Bacterial natural products

Prediction, regulation and characterization of biosynthetic gene clusters in Actinobacteria

Ana Ceniceros

Microbial Physiology

University of Groningen

The work described in this thesis was carried out in the Microbial Physiology group of the Groningen Biomolecular Sciences and Biotechnology Institute (GBB) of the University of Groningen, The Netherlands. The author gratefully acknowledges the financial support of University of Groningen and the GBB for printing this thesis

This research was supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organization for Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs, and by the University of Groningen.

Cover design: Javier Alonso-Olalla

Printed by: Ipskamp Drukkers, Enschede

ISBN: 978-90-367-9506-7

ISBN: 978-90-367-9505-0 (electronic version)

university of groningen

Bacterial natural products

Prediction, regulation and characterization of biosynthetic gene clusters in Actinobacteria

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Friday 17 February 2017 at 11.00 hours

by

Ana Ceniceros

born on 23 March 1986
in Pamplona, Spain

Supervisor
Prof. L. Dijkhuizen

Co-supervisor
Dr. J.M. Petrusma

Assessment Committee
Prof. O.P. Kuipers
Prof. A.J.M. Driessen
Prof. H.A.B. Wosten

A mis padres

Contents

CHAPTER 1

General introduction

Natural products and their physiological roles

Natural products are compounds produced by living organisms. They can be essential compounds such as vitamins, or secondary metabolites which are not essential for an organism but may provide an advantage in their natural environment. These metabolites are therefore also referred to as specialized compounds. Latex is an example of a natural product obtained from plants. It is formed by a mixture of different chemicals that include alkaloids and rubber, and proteins such as proteases or chitinases. This mixture acts as defence mechanism against insects by trapping them or intoxicating them. It is also thought to be a way to excrete waste compounds, coverage of damaged tissue and defence against pathogens [1]. Natural products can also be chelators, molecules that improve the availability of essential metals that are normally present in low amounts in the environment. Iron chelators (siderophores) are of great importance since iron is usually available in limiting amounts. In pathogenic bacteria like *Mycobacterium tuberculosis*, living as an intracellular parasite, siderophores are essential for their survival since the iron is sequestered by the host cell [2]. Pigments are another important type of natural product. They can function as photoreceptors in photosynthesis, light protector, antioxidants or can be involved in virulence, as is the case for carotenes [3]. Antibiotics are thought to play a role as defence mechanism by eliminating competing organisms from their habitat. But in many cases the physiological role of secondary metabolites is not fully understood. Some antimicrobial compounds are produced (under laboratory conditions) in too low amounts to be able to inhibit the growth of surrounding organisms and therefore are thought to have a different function in the producer strain, possibly as signalling molecules or involved in motility [4]. Many secondary metabolites are highly valued by humankind in a wide range of applications. The most relevant producers of secondary metabolites are plants, fungi and bacteria. Depending on the chemical nature of the precursor, they are classified in 5 major classes: saccharides, terpenes, alkaloids, peptides or

polyketides. The latter are synthesized by specialized enzymes named polyketide synthases (PKS). Peptide metabolites include ribosomally and nonribosomally synthesized ones. The last are synthesized by dedicated enzymes called nonribosomal peptide synthetases (NRPS) [5]. In many cases these are mixed types of metabolites synthesized from hybrid gene clusters, e.g. NRPS plus PKS [6, 7].

Saccharides

Saccharides, also known as carbohydrates, are a large group of molecules categorized by the number of monomers that form them as monosaccharides, oligosaccharides and polysaccharides. Saccharides may have very diverse functions and act either as primary or secondary metabolites. Polysaccharides for instance are a main component in biofilms, important for survival and colonization of surfaces [8]. These biofilms support survival of microorganisms in their environment. Biofilms are especially problematic in public health since under these conditions microorganisms are more resistant to antimicrobial agents and they are difficult to remove from contaminated instruments [9]. The best-known example of biofilm formation is dental plaque. But biofilms are especially dangerous when involving colonization by *Staphylococcus* spp. of surgical instruments in hospitals which results in many deaths every year [10] (Figure 1). Saccharides also form the O-antigen from lipopolysaccharides in the outer membrane of Gram-negative bacteria, which results in the different serotypes and are important for their pathogenicity and detection by the adaptive immune system [11, 12]. Furthermore, saccharides can also be bioactive molecules. This class of molecules has not been explored for bioactivity as much as other classes. Bioinformatics analysis has shown that they constitute a big proportion of the gene clusters present in *Actinobacteria*, the main bacterial source of secondary metabolites [13, 14]. This class of molecules is highly interesting and may hold a great number of novel compounds with bioactive properties.

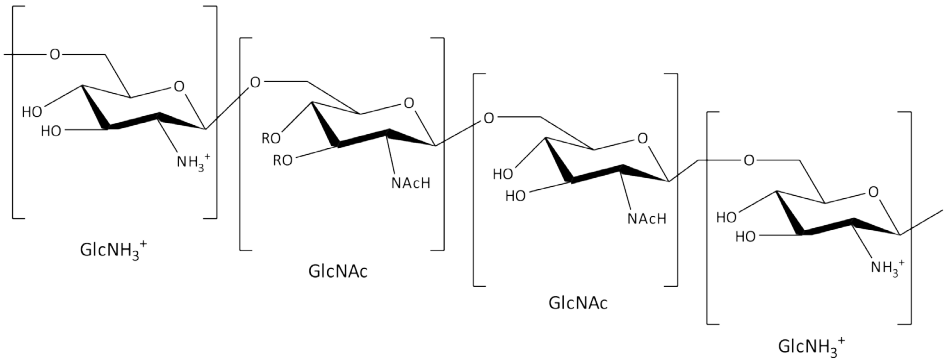


Figure 1 Representation of the structure of the intercellular polysaccharide adhesin (PIA) I from *Staphylococcus epidermidis* which is part of its biofilm and formed by β-1,6-linked 2-acetamido-2-deoxy-D-glucopyranosyl residues (GlcNAc). In some cases, they are not acetylated and therefore have a positive charge ($GlcNH_3^+$). Adapted from Rhode *et al*. [15].

Terpenes

Terpenes are one of the biggest chemical families of secondary metabolites. Their properties vary tremendously, from fragrances or pigments to hormones and bioactive compounds [16, 17]. This type of molecules are formed by terpene synthases which contain characteristic domains which have improved their annotation in the genomes of bacteria, fungi and plants [16]. The identification of terpene synthases in bacteria was challenging since they do not possess a high sequence similarity but was recently improved by Yamada et al. [16] by optimizing the Hidden Markov Models (HMM) [18] for their detection. Yamada et al. also describe a large number of previously unknown putative terpene biosynthesis gene clusters (BGCs) in different strains of bacteria [16]. Terpenoids are formed by C5 units (isopentenyl diphosphate or dimethylallyl diphosphate). Terpenes are classified by the number of isoprenoids units that form them as hemiterpene (1 unit), monoterpene (2 units), sesquiterpenes (3 units), diterpenes (4 units), sesterpenes (5 units), triterpenes (6 units) (Figure 2), tetraterpenes (8 units) and polyterpenes with more than 8 units [19].

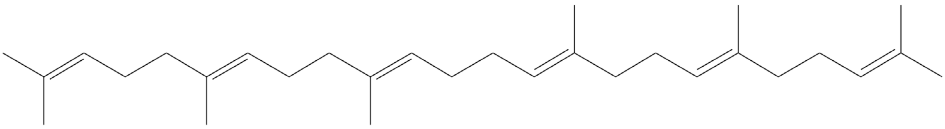


Figure 2 Structure of the triterpene squalene.

Alkaloids

Originally described from plants, alkaloid secondary metabolites are currently referred to as those organic compounds that contain basic nitrogen, which comprises a wide range of molecules with different chemical structures and functions, and that produce various physiological reactions. Some notable examples are cocaine, morphine, nicotine or caffeine [20]. A few alkaloids have been described in bacteria, such as pyreudiones A-D (Figure 3) which is a bicyclic pyrrolizidine alkaloid produced by a strain from *Pseudomonas fluorescens.* Pyreudiones are a defence mechanism against predation by amoebas. Their biosynthesis involves a nonribosomal peptide synthetase [21].

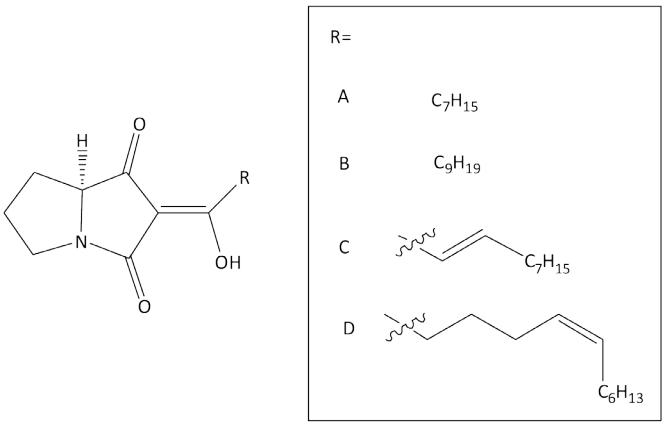


Figure 3 Structure of pyreudiones A-D. Adapted from Klapper *et al.* [21].

Polyketides

Polyketides (PKs) are another important class of secondary metabolites. Many biologically/industrially/medically important compounds are PKs such as the antibiotic erythromycin (Figure 4) [22]. They can be synthesized by three classes of polyketide synthases (PKSs), PKS-I, PKS-II or PKS-III.

PKS-I are modular enzymes, containing a loading module formed by an acyltransferase (AT) and an acyl carrier protein domain (ACP) that loads the substrate. The loading module is followed by one or several modules that will synthesize the product (synthesis modules). The synthesis modules contain at least a ketosynthase domain (KS), an acyltransferase domain (AT) and an ACP domain (Figure 5a). In some cases, they also contain a ketoreductase domain (KR) which transforms the keto group into a hydroxyl group, a dehydratase (DH) that creates a double bound or an enoylreductase (ER) that introduces a single bond [23]. These modules elongate the synthesized molecule by 2 carbon atoms at a time. The last module contains a thioesterase (TE) domain to release the final molecule. PKS-I enzymes may have a tremendous size depending on the number of modules that they contain. PKS-II are formed by multi-enzyme complexes that act iteratively. The minimal system is formed by two KS subunits and an ACP (Figure 5b). Normally they synthesize polycyclic aromatic compounds [24, 25]. PKS-III are single subunit enzymes that have an iterative condensing activity and that are ACP independent (Figure 5c) [23, 26, 27].

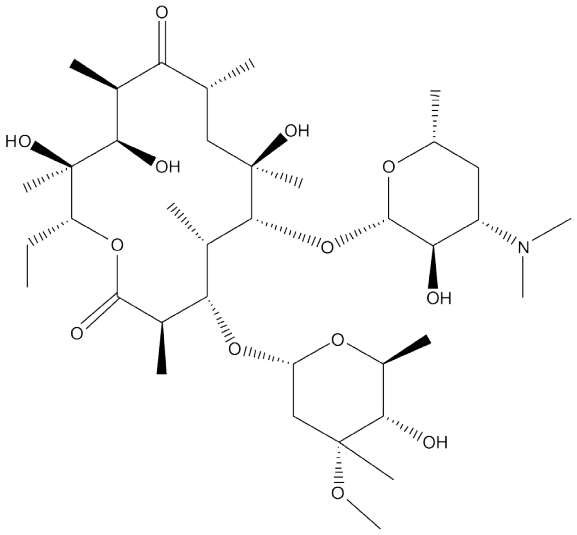


Figure 4 Structure of the antibiotic erythromycin (ChemSpider).

