate was detected as key biosynthetic genes were located upstream and were thus not expressed at a significant level to convert the excess of intermediate into the mature thiopeptide (Flinspach *et al*., 2014). Due to its small size and easy-to-control function, the *tcp*830 is a suitable tool for activation of BGCs; still a limitation lies on its poor capacity to activate clusters with multioperon structures. Then, it cannot be excluded that the compound obtained in our experiments may be an intermediate or an incomplete version of cpr lacking the tag sequence.

At this point, no further studies to confirm this new ion as the actual thiopeptide were conducted since the same signal was found in samples belonging to other projects, inferring that this signal could have resulted from contamination or carryover in the HPLC column.

3.2.5 Effect of the heterologous host on the expression of the *cpr* BGC

After testing refactored versions of the *cpr* BGC and different promoters, it was proposed that the thiopeptide could be produced and unequivocally identified by using a different host. Several *Streptomyces* strains are currently available for heterologous expression of NPs. However, their features are still suboptimal for expression of complex molecules and none of them fulfill the requirements to become a universal host. *Streptomyces coelicolor* M1152 was initially tested for production of cpr. This strain is widely used for heterologous expression since it provides low background and is suitable for conjugation and protoplast transformation. *S. coelicolor* M1152 and *S. coelicolor* M1154 are derivatives harboring deletions of actinorhodin (Act), undecylprodigiosin (Red), calcium-dependent antibiotic (CDA), and the type I modular PKS Cpk BGCs (Gomez-Escribano & Bibb, 2011). These improved strains have been successfully used for heterologous expression of RiPPs like kocurin (Linares-Otoya *et al*., 2017), GE2270 (Flinspach *et al*., 2014) and erythreapeptin (Völler *et al*., 2012). In the present thesis however, most of the expression experiments using *S. coelicolor* for production of cpr resulted unsuccessful. For that reason, *Streptomyces albus* J1074 was alternatively used. The J1074 strain is a mutant derived from *S. albus* (strain G) with the smallest chromosome and a faster developmental growth rate, a life cycle can be completed in just 4 days (Zaburannyi *et al*., 2014). *Streptomyces albus* J1074 was selected because of its extraordinary capacity as heterologous host, giving better yields than other *Streptomyces* strains. This competitive advantage is due to the existence of a second active *attB*

site within the chromosome at which integration of phiC31-based plasmid can occur (Bilyk & Luzhetskyy, 2014). This was supposed to be an advantage for the pSET-*cpr* constructs. Besides, *S. albus* J1074 has a more comparable genetic background to *S. griseus*.
Another tested host was *Streptomyces lividans* TK24. This strain can accept methylated DNA (reducing steps in the experimental workflow) and is suitable for production of recombinant proteins. Specifically, *S. lividans* TK24, which contains the mutation RpsL, improves the level of NPs production (Busche *et al.*, 2018). This variant was chosen as it presents a low protease activity, which is beneficial for heterologous expression of peptide natural products such as RiPPs.
To test the effect of heterologous hosts on the activation of cpr production, a third refactored construct (pSET:*ermE**:*cpr*BGC$_{r\text{-}3(ABC\text{-}transporters)}$) r-3 harboring two ABC transporters originally found in the vicinity of the *cpr* BGC was employed. Although promising results were obtained in the expression of the minimal cluster, the question whether the ABC transporters could enhance the pathway activation remained still unanswered. This was especially important since it was hypothesized the peak detected ($[M+H]^{2+}$ at *m/z* 1070.08) in previous experiments could correspond to an intermediate of cpr. Expression of some bioactive NPs requires specific transporters that not only mediate efflux of the mature metabolite, but also control its production by feedback or by controlling the expression of biosynthetic elements. Moreover, bioinformatic analysis of the *cpr* BGC showed some inconsistent predictions related to macrocyclization and leader removal thus being important the evaluation of the impact of native transporters.
The new construct (pSET-*ermE**-*cpr*BGC$_{r\text{-}3(ABC\text{-}transporters)}$) was transferred to *S. coelicolor* M1152, *S. lividans* TK24 and *S. albus* J1074 for heterologous expression. In the course of this experiment, a reduction in the number of ex-conjugants of *S. albus* J1074 was observed, although the transfer of the DNA was totally efficient (corroborated by screening of exconjugants by colony PCR). Surprisingly, when crude extracts obtained from these three independent experiments were analyzed by HPLC-MS/MS, the doubly charged ion at *m/z* 1070.08 in the heterologous expression in *S. albus* J1074 was detected (Figure 27). Comparison of MS/MS spectra of this peak with that obtained during previous experiments with the *tcp*830 promoter (section 3.2.4) confirmed the signals corresponded to the same chemical entity.

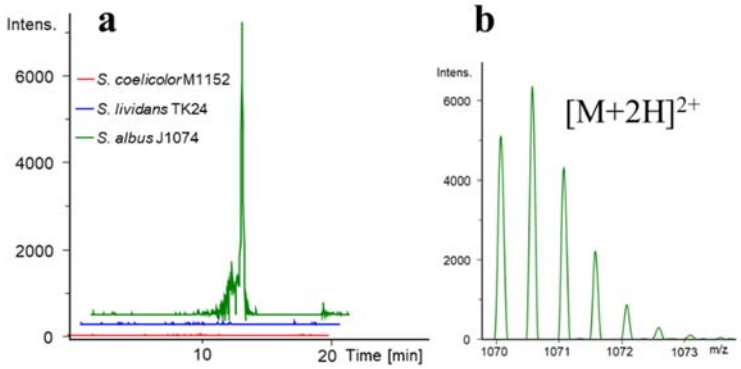


Figure 27. Effect of the heterologous host on the expression of the refactored *cpr* BGC. **(a)** LC/ESI-MS EIC of the $[M+2H]^{2+}$ ion at *m/z* 1070.08 obtained by heterologous expression of the *cpr* BGC (pSET-*ermE**-*cpr*BGC$_{r\text{-}3(ABC\text{-}transporters)}$) in three different *Streptomyces* strains. **(b)** HR-MS of the candidate ion observed in the expression sample in *S. albus* J1074. Retention time, monoisotopic mass and MS/MS fragmentation correspond to that obtained in the experiment shown in Figure 26.

The production of this compound in a different host harboring the *cpr* BGC (r-3) under control of *ermE** placed again this entity as the product of the *cpr* BGC. Due to time constraints and prioritization of other projects, no further chemical studies to confirm this ion as the thiopeptide cpr were conducted.

3.2.6 Discussion and outlook

In this project, heterologous expression in *Streptomyces* was applied to produce the thiopeptide (cpr) for further structure elucidation and biosynthetic studies. The first stage consisted of cloning the native *cpr* BGC. Gibson assembly was the *in vitro* cloning method selected. In this technique, long amplicon PCRs were used to generate BGC fragments of approximately 10 kb to be assembled into a vector. Given the high GC content found in the genomic DNA of *S. griseus* S4-7 used as template, amplification of such long DNA fragments required testing several high-fidelity DNA polymerases and optimization of thermocycling conditions. GC-rich DNA is fundamentally more stable due to stacking interactions and tends to form hairpin loops that stick around and accumulate. These secondary structures do not melt well at usual PCR denaturation temperatures and therefore require higher temperatures and the use of additives for optimal denaturation. Besides, primers used for amplification of GC-rich fragments can form homo- and hetero-dimers leading to unspecific PCR products (Green & Sambrook, 2019). The genomic DNA of S4-7 also contains large repeat sequences that could cause cross-recombination resulting in truncation of the BGC. To prevent this, a stepwise assembly of single fragments was preferred over the multi-insert assembly strategy. This allowed the verification of the correct assembly of the *cpr* BGC at each step and provided the basis for later refactoring experiments. Alternative *in vivo* methods are currently available for DNA circularization between a targeted BGC and a capture plasmid. These include the *S. cerevisiae* transformation-associated recombination (TAR)-based method where a capture plasmid with two homology arms is linearized and co-transformed with the microbial genomic DNA into the yeast for *in vivo* DNA circularization; the linear-linear homologous recombination (LLHR) method that employs a universal receiver vector containing homology arms obtained by PCR and the RE-digested genomic DNA for co-transformation into the GB05-dir *E. coli* strain (Wang *et al*., 2021); and the CAPTURE method, where the vector is split into two fragments with a *loxP* site at the ends, and the target BGC is released by Cas12 digestion of pure genomic DNA and pre-assembled *in vitro* by the T4 polymerase exo – fill-in DNA assembly strategy for subsequent transformation into *E. coli* for *in vivo* DNA circularization (Enghiad *et al.*, 2021). This last method has been tested in our laboratory for other projects, but it still requires some optimizations to achieve its successful application.

On the other hand, transformation of *E. coli* with the Gibson assembly mixture was a bottleneck in the cloning process. Despite the treatment of linearized plasmid DNA with rSAP (NEB) phosphatase to prevent religation, several negative clones were obtained during colony PCR. Screening of several colonies was then required for finding a few positive clones for plasmid isolation. Several *E. coli* strains are available for cloning. In this work, the *E. coli* DH10-β strain was preferred due to its high transformation efficiency and ability to take up and maintain large plasmids. Also, the use of highly pure DNA (insert and vector backbone) was crucial to enhance cloning efficiency.

After confirming correct assembly of the *cpr* BGC, the created plasmids were transferred to *Streptomyces* strains through biparental conjugation. Here, spores were selected as recipients as they are more efficient than mycelium. Exogenous DNA can be transferred from the donor *E. coli* to the recipient through the germ tube when spores are in the germination state after heat-shock activation (Song *et al.*, 2019).

After integration of the transferred plasmid DNA into the genome of the heterologous host, exconjugants carrying the same expression construct may present variability in the production levels of the target metabolite. This is because *Streptomyces* mycelium consists of compartments having several chromosomes, and not all of them necessarily harbor the gene cluster immediately after conjugation (Flinspach *et al*., 2010). To ensure chromosomal homogeneity, a second round of sporulation for isolation of two single colonies harboring the *cpr* BGC was performed in all heterologous expression experiments. Integration of exogenous DNA may also impact host physiology and alter DNA

replication, transcription and translation (Diaz Ricci & Hernández, 2000). Thus, a parallel experiment, where the empty vector was transferred to *Streptomyces*, was always included as negative control. This allowed for discrimination of those changes in the host metabolic profile induced by the vector itself and facilitated the detection of *cpr* BGC-specific metabolites. Analysis of the growth medium was also useful to identify components that were not completely metabolized during cultivation. All these aspects were critical for evaluating the output of the heterologous expression of the *cpr* BGC, especially because the thiopeptide was silent in the wild-type and there was no additional information besides the precursor peptide sequence that could aid the identification of the mature metabolite in the crude extracts.

Production of bacterial NPs in heterologous systems results from the co-expression of multiple genes including biosynthesis, regulation, transport, and self-resistance (Medema *et al*., 2015). An improper balance in gene transcription that resulted in poor or uncoordinated gene expression could have been the reason for the unsuccessful heterologous expression of the native *cpr* BGC. Later refactoring experiments (section 3.2.5) also failed to produce the BGC-specific compound, perhaps due to the same reason. Balanced gene expression is particularly critical for biosynthesis of a NP that requires the coordinated action of multiple processing enzymes that conduct a strictly organized cascade of PTMs. As promoters are responsible for the first stage of gene expression, they are crucial for activation and functional refactoring of silent BGCs in heterologous systems. Replacement of native transcriptionally silent promoters by well-characterized promoters provides a general strategy to disrupt native regulation to favor pathway activation. As described in section 3.2.4, the use of an inducible promoter for heterologous expression of refactored *cpr* BGC indeed led to the production of a new signal that was absent in the control samples, thus being a potential cluster-specific metabolite.

On the other hand, refactoring (section 3.2.3) and promoter replacement experiments (section 3.2.4) were achieved by Gibson assembly and traditional cloning. This was only possible because the modifications conducted in the *cpr* BGC architecture were minimal. However, more sophisticated genome engineering tools could be required in the future to truly rewrite the cpr pathway to optimize production and allow biosynthesis studies and generation of analogues. In this regard, the introduction of inducible promoters with varied strength in front of specific ORFs would be required to refactor the *cpr* BGC in a way that all biosynthetic elements can be effectively expressed at the right time point. For this purpose, advanced techniques like yeast homologous recombination (YHR) could be applied (Joska *et al.*, 2014). A YHR-based strategy could be used for simultaneous replacement of promoters with high efficiency and accuracy. Major BGC refactoring could be directed by transcriptomics; this would aid to determine the positions where promoters need to be inserted and which elements should be deleted to allow metabolite production.

Another critical aspect for activation of silent BGCs is the selection of the heterologous host. Certain organisms like *S. coelicolor* present low expression rates, in some cases even lower than the native producer. This limitation hinders the detection of target metabolites produced by heterologous BGCs. In the last set of experiments (3.2.5) here reported, a refactored construct (r-3) was expressed in three different *Streptomyces* strains. However, the cpr candidate ion was only found in the heterologous expression in *S. albus* J1074. It could be inferred that the presence of two phage attachment sites in *S. albus* J1074 allowed for higher expression of the *cpr* BGC deriving in production of the cpr candidate at a detectable level. Altogether, these results depict the importance of testing different promoters and heterologous hosts to achieve activation of silent BGCs.

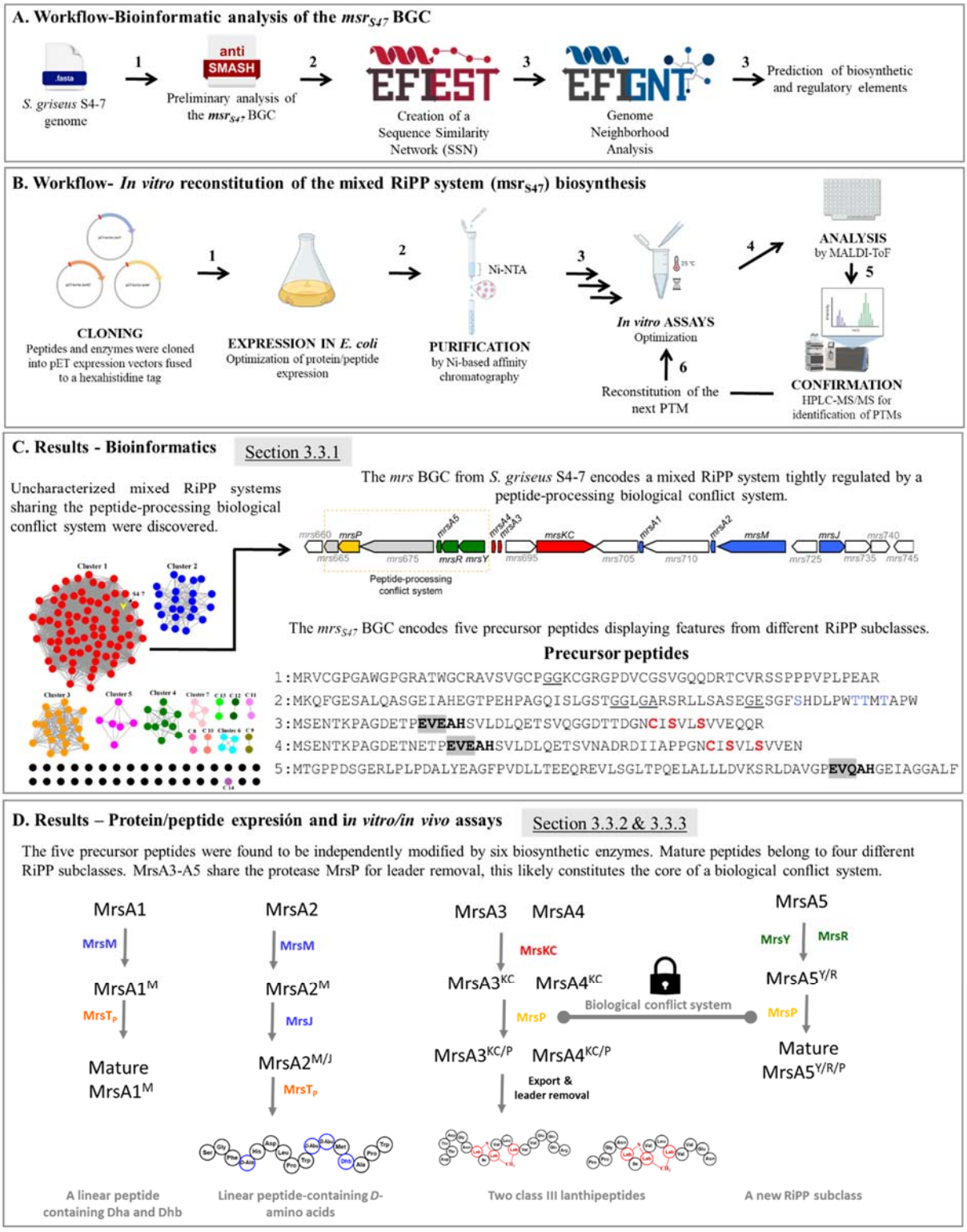
Further chemical and biological experiments need to be conducted to fully characterize the structure and biosynthesis of the thiopeptide cpr. First, the candidate ion obtained in this thesis needs to be confirmed as the actual cpr thiopeptide. Large-scale cultivation of the *Streptomyces* strains (*S. coelicolor* M1152/pSET:*ermE*:cpr*BGC r-1 or r-2, and *S. albus* 1074/pSET-*ermE**-*cpr*BGC r-3) followed by extraction and LC-MS/MS analysis will be required to validate the reproducibility of the experiment. Also, partial fractionation of crude extracts and subsequent analysis through NMR could

help to connect the cpr candidate with a thiopeptide structure. Once the cpr candidate has been confirmed, upscaling the production for purification of the mature compound will be possible to finally enable the evaluation of the biological activity of the thiopeptide.

Heterologous expression has traditionally been used to study the function of individual genes found in BGCs, to study biosynthesis of NPs in a more genetically tractable host, to produce metabolites at higher levels than in the natural producer, and even to generate new analogs by combinatorial biosynthesis. (Gomez-Escribano & Bibb, 2011; Gust, 2009). Nevertheless, results obtained in this project displayed some obstacles that still need to be overcome. As exemplified by most of the *cpr* heterologous expression experiments, BGCs may remain silent in a heterologous host, thus requiring extensive refactoring steps to achieve their activation. Despite the vast synthetic biology toolbox, heterologous expression still requires time-consuming trial and error processes, especially regarding the selection of the heterologous host, since there is no single strain that can be applied universally.

3.3 Studies on the *mrs*$_{S47}$ BGC revealed a novel family of mixed RiPP system BGCs from Actinobacteria

Graphical abstract



3.3.1 Genomic overview and bioinformatic analysis of the mixed RiPP system BGC from *S. griseus* S4-7 (mrs_{S47} BGC)

Given the presence of the *lanM* gene (*mrsM* hereafter) and after a first bioinformatic analysis, the mrs_{S47} cluster was thought to encode a class II lanthipeptide (Kim *et al*., 2019). Preliminary experiments aimed to obtain the mature lanthipeptide by co-expression of the MrsM synthetase and the originally predicted single precursor peptide (MrsA3 in Figure 28) in *E. coli*. Despite successful co-expression and some optimization experiments, no PTMs were observed in the precursor peptide after analysis by HPLC-MS/MS (Supplementary Figure S10). For that reason, more detailed and advanced bioinformatic analyses were conducted in our laboratory.
The mrs_{S47} BGC (Figure 28) was analyzed using the latest version of antiSMASH and other specialized tools for analysis of RiPP BGCs such as RiPPMiner (Agrawal *et al.*, 2017) and DeepRiPP (Merwin *et al*., 2020).

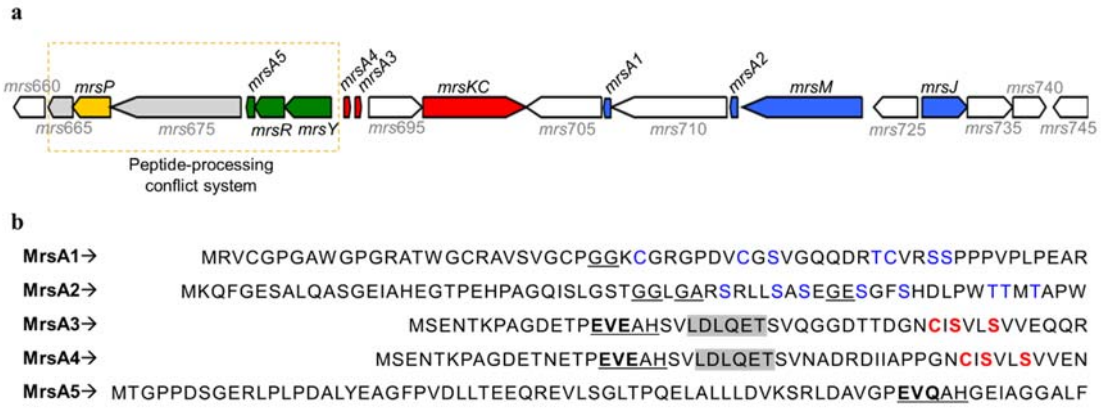


Figure 28. The mrs_{S47} BGC found in the genome of *S. griseus* S4-7 encodes a mixed RiPP system. **(a)** The mrs_{S47} BGC encodes five precursor peptides (MrsA1-A5), six biosynthetic enzymes (MrsM, MrsJ, MrsKC, MrsP, MrsY, and MrsR), four CSRs (here designated as *mrs*675, *mrs*705, *mrs*710, and *mrs*725) and other proteins of unknown function. A peptide-processing conflict system (indicated with a discontinued yellow box) was predicted to be part of the mixed RiPP system. *mrs*665 encodes a GAF domain while *mrs*675 encodes a protein with an *N*-terminal adenylate/guanylate cyclase domain and an AAA ATPase domain. **(b)** Precursor peptide sequences. MrsA1 and MrsA2 were predicted to be modified by the MrsM synthetase to give a class II lanthipeptide and a linear peptide containing dehydroamino acids; residues susceptible to be modified are indicated in blue color and the probable cleavage points are underlined. MrsA3-A4 were predicted to be the MrsKC substrates, and the residues involved in labionin formation are shown in red color; The MrsKC binding sequence and the hypothetical MrsP recognition element are indicated in a gray box and underlined, respectively. MrsA5 harbors an EVE-like motif for leader peptide removal via MrsP.

Results revealed that, in addition to the MrsM enzyme, a class III lanthipeptide synthetase (MrsKC) was encoded within the cluster. The mrs_{S47} BGC also unexpectedly encodes multiple putative precursor peptides (MrsA1-A5) and some of them are lacking the lanthipeptide hallmark (multiple Ser and Thr residues for dehydroamino acids formation and cysteines for thioether crosslinks). The existence of other modifying enzymes like a Rossmann fold-containing oxidoreductase (MrsJ) and a radical SAM maturase (MrsY) pointed out this BGC to be a new member of the previously described GCF146 (Zhang *et al*., 2015). Nevertheless, our analysis also displayed a set of modifying enzymes not previously reported for this family of clusters. These were predicted to be an S8 serine-like peptidase (MrsP) and an iron-containing redox enzyme (MrsR). Since no members of the GCF146 have been experimentally characterized and given the existence of additional biosynthetic elements previously overlooked, all encoded enzymes and

peptides were bioinformatically examined to predict their biosynthetic role (Table 4). For those enzymes whose function could be readily inferred based on literature and sequence homology (MrsM, MrsJ, MrsKC and MrsP), alignment with already characterized proteins was sufficient to confirm their structural integrity; conversely, modifying enzymes with no clear precedent in peptide maturation (MrsY and MrsR) were submitted to deeper analyses.

Table 4: Proposed function of open reading frames (ORFs) in the mrs_{S47} biosynthetic gene cluster.

ORF	Protein size (aa)	Predicted function	Hypothetical protein	Protein homolog (organism)	Identity /coverage (%/%)
*mrs*660	261	Unknown	Xylose isomerase	Xylose isomerase (*Streptomyces* sp. CB02058)	90
*mrs*665	205	Unknown	GAF domain containing-protein	Histidine kinase (*Streptomyces* sp. CB02366)	90
mrsP	312	Leader peptide removal and core of the conflict system	S8 peptidase	AmyP (*Bacillus amyloliquefaciens*)	23
*mrs*675	1079	Regulation	Adenylate/ guanylate cyclase	AAA ATPase domain (*Actinosynnema pretiosum* subsp. *pretiosum*)	30 / 55
*mrsA*5	72	-	Precursor peptide	-	-
mrsR	241	Peptide-processing	Iron-containing redox enzyme	PqqC (*Klebsiella pneumoniae*)	21
mrsY	382	Peptide processing	rSAM/SPASM	-	-
*mrsA*4	53	-	Precursor peptide	-	-
*mrsA*3	49	-	Precursor peptide	-	-
*mrs*695	457	Transport	Transporter	BotT (*Streptomyces bottropensis*)	30 / 93
mrsKC	872	Lanthipeptide biosynthesis	Class III lanthipeptide synthetase	LanKC (*Streptomyces viridochromogenes* DSM 40736)	39 / 97
*mrs*705	642	Regulation	SARP	Noc-10 (*Nocardia* sp. ATCC 202099)	40 / 38
*mrsA*1	61	-	Precursor peptide	-	-
*mrs*710	976	Regulation	LuxR protein	DnrI/RedD/AfsR-family transcriptional regulator (*S. clavuligerus* ATCC 27064)	24 / 41
*mrsA*2	65	-	Precursor peptide	-	-
mrsM	1014	Dehydratase	Class II lanthipeptide synthetase	GarM (*Actinoplanes garbadinensis*)	33 / 96
*mrs*725	371	Regulation	LacI protein	LacI family transcriptional regulator (*Verrucosispora* sp. FIM060022)	42 / 98
mrsJ	383	D-amino acid formation	Reductase	Gfo/Idh/MocA family oxidoreductase (*Paenibacillus polymyxa* E681)	27 / 59

*mrs*735	413	Unknown	DUF993	Dihydrodipicolinate synthase/N-acetylneuraminate lyase (*Streptomyces* sp. JS01)	90
*mrs*740	278	Unknown	Xylose isomerase-like TIM barrel	Sugar phosphate isomerase/epimerase (*Streptomyces* sp. Root1295)	90
*mrs*745	183	Unknown	DUF664	GacF (*Micromonospora echinospora*)	34 / 98

3.3.1.1 Class II lanthipeptide synthetase - MrsM

Class II lanthipeptide synthetases are known generally as LanMs. These bifunctional enzymes carry an *N*-terminal domain responsible for dehydration of Ser/Thr residues, and a *C*-terminal cyclase that forms (me)lanthionine bridges (Repka *et al*., 2017). To our knowledge, only two LanM proteins that catalyze dehydration but no cyclization have been reported. These were found in the *npn* BGC from the filamentous cyanobacterium *Nostoc punctiforme* PCC 73102, and in the *bsj* BGC from *Bacillus cereus* SJ1 (Zhang *et al*., 2014; Huo & van der Donk, 2016). NpnM is the modifying enzyme of a set of six peptides (named NpnA1-A6) that lack Cys residues in the core region. The products of NpnM are linear peptides devoid of thioether bridges. It has been proposed that absence of cysteines in the NpnA precursor peptides likely resulted from mutations in the sequence of ancestor LanAs, and that NpnM coevolved with its substrates (Zhang *et al*., 2014). The *bsj* BGC encodes the two-component lantibiotic bicereucin. Three precursor peptides BsjA1-A3 are found in this cluster; BsjA1 and BsjA3 share near identical sequences and have multiple Ser/Thr residues but lack cysteines. BsjM installs four Dha and four Dhb residues in BsjA1 and BsjA3 is giving the same linear product. BsjA2 has a different core peptide sequence containing a sole Cys; the same class II lanthipeptide synthetase BsjM produces four Dha, seven Dhb residues, and one lanthionine crosslink. Bicereucin is the only two-component lantibiotic in which one peptide contains a single thioether bridge and the second is a linear peptide (Huo & van der Donk, 2016).

The cyclase domain of MrsM was aligned to that of NpnM, BsjM and seven additional LanM proteins reported to form (me)lanthionine crosslinks (Figure 29), i.e. the actagardine synthetase GarM from *Actinoplanes garbadinensis* (Boakes *et al.*, 2009), CerM from *Bacillus cereus,* which is the modifying enzyme of seven peptides named cerecidins A1-A7 (Wang *et al*., 2014), HalM1 and HalM2 that produce the two-component lanthipeptide haloduracin (Lawton *et al*., 2007), LchM1 from lichenicidin α (Shenkarev *et al*., 2010), the processing enzyme (FlaM) of flavecin α and flavecin β from *Ruminococcus flavaciens* (Zhao & van der Donk, 2016), and LctM encoded in the lacticin BGC from *Lactococcus lactis* (McAuliffe *et al*., 2000).

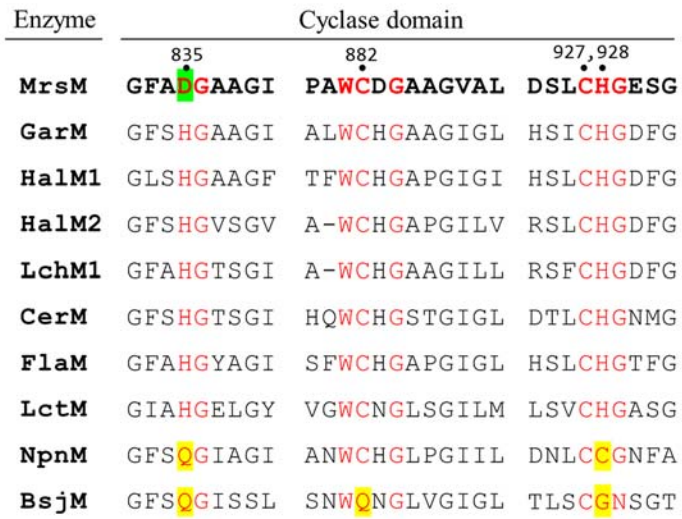


Figure 29. Sequence alignment of MrsM and selected LanM synthetases. Active site residues highly conserved among cyclase domains are in red color. Asp835 in MrsM replaces the conserved His residue required for cyclase activity in all the enzymes that form thioether crosslinks. Mutations in the active site in NpnM and BsjJ are highlighted in a yellow box.

MrsM was found to be a 1,014-amino acid protein with strong sequence similarity to referenced class II lanthipeptide synthetases. BlastP analysis clearly detected the zinc binding site (Cys882, Cys927, His928 in MrsM) that has been shown to be important for cyclization (Zhang *et al*., 2012). A CCH motif is found in all enzymes forming (me)-lanthionine bridges, but not in NpnM and BsjM that instead have a CCC and a QCG motif, respectively. It was proposed that the presence of three Cys ligands account in part for the high substrate tolerance of these enzymes in a similar way as for ProcM in prochlorosins biosynthesis (Zhang *et al*., 2014). The presence of a CCH in MrsM could be an indication of substrate specificity.

A highly conserved HG motif is found in almost all LanMs where the His is essential for cyclase activity. This motif is also present in the cyclase NisC thus being crucial for cyclization in all classes of lanthipeptides. In contrast, NpnM and BsjM present an impaired active site with a QG motif instead of HG (Zhang *et al*., 2014). MrsM was found to present a mutation in the active site but here, the His was replaced by Asp (Figure 29). Altogether, MrsM was expected to function as a dehydratase but its functionality as cyclase could only be confirmed experimentally. Conserved regions important for catalysis in lanthipeptide cyclases are red colored in Figure 29.

Within the five precursor peptides found in the *mrs*$_{S47}$ BGC, MrsA1-A2 were predicted to be the MrsM substrates. MrsA1 is 61 amino acids in length and contains the characteristic Gly-Gly motif for leader removal via a LanT$_P$. Considering this motif as the cleavage site, the core peptide is composed of 33 amino acids including three Cys, three Ser and one Thr residues. Hence, this precursor peptide was expected to form a class II lanthipeptide containing (me)lanthionin crosslinks. In contrast, MrsA2 is 64 amino acids in length and is rich in Thr and Ser but has no Cys residues in the *C*-terminal region. Based on the predictions regarding the GCF146 and the analysis of the MrsM lanthipeptide synthetases, this peptide was also proposed to be a MrsM substrate but was expected to remain linear. RiPPMiner was used to predict the cleavage site; however, its limited capacity for analyzing peptides lacking cysteines did not allow the prediction of potential cleavage sites. Still, the sequence contains three points where leader removal could take place through a LanT$_P$ protease (Figure 28b). Reconstitution of MrsM activity on MrsA1-A2 would confirm the actual cleavage point based on the modified residues.

3.3.1.2 Rossmann fold-containing oxidoreductase - MrsJ

As described in section 1.3.4.1, four different subclasses of oxidoreductases (LanJ) have been reported to stereoselectively reduce Dha and Dhb to form D-alanine and D-amino butyric acid in lanthipeptides (Lohans & Vederas, 2014). D-amino acids have also been found in the NpnA linear peptides and in the two-component bicereucin (Zhang *et al*., 2014; Huo & van der Donk, 2016). Even though both pathways (Npn and Bsj) yielded linear peptides, the corresponding LanJs belong to different subgroups. $NpnJ_A$ is a zinc-dependent dehydrogenase that uses NADPH as co-substrate to install one D-Ala in NpnA3 in a leader-independent manner (Huo & van der Donk, 2016). On the other hand, the flavin oxidoreductase $BsjJ_B$ converted one Dha into D-Ala in BsjA2 and formed one D-amino butyric acid and three D-Ala residues in BsjA2 (Zhang *et al*., 2014).
According to BLAST analysis, MrsJ contains an *N*-terminal Rossmann-fold NAD(P)H/NAD(P)(+) binding (NADB) domain, and is a member of the Gfo/Idh/MocA [GFO: glucose-fructose oxidoreductase, IDH: inositol 2-dehydrogenase and MOCA which catalyzes a dehydrogenase reaction involved in rhizopine catabolism (UniProt Consortium, 2023)] superfamily of oxidoreductases. Proteins belonging to this group have the *N*-terminal nucleotide-binding domain and a *C*-terminal α/β- domain important for substrate binding and catalysis (Taberman *et al*., 2016).
The primary sequence of MrsJ and twelve selected oxidoreductases involved in RiPP biosynthesis were organized into networks based on their sequence similarity (Figure 30). The web resource developed by the Enzyme Function Initiative (EFI, accessed https://efi.igb.illinois.edu) provides genomic enzymology tools for generation of sequence similarity networks (SSNs) for protein families, and for analysis and visualization of genomic context of the clustered proteins (Zallot *et al*., 2019). An SSN is a multidimensional network that displays the pairwise sequence similarity relationships among homologous proteins. In an SSN, each protein is represented by a node; two nodes are connected by an edge if they share pairwise sequence similarity above a specified threshold. The most related proteins are grouped together into clusters (Zallot *et al*., 2019). To create a sequence similarity network for MrsJ, protein sequences were submitted to the enzyme similarity tool (https://efi.igb.illinois.edu/efi-est/ option C: FASTA sequences as input (Zallot *et al*., 2019)) that calculated the similarity among sequences and connected them based an alignment score of 10 (Figure 30).

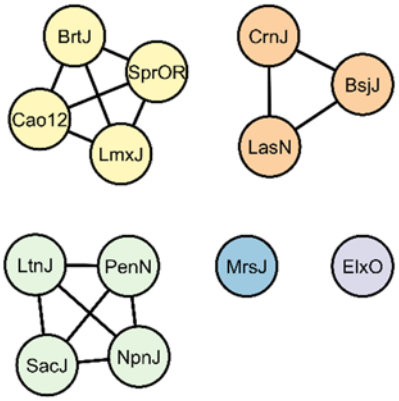


Figure 30: Sequence similarity network (SSN) of selected reductases (LanJs) involved in RiPP biosynthesis. The network was created by using a list of sequences in FASTA format. Protein sequences were used "as is", with an alignment score of 10. **(Green)** zinc-binding dehydrogenases ($LanJ_A$). NpnJ from *Nostoc punctiforme*; SacJ from *Staphylococcus aureus*; LtnJ from *Lactococcus lactis* and PenN from *Lactobacillaceae*. **(Orange)** flavin oxidoreductases ($LanJ_B$). BsjJ from *Bacillus* sp.; CrnJ from *Streptomyces* sp. A108 and LasN from *Latilactobacillus*. **(Yellow)** luciferase-like monooxygenases ($LanJ_C$)

and F450 H_2-dependent reductases (LanJ$_D$). BrtJ from *Streptomyces rimosus*; LmxJ from *Streptomyces rochei*; SprOR from *Streptomyces pristinaespiralis* and Cao12 from *Streptomyces cacaoi*. **(Lavender)** NADP-dependent oxidoreductase ElxO from *Staphylococcus* (this enzyme does not form D-amino acids). **(Blue)** Gfo/Idh/MocA oxidoreductase MrsJ found in the mrs$_{S47}$ BGC.

Despite members of the LanJ$_A$ and LanJ$_B$ groups used as references are from distant species, they formed their respective networks (green and orange). Analogously, BrtJ and LxmJ (subclasses C and D) clustered together along with the LLM class flavin-dependent oxidoreductases Cao12 and SprOR (network in yellow). Cao12 forms five D-Ala and two D-amino butyric acid residues in cacaoidin, which is the first member of the lanthidin family of RiPPs (Román-Hurtado *et al*., 2021). SprOR installs D-Ala and D-amino butyric acid residues in the class V lanthipeptide pristinin (Kloosterman *et al*., 2020). ElxO was not connected to any class of LanJ as this oxidoreductase is not involved in D-amino acid formation. Given the existence of a precursor peptide lacking Cys residues (MrsA2) and potentially a class II lanthipeptide (MrsA1) that can be modified by the MrsM synthetase, it was proposed MrsJ converts Dha/Dhb into D-amino acids in a similar mechanism to that of the known reductases; however, MrsJ may represent a new subclass of LanJs.

3.3.1.3 Class III lanthipeptide synthetase - MrsKC

Class III lanthipeptides are synthesized by trifunctional enzymes generically called LanKCs. These proteins are composed of three domains: a central kinase responsible for activation of Ser/Thr residues; an *N*-terminal lyase for elimination of the phosphate group; and a *C*-terminal cyclase that catalyzes the conjugate addition (Hegemann & van der Donk, 2018). LanKCs tend to form (me)labionin scaffolds when the precursor peptide contains the appropriate motif (S$_{X2}$S$_{X2-5}$C) that provides an additional Ser residue to be used as nucleophile in a second cyclization reaction (Meindl *et al.*, 2010).
To confirm the presence and integrity of the three catalytic domains in MrsKC, a sequence alignment with representative class III lanthipeptide synthetases was carried out (Figure 31). Enzymes used as references covered variations of PTMs that are not limited to thioether crosslinks. For example, LabKC installs two labionin structures and one disulfide bond yielding structures known as labyrinthopeptins (Meindl *et al*., 2010). AciKC forms two labionins in catenulipeptin; *in vitro* characterization of this enzyme provided evidence of the *C*-to-*N*-terminal directionality of cyclization events in class III lanthipeptides (Wang & van der Donk, 2012). EryKC forms a mixture of labionin and lanthionine in erythreapeptin, although it is unknown which conditions are leading to either one or the other structure. Finally, MicKC works interdependently with the FAD-dependent cysteine decarboxylase MicD to form an avionin structure in microvionin lipolanthines (Wiebach *et al*., 2020).
Sequence alignment demonstrated MrsKC shares homology with the reference enzymes along the three domains. Three conserved residues (Lys, Asp and Asn) in the kinase domain are important for activity (Hegemann & Süssmuth, 2020). The amine side chain of Lys is important for coordination of the α- and β-phosphates of ATP while Asp and Asn are needed for catalysis. The Asp acts as an acceptor during proton transfer whereas the Asn forms a canonical contact with the Asp and participates in Mg^{2+} coordination (Hegemann & Süssmuth, 2020). Sequence alignment confirmed the presence of Asp367 and Asn372 in the catalytic loop of the kinase domain in MrsKC; however, the conserved Lys was replaced by Arg257. Although it is unknown which impact Lys replacement could have on the MrsKC activity, it might be related to the affinity of the enzyme to a specific phosphate donor other than ATP (conserved residues are highlighted in red).

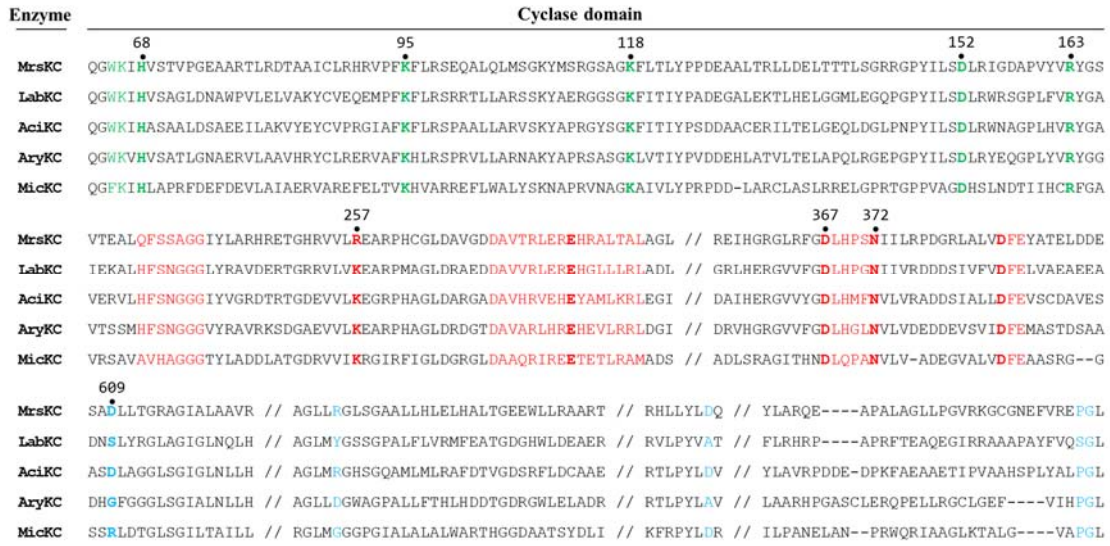


Figure 31. Sequence alignment of MrsKC and selected class III lanthipeptide synthetases. Conserved catalytic residues are shown in green (lyase), red (kinase), and blue (cyclase). Residues required for activity are in bold and residue numbers are based on MrsKC.

Five conserved residues critical for lyase activity (His68, Lys95, Lys118, Asp152 and Arg163) were identified in the MrsKC lyase domain. Based on the role of these conserved residues in the reference enzymes (Hegemann & Süssmuth, 2020), the Lys118 would act as the catalytic base to remove the α-proton of the pSer/pThr to initiate the elimination reaction. His68 would act as catalytic acid for protonation of the bridging oxygen of the phosphate leaving group, and the interaction with Asp152 would stabilize the protonated form of the His side chain for this function. Lastly, the highly conserved Asp residue crucial for cyclase activity was clearly identified in the cyclase domain of MrsKC (other important residues are highlighted in blue). Based on the presence of residues critical for catalysis, MrsKC was expected to be functional as a class III lanthipeptide synthetase installing thioether crosslinks and/or labionins.

Precursor peptides MrsA3 and MrsA4 found in the mrs_{S47} BGC were predicted to be MrsKC substrates. Both peptides (48 and 52 amino acids, respectively) share high homology along their whole sequences. They contain the LDLQET motif which is the conserved binding site for LanKC synthetases (Hegemann & Süssmuth, 2020). Labionins derive from an $S_{X2}S_{X2\text{-}5}C$ motif and are exclusively found in class III lanthipeptides. Interestingly, MrsA3 and MrsA4 show a $CS_{X1}S_{X2\text{-}5}$ motif which could be a variation of the labionin-forming sequence. Besides, the *N*-terminal regions of MrsA3-A4 are larger than extensively studied class III lanthipeptides (Supplementary Figure S11); these leader sequences might be required for binging additional modifying enzymes during peptide processing or could also have alternative roles. Concerning the cleavage site, analysis by RiPPMiner predicted core peptides 18 and 15 amino acids in length for MrsA3 and MrsA4, respectively.

3.3.1.4 An S8 serine-like protease for leader removal- MrsP

mrsP encodes a member of the S8 family of peptidases. Also known as subtilase family, S8 peptidases are the second largest group of nonspecific serine endopeptidases. These proteins have a catalytic triad Asp-His-Ser and are divided into two subfamilies S8A and S8B.

At the time this peptidase was studied in our laboratory, there were no reports on the participation of S8 peptidases in class III lanthipeptide biosynthesis. Nonetheless, in 2022 Zhang *et al*. reported a comprehensive characterization of the first S8 protease involved in maturation of class III

lanthipeptides (Zhang *et al*., 2022). Amylopeptins (Amy I-VII) are class III lantibiotics derived from the gut microbiota of Sprague-Dawley rats and exhibit narrow antimicrobial activity. The S8 protease AmyP was found to remove the leader peptide in a site-specific reaction that is preceded by labionin formation. Detailed characterization of AmyP catalysis identified the recognition element for the protease to bind its substrate. AmyP binds the (T/S/A)EVE motif in the leader peptide(s), which is located two amino acids *N*-terminally to the cleavage site, and its activity is independent of both modifications in the precursor peptides and the AmyKC binding site (VLELQQL) (Zhang *et al*., 2022). Comparative sequence analysis of AmyP identified residues Ser395, His194 and Asp160 as the potential catalytic triad. This was further confirmed by mutagenesis experiments where the proteolytic activity of AmyP was completely abolished when Ser395 was replaced by Ala (Zhang *et al*., 2022).