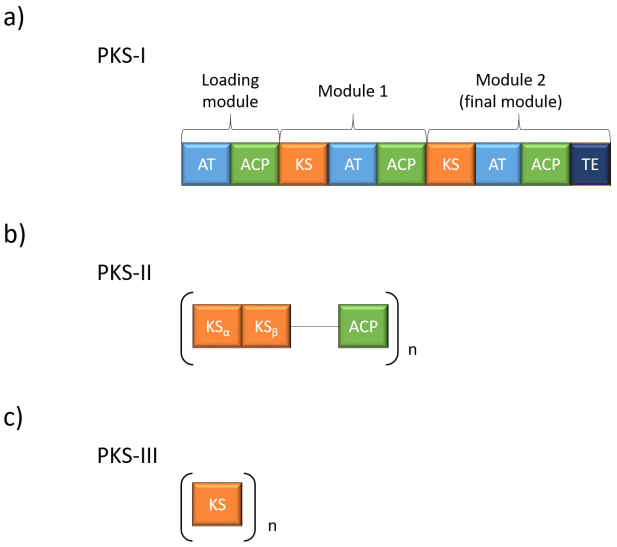


Figure 5 Domains found in the three types of polyketide synthases. a) PKSs from type I are modular enzymes with a loading module followed by a number of elongation modules. The last module contains a thioesterase domain (TE) that detaches the final product. b) Type II PKSs are formed by separate enzymes that act repeatedly to form the final product. At the minimum two ketosynthetase domains (KS_α and KS_β) and an acyl carrier protein (ACP) are involved. c) Type III PKSs are single subunits that act iteratively and do not need an ACP domain. AT: Acyltransferase. Adapted from [23, 26, 27].

Ribosomally synthesized and Post-translationally modified Peptides (RiPPs)

Another type of secondary metabolite is that of ribosomally synthetized peptides that undergo essential post-translational modifications. They often have a very narrow antimicrobial activity, in most cases only affecting closely related organisms. Some broader spectra RiPPs have also been found [28]. In most cases the synthesis of these molecules starts by a core protein encoded by one gene. This core protein often contains an N-terminal leader peptide that needs to be recognized by modifying proteins and by transport proteins. In a few cases the leader peptide is located at the C-terminus, as is the case for bottromycins, macrocyclic compounds with antibiotic activity [28]. Many different families of molecules are included in RiPPs, among them lanthipeptides which contain meso-lathionine and 3-methyllanthionine residues. Nisin is the better-known example of this family (Figure 6) [28].

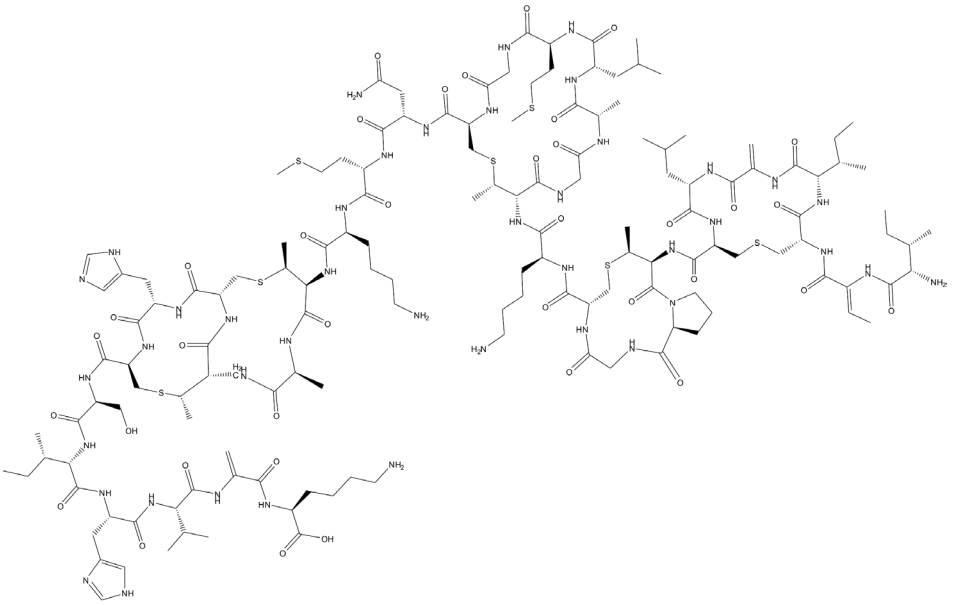


Figure 6 Structure of nisin (ChemSpider).

Nonribosomal peptides (NRPs)

Ribosomally synthetized peptides are formed by a sequence made up of the 20 canonical amino acids. The peptides formed by nonribosomal peptide synthetases (NRPS) however usually contain other residues as well. NRPSs can utilize about 500 different non-proteinogenic substrates, including D-amino acids [29], resulting in synthesis of a wide range of molecules. NRPSs are modular enzymes in which each module is in charge of adding a specific monomer to the final peptide molecule. Many important compounds are products of these enzymes such as the fungal antibiotic penicillin, the anticancer compound bleomycin or the immune-depressant cyclosporine A (Figure 7) [30]. Each module in an NRPS contains at least an adenylation domain (A) that has specificity for a substrate and will activate it to start the synthesis, a thiolation (T) or peptide carrier protein (PCP) domain where the activated residue is covalently bound and a condensation domain (C) that catalyzes the elongation of the peptidyl chain (Figure 8). The C domain shows specificity for the residue activated by the downstream A domain and can be of different types

depending on the condensation reaction they catalyze [31]. In some cases, an NRPS can have an initial C domain, called C-Starter, that acylates the first residue of the peptidyl chain [31]. The last module also contains a TE module that detaches the final peptide from the enzyme [29]. Depending on the number of modules needed, these enzymes may be very large, indicating a possibly very complex product. Their large gene sizes also make these NRPSs difficult to study, e.g. they are difficult to amplify, clone and express. However, the nature of these enzymes makes them easily detectable in a bioinformatics analysis, and the development of new molecular biology techniques may facilitate their study.

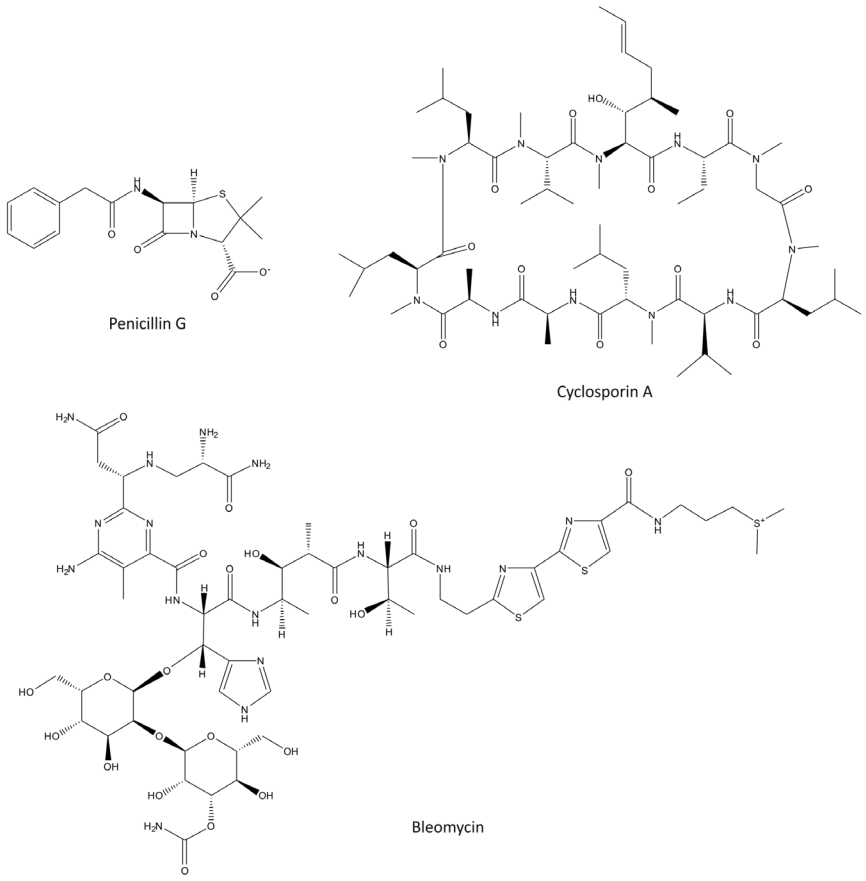


Figure 7 Structure of penicillin G, cyclosporine A and bleomycin, all representing examples of NRPs.

Figure 8 Domains most commonly present in NRPS modules. The last module contains an extra thioesterase domain which allows the detachment of the final product. A: adenylation domain. PCP: peptide carrier protein. C: Condensation domain. Adapted from [29]

Human uses of natural products

Natural products can be structurally very diverse molecules and may have very different functions for humans. Even one molecule may have different uses already. For instance, the red tripyrrole pigment prodigiosin, produced by a NRPS-PKS hybrid cluster [32] in different bacterial strains, such as the current industrial producer *Serratia marcescens*, acts as a pigment. It furthermore, has antibiotic activity against bacteria, can be used to treat cancer as an immunosuppressive agent, and also has anti-parasitic activity [33, 34]. Other natural products are used as sweetener, such as the Stevia glycosides isolated from the plant *Stevia rebaudiana*. These compounds have received a lot of attention in the last years due to their higher sweetening capacity than sucrose, providing a possible substitute for sucrose [35, 36]. Natural products have a very important role in human health as medical drugs, e.g. as antibiotics, anti-parasite drugs, antitumor drugs, etc. [37-39]. Antibiotics are one of the most important revolutions in human medicine. Illnesses such as the black death, leprosy or tuberculosis have caused innumerable deaths. The causative agents of these diseases are pathogenic bacteria that have been fought successfully with antibiotics. Most of the antibiotics that we use today were discovered and synthesized between the years 1930-1960s. Since then, only 3 antibiotic classes have been commercialized, pseudomonic acid (1985), oxazolidinone (2000) and the previously known antibiotic class lipopeptides (2003), which had not been commercialized until then [40]. The development of resistance to antibiotics in pathogenic bacteria has become one of the main threats in current times. Numerous

studies have analyzed the appearance of antibiotic resistance in bacteria and great effort has been done to understand this evolutionary process [41-43]. Resistance genes already existed in nature before the widespread use of antibiotics started. Producer strains are not affected by their own antibiotic compounds, which implies that they possess a mechanism to counteract their effects. These resistance mechanisms differ depending on the mode of action and / or structure of the antibiotic molecule [41]. The genes responsible for antibiotic resistance are susceptible for recombination and can be transferred to other bacteria sharing the same environment. The misuse of antibiotics (in animal feed, treatment of viral infections, taking them for a more limited time than specified…) favours the appearance and selection of resistant bacteria. Fleming already foresaw this problem in 1945 [44]. Even more threatening is the occurrence of multidrug resistance. Currently there are some bacteria resistant to every antibiotic available, e.g. some strains of the notorious causative agent of tuberculosis (TB) *M. tuberculosis* [45]. Different strategies are used to find or synthesize new antibiotics, or new ways to fight bacterial diseases: modification of the natural molecules to produce new semi-synthetic molecules [46], synthesis of completely new compounds with different targets [47], making use of bacteriophages [48, 49], or searching for novel antimicrobial compounds produced in nature. In recent years, it has been discovered that bacteria have an even greater potential to produce a wide range of previously unknown secondary metabolites than it was initially thought [50, 51]. The potential to synthesize these novel molecules is hidden in the newly sequenced genomes of many bacteria. Different computational tools have been developed to facilitate the identification of BGCs in genomes. Some examples are antiSMASH, which predicts BGCs and classifies them into the molecule class that the pathway they encode is predicted to produce, BAGEL, which is specialized in the prediction of bacteriocins, or NRPSpredictor [52-54]. Many studies now focus on activating the synthesis and elucidating the structure of these unknown compounds [55-58]. The work described in this PhD thesis adds to the knowledge of this

topic by the use of different techniques to predict and activate cryptic bacterial gene clusters that encode the enzyme machinery that could produce novel compounds.

Potential of bacterial natural product synthesis

Bacteria and fungi are the largest sources of anti-microbial natural products in use today [5]. One of the most important bacterial genera for synthesis of natural products is *Streptomyces*. When the genomes of the first strains of this genus were sequenced, it was observed that they possess large genomes, reaching over 10 Mb in size. The analysis of the genomes uncovered a bigger potential for producing natural products than previously expected. This genus is of great importance since 2/3 of the antibiotics that are available for human use have been isolated from various *Streptomyces* strains [59]. One characteristic of members of this genus is their complex life cycle, producing vegetative mycelia when the medium still has nutrients and aerial mycelia which will form spores when the nutrients in the medium become depleted. Secondary metabolism is activated when aerial mycelium develops. This complex life cycle is controlled by an equally complex regulatory network consisting of global regulators that control several aspects of primary and/or secondary metabolism, pathway specific regulators that control normally only the expression of their own clusters, and various other factors such as signalling molecules, e.g. γ-butyrolactones [60, 61]. This genus is known to be able to produce high yields of different types of molecules and to hold in their genomes the potential for producing many compounds that have remained unknown, making it one of the most interesting targets to activate cryptic BGCs. For this reason, we studied the synthesis of selected secondary metabolites by *Streptomyces clavuligerus* in **Chapters 2** and **3**. *S. clavuligerus* was predicted to contain around 50 BGCs, most of which are completely unknown [51]. This number goes up to 97 when its sequence is analyzed by antiSMASH using the ClusterFinder algorithm, which also detects BGCs from unknown classes [13]. This strain is most well-

known for the production of clavulanic acid and cephamycin C, its main secondary metabolites [62, 63].

Analysis of the rapidly growing number bacterial genome sequences available has uncovered a great potential for synthesis of natural products by strains that had never been studied for this purpose, and from newly isolated strains [13, 64]. Strains that live in unconventional habitats may also be sources of completely novel arrays of compounds [65-67]. Recently, Zipperer and collaborators discovered lugdinin, produced by *Staphylococcus lugdunensis,* a strain isolated from the human nasal microbiome. This NRP antibiotic impairs *Staphylococcus aureus* colonization [67]. A great potential for natural product synthesis has also been found in strains that are well-known and currently used for different purposes. One example is the genus *Rhodococcus,* which has been most extensively studied for its catabolic abilities [68-71]. It has a simpler life cycle than *Streptomyces* and, in general, is faster growing. It represents aerobic, non-sporulating and acid resistant strains containing mycolic acids in their cell walls, and it is closely related to the *Mycobacterium* genus which also contains a number of cryptic BGCs [64]. Studies have shown that *Rhodococcus* genomes are large, also reaching 10 Mb, and contain an astonishing number of putative natural product clusters, mostly NRPSs [64]. In **Chapter 4** we performed a more detailed bioinformatics analysis of the potential of the *Rhodococcus* genus to produce natural products. Interestingly, this genus has remained largely unexplored for its secondary metabolic potential. Only four *Rhodococcus* antibiotics have been purified and structurally characterized. These are rhodopeptines, lipopeptide antibiotics with antifungal activity isolated from *Rhodococcus* sp. Mer-N1033; lariatins A and B, anti-mycobacterial peptides with lasso structure produced by *Rhodococcus jostii* K01-B0171; aurachin RE which is a quinolone antibiotic active against Gram-positive bacteria and produced by *Rhodococcus erythropolis* JCM 6824 [72-74]; The recently described humimycins, active against methicillin-resistant *Staphylococcus aureus* (MRSA). This compound however, was not

isolated from bacterial broth. It is the result of the synthesis of the predicted peptide putatively synthetized by an NRPS [75]. Rhodopeptines, lariatins and humimycins are novel molecules that had never been described before although they are part of already known antibiotic classes. Aurachin A-D had already been described as a product of *Stigmatella aurantiaca* in 1987 [76]. Interestingly, aurachin RE was found to be more active than the previously described homologues. Its structure is very similar to that of aurachin C but with an extra hydroxyl group that seems to provide this improved activity [73]. The higher activity of aurachin RE compared to the previously described analogues, the activity of lariatins against *M. tuberculosis*, one of the biggest health threats that we are currently facing, the novel antifungal rhodopeptine, and the different bioactive compounds that have been detected in this genus but have not yet been characterized [77, 78], show how promising the study of *Rhodococcus* cryptic BGCs is. Another interesting feature of *Rhodococcus* is its possible use as host for heterologous gene expression. Kurosawa *et al*. [79] observed that *Rhodococcus fascians* DDO356 is able to exchange genetic material with *Streptomyces padanus* MITKK-103 and that *Streptomyces* genes are expressed in *Rhodococcus*, leading to the production of the novel aminoglycoside antibiotics rhodostreptomycins. *Rhodococcus* strains thus may be suitable for heterologous expression of *Streptomyces* cryptic gene clusters. In **Chapter 2** we demonstrated that this is possible: introduction of the putative indigoidine synthetase from the cryptic indigoidine gene cluster from *S. clavuligerus* controlled by a strong constitutive promoter in *R. jostii* RHA1 resulted in production of a blue pigment, identified as indigoidine. In **Chapter 5** we show that *R. jostii* RHA1, known for its ability to degrade polychlorinated biphenyls [69], potentially is able to produce bioactive compounds that had never been described before. *R. jostii* RHA1 also encodes and produces γ-butyrolactone signal molecules **(Chapter 5)** known to be involved in the regulation of secondary metabolism in *Streptomyces* [80].

Production of natural products under laboratory conditions

The study of natural products largely has been limited to those compounds that are produced under laboratory conditions and that are easily detected, as is the case of clavulanic acid or cephamycin in *S. clavuligerus*. The metabolites and their amounts produced by an organism differ depending on the media and growth conditions used [81, 82]. The isolation and identification of a specific compound from a culture broth can be challenging if the compound is produced in minor amounts or if there is no easy and fast way to detect them. The development of faster and cheaper genome sequencing methods and their computational analysis has led to the discovery of a vast number and large diversity of putative BGCs predicted to be involved in the synthesis of natural products. These clusters of genes often include transporters and regulators involved in the synthesis of these compounds. Genome analysis of strains known to produce secondary metabolites also has shown the presence of many more putative BGCs that are cryptic, either not produced under laboratory conditions (silent) or with unknown products. In the natural and wild type situation, the products of these clusters generally are produced in very small amounts only and also may be toxic for the host organism if produced in high quantities. As mentioned before, specific conditions of cultivation may be needed to activate expression of these cryptic gene clusters. The natural habitat of the bacterial species studied generally is completely different and much more complex than the conditions used in a laboratory. Under natural conditions bacteria are, for example, more subjected to physico/chemical changes in the environment, e.g. temperature changes, changes in the microbial community, etc. The regulatory systems used by bacteria to control gene expression and synthesis of these compounds are complex. A full understanding of the regulation of secondary metabolism therefore is still elusive. A current goal of the scientific community is to learn how to induce the expression of cryptic gene clusters and to stimulate production of these novel secondary metabolites. First, promising BGCs

are identified by the use of different bioinformatics tools. Once identified, different techniques can be used to attempt their activation: Changing the substrate availability by modifying the medium composition, manipulating known biosynthesis pathways to make precursors available for other, unknown pathways, studying strains that have never been explored before for natural products, co-cultivation of strains that may induce the production of a compound in one or both of them, heterologous expression of gene clusters in other strains that are easier to manipulate than the parent strain, or the manipulation of pathways through synthetic biology approaches. All these techniques will be discussed in the following sections.

Identification of secondary metabolism gene clusters: Bioinformatics tools

Different bioinformatics tools have been developed in order to be able to predict the secondary metabolite arsenal available in the genome of different strains as well as to identify the most promising BGCs to study, avoiding rediscovery [83]. Secondary metabolite gene clusters can be detected by a manual search for the main biosynthetic enzymes by tools like BLAST [84] and Hidden Markov Models (HMMer) [18]. To identify the rest of the genes in the clusters, the proteins encoded by the surrounding genes also need to be manually searched. MultiGeneBlast was developed recently to be able to perform BLAST searches for more than one enzyme, facilitating the search for complete gene clusters [85]. Other tools have been developed that can automatically find and predict complete gene clusters in a genome sequence query. Some examples of these programs are antiSMASH, BAGEL, NRPSPredictor and ClusterFinder [83]. The first three software programs look for genes encoding proteins containing conserved domains known to be involved in the synthesis of specific secondary metabolites. The corresponding gene cluster is then categorized in different classes of natural products. These programs thus are limited to detect BGCs belonging to known molecule classes. ClusterFinder, however, translates the nucleotide sequences into a

succession of adjoining Pfam domains [13]. Then it calculates the probabilities of each domain to be part of a cluster by calculating the frequency that it is found in one based on a database of different BGCs with known products and a list of regions which are not predicted to be part of a gene cluster yet [13]. Therefore, ClusterFinder can find characterized and uncharacterized gene clusters [13, 83]. In **Chapter 3** we made use of this program to study secondary metabolism in *Rhodococcus*.

Activation of cryptic BGCs

The selection of promising species and/or gene clusters as targets for further studies is an essential step to avoid as much as possible the rediscovery of known compounds or spending time and resources on the production of a compound lacking antimicrobial activity. Different techniques that have been developed in the last years can be applied in order to activate cryptic BGCs. When the objective is to find new compounds from a strain with a specific function that can be screened for, like antibiotic synthesis, general techniques such as changing growth conditions can be used. However, when the goal is to study the product of a specific cluster, then more specific strategies need to be used. Below we give an overview of the different strategies used to study cryptic BGCs.

Enhancing availability of precursors or intermediate compounds

The discovery of the large number of cryptic BGCs present in bacteria that may encode completely novel molecules has stimulated scientists to explore how to induce their activation. For this purpose, several different techniques have been developed in the last decades. The lack of production of a compound can be due to a lack of precursors or substrates or even to a lack of inducers. Changing the media components and, thereby providing different substrates has sometimes shown positive results, as in case of closthioamide, an antibiotic isolated from *Clostridium cellulolyticum* [86]. This technique is also useful to activate synthesis of compounds that are only produced in the presence or

absence of a specific nutrient, as is the case for siderophores which are normally activated by low levels of iron [87]. In some cases, precursors needed for the synthesis of a compound of choice are used by other, more active, pathways which can lower or even block its production [88]. In other instances, the intermediates of a pathway can be regulating the biosynthesis of another one, as is the case for holomycin production in *S. clavuligerus* which is thought to be regulated by an intermediate of the clavulanic acid synthesis [89]. As a result of the development of efficient molecular biology techniques for an increasing number of bacteria, we are able to manipulate genes and pathways in these species. Redirection of precursors has also proven effective in improving production of different compounds. Engineering the metabolism of a strain, with a focus on glucose catabolism, to enhance the amount of acyl-CoA available, or modulating fatty acid biosynthesis, can activate or improve the production of the target molecule [88, 90, 91]. This strategy was followed in **Chapter 3**, using a strain of *S. clavuligerus* with seven deletions in the clavulanic acid gene cluster, completely blocking clavulanic acid synthesis.

Co-cultivation

Cultivation of two or more strains together has shown to be effective in activating cryptic pathways [92]. In nature, bacterial strains live in complex communities and are therefore subjected to numerous stimuli that are not present in a pure culture in the laboratory. It is therefore a strong possibility that most of the silent clusters are not active in pure cultures because they are not needed under these conditions. This technique is also not directed to a specific gene cluster, but it has been shown useful to find several antimicrobials by simply screening for inhibition of growth of the indicator strain by the producer strain [92].

The techniques described above are designed to activate or improve random pathways, since there is no way yet to target specific pathways. Also, another undesirable pathway may take over. There is also a high risk

of rediscovery of an already known compound or a very closely related one. This was for instance the case with etamycin (a cyclic peptide antibiotic), which was described for the first time simultaneously in an unidentified *Streptomyces* species by Heinemann and from *Streptomyces griseus* by Bartz in the middle of the 1950´s and found again in 2010 from a marine actinomycete [93-95]. In the following sections, we discuss strategies available for activation of cryptic BGCs that are more directed.