To investigate if MrsP could be the peptidase responsible for leader removal in these class III lanthipeptides, its sequence was aligned with that of AmyP and other known proteases involved in lanthipeptide maturation. The proteins used as reference included ElxP which trims ElxA to expose an *N*-terminal Dhb that undergoes secondary PTMs to form epilancin 15x (Ortega *et al*., 2014), the protease of lichenicidin LicP (Tang *et al*., 2015), and NisP which is the extracellular subtilisin-like serine protease responsible for leader removal in nisin biosynthesis (Lagedroste *et al*., 2017). As seen in Figure 32, the three conserved residues critical for catalytic activity were present in MrsP, including the serine residue (Ser395 in AmyP and Ser251 in MrsP) that defines these proteins as serine proteases. Eight other residues were highly conserved in the reference sequences (highlighted in yellow in Figure 32); in MrsP however, these residues (Val65, Val127, Gly144, Val156, Gly230, Gln233, Thr244, and Arg26) were not conserved. The relevance of these residues in the proteolytic activity of LanPs is still unknown.

Extracellular lanthipeptide proteases like LicP, NisP and AmyP catalyze leader removal as the final biosynthetic step. These proteins contain a short *N*-terminal peptide chain (16-30 amino acids) which is common to proteins whose activity takes place extracellularly (Tang *et al*., 2015). This sequence is known as "signal peptide" and is crucial for LanPs maturation; such maturation processes involve self-cleavage and release of the translocated protein and signal peptide (Tang *et al*., 2015). ElxP conversely lacks the signal peptide, which could be related to its intracellular function in epilancin maturation. The primary sequence of MrsP was examined using the PrediSi (prediction of signal peptide) tool (Hiller *et al*., 2004); interestingly, no signal peptide for secretion was found.

The existence of residues crucial for leader removal in class III lanthipeptide suggested MrsP as the putative protease for the MrsKC products; nevertheless, the lack of a signal peptide for translocation could be a hint that MrsP does not perform the last maturation step in the biosynthesis of these lanthipeptides from the mrs_{S47} BGC.

In fact, both leader peptides MrsA3-A4 harbor the peptidase recognition element (EVE) found in AmyAs. Thus, the hypothesis of a leader removal by the S8 protease MrsP was further supported. Intriguingly, the position of MrsP and MrsKC binding sites in the leader peptides suggested that MrsP would trim off a portion of the LP maintaining the MrsKC binding site still attached to the core peptide (Figure 28b). This observation along with the fact that MrsP does not contain a signal peptide for translocation suggests that PTM(s) could precede the labionin formation by MrsKC, and leader peptide removal would take place in a two-step mechanism. Due to the high sequence similarity of MrsA3 and MrsA4 in the core region, and especially in the region where the labionin is expected, the possibility that both precursors give the same product being *mrsA3* and *mrsA4* a duplication to ensure gene expression cannot be excluded.

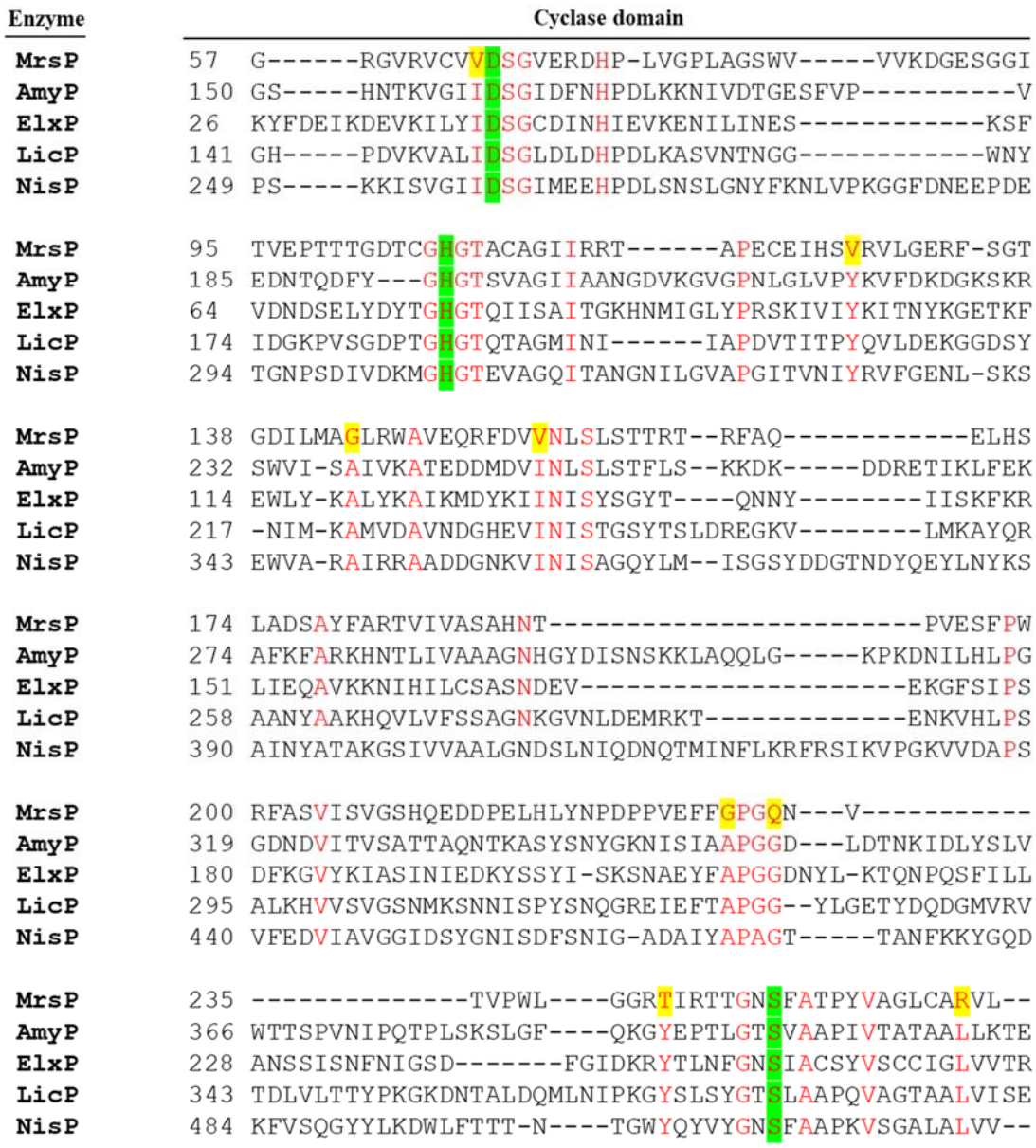


Figure 32. Sequence alignment of selected proteases involved in lanthipeptide maturation. Highly conserved residues among proteases are in red color and highlighted in green boxes. The eight catalytic residues in MrsP are highlighted in yellow boxes.

3.3.1.5 The radical SAM/SPASM protein MrsY and existence of other mixed RiPP ssytems

The radical SAM enzyme (MrsY) was initially predicted to install secondary PTMs in peptides that were previously modified by the lanthipeptide synthetases (Zhang *et al*., 2015). Since there was no precedent about the type of modification this enzyme could catalyze in the system, in this thesis, a detailed bioinformatic analysis for MrsY was performed. BLAST alignment confirmed the presence of the highly conserved CxxxCxxC motif for coordination of the Fe-S cluster. To get insights into the subfamily this protein belongs to, the amino acid sequence was submitted to RadicalSAM.org (https://radicalsam.org/explore.php). This resource enables identification of isofunctional groups of rSAM enzymes using sequence similarity networks (SSNs) and genomic context (Oberg *et al*., 2022). The MrsY protein was found to be a member of the large and functionally diverse rSAM/SPASM subfamily. As illustrated in section 1.3.5, several members of this family have been recently described to catalyze a number of chemical transformations in

RiPP biosynthesis. Thus, a deeper analysis of MrsY was required to obtain a more accurate function prediction.

The EFI-Enzyme Similarity Tool (EFI-EST) is used to create SSNs for a target protein and its closest homologs. A protein sequence is used as query for a search of the UniProt, UniRef90, or UniRef50 database using BLAST. The EFI-EST retrieves a maximum number of sequences (≤ 10,000, default: 1,000) and calculates their similarity. To generate the SSN, a minimum sequence similarity threshold should be specified by the user. This value indicates the sequence pairs connected by edges and determines the segregation of proteins into clusters (Zallot *et al*., 2019). Once the analysis is completed, the SSN is available for visualization and processing in Cytoscape. The SSN can be manually diced into isofunctional clusters by adjusting the similarity threshold, and fragmented or incomplete sequences can be removed before using the SSN for further analysis of genomic context. A genome neighborhood network (GNN) enables large-scale identification of functionally linked genes. EFI-Genome Neighborhood Tool (EFI-GNT) collects, analyzes, and illustrates genome neighborhood information for proteins in clusters of an SSN provided as input. EFI-GNT generates a colored GNN with genome context information and the corresponding diagrams (BGCs) are visualized in an interactive genome neighborhood diagram (GND) viewer (Zallot *et al*., 2019). Here, each sequence from the SSN is considered as a query. Information related to proteins that are neighbors of the input queries is collected from the database. Then, the co-occurrence frequencies of the identified neighboring Pfam families with the input queries are calculated for each cluster. Every GND displays a BGC that corresponds to a node in the colored GNN. To prevent confusion, the term “BGC” or “diagram” will be hereafter used to refer to biosynthetic gene clusters while “cluster” will refer to the groups generated in the SSN.

The primary sequence of MrsY was used as input to create an SSN (with 1,000 retrieved sequences and initial alignment threshold of 8). EFI-ETS built an SSN with 258 sequences (259 nodes) and 32,693 edges. The network was visualized and diced in Cytoscape. Nodes with fragmented sequences were removed, and the percentage of identity and alignment score were adjusted to 55% and 114, respectively. The filtered SSN was then submitted to the EFI-GNT with a neighborhood size of 10, which means 10 genes up- and downstream, and 20 % of minimal co-occurrence.

The colored GNN generated by EFI-GNT contained 14 clusters and 33 singletons (Figure 33). Visualization of interactive genome neighborhood diagrams (GNDs) allowed to explore the genomic context of MrsY in every cluster and singleton (twelve representative diagrams are depicted in Figure 34). Each cluster/singleton represented a unique pathway, and each node corresponded to a BGC or a set of identical or at least 55% similar BGCs containing the target rSAM protein in a given organism.

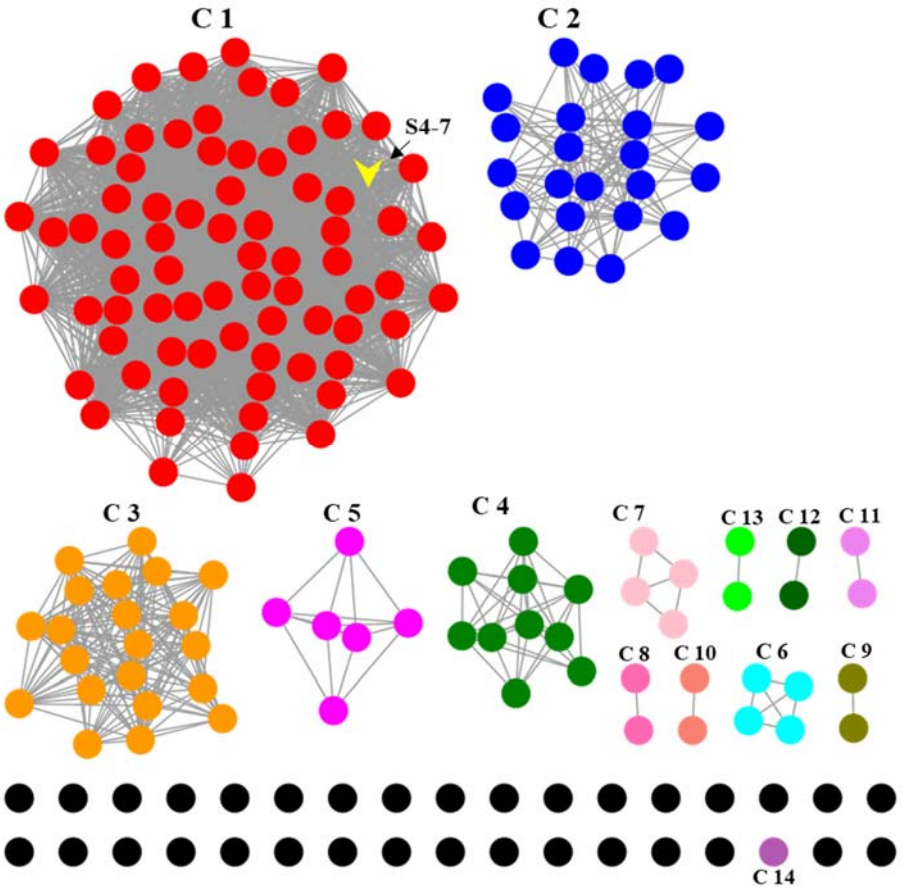


Figure 33. Sequence similarity network (SSN) of MrsY from *S. griseus* S4-7. A set of 258 sequences was retrieved from the UniProt database. Sequences were organized into 14 clusters and 33 singletons. The MrsY sequence was in cluster 1 (indicated in yellow). Sequences containing in a cluster share at least 55 % similarity and an alignment score (AS) of 114 was used to organize the sequences into clusters/singletons.

Regardless of the cluster or organism, all BGCs contained a highly conserved operon encoding a radical SAM/SPASM protein, an iron-containing redox enzyme, and an S8 serine-like protease. These operons were analogous to that found in the mrs_{S47} BGC composed by MrsY, MrsR and MrsP. Besides, these operons display a precursor peptide similar to MrsA5 found in the mrs_{S47} BGC. This precursor does not contain Cys residues and is poor in Ser and Thr residues across its whole sequence. Since this additional precursor peptide lacks the typical hallmark for lanthipeptides and has no similarity to any previously characterized RiPP, it was initially overlooked in the mrs_{S47} BGC and only recognized as part of the RiPP system after bioinformatic analysis of the rSAM protein described here. Peptide sequences from all BGCs found in the GNN were compared to identify common patterns. Amino acid sequences were highly conserved in all members of clusters 1 and 2. Little variations at the *C*-terminus were found across peptides from other clusters (Figure 34). The most intriguing feature in these precursors was a (E/D)V(E/Q/V/R) motif that seemed to be a variation of the EVE motif for binding of S8 proteases in class III lanthipeptides. The Val residue at position -4 was always conserved while the Glu residues at position -3 and -5 showed variability. Besides, the His residue at position -1 seemed to be essential for the enzyme to cleave the peptide exactly two residues *N*-terminally to the cleavage site. The presence of the EVE-like motif in the precursors suggested that the S8 peptidase encoded in these BGCs is responsible for leader removal in MrsA5-like peptides. This finding was especially relevant for members of cluster 1 where the class III lanthipeptides were predicted to be cleaved by the same protease. Hence, the S8 protease would represent a crucial intersection in the

maturation of these mixed RiPP systems and was the key to understand the whole biosynthetic pathway.

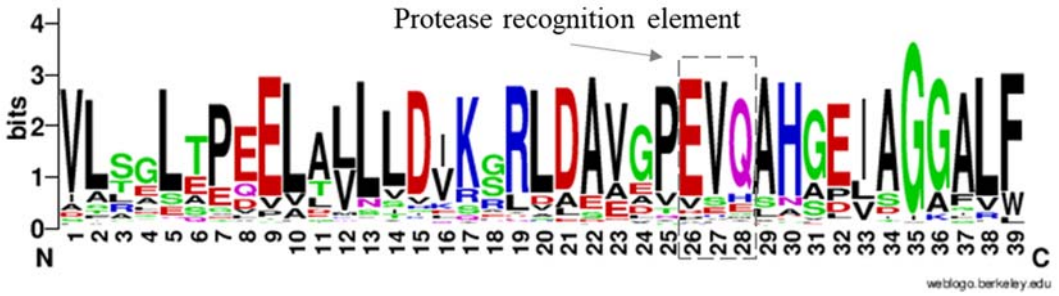


Figure 34. The sequence logo of MrsA5-like precursor peptides found in the BGCs retrieved during SSN analysis of MrsY. Sequences showed high similarity including the recognition element for S8 protease-mediated leader removal.

Peptide chains ranging from seven to thirteen residues in MrsA5-like precursors thus represented the core region. These contained a variable nonpolar aliphatic (Gly, Ala or Ile) or polar uncharged (Ser or Thr) residue at position 1 and an aromatic residue (either Phe or Trp) at the *C*-terminus. To better understand the organization of these systems, 11 representative clusters/singletons, from the SSN of MrsY shown in Figure 33, were selected for a deeper exploration (Figure 35). They were chosen based on the presence of biosynthetic elements that could provide additional information on the biosynthesis of these mixed RiPP systems, and on the presence of potential RiPP precursor peptides (confirmed by using RiPPMiner).

Cluster 1: *Streptomyces griseus* S4-7- this study (See Figure 28)

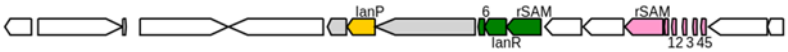
Cluster 2: *Streptomyces* sp. FR-008

Precursor peptides:
1:MSENTSPETPETVEAP**EVE**AHSASVLDLQGTTSVDQEHIADGNCISVLSVVENAK
2:MQIDDAATHLTDGAGESAGAHPVGAITLFTTGALGLRSRLLSASEGGASYSHELPWTTMTAPQ
3:MTPPQRPDLLTRLHEAGFAVDLLTEEQREVLTGLEPDELAVLLDIKGRLDAVGP**EVQ**AHGEIAGGALF

Cluster 3: *Streptomyces* sp. ADI91-18

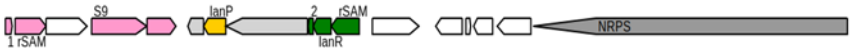
Precursor peptides:
1:MPENTNPETENEDI**EVV**AHREDEEAEAGCIVNNSHEM
2:MTEDTTPYEDASSGGSPSEPTLEALAAAGFELDTLSEEQYEVLRALSPEELALLVDIRGRLDAAAP**EV**
QAHADVAGGALF

Cluster 3: *Streptomyces* sp. 846.5



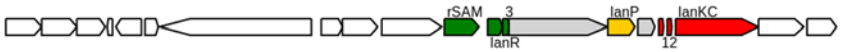
Precursor peptides:
1: MAYTDDEKPRIEDGTDDEP**EVV**AHGVLGEDDSFNLKCPLMLET
2: MAHIDDEKPQVEDEALEGQEAP**EVV**AHGIDDATLDADGKCTVYVKI
3: MAQTDDEKPQVEDEAQDAP**EVV**AHAVDDAELSEFECKIFVEI
4: MAQTDDEKPQVEDEAQDAP**EVV**AHGVDDDATLGEMDCKIFVEL
5: MAENTQPSSTDQPENQAVPALEEVEGADERP**EVV**AHMTSAIGIPQCNYYNEF
6: MNNHDTTLDRLQTAGFALDALNDEQREVLRALTPQELAVLIDIKGRLDELGP**EVQ**AHSEIAGGALF

Cluster 5: *Streptomyces* sp. M41



Precursor peptides:
1:MAENVEPNEETTETPAEETP**EVV**AHAASNVAWCGVHQESVDEP**EVV**AHAASNVAWCGVHQEHLDEP**EVV**AHSEEGEDTPWCGVHQEHNA
2: MTDPAPDRLRALEDAGFPLHSLSPEQREVLQDLSTEELTLLLGLKERLDAAEP**EVQ**AHSEIAGGALF

Cluster 7: *Kitasatospora cheerisanensis* KCTC 2395



Precursor peptides:
1:MSADQTPEDKAAETEAVLSLQEAPTAQTED**EVQ**AHI**S**TA**S**LEQ**C**SNQGTKTRI
2:MSNTPEDSATETEAVLSLQETPTAQSED**EVQ**AHI**S**TS**S**VHI**C**SPVGDKTE
3:MTDRTIGPDPALAALTAAGFDLSGLTEEQHAVLRSLTDQELSLLADLKARLEEAGP**EVQ**GHAEVAGGALF

Cluster 8: *Kitasatospora atroaurantiaca*

Precursor peptides:
1:MAENTEPTRDEA**DVV**AHSDEDTPWCVINTGTTKADEP**EVV**AHSDEDTPWCVINTGTTKADEETP**EVV**AHTADEDTPWCIINTGTSN
2:MTAPDPRTIAVLAAAGFPVDVLTDEQLQVFGSLSEEELALLLDIKARLDEVEP**EVQ**AHGTVAGGALF

Cluster 12: *Actinobacteria bacterium* OV320

Precursor peptides:
1:MAEAEQVPVQGDDDSTKVEKAAEVARAA**EVA**AHSLADAVLQLQRLDPEKEASIVTDPAGSGGSCHNISCF
2:MSEEIRPDAEPVEPAEAAEADG**DVE**AHLAAGQDAVLGLQKLTPEREASIISDPAGSGGSCHNISCF
3:MTPDSGEEGAVDGAGEHGAGSVLRALEQAGFPVRAFTGEQCAVFAGLTPQELELILDLKARLDAVEP**EVQ**AHTVVAGAALF

Cluster 16: *Streptomyces scabichelini*

Precursor peptides:
1: MPNPIEELRSAGVPVDSIPEEQREVLANLSEHEVGVLKGLKDRLGQ**DVQ**AHGAEGAEGIGVILW
2: MSKNDAVAALSAAGIPMDSLTDEQKSKLAQLSTEEIAVLAAVKGRLDAVT**DVQ**AHGAEGAEGTGIIIW
3: MAKNDALSALAAAGIPTDGLTDEQREQLSRLSPQEIAVLASVKGRLEAVS**DVQ**AHGAEGAEMGIIVW
4: MSDVERHPNLDALRAAGFPVDSIAQEQRHVLDSLTAGEVSVLIAVQERLQEAVP**EVQ**AHSAEPIIGGVLF

Cluster 22: *Catellatospora* sp. TT07R-123

Precursor peptides:
1:MTNIELTPGEQPEEVAALQETPGVEGDD**EVQ**GHI**S**TV**S**LAV**C**MAAD
2:MTNDELVARWKTPGAADAGHPAGEVLPRAVRPIGGRALLLAGQPINGAQTSGDSWTLSPSPFTMDFSL

Cluster 43: *Virgisporangium aliadipatigenens*

Precursor peptides:
1:MPENTEPQESEDVLALQETPAEDD**DVE**AHGVAPGEGAV**S**TL**S**LEH**C**S
2: MTDGAYTVEELEAHGFPVHDLTEEQRDVLRNLTPEEFAILVDVRARLEAVGP**EVQ**AHAEIAGAALF

Cluster 47: *Streptomyces* sp. R302

Precursor peptides:
1:MQISHSVTHVDGAQENTVQGHPAGEVTIGVGGRLGMRSRLLRASAADQAGYSADMPFTTFTITCAD
2:NQEIARRLHISPR**TVE**KHLASLLAKTGRADRAAL**C**AL**S**AE**S**AEQD
3:MSENTQPTEETP**EVE**AHSMVGENATAPEAFDDAAEIICGVYDKEVQA
4:MREPPDMLDTLYRAGFAVDLLTEEQRQVLTELTPEELAVLLDVKNRLDALGP**EVE**AHGEFAGGALF

Figure 35. Genomic Neighborhood Diagrams (GNDs) from twelve representative clusters encoding hypothetical mixed RiPP systems. Precursor peptide genes are marked in the same color as their predicted modifying enzyme. Genes encoding biosynthetic proteins related to mixed RiPP systems are highlighted. Blue: LanM synthetases and LanJ oxidoreductases. Red: LanKC lanthipeptide synthetases. Green: hypothetical peptide-processing system consisting of a radical SAM protein, an iron-containing redox enzyme and a MrsA5-like precursor peptide. Yellow: S8 protease. Pink: other enzymes potentially involve in peptide maturation. In cluster 5, a gene encoding a non-ribosomal peptide synthetase (NRPS) is also highlighted. For each cluster, precursor peptide sequences are displayed; cleavage sites are underlined, and the EVE-like recognition element is in bold. Residues forming the motif for labionin formation are marked in red color.

Cluster 1 contains 89 members homologous to the mrs_{S47} BGC from *S. griseus* S4-7. 88 of them are present in *Streptomyces* strains and only one is from another bacterium, *Kitasatospora albolonga*. In addition to four clearly identified precursor peptides (corresponding to MrsA1-A4), all members of this cluster encode a class II lanthipeptide synthetase (MrsM), a Gfo/Idh/MocA-like oxidoreductase (MrsJ), a class III lanthipeptide synthetase (MrsKC), and the highly conserved region encoding an rSAM/SPASM enzyme (MrsY), an iron-containing redox enzyme (MrsR), an S8 peptidase (MrsP), and a MrsA5-like precursor peptide.
Cluster 2 contains 35 members (all from *Streptomyces*) that seem overall closely related to those in cluster 1, the main difference is the oxidoreductase. Only six BGCs contain a Gfo/Idh/MocA-like oxidoreductase gene, but it is located on the opposite side of the BGC (Figure 35). In contrast, the other 25 BGCs encode a short-chain dehydrogenase, and a monooxygenase FAD-binding protein. Interestingly, these BGCs also encode a (2Fe-2S)-binding protein, which is another subfamily of rSAM proteins. Members of cluster 2 might represent a variation of the mixed RiPP system encoded in the mrs_{S47} BGC.
Cluster 3 displays 24 members that lack the class III lanthipeptide synthetase gene and harbor a different subclass of NADP-dependent oxidoreductase and a dehydrogenase encoding genes. These BGCs were particularly interesting as they additionally encoded a second rSAM protein, a dehydratase similar to that found in thiopeptide BCGs, a glycosyltransferase, and an S9 propyl endopeptidase. There were no potential precursor peptides annotated that could be modified by these enzymes. However, it is possible that these were not detected during analysis due to their short length.
It is possible that the peptide modified by the rSAM/SPASM and the iron-containing redox enzyme plays a regulatory role in the biosynthesis of co-clustered metabolites. However, the presence of a gene encoding a non-ribosomal peptide synthase (NRPS) in members of cluster 5

suggests this regulation is not limited to RiPPs (Figure 35). The BGCs in cluster 5 also encode a second rSAM protein, a cytochrome P450, and an S9 peptidase.

Cluster 7 clearly represents a two-component class III lanthipeptide system. These BGCs encode a LanKC synthetase and two precursor peptides. Both precursors harbor the LanKC binging site and the EVE-like recognition element for leader removal by the LanP protease. The peptides also contain the classic $S_{x2}S_{x2\text{-}5}C$ sequence for labionin formation and the spatial disposition of LanKC and LanP binding sequences suggests biosynthesis of these peptides follows a classic peptide processing where the LanKC installs a labionin and then the leader peptide is removed by the LanP protease.

The BGCs in cluster 8 do not encode lanthipeptides. The precursor peptide found in the representative diagram is enigmatic as it seems to harbor multiple EVE-like motifs along its whole sequence. For some RiPPs such as cyanobactins, dikaritins, amatoxins, orbitides, lyciumins, and cyclotides, the core peptide is flanked on one or both sides by short recognition sequences (could be also seen as leader sequences) that are important for post-translational modification. These classes of RiPPs have multiple copies of the core peptides that can be highly diverse in sequence and yield a group of different RiPPs (Montalbán-López *et al*., 2020). The peptide encoded in BGCs from cluster 8 is likely modified by the radical SAM protein and the cytochrome P450 monooxygenase encoded within the BGC. Modified peptides could also undergo modifications by the prolyl oligopeptidase encoded near the precursor peptide (Figure 35).

The cluster 12 BGCs seem to encode class III lanthipeptides. The two precursor peptides share near identical core sequences and lack the motif for labionin formation. It is then probable these peptides yield a single product containing (me)lanthionin crosslinks. A gene encoding a luciferase family oxidoreductase is found in these clusters; this enzyme could convert dehydroamino acids into D-amino acids thus being the first class III lanthipeptide containing D-amino acids.

Singleton 16 encodes multiple precursor peptides. Peptide sequences did not fit with a lanthipeptide precursor, still they contained the EVE-like motif.

The singleton number 22 clearly delineates a mixed RiPP system different to the *mrs*$_{S47}$. In this BGC, a class II lanthipeptide synthetase is encoded along with its cognate substrate which seems to yield a linear peptide, as no cysteines are present at the *C*-terminus. However, there is no oxidoreductase encoded in the neighborhood. A second precursor peptide is encoded next to its LanKC modifying enzyme. In contrast to the peptides found in the *mrs*$_{S47}$ BGC, this precursor contains the classic $S_{X2}S_{X2\text{-}5}C$ motif for labionin formation and the LanKC binding site is found *N*-terminally to the LanP recognition sequence suggesting a traditional lanthipeptide maturation process. Interestingly, a cytochrome P450 that could further modify any of the peptides is located within the BGC. Also, two S8 proteases and two radical SAM enzymes appear next to the iron-containing redox enzyme. The presence of these enzymes in the BGC is enigmatic since they cloud either be a repeat or they may represent an even more complex peptide-processing system.

Conversely, singleton 43 outlines a simpler RiPP system. This BGC is found in the genome of *Virgisporangium aliadipatigenens* and encodes a LanKC modifying enzyme and a precursor peptide harboring the *N*-terminal LanKC binding site, a central EVE-like motif, and the $S_{X2}S_{X2\text{-}5}C$ sequence for labionin formation. No additional biosynthetic elements were found in the neighborhood when the diagram was explored up to 20 ORFs up- and downstream the radical SAM-encoding gene.

The last interesting BGC explored here is the singleton 47 from *Streptomyces* sp. R302. It encodes a class II lanthipeptide synthetase and a precursor peptide devoid of Cys at the *C*-terminus. A second encoded precursor harbors the EVE-like motif; however, the core sequence does not correspond to a lanthipeptide. According to the enzymes encoded near this precursor gene, it is possible to infer that the final product derived from this peptide likely represents a new class of

RiPPs. These enzymes are a radical SAM protein, a glycosyltransferase and a class I SAM-dependent methyltransferase.
Some conclusions could be drawn based on the analysis of these representative BGCs. First, the bioinfomatic analysis based on the radical SAM protein MrsY displayed the widespread occurrence of mixed RiPP systems in Actinobacteria. In all BGCs, the highly conserved operon encoding the rSAM/SPASM enzyme, the iron-containing redox enzyme, the S8 protease, and the unique precursor peptide would play a pivotal role in the function of divergent mixed RiPP systems. Second, the radical SAM/SPASM enzyme and the iron-containing redox enzyme may catalyze unprecedented PTMs in its cognate substrate peptide that still needs to be experimentally characterized. This observation rejects the hypothesis that the radical SAM protein encoded in members of the GCF146 modifies the lanthipeptides (Zhang *et al*., 2014). Third, the consistent presence of an EVE-like motif in the MrsA5-like peptides and in other precursor peptides (in MrsA3-A4 in the mrs_{S47} BGC, for instance) suggests both classes of RiPPs involved a maturation step mediated by the S8 protease, thus reinforcing the hypothesis that this enzyme (MrsP) is the key point to understand the convergence of these pathways.

3.3.1.6 Iron-containing redox enzyme -MrsR

The iron-containing redox enzyme encoded in the mrs_{S47} BGC was provisionally named MrsR. This 241 amino acid protein (IPR016084 (Haem_Oase)) belongs to a superfamily of enzymes containing a multi-helical structural domain with two repeats of a 3-helical motif (UniProt Consortium, 2023). This domain is found in both eukaryotic and prokaryotic heme oxygenases including the coenzyme PQQ (pyrrolo-quinoline-quinone) biosynthesis protein C (PqqC) (Magnusson *et al*., 2004). This cofactorless oxidase catalyzes the last step in PQQ biosynthesis. The reaction involves a ring closure and overall, eight-electron oxidation of its substrate (Bonnot *et al*., 2013).
As shown in the bioinformatic analysis of the MrsY, the MrsR enzyme is conserved encoded next to the rSAM/SPASM enzyme and the MrsA5-like precursor peptide. Thus, it can be inferred that this enzyme collaborates with the rSAM/SPASM to modify the peptide. Post-translational modifications mediated by MrsR are not necessarily homologous to those occurring during PQQ biosynthesis. Reconstitution of MrsY and MrsR activity *in vitro* or *in vivo* would reveal their role on the biosynthesis of this novel and highly widespread peptide.

3.3.1.7 Further insights into the highly conserved operon found in the mixed RiPP systems

Bioinformatic analyses of genomic context based on MrsY showed the existence of other mixed RiPP systems in Actinobacteria. Their biosynthesis appears to comprise two or more sub-pathways interconnected by an S8 protease. One sub-pathway is not conserved and yields a variety of RiPP subclasses; in the case of the mrs_{S47} BGC, the class III lanthipeptides MrsA3-A4 are encoded. Conversely, the second sub-pathway is highly conserved and involves the processing of a peptide (MrsA5) by action of the rSAM/SPASM protein and the iron-containing redox enzyme. Some clusters displayed a third sub-pathway such as that in mrs_{S47} where MrsM and MrsJ would yield a classic type II lanthipeptide and a linear peptide likely containing D-amino acids; this last pathway however did not show a direct connection with the S8 protease-mediated peptide maturation.
The fact that all detected clusters contain the MrsA5 sub-pathway, and the high similarity of the precursor peptides suggest that this system might be biologically relevant for the producers. Deeper exploration of the operon encoding the MrsA5 peptide showed that all BGCs from the SSN (Figure 33) also encode MrsY, MrsR and MrsP homologs. In addition, two other genes

encoding an adenylate/guanylate cyclase and a GAF domain (*mrs*675 and *mrs*665 in S4-7) are also highly conserved in the operon. This might suggest that these ORFs could encode proteins with regulatory functions. Since the MrsA5 peptide belongs to a family of uncharacterized peptides known as FxSxx-COOH (Haft & Basu, 2011), it was hypothesized the operon could encode a biological conflict system. Several conflict systems with novel thresholding mechanisms including the so-called peptide-processing biological conflict systems have been recently bioinformatically described (Kaur *et al*., 2020). The central elements of these systems are a protease and a peptide of the FxSxx-COOH family. The protease releases the *C*-terminal peptide which is supposed to be modified by a radical SAM/SPASM protein; the modified small peptide is proposed to function as the effector of the biological conflict (Kaur *et al*., 2020). To date, little is known about the deployment and function of these systems as they have not been experimentally characterized. However, comparative genomics and sequence/structure analyses have shown that some biological conflict systems that are under threshold-dependent regulation seem to be deployed after generation of a nucleotide-derived signal produced by a member of the cyclic nucleotide (cNMP)-generating cyclase family (Burroughs *et al*., 2015). GAF sensory domains are often associated with the binging of cyclic GMP (Kalia *et al*., 2013).
Considering the available information regarding theoretical activation of biological conflict systems (Kaur *et al*., 2020) and the regulatory cascades controlling RiPP biosynthesis (Bartholomae *et al*., 2017), in this thesis, a mechanism for deployment of the peptide-processing conflict system in S4-7 was hypothesized (Figure 36). The activation might be initiated by sensing an "invader"; this would trigger conformational changes, eventual dimerization and ultimate production of signaling nucleotides (cAMP and cGMP) by the adenylate/guanylate cyclase located at the *N*-terminus of Mrs675. The GAF domain, encoded downstream the MrsP (Figure 28), could serve as the sensor within the bacterial cell, and would control downstream signaling for expression of MrsP which, in turn, will mediate peptide-processing. MrsA5 would be modified at its *C*-terminus by MrsR and MrsY and then cleaved off by the protease to yield the mature RiPP that will work as the final effector (likely as "toxin"). The protease would mediate leader removal of not only MrsA5, but also the class III lanthipeptides MrsA3-A4, thus its expression being the "master switch" of the system. The mature peptide itself or even the leader peptide might eventually function as feedback regulator.

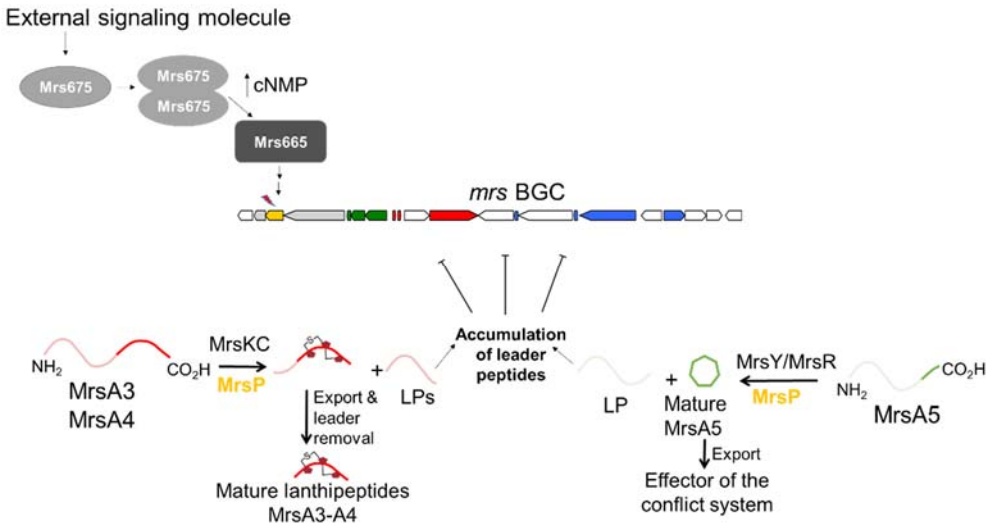


Figure 36: Proposed mechanism for activation of the peptide-processing biological conflict system found in the *mrs*$_{S47}$ BGC. After sensing of an "invader", the guanylate/adenylate cyclase Mrs675 will undergo a conformational change followed by eventual dimerization; this signaling hub provides nucleotides that will

be transmitted to the internal sensor GAF domain Mrs665. This would then mediate downstream transcription activation of biosynthetic genes. During biosynthesis, MrsP will cleave off peptides MrsA3-A5 thus releasing the leader peptide (or part of it in the case of MrsA3 and MrsA4). Accumulation of leader peptides could be sensed and used as a signal to stop biosynthesis thus impeding self-intoxication.

The study of the distribution of lanthipeptide-like BGCs in Actinobacteria proposed GCF146 to be a highly widespread hybrid class II/ class III lanthipeptide (Zhang *et al*., 2015). Since then, no experimental characterization of any member of this cluster family has been reported. Bioinformatic analyses conducted in the present thesis revealed these are not proper, "classic" lanthipeptide BGCs as they encode not only class II and class III lanthipeptides, but also linear peptides likely containing D-amino acids, and a peptide harboring unprecedented modifications likely installed by an rSAM/SPASM protein and an iron-containing enzyme, the last probably constituting a peptide-processing biological conflict.
Given the widespread of these complex mixed RiPP systems, the lack of understanding on their biosynthesis, and the relevance of the mrs_{S47} for the ecological role of *S. griseus* S4-7, the next step in this project was the *in vitro/in vivo* reconstitution of the main biosynthetic steps and the characterization of the final products.

3.3.2 Heterologous expression and purification of biosynthetic enzymes and precursor peptides in *Escherichia coli*

3.3.2.1 Expression and purification of lanthipeptide synthetases MrsM and MrsKC

mrsKC and *mrsM* were cloned into a pCDFDuet vector for expression in *E. coli* (section 6.2.2.6). Preliminary experiments using *E. coli* BL21 for expression of MrsKC yielded insoluble protein (according to section 6.2.5). The aggregation of expressed protein into inclusion bodies represented a challenge for further purification. Several strategies were proposed to obtain soluble protein. These included the usage of alternative *E. coli* strains such as Rosetta, which supplies tRNAs for AGG, AGA, ATA, CTA, CCC and GGA codons for expression of proteins that contain codons rarely used in *E. coli*. Another strain was Lemo21 which is particularly helpful for expression of challenging proteins. When the Lemo21 system is used, addition of L-rhamnose to expression cultures modulates the levels of lysozyme (*lysY*), the natural T7 RNA polymerase inhibitor, enabling fine tuning of T7 expression levels, and resulting in more soluble and correctly folded protein. However, after testing both strains and concentrations ranging from 0 to 2000 µM of rhamnose to induce the Lemo21 system, MrsKC was still only found in the insoluble fraction (data not shown).
Formation of inclusion bodies is indeed a common problem during recombinant protein expression in *E. coli*; multiple factors like high inducer concentrations, elevated temperature during expression, and the use of strong promoters that results in a high translational rate contribute towards formation of protein aggregates (Francis & Page, 2010). In this study, variation of temperature and IPTG concentration for induction of MrsKC expression were tested, but in all cases, it resulted in inclusion bodies formation (data not shown).
Alternatively, recovery and refolding of MrsKC from the insoluble fraction were tried following the protocols described by Singh *et al.*, 2015. In this case, cells were completely lysed, and inclusion bodies were recovered by centrifugation. Then, the pellet containing the isolated inclusion bodies was washed to remove intact cells that could contaminate the preparation. Protein was solubilized by adding high concentrations of urea and DTT; in this step however, the protein is desaturated. Several protocols are available for refolding; for MrsKC dialysis was preferred. Although the protein was isolated from inclusion bodies, MrsKC precipitated after recovery from

the dialysis chamber (data not shown). Refolding of denatured proteins by dialysis required long incubation periods to allow the protein to recover its tridimensional structure while denaturant is gradually removed from the buffer. Improper folding of MrsKC could lead to precipitation and the activity of the enzyme was not assured. Since refolding of MrsKC purified from inclusion bodies implied major optimization, no further efforts were done in this direction.
Finally, a new *E. coli* strain was tested for MrsKC expression. *E. coli* LOBSTR (Andersen *et al*., 2013) harbors extra copies of the *argU*, *ileY* and *leuW* tRNA genes for expression of proteins containing rare codons. Besides, LOBSTR presents mutations in ArnA and SlyD, the two most common *E. coli* contaminants, yielding proteins of higher purity. Even though expression of MrsKC in *E. coli* LOBSTR also resulted in inclusion body formation, a small fraction of soluble protein was obtained after optimization of some parameters such as IPTG concentration, OD_{600} and incubation time (Figure 37).

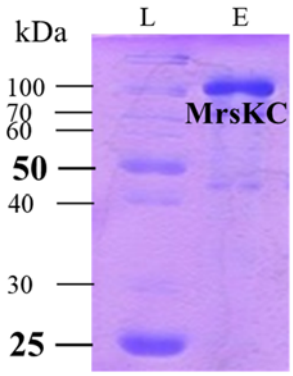


Figure 37: SDS-PAGE gel of purified MrsKC expressed in *E. coli* LOBSTR. **(E)** elution fraction. **(L)** Blue prestained protein standard (NEB).

Conversely, expression of MrsM in *E. coli* BL21 yielded soluble protein with no further optimization needed (Figure 38). MrsM was purified by affinity chromatography obtaining relatively low yields that can be attributed to the size of the protein (106 kDa). It has been long recognized that some targets, especially larger multidomain and membrane proteins, either fail to express or express and form inclusion bodies (Francis & Page, 2010). Still, production of MrsM through standard expression in *E. coli* was sufficient to obtain pure and functional enzyme to be used for enzymatic assays.

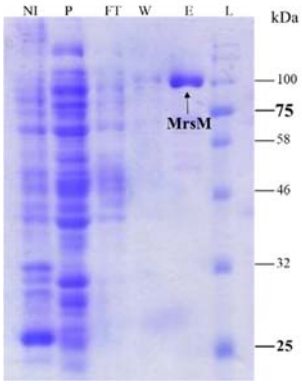


Figure 38. SDS-PAGE gel of purified MrsM expressed in *E. coli* BL21. **(NI)** Non-induced. **(P)** pellet (insoluble fraction). **(F)** flow-through. **(W)** wash I-II. **(E)** elution fraction. **(L)** Blue prestained protein standard (NEB).

Finally, both lanthipeptide synthetases were transferred to assay buffer containing 5% glycerol and stored at -70 ºC until use. Whenever possible, protein purification and *in vitro* assays were prepared on the same day since freeze-thawing and long-term storage resulted in protein degradation.

3.3.2.2 Expression and purification of MrsJ

Some proteins involved in the biosynthesis of D-amino acids have been studied in heterologous systems. In many cases however, reconstitution of their activity and substrate specificity *in vitro* have not been investigated due to unsuccessful attempts to express LanJs in *E. coli* (Yang & van der Donk, 2015). In this thesis, the gene encoding the oxidoreductase *mrsJ* was cloned into the pET28a vector and successfully expressed in *E. coli* BL21 under standard conditions (sections 6.2.2.6 and 6.2.5.1, respectively). The protein was purified from the soluble faction by affinity chromatography according to section 6.2.5.2 (Figure 39).

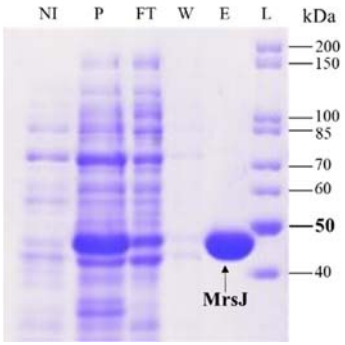


Figure 39. SDS-PAGE gel of purified MrsJ expressed in *E. coli* BL21. **(NI)** Non-induced. **(P)** pellet (insoluble fraction). **(F)** flow-through. **(W)** wash I-II. **(E)** elution fraction. **(L)** unstained protein standard, broad range (10-200 kDa), NEB.

3.3.2.3 Expression and purification of MrsP

The S8 protease MrsP was successfully cloned into pET28 vector and expressed in *E. coli* BL21. The most common IPTG concentrations for induction range from 0.1 to 1.0 mM. In our laboratory, a concentration of 40 µM IPTG for induction is used as standard concentration giving soluble proteins in most of the cases. However, this IPTG concentration for induction of MrsP expression resulted in a high transcription rate causing formation of inclusion bodies. Since T7 is a strong promoter, some translated proteins aggregate and accumulate in inclusion bodies because they fail to fold before encountering another unfolded protein (Francis & Page, 2010). To enhance the production of soluble protein, the use of weaker promoters could be tested and/or expression parameters optimized. In the case of MrsP, lowering the IPTG concentration to 0.1 mM was sufficient to obtain soluble protein (Figure 40), which was later proved to be active.

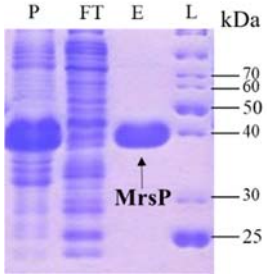


Figure 40. SDS-PAGE gel of purified MrsP expressed in *E. coli* BL21. **(P)** pellet (insoluble fraction). **(F)** flow-through. **(E)** elution fraction. **(L)** unstained protein standard, broad range (10-200 kDa), NEB.

3.3.2.4 Production of active rSAM protein MrsY

a) Expression of MrsY in *E. coli* under semi-anaerobic conditions

Although advances in recombinant technologies have greatly facilitated the expression of radical SAM proteins, their oxygen sensibility and complexity of their structures continue to make these enzymes challenging to work with. To express the radical SAM protein encoded in the mrs_{S47} BGC, several *E. coli* strains and culture parameters were tested without success. Since the nucleotide sequence contained multiple rare codons that are difficult to be used by *E. coli*, a codon optimized gene was synthesized (Supplementary Figure S12). The optimized gene was subcloned into pET28 and expression in *E. coli* was tested. Again, this resulted in no protein expression. As explained in section 1.3.5, radical SAM proteins require the installment of so-called Fe-S clusters for catalysis. To obtain folded protein containing such metalloclusters, radical SAM proteins are often co-expressed with either the SUF or ISC operon for *in vivo* assembly and loading of the clusters into the folded protein (Himes *et al*., 2016; Wecksler *et al*., 2009). Here, co-expression of the MrsY with the pPH151plasmid that contains the *E. coli suf ABCDSE* genes was tried (pPH151 was kindly provided by Dr. Petra Hänzelmann of the Rudolf Virchow Center for Experimental Biomedicine Institute for Structural Biology at the University of Würzburg). However, after several expression experiments, no protein was detected in fractions obtained from expression cultures.

Given their high oxygen sensibility, most radical SAM proteins require expression and downstream handling under anoxic conditions. There are however some exceptions like DarE, which is a radical SAM monooxygenase that uses oxygen as a co-substrate to install two ether crosslinks in darobactin biosynthesis (Nguyen *et al*., 2022). To investigate the oxygen sensibility of MrsY, small expression cultures under aerobic conditions were grown; after addition of IPTG, Erlenmeyer flasks were purged with argon to produce an oxygen-deficient environment. Incubation for protein expression was carried out overnight at 16 ºC. The MrsY protein was ultimately expressed only in the culture where argon was applied, confirming MrsY needs an anoxic environment to be expressed (Figure 41). To evaluate the effect of temperature in protein production, the same experiment was performed at 25 ºC and 30 ºC. While protein expression was completely abolished at 23 ºC, expressed protein was detected only in the insoluble fraction when temperature was increased to 30 ºC (Supplementary Figure S13). Production of MrsY was eventually upscaled to 500 ml cultures to produce sufficient protein for *in vitro* assays.

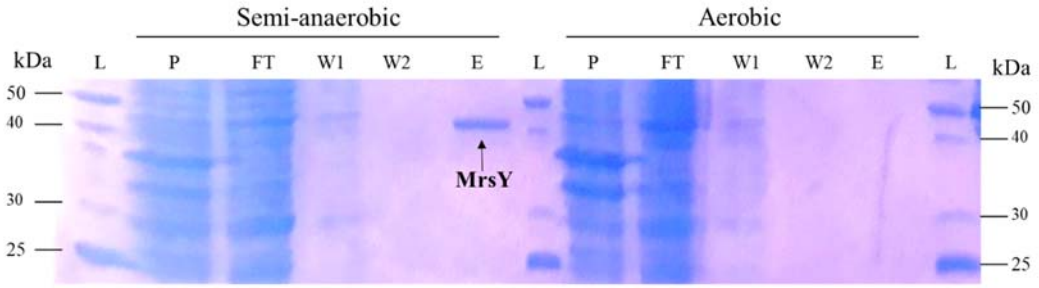


Figure 41. SDS-PAGE of purified radical SAM protein MrsY expressed in *E. coli* Rosetta under aerobic and semi-anaerobic conditions. **(P)** pellet (insoluble fraction). **(F)** flow-through. **(W1)** wash I. **(W2)** wash II. **(E)** elution fraction. **(L)** unstained protein standard, broad range (10-200 kDa), NEB.

b) Protein purification and reconstitution of Fe-S clusters under anaerobic conditions

Given the instability against oxygen and metal-chelating agents, rSAM proteins can be purified only in their *apo* form after expression in *E. coli*. Then, the [4Fe-4S] clusters must be faithfully

reconstructed under strict anaerobic conditions. To study the function of the radical SAM protein MrsY in the mrs_{S47}, the protein was purified under strict anaerobic conditions and [4Fe-4S] clusters were chemically reconstituted (in collaboration with Prof. Dr. Dahl at the Institute for Microbiology & Biotechnology, University of Bonn). Protein purification was carried out according to section 6.2.5.3. To avoid degradation and precipitation, pure protein was transferred to assay buffer and stored at 4 ºC for a maximum of one day before reconstitution. Chemical reconstitution consisted of two main steps: a prolonged reduction of apoprotein with DTT, and iron-sulfur cluster assembly by addition of $FeCl_3$ and Li_2S followed by incubation for 2-3 hours (section 6.2.5.4). Figure 42 shows the UV spectra before and after reconstitution where the peak at 420 nm indicates the presence of Fe-S clusters after 1.5 hours. Unbound iron and sulfide were removed via desalting with a PD-10 column using assay buffer, and the holoprotein was concentrated to be used for *in vitro* assays.

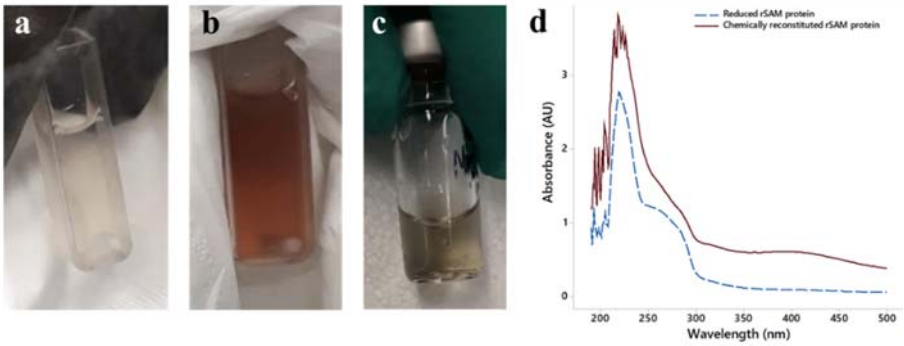


Figure 42. Chemical reconstitution of the Fe-S cluster in the rSAM protein MrsY. **(a)** protein upon reduction with DTT, **(b)** protein after addition of $FeCl_3$, **(c)** reconstituted MrsY, and **(d)** UV-visible absorption spectra of reduced protein (discontinuous blue line), and Fe-S cluster-reconstituted protein (brown).

3.3.2.5 Expression and purification of the iron-containing redox enzyme MrsR

The iron-containing redox enzyme MrsR was cloned into the pET28 expression vector. *E. coli* LOBSTR and Rosetta were tested as hosts for expression under aerobic or semi-anaerobic conditions. Determining if MrsR is produced under these conditions was important to design further experiments where the activity of MrsR and MrsY could be studied *in vivo*. No differences were observed in biomass production and the protein was recovered in the soluble fraction after purification (Figure 43).

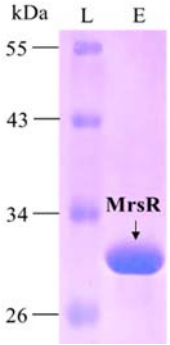


Figure 43: SDS-PAGE gel of purified MrsR expressed in *E. coli* BL21. **(E)** elution fraction. **(L)** color prestained protein standard, broad range (10-250 kDa), NEB.

3.3.2.6 Precursor peptide production, purification, and mass analysis

To obtain the peptides to be used as substrate for *in vitro* assays, precursor genes were individually cloned into pET28a vectors fused to a poly-histidine tag for purification by affinity chromatography (section 6.2.2.6). Several *E. coli* strains were initially tested for efficient expression to obtain the highest titers. Peptides were produced under all conditions, but yields varied depending on the strains and systems used. For example, induction of the Lemo21 system for expression of MrsA3 showed a decline in the expression levels when high concentrations of rhamnose were applied (Figure 44). Peptides were ultimately purified from the soluble fraction with no additional optimization. During purification, peptides were eluted using high concentrations of imidazole; as this can interfere in the downstream assays, peptides were transferred to an appropriate buffer by using a PD-10 column. The mass of each peptide was experimentally determined by MALDI-ToF MS (matrix-assisted laser desorption/ionization time-of-flight mass spectrometry) measurements in linear mode (Supplementary Figure S14). In all cases, observed masses were in accordance with the theoretical average masses.

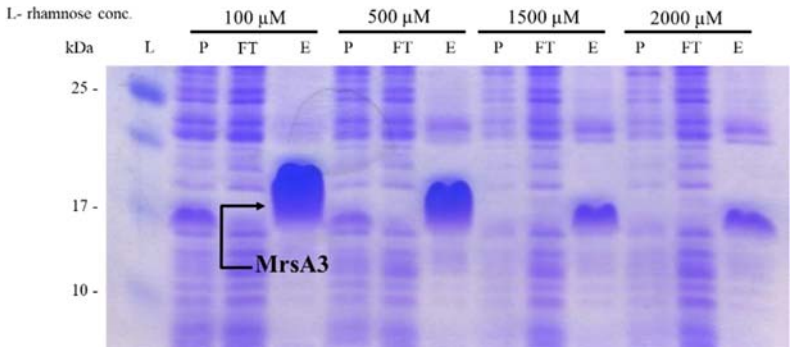


Figure 44. Optimization of precursor peptide expression in *E. coli*. SDS-PAGE gel of purified MrsA3 expressed using different L-rhamnose concentrations for induction of the Lemo21 system. Increased concentration of inducer caused a dramatic drop in peptide expression. **(S)** soluble fraction. **(P)** non-soluble fraction. **(E)** elution fraction. **(L)** unstained protein standard, broad range (10-200 kDa), NEB.

3.3.3 *In vitro* assays for reconstitution of the mrs_{S47} biosynthesis

3.3.3.1 MrsM modifies MrsA1 and MrsA2 to produce two subclasses of RiPPs

The first *in vitro* assays aimed to prove that MrsM is the modifying enzyme of MrsA1 and MrsA2. Pure MrsM synthetase was mixed with MrsA1 or MrsA2 in a reaction buffer containing magnesium chloride and ATP (section 6.2.6). Some precursor peptides enable direct cleavage by commercial proteases to remove most of the leader peptide without proteolysis in the core region, but others require the insertion of an artificial cleavage site. To demonstrate that modifications by MrsM occurred in the core peptide and to identify the positions where the modifications were installed, an artificial cleavage site for leader removal using Factor Xa was inserted after the Gly-Gly motif in MrsA1 (Figure 45A). For MrsA2, trypsin was used to obtain a fragment of 2,620.20 Da which included the core peptide (Figure 46A; Supplementary Table S1). An $[M+4H]^{4+}$ ion at *m/z* 844.4092 was obtained after incubation of MrsA1 and MrsM. This mass corresponded to the four-fold dehydrated core peptide ($MrsA1^{M}$*); comparison of MS/MS spectrum with that of the unmodified core peptide (MrsA1*) used as control (Figure 45B) confirmed this ion as the modified peptide (Figure 45C). The modified peptide was then treated with iodoacetamide (IAA) to identify the number of thioether crosslinks. LC-MS analysis of the derivatized peptide ($MrsA1^{M}$*-IAA) showed an ion indicating the alkylation of three thiol groups (Figure 45D). These results indicated that cyclization did not take place despite dehydroamino acids were formed.

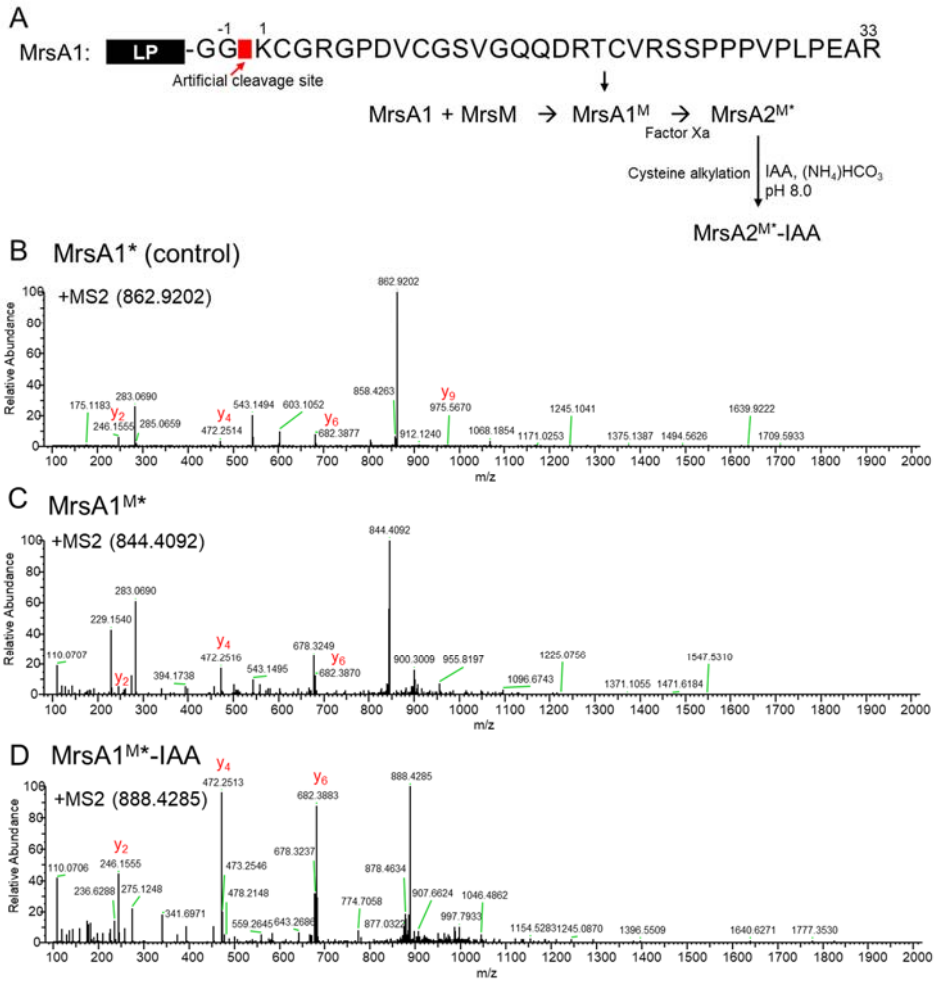


Figure 45: The class II lanthipeptide synthetase MrsM installs four dehydroamino acids in MrsA1. (**A)** Amino acid sequence of MrsA1 (artificial cleavage site for Factor Xa is indicated) and order of *in vitro* reactions. (**B)** MS/MS spectrum of the unmodified MrsA1* core peptide used as control (calculated $[M+4H]^{4+}$ ion at *m/z*

862.93). **(C)** MS/MS spectrum of the MrsM reaction product MrsA1^{M}* core peptide (calculated [M+4H]$^{4+}$ ion at *m/z* 844.92). (D) MS/MS spectrum of the derivatized MrsA1^{M}*-IAA (calculated [M+4H]$^{4+}$ ion at *m/z* 887.67). The * indicated the peptide was treated with Factor Xa prior LC-MS/MS analysis.

On the other hand, incubation of MrsA2 with MrsM for six hours at room temperature caused a mass shift that indicated the loss of three molecules of water. Nonetheless, when incubation was extended to 12 hours, one additional dehydration was detected. Different experimental parameters such as buffer system, incubation time and protein:substrate ratio were tested to ensure full conversion of MrsA2. Formation of four dehydroamino acids to yield MrsA2^{M} was ultimately confirmed by analysis of the reaction product through mass spectrometry (Figure 46B-C). The MS/MS spectrum of MrsA2^{M} after tryptic digest (indicated with an asterisk) confirmed the formation of one dehydroalanine (Dha) and three dehydrobutyrines (Dhb); their corresponding positions in the core peptide were assigned according to the fragmentation pattern observed (Figure 46D).

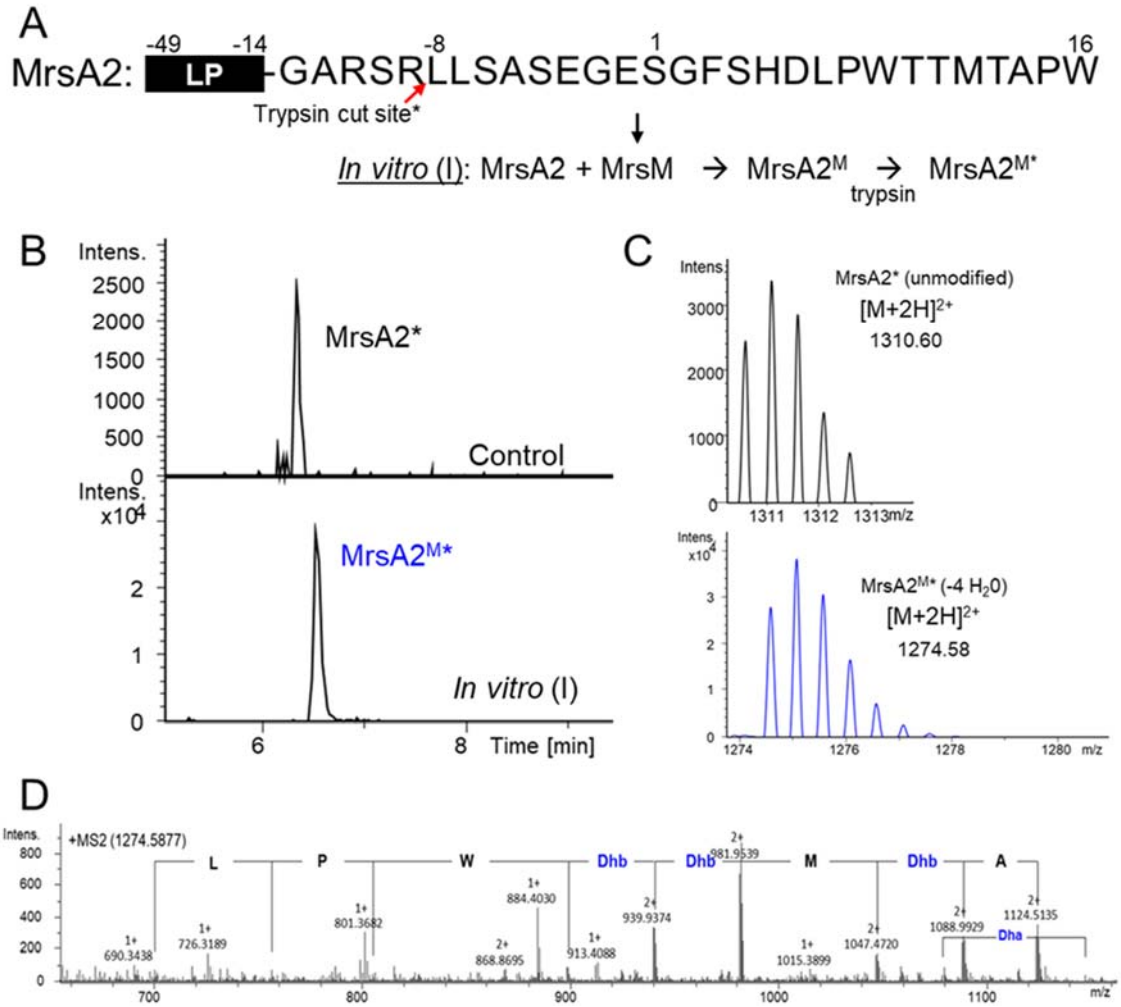


Figure 46. The class II lanthipeptide synthetase MrsM installs four dehydroamino acids in MrsA2. (**A)** Amino acid sequence of MrsA2 (trypsin cut site is indicated) and order of *in vitro* reactions. (**B)** Extracted ion chromatogram (EIC) of MrsA2* used as control (upper panel) and the MrsM reaction product (MrsA2^{M}*, lower panel). The asterisk indicates that the peptides were digested with trypsin prior to analysis. (**C)** High resolution mass spectra of MrsA2* and MrsA2^{M}* showing a mass shift corresponding to the loss of four molecules of water. (**D)** MS/MS spectrum of MrsA2^{M}* displays the residues converted into Dha and Dhb upon incubation with MrsM. (Calculated monoisotopic masses for MrsA2*: 1310.60 and MrsA2^{M}*: 1274.58. For indicated ion fragments *z*=2).

3.3.3.2 MrsJ converts Dha and Dhb into D-ala and D-amino butyric acid

In vitro reconstitution of MrsJ activity consisted of incubation of MrsA2^{M} with purified MrsJ at room temperature (section 6.2.6). A second sample lacking MrsJ was prepared as negative control (Figure 47A). After 12 hours of incubation, samples were treated with trypsin and analyzed by HPLC-MS/MS. A new peak that had increased in mass by 6 Da was observed indicating the reduction of three dehydroamino acids (Figure 47B-D). Further analysis of the MS/MS spectrum confirmed the reduction of one Dha and two Dhb residues (Figure 47D). This modified peptide was named MrsA2$^{M/J}$.

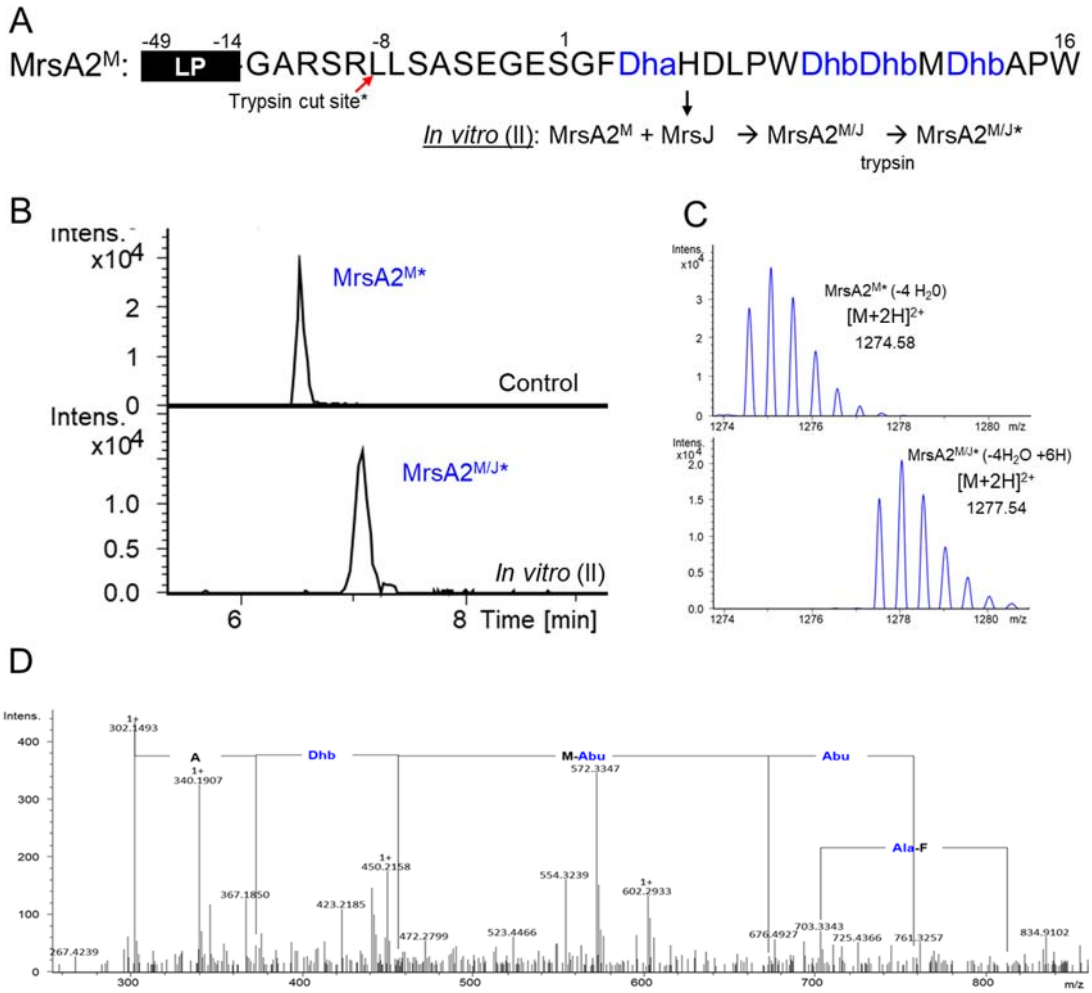


Figure 47. MrsJ reduces four side chains in MrsA2^{M}. **(A)** Amino acid sequence of MrsA2^{M} showing the dehydroamino acids formed by MrsM (trypsin cut site is indicated) and order of *in vitro* reactions. **(B)** Extracted ion chromatogram (EIC) of MrsA2^{M} used as control (upper panel) and the MrsJ reaction product (MrsA2$^{M/J}$, lower panel). The asterisk indicates that the peptide was digested with trypsin prior to analysis. **(C)** High resolution mass spectra of MrsA2^{M}* and MrsA2$^{M/J}$* showing a mass shift corresponding to +6 Da. **(D)** MS/MS spectrum of MrsA2$^{M/J}$* displays the residues that were converted into D-Ala and D-amino butyric acid by the action of MrsJ. (Calculated monoisotopic masses for MrsA2^{M}*: 1274.58 and MrsA2$^{M/J}$*: 1277.58. For indicated ion fragments z=1).

According to the residues modified by MrsM and MrsJ, a position for insertion of an artificial cleavage site in MrsA2 was selected. The coding sequence for the sequence IEGR was inserted at that position for leader removal using the commercial protease Factor Xa in the same way as for MrsA1. The new precursor peptide variant was called MrsA2$_{xa}$. The presence of a Factor Xa recognition sequence in MrsA2$_{xa}$ did not affect expression nor yield of pure peptide, and proteolysis with Factor Xa was overall

efficient. To increase the yields of modified peptide and to shorten the experimental workflow, MrsM and $MrsA2_{xa}$ were co-expressed in *E. coli* LOBSTR. The modified peptide was purified as described in section 6.2.2.6 and then incubated with Factor Xa for leader removal. Analysis through MALDI-ToF MS of the digested peptide (core-$MrsA2_{xa}^{M}$) proved the installment of four dehydroamino acids and the correct leader removal (Supplementary Figure S15). Formation of four dehydroamino acids *in vivo* supported the results obtained during *in vitro* assays. The modified variant ($MrsA2_{xa}^{M}$) was employed as new substrate for confirmation of MrsJ activity. After *in vitro* reaction of $MrsA2_{xa}^{M}$ with MrsJ, the product $MrsA2_{xa}^{M/J}$ was mixed with 2 mM $CaCl_2$ and 1 µl of factor Xa, and incubation continued for 6 hours, that was the optimal incubation time for Factor Xa to remove the leader. The reaction product was named core- $MrsA2_{xa}^{M/J}$ and was immediately analyzed by MALDI-ToF MS (Supplementary Figure S15). Results confirmed the reduction of three dehydroamino acids and provided the mass of the final product. Even though the presence of the Factor Xa sequence did not abrogate MrsJ activity, a decrease in the conversion ratio was clearly observed. After several attempts to produce sufficient modified peptide, it was not possible to enhance the yields, thus impeding the purification of sufficient amount for determining the stereochemistry of the MrsJ products. It was inferred that the presence of the artificial cleavage site may have reduced the affinity of the substrate, leading to a decrease in the production of the final peptide. This also affected peak resolution on the HPLC and prevented the generation of a MS2 spectrum for further confirmation. For this reason, the MrsA2 peptide (without the artificial cleavage sequence) was preferred for upscaling production, purification and ultimate analysis of configuration.

The stereochemistry of alanine and aminobutyric acid residues formed in $MrsA2^{M/J}$ by the reductase MrsJ will be determined by Marfey's assay. FDAA (1-fluoro-2-4-dinitrophenyl-5-L-alanine amide) is a derivatization reagent that reacts with primary amines for UV detection. Derivatives can be detected at 340 nm with nanomole sensitivity (Marfey, 1984). FDAA derivatives of D-amino acids exhibit strong intramolecular binding thus reducing their polarity relative to the corresponding L-amino acids. Due to the difference in polarity, D-derivatives are retained on the column and elute later than the corresponding L-derivative; thus, FDAA derivatized amino acids can be separated by HPLC with no need of a chiral column (Figure 48).

$MrsA2^{M/J*}$ will be purified by RP-HPLC and then subjected to acid hydrolysis. Derivatization of free amino acids and the corresponding standards (L-Ala, D-Ala, L-Abu and D-Abu) as well as the analysis by HPLC-DAD-MS is currently in progress in our laboratory.

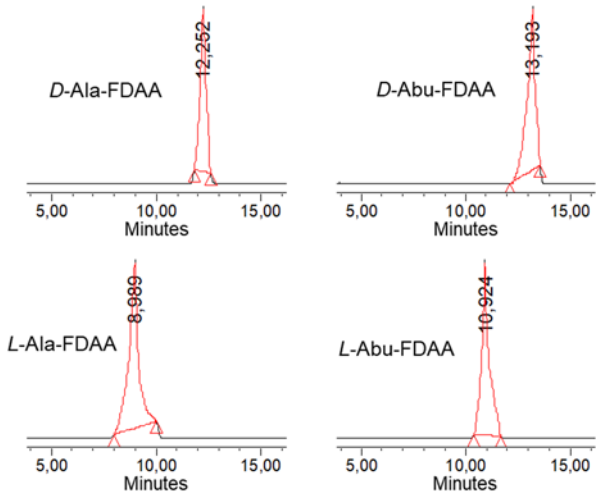


Figure 48: Extracted chromatograms of derivatized amino acids. Standards were derivatized with 1-fluoro-2-4-dinitrophenyl-5-L-alanine amide (FDAA) according to Marfey (1984) and then separated on a C18 HPLC column and detected at 340 nm.

3.3.3.4 MrsKC catalyzes labionin formation in MrsA3-A4

The labionin formation in MrsA3-A4 by MrsKC was reconstituted in a reaction that required a mixture of NTPs and magnesium chloride. Linear peptides MrsA3 and MrsA4 (Figure 49A) were incubated independently with MrsKC; reactions lacking the enzyme were prepared as negative control (Figure 49B). After six hours of incubation at room temperature, pure MrsP was added to the reaction and incubation continued for one hour. Of note, reconstitution of labionin formation was carried out after the activity of MrsP was studied. Therefore, the MrsP could generically be used as protease for leader removal while studying MrsKC activity. The reaction products were analyzed by HPLC-MS/MS to confirm labionin formation (Figure 49C-D). As anticipated, MrsP was able to cleave MrsA3-A4 after labionin formation. Cyclization was confirmed by alkylation with iodoacetamine (IAA) of MrsA3^{KC}. No mass shift was observed in the modified peptide after one hour of incubation with 20 mM of IAA (Supplementary Figure S16). MrsA3 contains two threonine residues that are also susceptible for dehydration, but none of them were modified. Labionin formation was further confirmed by co-expression of MrsKC with a precursor peptide carrying the Factor Xa cleavage site for removal of the whole leader peptide. MS/MS fragmentation indeed showed labionin formation *in vivo* (Supplementary Figure S16).

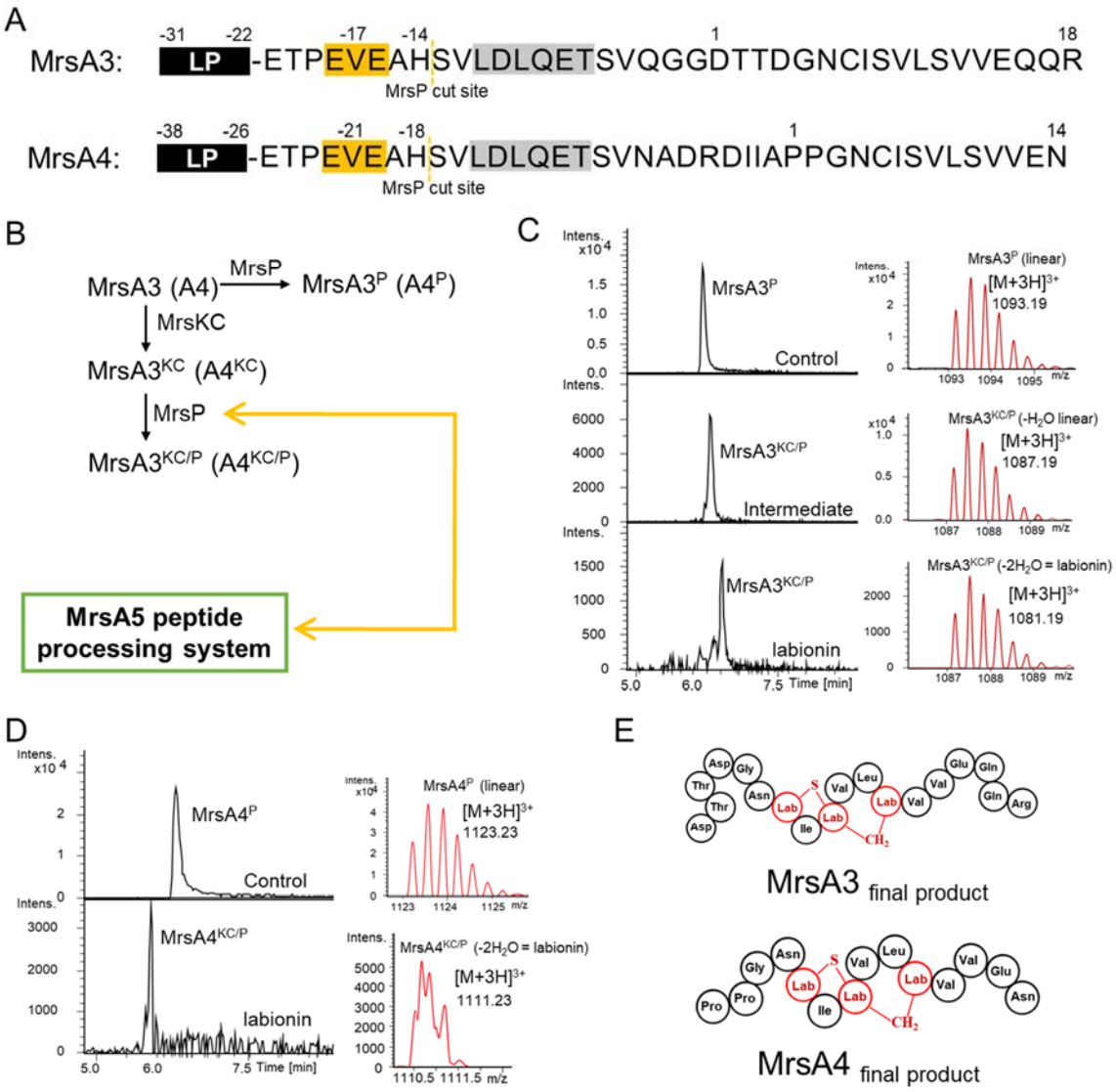


Figure 49. *In vitro* reconstitution of labionin formation in MrsA3 and MrsA4. **(A)** Precursor peptide sequences indicating the MrsP recognition element (yellow box) and cut site. MrsKC binding site is highlighted in a gray

box. **(B)** Order of reaction for reconstitution of labionin formation and proteolysis. **(C)** Extracted ion chromatograms (left panel) and high-resolution mass spectra (right panel) of linear MrsA3, intermediate containing one Dha, and peptide containing labionin (MrsA3^{KC}). **(D)** Extracted ion chromatograms (left panel) and high-resolution mass spectra (right panel) of linear MrsA4 and peptide containing labionin (MrsA4^{KC}). In all cases, MrsP was used to trim off the leader peptides (indicated with a ^{P}). **(E)** Structures proposed for the final products of MrsA3 and MrsA4. (Calculated monoisotopic masses: 1093.19 Da for MrsA3^{P}, 1081.19 Da for MrsA3$^{KC/P}$, 1123.23 Da for MrsA4^{P}, and 1111.23 Da for MrsA4$^{KC/P}$).

Throughout MrsA3 modification, an intermediate harboring a single dehydroamino acid at position 8 was detected (Figure 49c). Labionin formation most likely follows a *C*-to-*N*-terminal directionality where the Ser8 residue is the first to be converted into Dha. This would favor thioether formation and the subsequent nucleophilic attack to the second Dha. The *C*-to-*N* directionality also aligns with the inverted motif (CSx_2S_{X2-5}) observed in these class III lanthipeptides (Figure 49e).

3.3.3.5 MrsP removes leader peptide in class III lanthipeptides MrsA3-A4 and in MrsA5

To date, only one S8 protease has been reported to participate in class III lanthipeptide maturation. AmyP removes LP in AmyA peptides in a site-specific manner, and its substrate recognition is independent of the modifications of the precursor peptide (Zhang *et al*., 2022).

The presence of the EVE motif in MrsA3 and MrsA4 suggested that MrsP was involved in leader removal during the biosynthesis of these class III lanthipeptides. Hence, unmodified MrsA3-A4 were incubated with MrsP at room temperature for 2, 12, and 20 hours. A reaction containing MrsA2 as substrate was also included to further prove the specificity of MrsP. Samples were visualized on acrylamide gels where two peptide fragments in MrsA3 and MrsA4 but not in MrsA2 samples were clearly observed (Supplementary Figure S17). MALDI-ToF MS analysis confirmed the excision of MrsA3-A4 and allowed the identification of the cleavage site (Figure 50). MrsP cleaved the peptides exactly two amino acids after the EVE motif, being in accordance with the mechanism described for AmyP. Further variations of enzyme:substrate ratio and incubation time showed that MrsP fully digested linear peptides even after 15 minutes of incubation. These findings promoted MrsP as the peptidase responsible for leader removal in MrsA3-A4. *In vitro* reconstitution of MrsP activity also expanded the tolerance of the amino acid preceding the EVE motif since MrsA3-A4 here harbor a proline residue instead of T/S/A. The activity of MrsP on peptides containing labionin (MrsA3^{KC} and MrsA4^{KC}) described in section 3.3.3.4 also aligned with these results and confirmed that MrsP proteolysis is independent of the modifications in the core peptide.

As predicted during the bioinformatics analysis, MrsP trims off the peptides by keeping the MrsKC binding site attached to the core; this observation along with the fact that leader peptides in MrsA3-A4 are unusually large evoked the questions of what the actual order of PTMs was, and whether additional modifications precede labionin formation. This phenomenon could indeed be related to the biosynthesis of MrsA5.

As mentioned in section 3.3.1.5, MrsA5 was only recognized as part of the mixed RiPP system when bioinformatics analysis of the rSAM protein MrsY showed this peptide to be conserved across the BGCs. The presence of the EVE-like sequence in MrsA5 was hypothesized to be a variation of the MrsP binding site. To investigate whether MrsP could recognize this peptide as substrate, pure MrsA5 was incubated with MrsP for one hour, and the reaction products were directly analyzed through MALDI-ToF MS (Figure 50). Results confirmed that cleavage of MrsA5 occurred two residues after the EVE-like motif giving a core peptide with a mass of 833.43 Da. These results further expand the role of S8 proteases in RiPPs maturation and put forward leader removal as the crucial step in the biosynthesis of the newly discovered mixed RiPP systems.

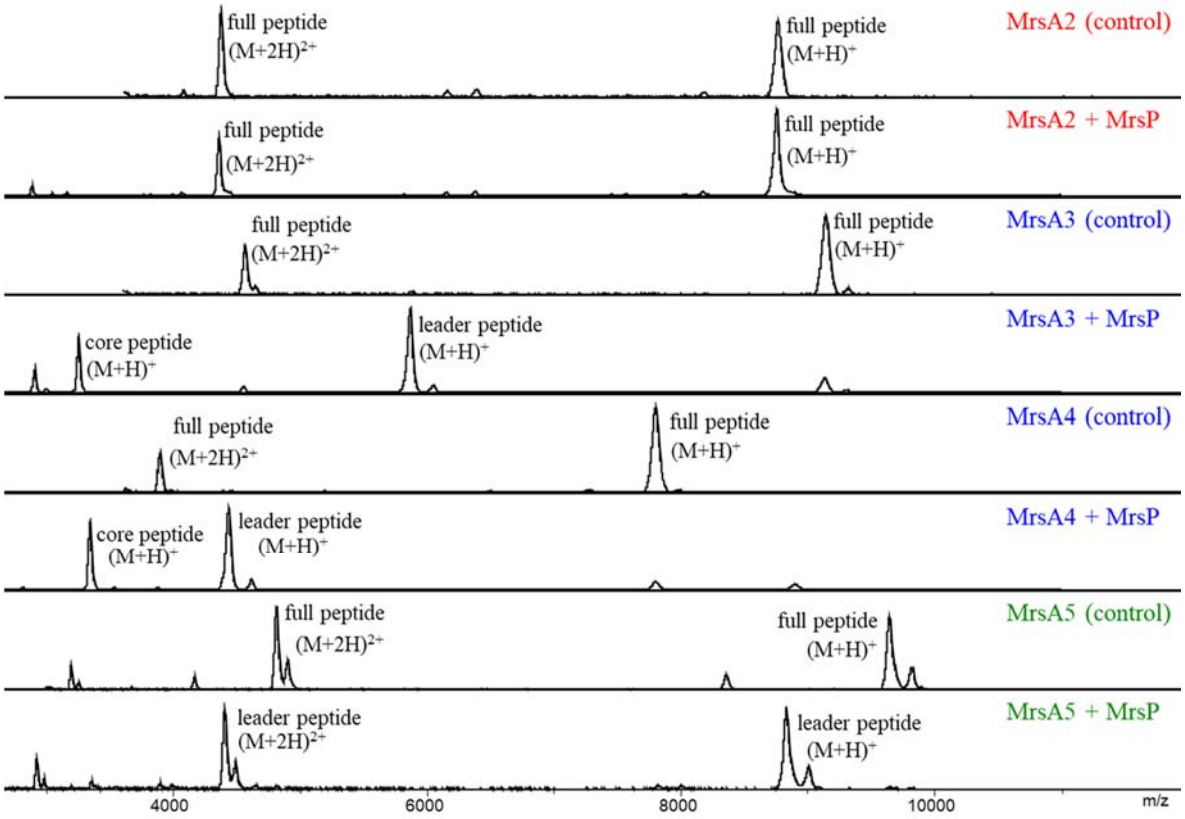


Figure 50. *In vitro* reconstitution of MrsP-mediated proteolysis. MALDI-ToF mass spectra obtained in linear mode showed that MrsA2 remained intact after incubation with the S8 protease MrsP while linear peptides MrsA3, MrsA4 and MrsA5 were trimmed off. For MrsA5, only the leader peptide is shown as the core peptide has a mass of 833.73 Da and can only be measured via reflector mode.

3.3.3.6 *In vitro* reconstitution of the rSAM/SPASM MrsY and MrsR activity

The activity of MrsY was initially tested *in vitro*. After expression in *E. coli*, the protein was purified under strict anaerobic conditions followed by chemical reconstitution of the Fe-S cluster (section 3.3.2.4). *In vitro* reactions were prepared and incubated at room temperature under anaerobic conditions. These contained MrsY and MrsA5 (1:10) mixed in a reaction buffer supplemented with sodium dithionite and 5 µM of *S*-adenosyl-L-methionine (SAM). After five hours of incubation, MrsP was added for leader removal. Results from the LC-MS/MS analysis showed an ion corresponding to the linear MrsA5 core peptide, but no modifications were detected. Multiple variables such as the buffer system, enzyme: substrate ratio and incubation time are critical for *in vitro* reconstitution of enzymatic activity. One factor that dramatically influences the performance of radical SAM proteins is the reducing agent. Although sodium dithionite is routinely used as chemical reductant, studies have demonstrated that the flavodoxin, flavodoxin reductase and NADPH system is more effective (Feng *et al*., 2013). This natural reduction system, for instance, led to a 100% increase of NosL activity *in vitro* (Zhang et al., 2011). The activity of the radical SAM decarboxylase BlsE involved in the biosynthesis of blasticidin S increased by 150% when the flavodoxin reductase system was used (Feng *et al*., 2013). Iron-sulfur clusters can be observed in four different oxidation states. The intact cluster exists mainly as $[4Fe\text{-}4S]^{2+}$ in the purified enzyme, and as $[4Fe\text{-}4S]^{3+}$ as minor component. When the enzyme is treated with sodium dithionite in the presence of SAM, clusters are reduced to $[4Fe\text{-}4S]^{1+}$, which is the catalytic form of all radical SAM enzymes. Catalysis is then initiated by one electron transfer from the $[4Fe\text{-}4S]^{1+}$ cluster to SAM; this causes the hemolytic cleavage of the C-S bond of SAM to generate methionine and a 5'-dAdo$^{\cdot}$ radical. The ultimate abstraction of a hydrogen from the substrate by the

highly reactive radical prompts the reaction cascade (Szu *et al*., 2009). Hence, effective reduction of the iron-sulfur cluster is crucial to trigger the chemical transformation.
Another factor that could have prevented the processing of MrsA5 is the absence of biosynthetic elements required for peptide recognition or binding. In fact, certain rSAM/SPASM proteins do require a small protein that serves as chaperone to deliver the peptide to the enzyme. During the biosynthesis of the cofactor PQQ, for instance, the activity of the rSAM protein PqqE is dependent on the chaperone PqqD, which binds PqqA tightly, thus mediating the formation of a carbon-carbon bond between two amino acid side chains on the precursor peptide PqqA (Barr *et al*., 2016). Similarly, the accessory protein MftB facilitates the binding of the precursor peptide MftA with the rSAM protein MftC to catalyze the decarboxylation of the *C*-terminal tyrosine (Khaliullin *et al*., 2016). Although none of the ORFs identified in the mrs_{S47} BGC encoded a protein with a chaperone architecture, it was not excluded that the activity of MrsY in the biosynthesis of the mixed RiPP system could be dependent of MrsR. Further experiments explored the possible role of this enzyme as chaperone; here, the *in vitro* assays were conducted in the presence of MrsR. However, no modifications on the MrsA5 peptide were detected (data not shown).