Targeting biosynthetic genes and manipulating regulatory networks

The risk of activating synthesis of a known compound and the need to target specific gene clusters has led to the development of more specific strategies. One of these approaches is to overexpress the genes for the main biosynthetic enzymes of a pathway, if it is possible to predict these, as is usually the case for NRPSs or PKSs [23, 26, 27, 29]. The exchange of the promoter region of the main biosynthetic genes by a known strong promoter may be a viable alternative strategy, although in many strains promoter regions are difficult to determine. In other cases, the expression of the main biosynthetic gene is not the limiting factor, which can be checked by expression analysis or by proteomics, but it is of course difficult to know which genes/enzymes should be targeted when very little is known about the cluster. Other genes in the cluster may be essential for the correct molecular configuration of the product, or the substrate may be limited or used for another biosynthetic pathway. Another way to activate cryptic gene clusters is to alter their regulation. But in most cases the regulation is still unknown, even in identified gene clusters. Where known, overexpressing activators or deletion/disruption of repressors may result in activation of one or more biosynthetic gene clusters [96, 97]. Olano and co-workers [56] were able to activate five cryptic gene clusters from *Streptomyces albus* J1074 by using the methods mentioned in this section. The choice of one strategy or the other depends on the targeted cluster. In case the main biosynthetic genes are not clearly distinguishable, or the gene cluster is too big, several promoter insertions may be needed to activate the complete gene

cluster. Modifying regulatory networks may activate the whole biosynthetic pathway, but in many strains not much is known about their regulation. Several global regulators are known to be involved in the synthesis of secondary metabolites in *Streptomyces*, such as (p)ppGpp synthetase, RelA or CRP [97, 98]. These regulators cannot be used to target one specific BGC but gene clusters are also in many cases regulated by pathway specific regulators like the denominated *Streptomyces* Antibiotic Regulatory Proteins (SARP). These SARPs were first described in the genus *Streptomyces* but subsequently also have been found in other genera [99]. The manipulation of these regulators provides a more targeted activation. Another regulation level that has been described for secondary metabolites are signalling molecules such as γ-butyrolactones which have also been related to secondary metabolism and morphogenesis in *Streptomyces* [100]. These small molecules bind to their receptors belonging to the TetR transcriptional regulators family, which are in most cases repressors, and change their conformation, thus interrupting repression [80, 101, 102]. In **Chapter 5** we identified these γ-butyrolactone signalling molecules in rhodococci , which is the first report of these molecules outside the *Streptomyces* genus. The γ-butyrolactone system may also be involved in the regulation of secondary metabolism in rhodococci, which has a great potential for their production, as shown in **Chapter 4**. Regulatory pathways, however, consist of multiple levels that interact with each other forming a complex regulatory network that is barely understood, which complicates the identification of genes that should be targeted to activate a specific pathway. Further studies on regulation systems are needed to learn how to better target these cryptic gene clusters.

Heterologous expression of putative cryptic biosynthetic gene clusters

The techniques described in the previous section can only be applied in organisms which can be genetically modified and that can be properly cultivated for the production of the desired compound. In many cases, the tools for manipulating a specific strain are scarce or are not yet

developed, or the strain grows very slowly, or cannot be cultured for fermentation, or cultured at all. One alternative is to express the gene clusters of interest into a better-known strain, also allowing their further engineering [56-58]. This is also a promising strategy for studying the BGCs detected in human pathogens, e.g. in Mycobacteria which can be a source of many natural products [64], including *M. tuberculosis*, as also shown in **Chapter 4**. Even when this pathogen can be genetically manipulated, the question remains whether it is wise to force expression of unknown BGCs that may help them becoming even more virulent. The heterologous expression approach has in some cases been successful, without requiring further engineering of the sequences, e.g. in case of chloramphenicol and congocidine [57, 103]. Heterologous expression faces however, different challenges, from codon optimization, especially when the GC/AT content of the genome is very different between both strains [104], to unknown regulatory features [105]. Several strains are currently available for heterologous expression studies of BGCs, mainly from the *Streptomyces* genus and *E. coli* [58, 106-109], but the number of hosts is still very limited and the expression systems available are not effective in many cases. A greater array of possible hosts and a better understanding of the complex regulatory networks in these strains are needed to achieve a higher success rate in the expression and production of new natural products.

Production of natural compounds using synthetic biology

Another approach to activate cryptic gene clusters is to use the emerging discipline of synthetic biology. The goal of this strategy is to improve or modify existing biosynthetic pathways by exchanging genes encoding transporters, regulators and any other protein/enzyme necessary for the biosynthesis by others that will induce the production of higher titres of the desired compound or a modified derivative [110, 111]. These “improved” proteins/enzymes are called building blocks. Currently most effort is made in generating such building blocks. This is not an easy task since there is still a lot that we don’t know about regulation or metabolism in

donor and host strains. This strategy is very promising and will presumably provide us with novel compounds in the near future.

Scope of this thesis

In this thesis, we have explored the potential of two genera of Actinomycetales for natural product synthesis, paving the way for the discovery of novel compounds. **Chapter 1** provides a literature review of natural product synthesis with emphasis on bioinformatics analysis of biosynthetic gene clusters (BGCs) and attempts to activate expression of cryptic clusters using diverse approaches. In **Chapters 2** and **3** we attempted to activate cryptic BGCs of *S. clavuligerus*. In **Chapter 2** we were able to produce indigoidine by heterologous expression of the predicted homologue of the indigoidine synthetase IndC [51, 112] in different strains of *Streptomyces*, and in *R. jostii* RHA1. We also studied the function of an extra 4-oxalocrotonate tautomerase-like domain IndD that in *S. clavuligerus* ATCC 27064 is fused to the indigoidine synthetase. In **Chapter 3** we constructed a deletion strain of *S. clavuligerus* ATCC 27064 lacking the first seven genes in the clavulanic acid biosynthesis cluster (strain Δ7) to block this pathway and redirect precursors. This resulted in the (partial) characterization of tunicamycin-like MM 19290 antibiotics [113] isolated from *S. clavuligerus* strain Δ7. In **Chapter 4** we performed a bioinformatics analysis of the BGCs present in the genome of 20 different *Rhodococcus* strains, 7 *Mycobacterium* strains and 1 *Amycolicicoccus subflavus* strain, revealing an impressive potential for secondary metabolite synthesis by strains of these 3 genera. In **Chapter 5** we studied the predicted γ-butyrolactone signalling system in *R. jostii* RHA1, one of the *Rhodococcus* strains with a relatively high number of putative BGCs (Chapter 4). This *R. jostii* γ-butyrolactone biosynthetic gene cluster (RJB) encodes a molecule with the same structure as the predicted precursor of one of the γ-butyrolactone molecules produced by *S. coelicolor* [114].

References

1. Konno, K. Plant latex and other exudates as plant defense systems: roles of various defense chemicals and proteins contained therein. *Phytochemistry* **72**, 1510-1530 (2011).

2. Sritharan, M. Iron homeostasis in *Mycobacterium tuberculosis*: mechanistic insights into siderophore-mediated iron uptake. *J. Bacteriol.* **198**, 2399-2409 (2016).

3. Mohammadi, M., Burbank, L. & Roper, M. C. Biological role of pigment production for the bacterial phytopathogen *Pantoea stewartii* subsp. *stewartii*. *Appl. Environ. Microbiol.* **78**, 6859-6865 (2012).

4. Raaijmakers, J. M. & Mazzola, M. Diversity and natural functions of antibiotics produced by beneficial and plant pathogenic bacteria. *Annu. Rev. Phytopathol.* **50**, 403-424 (2012).

5. O'Connor, S. E. Engineering of secondary metabolism. *Annu. Rev. Genet.* **49**, 71-94 (2015).

6. Gallagher, K. A., Fenical, W. & Jensen, P. R. Hybrid isoprenoid secondary metabolite production in terrestrial and marine actinomycetes. *Curr. Opin. Biotechnol.* **21**, 794-800 (2010).

7. Bergmann, S. *et al*. Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans*. *Nat. Chem. Biol.* **3**, 213-217 (2007).

8. Donlan, R. M. Biofilms: microbial life on surfaces. *Emerg. Infect. Dis.* **8**, 881-890 (2002).

9. Donlan, R. M. Biofilm formation: a clinically relevant microbiological process. *Clin. Infect. Dis.* **33**, 1387-1392 (2001).

10. Sun, D. *et al*. Inhibition of *Staphylococcus aureus* biofilm by a copper-bearing 317L-Cu stainless steel and its corrosion resistance. *Mater. Sci. Eng. C. Mater. Biol. Appl.* **69**, 744-750 (2016).

11. Duan, Z. *et al*. Genetic diversity of O-antigens in *Hafnia alvei* and the development of a suspension array for serotype detection. *PLoS One* **11**, e0155115 (2016).

12. Ramachandran, G. Gram-positive and gram-negative bacterial toxins in sepsis: a brief review. *Virulence* **5**, 213-218 (2014).

13. Cimermancic, P. *et al*. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412-421 (2014).

14. McCranie, E. K. & Bachmann, B. O. Bioactive oligosaccharide natural products. *Nat. Prod. Rep.* **31**, 1026-1042 (2014).

15. Rohde, H., Frankenberger, S., Zähringer, U. & Mack, D. Structure, function and contribution of polysaccharide intercellular adhesin (PIA) to *Staphylococcus epidermidis* biofilm formation and pathogenesis of biomaterial-associated infections. *Eur. J. Cell Biol.* **89**, 103-111 (2010).

16. Yamada, Y. *et al*. Terpene synthases are widely distributed in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 857-862 (2015).

17. Elissawy, A. M., El-Shazly, M., Ebada, S. S., Singab, A. B. & Proksch, P. Bioactive terpenes from marine-derived fungi. *Mar. Drugs* **13**, 1966-1992 (2015).

18. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).

19. Singh, B. & Sharma, R. A. Plant terpenes: defense responses, phylogenetic analysis, regulation and clinical applications. *3 Biotech* **5**, 129-151 (2015).

20. Kishimoto, S., Sato, M., Tsunematsu, Y. & Watanabe, K. Evaluation of biosynthetic pathway and engineered biosynthesis of alkaloids. *Molecules* **21**, 10.3390/molecules21081078 (2016).

21. Klapper, M., Götze, S., Barnett, R., Willing, K. & Stallforth, P. Bacterial alkaloids prevent amoebal predation. *Angew. Chem. Int. Ed Engl.* **55**, 8944-8947 (2016).

22. Cane, D. E. Programming of erythromycin biosynthesis by a modular polyketide synthase. *J. Biol. Chem.* **285**, 27517-27523 (2010).

23. Dutta, S. *et al*. Structure of a modular polyketide synthase. *Nature* **510**, 512-517 (2014).

24. Hertweck, C., Luzhetskyy, A., Rebets, Y. & Bechthold, A. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat. Prod. Rep.* **24**, 162-190 (2007).

25. Bao, W., Wendt-Pienkowski, E. & Hutchinson, C. R. Reconstitution of the iterative type II polyketide synthase for tetracenomycin F2 biosynthesis. *Biochemistry* **37**, 8132-8138 (1998).

26. Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **7**, 285-295 (2003).

27. Robbins, T., Liu, Y. C., Cane, D. E. & Khosla, C. Structure and mechanism of assembly line polyketide synthases. *Curr. Opin. Struct. Biol.* **41**, 10-18 (2016).

28. Cotter, P. D., Ross, R. P. & Hill, C. Bacteriocins - a viable alternative to antibiotics? *Nat. Rev. Microbiol.* **11**, 95-105 (2013).

29. Strieker, M., Tanović, A. & Marahiel, M. A. Nonribosomal peptide synthetases: structures and dynamics. *Curr. Opin. Struct. Biol.* **20**, 234-240 (2010).

30. Felnagle, E. A. *et al*. Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Mol. Pharm.* **5**, 191-211 (2008).

31. Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D. H. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.* **7**, 78 (2007).

32. Garneau-Tsodikova, S., Dorrestein, P. C., Kelleher, N. L. & Walsh, C. T. Protein assembly line components in prodigiosin biosynthesis: characterization of PigA,G,H,I,J. *J. Am. Chem. Soc.* **128**, 12600-12601 (2006).

33. Marchal, E. *et al*. Synthesis and antimalarial activity of prodigiosenes. *Org. Biomol. Chem.* **12**, 4132-4142 (2014).

34. Kumar, A., Vishwakarma, H. S., Singh, J., Dwivedi, S. & Kumar, M. Microbial pigments: production and their applications in various industries. *IJPCBS* **5**, 203-212 (2015).

35. Pal, P. K. *et al*. Crop-ecology and nutritional variability influence growth and secondary metabolites of *Stevia rebaudiana Bertoni*. *BMC Plant Biol.* **15**, 67-015-0457-x (2015).