3.3.3.7 Activity of MrsY and MrsR *in vivo*

Since *in vitro* reconstitution of the MrsY activity required several optimizations and the actual role of MrsR could not be verified, an *in vivo* approach was proposed. MrsY and MrsR were co-expressed with MrsA5 under semi-anaerobic conditions and the peptide was purified aerobically after 20 hours of incubation (section 6.2.5.2). MrsA5 alone was expressed under the same conditions to be used as negative control. After peptide purification and buffer exchange (section 6.2.5.5), MrsP was used to remove the leader peptide for subsequent identification of post-translational modifications (Figure 51A).
When the digested peptide was measured by MALDI-ToF MS, it was observed that incubation of MrsA5 with MrsP generated two fragments; one corresponding to the core sequence predicted based on bioinformatic analysis (*m/z* 834.73) and another (at *m/z* 777.71) that lacked the glycine residue at the *N*-terminus (Supplementary Figure S18A). Formation of the shorter fragments lacking the Gly residue likely resulted from extended incubation of the peptide with MrsP. Analysis by MALDI-ToF MS of the co-expression experiments indeed showed the presence of two new ions at *m/z* 753.70 and *m/z* 810.72 (later confirmed to be $[M+Na]^+$ adducts) only in the co-expression sample (Supplementary Figure S18B). Further analysis by LC-MS/MS showed a signal that corresponded to a MrsA5-related peptide whose *m/z* value indicated a mass shift of -46 Da (Figure 51B-D). For both peptides, the MS/MS analysis was not sufficient to fully identify the modified residues. It was only possible to observe fragments corresponding to a section of the peptide, which suggested a possible cyclization involving the *C*-terminal residue and a central side chain.
On the other hand, after co-expression of MrsY, MrsR and MrsA5 in *E. coli*, the SDS-PAGE gel did not show any prominent band in the flow through that corresponded to the size of MrsR (which was not fused to a His-tag); hence, it was not clear if the protein was expressed at a sufficient titers and the assignment of this as either chaperone or modified enzyme could not be confidently established. To confirm the contribution of MrsR on the modification of MrsA5, this protein was expressed and purified individually. Then, $MrsA5^{Y/R}$ obtained from the co-expression experiment was incubated with pure MrsR for 12 hours at room temperature. Interestingly, the intensity of the previously obtained signal *m/z* 810.72 significantly increased (Figure 51c). This last result suggested a greater conversion of the linear MrsA5 into the modified peptide upon incubation with MrsR under aerobic conditions. Besides, a new signal at *m/z* 826.71 was also detected by MALDI-ToF MS. Still, no further analysis of this new compound could be performed due to the lack of MS/MS spectra. Even though the role of MrsR was

the least understood in the mixed RiPP system mrs_{S47}, its similarity to PqqC suggests this iron-containing redox enzyme would catalyze an analogous reaction. It was initially proposed that MrsR could take the MrsA5 peptide previously modified by MrsY as substrate. However, the results obtained in this thesis suggest these two enzymes might work together.
Structure elucidation of the MrsA5$^{Y/R}$ modified peptide will be essential to draw a conclusion about the biosynthesis of this newly discovered RiPP.

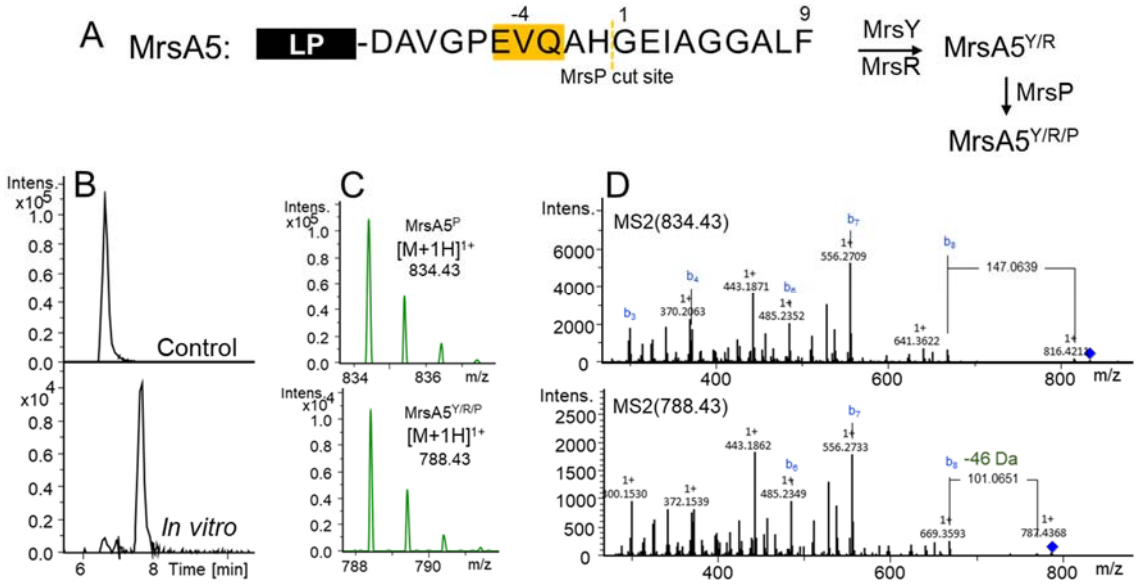


Figure 51. *In vitro* reconstitution of MrsA5$^{Y/R/P}$ biosynthesis. **(A)** MrsA5 peptide sequence indicating the MrsP recognition element and the order of peptide processing. **(B)** Extracted ion chromatogram (EIC) of the unmodified peptide MrsA5^{P} (upper panel) and mature peptide (MrsA5$^{Y/R/P}$) after modification by the radical SAM protein MrsY and the iron-containing redox enzyme MrsR (lower panel); both peptides were treated with MrsP for leader removal prior to LC-MS/MS analysis. **(C)** High-resolution mass spectra of the unmodified core peptide MrsA5^{P} (upper panel) and mature peptide MrsA5$^{Y/R/P}$ (lower panel). Calculated masses: *m/z* at 834.43 for the linear core peptide and *m/z* at 788.43 for the modified peptide. **(D)** MS/MS spectra of the corresponding peptides showing some b ions. The mass decrease by 46 Da is indicated for the modified peptide.

The MrsA5 $^{Y/R}$ metabolite was produced on a larger scale to be purified by reverse-phase HPLC where a total of five fractions were collected. Chromatogram extracted at 280 nm displayed two prominent peaks with retention times at 11.85 min and 16.81 min with similar peak intensities (Supplementary Figure S19). The compounds were collected in fraction 2 and 3. Analysis by MALDI-ToF MS confirmed fraction 2 as the unmodified core peptide while fraction 3 contained the mature peptide (Supplementary Figure S20). Purification of sufficient amounts of pure compound for NMR and bioactivity analysis is currently in progress in our laboratory.

3.3.4 Discussion and outlook

In this section, detailed studies on the $mrs_{S4\text{-}7}$ BGC containing the molecular marker "*lanM*" (*mrsM*) gene from *S. griseus* S4-7 were described. The combined application of bioinformatic tools such as antiSMASH, RiPPMiner and the enzyme similarity tools (EFI-EST and EFI-GNT) allowed the identification of the $mrs_{S4\text{-}7}$ BGC as member of a particular family of gene clusters frequently occurring in Actinobacteria. These BGCs, here termed mixed RiPP systems (*mrs*), encode a set of peptide natural products belonging to at least two different subclasses of RiPPs (Figure 35). The $mrs_{S4\text{-}7}$ BGC was found

to encode six biosynthetic enzymes (MrsM, MrsJ, MrsKC, MrsP, MrsY, and MrsR) and five precursor peptides (MrsA1-A5) that all together give mature peptides belonging to different RiPP subclasses.
To obtain insights into the biosynthesis of this mixed RiPP system (*mrs*$_{S4-7}$), all processing enzymes and precursor peptides were expressed in *E. coli* and purified by affinity chromatography. *In vitro* and *in vivo* assays allowed the reconstitution of the main biosynthetic steps. The results obtained showed that the peptides MrsA1 and MrsA2 are modified by the class II lanthipeptide synthetase MrsM. Dehydroamino acids formed in MrsA2 were further reduced by the oxidoreductase MrsJ (sections 3.3.3.1 and 3.3.3.2). These peptides were transformed into linear peptides containing Dha and Dhb residues. Like most class II lanthipeptide synthetases, MrsM contains a conserved CCH motif as zinc ligand which is important for cyclization (Figure 29). A highly conserved histidine found in LanM cyclase domains is believed to function as the acid that protonates the enolate formed during cyclization (Huo & van der Donk, 2016). The replacement of such His by an Asp residue in MrsM suggested it could have an impaired activity. This hypothesis was experimentally confirmed since MrsM was not able to catalyze thioether formation in MrsA1. Mutations in the same His residue have been reported for other synthetases like NpnM and BsjM, where a Gln replaced the His residue (Zhang *et al*., 2014; Huo & van der Donk, 2016). These two enzymes also produce linear peptides lacking (methyl)lanthionine bridges. Bicereucin (Bsj) is the only two-component lantibiotic in which one of the peptides contains a single lanthionine crosslink and the second is devoid of any bridges (Huo & van der Donk, 2016). Thus, it can be concluded that the replacement of His by Asp also abrogates the cyclase activity.
The next PTM in MrsA2^{M} is the reduction of Dha and Dhb to yield D-alanine and D-aminobutyric acid (Abu), respectively. This reaction is catalyzed by the oxidoreductase MrsJ. Reduction of dehydoamino acids in RiPP biosynthesis has been reported to be a stereoselective reaction catalyzed by reductases, generically called LanJs. It is inferred that the MrsJ-mediated reduction of dehydroamino acids in MrsA2^{M} would follow a similar mechanism. However, since MrsJ belongs to a different subfamily of reductases, further studies are required to confirm the configuration of the alanine and aminobutyric acid residues found in MrsA2$^{M/J}$. In this case, Marfey´s assay for chiral analysis of alanine and amino butyric acid should undoubtedly confirm the configuration of the modified residues (Sethi *et al*., 2021). This experiment is currently being conducted in our laboratory. The activity of MrsJ on MrsA1^{M} was also tested but no modifications were observed.
Under native conditions, the last step in the biosynthesis of class II lanthipeptides is the leader removal which occurs concomitant with the transport of the mature peptide. This process is catalyzed by a papain-like protease domain of a LanT$_{P}$ that requires the presence of a GG/A/S/E motif in the leader for cleavage. Likewise, it is proposed the last step in MrsA1^{M} and MrsA2$^{M/J}$ maturation is a LanT$_{P}$-mediated leader removal. This inference is supported by the existence of a gene encoding an ABC transporter ATP-binding protein containing both ATPase and permease components downstream the *mrs*$_{S4-7}$ BGC. To facilitate the experimental workflow, commercial proteases were employed to remove the leader peptide for analysis of MrsA1^{M} and MrsA2$^{M/J}$. In the case of MrsA1^{M}, the cleavage position was clearly identified. The recognition sequence for the Factor Xa protease was inserted at the leader peptide termini. Conversely, the amino acid sequence of MrsA2$^{M/J}$ allowed the use of trypsin to digest the peptide with no need to modify the primary sequence. For this peptide, the leader and core were not entirely defined and three probable positions where the cleavage could occur were initially identified. The most likely cleavage point was deduced based on the positions of the modified residues after reconstitution of the MrsM and MrsJ activities.
Labionin formation in MrsA3-A4 by the MrsKC synthetase was also proven. So far, labionins have exclusively been found in class III lanthipeptides deriving from a shared Sx$_2$Sx$_{2-5}$C motif. Dehydration and cyclization through certain LanKC enzymes are known to follow a *C*- to-*N*-terminal course (Hegemann & van der Donk, 2018). However, the detection of an intermediate during labionin

formation in MrsA3 suggested an *N*-to-*C*-terminal directionality that could be attributed to the orientation of the labionin motif (CS_XS_{X2-5}). This variation had not been previously reported and it seems to be exclusive of class III lanthipeptides found in some mixed RiPP systems. Regarding leader removal, the protease MrsP was confirmed to trim off not only unmodified precursors but also peptides containing the labionin. The observed mechanism was in accordance with that reported for AmyP (Zhang *et al*., 2022). Some intriguing features in MrsA3-A4 were the disposition of the MrsKC binding site, the MrsP recognition element, and the unusually large *N*-terminal leader sequences (Figure 28). Formation of (methyl)lanthionine/(methyl)labionin in class III lanthipeptides generally precedes leader removal. Proteolysis can occur in a single step mediated by a LanP protease, or through a mechanism where the peptide is first trimmed off by an endopeptidase and then the leader is completely removed extracellularly (Chen *et al.*, 2019; Zhao *et al*., 2022). The MrsP protease was proved to cleave MrsA3-A4 leaving the MrsKC binding site attached to the core. This was a hint that the function of MrsP could go beyond peptide maturation. In this scenario, MrsP-mediated cleavage could favour labionin formation by exposing the MrsKC binding site; the modified peptides MrsA3$^{KC/P}$-A4 $^{KC/P}$ might thus undergo their leader removal during or after export. The importance of MrsP in the Mrs$_{S4-7}$ system was further confirmed when the biosynthesis of MrsA5 was reconstituted.

MrsA5 was the fifth RiPP of the *mrs*$_{s4-7}$ studied in this thesis. This nonlanthipeptide metabolite was found to be processed by the radical SAM enzyme MrsY and the iron-containing redox enzyme MrsR. The modified peptide was cleaved off by the MrsP protease to yield the mature product. Although LC-MS/MS analysis confirmed MrsA5$^{Y/R}$ was modified, the MS/MS fragmentation pattern did not allow to determine the type of modifications and the positions of the modified residues. Still a decarboxylation and a crosslink could be expected based on the mass difference between the unmodified core peptide and the final product. Based on the conservation of residues in the MrsA5-like peptides (Figure 34), the crosslink could occur between the aromatic residue located at the *C*-terminal and a central side chain, most likely a glycine. The decarboxylation and an eventual rearrangement in MrsA5 are expected to be catalyzed by MrsY and MrsR. This hypothesis derives from the results of bioinformatic analyses where MrsR was found to share high similarity to PqqC. This enzyme catalyzes the final step in PQQ biosynthesis in a reaction that involves ring cyclization and eight-electron oxidation of 3a-(2-amino-2-carboxyethyl)-4,5-dioxo-4,5,6,7,8,9-hexahydroquinoline-7,9-dicarboxylic-acid to PQQ. Besides, MrsY was clustered with other RiPP maturases like PqqE and MftC in the sequence similarity network. Purification of sufficient titers of compound for NMR is currently in progress in our laboratory. This will enable the structure elucidation and the assignment of a RiPP subclass to this new peptide.

The possibility of a crosslink between the aromatic ring and the glycine in MrsA5 could be further supported, at least in part, by the triceptide formation in RiPPs recently reported (Ma *et al*., 2021; Nguyen *et al*., 2020). *In silico* and *in vivo* analyses of rSAM-dependent RiPPs identified a large and diverse group of radical SAM proteins that produce cyclophane moieties by linking an aromatic residue and a non-aromatic side chain (Ma *et al*., 2021). These enzymes were classified into three clades based on the characteristic motif of their substrates (Nguyen *et al*., 2020). One of these groups, the Fxs (TIGR04268), contained a highly conserved substrate tail FxxSxx-COOH that corresponds to the same family of MrsA5. Genome context analysis of the Fxs group set this as the largest group of maturases forming triceptides and was found to be broadly distributed among Actinobacteria and most prevalent in *Streptomyces* (Nguyen *et al*., 2020). Although MrsA5 apparently belongs to the same family, the amino acid sequence indicates this might constitute a novel structure not previously bioinformatically or experimentally characterized.

On the other hand, several studies to experimentally characterize the proposed peptide-processing conflict system need to be conducted. The hypothetical role of the GAF domain as intracellular sensor during activation of the conflict system could be supported by the existence of GAF-containing proteins that regulate antibiotic production. Biosynthesis of the antibiotic bitespiramycin by *Streptomyces*

spiramyceticus WSJ-1, for instance, was found to be controlled by four putative regulatory proteins including the GAF domain Bsm23 (Dai *et al*., 2020). The Bsm23 protein was shown to be essential for positive regulation of antibiotic production in the wild-type strain; interestingly, its overexpression resulted in repression of the whole biosynthetic pathway (Dai *et al*., 2020). This could indicate that overstimulation of the GAF domain causes a negative effect for the producer leading to pathway silencing as a self-protection mechanism.

Leader peptide sequences have been reported to serve as ligands modulating CSRs activity in RiPP biosynthesis (Li *et al*., 2018). The possible regulation through the leader peptide of the MrsP substrates (MrsA3-A5) also requires future investigation (Figure 36). Proving this hypothesis could explain the unusually large leader peptide sequences in MrsA3 and MrsA4, and the intriguing position of the MrsP recognition element with respect to the MrsKC binding site. The MrsA5 leader peptide could also have a regulatory function, especially because MrsY and MrsR activities seem to be leader independent. Leader peptide-mediated regulation could serve as the lowest level of regulation in Mrs biosynthesis and could also explain the shared utilization of MrsP to cleave peptides from two different sub-pathways in Mrs$_{S47}$.

Extensive investigations are still required to confirm and fully characterize the hypothetical peptide-processing biological conflict system. Experiments including electrophoretic mobility shift assays could help to further prove the participation of Mrs675 and the GAF domain Mrs665 in the activation of the conflict system in S4-7, and to provide evidence of the role of the leader peptides as downstream regulators.

Further experiments in the laboratory aiming to “unlock” Mrs biosynthesis in *S. griseus* S4-7 may be challenging given the lack of information about the order of events in the deployment cascade, and also due to the eventual lethal consequences derived from the production of the effector. In this concern, preliminary information could be collected through modelling of protein/peptide interactions. Also, transcriptomic analyses of the proposed ORFs implicated in the conflict system could help to confirm their participation in the system deployment.

4. Summary and future perspectives

Root-colonizing bacteria have been intensively investigated for their intimate relationships with plants and insects, and their manifold beneficial activities including antibiotic production. In this thesis, the structures and biosyntheses of ribosomal peptide natural products implicated in the protective effect of *Streptomyces griseus* S4-7 for strawberry plants and pollinators were investigated. These metabolites were encoded in two BGCs (*cpr* and mrs_{S47}) with no chemical structures connected. Selection of these cryptic BGCs was carried out based on previous antagonism tests where individual mutants (*ΔcprE* and *ΔmrsM*) lost activity against *Fusarium oxysporum* (Cha *et al*., 2016; Kim *et al*., 2019). Besides, the mrs_{S47} BGC contains the *mrsM* gene (*lanM*) that was previously established as the molecular marker of *Fusarium* suppressiveness in soil (Kim *et al*., 2019). Since initial cultivation of the wild type strain under laboratory conditions failed to produce the BGC-related compounds, various strategies for pathway activation and heterologous expression were performed in this thesis.

Natural host-based strategies targeted pleiotropic and cluster-situated regulators in *S. griseus S4-7* (section 3.1). Although challenging bacteria with a small number of physical, chemical and biological elicitors was not successful for activating the clusters of interest, two new griseopeptins were detected in the crude extracts. This type III lanthipeptide is encoded in the genome of S4-7 (BGC 11 in Table 1) and its structure, biosynthesis and role as morphogen have been previously studied (Völler *et. al*., 2012). Nonetheless, the biosynthesis of new griseopeptins and the effect of elicitation in their production depicted the effect of environmental stimuli not only in the activation/repression of pathways, but also in the behavior of biosynthetic enzymes responsible for final maturation of metabolites related to morphological differentiation.

The inhibitory effect of S4-7 to *F. oxysporum* was only observed during co-cultivation. However, crude extracts did not retain activity and no potential ions belonging to the target pathways could be detected during LC-MS/MS analyses. It can thus be inferred that bioactive peptides might have been produced in very tiny amounts during co-cultivation and that they were lost during extract preparation.

Pathway activation in *S. griseus* S4-7 was also attempted through overexpression of the validated SARP-like regulator PapR2. Only one crude extract retained anti-*Fusarium* activity, still no potential ions could be connected to the ribosomal peptide natural products of interest; this, in most of the cases, likely due to the low-signal intensities that resulted in the lack of MS/MS data.

Overexpression of cluster-situated regulators (CSRs) was also tested for activation of the *cpr* and mrs_{S47} BGCs. Independent overexpression of the SARP and LuxR regulatory genes *cprB* and *cprO* was insufficient for activation of cpr production. Although both regulatory proteins were predicted to be activators, results suggested that the negative effect of the TetR repressor CprR surpasses that of the activators, thus impeding pathway activation. These three proteins might thus form a concerted regulatory network controlling timing and expression levels of the thiopeptide cpr.

Analogously, none of the CSRs found in the mrs_{S47} BGC activated the pathway. This result could be partially attributed to the operon architecture and the complexity of the mixed RiPP system. Despite some potential ions within the target range of masses were detected in the crude extracts, none of them could be connected to any of the peptides encoded in the mrs_{S47} BGC. Overall, it was clear that the expression of the mrs_{S47} BGC is tightly controlled by an interplay of multiple regulatory levels and that the CSR Mrs675 is of relevance in this network. The decrease in the signal intensity of certain ions after individual overexpression of *mrs*675, the difficulty of generating a construct carrying all *mrs* CSRs, and the predicted architecture of the unusually large Mrs675 suggested this protein has a function not previously anticipated. This hypothesis was later corroborated through bioinformatics analyses of the mrs_{S47} BGC.

Alternatively, a heterologous host-based strategy and an *in vitro* reconstitution approach were developed to investigate the structures and the biosyntheses of metabolites encoded in the *cpr* and mrs_{S47} BGCs, respectively.

The native *cpr* BGC was cloned into an integrative vector and subsequently refactored for heterologous expression in *Streptomyces*. Constitutive and inducible promoters (*ermE** and *tcp*830, respectively) as well as three heterologous hosts (*S. coelicolor* M1152, *S. albus* J1074 and *S. lividans* TK24) were tested for thiopeptide production (Section 3.2). A $[M+2H]^{2+}$ ion at *m/z* 1070.08 was obtained when two refactored versions of the *cpr* BGC were expressed under control of the *tcp*830 promoter in *S. coelicolor* M1152. This ion was also detected when a third refactored version of the cluster was expressed in *S. albus* J1074. Further studies will be required to confirm the candidate ion as the thiopeptide encoded in the *cpr* BGC. Overall, the obtained results showed the importance of testing different heterologous systems to ensure optimal conditions for expressing a functional thiopeptide BGC.

On the other hand, *in vitro* reconstitution of the Mrs provided a deep understanding of the biosynthesis and structures of this member of a newly discovered family of mixed RiPP systems (Section 3.3). The structures of the peptides and the order of post-translational modifications were proposed after reconstituting the activity of the six biosynthetic enzymes on the purified precursors (section 3.3.3). The observation that MrsP participates in the maturation of MrsA3-A5, peptides that belong to two different RiPP subclasses, led to the proposal of a regulatory function for this peptidase. Further genome neighborhood analyses depicted the existence of other mixed RiPP systems in Actinobacteria where the genes encoding an apparent biological conflict system which seems to be the main regulatory system of this pathway, are highly conserved. Little is known about conflict systems and their deployment, but results obtained in this thesis provide the first experimental evidence on how they can regulate the biosynthesis of multiple RiPPs. The protease MrsP was thus proposed to be the core of the biological conflict controlling the RiPP maturation; one of the peptides, MrsA5, which has not yet been fully structurally characterized, but is highly conserved and appears to be the founding member of a novel, widespread RiPP family, likely represents the final effector of the conflict system.

The potentially lethal effectors in biological conflicts have an evolutionary enigma, yet it is known they target the flow of genetic information (replication, transcription and translation). Effectors have to be kept under a tight leash by an array of regulatory mechanisms that ensure their deployment at the right time. The existence of a biological conflict system in the Mrs pathway might explain the failures of activating the BGC when the classic levels of regulation such as pleiotropic and CSRs were targeted in the natural producer S4-7 (section 3.1). The operon encoding the proposed conflict system includes two proteins (Mrs660 and Mrs675) that seem to be part of the machinery that regulates the deployment of the biological conflict. The detrimental effect observed when *mrs*675 was tried to be cloned and overexpressed in S4-7 could be a hint of its role as a mediator in the conflict system deployment.

Further experiments are required to determine the mechanism for activation of the conflict system and its implications on the expression of the whole mixed RiPP system. Given the inherent complexity of the regulatory cascade, reconstruction of the *Streptomyces*-plants-bees ecosystem under the presence of *Fusarium oxysporum* could aid to identify the set of proteins involved in the conflict system deployment. Determining the timing and expression levels of individual genes encoded in the conflict system operon could provide information about the order of events during system activation. Then, the obtained data could be used to establish experiments in the laboratory to fully characterize the system.

The pharmacological and agricultural potential of this newly discovered set of ribosomal peptide natural products is still required. Antimicrobial activity tests of pure peptides their mixtures against a panel of clinically relevant bacteria and phytopathogens are still required to guide their potential application. It is also important to investigate whether these metabolites have synergistic effects and what are their mechanisms of action.

On the other hand, the discovery of mixed RiPP system BGCs in Actinobacteria brings potentially novel chemical structures to be characterized. In fact, experimental studies of mixed RiPP system BGCs (Figure 35) from microorganisms with specific ecological contexts could shed light on the function of these sets of peptides and could help to decipher their regulation via a biological conflict deployment.

5. Concluding remarks

The present study provided evidence of the remarkable potential of *Streptomyces* as an inexhaustible source of novel metabolites and their implications for plant-microbe interactions. The unsuccessful results on activating metabolic pathways in the natural producer S4-7, and all the setbacks that appeared when trying to develop a heterologous system for cpr production depicted some of the challenges that still need to be overcome in the field of natural products to unveil the full potential of this "treasure trove" for bioactive molecules.

On the other hand, the characterization of the mixed RiPP system, the insights into its regulation through a biological conflict system, and the discovery of a widespread family of mixed RiPP systems in Actinobacteria reinforces the necessity of studying targeted microorganisms from a more holistic perspective. This would enable a better understanding of the purpose of these molecules for the natural producers and could help to more easily identify potential pharmacological targets which could, in turn, give new candidates for drug development.

Results described in the present thesis further highlight the potential of *Streptomyces griseus* S4-7 as a bacterial strain producing a set of ribosomal peptide NPs with novel structures, and provide evidence of their biosynthesis and regulation.

6. Material and methods

6.1 Material

6.1.1 Chemicals and consumables

Unless otherwise stated, all common reagents and consumables were purchased from Sigma-Aldrich (St. Louis, USA), Carl Roth (Karlsruhe, Germany), Merck (Darmstadt, Germany), Thermo Fisher Scientific (Waltham, USA), VWR (Darmstadt, Germany), Eppendorf (Hamburg, Germany), Sarstedt (Nümbrecht, Germany), and Serva (Heidelberg, Germany).

6.1.2 Bacterial strains

Table 5 summarizes bacterial strains and their relevant information.

Table 5: Bacterial organisms used in this work, their relevant characteristics and intended use.

Strain	Genotype	Origin	Intended use
Bacillus megaterium DSM 32	Narsing Rao *et al*., 2019	German collection of microorganisms and cell cultures (DSMZ)	Test microorganism for agar-diffusion assay
Eurotium rubrum	--	German collection of microorganisms and cell cultures (DSMZ)	Test microorganism for agar-diffusion assay
Escherichia coli BL21 (DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm]* Δ*hsdS λ DE3 = λ sBamHIo* Δ*EcoRI-B int::(lacI::PlacUV5::T7 gene1) i21* Δ*nin5*	New England Biolabs	Protein expression for *in vitro* reconstitution of the mrs_{S47} BGC
Escherichia coli ET12567	F– *dam*-13::Tn9 *dcm*-6 *hsdM hsdR*, Cmr	MacNeil *et al*., 1992	Donor for conjugation
Escherichia coli Lemo21(DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm]* Δ*hsdS/ pLemo*(Cam^R) *λ DE3 = λ sBamHIo* Δ*EcoRI-B int::(lacI::PlacUV5::T7 gene1) i21* Δ*nin5* pLemo = pACYC184-*PrhaBAD-lysY*	New England Biolabs	Protein expression for *in vitro* reconstitution of the mrs_{S47} BGC
Escherichia coli LOBSTR	Derived from *E. coli* BL21(DE3). Δ*arnA* Δ*slyD*	Andersen *et al*., 2013	Protein expression for *in vitro* reconstitution of the mrs_{S47} BGC
Escherichia coli NEB® 10-beta	Δ*(ara-leu) 7697 araD139 fhuA* Δ*lacX74 galK16 galE15 e14-* $\phi 80$d*lacZ*Δ*M15 recA1 relA1 endA1 nupG rpsL* (Str^R) *rph spoT1* Δ*(mrr-hsdRMS-mcrBC)*	New England Biolabs	Cloning of large constructs and storage of plasmids
Escherichia coli Rosetta	F^- *ompT* $hsdS_B$(r_B^- m_B^-) *gal dcm* (DE3) pRARE	New England Biolabs	Protein expression for *in vitro* reconstitution of the mrs_{S47} BGC

Escherichia coli α-select silver	F$^-$ *deo*R *end*A1 *rec*A1 *rel*A1 *gyr*A96 *hsd*R17(r_k^-, m_k^+) *sup*E44 *thi*-1 *pho*A Δ(*lac*ZYA *arg*F)U169 Φ80*lac*ZΔM15λ$^-$	Bioline	Cloning and storage of plasmids
Fusarium oxysporum strain 588	--	Institute for Pharmaceutical Biology, University of Bonn	Test microorganism for agar-diffusion assay
Streptomyces albus J1074	--	Prof. Dr. Till F. Schäberle. University of Giessen	Heterologous expression of the thiopeptide BGC
Streptomyces coelicolor M1152	Δ*act* Δ*red* Δ*cpk* Δ*cda* *rpoB*[C1298T] *rpsL*[A262G C271T]		Heterologous expression of the thiopeptide BGC
Streptomyces griseus S4-7	--	Cha *et al.* 2015	This study
Streptomyces lividans TK24	--	Prof. Dr. Till F. Schäberle. University of Giessen	Heterologous expression of the thiopeptide BGC

6.1.3 Vectors

Vectors used for cloning and for gene transfer during biparental conjugation were maintained and propagated in *E. coli*. Glycerol stocks were kept at -70 °C for long term storage.

Table 6: Plasmids used and their relevant characteristics including reference. Pure plasmids were stored at – 20 °C.

Vector	Characteristics	Origin
pRM4	--	Prof. Dr. Yvonne Mast, Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany
pSET152_*ermE**	Apr^R, *Streptomyces* integrative vector, *lacZα*, MCS, *rep*^{PUZ}, *ermE** promotor, *oriT*	Institute for Pharmaceutical Biology, University of Bonn
pGM1202	pGM1202 was a gift from Günther Muth (Addgene plasmid # 69615; http://n2t.net/addgene:69615; RRID: Addgene_69615)	Prof. Dr. Till F. Schäberle, University of Giessen
pUZ8002	Kan^R, RK2 derivative, defective *oriT* (*aph*)	Institute for Pharmaceutical Biology, University of Bonn
pET28a(+)	Kan^R, bacterial expression vector with T7 lac promoter, *N*-terminal His tag, MCS-1, thrombin cleavage site, internal T7 epitope tag, *C*-terminal His tag, pBR322 ori	Institute for Pharmaceutical Biology, University of Bonn

pCDFDuet-1	Apr^R, bacterial expression vector, *lac* operator, RBS, *N*-terminal His tag, MCS, pCloDF12 ori	Institute for Pharmaceutical Biology, University of Bonn
pPH151	Contains the *E. coli suf ABCDSE* operon for biogenesis and assembly of Fe-S clusters	Dr. Petra Hänzelmann, Institute for Structural Biology, University of Würzburg

6.1.4 Enzymes

Enzymes were generally stored at -20 °C and maintained in a cool rack while handling.

Table 7: Enzymes used, their enzyme type, if any, their restriction site, including information on the company (headquarters).

Enzyme	Type of enzyme	Cut site	Company
*Bgl*II	restriction endonuclease	5'…A▼GATCT…3'	New England Biolabs (Ipswich, USA)
*Dpn*I	restriction endonuclease	5'…GA▼TC…3'	New England Biolabs (Ipswich, USA)
*Eco*RI-HF	high fidelity restriction endonuclease	5'…G▼AATTC…3'	New England Biolabs (Ipswich, USA)
Factor Xa	Protease	Ile-Glu/Asp-Gly-Arg▼X-X-X	New England Biolabs (Ipswich, USA)
GoTaq G2 Flexi DNA Polymerase	Polymerase	-	Promega (Fitchburg, USA)
*Hin*dIII -HF	high fidelity restriction endonuclease	5'…A▼AGCTT…3'	New England Biolabs (Ipswich, USA)
*Nde*I	restriction endonuclease	5'…CA▼TATG…3'	New England Biolabs (Ipswich, USA)
*Nhe*I	restriction endonuclease	5'…G▼CTAGC…3'	New England Biolabs (Ipswich, USA)
*Not*I-HF	restriction endonuclease	5'…GC▼GGCCGC…3'	New England Biolabs (Ipswich, USA)
Phusion High-Fidelity DNA polymerase	Polymerase	-	New England Biolabs (Ipswich, USA)
*Pme*I	restriction endonuclease	5'…GTTT▼AAAC…3'	New England Biolabs (Ipswich, USA)
PrimeSTAR GXL polymerase	Polymerase	-	Takara (San Jose CA, USA)
Q5 High-Fidelity DNA Polymerase	Polymerase	-	New England Biolabs (Ipswich, USA)

rSAP	shrimp alkaline phosphatase	-	New England Biolabs (Ipswich, USA)
*Sac*I	restriction endonuclease	5'…GAGCT▼C…3'	New England Biolabs (Ipswich, USA)
*Spe*I	restriction endonuclease	5'…A▼CTAGT…3'	New England Biolabs (Ipswich, USA)
T4 DNA ligase	Ligase	-	New England Biolabs (Ipswich, USA)
T5 exonuclease	Exonuclease	-	New England Biolabs (Ipswich, USA)
Taq DNA ligase	Ligase	-	New England Biolabs (Ipswich, USA)
Trypsin	Protease	Lys/Arg▼X-X-X	Not available
*Xho*I	restriction endonuclease	5'…C▼TCGAG…3'	New England Biolabs (Ipswich, USA)

6.1.5 Oligonucleotides

Oligonucleotides were ordered from Eurofins Genomics (Ebersberg, Germany).

Table 8: Primers used, their sequence from 3' to 5' end, as well as their melting temperature (T_M) in °C. Oligonucleotides are listed according to their intended use.

Primer	Sequence 3' → 5'	T_M(°C)
For construction of plasmids for overexpression of CSRs found in cluster 16		
ga-*lan*675-fwd	GTCAGAGAAGGGAGCGGACAGATGAACTGCCTCTCGTGC	> 75
ga-*lan*675-rev	GGCCAGGGCGGGAGATTTCATCAGCTCCTCTCCGTCGTC	> 75
ga-*lan*705-fwd	GTCAGAGAAGGGAGCGGACACATGGGGGCCGGGACGGG	> 75
ga-*lan*705-rev	GGCCAGGGCGGGAGATTTCATCACCCTGGGCTGAGGGC	> 75
ga-*lan*710-fwd	GTCAGAGAAGGGAGCGGACAGATGCGAGCCACTTCGCCGGTC	> 75
ga-*lan*710-rev	GGCCAGGGCGGGAGATTTCATCAGGCCGGCTCGGCCGC	> 75
*lan*725-*Nde*I-fwd	GTCACATATGACGCTAGCCTGCCGA	66
*lan*725-*Eco*RI-rev	ATTAGAATTCTCAGCCCCTCGGCGCCG	69
For construction of plasmids for overexpression of CSRs from cluster 19		
GA-*cpr*SARP-fwd	AGCAACGGAGGTACGGAAGGATGCGGTTCCAGCTTCTG	60
GA-*cpr*SARP-vev	GCAGGTCGACTCTAGAGAGGTCACACCTTCTCTGTGCG	54
ga-cSARP2-fwd	GTCAGAGAAGGGAGCGGACAGATGCGGTTCCAGCTTCTGG	57
ga-cSARP2-rev	GCTCGGTACCCGGGGATCTATACGGCGGGTGTGACGGG	61
*cpr.luxR-Nde*I-fwd	TGTACATATGTCTGTCTTCGGGGTC	58
*cpr.luxR-Hin*dIII-rev	TACAAAGCTTTCAGATCCAGCCGCGTTC	65
For construction of plasmids for heterologous expression of the cpr BGC		
pSET152_fwd	CCTCTCTAGAGTCGACCTGCAGC	62

pSET152_rev	CCTTCCGTACCTCCGTTGCT	62
CprA-For (for *cprL*)	AGCAACGGAGGTACGGAAGGACTAGTATGACCCCCAAGACCGAA	> 75
1-GA-*cpr*3-R (*cprL*)	GCTGCAGGTCGACTCTAGAGAGGGTTTAAACTCTCAGGCGCTGC AGGAG	> 75
2-GA-*cpr*1-F	CCTGCAGCGCCTGAGAGTTTAGGGGAACACGTGTCAAGGT	74
2-GA-*cpr*1-R	GCAGGTCGACTCTAGAGAGGGTTTAAACTCAGGCGATTTCGTAG CGG	62
3-GA-*cpr*2-F	CTACGAAATCGCCTGAGTTTGCGGGGACACGACGATGA	70
*cpr*2_r1_GA	GCTGCAGGTCGACTCTAGAGAGGGTTTAAACCGTCATGACGCCG TCCC	74
*tcp*830_*Eco*RI_For	CGTATTGAATTCTCCGTC GACCTGCAGTTC	61
*tcp*830_*Spe*I_Rev	GCTGCGACTAGTCCTCCCAGATCTCTATCACT	60
GA_*cpr*C_fwd	CCTGCAGCGCCTGAGAGTTTGAAGGTGTGATCCCCATG	72
*cpr*4-rev2	GTCGACTCTAGAGAGGGTTTAAACTCAGGCGGTCAGCAGATGG TG	71
*cpr*5_fwd	TCTGCTGACCGCCTGAGTTTGAGCGACAAGTCATGACG	72
LA-*cpr*-GA-fwd	CCTGCAGCGCCTGAGAGTTTAGCGACGTCCAGTCATCGAC	74
*cpr*ABC-GA-Rev	GTCGACTCTAGAGAGGGTTTAAATGCCTCCCGCCGCGGCTG	74

For construction of plasmids for protein/peptide expression in E. coli

*lan*Core For1	TGGTGGTGCTCGAGTGCTGGTTCAGCGCTCAGTT	74
*lan*Leader Rev 2	ATCCGAATTCGAGCTCATGTCCGAGAACACGA	67
*lan*Leader For1	GCGACCTTCTATGCCGCCCTGGACGGAG	70
*lan*Core Rev1	CATAGAAGGTCGCGACACCACCGACGGCAAC	70
*lan*A2_*Nde*I_FP	CTGGCATATGATGTCCGAGAACACGAAG	58
*lan*A2_*Hind*III-RP	ATGGAAGCTTTCAGTTCTCGACGACGCT	64
*lan*A3_*Nde*I_FP	TAGACATATGAAGCAGTTCGGCGAAT	60
*lan*A3_*Hind*III_RP	TAGAAGCTTCCGGATCTGGTCACCAC	63
ga-CP-*lan*A3-fwd	TGGTGGTGCTCGAGTGCTCACCACGGGGCGGTCAT	77
ga-CPGE-*lan*A3-rev	CGAGATAGAAGGTCGCAGCGGGTTCAGCCACGA	70
ga-LPGE-*lan*A3-fwd	GCGACCTTCTATCTCGCCCTCCGAAGCG	70
ga-LP-*lan*A3-rev	ATCCGAATTCGAGCTCATGAAGCAGTTCGGCGAAT	69
*lan*A4-*Nde*I-fwd	CGTAATCATATGACCGGACCACCCGATAGC	62
*lan*A4-*Hind*III-rev	GTTAATAAGCTTGGTCAGAACAGCGCGCC	66
LM1 For (*Eco*RI)	GTTTAGGAATTCGATGGCGGAGAAGAACCTG	61
LM1 Rev (*Hind*III)	GTTATCAAGCTTTCAGGGGGCGCGTGAG	59
*lanO_Nde*I_FP	TAGACATATGACACGCAGGACAGTGC	61
*lanO_Hind*III_RP	TAGAAGCTTTCACAGCGTCAGCTCCG	64
*lanKC_Nde*I_FP	TAGACATATGCCCGGCATCCAGGAG	63
*lanKC_Hind*III_RP	TAGAAGCTTTCAGCCCAGGGTGAGCAG	66
*lanP_Nde*I_FP	TAGCCATATGACGACCACGGAACCG	65
*lanP_Hind*III_RP	GACTAAGCTTTCAGTTCTCGGAGTCATC	59
*lan*rS_*Bam*HI_FP	GTACGGATCCTATGGCAGCGGTATCGGGC	66
*lan*rS_*Hind*III_RP	GATAAGCTTTCACCGGAGGTCCCGTCC	66
*lan*ox_*Eco*RI_fwd	GATCTAGAATTCGTGAGCGCCGGGGCATC	66
*lan*ox_*Hind*III_rev	CTGATAAGCTTTCATGGTGTGACTCCCATCGGTG	66

For colony PCR

screen-*lan*705-fwd	ACCTGCTGGACGTGTTG	55
screen-*lan*675-fwd	GTCCTTCCGGTGTTCGA	55
pSET-*lanKC*-Fwd	CACCTCCTCTACCTCGAC	55
pET-RP	CTAGTTATTGCTCAGCGG	53

T7	TAATACGACTCACTATAGGG	49
DuetDOWN1	GATTATGCGGCCGTGTACAA	57
ACYCDuetUP1	GGATCTCGACGCTCTCCCT	60
pSET-*cpr*1_For	GCGTCGATTTTTGTGATGCTCG	60
pSET-*cpr*1_Rev	TGACGCCGTGCTTGGAGAAC	62
pSET-*cpr*2_For1	GAGCGGGACTTCACCAAG	57
pSET-*cpr*2_Rev1	GTCAGTAGACGACAAGCG	55
pSET-*cpr*E_Fwd	TACGCGTTCGTGAGCCA	56
pSET-*cpr*F_Rev	GGTTGTTGAAGACGTCGC	55
M13_FP	GTAAAACGACGGCCAGT	56
M13_RP	CAGGAAACAGCTATGAC	53
pGM1202_screen_rev	CTATTCTCACTCCGCTGA	53

6.1.6 Media

All media were prepared with demineralized water and autoclaved immediately at 121 ºC for 20 min; all non-autoclavable ingredients were sterile filtered. pH was adjusted using 3 M NaOH or 3 M HCl accordingly before autoclaving. Agar was cooled down at 50 ºC and depending on the intended antibiotic was added before pouring out.

Table 9: Media used for cultivation of bacterial strains. Ingredients are indicated to prepare 1 liter of medium.

Medium	Ingredient	pH
10x TB salts	0.17 mM KH_2PO_4 0.72 mM K_2HPO_4	-
2x YT	1 g/L yeast extract 1.6 g/L tryptone 0.5 g/L NaCl	7.0
CASO	3 g/L peptone ex casein 1 g/L soy peptone 1 g/L NaCl	7.3
ISP-2	4 g/L yeast extract 10 g/L malt extract 4 g/L dextrose	-
Luria-Broth (LB) agar	10 g/L tryptone 5 g/L yeast extract 5 g/L NaCl 15 g/L agar	7.5
Luria-Broth (LB)	10 g/L tryptone 5 g/L yeast extract 10 g/L NaCl	7.5
Mannitol Soya flour (MS) agar (also known as NL19)	20 g/L mannitol 20 g/L soy flour 20 g/L agar	-
MPY agar	20 g/L malt extract 2.5 g/L peptone ex meat 2.5 g/L yeast extract 12 g/L agar	-
NB agar	7.8 g/L peptone ex meat 7.8 g/L peptone ex casein 2.8 g/L yeast extract 5.6 g/L NaCl	7.5

	1 g/L D-glucose 12 g/L agar	
NL300	20 g mannitol 20 g cotton seed	7.5
NL410	10 g/L glucose 10 g/L glycerol 5 g/L oatmeal 10 g/L soy flour 5 g/L yeast extract 5 g/L bacto casamino acids 1 g/L $CaCO_3$	7.0
NL800	5 g/L glucose 10 g/L glycerol 10 g/L soluble starch 5 g/L soy flour, full fat 2 g/L yeast extract 1 g/L NaCl 1 g/L $CaCO_3$	7.2
PDK agar	10 g/L soy peptone 10 g/L potato dextrose broth 15 g/L agar	7.0
R5	103 g/L saccharose 10 g/L glucose 0,25 g/L K_2SO_4 10,12 g/L $MgCl_2$ 0,1 g/L casamino acids 5 g/L yeast extract 5,73 g/L TES 2 ml trace element solution in 955 ml $H_2O_{deion.}$ after autoclaving separate addition of: 20 ml 1 M $CaCl_2$, 10 ml 0,54% KH_2PO_4, 15 ml 20% L-proline	7.2
SM13	20 g/L D-glucose 20 g/L soluble starch 10 g/L glycerol 10 g/L corn –steep liquor 10 g/L pharmamedia 5 g/L polypeptone 5 g/L NaCl	7.0
SOC	10 g/L tryptone 5 g/L yeast extract 0.5 g/L NaCl 2.5 mL 1M KCl 20 mL 1M D-glucose[1]	7.5
Terrific Broth (TB)	12 g/L tryptone 24 g/L yeast extract 5 g/L glycerol 100 mL/L 10x TB-salts[1]	-
Tryptic Soy Broth (TSB)	17 g/L tryptone 3 g/L soy peptone 2.5 g/L D-glucose 5 g/L NaCl 2.5 g/L K_2HPO_4	7.3

YEME	3 g/L yeast extract 3 g/L malt extract 5 g/L peptone 10 g/L glucose 340 g/L sucrose (34% final concentration) 5 mM $MgCl_2*6H_2O$[1]	-

[1]Added after autoclaving
[2]Liquid media was prepared omitting the agar

6.1.7 Buffer

All buffers were prepared with demineralized water and stored at 4 ºC until use. Unless otherwise indicated, pH was adjusted using 3 N NaOH/KOH or 3 M HCl accordingly.

Table 10: Buffers used, their composition and pH.

Buffer	Ingredients	pH
50x TAE	2 M Tris 1 M acetic acid (glacial) 50 mM EDTA	8.0
5x isothermal (ISO) buffer	500 mM Tris-HCl 50 mM $MgCl_2$ 1 mM each of the 4 dNTPs 50 mM DTT 25% PEG-8000 5 mM NAD	7.5
Elution buffer	50 mM NaH_2PO_4 300 mM NaCl 250 mM imidazole	8.0
Factor Xa buffer	20 mM Tris 100 mM NaCl 2 mM $CaCl_2$	8.0
HEPES buffer	50 mM HEPES	7.8
LanO buffer	100 mM HEPES 500 mM NaCl	8.0
LanP Buffer	20 mM HEPES 1 M NaCl	7.5
Lysis buffer	50 mM NaH_2PO_4 300 mM NaCl 10 mM imidazole	8.0
Reconstitution buffer	50 mM Tris 300 mM NaCl 1 mM DTT	8.0
SDS-PAGE running buffer	0.1% SDS 1.44% glycine 0.3% Tris	-
Separating buffer	0.4% SDS 1.5 M Tris-HCl	8.8
Stacking buffer	0.4% SDS	6.8

	0.5 M Tris-HCl	
Tris-HCl	50 mM Tris	7.5
Urea buffer	10 mM Tris 100 mM NaH_2PO_4 8 M Urea	8.0
Wash buffer I	50 mM NaH_2PO_4 300 mM NaCl 20 mM imidazole	8.0
Wash buffer II	50 mM NaH_2PO_4 300 mM NaCl 35 mM imidazole	8.0

6.2 Methods

6.2.1 Microbiological methods

6.2.1.1 Cultivation of bacterial strains

a) *Escherichia coli*

Cultivation of *E. coli* strains was carried out under aerobic conditions at 37 ºC and 200 rpm. For pre-cultures, 5 mL of LB medium containing the appropriate antibiotics were inoculated either with 50 µL of glycerol stock or with a colony picked with a sterile toothpick.

b) *Streptomycetes*

For cultivation of *Streptomyces griseus* S4-7 and the different strains used as heterologous hosts, a seed culture was prepared inoculating 20 mL of medium with 50 µL of glycerol stock followed by incubation at 30 ºC for 72 hours. Main cultures were prepared with 1% v/v of seed culture and incubated (30 ºC, 180-200 rpm) for 7-12 d.

6.2.1.2 Bacterial transformation

Transformation is the process by which foreign DNA is transferred into a cell. To make competent cells, the permeability of the cell walls is increased by means of chemical treatment with calcium chloride or by a short electrical pulse which allows DNA to diffuse into the cells.
In this work, chemical transformation was used to introduce pure plasmids into *E. coli* strains used as donor for conjugation, and for cloning of small constructs. When large constructs (> 10 kb) were cloned, electroporation was preferred, as it is more efficient.

a) Chemical transformation

Chemically competent *E. coli* cells were prepared inoculating 50 mL LB medium with 500 µL overnight culture followed by incubation at 37 ºC and 200 rpm until the OD_{600} was 0.4 - 0.6. Cells were then harvested by centrifugation (5000 *x g*, 3 min, 4 ºC) and resuspended in 3.5 mL of 70 mM $CaCl_2$, 20 mM $MgSO_4$. After incubation on ice for 30 min, a volume of 875 µL sterile glycerol was added to the bacterial suspension. 50 µL aliquots were pipetted in 1.5 mL Eppendorf tubes and stored at -80 ºC until use.
For transformation, competent cells were thawed on ice for 15 min and then 5 µL of sample DNA were added to the tube flicking to mix the cells and DNA. The tube was placed on ice for 30 min and heat

shock was done at 42 ºC for 60 s followed by incubation on ice for 2 min (5 min for 10-beta *E. coli*). Transformed cells were recovered by adding 950 µL of SOC medium and incubating for 1 h at 37 ºC. An aliquot of 100 µL of cells suspension was plated onto a first LB agar plate containing the appropriate antibiotic while the rest was centrifuged, the pellet resuspended in the residual medium and spread onto a second agar plate. Both plates were incubated at 37 ºC overnight.

b) Electroporation

Electrocompetent cells were prepared from a 20 mL culture at OD_{600} = 0.5. Cells were harvested by centrifugation (5000 *x g*, 3 min, 4 ºC) and washed twice with 10 mL glycerol (10% v/v). Pellet was then resuspended in 1 mL of glycerol and aliquoted to 70 µL in 1.5 mL tubes.
For electroporation, 3 µL of ligation or Gibson assembly reaction product were mixed with one aliquot of competent cells. The mixture was transferred to a UV-sterilized electroporation cuvette and exposed to a voltage of 2.5 kV for 6.5 ms (MicroPulser, BioRad). Immediately after pulse application, 900 µL of SOC medium were added and recovery took place at 37 ºC for 1 h. Transformed cells were plated out onto selective plates as described for chemical transformation.

6.2.1.3 Storage of bacterial strains as cryostocks

All newly created strains were stored as cryocultures for long-time preservation. 700 µL of culture was mixed with one volume of 50% sterile glycerol in a cryogenic storage vial. Tubes were stored at -80 ºC and used as needed for the preparation of precultures.

6.2.1.4 Storage of *Streptomyces* spores as glycerol stocks

Glycerol stocks of *Streptomyces* spores are useful for reliable long-term storage of strains, for conjugation protocols and to obtain homogeneous cultures. Stocks of *Streptomyces* strains were prepared according to Shepherd *et al*., 2010. Briefly, spores were grown in MS agar for 7-12 d at 30 ºC. Spores were rubbed to the surface by using a sterile loop and resuspended into sterile water. The spore chains were broken up by vortexing until a homogeneous suspension was obtained. Then, the suspension was passed through a sterile cotton filter and collected in a sterile 50 mL screw top vial. Spores were concentrated by centrifugation (10 min, 2000 *x g*) and resuspended in 1 mL of fresh water. Sterile glycerol was added to a final concentration of 25% v/v and spore stocks were finally frozen and stored at -80 ºC.

6.2.1.4 Intergeneric conjugation between *E. coli* and *Streptomyces*

The methylation-deficient *E. coli* ET12567 harboring the helper plasmid pUZ8002 was used as donor strain while *Streptomyces* spores were used as recipient.
The plasmid to be transferred was transformed into electrocompetent *E. coli* ET12567/pUZ2008 cells and plated out on LB agar containing chloramphenicol (25 µg/mL), kanamycin (50 µg/mL) and the appropriate antibiotic for selection of the plasmid of interest (apramycin 50 µg/mL in most of the cases). After overnight incubation, a single colony was inoculated in 5 mL of LB medium with antibiotics and incubated for 12 h to be used as pre-culture. On the following day, 10 mL of LB medium was inoculated by 100 µL of pre-culture and incubated until an OD_{600} of 0.5 was reached. To prevent further growth, all following steps were conducted on ice. Cells were harvested by centrifugation (5000 rpm, 5 min, 4 ºC) and washed twice with LB containing 10 mM of $MgCl_2$ to remove antibiotics. Cells were then resuspended in 500 µL of LB and stored on ice until use for conjugation.

For preparation of the recipient, an aliquot of 100 µL of spores was thawed and washed twice with 1 mL of 2X YT medium. Spores were then resuspended in 500 µL of fresh 2X YT medium and activated by heating to 50 ºC for 10 min.
For conjugation, donor and recipient were mixed in a 1.5 ml tube and vortexed for 30 s. Bacterial mixture was centrifuged (13 000 x *g* for 1 min) and the pellet resuspended in 100 µL of medium. Finally, one MS agar plate containing 50 mM of $MgCl_2$ was used to spread the bacterial mixture and incubated at 30 ºC. After 20 h, the plate was overlaid with 0.5 mg of nalidixic acid and 1 mg of apramycin and incubation continued until ex-conjugants appeared.
Single colonies were picked and transferred to new MS agar plates containing apramycin and nalidixic acid (50 and 25 µg/mL, respectively) to confirm the ex-conjugants. After 5-7 days of incubation, five selected colonies were validated by colony PCR and two positive clones were used to inoculate 20 mL of TSB to be used as seed culture for heterologous production.

6.2.1.5 Antagonism test against *Fusarium*

To evaluate the *Fusarium* growth inhibition from *Streptomyces* strains, the antagonism test described by Kim *et al*., 2019 was used. 10 µL of mycelium was inoculated onto an agar plate as shown in Figure 53. Plates were incubated at 30 ºC for three days to allow bacterial growth. Then, a piece of agar containing *Fusarium* mycelia was placed at the center of the plate and incubation followed at 25 ºC for additional five days. Antifungal activity was determined based on the size of the inhibition zone.

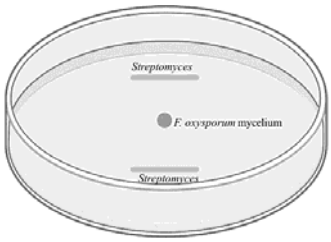


Figure 52. Exemplification of an antagonism test of *Streptomyces* to *Fusarium oxysporum*. *Streptomyces* are first applied onto an agar plate and incubated at 30 ºC for 3 d to allow cell growth and metabolite production. *F. oxysporum* mycelium is then placed in the center and incubation continues for 5 d. Inhibition of *Fusarium* development is measured according to a control strain.

6.2.1.6 Agar diffusion test

The agar diffusion test is the primary method to qualitatively assess the antimicrobial activity of a sample (Murray, 2015). This technique was used to screen crude extracts obtained from different cultivation and heterologous expression experiments. *Bacillus megaterium* was chosen as most of the known RiPPs are active against gram-positive bacteria. Also, antifungal activity tests against *Fusarium oxysporum* 588 were conducted to monitor the production of the compounds of interest.
50 µL of crude extract dissolved in methanol at 5 mg/mL were applied to a paper disk which was placed in the center of a petri dish containing the appropriate agar. Then, plates were covered with a bacterial spray and incubated according to Table 11. The inhibition zone was measured and compared with the antibiotics used control: ampicillin (200 µg/mL), benzylpenicillin (1 mg/mL), miconazole (500 µg/mL) and streptomycin (1 mg/mL).

Table 11: Growth conditions for microorganisms used for agar diffusion test.

Test microorganism	Medium	Growth conditions
B. megaterium de Bary 1884	NB agar	37 ºC for 24 h
F. oxysporum 588	MPY agar	25 ºC for 5-7 d

6.2.2 Molecular biological methods

6.2.2.1 DNA isolation

a) Isolation of genomic DNA

Genomic DNA from *Streptomyces griseus* S4-7 was obtained using the "Wizard® Genomic DNA Purification Kit (Promega/Fitchburg, USA) following the manufacturer´s instructions for isolation of genomic DNA from Gram-positive bacteria. Samples of 1.5 mL of mycelial culture were used for purification. Pure genomic DNA was eluted with 100 µL of sterile water, aliquoted to 20 µL and stored at -20 ºC. To prevent DNA fragmentation, freezing and thawing was avoided using the same aliquot maximum three times.

b) Plasmid isolation

Plasmids were isolated from stationary liquid *E. coli* cultures using the "FastGene Plasmid Mini Kit" (Nippon Genetics) for high copy plasmids and the "Pure Yield TM Plasmid Miniprep System" (Promega/Fitchburg, USA) for low copy plasmids. In all cases, elution buffer was substituted by 20-50 µL sterile water to avoid the presence of components from the buffer that could inhibit the activity of enzymes used in subsequent steps.
In all cases, DNA concentration was measured with an Eppendorf BioSpectrometer® basic, 100 – 240 V ±10 % (50 – 60 Hz) (Eppendorf AG Hamburg, Germany)

6.2.2.2 DNA amplification by PCR

a) Amplification of vectors and inserts up to 3 kb using Q5 polymerase

Due to the complexity of the template (genomic DNA from *Streptomyces*) used in this work, Q5® High-Fidelity DNA Polymerase (NEB) was used for amplification of DNA fragments up to 3 kb and for vectors. Reaction setup and thermocycling conditions are described in Table 12 and 13, respectively. All components were thawed, mixed on ice and quickly transferred to a cycler preheated to 98 ºC, which is the denaturation temperature.

Table 12: Reaction setup for one PCR using Q5 polymerase.

Component	1x 25 µL reaction	Final concentration
5x Q5 High GC Enhancer	5 µL	1x
5x Q5 reaction buffer	5 µL	1x
dNTPs mixture (10 mM each)	0.5 µL	200 µM
10 µM Primer mix	1.25 µL	0.5 µM
Template DNA	Variable	< 1 µg
Q5-High-Fidelity DNA polymerase	0.25 µL	0.02 U/ µL
Nuclease-free water	To 25 µL	-

Table 13: Thermocycling conditions for PCRs to amplify DNA using Q5 polymerase.

Step	Temperature	Time
Initial denaturation	98 ºC	30 s
	98 ºC	10 s
25 – 30 cycles	50–72 ºC*	15 s
	72 ºC	30 s/kb
Final extension	72 ºC	2 min
Hold	10 ºC	∞

*Optimal temperature was experimentally determined by gradient PCR.
The lid temperature was set to 98 °C.

b) Amplification of long DNA fragments using PrimeSTAR GXL polymerase

PrimeSTAR GXL polymerase is robust high-fidelity enzyme recommended for amplification of difficult targets such as GC-rich sequences or long amplicons. This enzyme was particularly used when genomic DNA from *S. griseus* S4-7 was used as template. Tables 14 and 15 display the reaction setup and the thermocycling conditions.

Table 14: Reaction setup for one PCR using PrimeSTAR GXL polymerase.

Component	1x 50 µL reaction	Final concentration
5x PrimeSTAR GXL buffer	10 µL	1x
dNTPs mixture (2.5 mM each)	4 µL	200 µM each
10 µM Primer mix	1 µL	0.2 µM
DMSO	2.5 µL	5% v/v
Template DNA	Variable	Variable
PrimeSTAR GXL DNA polymerase	1 µL	0.025 U/µL
Nuclease-free water	To 50 µL	-

Table 15: Thermocycling conditions for two-step PCRs to amplify DNA using PrimeSTAR GXL polymerase.

Step	Temperature	Time
Initial denaturation	98 ºC	10 s
30 cycles	68 ºC	1 min/kb
Hold	10 ºC	∞

*Optimal temperature was experimentally determined by gradient PCR
The lid temperature was set to 98 °C.

c) Colony PCR

To determine the presence or absence of insert DNA in plasmid constructs, colony PCR was used as a first method. Individual transformants were picked with a sterile toothpick and resuspended in 5 µL of autoclaved water to be used as template. Then, 20 µL of PCR master mix prepared as shown in Table 16 were added and amplification took place under the cycling conditions described in Table 17. Annealing temperature was experimentally determined through gradient PCR using the ligation or Gibson assembly product, which served at the same time, as a positive control.

Table 16: Reaction setup for one PCR using Go Taq polymerase.

Component	1x 25 µL reaction	Final concentration
5x Green GoTaq reaction buffer	5 µL	1x
25 mM $MgCl_2$ solution	2.5 µL	2.5 mM
dNTPs mixture (10 mM each)	1 µL	200 µM each
10 µM Primer mix	1 µL	0.4 µM
DMSO	1.25 µL	5%
Template DNA	Variable	Variable
GoTaq DNA polymerase	0.125 µL	1.25 U/ µL
Nuclease-free water	To 25 µL	-

Table 17: Thermocycling conditions for colony PCRs using GoTaq polymerase.

Step	Temperature	Time
Initial denaturation	95 ºC	2 min
	95 ºC	30 s
25 – 30 cycles	50–72 ºC*	30 s
	72 ºC	1 min/kb
Final extension	72 ºC	5 min
Hold	10 ºC	∞

*Optimal temperature was experimentally determined by gradient PCR. The lid temperature was set to 95 °C.

6.2.2.3 Agarose gel electrophoresis

Agarose gel electrophoresis was used for analytical and preparative separation of DNA. Agarose was dissolved in 1x TAE buffer (1% w/v) in a microwave and stored at 60 ºC until use. This solution was poured into a horizontal cast and a comb was placed into the solution to form the wells. After polymerization of the agarose, the comb was removed, and the chamber was filled with 1x TAE buffer. 5-10 µL of DNA ladder was loaded in the first well as reference and samples were mixed with 6X Gel loading dye purple (NEB) to be pipetted into the pockets. The electrophoretic separation was carried out at 120 v for 35 min and the result was visualized under UV light after staining with ethidium bromide.

6.2.2.4 Extraction and purification of DNA from agarose gel

PCR products were purified from agarose gels to be used as inserts for cloning purposes. Samples were loaded onto an agarose gel and separated according to section 5.2.2.3. Then, the gel was cut from the sides and only the sections containing the standard DNA were stained and exposed to the UV light to avoid damage of DNA. The desired DNA bands were cut out of the gel using the standard as size reference. Purification followed using the Zymoclean Gel DNA Recovery Kit (Zymo Research/Irvine, USA). Pure DNA was eluted in 7-10 µL of sterile water and stored at -20 ºC until use.

6.2.2.5 Enzymatic modification of DNA

a) Restriction enzyme digestion

Restriction enzymes or endonucleases are DNA-cutting enzymes that recognize a specific target sequence and cut at or near that sequence. In this work, modification of DNA with endonucleases was carried out for linearization of plasmids to be used as vector backbones during cloning, and for analysis of new-cloned constructs.
In all cases, a 50 µL reaction was set as described in Table 18. For double digest, 1 µL of each restriction enzyme was used and the water volume was adjusted accordingly. Depending on the intended use, samples were incubated at 37 ºC for 3 h (when the linear product was to be used for cloning) or one hour for verification of constructs. Inactivation of the enzymes took place for 20 min at the temperature recommended by the manufacturer.

Table 18: Standard restriction digest provided by New England BioLabs.

Component	1x 50 µL reaction
10x NEBuffer*	5 µL (1x)
DNA	1 µg
Restriction enzyme	1 µL (10 units)
Nuclease-free water	To 50 µL

*CutSmart buffer was used for all the enzymes except for *Bgl*II that required the use of 3.1 buffer (NEB).

b) Dephosphorylation of linearized plasmid DNA

Dephosphorylation of linear DNA is the process by which a phosphatase removes the 5´phosphate to reduce the occurrence of re-circularized vector during ligation. As the shrimp alkaline phosphatase (NEB/Ipswich, USA) used in this work is compatible with CutSmart buffer, 1 µL of rSAP was directly added to the restriction digest after 2 h, and incubation at 37 ºC continued for one additional hour.

6.2.2.6 Cloning

a) Traditional cloning

Restriction enzyme-based cloning was selected to create all the constructs for heterologous expression of proteins and precursor peptides used for *in vitro* reconstitution of the hybrid RiPP system. Primers containing unique restriction sites compatible to the vector were used for amplification of inserts. Then, PCR product and vector were digested according to section 5.2.2.5a) and concentrated using the DNA Clean & Concentrator-5 (Zymo Research/Irvine, USA). Vector and insert were mixed at a ratio of 1:5 (vector: insert) and incubated for 30 min at 23 ºC in the presence of the T4 ligase (NEB/Ipswich, USA) as indicated in Table 19. Enzyme was heat inactivated at 65 ºC for 10 min.

Table 19: Components for a ligation reaction with T4 DNA ligase.

Component	1x 20 µL reaction
10x T4 DNA ligase buffer	2 µL (1x)
Vector DNA	variable*
Insert DNA	variable*
T4 DNA ligase buffer	1 µL
Nuclease-free water	To 20 µL

*volume of vector + insert was always adjusted to 8 µL according to the DNA length and concentration.

On the other hand, heterologous expression of the MrsY protein required codon optimization. The gene encoding the rSAM sequence was optimized for expression in *E. coli* using the GeneArt gene synthesis services (Thermo Fisher Scientific). Restriction sites for double digest with *Bam*HI and *Hin*dIII were added at the 3´and 5´ The optimized DNA sequence was then subcloned into pET28a following the protocol described above.

b) Gibson assembly

Gibson assembly is an exonuclease-based method that allows the seamless assembly of DNA under isothermal conditions. In this method, linear insert and vector are prepared by PCR and/or restriction digest. The reaction involves three enzymes: a 5´exonuclease for generation of long overhangs, a polymerase which fills in the gaps of the annealed single strand ends, and a DNA ligase to seal the nicks of the annealed and filled-in gaps.
In this thesis, Gibson assembly was employed to generate all the constructs used for heterologous expression of the thiopeptide BGC and for the CSR-guided pathway activation. In the first case, a stepwise approach was chosen to facilitate the amplification and assembly of the native BGC (> 30 kb) and to enable the cluster refactoring. Unless otherwise stated, insert and vector (5 µL in total) were mixed at a ratio of 1:3 (vector: insert) with a 15 µL aliquot of Gibson assembly master mix (Table 20). Incubation was at 50 ºC for 15-60 min, 15 min for one-step and 1 h for stepwise cloning.

Table 20: Master mixture used and their components for Gibson assembly.

Component	113x 15 µL reaction
5x ISO buffer	320 µL (1x)
10 U/µL T5 exonuclease	0.64 µL
2 U/µL Phusion polymerase	20 µL
40 U/µL Taq ligase	160 µL
Nuclease-free water	1.2 mL

Gibson assembly master mix was prepared, aliquoted to 15 µL and stored at -20 ºC until use.

6.2.2.7 Primer design and primer preparation

Depending on the intended use, primers were designed with SnapGene software or with the NEBuilder Assembly Tool (NEB/Ipswich, USA).
Lyophilized oligonucleotides were dissolved in autoclaved water to obtain a concentration of 100 µM. This stock solutions were used to prepare a primer mix containing forward and reverse primer at 10 µM each to be used for PCRs.

a) Primers for Gibson assembly

Gibson assembly primers containing 20 bp flanking the insert plus 20 bp homologous to the vector were used to amplify inserts. For stepwise Gibson assembly, a *Pme*I restriction site was inserted between the insert and overhang sequence so that the new-created construct could be linearized and used as a vector backbone in the following cloning step.

b) Primers for traditional cloning

Primers for traditional cloning consisted of 18-20 bp homologous to the insert to be amplified. Unique restriction sites were incorporated at the 5´ along with 4-6 additional random base pairs to ensure proper enzyme performance as many restriction enzymes present lower efficiency when cleaving close to the end of DNA fragments. To allow the purification of proteins and peptides after expression, single genes were cloned in frame with the *N*-terminal hexahistidine tag.

c) Primer for screening

Specific primers flanking the vector backbone and the end of the insert were designed for screening of colonies harboring a cloned plasmid. The length of the primers was 18-23 nucleotides with a 40-60 % GC-content and the absence of secondary binding sequences was verified during the cloning *in silico*.

6.2.2.8 Sanger sequencing

To confirm the correct assembly of insert and vector, cloned constructs were sent to Eurofins Genomics (Ebersberg, Germany) for Sanger sequencing. Primers used for each project are indicated in Table 8. Results were received as FASTA files and verified by annealing them to the *in silico* cloned construct using SnapGene software.

6.2.3 Activation of secondary metabolism in *S. griseus* S4-7

6.2.3.1 Standard fermentation

Standard cultivation of *S. griseus* S4-7 for production of secondary metabolites was carried out in PDK and MS liquid media and agar. Seed cultures were prepared as described in section 6.2.1.1b) and then used to inoculate the main cultures. For cultivation in liquid media, 100 mL of medium was prepared in 500 mL flasks. The same volume of agar was prepared and poured out into Petri dishes.

6.2.3.2 Chemical elicitation

PDK medium was used for all experiments involving the application of elicitors. 100 mL of medium was inoculated by 1 mL of seed culture as described before and elicitors were applied according to Table 21. Main cultures were incubated for 7 d and then extracted as described in section 6.2.7.1.

Table 21: Overview of elicitors used for activation of secondary metabolite production in *S. griseus* S4-7.

Type of elicitation	Elicitor	Comments
Biological	Co-cultivation with *F. oxysporum*	A piece of agar containing *F. oxysporum* was added to the culture on the third day of cultivation
Biological	Co-cultivation with *F. oxysporum* (in agar)	A piece of agar containing *F. oxysporum* was placed into the plate on the third day of cultivation
Biological	*S. coelicolor*	-
Biological	*B. subtilis* lysate	100 µL of bacterial lysate was added 24 h after inoculation
Chemical	Ethanol	At 2% v/v added 24 h after inoculation
Chemical	Tetracycline	At 100 ng/mL added 24 h after inoculation
Chemical	Zinc	-
Chemical	$LaCl_3$	-
Physical	Cotton	1.5 g of cotton were added to the flask and autoclaved with the medium
Physical	pH shock- pH 6	pH was adjusted to 6 by adding HCl
Physical	pH shock- pH 8.5	pH was adjusted to 8.5 by adding NaOH
Physical	Temperature shift- 37 ºC	-
Physical	Temperature shift- 16 ºC	-

6.2.3.3 Regulator based semi-targeted pathway activation

The SARP-type regulator PapR2 found in the pristinamycin BGC from *Streptomyces pristinaespiralis* was tested for activation of metabolic pathways in *S. griseus* S4-7 according to Krause *et al*., 2020. A plasmid harboring the *papR2* gene under control of the *ermE** constitutive promoter and the empty vector pRM4 to be used as negative control were kindly provided by Prof. Dr. Yvonne Mast (Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany). Pure plasmids were transferred to *S. griseus* S4-7 via conjugation as described in section 5.2.1.4; confirmed ex-conjugants were cultivated in eight different systems (seed culture/main culture): (i) NL410/NL19; (ii) R5/R5; (iii) R5/NL300; (iii) R5/NL800, in liquid and solid media. Fermentation took place at 30 ºC for 10 days.

6.2.3.4 Overexpression of CSRs

CSRs predicted to be activators found in the *cpr* BGC and mrs_{S47} BGC from *S. griseus* S4-7 were cloned into the pGM1202:*tipA* vector by Gibson assembly according to 6.2.2.6. Genes were cloned to be under the control of the inducible promoter *tipA*. Plasmids were transferred to *S. griseus* S4-7 via conjugation.

50 mL main cultures were prepared using six media (PDK, CASO, YEME, ISP-2, SM13, and TSB) inoculated by 500 µL seed culture. Overexpression of CSRs was induced by addition of 12.5 µg/mL thiostrepton.

6.2.4 Refactoring of the *cpr* BGC for heterologous expression in *Streptomyces*

6.2.4.1 Construction of the plasmid for expression of the native *cpr* BGC.

Figure 54 summarizes the workflow for generation of the plasmid carrying the native *cpr* BGC. In all cases, PCR, restriction enzyme digest, Gibson assembly, and other molecular methods were carried out as described in section 6.2.2.

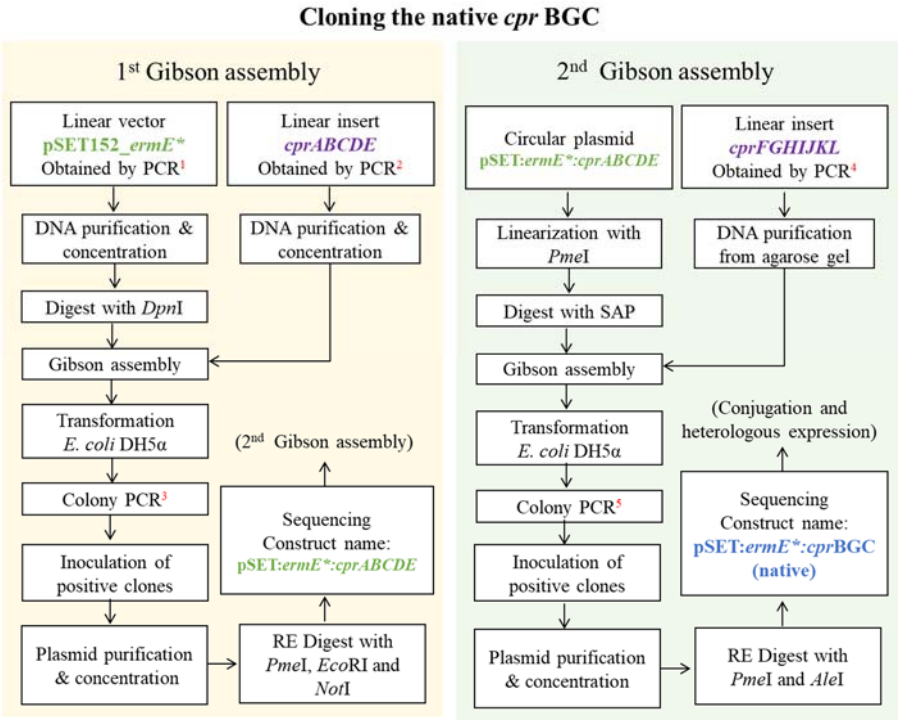


Figure 53. Workflow for cloning the native *cpr* BGC. A two-step Gibson assembly (GA) strategy was used to insert *cprABCDEFGHIJKL* into the pSET vector to be under control of the constitutive *ermE** promoter. Numbers in red correspond to the primer pairs used for each PCR (oligonucleotide sequences are listed in Table 8). General molecular methods are described in section 6.2.2. The final construct was named pSET:*ermE*:cpr*BGC$_{(native)}$. PCR[1]: pSET152_fwd / pSET152_rev. PCR[2]: 2-GA-Cpr1-F / 2-GA-Cpr1-R. PCR[3]: pSET-cpr1 For / pSET-cpr1 Rev. PCR[4]: 3-GA-cpr2-F / cpr2_r1_GA. PCR[5]: pSET-cpr2_For1 / pSET-cpr2_Rev1.

6.2.4.2 Construction of the plasmid for expression of the refactored *cpr* BGC (r-1)

The first refactoring step consisted of relocating the precursor peptide gene *cprL* next to the promoter to ensure correct transcription. Figure 55 shows the workflow for refactoring the *cpr* BGC to generate the pSET:*ermE*:cpr*BGC$_{r\text{-}1\ (cprL)}$. In all cases, PCR, restriction enzyme digest, Gibson assembly, and other molecular methods were carried out as described in section 6.2.2.

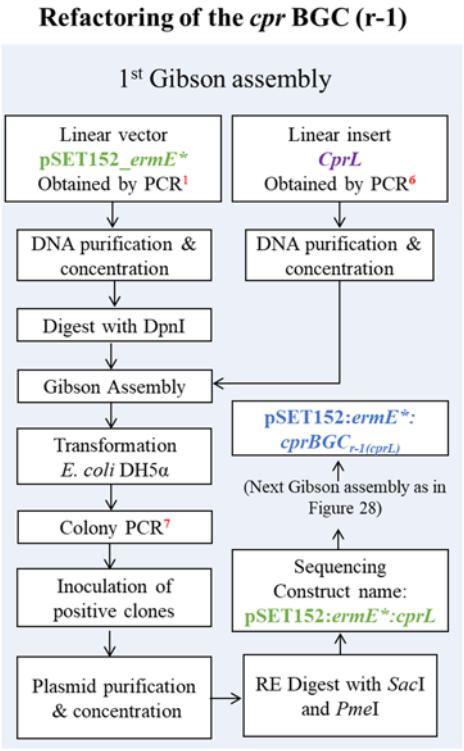


Figure 54. Experimental workflow for refactoring of *cpr* BGC (r-1). Precursor peptide *cprL* was inserted in front of the *ermE** promoter in the first Gibson assembly step. Then, the obtained construct was used as vector backbone for insertion of the *cpr* BGC obtained through the procedure described in Figure 54. Numbers in red correspond to the primer pairs used for each PCR (oligonucleotide sequences are listed in Table 8). General molecular methods are described in section 6.2.2. The final construct was named pSET:*ermE**:*cpr*BGC$_{r\text{-}1(cprL)}$. PCR[1]: pSET152_fwd / pSET152_rev. PCR[6]: *cprA*-For / 1-GA-*cpr*3-R. PCR[7]: M13-FP / M13-RP.

6.2.4.3 Construction of the plasmid for expression of the minimal *cpr* BGC (r-2)

A second refactoring approach for the *cpr* BGC consisted of the creation of a "minimal gene cluster" that contained the precursor peptide and modifying enzymes only. For this purpose, multiple cloning steps were required. Figure 56 summarizes the main steps.

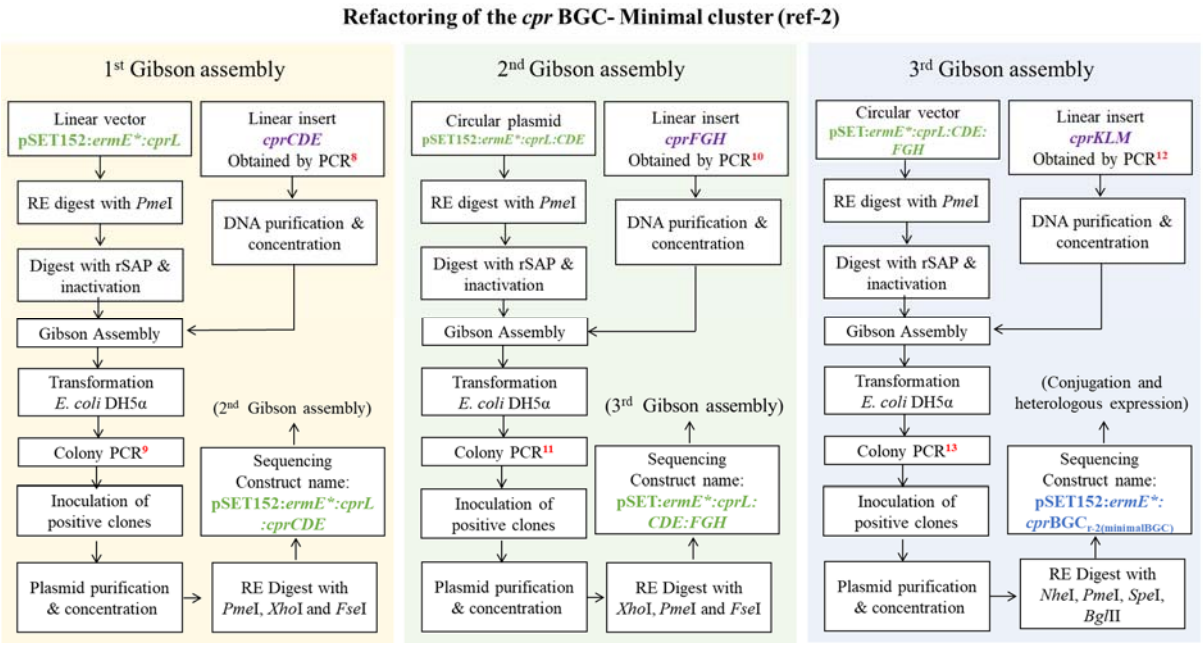


Figure 55. Experimental workflow for refactoring of the *cpr* BGC (r-2). A minimal *cpr* BGC was created by cloning the biosynthetic genes into the pET:*ermE** in a three-step Gibson assembly. The construct pET:*ermE**:*cprL* obtained as described in Figure 55 was taken as initial vector backbone. Inserts in length of 6 kbp, 5 kbp and 4 kpb were added stepwise. Numbers in red correspond to the primer pairs used for each PCR (oligonucleotide sequences are listed in Table 8). General molecular methods are described in section 6.2.2. The final construct was named pSET:*ermE**:*cpr*BGC$_{r\text{-}2(minimal\ BGC)}$. PCR[8]: GA_*cprC*_fwd / 2-GA-*cpr*1_R. PCR[9]: pSET-*cprE*-Fwd / M13-FP. PCR[10]: 3-GA-*cpr*2-F / *cpr*4-rev2. PCR[11]: pSET-*cprE*-Fwd / pSET-*cprF*-Rev. PCR[12]: *cpr*5_fwd / *cpr*2_r1_GA. PCR[13]: pSET-*cpr*2_For1 / M13-FP.

6.2.4.4 Promoter replacement for heterologous expression of refactored (r-1 and r-2) *cpr* BGC

Different promoters were tested to activate the heterologous expression of the *cpr* BGC. Figure 57 shows the procedure to replace the *ermE** promoter for the inducible *tcp*830 promoter in the previously obtained refactored constructs.

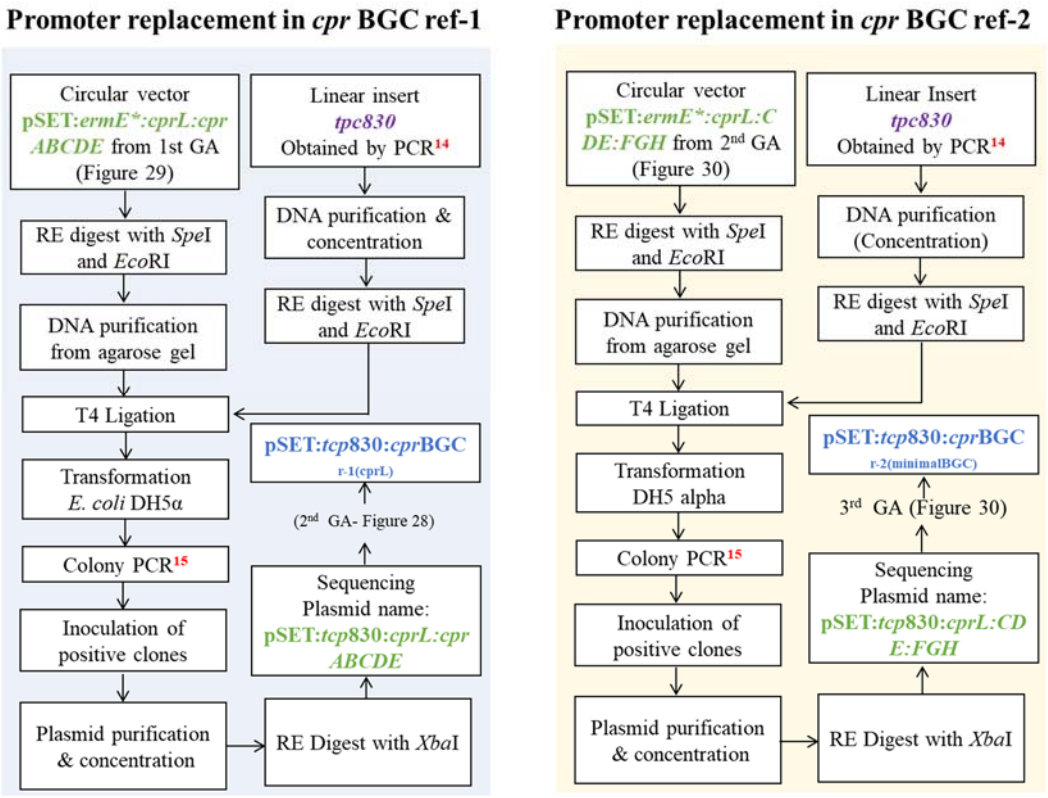


Figure 56. Experimental workflow for promoter replacement in the pSET:*ermE**:*cpr*BGC$_{r\text{-}1(cprL)}$ (left panel) and pSET:*ermE**:*cpr*BGC$_{r\text{-}2(\text{minimal BGC})}$ (right panel). **Left:** the pSET:*ermE**:*cprL:cprABCDE* obtained from first GA (Figure 55) was linearized by restriction digest and the DNA fragment containing the *cpr* BGC was ligated to a second DNA fragment containing the *tcp830* promoter obtained by PCR and digested with compatible restriction enzymes. Cloning workflow continued as described in Figure 55. The final construct was named pSET:*tcp830:cprBGC*$_{r\text{-}1(cprL)}$. PCR[14]: *tcp*830_*Eco*RI _For / *tcp*830_*Spe*I_ Rev. PCR[15]: pSET-*cpr*1 For / pSET-*cpr*1 Rev. **Right:** the pSET:*ermE*:cprL:CDE:FGH* obtained from second GA (Figure 56) was linearized by restriction digest and the DNA fragment containing the *cpr* BGC was ligated to a second DNA fragment containing the *tcp830* promoter obtained by PCR and digested with compatible restriction enzymes. Cloning workflow continued as described in Figure 56. The final construct was named pSET:*tcp*830:*cpr*BGC$_{r\text{-}2(\text{minimal BGC})}$. PCR[14]: *tcp*830_*Eco*RI _For / *tcp*830_*Spe*I_ Rev. PCR[15]: pSET-*cpr*1 For / pSET-*cpr*1 Rev.

6.2.4.5 Heterologous expression of the *cpr* BGC

Streptomyces coelicolor M1152, *S. albus* J1074 and *S. lividans* TK24 were used as heterologous hosts for expression of native and refactored *cpr* BGC. Main cultures were prepared by inoculating 100 mL of medium with 1 mL of seed culture and incubated at 30 ºC (200 rpm) for 7-12 d. As *Streptomyces* need good aeration for optimal growth, 500 mL baffled flasks or Erlenmeyer containing 1 g of glass beads were used. Since the pSET125 is integrative, no antibiotic was added to the main fermentation cultures. When pGM1202 was used, 12.5 µg/mL of thiostrepton were added to induce the expression by activation of the *tipA* promoter.

6.2.5 Overexpression and purification of proteins and precursor peptides in *E. coli*

6.2.5.1 Protein and peptide expression

Protein expression was carried out in 100 mL of TB-medium containing the appropriate antibiotic. Pre-cultures of *E. coli* strains (see Table 22) harboring the expression vector were prepared as described in section 5.2.1.1. Cultures were inoculated with 1 mL of pre-culture and incubated at 37 ºC in a rotary shaker at 220 rpm. When the cells reached an optical density of 0.8-1.0, the flasks were placed into ice. After 30 min, expression was induced with IPTG. Incubation continued for 15 h to 16 ºC.

Table 22: Summary of conditions for protein/peptide expression in *E. coli*.

Protein/peptide	Construct name	Heterologous host	Expression conditions
MrsA1xa	pET28a:*mrsA1xa*	*E. coli* BL21	400 µM IPTG, aerobic
MrsA2	pET28a:*mrsA2*	*E. coli* BL21	400 µM IPTG, aerobic
MrsA2xa	pET28a:*mrsA2xa*	*E. coli* BL21	400 µM IPTG, aerobic
MrsA3	pET28a:*mrsA3*	*E. coli* BL21	400 µM IPTG, aerobic
MrsA3xa	pCDFDuet:*mrsA3xa*	*E. coli* BL21	400 µM IPTG, aerobic
MrsA4	pET28a:*mrsA4*	*E. coli* BL21	400 µM IPTG, aerobic
MrsA5	pET28a:*mrsA5*	*E. coli* BL21	400 µM IPTG, aerobic
MrsKC	pET28a:*mrsKC*	*E. coli* LOBSTR	Final OD_{600} 1.0; 400 µM IPTG, aerobic; 20 h incubation
MrsM	pCDFDuet:*mrsM*	*E. coli* LOBSTR	Final OD_{600} 1.0; 400 µM IPTG, aerobic; 20 h incubation
MrsJ	pET28a:*mrsJ*	*E. coli* BL21	400 µM IPTG, aerobic
MrsP	pET28a:*mrsP*	*E. coli* BL21	100 µM IPTG, aerobic
MrsR	pET28a:*mrsR*	*E. coli* LOBSTR	400 µM IPTG, aerobic
MrsY	pCDFDuet:*mrsY(op)*	*E. coli* Rosetta	Co-expression with pPH151; final OD_{600} 0.9; 1000 µM IPTG, anaerobic (after addition of IPTG, flasks were purged with argon)

6.2.5.2 Purification of soluble 6xHis-tagged proteins and peptides under aerobic conditions

Cells were harvested from expression cultures by centrifugation (10,000 rpm, 4 ºC, 2 min). Supernatant was discarded and the pellet was resuspended in 10 mL of lysis buffer by vortexing. Cells lysis was performed with a sonicator in 10-second intervals, and soluble fraction was recovered by centrifugation (10,000 rpm, 4 ºC, 10 min). In the meantime, 600 µL of Ni-NTA-agarose were washed twice with lysis buffer. Supernatant and Ni-NTA agarose were mixed into a new 50 mL screw cap tube and incubated for one hour (on ice with slight agitation) to allow the His-tagged proteins to bind the agarose.
Suspension was then poured onto a polypropylene column (Qiagen/Hilden, Germany) and the liquid was eluted by gravity. The proteins attached to the agarose were rinsed with 4 mL of wash buffer I and 4 mL of wash buffer II. Elution of the pure proteins was done by adding 2.5 mL of elution buffer. All fractions obtained in each step were stored for later analysis by SDS-PAGE.