36. Torri, L. *et al*. Comparison of reduced sugar high quality chocolates sweetened with stevioside and crude stevia "green" extract. *J. Sci. Food Agric.* (2016).

37. Mousa, W. K. & Raizada, M. N. The diversity of anti-microbial secondary metabolites produced by fungal endophytes: an interdisciplinary perspective. *Front. Microbiol.* **4**, 65 (2013).

38. Onguéné, P. A. *et al*. The potential of anti-malarial compounds derived from African medicinal plants, part III: an in silico evaluation of drug metabolism and

pharmacokinetics profiling. *Org. Med. Chem. Lett.* **4**, 6-014-0006-x. Epub 2014 Sep 5 (2014).

39. Crotti, S. *et al*. Mass spectrometry in the pharmacokinetic studies of anticancer natural products. *Mass Spectrom Rev.* (2015).

40. Silver, L. L. Challenges of antibacterial discovery. *Clin. Microbiol. Rev.* **24**, 71-109 (2011).

41. Blair, J. M., Webber, M. A., Baylay, A. J., Ogbolu, D. O. & Piddock, L. J. Molecular mechanisms of antibiotic resistance. *Nat. Rev. Microbiol.* **13**, 42-51 (2015).

42. Nguyen, L. Antibiotic resistance mechanisms in *M. tuberculosis*: an update. *Arch. Toxicol.* **90**, 1585-1604 (2016).

43. Davies, J. & Davies, D. Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* **74**, 417-433 (2010).

44. Rosenblatt-Farrell, N. The landscape of antibiotic resistance. *Environ. Health Perspect.* **117**, A244-50 (2009).

45. Parida, S. K. *et al*. Totally drug-resistant tuberculosis and adjunct therapies. *J. Intern. Med.* **277**, 388-405 (2015).

46. Butler, M. S. & Buss, A. D. Natural products–the future scaffolds for novel antibiotics? *Biochem. Pharmacol.* **71**, 919-929 (2006).

47. Swier, L. J. *et al*. Structure-based design of potent small-molecule binders to the S-component of the ECF transporter for thiamine. *Chembiochem* **16**, 819-826 (2015).

48. Sybesma, W. *et al*. Bacteriophages as potential treatment for urinary tract infections. *Front. Microbiol.* **7**, 465 (2016).

49. Yosef, I., Manor, M. & Qimron, U. Counteracting selection for antibiotic-resistant bacteria. *Bacteriophage* **6**, e1096996-Mar (2016).

50. Bentley, S. D. *et al*. Complete genome sequence of the model Actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141-147 (2002).

51. Medema, M. H. *et al*. The sequence of a 1.8-Mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. *Genome Biol. Evol.* **2**, 212-224 (2010).

52. de Jong, A., van Hijum, S. A., Bijlsma, J. J., Kok, J. & Kuipers, O. P. BAGEL: a web-based bacteriocin genome mining tool. *Nucleic Acids Res.* **34**, W273-9 (2006).

53. Röttig, M. *et al*. NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity . *Nucleic Acids Res.* **39**, W362-7 (2011).

54. Medema, M. H. *et al*. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39**, W339-46 (2011).

55. Cano-Prieto, C. *et al*. Genome mining of *Streptomyces sp.* Tu 6176: characterization of the nataxazole biosynthesis pathway. *Chembiochem* **16**, 1461-1473 (2015).

56. Olano, C. *et al*. Activation and identification of five clusters for secondary metabolites in *Streptomyces albus* J1074. *Microb. Biotechnol.* **7**, 242-256 (2014).

57. Komatsu, M. *et al*. Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites. *ACS Synth. Biol.* **2**, 384-396 (2013).

58. Gómez-Escribano, J. P. & Bibb, M. J. *Streptomyces coelicolor* as an expression host for heterologous gene clusters. *Meth. Enzymol.* **517**, 279-300 (2012).

59. Chater, K. F. *Streptomyces* inside-out: a new perspective on the bacteria that provide us with antibiotics. *Philos. Trans. R. Soc. Lond. , B, Biol. Sci.* **361**, 761-768 (2006).

60. Liu, G., Chater, K. F., Chandra, G., Niu, G. & Tan, H. Molecular regulation of antibiotic biosynthesis in *Streptomyces*. *Microbiol. Mol. Biol. Rev.* **77**, 112-143 (2013).

61. Chandra, G. & Chater, K. F. Developmental biology of *Streptomyces* from the perspective of 100 actinobacterial genome sequences. *FEMS Microbiol. Rev.* **38**, 345-379 (2014).

62. Jacks, T. M., Schleim, K. D., Judith, F. R. & Miller, B. M. Cephamycin C treatment of induced enterotoxigenic colibacillosis (scours) in calves and piglets. *Antimicrob. Agents Chemother.* **18**, 397-402 (1980).

63. Leigh, D. A., Bradnock, K. & Marriner, J. M. Augmentin (amoxycillin and clavulanic acid) therapy in complicated infections due to beta-lactamase producing bacteria. *J. Antimicrob. Chemother.* **7**, 229-236 (1981).

64. Doroghazi, J. R. & Metcalf, W. W. Comparative genomics of Actinomycetes with a focus on natural product biosynthetic genes. *BMC Genomics* **14**, 611-2164-14-611 (2013).

65. Pettit, R. K. Culturability and secondary metabolite diversity of extreme microbes: expanding contribution of deep sea and deep-sea vent microbes to natural product discovery. *Mar. Biotechnol. (NY)* **13**, 1-11 (2011).

66. Tedesco, P. *et al*. Antimicrobial activity of monoramnholipids produced by bacterial strains isolated from the Ross sea (Antarctica). *Mar. Drugs* **14**, 10.3390/md14050083 (2016).

67. Zipperer, A. *et al*. Human commensals producing a novel antibiotic impair pathogen colonization. *Nature* **535**, 511-516 (2016).

68. Curragh, H. *et al*. Haloalkane degradation and assimilation by *Rhodococcus rhodochrous* NCIMB 13064. *Microbiology* **140 (Pt 6)**, 1433-1442 (1994).

69. McLeod, M. P. *et al*. The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 15582-15587 (2006).

70. Taylor, C. R. *et al*. Isolation of bacterial strains able to metabolize lignin from screening of environmental samples. *J. Appl. Microbiol.* **113**, 521-530 (2012).

71. van der Geize, R. & Dijkhuizen, L. Harnessing the catabolic diversity of Rhodococci for environmental and biotechnological applications. *Curr. Opin. Microbiol.* **7**, 255-261 (2004).

72. Iwatsuki, M. *et al*. Lariatins, antimycobacterial peptides produced by *Rhodococcus* sp. K01-B0171, have a lasso structure. *J. Am. Chem. Soc.* **128**, 7486-7491 (2006).

73. Kitagawa, W. & Tamura, T. A quinoline antibiotic from *Rhodococcus erythropolis* JCM 6824. *J. Antibiot. (Tokyo)* **61**, 680-682 (2008).

74. Chiba, H., Agematu, H., Sakai, K., Dobashi, K. & Yoshioka, T. Rhodopeptins, novel cyclic tetrapeptides with antifungal activities from *Rhodococcus sp*. III. Synthetic study of rhodopeptins. *J. Antibiot. (Tokyo)* **52**, 710-720 (1999).

75. Chu, J. *et al*. Discovery of MRSA active antibiotics using primary sequence from the human microbiome. *Nat. Chem. Biol.* **12**, 1004-1006 (2016).

76. Kunze, B., Höfle, G. & Reichenbach, H. The aurachins, new quinoline antibiotics from myxobacteria: production, physico-chemical and biological properties. *J. Antibiot. (Tokyo)* **40**, 258-265 (1987).

77. Borisova, R. B. Isolation of a *Rhodococcus* soil bacterium that produces a strong antibacterial compound. *Electronic Theses and Dissertations. East Tennessee State University* **Paper 1388** (2011).

78. Crabtree, M. N. Investigating potential bioactive compounds from *Rhodococcus* and their effects on MCF7 breast cancer cells. *Electronic Theses and Dissertations. East Tennessee State University* **Paper 2278** (2013).

79. Kurosawa, K. *et al*. Rhodostreptomycins, antibiotics biosynthesized following horizontal gene transfer from *Streptomyces padanus* to *Rhodococcus fascians*. *J. Am. Chem. Soc.* **130**, 1126-1127 (2008).

80. Takano, E., Chakraburtty, R., Nihira, T., Yamada, Y. & Bibb, M. J. A complex role for the gamma-butyrolactone SCB1 in regulating antibiotic production in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **41**, 1015-1028 (2001).

81. Pfefferle, C., Theobald, U., Gürtler, H. & Fiedler, H. Improved secondary metabolite production in the genus *Streptosporangium* by optimization of the fermentation conditions. *J. Biotechnol.* **80**, 135-142 (2000).

82. Ser, H. L. *et al*. Fermentation conditions that affect clavulanic acid production in *Streptomyces clavuligerus*: a systematic review. *Front. Microbiol.* **7**, 522 (2016).

83. Weber, T. & Kim, H. U. The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production. *Syst. Synth. Biol.* **1**, 69-79 (2016).

84. Altschul, S. F. *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402 (1997).

85. Medema, M. H., Takano, E. & Breitling, R. Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol. Biol. Evol.* **30**, 1218-1223 (2013).

86. Lincke, T., Behnken, S., Ishida, K., Roth, M. & Hertweck, C. Closthioamide: an unprecedented polythioamide antibiotic from the strictly anaerobic bacterium *Clostridium cellulolyticum*. *Angew. Chem. Int. Ed. Engl.* **49**, 2011-2013 (2010).

87. Rodriguez, G. M., Voskuil, M. I., Gold, B., Schoolnik, G. K. & Smith, I. *ideR*, an essential gene in *Mycobacterium tuberculosis*: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. *Infect. Immun.* **70**, 3371-3381 (2002).

88. Olano, C., Lombo, F., Méndez, C. & Salas, J. A. Improving production of bioactive secondary metabolites in Actinomycetes by metabolic engineering. *Metab. Eng.* **10**, 281-292 (2008).

89. de la Fuente, A., Lorenzana, L. M., Martín, J. F. & Liras, P. Mutants of *Streptomyces clavuligerus* with disruptions in different genes for clavulanic acid biosynthesis produce large amounts of holomycin: possible cross-regulation of two unrelated secondary metabolic pathways. *J. Bacteriol.* **184**, 6559-6565 (2002).

90. Ochi, K. & Hosaka, T. New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters. *Appl. Microbiol. Biotechnol.* **97**, 87-98 (2013).

91. Kim, H. S. & Park, Y. I. Isolation and identification of a novel microorganism producing the immunosuppressant tacrolimus. *J. Biosci. Bioeng.* **105**, 418-421 (2008).

92. Marmann, A., Aly, A. H., Lin, W., Wang, B. & Proksch, P. Co-cultivation–a powerful emerging tool for enhancing the chemical diversity of microorganismsm. *Mar. Drugs* **12**, 1043-1065 (2014).

93. Wietz, M., Månsson, M., Vynne, N. G. & Gram, L. in *Marine Microbiology: Bioactive Compounds and Biotechnological Applications* (ed Kim, S. K.) (Wiley-VCH, Weinheim, 2013).

94. Haste, N. M. *et al*. Activity of the streptogramin antibiotic etamycin against methicillin-resistant *Staphylococcus aureus*. *J. Antibiot. (Tokyo)* **63**, 219-224 (2010).

95. Garcia-Mendoza, C. Studies on the mode of action of etamycin (viridogrisein). *Biochim. Biophys. Acta* **97**, 394-396 (1965).

96. Gottelt, M., Kol, S., Gómez-Escribano, J. P., Bibb, M. & Takano, E. Deletion of a regulatory gene within the *cpk* gene cluster reveals novel antibacterial activity in *Streptomyces coelicolor* A3(2). *Microbiology* **156**, 2343-2353 (2010).

97. Gao, C., Hindra, Mulder, D., Yin, C. & Elliot, M. A. Crp is a global regulator of antibiotic production in *Streptomyces*. *MBio* **3**, 10.1128/mBio.00407-12 (2012).