6.2.5.3 Purification of soluble 6xHis-tagged proteins under anaerobic conditions

The radical SAM protein MrsY was purified under strict anaerobic conditions. All buffers and materials were placed into the glove box at least 20 h before use. Protein purification was carried out according

to section 5.2.5.2 with some modifications. Briefly, cell lysis was done by sonication (95% power) for 20 min and all purification steps were done at room temperature. The lysate was aliquoted to 2 mL and centrifuged for 10 min at 13,000 x g.

6.2.5.4 Chemical reconstitution of Fe-S clusters

Chemical reconstitution of Fe-S clusters in the MrsY radical SAM protein was carried out following the procedure developed by Freibert *et al*. (2018). Briefly, 3 mL of assay buffer containing 76 µM of pure protein were supplemented with a freshly prepared 100 mM DTT solution to reach a final DTT concentration of 8 mM. Reduction of the apoprotein was done at 4 ºC under strict anaerobic conditions for 3 h. For subsequent Fe-S reconstitution, eightfold molar excess of $FeCl_3$ were added to the reduced protein followed by incubated for 5 min. Lithium sulfide was then added at the same molar concentration as iron and the mixture was incubated for 1.5 h. To prevent precipitation, $FeCl_3$ and Li_2S were added slowly in two aliquots. Absorbance was measured every 30 min to monitor the Fe-S cluster assembly. The reconstituted protein was centrifuged for 20 min at 14,000 x g to remove eventual precipitants. Unbound iron and sulfide were finally removed via desalting and concentrator columns (Vivaspin 500, 30 kDa cutoff. Sartorius/Göttingen, Germany), and the concentrated holoprotein was concentrated in assay buffer to be used for *in vitro* assays (Freibert *et al*., 2018).

6.2.5.5 Desalting and concentration of pure proteins

Since the high concentration of imidazole in the elution buffer may lead to protein degradation and lack of activity, buffer exchange was carried out with PD-10 columns (Cytiva/Marlborough, Massachusetts) following the gravity protocol. Buffer utilized for this purpose depended on the enzyme requirements and the intended use. Proteins were then concentrated using a Vivaspin 500 filters (Merk/Darmstadt, Germany) with a maximum molecular weight cut-off of half the proteins size.

6.2.5.6 Protein concentration measurements

Protein concentration was determined by the method of Gill and Hippel (Gill & Hippel, 1989). The molar extinction coefficient at λ=280 nm was estimated by the number of tryptophans, tyrosines and disulfide bonds in the protein. Absorbance was measured with an Eppendorf Biophotometer kinetic in a 1 mm quartz cuvette. The photometer calculated the protein concentration in mg/mL after measurement from the provided molar extinction coefficient of the protein, calculated by the method of Gill and von Hippel.

6.2.5.7 Storage of proteins

After purification and buffer exchange, proteins and peptides were mixed with sterile glycerol (final concentration of 5% v/v) and stored at -80 ºC for no longer than 3 months. Whenever possible, proteins were used fresh for *in vitro* assays.

6.2.5.8 Visualization of proteins in SDS PAGE gels

Qualitative analysis of protein expression was carried out by SDS-PAGE electrophoresis. 12% acrylamide gels were used for analysis of proteins, while 18% gels were preferred to visualize the precursor peptides. Aliquots of 30 µL from each fraction obtained in section 5.2.5.2 were mixed with 10 µL of NuPAGE LDS Sample Buffer (4X) and 4 µL of NuPAGE Sample Reducing Agent (10X). Samples were boiled for 10 min at 70 ºC, then loaded onto the acrylamide gel and run with a voltage of

120 for 1.5 to 2 h (PowerPac, Bio Rad). After protein separation, the gel was immersed in a staining solution containing 0.1% w/v brilliant blue, 1% acetic acid and 4% methanol for 15 min. The excess of colorant was removed by incubating the gel in destaining solution (1% acetic acid and 2% methanol) overnight. Unstained protein standard, Broad Range (10-200 kDa) (NEB/Ipswich, USA) was used as reference to verify the correct size of the target proteins and peptides.

6.2.6 *In vitro* assays for reconstitution of the mrs biosynthesis

Unless otherwise stated, all reactions were mixed on ice and immediately transferred to a thermocycler preheated to 25 ºC. After incubation, reactions were stored at -20 ºC until analysis.

MrsM

Reconstitution of MrsM reaction consisted of 5 µM of MrsM and 50 µM of substrate (MrsA1 or MrsA2) in Tris buffer (50 mM, pH 7.8) containing 10 mM $MgCl_2$, 2 mM DTT, 2.5 mM ATP. The reaction volume was 100 µL and incubation time ranging 1 to 12 h were tested.

MrsJ

To reconstitute the PTMs installed by MrsJ, modified $MrsA1^M$ or $MrsA2^M$ which contained the dehydroamino acids was used as substrate. Enzyme substrate (2 µM and 100 µM for MrsJ and modified peptide, respectively) were mixed in a 50 µL containing 2.5 mM of co-factor. Six buffer systems were tested for optimal activity as shown in Table 23.

Table 23: Buffer composition for optimization of MrsJ activity.

Buffer composition	
Buffer 1. HEPES 100 mM NaCl 500 mM pH 7.5	Buffer 2. Tris 100 mM NaCl 500 mM pH 7.5
Buffer 3. HEPES 100 mM NaCl 500 mM pH 8.0	Buffer 4. Tris 100 mM NaCl 500 mM pH 8.0
Buffer 5. HEPES 100 mM NaCl 500 mM pH 7.0	Buffer 6. Tris 100 mM NaCl 500 mM pH 7.0

MrsKC

The reaction was prepared in a final volume of 100 µL. The reaction consisted of HEPES buffer (50 mM, pH 7.5) containing 2 mM DTT, 10 mM $MgCl_2$ and NTPs mixture (2.5 mM each). MrsKC and substrate peptide (MrsA3 and MrsA4) were added at 5 µM and 50 µM, respectively. Incubation was for 6 h.

MrsP

Protease activity of MrsP was tested by incubation of 5 µM of protein with individual substrate peptide: 80 µM of MrsA3, 75 µM of MrsA4, 100 µM of MrsA2, and 100 µM of MrsA5. 50 µL reactions were prepared in HEPES buffer (50 mM, pH 7.5) supplemented with 2 mM DTT. Incubation for 2, 12 and 24 h were tested.

MrsR

The purified MrsA5^{Y} was incubated under aerobic conditions with MrsR in Tris buffer (pH 8) for 12 h.

Factor Xa

Engineered peptides containing the Factor Xa recognition site were incubated with 1 µL of Factor Xa for 1 to 6 h at 23 ºC. Protease and $CaCl_2$ at final concentration of 2 mM were added directly to the main reaction for leader removal.

Trypsin

Trypsin was used to proteolyze MrsA2 after reconstitution of MrsM and MrsJ activity. For this purpose, 1 µL of trypsin (10 µg/mL) was added to the tube containing the modified peptide and incubated for 3 h at 25 ºC.

6.2.7 Analytical methods

6.2.7.1 Preparation of crude extracts from *Streptomyces* cultures

Extraction of crude extracts from *S. griseus* S4-7 was initially performed with different solvents like *n*-butanol, ethyl acetate, acetone and methanol to determine the best solvent for extraction of bioactive compounds. All samples obtained from heterologous expression and elicitation experiments were extracted with *n*-butanol.

a) Extraction of liquid cultures

Extraction was done with organic solvents. After fermentation, one volume of solvent was added directly to the flask containing the culture and incubated overnight at room temperature with slight shaking (120 rpm). On the following day, cell debris and aqueous phase were removed by centrifugation (5,000 rpm, 10 min, room temperature) and the organic phase was collected in a round flask. Extracts were recovered by solvent removal via rotary evaporation. Samples were resuspended with methanol, dried until constant weight and stored at -20 ºC until analysis.

b) Extraction of solid culture

Agar was cut into small pieces and transferred to an Erlenmeyer flask. Organic solvent was added to cover the agar pieces and the following steps were according to section 3.2.7.1 a).

6.2.7.2 MALDI measurements

Peptides obtained from *in vitro* assays were analyzed by MALDI-TOF mass spectrometry at the Mass Spectrometry Core Facility in the Institute of Biochemistry and Molecular Biology, University of Bonn. Samples were dialyzed for 10-30 min using MCE membranes (0.025 µm, 13 mm) or by using Ziptips C18. Then, 1 µL of sample was mixed with 1 µL of HCCA matrix solution (20 g/l HCCA in 50% ACN) and immediately transferred to the center of a spot on a ground steel MALDI target plate. After drying and cocrystallization of the matrix, the plate was inserted into the spectrometer. Measurements were done with a linear mode method for undigested peptides while those where leader peptide was already removed were analyzed on reflective mode. Protein Standard II (Bruker/Madison, United States) and peptide calibration standard (Bruker/Madison, United States) were used for calibration of method parameters for linear and reflective detector, respectively.

6.2.7.3 HPLC-MS analysis

HPLC-MS was carried out using a micrOTOF-QII mass spectrometer (Bruker) with ESI-source coupled with an HPLC Dionex Ultimate 3000 (Thermo Scientific) using an EC10/2 Nucleoshell C18 2.7 µm column (Macherey-Nagel). The column temperature was 25 °C. MS data were acquired over a range from 100 to 3000 *m/z* in positive mode. Auto MS/MS fragmentation was achieved with rising collision energy (35−50 keV over a gradient from *m/z* 500 to 2000) with a frequency of 4 Hz for all ions over a threshold of 100. HPLC was performed with a gradient with an initial mobile phase composition of 90% H_2O containing 0.1% AcOH. The gradient started after 1 min to 100% acetonitrile (0.1% AcOH) in 20 min. A 5 µL amount of a 2 mg/mL sample solution (MeOH) was injected at a flow of 0.3 mL/min. Data analysis was performed using Bruker Compass DataAnalysis version 4.2 (Build: 383.1).

6.2.8 Data analysis

Analysis of the HPLC-MS/MS data was carried out by using Data Analysis 4.2 (Bruker Daltonik GmbH) or MassHunter Qualitative Analysis B.05.00 (Agilent/California, United States).

6.2.8.1 Molecular networking

Molecular networks were created using the online workflow (https://ccms-ucsd.github.io/GNPSDocumentation/) on the GNPS website (http://gnps.ucsd.edu). The data was filtered by removing all MS/MS fragment ions within +/- 17 Da of the precursor m/z. MS/MS spectra were window filtered by choosing only the top 6 fragment ions in the +/- 50 Da window throughout the spectrum. The precursor ion mass tolerance was set to 2.0 Da and a MS/MS fragment ion tolerance of 0.5 Da. A network was then created where edges were filtered to have a cosine score above 0.7 and more than 6 matched peaks. Further, edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. Finally, the maximum size of a molecular family was set to 100, and the lowest scoring edges were removed from molecular families until the molecular family size was below this threshold. The spectra in the network were then searched against GNPS' spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 6 matched peaks. Visualization of networks was carried out using the Cytoscape 3.9.1 software (cytoscape.org).

6.2.9 Bioinformatic methods

6.2.9.1 AntiSMASH

AntiSMASH is a free online tool that allows rapid identification, annotation and analysis of biosynthetic gene clusters in bacterial genomes (Blin *et al*., 2021). After uploading a DNA-sequence in .gbk, .embl or .fasta-format, the software employs a large number of *in silico* tools, to perform an in-depth analysis of genes. Results can be investigated in a browser-supported surface or downloaded in .gbk-format. AntiSMASH is powered by several open-source tools such as NCBI BLAST+, HMMer 3, Muscle 3, FastTree, PySVG and JQuery SVG. https://antismash.secondarymetabolites.org/#!/start

6.2.9.2 RadicalSAM.org

RadicalSAM.org is a resource that helps to interpret sequence-function space and discover new radical SAM enzyme chemistry. This web-based tool enables a user-friendly genomic enzymology strategy to explore sequence-function space in the radical SAM superfamily. By using sequence similarity

networks (SSNs) and the genome context of bacterial, fungal and archaeal members, RadicalSAM.org allows the recognition of isofunctional groups radical SAM proteins (Oberg *et al.*, 2022).
The primary sequence of a target protein is submitted to RadicalSAM.org. The tool displays the cluster or megacluster that contains the most similar sequences to the input. Then, the cluster can be manually diced to obtain an isofunctional cluster of radical SAM protein sequences.

6.2.9.3 Enzyme Function Initiative (EFI) tools

Enzyme Function Initiative (EFI) provides a collection of webtools that use NCBI BLAST and CD-HIT for creating sequence similarity networks (SSNs) and genome neighborhood networks (GNNs). These resources aim to facilitate the discovery of *in vitro* enzymatic and *in vivo* metabolic/physiological functions of unknown enzymes discovered in genome projects. The tools use the CGFP programs from the Balskus Lab (https://bitbucket.org/biobakery/cgfp/src) and ShortBRED from the Huttenhower Lab (http://huttenhower.sph.harvard.edu/shortbred). The data used is obtained from the UniProt Consortium databases (UniProt Consortium, 2023), and the InterPro and ENA databases from EMBL-EBI (Madeira *et al*., 2022).

7. References

Abdelmohsen, U. R., Grkovic, T., Balasubramanian, S., Kamel, M. S., Quinn, R. J., & Hentschel, U. (2015). Elicitation of secondary metabolism in actinomycetes. *Biotechnology Advances*, *33*(6 Pt 1), 798–811. https://doi.org/10.1016/j.biotechadv.2015.06.003

Abdel-Razek, A. S., El-Naggar, M. E., Allam, A., Morsy, O. M., & Othman, S. I. (2020). Microbial natural products in drug discovery. *Processes, 8*(4), 470. http://dx.doi.org/10.3390/pr8040470

Agrawal, P., Khater, S., Gupta, M., Sain, N., & Mohanty, D. (2017). RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links. *Nucleic Acids Research, 45*(W1), W80–W88. https://doi.org/10.1093/nar/gkx408

Akhter, N., Liu, Y., Auckloo, B. N., Shi, Y., Wang, K., Chen, J., Wu, X., & Wu, B. (2018). Stress-driven discovery of new angucycline-type antibiotics from a marine *Streptomyces pratensis* NA-ZhouS1. *Marine Drugs, 16*(9), 331. https://doi.org/10.3390/md16090331

Andersen, K. R., Leksa, N. C., & Schwartz, T. U. (2013). Optimized *E. coli* expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins, 81*(11), 1857–1861. https://doi.org/10.1002/prot.24364

Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K. D., Fischbach, M. A., Garavelli, J. S., Göransson, U., … van der Donk, W. A. (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Natural Product Reports, 30*(1), 108–160. https://doi.org/10.1039/c2np20085f

Aron, A. T., Gentry, E. C., McPhail, K. L., Nothias, L. F., Nothias-Esposito, M., Bouslimani, A., Petras, D., Gauglitz, J. M., Sikora, N., Vargas, F., van der Hooft, J. J. J., Ernst, M., Kang, K. B., Aceves, C. M., Caraballo-Rodríguez, A. M., Koester, I., Weldon, K. C., Bertrand, S., Roullier, C., Sun, K., … Dorrestein, P. C. (2020). Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nature Protocols*, *15*(6), 1954–1991. https://doi.org/10.1038/s41596-020-0317-5

Atanasov, A. G., Zotchev, S. B., Dirsch, V. M., International Natural Product Sciences Taskforce, & Supuran, C. T. (2021). Natural products in drug discovery: advances and opportunities. *Nature Reviews. Drug Discovery, 20*(3), 200–216. https://doi.org/10.1038/s41573-020-00114-z

Barbosa, J. C., Mösker, E., Faria, R., Süssmuth, R. D., Mendo, S., & Caetano, T. (2023). Class II two-peptide lanthipeptide proteases: exploring LicTP for biotechnological applications. *Applied Microbiology and Biotechnology*, 10.1007/s00253-023-12388-5. Advance online publication. https://doi.org/10.1007/s00253-023-12388-5

Barr, I., Latham, J. A., Iavarone, A. T., Chantarojsiri, T., Hwang, J. D., & Klinman, J. P. (2016). Demonstration that the radical *S*-Adenosylmethionine (SAM) enzyme PqqE catalyzes de novo carbon-carbon cross-linking within a peptide substrate PqqA in the presence of the peptide chaperone PqqD. *The Journal of Biological Chemistry, 291*(17), 8877–8884. https://doi.org/10.1074/jbc.C115.699918

Bartholomae, M., Buivydas, A., Viel, J. H., Montalbán-López, M., & Kuipers, O. P. (2017). Major gene-regulatory mechanisms operating in ribosomally synthesized and post-translationally modified

peptide (RiPP) biosynthesis. *Molecular Microbiology, 106*(2), 186–206. https://doi.org/10.1111/mmi.13764

Bibb, M. J., Janssen, G. R., & Ward, J. M. (1985). Cloning and analysis of the promoter region of the erythromycin resistance gene (*ermE*) of *Streptomyces erythraeus*. *Gene, 38*(1-3), 215–226. https://doi.org/10.1016/0378-1119(85)90220-3

Bilyk, B., & Luzhetskyy, A. (2014). Unusual site-specific DNA integration into the highly active pseudo-*attB* of the *Streptomyces albus* J1074 genome. *Applied Microbiology and Biotechnology, 98*(11), 5095–5104. https://doi.org/10.1007/s00253-014-5605-y

Blanco, A. G., Sola, M., Gomis-Rüth, F. X., & Coll, M. (2002). Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator. *Structure* (London, England: 1993), *10*(5), 701–713. https://doi.org/10.1016/s0969-2126(02)00761-x

Bleich, R., Watrous, J. D., Dorrestein, P. C., Bowers, A. A., & Shank, E. A. (2015). Thiopeptide antibiotics stimulate biofilm formation in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America, 112*(10), 3086–3091. https://doi.org/10.1073/pnas.1414272112

Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., van Wezel, G. P., Medema, M. H., & Weber, T. (2021). AntiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research, 49*(W1), W29–W35. https://doi.org/10.1093/nar/gkab335

Boakes, S., Cortés, J., Appleyard, A. N., Rudd, B. A., & Dawson, M. J. (2009). Organization of the genes encoding the biosynthesis of actagardine and engineering of a variant generation system. *Molecular Microbiology, 72*(5), 1126–1136. https://doi.org/10.1111/j.1365-2958.2009.06708.x

Bonnot, F., Iavarone, A. T., & Klinman, J. P. (2013). Multistep, eight-electron oxidation catalyzed by the cofactorless oxidase, PqqC: identification of chemical intermediates and their dependence on molecular oxygen. *Biochemistry, 52*(27), 4667–4675. https://doi.org/10.1021/bi4003315

Bothwell, I. R., Cogan, D. P., Kim, T., Reinhardt, C. J., van der Donk, W. A., & Nair, S. K. (2019). Characterization of glutamyl-tRNA-dependent dehydratases using nonreactive substrate mimics. *Proceedings of the National Academy of Sciences of the United States of America, 116*(35), 17245–17250. https://doi.org/10.1073/pnas.1905240116

Burkhardt, A., Henry, P. M., Koike, S. T., Gordon, T. R., & Martin, F. (2019). Detection of *Fusarium oxysporum* f. sp. fragariae from infected strawberry plants. *Plant Disease, 103*(5), 1006–1013. https://doi.org/10.1094/PDIS-08-18-1315-RE

Burroughs, A. M., Zhang, D., Schäffer, D. E., Iyer, L. M., & Aravind, L. (2015). Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Research, 43*(22), 10633–10654. https://doi.org/10.1093/nar/gkv1267

Busche, T., Tsolis, K. C., Koepff, J., Rebets, Y., Rückert, C., Hamed, M. B., Bleidt, A., Wiechert, W., Lopatniuk, M., Yousra, A., Anné, J., Karamanou, S., Oldiges, M., Kalinowski, J., Luzhetskyy, A., & Economou, A. (2018). Multi-omics and targeted approaches to determine the role of cellular proteases in *Streptomyces* protein secretion. *Frontiers in Microbiology, 9*, 1174. https://doi.org/10.3389/fmicb.2018.01174

Bushin, L. B., Clark, K. A., Pelczer, I., & Seyedsayamdost, M. R. (2018). Charting an unexplored streptococcal biosynthetic landscape reveals a unique peptide cyclization motif. *Journal of the American Chemical Society, 140*(50), 17674–17684. https://doi.org/10.1021/jacs.8b10266

Caruso, A., Martinie, R. J., Bushin, L. B., & Seyedsayamdost, M. R. (2019). Macrocyclization via an arginine-tyrosine crosslink broadens the reaction scope of radical *S*-Adenosylmethionine enzymes. *Journal of the American Chemical Society, 141*(42), 16610–16614. https://doi.org/10.1021/jacs.9b09210

Cha, J. Y., Han, S., Hong, H. J., Cho, H., Kim, D., Kwon, Y., Kwon, S. K., Crüsemann, M., Bok Lee, Y., Kim, J. F., Giaever, G., Nislow, C., Moore, B. S., Thomashow, L. S., Weller, D. M., & Kwak, Y. S. (2016). Microbial and biochemical basis of a *Fusarium* wilt-suppressive soil. *International Society for Microbial Ecology Journal, 10*(1), 119–129. https://doi.org/10.1038/ismej.2015.95

Chang, P. T., Rao, K., Longo, L. O., Lawton, E. S., Scherer, G., & Van Arnam, E. B. (2020). Thiopeptide defense by an ant's bacterial symbiont. *Journal of Natural Products, 83*(3), 725–729. https://doi.org/10.1021/acs.jnatprod.9b00897

Charusanti, P., Fong, N. L., Nagarajan, H., Pereira, A. R., Li, H. J., Abate, E. A., Su, Y., Gerwick, W. H., & Palsson, B. O. (2012). Exploiting adaptive laboratory evolution of *Streptomyces clavuligerus* for antibiotic discovery and overproduction. *PloS One, 7*(3), e33727. https://doi.org/10.1371/journal.pone.0033727

Chater K. F. (1972). A morphological and genetic mapping study of white colony mutants of *Streptomyces coelicolor. Journal of General Microbiology, 72*(1), 9–28. https://doi.org/10.1099/00221287-72-1-9

Chen, S., Xu, B., Chen, E., Wang, J., Lu, J., Donadio, S., Ge, H., & Wang, H. (2019). Zn-dependent bifunctional proteases are responsible for leader peptide processing of class III lanthipeptides. *Proceedings of the National Academy of Sciences of the United States of America, 116*(7), 2533–2538. https://doi.org/10.1073/pnas.1815594116

Čihák, M., Kameník, Z., Šmídová, K., Bergman, N., Benada, O., Kofroňová, O., Petříčková, K., & Bobek, J. (2017). Secondary metabolites produced during the germination of *Streptomyces coelicolor*. *Frontiers in Microbiology, 8*, 2495. https://doi.org/10.3389/fmicb.2017.02495

Clark, K. A., & Seyedsayamdost, M. R. (2022). Bioinformatic atlas of radical SAM enzyme-modified RiPP natural products reveals an isoleucine-tryptophan crosslink. *Journal of the American Chemical Society, 144*(39), 17876–17888. https://doi.org/10.1021/jacs.2c06497

Clark, K. A., Bushin, L. B., & Seyedsayamdost, M. R. (2019). Aliphatic ether bond formation expands the scope of radical SAM enzymes in natural product biosynthesis. *Journal of the American Chemical Society, 141*(27), 10610–10615. https://doi.org/10.1021/jacs.9b05151

Commissariat à l'énergie atomique et aux énergies alternatives (CEA). (2021). The Fe-S radical-SAM enzymes. *Chemistry and Biology of Metals Laboratory*. https://www.cbm-lab.fr/en/Pages/Biocat/Thema_02.aspx

Cotter, P. D., O'Connor, P. M., Draper, L. A., Lawton, E. M., Deegan, L. H., Hill, C., & Ross, R. P. (2005). Posttranslational conversion of L-serines to D-alanines is vital for optimal production and activity of the lantibiotic lacticin 3147. *Proceedings of the National Academy of Sciences of the United States of America, 102*(51), 18584–18589. https://doi.org/10.1073/pnas.0509371102

Craney, A., Ahmed, S., & Nodwell, J. (2013). Towards a new science of secondary metabolism. *The Journal of Antibiotics, 66*(7), 387–400. https://doi.org/10.1038/ja.2013.25

Crüsemann, M., O'Neill, E. C., Larson, C. B., Melnik, A. V., Floros, D. J., da Silva, R. R., Jensen, P. R., Dorrestein, P. C., & Moore, B. S. (2017). Prioritizing natural product diversity in a collection of 146 bacterial strains based on growth and extraction protocols. *Journal of Natural Products, 80*(3), 588–597. https://doi.org/10.1021/acs.jnatprod.6b00722

Cuthbertson, L., & Nodwell, J. R. (2013). The TetR family of regulators. *Microbiology and Molecular Miology Reviews. 77*(3), 440–475. https://doi.org/10.1128/MMBR.00018-13

Dai, J., Wang, Y., Liu, J., & He, W. (2020). The regulatory genes involved in spiramycin and bitespiramycin biosynthesis. *Microbiological Research, 240*, 126532. https://doi.org/10.1016/j.micres.2020.126532

Dangel, V., Westrich, L., Smith, M. C., Heide, L., & Gust, B. (2010). Use of an inducible promoter for antibiotic production in a heterologous host. *Applied Microbiology and Biotechnology, 87*(1), 261–269. https://doi.org/10.1007/s00253-009-2435-4

Davis, K. M., Schramma, K. R., Hansen, W. A., Bacik, J. P., Khare, S. D., Seyedsayamdost, M. R., & Ando, N. (2017). Structures of the peptide-modifying radical SAM enzyme SuiB elucidate the basis of substrate recognition. *Proceedings of the National Academy of Sciences of the United States of America, 114*(39), 10420–10425. https://doi.org/10.1073/pnas.1703663114

De Simeis, D., & Serra, S. (2021). Actinomycetes: A Never-Ending Source of Bioactive Compounds-An Overview on Antibiotics Production. *Antibiotics* (Basel, Switzerland), *10*(5), 483. https://doi.org/10.3390/antibiotics10050483

Dias, D. A., Urban, S., & Roessner, U. (2012). A historical overview of natural products in drug discovery. *Metabolites, 2*(2), 303–336. https://doi.org/10.3390/metabo2020303

Diaz Ricci, J. C., & Hernández, M. E. (2000). Plasmid effects on *Escherichia coli* metabolism. *Critical Reviews in Biotechnology, 20*(2), 79–108. https://doi.org/10.1080/07388550008984167

Duan, L., Wang, S., Liao, R., & Liu, W. (2012). Insights into quinaldic acid moiety formation in thiostrepton biosynthesis facilitating fluorinated thiopeptide generation. *Chemistry & Biology, 19*(4), 443–448. https://doi.org/10.1016/j.chembiol.2012.02.008

Dunbar, K. L., Tietz, J. I., Cox, C. L., Burkhart, B. J., & Mitchell, D. A. (2015). Identification of an auxiliary leader peptide-binding protein required for azoline formation in ribosomal natural products. *Journal of the American Chemical Society, 137*(24), 7672–7677. https://doi.org/10.1021/jacs.5b04682

Enghiad, B., Huang, C., Guo, F., Jiang, G., Wang, B., Tabatabaei, S. K., Martin, T. A., & Zhao, H. (2021). Cas12a-assisted precise targeted cloning using *in vivo* Cre-lox recombination. *Nature Communications, 12*(1), 1171. https://doi.org/10.1038/s41467-021-21275-4

Feng, J., Wu, J., Dai, N., Lin, S., Xu, H. H., Deng, Z., & He, X. (2013). Discovery and characterization of BlsE, a radical *S*-adenosyl-*L*-methionine decarboxylase involved in the blasticidin S biosynthetic pathway. *PloS One, 8*(7), e68545. https://doi.org/10.1371/journal.pone.0068545

Flinspach, K., Kapitzke, C., Tocchetti, A., Sosio, M., & Apel, A. K. (2014). Heterologous expression of the thiopeptide antibiotic GE2270 from *Planobispora rosea* ATCC 53733 in *Streptomyces coelicolor*

requires deletion of ribosomal genes from the expression construct. *PloS One, 9*(3), e90499. https://doi.org/10.1371/journal.pone.0090499

Flinspach, K., Westrich, L., Kaysser, L., Siebenberg, S., Gomez-Escribano, J. P., Bibb, M., Gust, B., & Heide, L. (2010). Heterologous expression of the biosynthetic gene clusters of coumermycin A(1), clorobiocin and caprazamycins in genetically modified *Streptomyces coelicolor* strains. *Biopolymers, 93*(9), 823–832. https://doi.org/10.1002/bip.21493

Francis, D. M., & Page, R. (2010). Strategies to optimize protein expression in *E. coli*. *Current Protocols in Protein Science, Chapter 5*(1), 5.24.1–5.24.29. https://doi.org/10.1002/0471140864.ps0524s61

Freibert, S. A., Weiler, B. D., Bill, E., Pierik, A. J., Mühlenhoff, U., & Lill, R. (2018). Biochemical reconstitution and spectroscopic analysis of iron-sulfur proteins. *Methods in Enzymology, 599*, 197–226. https://doi.org/10.1016/bs.mie.2017.11.034

Gallia, J., Lavrich, K., Tan-Wilson, A., & Madden, P. H. (2013). Filtering of MS/MS data for peptide identification. *BMC Genomics,* 14 Suppl 7(Suppl 7), S2. https://doi.org/10.1186/1471-2164-14-S7-S2

Genilloud O. (2018). Mining Actinomycetes for novel antibiotics in the omics era: are we ready to exploit this new paradigm?. *Antibiotics* (Basel, Switzerland), *7*(4), 85. https://doi.org/10.3390/antibiotics7040085

Genome *Streptomyces*. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004 – [cited 2023 Jun 17]. Available from: https://www.ncbi.nlm.nih.gov/gene/

Gill, S. C., & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry, 182*(2), 319–326. https://doi.org/10.1016/0003-2697(89)90602-7

Gomez-Escribano, J. P., & Bibb, M. J. (2011). Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microbial Biotechnology, 4*(2), 207–215. https://doi.org/10.1111/j.1751-7915.2010.00219.x

Gomez-Escribano, J. P., & Bibb, M. J. (2012). *Streptomyces coelicolor* as an expression host for heterologous gene clusters. *Methods in Enzymology, 517*, 279–300. https://doi.org/10.1016/B978-0-12-404634-4.00014-0

Green, M. R., & Sambrook, J. (2019). Polymerase chain reaction (PCR) amplification of GC-rich templates. *Cold Spring Harbor Protocols, 2019*(2), 10.1101/pdb.prot095141. https://doi.org/10.1101/pdb.prot095141

Guo, J., Zhang, X., Lu, X., Liu, W., Chen, Z., Li, J., Deng, L., & Wen, Y. (2018). SAV4189, a MarR-family regulator in *Streptomyces avermitilis*, activates avermectin biosynthesis. *Frontiers in Microbiology, 9*, 1358. https://doi.org/10.3389/fmicb.2018.01358

Gust B. (2009). Chapter 7. Cloning and analysis of natural product pathways. *Methods in Enzymology, 458*, 159–180. https://doi.org/10.1016/S0076-6879(09)04807-1

Haft, D. H., & Basu, M. K. (2011). Biological systems discovery in silico: radical S-adenosylmethionine protein families and their target peptides for posttranslational modification. *Journal of Bacteriology*, *193*(11), 2745–2755. https://doi.org/10.1128/JB.00040-11

Håvarstein, L. S., Diep, D. B., & Nes, I. F. (1995). A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. *Molecular Microbiology, 16*(2), 229–240. https://doi.org/10.1111/j.1365-2958.1995.tb02295.x

He, B. B., Cheng, Z., Zhong, Z., Gao, Y., Liu, H., & Li, Y. X. (2022). Expanded sequence space of radical *S*-Adenosylmethionine-dependent enzymes involved in post-translational macrocyclization. *Angewandte Chemie* (International ed. in English), *61*(48), e202212447. https://doi.org/10.1002/anie.202212447

Hegemann, J. D., & Süssmuth, R. D. (2020). Matters of class: coming of age of class III and IV lanthipeptides. *Royal Society of Chemistry. Chemical Biology, 1*(3), 110–127. https://doi.org/10.1039/d0cb00073f

Hegemann, J. D., & van der Donk, W. A. (2018). Investigation of substrate recognition and biosynthesis in class IV lanthipeptide systems. *Journal of the American Chemical Society, 140*(17), 5743–5754. https://doi.org/10.1021/jacs.8b01323

Hegemann, J. D., Bobeica, S. C., Walker, M. C., Bothwell, I. R., & van der Donk, W. A. (2019). Assessing the flexibility of the prochlorosin 2.8 scaffold for bioengineering applications. *American Chemical Society. Synthetic Biology, 8*(5), 1204–1214. https://doi.org/10.1021/acssynbio.9b00080

Hibbing, M. E., Fuqua, C., Parsek, M. R., & Peterson, S. B. (2010). Bacterial competition: surviving and thriving in the microbial jungle. *Nature Reviews. Microbiology, 8*(1), 15–25. https://doi.org/10.1038/nrmicro2259

Higo, A., Hara, H., Horinouchi, S., & Ohnishi, Y. (2012). Genome-wide distribution of AdpA, a global regulator for secondary metabolism and morphological differentiation in *Streptomyces*, revealed the extent and complexity of the AdpA regulatory network. *DNA Research: an International Journal for Rapid Publication of Reports on Genes and Genomes, 19*(3), 259–273. https://doi.org/10.1093/dnares/dss010

Hiller, K., Grote, A., Scheer, M., Münch, R., & Jahn, D. (2004). PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Research, 32*(Web Server issue), W375–W379. https://doi.org/10.1093/nar/gkh378

Himes, P. M., Allen, S. E., Hwang, S., & Bowers, A. A. (2016). Production of sactipeptides in *Escherichia coli*: probing the substrate promiscuity of subtilosin A biosynthesis. *American Chemical Society. Chemical Biology, 11*(6), 1737–1744. https://doi.org/10.1021/acschembio.6b00042

Hoshino, S., Okada, M., Wakimoto, T., Zhang, H., Hayashi, F., Onaka, H., & Abe, I. (2015). Niizalactams A-C, multicyclic macrolactams isolated from combined culture of *Streptomyces* with mycolic acid-containing bacterium. *Journal of Natural Products, 78*(12), 3011–3017. https://doi.org/10.1021/acs.jnatprod.5b00804

Hudson, G. A., & Mitchell, D. A. (2018). RiPP antibiotics: biosynthesis and engineering potential. *Current Opinion in Microbiology, 45*, 61–69. https://doi.org/10.1016/j.mib.2018.02.010

Huo, L., & van der Donk, W. A. (2016). Discovery and characterization of bicereucin, an unusual *d*-amino acid-containing mixed two-component lantibiotic. *Journal of the American Chemical Society, 138*(16), 5254–5257. https://doi.org/10.1021/jacs.6b02513

Huo, L., Rachid, S., Stadler, M., Wenzel, S. C., & Müller, R. (2012). Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. *Chemistry & Biology, 19*(10), 1278–1287. https://doi.org/10.1016/j.chembiol.2012.08.013

Iorio, M., Sasso, O., Maffioli, S. I., Bertorelli, R., Monciardini, P., Sosio, M., Bonezzi, F., Summa, M., Brunati, C., Bordoni, R., Corti, G., Tarozzo, G., Piomelli, D., Reggiani, A., & Donadio, S. (2014). A glycosylated, labionin-containing lanthipeptide with marked antinociceptive activity. *American Chemical Society. Chemical Biology, 9*(2), 398–404. https://doi.org/10.1021/cb400692w

Jia, B., Jin, Z. H., Lei, Y. L., Mei, L. H., & Li, N. H. (2006). Improved production of pristinamycin coupled with an adsorbent resin in fermentation by *Streptomyces pristinaespiralis*. *Biotechnology Letters, 28*(22), 1811–1815. https://doi.org/10.1007/s10529-006-9157-9

Jiang, J., Sun, Y. F., Tang, X., He, C. N., Shao, Y. L., Tang, Y. J., & Zhou, W. W. (2018). Alkaline pH shock enhanced production of validamycin A in fermentation of *Streptomyces hygroscopicus*. *Bioresource Technology, 249*, 234–240. https://doi.org/10.1016/j.biortech.2017.10.012

Jones, S. E., & Elliot, M. A. (2018). 'Exploring' the regulation of *Streptomyces* growth and development. *Current Opinion in Microbiology, 42*, 25–30. https://doi.org/10.1016/j.mib.2017.09.009

Jones, S. E., Ho, L., Rees, C. A., Hill, J. E., Nodwell, J. R., & Elliot, M. A. (2017). *Streptomyces* exploration is triggered by fungal interactions and volatile signals. *eLife, 6*, e21738. https://doi.org/10.7554/eLife.21738

Joska, T. M., Mashruwala, A., Boyd, J. M., & Belden, W. J. (2014). A universal cloning method based on yeast homologous recombination that is simple, efficient, and versatile. *Journal of Microbiological Methods*, 100, 46–51. https://doi.org/10.1016/j.mimet.2013.11.013

Kalia, D., Merey, G., Nakayama, S., Zheng, Y., Zhou, J., Luo, Y., Guo, M., Roembke, B. T., & Sintim, H. O. (2013). Nucleotide, c-di-GMP, c-di-AMP, cGMP, cAMP, (p)ppGpp signaling in bacteria and implications in pathogenesis. *Chemical Society Reviews, 42*(1), 305–341. https://doi.org/10.1039/c2cs35206k

Kalkreuter, E., Pan, G., Cepeda, A. J., & Shen, B. (2020). Targeting bacterial genomes for natural product discovery. *Trends in Pharmacological Sciences, 41*(1), 13–26. https://doi.org/10.1016/j.tips.2019.11.002

Karbalaei-Heidari, H. R., & Budisa, N. (2020). Combating antimicrobial resistance with new-to-nature lanthipeptides created by genetic code expansion. *Frontiers in Microbiology, 11*, 590522. https://doi.org/10.3389/fmicb.2020.590522

Katz, L., & Baltz, R. H. (2016). Natural product discovery: past, present, and future. *Journal of Industrial Microbiology & Biotechnology, 43*(2-3), 155–176. https://doi.org/10.1007/s10295-015-1723-5

Kaur, G., Burroughs, A. M., Iyer, L. M., & Aravind, L. (2020). Highly regulated, diversifying NTP-dependent biological conflict systems with implications for the emergence of multicellularity. *eLife, 9*, e52696. https://doi.org/10.7554/eLife.52696

Kersten, R. D., Yang, Y. L., Xu, Y., Cimermancic, P., Nam, S. J., Fenical, W., Fischbach, M. A., Moore, B. S., & Dorrestein, P. C. (2011). A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nature Chemical Biology, 7*(11), 794–802. https://doi.org/10.1038/nchembio.684

Khaliullin, B., Aggarwal, P., Bubas, M., Eaton, G. R., Eaton, S. S., & Latham, J. A. (2016). Mycofactocin biosynthesis: modification of the peptide MftA by the radical *S*-adenosylmethionine protein MftC. *FEBS Letters, 590*(16), 2538–2548. https://doi.org/10.1002/1873-3468.12249

Khaliullin, B., Ayikpoe, R., Tuttle, M., & Latham, J. A. (2017). Mechanistic elucidation of the mycofactocin-biosynthetic radical *S*-adenosylmethionine protein, MftC. *The Journal of Biological Chemistry, 292*(31), 13022–13033. https://doi.org/10.1074/jbc.M117.795682

Kim, D. R., Cho, G., Jeon, C. W., Weller, D. M., Thomashow, L. S., Paulitz, T. C., & Kwak, Y. S. (2019). A mutualistic interaction between *Streptomyces* bacteria, strawberry plants and pollinating bees. *Nature Communications, 10*(1), 4802. https://doi.org/10.1038/s41467-019-12785-3

Kim, D. R., Jeon, C. W., Shin, J. H., Weller, D. M., Thomashow, L., & Kwak, Y. S. (2019). Function and distribution of a lantipeptide in strawberry *Fusarium* wilt disease-suppressive soils. *Molecular Plant-Microbe Interactions MPMI, 32*(3), 306–312. https://doi.org/10.1094/MPMI-05-18-0129-R

Kloosterman, A. M., Cimermancic, P., Elsayed, S. S., Du, C., Hadjithomas, M., Donia, M. S., Fischbach, M. A., van Wezel, G. P., & Medema, M. H. (2020). Expansion of RiPP biosynthetic space through integration of pan-genomics and machine learning uncovers a novel class of lanthipeptides. *PLoS Biology, 18*(12), e3001026. https://doi.org/10.1371/journal.pbio.3001026

Kodani, S., Hudson, M. E., Durrant, M. C., Buttner, M. J., Nodwell, J. R., & Willey, J. M. (2004). The SapB morphogen is a lantibiotic-like peptide derived from the product of the developmental gene ramS in *Streptomyces coelicolor*. *Proceedings of the National Academy of Sciences of the United States of America, 101*(31), 11448–11453. https://doi.org/10.1073/pnas.0404220101

Kodani, S., Lodato, M. A., Durrant, M. C., Picart, F., & Willey, J. M. (2005). SapT, a lanthionine-containing peptide involved in aerial hyphae formation in the Streptomycetes. *Molecular Microbiology, 58*(5), 1368–1380. https://doi.org/10.1111/j.1365-2958.2005.04921.x

Kong, D., Wang, X., Nie, J., & Niu, G. (2019). Regulation of antibiotic production by signaling molecules in *Streptomyces. Frontiers in Microbiology, 10*, 2927. https://doi.org/10.3389/fmicb.2019.02927

Kostenko, A., Lien, Y., Mendauletova, A., Ngendahimana, T., Novitskiy, I. M., Eaton, S. S., & Latham, J. A. (2022). Identification of a poly-cyclopropylglycine-containing peptide via bioinformatic mapping of radical *S*-adenosylmethionine enzymes. *The Journal of Biological Chemistry, 298*(5), 101881. https://doi.org/10.1016/j.jbc.2022.101881

Krause, J., Handayani, I., Blin, K., Kulik, A., & Mast, Y. (2020). Disclosing the potential of the SARP-type regulator PapR2 for the activation of antibiotic gene clusters in *Streptomycetes*. *Frontiers in Microbiology*, 11, 225. https://doi.org/10.3389/fmicb.2020.00225

Krawczyk, B., Ensle, P., Müller, W. M., & Süssmuth, R. D. (2012). Deuterium labeled peptides give insights into the directionality of class III lantibiotic synthetase LabKC. *Journal of the American Chemical Society, 134*(24), 9922–9925. https://doi.org/10.1021/ja3040224

Lagedroste, M., Smits, S. H. J., & Schmitt, L. (2017). Substrate specificity of the secreted nisin leader peptidase NisP. *Biochemistry, 56*(30), 4005–4014. https://doi.org/10.1021/acs.biochem.7b00524

Lawton, E. M., Cotter, P. D., Hill, C., & Ross, R. P. (2007). Identification of a novel two-peptide lantibiotic, haloduracin, produced by the alkaliphile *Bacillus halodurans* C-125. *Federation of*

European Microbiological Societies. Microbiology Letters, 267(1), 64–71. https://doi.org/10.1111/j.1574-6968.2006.00539.x

Lee, J. C., Park, H. R., Park, D. J., Lee, H. B., Kim, Y. B., & Kim, C. J. (2003). Improved production of teicoplanin using adsorbent resin in fermentations. *Letters in Applied Microbiology, 37*(3), 196–200. https://doi.org/10.1046/j.1472-765x.2003.01374.x

Leitner, A., Foettinger, A., & Lindner, W. (2007). Improving fragmentation of poorly fragmenting peptides and phosphopeptides during collision-induced dissociation by malondialdehyde modification of arginine residues. *Journal of Mass Spectrometry: JMS, 42*(7), 950–959. https://doi.org/10.1002/jms.1233

Lesueur, L. L., Mir, L. M., & André, F. M. (2016). Overcoming the specific toxicity of large plasmids electrotransfer in primary cells *in vitro*. *Molecular Therapy. Nucleic Acids, 5*(3), e291. https://doi.org/10.1038/mtna.2016.4

Lewis, J. K., Jochimsen, A. S., Lefave, S. J., Young, A. P., Kincannon, W. M., Roberts, A. G., Kieber-Emmons, M. T., & Bandarian, V. (2021). New role for radical SAM enzymes in the biosynthesis of thio(seleno)oxazole RiPP natural products. *Biochemistry, 60*(45), 3347–3361. https://doi.org/10.1021/acs.biochem.1c00469

Li, C., Alam, K., Zhao, Y., Hao, J., Yang, Q., Zhang, Y., Li, R., & Li, A. (2021). Mining and biosynthesis of bioactive lanthipeptides from microorganisms. *Frontiers in Bioengineering and Biotechnology*, 9, 692466. https://doi.org/10.3389/fbioe.2021.692466

Li, J., Li, Y., Niu, G., Guo, H., Qiu, Y., Lin, Z., Liu, W., & Tan, H. (2018). NosP-regulated nosiheptide production responds to both peptidyl and small-molecule ligands derived from the precursor peptide. *Cell Chemical Biology, 25*(2), 143–153.e4. https://doi.org/10.1016/j.chembiol.2017.10.012

Li, L., Maclntyre, L. W., & Brady, S. F. (2021). Refactoring biosynthetic gene clusters for heterologous production of microbial natural products. *Current Opinion in Biotechnology*, 69, 145–152. https://doi.org/10.1016/j.copbio.2020.12.011

Linares-Otoya, L., Linares-Otoya, V., Armas-Mantilla, L., Blanco-Olano, C., Crüsemann, M., Ganoza-Yupanqui, M. L., Campos-Florian, J., König, G. M., & Schäberle, T. F. (2017). Identification and heterologous expression of the kocurin biosynthetic gene cluster. *Microbiology* (Reading, England), *163*(10), 1409–1414. https://doi.org/10.1099/mic.0.000538

Lohans, C. T., Li, J. L., & Vederas, J. C. (2014). Structure and biosynthesis of carnolysin, a homologue of enterococcal cytolysin with *D*-amino acids. *Journal of the American Chemical Society, 136*(38), 13150–13153. https://doi.org/10.1021/ja5070813

Ma, S., Chen, H., Li, H., Ji, X., Deng, Z., Ding, W., & Zhang, Q. (2021). Post-Translational Formation of Aminomalonate by a Promiscuous Peptide-Modifying Radical SAM Enzyme. *Angewandte Chemie* (International ed. in English), *60*(36), 19957–19964. https://doi.org/10.1002/anie.202107192

MacNeil, D. J., Gewain, K. M., Ruby, C. L., Dezeny, G., Gibbons, P. H., & MacNeil, T. (1992). Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene, 111*(1), 61–68. https://doi.org/10.1016/0378-1119(92)90603-m

Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., & Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in

2022. *Nucleic Acids Research, 50*(W1), W276–W279. Advance online publication. https://doi.org/10.1093/nar/gkac240

Magnusson, O. T., Toyama, H., Saeki, M., Rojas, A., Reed, J. C., Liddington, R. C., Klinman, J. P., & Schwarzenbacher, R. (2004). Quinone biogenesis: structure and mechanism of PqqC, the final catalyst in the production of pyrroloquinoline quinone. *Proceedings of the National Academy of Sciences of the United States of America, 101*(21), 7913–7918. https://doi.org/10.1073/pnas.0402640101

Maiti, P. K., Das, S., Sahoo, P., & Mandal, S. (2020). *Streptomyces* sp. SM01 isolated from Indian soil produces a novel antibiotic picolinamycin effective against multi drug resistant bacterial strains. *Scientific Reports, 10*(1), 10092. https://doi.org/10.1038/s41598-020-66984-w

Major, D., Flanzbaum, L., Lussier, L., Davies, C., Caldo, K., & Acedo, J. Z. (2021). Transporter protein-guided genome mining for head-to-tail cyclized bacteriocins. *Molecules* (Basel, Switzerland), *26*(23), 7218. https://doi.org/10.3390/molecules26237218

Malcolmson, S. J., Young, T. S., Ruby, J. G., Skewes-Cox, P., & Walsh, C. T. (2013). The posttranslational modification cascade to the thiopeptide berninamycin generates linear forms and altered macrocyclic scaffolds. *Proceedings of the National Academy of Sciences of the United States of America, 110*(21), 8483–8488. https://doi.org/10.1073/pnas.1307111110

Marfey, P. (1984). Determination of *D*-amino acids. II. Use of a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzene. *Carlsberg Research Communications. 49*, 591. https://doi.org/10.1007/BF02908688

Martínez-Burgo, Y., Santos-Aberturas, J., Rodríguez-García, A., Barreales, E. G., Tormo, J. R., Truman, A. W., Reyes, F., Aparicio, J. F., & Liras, P. (2019). Activation of secondary metabolite gene clusters in *Streptomyces clavuligerus* by the PimM regulator of *Streptomyces natalensis*. *Frontiers in Microbiology*, 10, 580. https://doi.org/10.3389/fmicb.2019.00580

Mast, Y., Guezguez, J., Handel, F., & Schinko, E. (2015). A complex signaling cascade governs pristinamycin biosynthesis in *Streptomyces pristinaespiralis*. *Applied and Environmental Microbiology, 81*(19), 6621–6636. https://doi.org/10.1128/AEM.00728-15

McAuliffe, O., Hill, C., & Ross, R. P. (2000). Each peptide of the two-component lantibiotic lacticin 3147 requires a separate modification enzyme for activity. *Microbiology* (Reading, England), *146* (Pt 9), 2147–2154. https://doi.org/10.1099/00221287-146-9-2147

Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., de Bruijn, I., Chooi, Y. H., Claesen, J., Coates, R. C., Cruz-Morales, P., Duddela, S., Düsterhus, S., Edwards, D. J., Fewer, D. P., Garg, N., Geiger, C., Gomez-Escribano, J. P., Greule, A., Hadjithomas, M., … Glöckner, F. O. (2015). Minimum information about a biosynthetic gene cluster. *Nature Chemical Biology, 11*(9), 625–631. https://doi.org/10.1038/nchembio.1890

Meindl, K., Schmiederer, T., Schneider, K., Reicke, A., Butz, D., Keller, S., Gühring, H., Vértesy, L., Wink, J., Hoffmann, H., Brönstrup, M., Sheldrick, G. M., & Süssmuth, R. D. (2010). Labyrinthopeptins: a new class of carbacyclic lantibiotics. *Angewandte Chemie* (International ed. in English), *49*(6), 1151–1154. https://doi.org/10.1002/anie.200905773

Menegatti, C., Lourenzon, V. B., Rodríguez-Hernández, D., da Paixão Melo, W. G., Ferreira, L., Andricopulo, A. D., do Nascimento, F. S., & Pupo, M. T. (2020). Meliponamycins: antimicrobials from stingless bee-associated *Streptomyces* sp. *Journal of Natural Products, 83*(3), 610–616. https://doi.org/10.1021/acs.jnatprod.9b01011

Merrick M. J. (1976). A morphological and genetic mapping study of bald colony mutants of *Streptomyces coelicolor*. *Journal of General Microbiology, 96*(2), 299–315. https://doi.org/10.1099/00221287-96-2-299

Merwin, N. J., Mousa, W. K., Dejong, C. A., Skinnider, M. A., Cannon, M. J., Li, H., Dial, K., Gunabalasingam, M., Johnston, C., & Magarvey, N. A. (2020). DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products. *Proceedings of the National Academy of Sciences of the United States of America, 117*(1), 371–380. https://doi.org/10.1073/pnas.1901493116

Mingyar, E., Mühling, L., Kulik, A., Winkler, A., Wibberg, D., Kalinowski, J., Blin, K., Weber, T., Wohlleben, W., & Stegmann, E. (2021). A regulator based "semi-targeted" approach to activate silent biosynthetic gene clusters. *International Journal of Molecular Sciences, 22*(14), 7567. https://doi.org/10.3390/ijms22147567

Mo, S., Kim, J. H., & Oh, C. H. (2013). Different effects of acidic pH shock on the prodiginine production in *Streptomyces coelicolor* M511 and SJM1 mutants. *Journal of Microbiology and Biotechnology, 23*(10), 1454–1459. https://doi.org/10.4014/jmb.1307.07067

Mohimani, H., Gurevich, A., Mikheenko, A., Garg, N., Nothias, L. F., Ninomiya, A., Takada, K., Dorrestein, P. C., & Pevzner, P. A. (2017). Dereplication of peptidic natural products through database search of mass spectra. *Nature Chemical Biology, 13*(1), 30–37. https://doi.org/10.1038/nchembio.2219

Mohimani, H., Gurevich, A., Shlemov, A., Mikheenko, A., Korobeynikov, A., Cao, L., Shcherbin, E., Nothias, L. F., Dorrestein, P. C., & Pevzner, P. A. (2018). Dereplication of microbial metabolites through database search of mass spectra. *Nature Communications, 9*(1), 4035. https://doi.org/10.1038/s41467-018-06082-8

Mohimani, H., Kersten, R. D., Liu, W. T., Wang, M., Purvine, S. O., Wu, S., Brewer, H. M., Pasa-Tolic, L., Bandeira, N., Moore, B. S., Pevzner, P. A., & Dorrestein, P. C. (2014). Automated genome mining of ribosomal peptide natural products. *American Chemical Society. Chemical Biology, 9*(7), 1545–1551. https://doi.org/10.1021/cb500199h

Montalbán-López, M., Scott, T. A., Ramesh, S., Rahman, I. R., van Heel, A. J., Viel, J. H., Bandarian, V., Dittmann, E., Genilloud, O., Goto, Y., Grande Burgos, M. J., Hill, C., Kim, S., Koehnke, J., Latham, J. A., Link, A. J., Martínez, B., Nair, S. K., Nicolet, Y., Rebuffat, S., … van der Donk, W. A. (2021). New developments in RiPP discovery, enzymology and engineering. *Natural Product Reports, 38*(1), 130–239. https://doi.org/10.1039/d0np00027b

Murakami, T., Holt, T. G., & Thompson, C. J. (1989). Thiostrepton-induced gene expression in *Streptomyces lividans*. *Journal of Bacteriology, 171*(3), 1459–1466. https://doi.org/10.1128/jb.171.3.1459-1466.1989

Murray P. R. (2015). The clinician and the microbiology laboratory. *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, 191–223. https://doi.org/10.1016/B978-1-4557-4801-3.00016-3

Myronovskyi, M., & Luzhetskyy, A. (2019). Heterologous production of small molecules in the optimized *Streptomyces* hosts. *Natural Product Reports, 36*(9), 1281–1294. https://doi.org/10.1039/c9np00023b

Narsing Rao, M. P., Dong, Z. Y., Liu, G. H., Li, L., Xiao, M., & Li, W. J. (2019). Reclassification of *Bacillus aryabhattai* Shivaji *et al*. 2009 as a later heterotypic synonym of *Bacillus megaterium* de Bary

1884 (Approved Lists 1980). *Federation of European Microbiological Societies Microbiology Letters, 366*(22), fnz258. https://doi.org/10.1093/femsle/fnz258

Nett, M., Ikeda, H., & Moore, B. S. (2009). Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Natural Product Reports, 26*(11), 1362–1384. https://doi.org/10.1039/b817069j

Newman, D. J., & Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *Journal of Natural Products, 83*(3), 770–803. https://doi.org/10.1021/acs.jnatprod.9b01285

Nguyen, H., Made Kresna, I. D., Böhringer, N., Ruel, J., Mora, E., Kramer, J. C., Lewis, K., Nicolet, Y., Schäberle, T. F., & Yokoyama, K. (2022). Characterization of a radical SAM oxygenase for the ether crosslinking in darobactin biosynthesis. *Journal of the American Chemical Society, 144*(41), 18876–18886. https://doi.org/10.1021/jacs.2c05565

Nguyen, T. Q. N., Tooh, Y. W., Sugiyama, R., Nguyen, T. P. D., Purushothaman, M., Leow, L. C., Hanif, K., Yong, R. H. S., Agatha, I., Winnerdy, F. R., Gugger, M., Phan, A. T., & Morinaka, B. I. (2020). Post-translational formation of strained cyclophanes in bacteria. *Nature Chemistry, 12*(11), 1042–1053. https://doi.org/10.1038/s41557-020-0519-z

Oberg, N., Precord, T. W., Mitchell, D. A., & Gerlt, J. A. (2022). RadicalSAM.org: a resource to interpret sequence-function space and discover new radical SAM enzyme chemistry. *ACS Bio & Med Chem Au, 2*(1), 22–35. https://doi.org/10.1021/acsbiomedchemau.1c00048

Okada, B. K., & Seyedsayamdost, M. R. (2017). Antibiotic dialogues: induction of silent biosynthetic gene clusters by exogenous small molecules. *Federation of European Microbiological Societies Microbiology Reviews, 41*(1), 19–33. https://doi.org/10.1093/femsre/fuw035

Ongey, E. L., & Neubauer, P. (2016). Lanthipeptides: chemical synthesis versus *in vivo* biosynthesis as tools for pharmaceutical production. *Microbial Cell Factories*, 15, 97. https://doi.org/10.1186/s12934-016-0502-y

Ordóñez-Robles, M., Rodríguez-García, A., & Martín, J. F. (2016). Target genes of the *Streptomyces tsukubaensis* FkbN regulator include most of the tacrolimus biosynthesis genes, a phosphopantetheinyl transferase and other PKS genes. *Applied Microbiology and Biotechnology, 100*(18), 8091–8103. https://doi.org/10.1007/s00253-016-7696-0

Ortega, M. A., Velásquez, J. E., Garg, N., Zhang, Q., Joyce, R. E., Nair, S. K., & van der Donk, W. A. (2014). Substrate specificity of the lanthipeptide peptidase ElxP and the oxidoreductase ElxO. *American Chemical Society. Chemical Biology, 9*(8), 1718–1725. https://doi.org/10.1021/cb5002526

Pérez-Rueda, E., & Collado-Vides, J. (2001). Common history at the origin of the position-function correlation in transcriptional regulators in archaea and bacteria. *Journal of Molecular Evolution, 53*(3), 172–179. https://doi.org/10.1007/s002390010207

Pham, J. V., Yilma, M. A., Feliz, A., Majid, M. T., Maffetone, N., Walker, J. R., Kim, E., Cho, H. J., Reynolds, J. M., Song, M. C., Park, S. R., & Yoon, Y. J. (2019). A review of the microbial production of bioactive natural products and biologics. *Frontiers in Microbiology*, 10, 1404. https://doi.org/10.3389/fmicb.2019.01404

Pham, V., Nguyen, H. T., Nguyen, C. T., Choi, Y. S., Dhakal, D., Kim, T. S., Jung, H. J., Yamaguchi, T., & Sohng, J. K. (2021). Identification and enhancing production of a novel macrolide compound in

engineered *Streptomyces peucetius*. *Royal Society of Chemistry Advances, 11*(5), 3168–3173. https://doi.org/10.1039/d0ra06099b

Phan, C. S., & Morinaka, B. I. (2022). A prevalent group of actinobacterial radical SAM/SPASM maturases involved in triceptide biosynthesis. *American Chemical Society. Chemical Biology, 17*(12), 3284–3289. https://doi.org/10.1021/acschembio.2c00621

Pitt J. J. (2009). Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *The Clinical Biochemist. Reviews, 30*(1), 19–34.

Qiu, J., Zhuo, Y., Zhu, D., Zhou, X., Zhang, L., Bai, L., & Deng, Z. (2011). Overexpression of the ABC transporter AvtAB increases avermectin production in *Streptomyces avermitilis*. *Applied Microbiology and Biotechnology, 92*(2), 337–345. https://doi.org/10.1007/s00253-011-3439-4

Raaijmakers, J. M., De Bruijn, I., Nybroe, O., & Ongena, M. (2010). Natural functions of lipopeptides from *Bacillus* and *Pseudomonas*: more than surfactants and antibiotics. *Federation of European Microbiological Societies. Microbiology Reviews, 34*(6), 1037–1062. https://doi.org/10.1111/j.1574-6976.2010.00221.x

Raffatellu M. (2018). Learning from bacterial competition in the host to develop antimicrobials. *Nature Medicine, 24*(8), 1097–1103. https://doi.org/10.1038/s41591-018-0145-0

Ranieri, M., Chan, D., Yaeger, L. N., Rudolph, M., Karabelas-Pittman, S., Abdo, H., Chee, J., Harvey, H., Nguyen, U., & Burrows, L. L. (2019). Thiostrepton hijacks pyoverdine receptors to inhibit growth of *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy, 63*(9), e00472-19. https://doi.org/10.1128/AAC.00472-19

Ravcheev, D. A., Khoroshkin, M. S., Laikova, O. N., Tsoy, O. V., Sernova, N. V., Petrova, S. A., Rakhmaninova, A. B., Novichkov, P. S., Gelfand, M. S., & Rodionov, D. A. (2014). Comparative genomics and evolution of regulons of the LacI-family transcription factors. *Frontiers in Microbiology*, 5, 294. https://doi.org/10.3389/fmicb.2014.00294

Rea, M. C., Sit, C. S., Clayton, E., O'Connor, P. M., Whittal, R. M., Zheng, J., Vederas, J. C., Ross, R. P., & Hill, C. (2010). Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against *Clostridium difficile*. *Proceedings of the National Academy of Sciences of the United States of America, 107*(20), 9352–9357. https://doi.org/10.1073/pnas.0913554107

Ren, H., Shi, C., Bothwell, I. R., van der Donk, W. A., & Zhao, H. (2020). Discovery and characterization of a class IV lanthipeptide with a nonoverlapping ring pattern. *American Chemical Society. Chemical Biology, 15*(6), 1642–1649. https://doi.org/10.1021/acschembio.0c00267

Ren, H., Wang, B., & Zhao, H. (2017). Breaking the silence: new strategies for discovering novel natural products. *Current Opinion in Biotechnology, 48*, 21–27. https://doi.org/10.1016/j.copbio.2017.02.008

Repka, L. M., Chekan, J. R., Nair, S. K., & van der Donk, W. A. (2017). Mechanistic understanding of lanthipeptide biosynthetic enzymes. *Chemical Reviews, 117*(8), 5457–5520. https://doi.org/10.1021/acs.chemrev.6b00591

Rodríguez-García, A., Combes, P., Pérez-Redondo, R., Smith, M. C., & Smith, M. C. (2005). Natural and synthetic tetracycline-inducible promoters for use in the antibiotic-producing bacteria *Streptomyces*. *Nucleic Acids Research, 33*(9), e87. https://doi.org/10.1093/nar/gni086

Román-Hurtado, F., Sánchez-Hidalgo, M., Martín, J., Ortiz-López, F. J., & Genilloud, O. (2021). Biosynthesis and heterologous expression of cacaoidin, the first member of the lanthidin family of RiPPs. *Antibiotics* (Basel, Switzerland), *10*(4), 403. https://doi.org/10.3390/antibiotics10040403

Russell, A. H., Vior, N. M., Hems, E. S., Lacret, R., & Truman, A. W. (2021). Discovery and characterisation of an amidine-containing ribosomally-synthesised peptide that is widely distributed in nature. *Chemical Science, 12*(35), 11769–11778. https://doi.org/10.1039/d1sc01456k

Salazar, B., Ortiz, A., Keswani, C., Minkina, T., Mandzhieva, S., Pratap Singh, S., Rekadwad, B., Borriss, R., Jain, A., Singh, H. B., & Sansinenea, E. (2022). *Bacillus* spp. as bio-factories for antifungal secondary metabolites: innovation beyond whole organism formulations. *Microbial Ecology*. https://doi.org/10.1007/s00248-022-02044-2

Schwalen, C. J., Hudson, G. A., Kille, B., & Mitchell, D. A. (2018). Bioinformatic expansion and discovery of thiopeptide antibiotics. *Journal of the American Chemical Society, 140*(30), 9494–9501. https://doi.org/10.1021/jacs.8b03896

Servant, P., & Mazodier, P. (2001). Negative regulation of the heat shock response in *Streptomyces*. *Archives of Microbiology, 176*(4), 237–242. https://doi.org/10.1007/s002030100321

Sethi, S., Martens, J., & Bhushan, R. (2021). Assessment and application of Marfey's reagent and analogs in enantioseparation: a decade's perspective. *Biomedical Chromatography: BMC*, *35*(1), e4990. https://doi.org/10.1002/bmc.4990

Shadel, G. S., & Baldwin, T. O. (1991). The *Vibrio fischeri* LuxR protein is capable of bidirectional stimulation of transcription and both positive and negative regulation of the *luxR* gene. *Journal of Bacteriology, 173*(2), 568–574. https://doi.org/10.1128/jb.173.2.568-574.1991

Sharan, S. K., Thomason, L. C., Kuznetsov, S. G., & Court, D. L. (2009). Recombineering: a homologous recombination-based method of genetic engineering. *Nature Protocols, 4*(2), 206–223. https://doi.org/10.1038/nprot.2008.227

Sharma, R., Jamwal, V., Singh, V. P., Wazir, P., Awasthi, P., Singh, D., Vishwakarma, R. A., Gandhi, S. G., & Chaubey, A. (2017). Revelation and cloning of valinomycin synthetase genes in *Streptomyces lavendulae* ACR-DA1 and their expression analysis under different fermentation and elicitation conditions. *Journal of Biotechnology*, 253, 40–47. https://doi.org/10.1016/j.jbiotec.2017.05.008

Shen, Y. Q., Bonnot, F., Imsand, E. M., RoseFigura, J. M., Sjölander, K., & Klinman, J. P. (2012). Distribution and properties of the genes encoding the biosynthesis of the bacterial cofactor, pyrroloquinoline quinone. *Biochemistry, 51*(11), 2265–2275. https://doi.org/10.1021/bi201763d

Shenkarev, Z. O., Finkina, E. I., Nurmukhamedova, E. K., Balandin, S. V., Mineev, K. S., Nadezhdin, K. D., Yakimenko, Z. A., Tagaev, A. A., Temirov, Y. V., Arseniev, A. S., & Ovchinnikova, T. V. (2010). Isolation, structure elucidation, and synergistic antibacterial activity of a novel two-component lantibiotic lichenicidin from *Bacillus licheniformis* VK21. *Biochemistry, 49*(30), 6462–6472. https://doi.org/10.1021/bi100871b

Shepherd, M. D., Kharel, M. K., Bosserman, M. A., & Rohr, J. (2010). Laboratory maintenance of *Streptomyces* species. *Current Protocols in Microbiology*, Chapter 10, Unit–10E.1. https://doi.org/10.1002/9780471729259.mc10e01s18

Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M., & Panda, A. K. (2015). Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. *Microbial Cell Factories, 14*, 41. https://doi.org/10.1186/s12934-015-0222-8

Song, Z. Q., Liao, Z. J., Hu, Y. F., Ma, Z., Bechthold, A., & Yu, X. P. (2019). Development and optimization of an intergeneric conjugation system and analysis of promoter activity in *Streptomyces rimosus* M527. *Journal of Zhejiang University. Science. B, 20*(11), 891–900. https://doi.org/10.1631/jzus.B1900270

Szu, P. H., Ruszczycky, M. W., Choi, S. H., Yan, F., & Liu, H. W. (2009). Characterization and mechanistic studies of DesII: a radical *S*-adenosyl-*L*-methionine enzyme involved in the biosynthesis of TDP-D-desosamine. *Journal of the American Chemical Society*, *131*(39), 14030–14042. https://doi.org/10.1021/ja903354k

Taberman, H., Parkkinen, T., & Rouvinen, J. (2016). Structural and functional features of the NAD(P) dependent Gfo/Idh/MocA protein family oxidoreductases. *Protein Science: a Publication of the Protein Society, 25*(4), 778–786. https://doi.org/10.1002/pro.2877

Tan, S., Moore, G., & Nodwell, J. (2019). Put a bow on it: knotted antibiotics take center stage. *Antibiotics* (Basel, Switzerland), *8*(3), 117. https://doi.org/10.3390/antibiotics8030117

Tanaka, A., Takano, Y., Ohnishi, Y., & Horinouchi, S. (2007). AfsR recruits RNA polymerase to the *afsS* promoter: a model for transcriptional activation by SARPs. *Journal of Molecular Biology, 369*(2), 322–333. https://doi.org/10.1016/j.jmb.2007.02.096

Tanaka, Y., Hosaka, T., & Ochi, K. (2010). Rare earth elements activate the secondary metabolite-biosynthetic gene clusters in *Streptomyces coelicolor* A3(2). *The Journal of Antibiotics, 63*(8), 477–481. https://doi.org/10.1038/ja.2010.53

Tang, W., Bobeica, S. C., Wang, L., & van der Donk, W. A. (2019). CylA is a sequence-specific protease involved in toxin biosynthesis. *Journal of Industrial Microbiology & Biotechnology, 46*(3-4), 537–549. https://doi.org/10.1007/s10295-018-2110-9

Tang, W., Dong, S. H., Repka, L. M., He, C., Nair, S. K., & van der Donk, W. A. (2015). Applications of the class II lanthipeptide protease LicP for sequence-specific, traceless peptide bond cleavage. *Chemical Science, 6*(11), 6270–6279. https://doi.org/10.1039/c5sc02329g

Terlouw, B. R., Blin, K., Navarro-Muñoz, J. C., Avalon, N. E., Chevrette, M. G., Egbert, S., Lee, S., Meijer, D., Recchia, M. J. J., Reitz, Z. L., van Santen, J. A., Selem-Mojica, N., Tørring, T., Zaroubi, L., Alanjary, M., Aleti, G., Aguilar, C., Al-Salihi, S. A. A., Augustijn, H. E., Avelar-Rivas, J. A., … Medema, M. H. (2023). MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Research, 51*(D1), D603–D610. https://doi.org/10.1093/nar/gkac1049

Thoendel, M., & Horswill, A. R. (2010). Biosynthesis of peptide signals in Gram-positive bacteria. *Advances in Applied Microbiology*, 71, 91–112. https://doi.org/10.1016/S0065-2164(10)71004-2

Tietz, J. I., Schwalen, C. J., Patel, P. S., Maxson, T., Blair, P. M., Tai, H. C., Zakai, U. I., & Mitchell, D. A. (2017). A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nature Chemical Biology, 13*(5), 470–478. https://doi.org/10.1038/nchembio.2319

Tomm, H. A., Ucciferri, L., & Ross, A. C. (2019). Advances in microbial culturing conditions to activate silent biosynthetic gene clusters for novel metabolite production. *Journal of Industrial Microbiology & Biotechnology, 46*(9-10), 1381–1400. https://doi.org/10.1007/s10295-019-02198-y

Ueda, K., Oinuma, K., Ikeda, G., Hosono, K., Ohnishi, Y., Horinouchi, S., & Beppu, T. (2002). AmfS, an extracellular peptidic morphogen in *Streptomyces griseus*. *Journal of Bacteriology, 184*(5), 1488–1492. https://doi.org/10.1128/JB.184.5.1488-1492.2002

UniProt Consortium (2023). UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Research, 51*(D1), D523–D531. https://doi.org/10.1093/nar/gkac1052

Urem, M., van Rossum, T., Bucca, G., Moolenaar, G. F., Laing, E., Świątek-Połatyńska, M. A., Willemse, J., Tenconi, E., Rigali, S., Goosen, N., Smith, C. P., & van Wezel, G. P. (2016). OsdR of *Streptomyces coelicolor* and the dormancy regulator DevR of *Mycobacterium tuberculosis* Control Overlapping Regulons. *mSystems, 1*(3), e00014-16. https://doi.org/10.1128/mSystems.00014-16

van der Heul, H. U., Bilyk, B. L., McDowall, K. J., Seipke, R. F., & van Wezel, G. P. (2018). Regulation of antibiotic production in Actinobacteria: new perspectives from the post-genomic era. *Natural Product Reports, 35*(6), 575–604. https://doi.org/10.1039/c8np00012c

van Staden, A., van Zyl, W. F., Trindade, M., Dicks, L., & Smith, C. (2021). Therapeutic application of lantibiotics and other lanthipeptides: old and new findings. *Applied and Environmental Microbiology, 87*(14), e0018621. https://doi.org/10.1128/AEM.00186-21

van Wezel, G. P., & McDowall, K. J. (2011). The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances. *Natural Product Reports, 28*(7), 1311–1333. https://doi.org/10.1039/c1np00003a

Vinogradov, A. A., & Suga, H. (2020). Introduction to thiopeptides: biological activity, biosynthesis, and strategies for functional reprogramming. *Cell Chemical Biology, 27*(8), 1032–1051. https://doi.org/10.1016/j.chembiol.2020.07.003

Völler, G. H., Krawczyk, B., Ensle, P., & Süssmuth, R. D. (2013). Involvement and unusual substrate specificity of a prolyl oligopeptidase in class III lanthipeptide maturation. *Journal of the American Chemical Society, 135*(20), 7426–7429. https://doi.org/10.1021/ja402296m

Völler, G. H., Krawczyk, J. M., Pesic, A., Krawczyk, B., Nachtigall, J., & Süssmuth, R. D. (2012). Characterization of new class III lantibiotics--erythreapeptin, avermipeptin and griseopeptin from *Saccharopolyspora erythraea*, *Streptomyces avermitilis* and *Streptomyces griseus* demonstrates stepwise *N*-terminal leader processing. *Chembiochem: a European Journal of Chemical Biology, 13*(8), 1174–1183. https://doi.org/10.1002/cbic.201200118

Wagenaar M. M. (2008). Pre-fractionated microbial samples--the second-generation natural products library at Wyeth. *Molecules* (Basel, Switzerland), *13*(6), 1406–1426. https://doi.org/10.3390/molecules13061406

Walker, M. C., Eslami, S. M., Hetrick, K. J., Ackenhusen, S. E., Mitchell, D. A., & van der Donk, W. A. (2020). Precursor peptide-targeted mining of more than one hundred thousand genomes expands the lanthipeptide natural product family. *BMC Genomics, 21*(1), 387. https://doi.org/10.1186/s12864-020-06785-7

Wang, H., & van der Donk, W. A. (2012). Biosynthesis of the class III lantipeptide catenulipeptin. *American Chemical Society. Chemical Biology, 7*(9), 1529–1535. https://doi.org/10.1021/cb3002446

Wang, J., Zhang, L., Teng, K., Sun, S., Sun, Z., & Zhong, J. (2014). Cerecidins, novel lantibiotics from *Bacillus cereus* with potent antimicrobial activity. *Applied and Environmental Microbiology, 80*(8), 2633–2643. https://doi.org/10.1128/AEM.03751-13

Wang, K., Liu, N., Shang, F., Huang, J., Yan, B., Liu, M., & Huang, Y. (2021). Activation of secondary metabolism in red soil-derived *Streptomycetes* via co-culture with mycolic acid-containing bacteria. *Microorganisms, 9*(11), 2187. https://doi.org/10.3390/microorganisms9112187

Wang, L., Tian, X., Wang, J., Yang, H., Fan, K., Xu, G., Yang, K., & Tan, H. (2009). Autoregulation of antibiotic biosynthesis by binding of the product to an atypical response regulator. *Proceedings of the National Academy of Sciences of the United States of America, 106*(21), 8617–8622. https://doi.org/10.1073/pnas.0900592106

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D., Watrous, J., Kapono, C. A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M. J., Liu, W. T., Crüsemann, M., Boudreau, P. D., Esquenazi, E., Sandoval-Calderón, M., Kersten, R. D., … Bandeira, N. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology, 34*(8), 828–837. https://doi.org/10.1038/nbt.3597

Wang, W., Zheng, G., & Lu, Y. (2021). Recent advances in strategies for the cloning of natural product biosynthetic gene clusters. *Frontiers in Bioengineering and Biotechnology, 9*, 692797. https://doi.org/10.3389/fbioe.2021.692797

Wecksler, S. R., Stoll, S., Tran, H., Magnusson, O. T., Wu, S. P., King, D., Britt, R. D., & Klinman, J. P. (2009). Pyrroloquinoline quinone biogenesis: demonstration that PqqE from *Klebsiella pneumoniae* is a radical *S*-adenosyl-*L*-methionine enzyme. *Biochemistry, 48*(42), 10151–10161. https://doi.org/10.1021/bi900918b

Werren J. H. (2011). Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 10863–10870. https://doi.org/10.1073/pnas.1102343108

Wiebach, V., Mainz, A., Schnegotzki, R., Siegert, M. J., Hügelland, M., Pliszka, N., & Süssmuth, R. D. (2020). An amphipathic alpha-helix guides maturation of the ribosomally-synthesized lipolanthines. *Angewandte Chemie* (International ed. in English), *59*(38), 16777–16785. https://doi.org/10.1002/anie.202003804

Wong, K. S., & Houry, W. A. (2012). Novel structural and functional insights into the MoxR family of AAA+ ATPases. *Journal of Structural Biology, 179*(2), 211–221. https://doi.org/10.1016/j.jsb.2012.03.010

Xia, H., Li, X., Li, Z., Zhan, X., Mao, X., & Li, Y. (2020). The application of regulatory cascades in *Streptomyces*: yield enhancement and metabolite mining. *Frontiers in Microbiology, 11*, 406. https://doi.org/10.3389/fmicb.2020.00406

Xu, F., Nazari, B., Moon, K., Bushin, L. B., & Seyedsayamdost, M. R. (2017). Discovery of a cryptic antifungal compound from *Streptomyces albus* J1074 using high-throughput elicitor screens. *Journal of the American Chemical Society, 139*(27), 9203–9212. https://doi.org/10.1021/jacs.7b02716

Xu, F., Wu, Y., Zhang, C., Davis, K. M., Moon, K., Bushin, L. B., & Seyedsayamdost, M. R. (2019). A genetics-free method for high-throughput discovery of cryptic microbial metabolites. *Nature Chemical Biology, 15*(2), 161–168. https://doi.org/10.1038/s41589-018-0193-2

Xu, M., Zhang, F., Cheng, Z., Bashiri, G., Wang, J., Hong, J., Wang, Y., Xu, L., Chen, X., Huang, S. X., Lin, S., Deng, Z., & Tao, M. (2020). Functional genome mining reveals a class V lanthipeptide containing a *d*-amino acid introduced by an F420 H2 -dependent reductase. *Angewandte Chemie* (International ed. in English), *59*(41), 18029–18035. https://doi.org/10.1002/anie.202008035

Xu, Z., Ji, L., Tang, W., Guo, L., Gao, C., Chen, X., Liu, J., Hu, G., & Liu, L. (2022). Metabolic engineering of *Streptomyces* to enhance the synthesis of valuable natural products. *Engineering Microbiology, 2*(2). https://doi.org/10.1016/j.engmic.2022.100022

Xue, D., Older, E. A., Zhong, Z., Shang, Z., Chen, N., Dittenhauser, N., Hou, L., Cai, P., Walla, M. D., Dong, S. H., Tang, X., Chen, H., Nagarkatti, P., Nagarkatti, M., Li, Y. X., & Li, J. (2022). Correlational networking guides the discovery of unclustered lanthipeptide protease-encoding genes. *Nature Communications, 13*(1), 1647. https://doi.org/10.1038/s41467-022-29325-1

Yagüe, P., Rodríguez-García, A., López-García, M. T., Rioseras, B., Martín, J. F., Sánchez, J., & Manteca, A. (2014). Transcriptomic analysis of liquid non-sporulating *Streptomyces coelicolor* cultures demonstrates the existence of a complex differentiation comparable to that occurring in solid sporulating cultures. *PloS One, 9*(1), e86296. https://doi.org/10.1371/journal.pone.0086296

Yagüe, P., Willemse, J., Koning, R. I., Rioseras, B., López-García, M. T., Gonzalez-Quiñonez, N., Lopez-Iglesias, C., Shliaha, P. V., Rogowska-Wrzesinska, A., Koster, A. J., Jensen, O. N., van Wezel, G. P., & Manteca, Á. (2016). Subcompartmentalization by cross-membranes during early growth of *Streptomyces* hyphae. *Nature Communications*, 7, 12467. https://doi.org/10.1038/ncomms12467

Yang, X., & van der Donk, W. A. (2015). Post-translational introduction of D-alanine into ribosomally synthesized peptides by the dehydroalanine reductase NpnJ. *Journal of the American Chemical Society, 137*(39), 12426–12429. https://doi.org/10.1021/jacs.5b05207

Yu, Y., Duan, L., Zhang, Q., Liao, R., Ding, Y., Pan, H., Wendt-Pienkowski, E., Tang, G., Shen, B., & Liu, W. (2009). Nosiheptide biosynthesis featuring a unique indole side ring formation on the characteristic thiopeptide framework. *American Chemical Society. Chemical Biology, 4*(10), 855–864. https://doi.org/10.1021/cb900133x

Zaburannyi, N., Rabyk, M., Ostash, B., Fedorenko, V., & Luzhetskyy, A. (2014). Insights into naturally minimized *Streptomyces albus* J1074 genome. *BMC Genomics, 15*, 97. https://doi.org/10.1186/1471-2164-15-97

Zallot, R., Oberg, N., & Gerlt, J. A. (2019). The EFI web resource for genomic enzymology tools: leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry, 58*(41), 4169–4182. https://doi.org/10.1021/acs.biochem.9b00735

Zhang, P., Wu, H., Chen, X. L., Deng, Z., Bai, L., & Pang, X. (2014). Regulation of the biosynthesis of thiopeptide antibiotic cyclothiazomycin by the transcriptional regulator SHJG8833 in *Streptomyces hygroscopicus* 5008. *Microbiology* (Reading, England), *160*(Pt 7), 1379–1392. https://doi.org/10.1099/mic.0.076901-0

Zhang, Q., Doroghazi, J. R., Zhao, X., Walker, M. C., & van der Donk, W. A. (2015). Expanded natural product diversity revealed by analysis of lanthipeptide-like gene clusters in actinobacteria. *Applied and Environmental Microbiology, 81*(13), 4339–4350. https://doi.org/10.1128/AEM.00635-15

Zhang, Q., Li, Y., Chen, D., Yu, Y., Duan, L., Shen, B., & Liu, W. (2011). Radical-mediated enzymatic carbon chain fragmentation-recombination. *Nature Chemical Biology, 7*(3), 154–160. https://doi.org/10.1038/nchembio.512

Zhang, Q., Ortega, M., Shi, Y., Wang, H., Melby, J. O., Tang, W., Mitchell, D. A., & van der Donk, W. A. (2014). Structural investigation of ribosomally synthesized natural products by hypothetical structure enumeration and evaluation using tandem MS. *Proceedings of the National Academy of Sciences of the United States of America, 111*(33), 12031–12036. https://doi.org/10.1073/pnas.1406418111

Zhang, Q., Yang, X., Wang, H., & van der Donk, W. A. (2014). High divergence of the precursor peptides in combinatorial lanthipeptide biosynthesis. *American Chemical Society. Chemical Biology, 9*(11), 2686–2694. https://doi.org/10.1021/cb500622c

Zhang, Q., Yu, Y., Vélasquez, J. E., & van der Donk, W. A. (2012). Evolution of lanthipeptide synthetases. *Proceedings of the National Academy of Sciences of the United States of America, 109*(45), 18361–18366. https://doi.org/10.1073/pnas.1210393109

Zhang, S., Chen, Y., Zhu, J., Lu, Q., Cryle, M. J., Zhang, Y., & Yan, F. (2022). Structural diversity, biosynthesis, and biological functions of lipopeptides from *Streptomyces*. *Natural Product Reports*, 10.1039/d2np00044j. Advance online publication. https://doi.org/10.1039/d2np00044j

Zhang, Y., Hong, Z., Zhou, L., Zhang, Z., Tang, T., Guo, E., Zheng, J., Wang, C., Dai, L., Si, T., & Wang, H. (2022). Biosynthesis of gut-microbiota-derived lantibiotics reveals a subgroup of S8 family proteases for class III leader removal. *Angewandte Chemie* (International ed. in English), *61*(6), e202114414. https://doi.org/10.1002/anie.202114414

Zhao, C., Sheng, W., Wang, Y., Zheng, J., Xie, X., Liang, Y., Wei, W., Bao, R., & Wang, H. (2022). Conformational remodeling enhances activity of lanthipeptide zinc-metallopeptidases. *Nature Chemical Biology, 18*(7), 724–732. https://doi.org/10.1038/s41589-022-01018-2

Zhao, X., & Kuipers, O. P. (2021). Nisin- and ripcin-derived hybrid lanthipeptides display selective antimicrobial activity against *Staphylococcus aureus*. *American Chemical Society. Synthetic Biology, 10*(7), 1703–1714. https://doi.org/10.1021/acssynbio.1c00080

Zhao, X., & van der Donk, W. A. (2016). Structural characterization and bioactivity analysis of the two-component lantibiotic Flv system from a ruminant bacterium. *Cell Chemical Biology, 23*(2), 246–256. https://doi.org/10.1016/j.chembiol.2015.11.014

Zhou, Q., Ning, S., & Luo, Y. (2020). Coordinated regulation for nature products discovery and overproduction in *Streptomyces*. *Synthetic and Systems Biotechnology, 5*(2), 49–58. https://doi.org/10.1016/j.synbio.2020.04.002

Zhu, D., Wang, Y., Zhang, M., Ikeda, H., Deng, Z., & Cane, D. E. (2013). Product-mediated regulation of pentalenolactone biosynthesis in *Streptomyces* species by the MarR/SlyA family activators PenR and PntR. *Journal of Bacteriology, 195*(6), 1255–1266. https://doi.org/10.1128/JB.02079-12

Zong, G., Fu, J., Zhang, P., Zhang, W., Xu, Y., Cao, G., & Zhang, R. (2022). Use of elicitors to enhance or activate the antibiotic production in *Streptomyces. Critical Reviews in Biotechnology, 42*(8), 1260–1283. https://doi.org/10.1080/07388551.2021.1987856

8. Abbreviations and acronyms

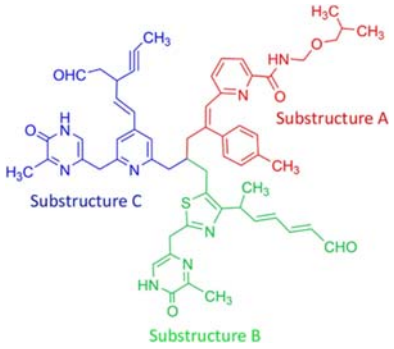
AAA	ATPase domain
ABC	ATP-binding cassette
AdpA	A-factor-dependent protein
A-factor	2-isocapryloyl-3r-hydroxymethyl-γ-butyrolactone
AIP	auto inducing peptide
Ama	aminomalonate
AntiSMASH	antibiotics and secondary metabolite analysis shell
ARTS	antibiotic resistant target seeker
ATP	adenosine triphosphate
BAC	bacterial artificial chromosome
BGC	biosynthetic gene cluster
BPC	base peak chromatogram
BTAD	bacterial transcriptional activator domain
cAMP	cyclic adenosine monophosphate
CAPTURE	Cas12a-assisted precise targeted cloning using in vivo Cre-*lox* recombination
CASO	casein-peptone soymeal-peptone
CDA	calcium-dependent antibiotic
CP	core peptide
Cpr	conprimycin
CSR	cluster-situated regulator
CYP 450	cytochrome P450
Dha	dehydroalanine
Dhb	dehydrobutyrine
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EFI	enzyme function initiative
EIC	extracted ion chromatogram
ESI	enzyme similarity tool
FAD	flavin adenine dinucleotide
FDAA	1-fluoro-2-4-dinitrophenyl-5-L-alanine amide
Fgly	formylglycine
FMN	flavin mononucleotide
GBL	γ-butyrolactones
GCF	gene cluster family

GMP	guanosine monophosphate
GNN	genome neighborhood network
GNPS	global natural product social molecular networking
GNT	genome neighborhood tool
Gri	griseopeptin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HiTES	high-throughput elicitor screening
HPLC	high-performance liquid chromatography
HTH	helix-turn-helix
IAA	iodoacetamine
IPTG	isopropyl-β-D-thiogalactopyranosid
ISP2	international *Streptomyces* project-2 medium
Lan	lanthipeptide
LAP	linear azole containing peptide
LLHR	linear-linear homologous recombination
LP	leader peptide
MALDI	matrix assisted laser desorption/ionization
MFS	major facilitator superfamily
MIBiG	minimum information about a biosynthetic gene cluster
Mrs	mixed RiPP system
MS	mass spectrometry
MS2 or MS/MS	tandem mass spectrometry
NAD	nicotinamide adenine dinucleotide
NADP	nicotinamide adenine dinucleotide phosphate
NGD	genome neighborhood diagram
NP	natural product
NRPS	non-ribosomal peptide synthetase
NTP	nucleoside triphosphate
OD	optical density
OE	overexpression
ORF	open reading frame
OSMAC	one strain many compounds
PCR	polymerase chain reaction
PKS	polyketide synthase
PP	precursor peptide
PQQ	pyrroloquinoline quinone

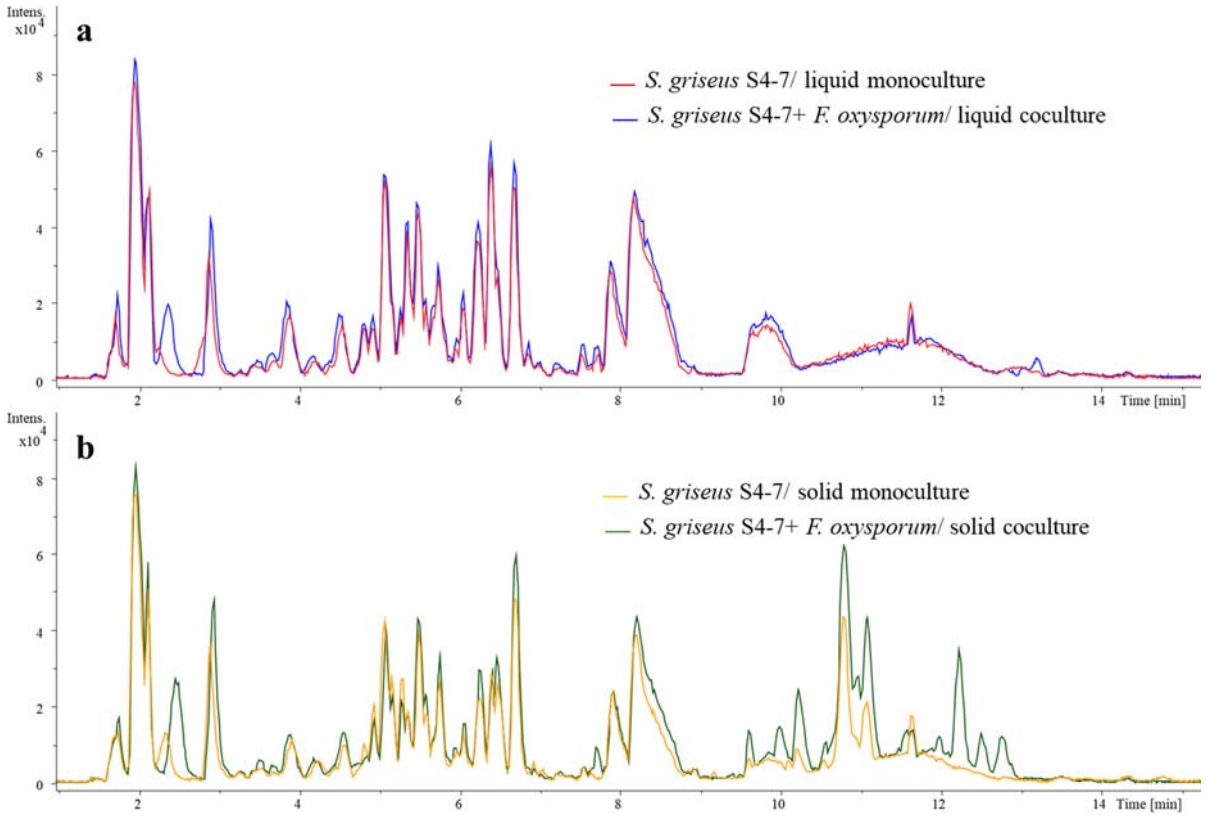
PTM	post-translational modification
QS	quorum sensing
RBS	ribosome binding site
RE	restriction enzyme
RiPP	ribosomally synthesized and post-translationally modified peptides
RNA	ribonucleic acid
RRE	RiPP recognition element
rSAM	radical *S*-adenosyl methionine
S4-7	*Streptomyces griseus* S4-7
SAM	*S*-adenosyl methionine
SAP	shrimp alkaline phosphatase
SARP	*Streptomyces* antibiotic regulatory protein
SM13	screening medium
SPASM	subtilosin A, pyrroloquinoline quinone, anaerobic sulfatase, and mycofactocin
SSN	sequence similarity network
SUF	multiprotein machinery for Fe–S cluster biogenesis
T1PKS	type I polyketide synthase
Ta	annealing temperature
TAR	transformation-associated recombination
TCS	two-component system
TIR	toll-interleukin receptor
Tm	melting temperature
TMS	transmembrane segment
TPR	tetratricopeptide repeat
Trans-reg	transcriptional regulator protein (*C*-terminal)
TRCF	transcription-repair coupling factor
TR-PCR	reverse transcription polymerase chain reaction
TSB	tryptic soy broth
TSS	transcription start site
vWA	von Willebrand factor A
Wt	wild type
YEME	yeast extract-malt extract
YHR	yeast homologous recombination

9. Appendix

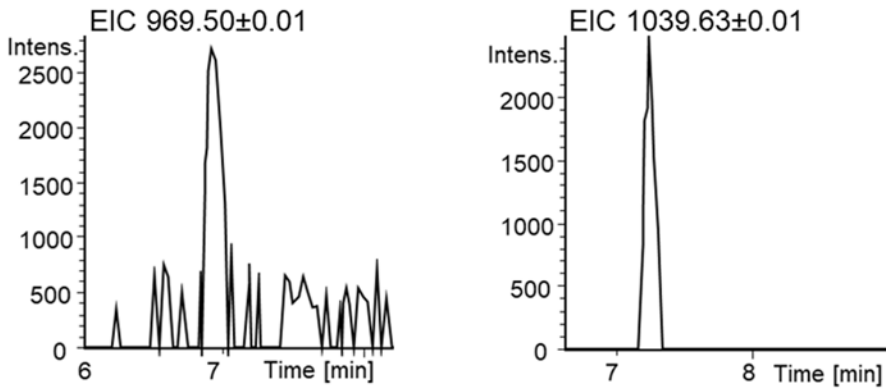
9.1 Supplementary Figures



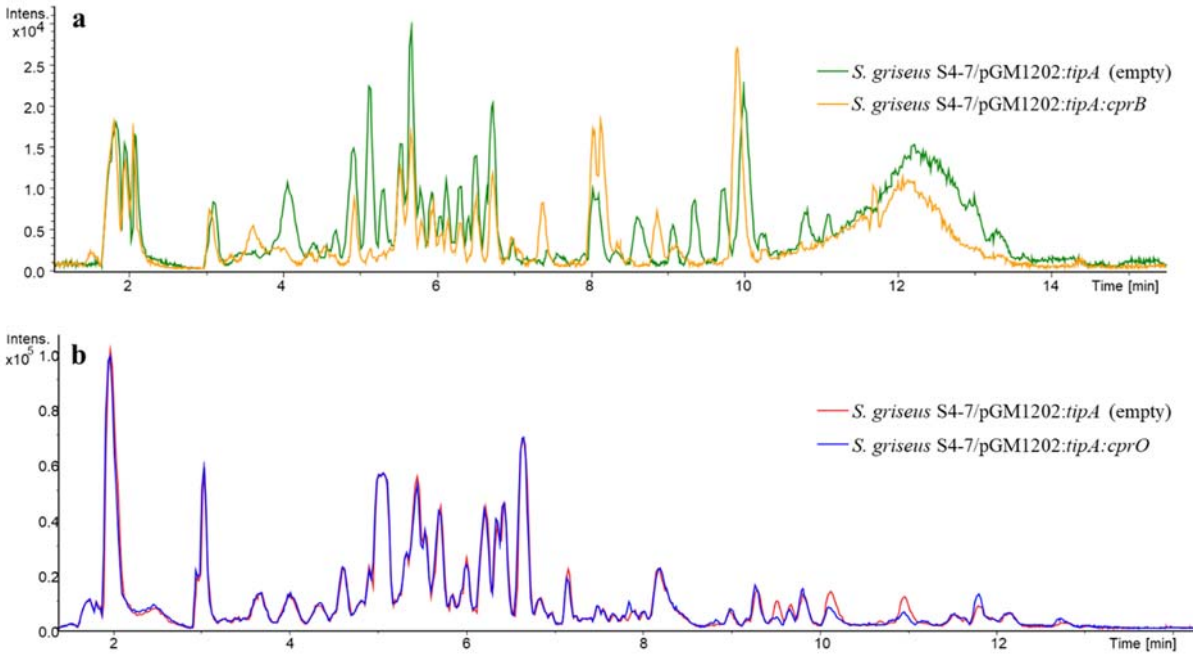
Supplementary Figure S1. Structure of Picolinamycin. Taken from Maiti *et al.*, 2020.