98. Martínez-Costa, O. H. *et al*. A *relA/spoT* homologous gene from *Streptomyces coelicolor* A3(2) controls antibiotic biosynthetic genes. *J. Biol. Chem.* **271**, 10627-10634 (1996).

99. Li, S., Lu, C., Chang, X. & Shen, Y. Constitutive overexpression of asm18 increases the production and diversity of maytansinoids in *Actinosynnema pretiosum*. *Appl. Microbiol. Biotechnol.* **100**, 2641-2649 (2016).

100. Kato, J. Y., Funa, N., Watanabe, H., Ohnishi, Y. & Horinouchi, S. Biosynthesis of gamma-butyrolactone autoregulators that switch on secondary metabolism and morphological development in *Streptomyces*. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 2378-2383 (2007).

101. Takano, E. Gamma-butyrolactones: *Streptomyces* signalling molecules regulating antibiotic production and differentiation. *Curr. Opin. Microbiol.* **9**, 287-294 (2006).

102. Li, X. *et al*. ScbR- and ScbR2-mediated signal transduction networks coordinate complex physiological responses in *Streptomyces coelicolor*. *Sci. Rep.* **5**, 14831 (2015).

103. Gómez-Escribano, J. P. & Bibb, M. J. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechnol.* **4**, 207-215 (2011).

104. Gustafsson, C., Govindarajan, S. & Minshull, J. Codon bias and heterologous protein expression. *Trends Biotechnol.* **22**, 346-353 (2004).

105. Bekiesch, P., Basitta, P. & Apel, A. K. Challenges in the heterologous production of antibiotics in *Streptomyces*. *Arch. Pharm. (Weinheim)* **349**, 594-601 (2016).

106. Baltz, R. H. *Streptomyces* and *Saccharopolyspora* hosts for heterologous expression of secondary metabolite gene clusters. *J. Ind. Microbiol. Biotechnol.* **37**, 759-772 (2010).

107. Komatsu, M., Uchiyama, T., Omura, S., Cane, D. E. & Ikeda, H. Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2646-2651 (2010).

108. Stevens, D. C. *et al*. Alternative sigma factor over-expression enables heterologous expression of a type II polyketide biosynthetic pathway in *Escherichia coli*. *PLoS One* **8**, e64858 (2013).

109. Gressler, M., Hortschansky, P., Geib, E. & Brock, M. A new high-performance heterologous fungal expression system based on regulatory elements from the *Aspergillus terreus* terrein gene cluster. *Front. Microbiol.* **6**, 184 (2015).

110. Sorg, R. A., Kuipers, O. P. & Veening, J. W. Gene expression platform for synthetic biology in the human pathogen *Streptococcus pneumoniae*. *ACS Synth. Biol.* **4**, 228-239 (2015).

111. Breitling, R. & Takano, E. Synthetic biology of natural products. *Cold Spring Harb. Perspect. Biol.* **8**, 10.1101/cshperspect.a023994 (2016).

112. Takahashi, H. *et al*. Cloning and characterization of a *Streptomyces* single module type non-ribosomal peptide synthetase catalyzing a blue pigment synthesis. *J. Biol. Chem.* **282**, 9073-9081 (2007).

113. Kenig, M. & Reading, C. Holomycin and an antibiotic (MM 19290) related to tunicamycin, metabolites of *Streptomyces clavuligerus*. *J. Antibiot. (Tokyo)* **32**, 549-554 (1979).

114. Martín-Sánchez, L. Quorum sensing in *Streptomyces coelicolor*: Regulation of the SCB signalling system that controls the synthesis of antibiotics. (2016).

CHAPTER 2

Characterization of the *Streptomyces clavuligerus* indigoidine synthetase and its associated tautomerase

Ana Ceniceros[1*], Andriy Kovalchuk[1,6*], Harshwardhan Poddar[2], Kirstin Scherlach[3], Marnix Medema[1,4], Gerrit J. Poelarends[2], Mirjan Petrusma[1], Lubbert Dijkhuizen[1], and Eriko Takano[1,5]

*Shared first authors

1. Microbial Physiology, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands
2. Chemical and Pharmaceutical Biology, Groningen Research Institute of Pharmacy (GRIP), University of Groningen, Antonius Deusinglaan 1, 9713 AV Groningen, The Netherlands
3. Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute Beutenbergstr. 11a, 07745 Jena, Germany
4. Current address: Bioinformatics Group, Wageningen University, Droevendaalsesteeg 1, 6708 PB, Wageningen, The Netherlands
5. Current address: Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN United Kingdom
6. Current address: Department of Forest Sciences, University of Helsinki, P.O. Box 27, 00014, Helsinki, Finland.

Abstract

Indigoidine is a blue pigment that is produced by diverse bacteria and is thought to have anti-oxidative and antimicrobial properties. Indigoidine is known to be synthesized by a single-module nonribosomal peptide synthetase (IndC). This indigoidine synthetase is part of a larger gene cluster that usually also includes a gene encoding a 4-oxalocrotonate tautomerase-like protein (IndD), with unknown function. The end-product of this extended pathway is unknown.

We have identified a putative indigoidine gene cluster in the *Streptomyces clavuligerus* genome sequence, but it is cryptic and no indigoidine is produced under laboratory growth conditions. In *S. clavuligerus*, the putative *indC* and *indD* genes are fused. Heterologous expression of this *S. clavuligerus indC(D)* plus flanking genes in several *Streptomyces* strains did not result in blue pigment production. Only expression of *indC(D)* alone, controlled by a strong promoter, resulted in production of indigoidine in *Streptomyces coelicolor*, *Rhodococcus jostii*, and *Escherichia coli*. Interestingly, separate expression of the *S. clavuligerus* IndC protein yielded more indigoidine than IndC(D) expression. Also, a truncated *S. clavuligerus* gene, encoding only the IndD domain, was successfully expressed in *E. coli* and purified as an active enzyme, catalyzing a promiscuous Michael-type addition reaction. IndD was inactive, however, with a range of known tautomerase substrates.

The data also shows that *R. jostii* RHA1 and other members of the genus *Rhodococcus* may provide interesting alternative expression hosts for *Streptomyces* secondary metabolism gene clusters and individual genes.

Introduction

The blue pigment indigoidine was first described more than 100 years ago [1], and isolated for the first time in 1939 [2]. It took three decades until the structure of the pigment was elucidated, showing that it belongs to the group of 3,3'-bipyridyl pigments [3] (Figure 1). The capacity to produce indigoidine is widespread among bacteria belonging to distant systematic groups, e.g. Alpha, Beta- and Gamma-proteobacteria and Actinobacteria [3-7]. Indigoidine is believed to protect the producing bacteria against oxidative stress and to be involved in pathogenicity of plant pathogens [4]. It is also thought to have antimicrobial activity and to facilitate surface colonization [8]. Cude et al. [9] showed that the indigoidine-synthesizing bacterium *Phaeobacter sp.* strain Y4I inhibits surface colonization by *Vibrio fischeri;* strain Y41 wild-type itself colonizes surfaces better than a derived mutant strain deficient in indigoidine production. In view of its extracellular localization and extremely low solubility, the role of indigoidine in scavenging free radicals needs additional experimental proof. It has been hypothesized that the blue pigment indigoidine is not the end-product of the biosynthetic pathway, but that the compound is reduced to the soluble and colourless leucoindigoidine (Figure 1a). However, bioactivity of leucoindigoidine has not been proven due to the strong acidic conditions needed to chemically reduce indigoidine [9]. Thapa et al. [10] have suggested a completely different pathway, hypothesising that the final product is the blue soluble pigment indochrome B1 (Figure 1b).

Indigoidine was produced successfully by expressing the single module NRPS IndC from *Streptomyces lavandulae* ATCC 11924 in *E. coli*, together with a phosphopantetheinyl transferase (PPTase) from *Streptomyces verticillus* (Svp)[7]. PPTase catalyzes the post-translational attachment of the 4'-PP (4'-phosphopantetheine) moiety to the peptide carrier domain, thereby activating the NRPS [11]. As shown in Figure 1c, IndC is in most cases part of a gene cluster also encoding a pseudouridine-5'-phosphate

glycosidase (IndA), a phosphatase (IndB), a transport protein and a 4-oxalocrotonate tautomerase-like protein (IndD). In many cases, also a hypothetical protein (DUF4243) [12] and a phosphoribosyl transferase-type I domain protein are present. The role of these enzymes in indigoidine biosynthesis, or modification, yielding the final product(s) of the pathway, is as yet unknown. Also, the regulation of indigoidine synthesis is poorly understood and varies between strains. Involvement of γ-butyrolactone signalling molecules, SARP activator proteins or other regulators such as *pecS*, has been reported in some cases [4, 13, 14].

IndC has five domains; three of these are commonly found in NRPS modules, for adenylation, thiolation or peptide carrier protein and a thioesterase. The NRPS condensation domain is absent in these IndC enzymes. IndC also contains an oxidation and a second adenylation domain [7]. Interestingly, *S. clavuligerus* is one of the few strains reported in which IndC has an additional C-terminal domain with highest sequence identity to 4-oxalocrotonate tautomerase (4-OT) [15, 16]. The indigoidine synthetase from *S. clavuligerus* will be referred to as IndC(D) in this study. In most bacteria, this putative tautomerase is encoded by an independent *indD* tautomerase gene in the *ind* gene cluster. IndA is a predicted pseudouridine-5'-phosphate glycosidase and where tested, lacks activity on indigoidine. It has been suggested that IndA acts on the monomer of indigoidine before it dimerizes and adds a D-ribose-5-phosphate residue on carbon 4 (Figure 1b). IndB is then hypothesized to eliminate the phosphate from the ribose, after which the molecule dimerizes to form the water-soluble blue pigment indochrome B1 as final product [10]. DUF4243 is suggested to have a regulatory role, although its mode of action is unknown. The family of 4-OTs includes enzymes with different activities, including tautomerase, dehalogenase, isomerase and promiscuous C–C bond-forming aldol and Michael-type addition activities [17, 18]. In general, 4-OT enzymes possess a β-α-β structure that ends in a β-hairpin and are active as hexamers [17]. First, they form dimers by the interaction of β-sheets and α-helices of one monomer with

another in antiparallel direction. Hexamers are formed by association of 3 dimers through their C-terminal hairpins [17, 19]. These enzymes retain a large structural similarity, but may have a low amino acid identity (as low as 10%) [20]. After the cleavage of the initial methionine, 4-OTs contain an N-terminal proline (Pro-1) that is known to be essential for enzyme activity, acting as catalytic base [21]. Failure to eliminate the initial methionine results in a significant decrease in enzyme activity [22]. Three other essential residues have been described in these enzymes but they are not as conserved. Depending on the amino acid identity of these essential residues, these enzymes have similar activity or improved tautomerase, dehalogenase, isomerase or Michaelase activities [17, 23]. The role of the IndD enzyme in synthesis or modification of indigoidine is unknown. Cude et al. [9] hypothesise that IndD may convert the blue pigment indigoidine to its reduced state, leucoindigoidine, which is colourless and water-soluble (Figure 1a). Heterologous expression of the complete gene cluster of *Photorhabdus luminescens* in *E. coli* did not result in indigoidine production. Deletion of IndD from the gene cluster of *P. luminescens* in *E. coli* resulted in awakening of the gene cluster and a weak production of indigoidine [12]. This suggested that IndD may have a repressing and/or inhibitory effect on indigoidine production. A transporter is usually present in indigoidine gene clusters, but the type of transporter varies between bacterial species [4, 8, 24].

Previously we have shown that the *Streptomyces clavuligerus* genome encodes many secondary metabolite gene clusters [25]. Also, a putative indigoidine gene cluster has been identified in its genome sequence, but no indigoidine is produced by *S. clavuligerus* under laboratory growth conditions. In this study, we have analyzed the cryptic putative indigoidine gene cluster of *S. clavuligerus* in more detail.