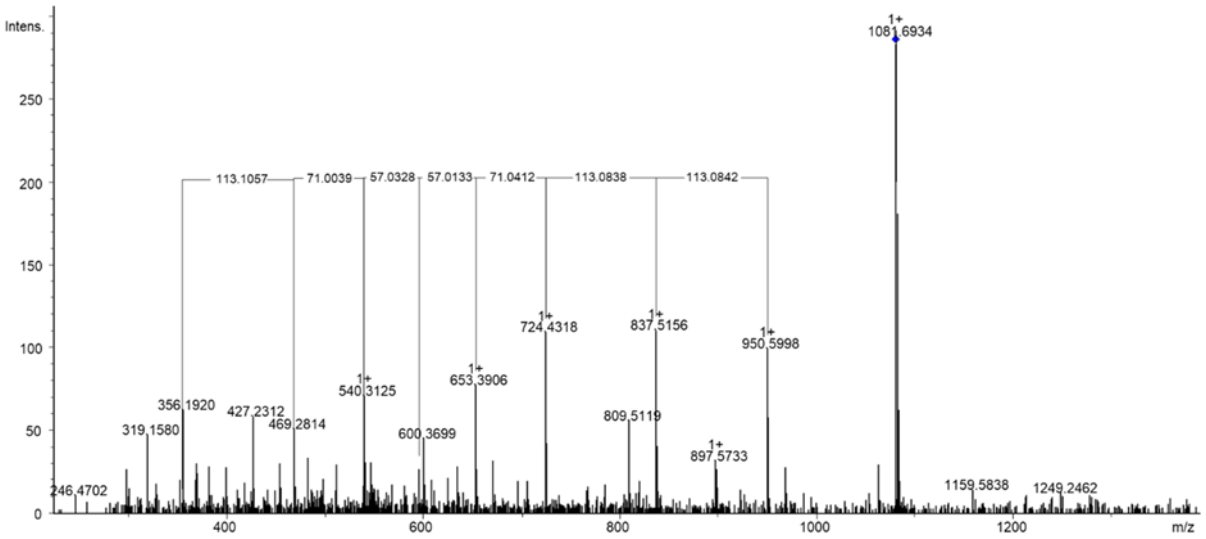
Supplementary Figure S2. Chromatographic profile of *S. griseus* S4-7 grown in liquid PDK medium **(a)** or in PDK agar **(b)** and effect of cocultivation with *Fusarium oxysporum*.



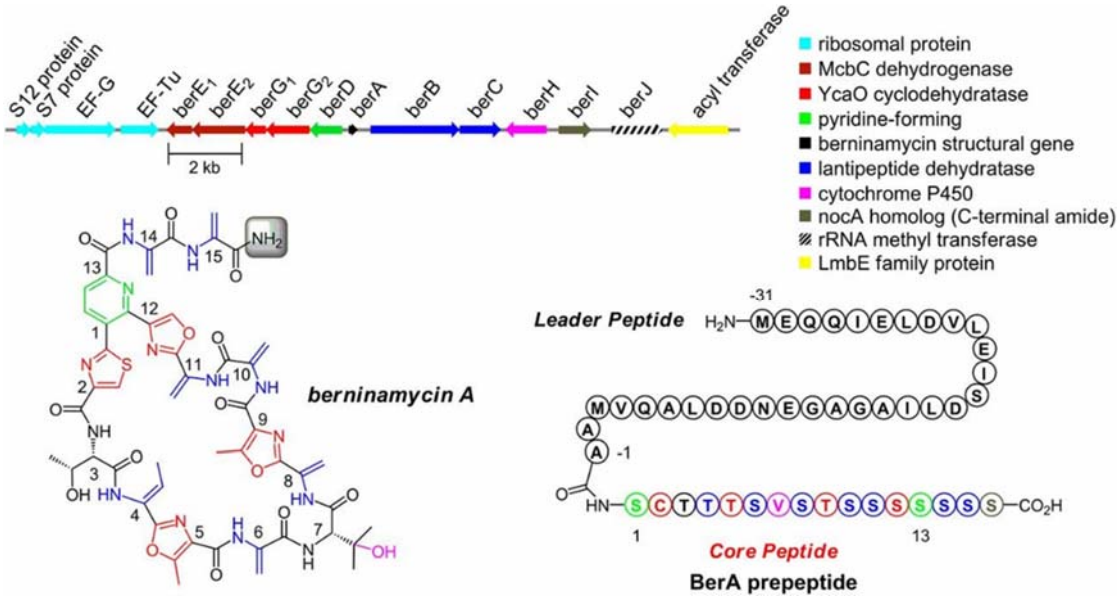
Supplementary Figure S3. Extracted ion chromatograms of two $[M+H]^{1+}$ ions at *m/z* 969.50 and *m/z* 1039.63 obtained from the overexpression of the PapR2 regulator in *S. griseus* S4-7 cultivated in NL800 medium.



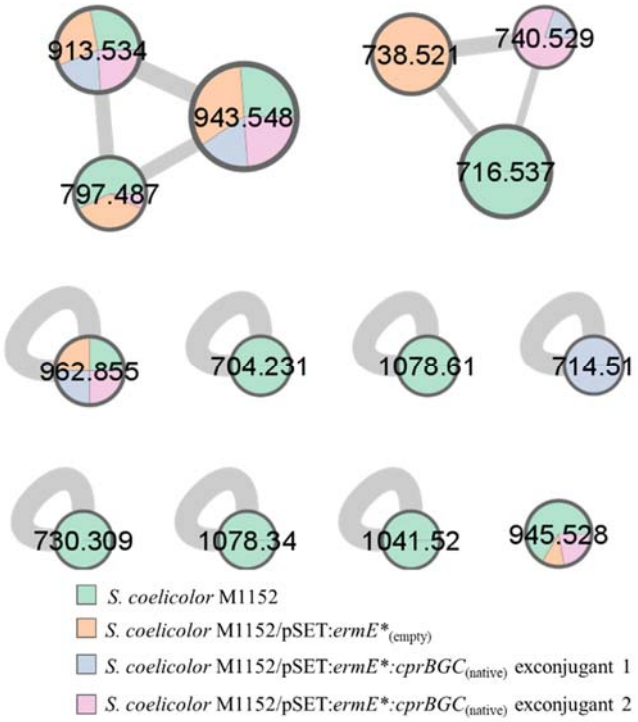
Supplementary Figure S4. Overexpression of CSRs from the *cpr* BGC in *S. griseus* S4-7. **(a)** Total ion chromatograms showed no differences between the control strain carrying the empty pGM1202:*tipA* vector and the strain overexpressing the CprB SARP regulator. **(b)** Total ion chromatograms from the overexpression of *cprO*. The MS/MS spectra of those ions present only in the overexpression sample were manually inspected, but none of them could be connected to the cpr thiopeptide.



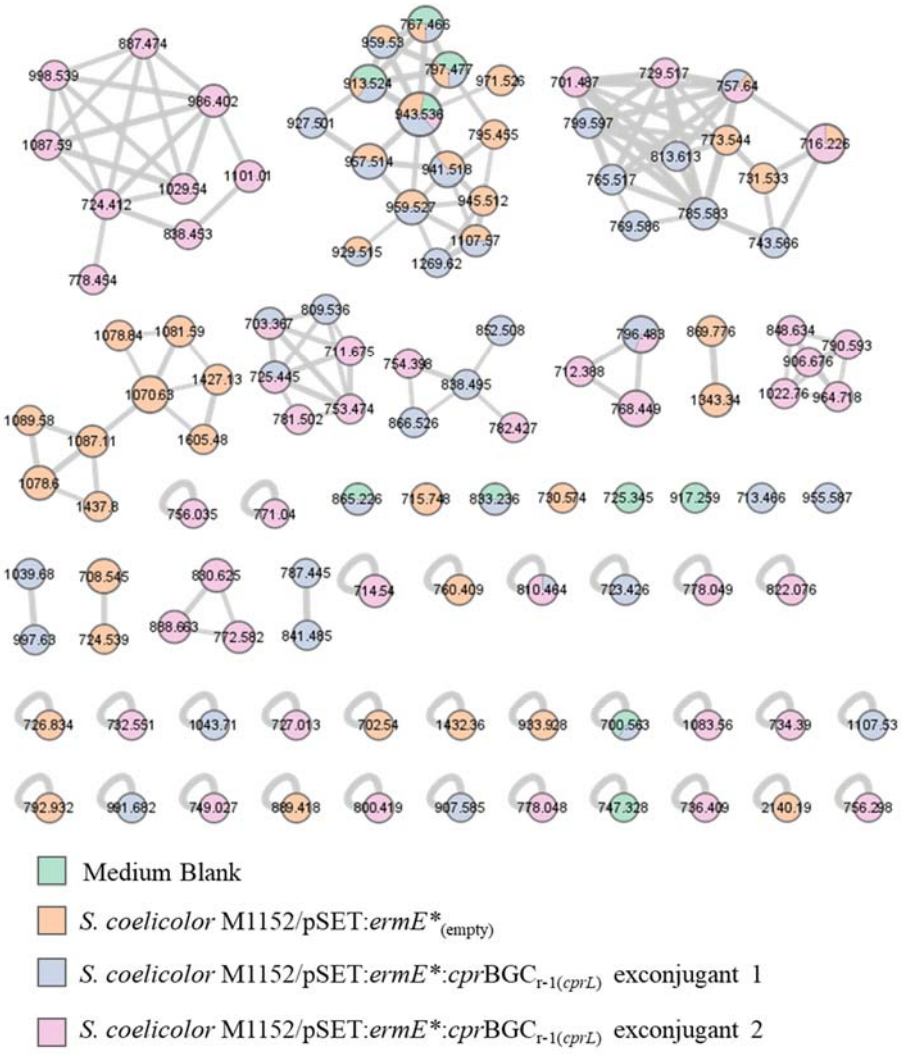
Supplementary Figure S5. High resolution MS/MS spectrum corresponding of the $[M+H]^+$ ion at *m/z* 1081.69 observed in the overexpression of *mrs*675 in *S. griseus* S4-7. The corresponding peptide was found in lower abundance in the overexpression sample compared to the negative control.



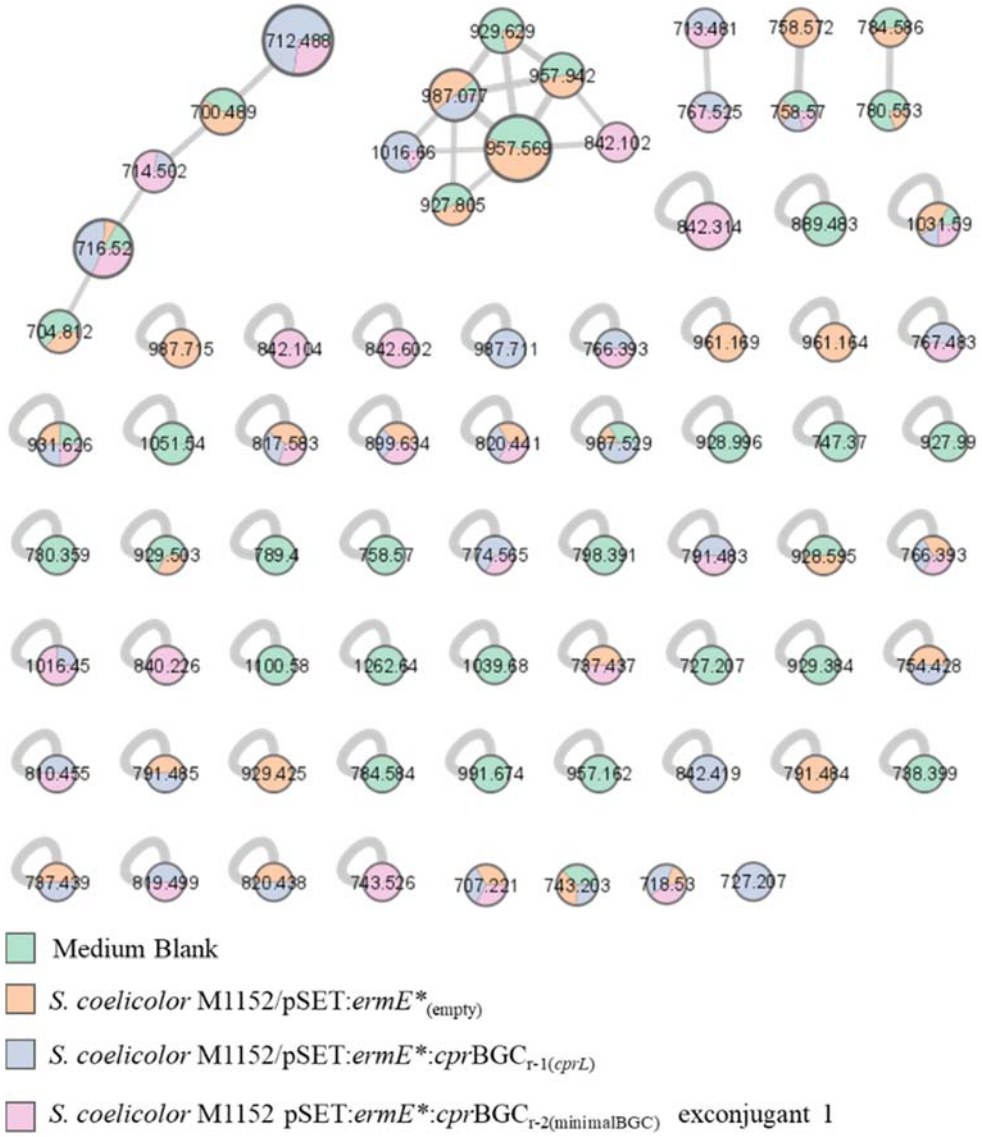
Supplementary Figure S6. Berninamycin A biosynthetic gene cluster and structure. Berninamycin A contains three oxazoles whose formation is catalyzed by two independent YcaO cyclodehydratases BerG1 and BerG2. (Taken from Malcolmson *et al.*, 2013).



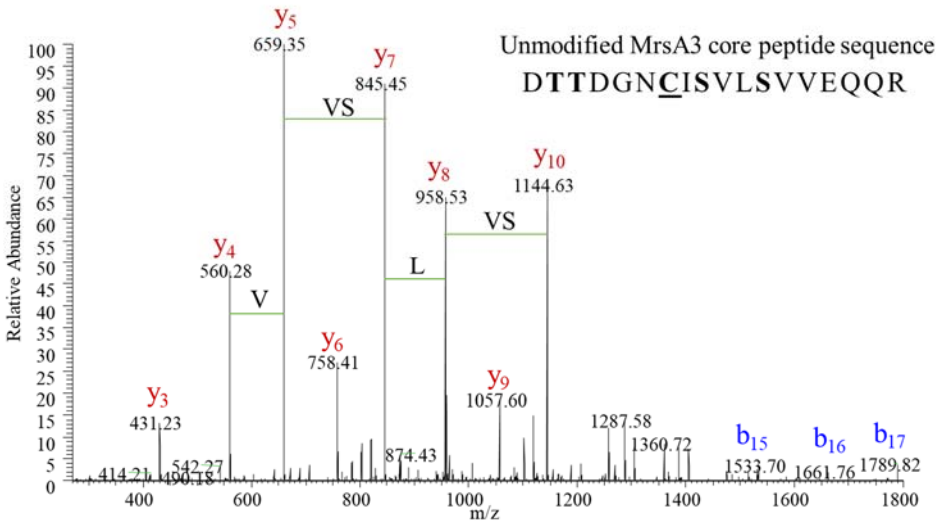
Supplementary Figure S7. Heterologous expression of the native *cpr* BGC in *Streptomyces coelicolor* M1152. The molecular network was generated using GNPS according to the parameters described in section 6.2.8.1. Nodes represent the respective *m/z* values, while bridges indicate their possible correlations (edge width displays the cosine score). Node size corresponds to the number of spectra. Nodes were filtered and masses out of the target range (700-2500 Da) were removed to facilitate visualization of potential cpr candidates. MS2 spectra corresponding to those masses present only in the heterologous expression experiments were manually verified.



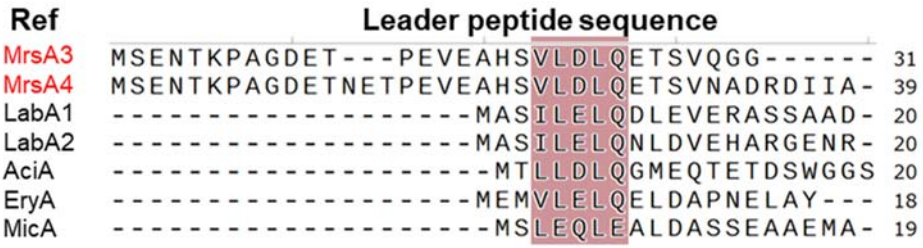
Supplementary Figure S8. Heterologous expression of the refactored *cpr* BGC (r-1) in *Streptomyces coelicolor* M1152. The molecular network was generated using GNPS according to the parameters described in section 6.2.8.1. Nodes represent the respective *m/z* values, while bridges indicate their possible correlations (edge width displays the cosine score). Node size corresponds to the number of spectra. Nodes corresponding to masses out of the target range (700-2500 Da) were removed to facilitate visualization of potential cpr candidates. MS2 spectra corresponding to those masses present only in the heterologous expression experiments were manually verified.



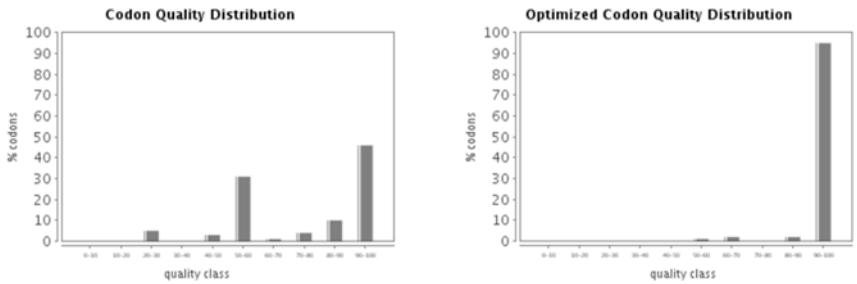
Supplementary Figure S9. Heterologous expression of the refactored minimal *cpr* BGC (r-2) in *Streptomyces coelicolor* M1152. The molecular network was generated using GNPS according to the parameters described in section 6.2.8.1. Data obtained from heterologous expression of the *cpr* BGC r-1 was also included. Nodes represent the respective *m/z* values, while bridges indicate their possible correlations (edge width displays the cosine score). Node size corresponds to the number of spectra. Nodes corresponding to masses out of the target range (700-2500 Da) were removed to facilitate visualization of potential cpr candidates MS2 spectra corresponding to those masses present only in the heterologous expression experiments were manually verified.



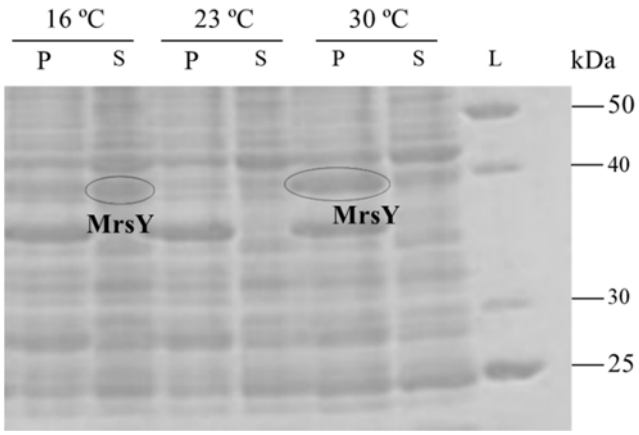
Supplementary Figure S10. High resolution MS/MS spectrum obtained from the MrsA3 core peptide. After co-expression of MrsA3 and MrsM, the peptide was treated with Factor Xa protease for leader removal. Fragments showed the presence of intact residues and the observed monoisotopic mass corresponded to that calculated for the linear core peptide. Monoisotopic mass calculated: 1962.93 Da.



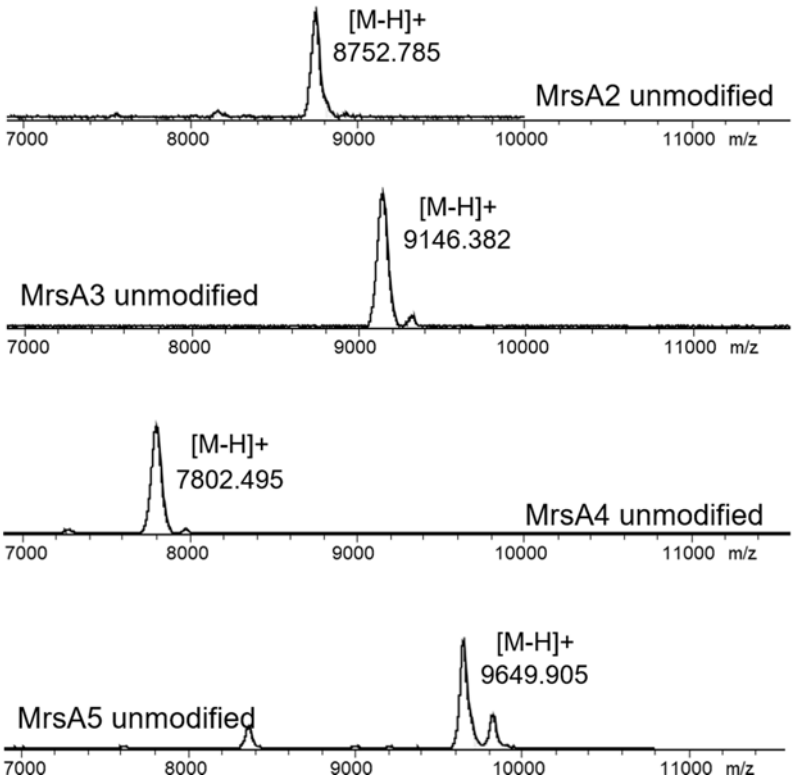
Supplementary Figure S11. Alignment of MrsA3-A4 leader peptides with selected class III lanthipeptides. MrsA3-A4 show an unusually long leader sequence. Class III lanthipeptide synthetase recognition sequences are highlighted in a red box.



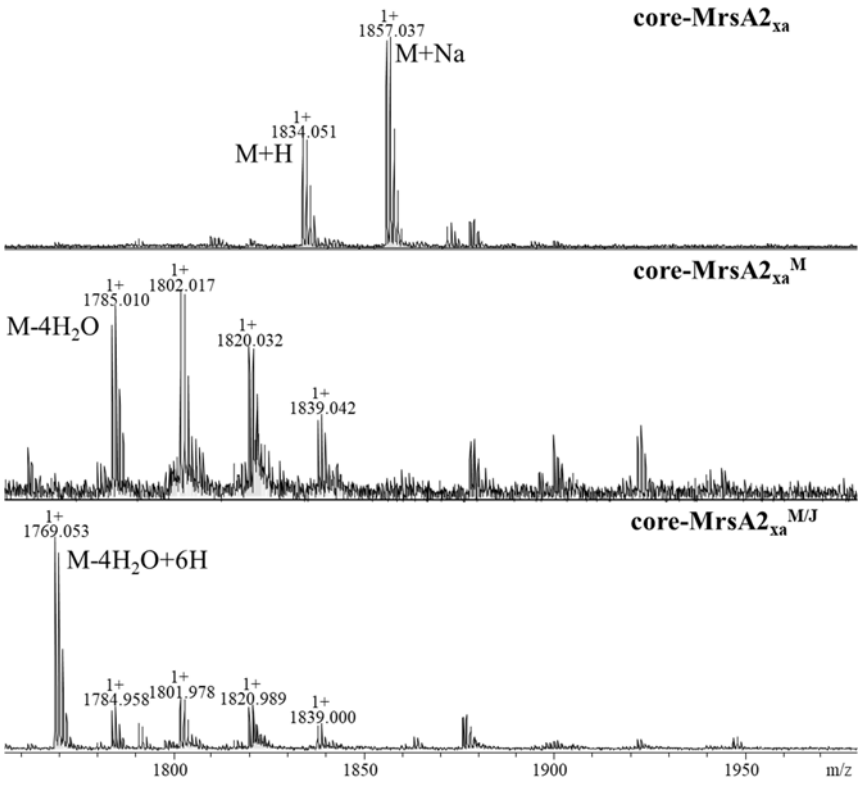
Supplementary Figure S12. Codon optimization of the radical SAM protein MrsY for expression in *E. coli*. The MrsY obtained from the genome of *S. griseus* S4-7 was submitted to codon optimization following the instructions of the GenArt gene synthesis from Thermo Fisher.



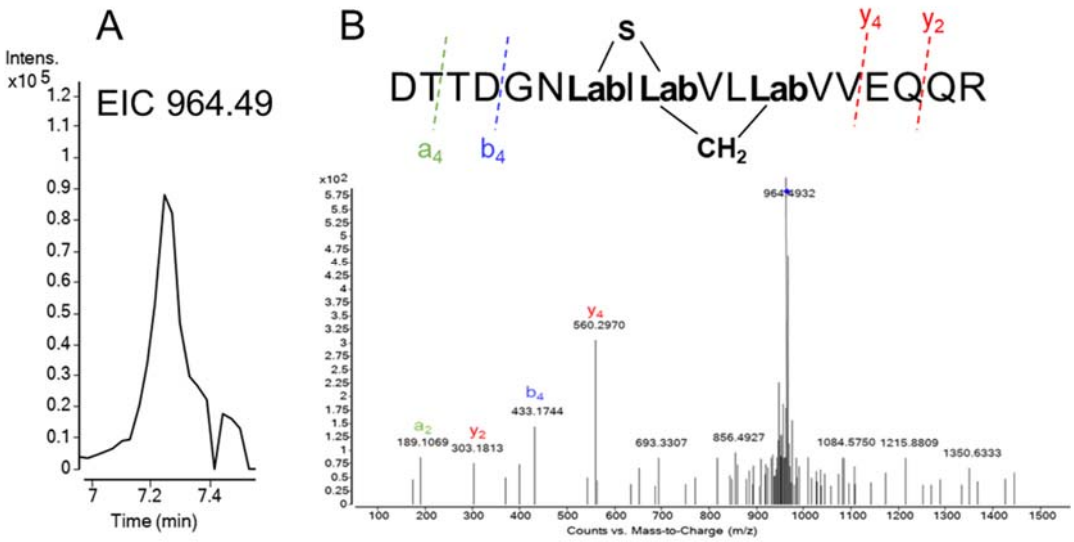
Supplementary Figure S13. Effect of the temperature on the radical SAM protein expression. SDS-PAGE gel of MrsY expressed in *E. coli* Rosetta under semi-anaerobic conditions at different temperatures. P (pellet): insoluble fraction. S: soluble fraction. L: unstained protein standard, broad range (10-200 kDa), NEB.



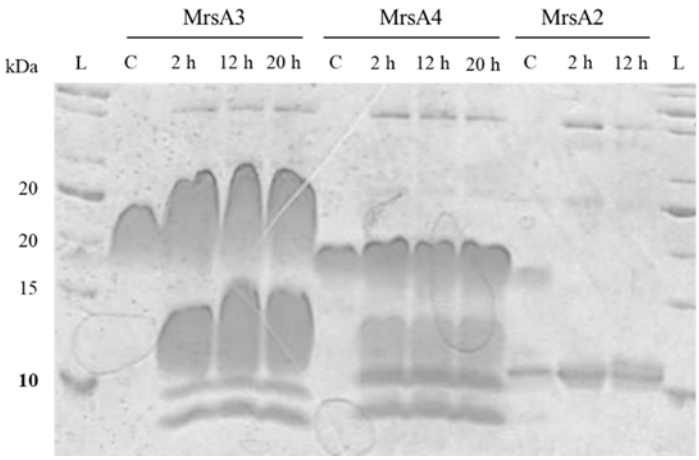
Supplementary Figure S14. MALDI-ToF mass spectra of unmodified precursor peptides MrsA1-A5 measured in linear mode. Calculated average masses: 10565.89 Da for MrsA1; 8757.50 Da for MrsA2; 9134.85 Da for MrsA3; 7801.44 Da for MrsA4, and 9644.73 Da for MrsA5.



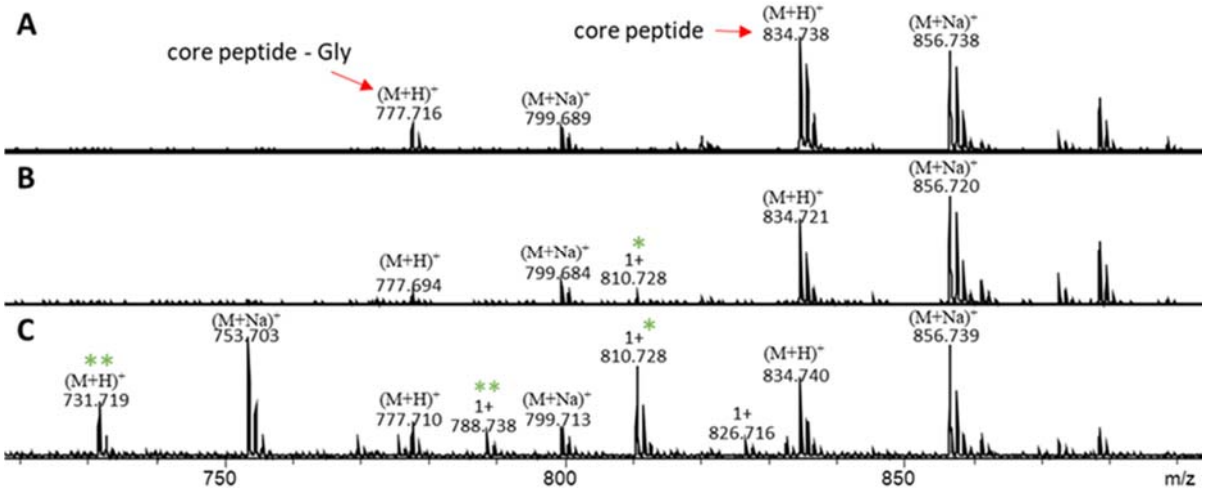
Supplementary Figure S15. *In vitro* reconstitution of MrsM and MrsJ activity in $MrsA2_{xa}$. **Upper panel:** MALDI-ToF mass spectrum of unmodified core $MrsA2_{xa}$ used as control. **Central panel:** MALDI-ToF mass spectrum of the core $MrsA2_{xa}$ that was obtained by co-expression with MrsM (core $MrsA2_{xa}^{M}$). **Lower panel:** MALDI-ToF mass spectrum of core $MrsA2_{xa}^{M/J}$ showing the 6 Da increase upon treatment of dehydrated $MrsA2_{xa}^{M}$ with MrsJ. After *in vitro* reaction, samples were incubated with Factor Xa for leader peptide removal (Calculated and observed average masses are summarized in Supplementary Table S3).



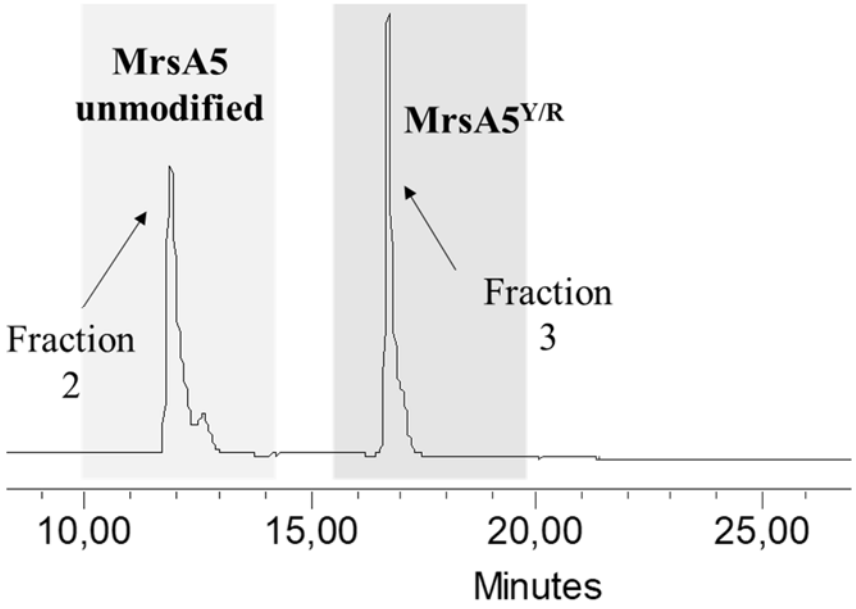
Supplementary Figure S16. Confirmation of labionin formation. **A:** EIC of $MrsA3_{xa}^{KC}$ after treatment with iodoacetamide (IAA) did not show the indicative mass shift of thiol alkylation. **B:** High resolution MS/MS spectrum of the corresponding $MrsA3_{xa}^{KC}$ peptide. For this experiment, a precursor peptide carrying the Factor Xa cleavage sequence was used.



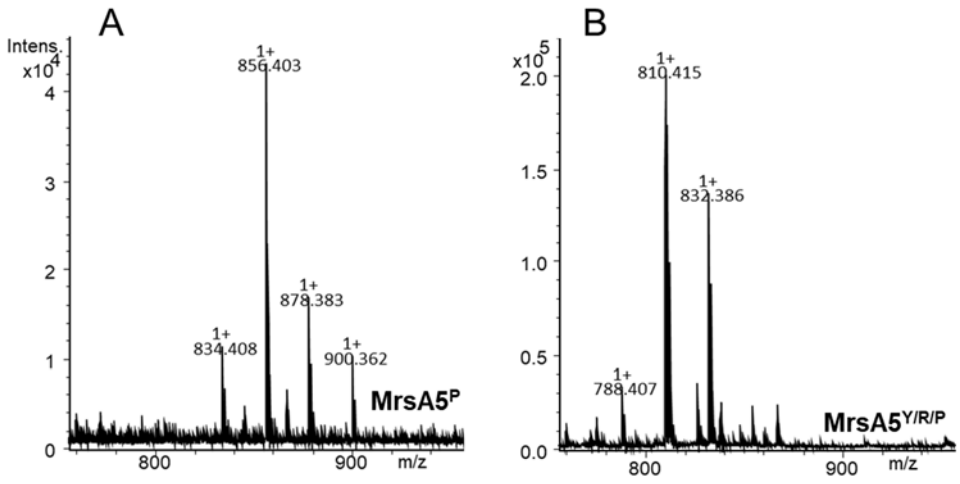
Supplementary Figure S17. SDS-PAGE gel of linear MrsA2-A4 incubated with MrsP at room temperature. C: control. L: unstained protein standard, broad range (10-200 kDa), NEB.



Supplementary Figure S18. Modification of MrsA5 by MrsY and MrsR *in vivo*. **A:** MALDI-ToF mass spectrum of unmodified MrsA5 used as control (MrsA5^{P}). Peptide digestion with MrsP yielded two fragments: one (*m/z* 834.71) corresponded to the predicted fragment while that at *m/z* 777.71 indicated the loss of a glycine residue. **B:** MALDI-ToF mass spectrum of MrsA5 co-expressed with MrsY and MrsR in *E. coli* showing a new peak at *m/z* 810.72 corresponding to the $[M+Na]^+$ ion. **C:** MALDI-ToF mass spectrum of MrsA5$^{Y/R/P}$ after further incubation with MrsR under aerobic conditions showed an increase in the intensity of the $[M+Na]^+$ ion at *m/z* 810.72. Note that the *m/z* at 731.71 corresponds to the decarboxylated peptide minus the *C*- terminal Gly.



Supplementary Figure S19. Chromatogram extracted at 280 nm corresponding to the purification of MrsA5$^{Y/R}$. After reconstitution of the MrsY and MrsR activity on the precursor MrsA5, MrsP was added to the reaction for leader removal. The sample was injected on a HPLC-DAD and two fractions (2 and 3) were collected.



Supplementary Figure S20. (A) MALDI-ToF mass spectra of the unmodified MrsA5 core peptide purified by HPLC in fraction 2 and **(B)** mature MrsA5$^{Y/R/P}$ collected in fraction 3.

9.2 Supplementary Tables

Supplementary Table S1: Fragments and masses expected upon tryptic digestion of MrsA2 precursor peptide. Mw (average mass): 8756.50 / Mw (monoisotopic mass): 8751.15. The fragment where the modifications would occur is in red color and residues susceptible for modification in bold.

Mass $[M+H]^+$	Position	Peptide sequence
3617.75	22-58	QFGESALQASGEIAHEGTPE HPAGQISLGSTGGLGAR
2620.20	61-84	LL**S**A**S**EGE**S**GF**S**HDLPW**TT**M**T**APW
1768.84	1-16	GSSHHHHHHSSGLVPR
559.26	17-21	GSHMK

Supplementary Table S2: Calculated and observed monoisotopic masses of MrsA2 after *in vitro* reconstitution of MrsM and MrsJ activity. Masses correspond to the fragment obtained after tryptic digestion.

Peptide	Calculated Mass $(M+2H)^{2+}$	Observed Mass $(M+2H)^{2+}$
MrsA2 (trypsin)	1310.6049	1310.60
$MrsA2^{M}$ - $4H_2O$ (trypsin)	1274.60	1274.58
$MrsA2xa^{MO}$ - $4H_2O$+6H (trypsin)	1277.60	1277.54

Supplementary Table S3: Calculated and observed average masses of core MrsA2 after *in vitro* reconstitution of MrsM and MrsJ activity. Leader peptide was removed by incubation with factor Xa. Observed masses were obtained by MALDI measurements in reflective mode.

Peptide	Calculated Mass		Observed Mass	
	$(M+H)^+$	$(M+Na)^+$	$(M+H)^+$	$(M+Na)^+$
core MrsA2xa	1835.04	1857.03	1835.05	1857.03
core $MrsA2xa^{M}$ ($-4H_2O$)	1763.04	1785.03	--	1785.01
core $MrsA2xa^{MO}$ ($-4H_2O$+6H)	1769.04	1791.03	1769.05	1791.02

Supplementary Table S4: Calculated and observed monoisotopic masses of MrsA3-A4 after *in vitro* reconstitution of MrsKC activity. MrsP was used to partially remove the leader peptide; thus, masses correspond to the core region obtained after incubation with MrsP.

Peptide	Calculated Mass $(M+3H)^{3+}$	Observed Mass $(M+3H)^{3+}$
MrsA3 (MrsP)	1093.19	1093.20
$MrsA3^{KC}$ (MrsP)	1081.19	1081.19
MrsA4 (MrsP)	1123.23	1123.23
$MrsA4^{KC}$ (MrsP)	1111.23	1110.86

Supplementary Table S5: Calculated and observed average masses of linear MrsA2-A5 upon incubation with MrsP. Observed masses were obtained by MALDI measurements in linear mode.

Peptide	Calculated Mass		Observed Mass	
	$(M+H)^+$	$(M+2H)^{2+}$	$(M+H)^+$	$(M+2H)^{2+}$
Full MrsA2	8757.50	4379.25	8769.05	4378.99
Full MrsA3	9134.85	5467.93	9146.48	4568.41
Core-MrsA3	3279.55	1640.28	3260.50	--
Leader-MrsA3	5874.32	2937.66	5873.73	--
Full MrsA4	7801.44	3901.22	7802.54	3897.41
Core-MrsA4	3369.76	1685.38	3350.75	--
Leader-MrsA4	4450.69	2225.85	4442.58	--

10. Acknowledgements

First and foremost, I would like to thank Prof. Dr. Gabriele König and Dr. Max Crüsemann for giving me the opportunity to join the working group, and for guiding me through all the journey to get my PhD. Thank you, Prof. König, for your reply to my very first email in 2017 and for your support in the critical moments. Thank you Max, for this amazing project, for your counseling, patience, confidence, and support at all times.

Many thanks to those that contributed to this research project:

Linda Brandt, Jessica Minich and Francesca Pent for conducting some of the cloning and heterologous expression experiments.

Prof. Dr. Christiane Dahl and Martina Grosser for your contribution to the reconstitution of the MrsY Fe-S clusters.

Dr. Marc Sylvester and his mass spec expertise were crucial in the development of the mrs project.

My deepest gratitude to all current and former members of the AG König/Crüsemann for your constant support, solidarity, and willingness to help.

I would also like to acknowledge the German Academic Exchange Service (DAAD) for financially supporting my PhD with the program “Research Grants – Doctoral Programmes in Germany”, and to the University of Bonn for the “get finished” study completion grant. Without their support, none of these would have been possible.