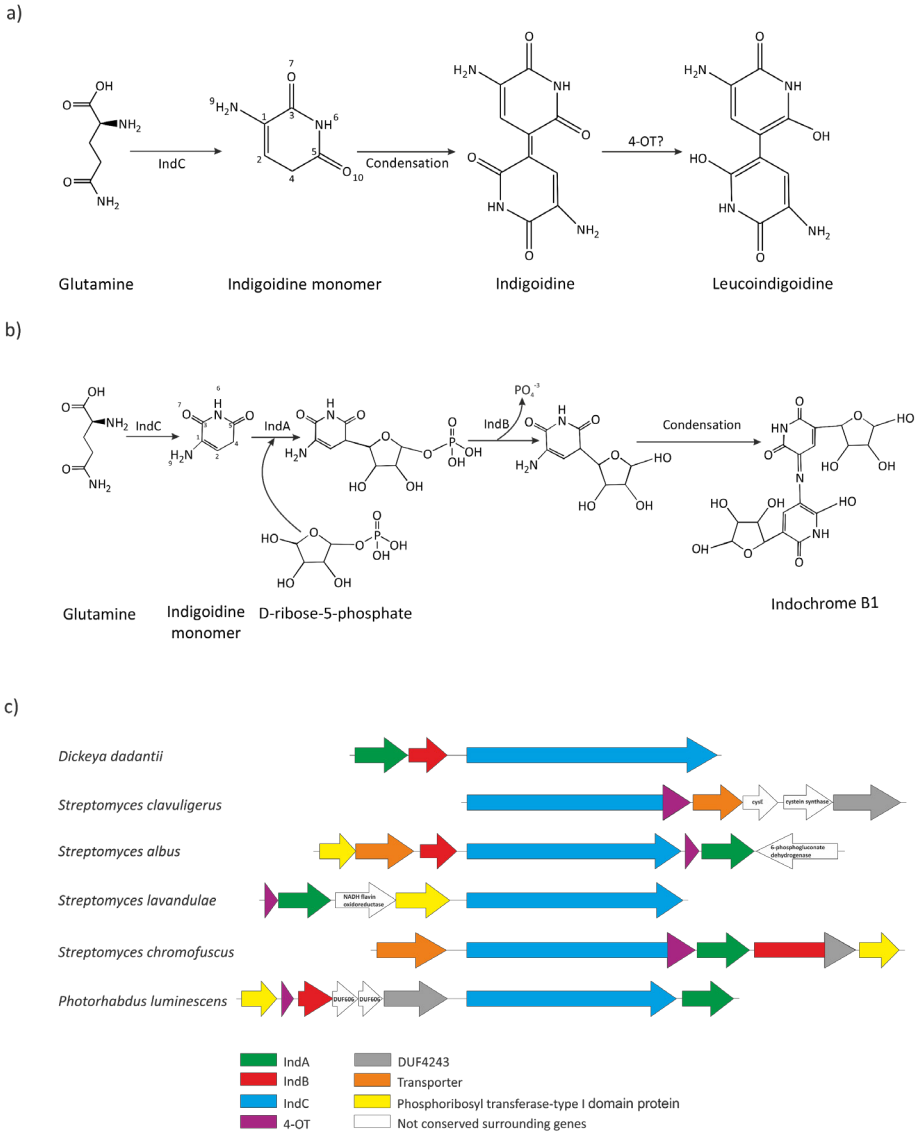


Figure 1. a) Predicted functions of IndC and IndD in indigoidine biosynthesis, adapted from Cude et al. [9]. b) Biosynthesis pathways predicted by Thapa et al. [10]. c) Representation of examples of indigoidine gene clusters, from *Dickeya dadantii, S. clavuligerus, Streptomyces albus, Streptomyces lavandulae, Streptomyces chromofuscus and P. luminescens* based on Reverchon et al. [4], Takahashi et al. [7], Medema et al. [25], Olano et al. [16], Yu et al. [26] and Thapa et al. [10]. IndA, IndB, IndC, a phosphoribosyl transferase-type I domain protein and the 4-OT-like protein (IndD) are the most conserved enzymes encoded in these clusters; only IndC is always present. DUF4243 is present in *S. clavuligerus*, *S. chromofuscus* (as a domain of IndB) and in *P. luminescens*.

Materials and methods

Strains, media and culture conditions

Escherichia coli DH5α was used for DNA cloning. The methylation-deficient strain *E. coli* ET12567 carrying the plasmid pUZ8002 was used to introduce plasmid DNA into *Streptomyces* cells by inter-generic conjugation. *Streptomyces coelicolor* M1146 [27], *S. coelicolor* M1152 [27], *Streptomyces lividans* WT and *Streptomyces avermitilis* WT, all lacking an indigoidine gene cluster, were used for the heterologous expression of the *S. clavuligerus* indigoidine gene cluster and indigoidine synthetase. *S. coelicolor* M1146, *R. jostii* RHA1 and *E. coli* BL21 (DE3) were used as hosts for heterologous expression of IndC(D) and IndC.

E. coli cells were grown in Luria-Bertani (LB) medium [28] supplemented when necessary with the following antibiotics: apramycin (Apra) (50 µg/ml), ampicillin (Amp) (50 µg/ml), kanamycin (Km) (50 µg/ml), chloramphenicol (Cm) (25 µg/ml). *S. clavuligerus* ATCC27064 was grown at 28^{o}C in TSB-YEME medium [29] for genomic DNA isolation. Soya flour mannitol agar (SFM), supplemented minimum media solid (SMMS), yeast extract-malt extract agar supplemented with 100 mM glutamate (YEME + 100 mM glutamate), R5 and trypton soya agar (TSA) were used to test the production of indigoidine by *Streptomyces* strains containing IndC(D), IndC or the cosmid harbouring the complete indigoidine gene cluster (Table 1). The compositions of these media were taken from Kieser *et al.* [29]. *R. jostii* RHA1 strains were grown at 30^{o}C on LB media.

Table 1. Strains used in this work.

Strain	Description	Reference
Escherichia coli DH5α		[30]
E. coli BL21(DE3)	Heterologous expression host	[31]
Streptomyces clavuligerus ATCC27064	Wild-type and parent strain	
Streptomyces coelicolor M1146	Heterologous expression host	[27]
S. coelicolor M1152	Heterologous expression host	[27]
Streptomyces avermitilis	Wild-type *S. avermitilis* strain; Heterologous expression host	
Streptomyces lividans WT	Wild-type *S. lividans* strain; Heterologous expression host	
Rhodococcus jostii RHA1	Wild-type *R. jostii* RHA1 strain; Heterologous expression host	[32]

DNA manipulations

Extraction of genomic DNA from *Streptomyces* strains was performed with the GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich). Amplification of DNA fragments by PCR was done using Phire Hot Start II PCR MasterMix (Thermo Fisher Scientific) and following the manufacturer instructions used with a longer initial denaturation of 1 min 30 s and a longer denaturation step in each cycle of 15 s. DNA fragments were purified from agarose gel or after PCR reaction with either illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) or Qiaex II Gel extraction kit (Qiagen). Restriction endonucleases and other DNA modifying enzymes were purchased from Thermo Fisher Scientific. DNA isolation and cloning procedures were performed according to [28]. DNA sequencing service was provided by GATC Biotech.

Expression of *S. clavuligerus* indigoidine synthetase and IndD in *E. coli*

For the heterologous expression in *E. coli*, IndC(D) and IndC versions of the gene SCLAV_p1474 were amplified by PCR from genomic DNA of *S. clavuligerus* ATCC27064 with primer pairs ind-EcoRI-forw/ind-XhoI-rev and ind-EcoRI-forw/ind-trunc-XhoI-rev (Table 2), respectively. Both

products were cloned into the EcoRI and XhoI sites of pET-28b (+), resulting in plasmids pTE491 and pTE492, respectively.

The *S. clavuligerus* phosphopantetheinyl transferase (PPTase)-encoding gene SCLAV_0102 was expressed together with IndC(D) and IndC to induce post-translational attachment of the 4′-PP (4′-phosphopantetheine) moiety to the carrier domain of the NRPS enzyme which is essential for its activity. SCLAV_0102 has 56% identity with the known PPTase from *S. verticillus* (Svp) which is known to be able to activate the indigoidine synthetase from *S. lavandulae* [7, 11]. The *S. clavuligerus* PPTase encoding gene was amplified using primers Scl0102-NdeI-forw/Scl0102-HindIII-rev and cloned in the NdeI and HindIII sites of pET-20b (+) resulting in plasmid pTE495. For co-expression with IndC(D) and IndC, Scl0102 was sub-cloned from pTE495 using BglII and PdiI into the BamHI and EcoRV sites from pACYC184 which has a compatible replication origin with pET vectors.

The tautomerase domain of SCLAV_p1474 (IndD) was amplified with primer pairs Tau-NdeI-forw/Tau-HindIII-rev and cloned in the NdeI and HindIII sites of pET-20b (+) resulting in plasmid pTE493.

Table 2. Primers used for DNA amplifications.

Primer	Sequence
ind-f	CAGCGAATTCCATGCTGATCTCCACAGAAG
ind-r2	GACTTCTAGACTCTCCCTCATGGATCATTG
ind-trunc-f	GTTCTCCTGCTCAGGTGTCTCCCCCGTTGC
ind-trunc-r	GGGAGACACCTGAGCAGGAGAACGAGAAACAATG
ind-EcoRI-forw	CAGTGAATTCGCACGCCACAGTGCTCAATCAATC
ind-XhoI-rev	GATCCTCGAGCTCTCCCTCATGGATCATTG
ind-trunc-XhoI-rev	GTATCTCGAGTCAGGTGTCTCCCCCGTTG
Tau-NdeI-forw	CAGTCATATGCCGCACATCAACATCAAG
Tau-HindIII-rev	CTAGAAGCTTGTAGTTCGGTGACTTGTGCAG
Scl0102-NdeI-forw	CTGACATATGATCGAGTCCCTGCTGC
Sc0102-HindIII-rev	CATGAAGCTTGGCTAAGCGGGGGACGGTC
Apra-Int Fw 2	TATATATGAGTAAACTTGGTCTGACAGTCAGGCGCCGGGGGTTC ATGTGCAGCTCCATCAG
Apra-Int Rv 2	AATAATATTGAAAAAGGAAGAGTAAGTTCCCGCCAGCCTGAGAT TCTTCGCCCTGCGAGAG

Expression of *S. clavuligerus* indigoidine synthetase in *S. coelicolor* M1146 and *R. jostii* RHA1

For the heterologous expression of the *S. clavuligerus* indigoidine synthetase IndC(D), the gene SCLAV_p1474 was amplified from genomic DNA of *S. clavuligerus* ATCC27064 by PCR with the primers ind-f and ind-r2 (Table 2). The product was cloned into the XbaI and EcoRI sites of pSET152-ermE* [33] resulting in plasmid pTE474. Plasmid pTE475, carrying only the *indC* part of the gene SCLAV_p1474 (truncated at base position 3849), was created by PCR amplification from the pTE474 template with the primers ind-trunc-f and ind-trunc-r. The full pTE474 thus was amplified, except the tautomerase region from IndC(D). Primers were designed with 23 base pairs complementary to each other to be able to circularize the PCR product resulting in pTE475, which was then cloned in *E. coli* DH5α. Plasmids pTE474 and pTE475 were introduced into *S. coelicolor* M1146 via conjugation with *E. coli* ET12567/pUZ8002 as described in Gust et al. [34] and in *R. jostii* RHA1 by electroporation following the protocol described in Arenskotter et al. [35]. *Streptomyces* ex-conjugants were selected and maintained on SFM plates supplemented

with apramycin (50 μg/ml). For the preparative production of blue pigment, *S. coelicolor* transformants were grown at 28°C in 450 ml of liquid YEME medium supplemented with 100 mM sodium glutamate in 1 L flasks for 4 days. Next, the culture was incubated for 3 more days at 25°C to obtain a higher yield of blue pigment. Purification of blue pigment was performed as described in Takahashi et al. [7]. NMR spectra of the purified compound from *S. coelicolor* M1146/IndC(D) were recorded on a Bruker Avance DRX 600 instrument. Spectra were referenced to the residual solvent signals. *R. jostii* strains were grown in LB at 30 °C for 2-4 days at 220 rpm or on LBA at 25 °C or 30 °C for 2-4 days.

Protein purification

Cells (0.5-L culture; 4.1 g) producing His-tagged IndD were suspended in 8 mL of buffer A (10 mM NaH_2PO_4 buffer, pH 7.3) and disrupted by sonication, after which unbroken cells and debris were removed by centrifugation (10,000 *g*, 45 min). The IndD protein was purified using the standard Ni-NTA chromatography protocol involving 3 wash steps with 25, 50 and 100 mM imidazole in buffer A. Retained IndD protein was eluted with 250 mM imidazole in buffer A. Fractions (1.5 mL) were analyzed by SDS-PAGE, and those that contained purified IndD protein were combined, concentrated using a vivaspin column, and the buffer was exchanged against buffer A using a pre-packed PD10 Sephadex G-25 gel filtration column. The Waddell method was used for estimating protein concentrations [36]. The purified protein was flash-frozen in liquid nitrogen and stored at -20°C until further use.

Expression of IndD in *R. jostii* RHA1

The IndD gene was subcloned from pTE493 and introduced in the NdeI and HindIII sites of pTip-QC1 [37] resulting in construct pTip-Tau. Electrotransformation was performed to introduce pTip-Tau into *R. jostii* IndC.

Expression of the complete indigoidine gene cluster from *S. clavuligerus* in *S. coelicolor* M1146, *S. coelicolor* M1152, *S. lividans* and *S. avermitilis*

One of the cosmids from a *S. clavuligerus* library [25] has a DNA fragment of 37286 nucleotides (position 1670506 to 1707791 in *S. clavuligerus* plasmid pSCL4, genes SCLAV_p1456-SCLAV_p1488) IndC(D), a downstream transporter and DUF4243) gene cluster predicted in Medema et al. [25] (SCLAV_p1474, SCLAV_p1475 and SCLAV_p1478). The hygromycin resistance gene from the cosmid was replaced by the Cφ31 attachment site and apramycin resistance gene from pSET152 using *E. coli* BW25113/pIJ790 as described in the ReDirect protocol [34]. The primers used for this strategy were Apra-Int Fw2 and Apra-Int Rv2 (Table 2). The DNA fragment was amplified using Phire Hot Start II PCR MasterMix from Thermo Fisher Scientific, using 60^{o}C as annealing temperature. The modified cosmid was introduced in the different *Streptomyces* strains mentioned in Table 1 through conjugation using the ReDirect protocol [34].

Table 3. Constructs used in this work.

Plasmid	Description
pTE474	pSET152-ermE* vector carrying full-length *S. clavuligerus indC(D)* gene
pTE475	pSET152-ermE* vector carrying *S. clavuligerus indC(D)* gene truncated at position 3850
pTE491	pET-28b (+) vector carrying full-length *S. clavuligerus indC(D)* gene
pTE492	pET-28b (+) vector carrying *S. clavuligerus indC*
pTE493	pET-20b (+) vector carrying the fragment of the *S. clavuligerus indC(D)* gene (positions 3850 - 4074) encoding IndD
pTE495	pET-20b (+) vector carrying the *S. clavuligerus* PPTase SCLAV_0102 gene
pTE496	pACYC184 vector carrying the BglII-PdiI fragment of plasmid pTE495
pTip-IndD	pTip-QC1 vector carrying the fragment of the *S. clavuligerus* IndD
Cosmid 4H02	Cosmid containing a 37286 nt fragment from *S. clavuligerus* plasmid pSCL4 (from 1670506-1707791) (DSM cosmid collection)
Cosmid 4H02A	Cosmid 4H02 after the exchange of the hygromycin resistance gene by integrase for Cφ31 attachment site and apramycin resistance gene

IndD structure prediction

IndD protein secondary structure prediction was performed with Phyre 2.0 using standard parameters [38].

Activity assays

Activity assays with IndD were performed as previously described [39, 40].

ESI-MS performed on the tautomerase domain

The mass of the IndD protein was determined using an LCQ electrospray mass spectrometer (Applied Biosystems, Foster City, CA), housed in the Mass Spectrometry Facility Core in the Groningen Research Institute of Pharmacy at the University of Groningen. Protein samples (25 μM) were prepared in 5 mM NH_4HCO_2 buffer, pH 7.5.

Results

Analysis of the indigoidine gene cluster of *S. clavuligerus*

Indigoidine gene clusters contain an indigoidine synthetase encoded by *indC* surrounded by genes that vary between species (Figure 1). Genome analysis of *S. clavuligerus* revealed the presence of the SCLAV_p1474 gene on plasmid pSCL4 [25] encoding a homologue of IndC(D) showing 57% amino acid identity with IndC from *D. dantii* and 59% with indC of *S. lavandulae*. In close proximity to *indC(D)* the SCLAV_p1478 gene encodes a protein with homology to DUF4243. A gene encoding a predicted protein from the drug/metabolite transporter family is located downstream of *indC(D)* (SCLAV_p1475) (Figure 1c). This transporter is from the same family as *pecM*, which together with *pecS* is involved in the regulation of the production of pectinase, cellulase and indigoidine in *D. dadantii* [41]. Also, an IndA domain encoding gene is present in *S. clavuligerus.* The *indA* gene however is not situated in close proximity to *indC(D)* but is located in the chromosome of *S. clavuligerus* (gene SCLAV_1172). No close homologue to the phosphatase IndB commonly present in these clusters was found in *S. clavuligerus.* The tautomerase part of the indigoidine gene cluster in *S. clavuligerus* is not present as a separate gene but instead this *indD* is a C-terminal domain of *indC* (Figure 1c) [15]. The amino acid sequence identity of 4-OTs is low but the structure is conserved [20]. The protein identity of the *S. clavuligerus* IndD is more similar to that of YdcE from *E. coli* (37% identity, 100% coverage) [42] than to the *Pseudomonas putida* 4-OT (29% identity, 72% coverage) [43]. Prediction of the secondary structure suggests that IndD of *S. clavuligerus* has a similar structure as YdcE from *E. coli* (Figure 2). This YdcE enzyme has been structurally characterized and is known to contain a C-terminal helical structure instead of the β-sheet hairpin more commonly found in 4-OT enzymes [44]. YdcE is also known to be active as a dimer instead of a hexamer, as described for most of the 4-OT enzymes. The essential catalytic residues in YdcE are Pro-1, Leu-11 (in some other IndD homologues substituted by Ile), Ser-39 and Trp-51; these residues are

also conserved in the *S. clavuligerus* IndD. The catalytic Pro-1 residue needs to be N-terminal to provide enzyme activity. In *S. clavuligerus* IndD this Pro-1 residue is fused to the IndC NRPS, however. To date, only 2 *indC* genes, including *indC(D)* of *S. clavuligerus*, are reported to encode a C-terminal tautomerase domain [15, 26]. BLAST analysis, however, revealed that at least 30 different strains belonging to Actinobacteria and Proteobacteria contain a homologue of the *indC(D)* gene of *S. clavuligerus*. These IndD 4-OT domains also possess all essential catalytic residues, and have a similar predicted C-terminal α-helix; in some cases, the last 2-3 amino acids are predicted to form a β-sheet (Figure 2).

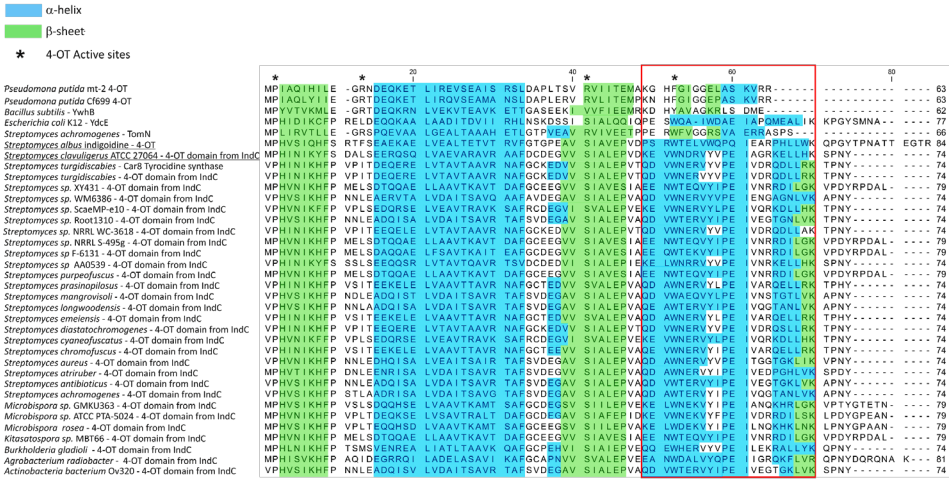


Figure 2. Multiple sequence alignment of the C-terminal IndD 4-OT domains of homologues of the indigoidine synthetase of *S. clavuligerus* IndC(D). The tautomerase domains were also aligned with 5 well-characterized members of the 4-OT family of tautomerases: *P. putida* mt-2 4-OT [43], its close homologue from *Pseudomonas* sp. CF600 [19], YwhB from *B. subtilis* [42], YdcE from *E. coli* [44] and TomN from *Streptomyces achromogenes* [15]. The catalytic and substrate binding residues from 4-OT are marked by black stars above the sequences. All YdcE essential residues are conserved in the different IndD 4-OT domains. Leu-11 sometimes is substituted by Ile. Secondary structure elements predicted to form a β-sheet are shown in green. Secondary structure elements predicted to form an α-helix are shown in blue. The red box highlights the residues involved in the formation of C-terminal β-sheets or α-helices, which differentiate 4-OT domains that form hexamers or dimers, respectively [44].

Heterologous expression of the indigoidine gene cluster of *S. clavuligerus*

The indigoidine gene cluster of *S. clavuligerus* is cryptic under laboratory growth conditions. Indigoidine has never been reported to be produced by *S. clavuligerus*. Heterologous expression of the indigoidine gene cluster was attempted by introducing a cosmid with a 37 Kb fragment of the *S. clavuligerus* genome including *indC(D)* and flanking genes into *S. coelicolor* M1146, *S. coelicolor* M1152, *S. avermitilis*, and *S. lividans.* All 4 strains became resistant to apramycin, confirming the presence of the cosmid. Production of blue pigment was tested by following growth on a range of solid media at 30°C and 25°C, but no production was observed.

Heterologous expression of *S. clavuligerus indC(D)* in *S. coelicolor* M1146, *R. jostii* RHA1 and *E. coli* Bl21(DE3)

To analyse the activity of IndC(D) of *S. clavuligerus*, its gene was heterologously expressed in *S. coelicolor* M1146, *R. jostii* RHA1 and *E. coli BL21*(DE3), resulting in blue pigment synthesis in all 3 strains. As expected, only in case of *E. coli* co-expression of the SCLAV_0102 phosphatase from *S. clavuligerus* with *indC(D)* was needed to activate the NRPS. Commonly used *E. coli* strains are not able to activate heterologous NRPSs [45]. *S. coelicolor* M1146 produced blue pigment in the absence of SCLAV_0102 (see also Takahashi et al.) [7]. In our experiments, *R. jostii* RHA1 also did not require addition of the SCLAV_0102 PPTase to produce blue pigment, which suggests that IndC(D) is activated by endogenous phosphatases; this may reflect the high content of NRPS clusters in the *R. jostii* RHA1 genome [46]. Blue pigment was successfully produced in all three strains. Only co-expression of the *S. clavuligerus* IndC(D) and PPTase in *E. coli* resulted in blue pigment production; no production of coloured substances was observed when either indigoidine synthetase or PPTase were expressed separately. Ex-conjugants from *S. coelicolor* M1146 carrying the *S. clavuligerus indC(D)* gene were able to produce the blue product on all solid media tested at 25°C. Addition of 100 mM

glutamate to the medium greatly enhanced the synthesis of the blue compound by *S. coelicolor* M1146 on all tested media. Glutamine synthase uses L-glutamate as substrate to form glutamine [47]. Glutamine is the substrate for IndC(D) indigoidine synthesis (see Figure 1). Addition of glutamate to the media did not stimulate production of blue pigment in *R. jostii* RHA1. The production of the pigment in *S. coelicolor* M1146 cells appeared to be temperature-dependent, as hardly any blue pigmentation was observed at 30°C (results not shown), whereas it is clearly visible when cells were grown at 25°C. *R. jostii* RHA1 cells grown at 30°C clearly produced the blue compound, although also here lower temperatures favoured its synthesis. The overall production in *E. coli* was significantly lower compared to *S. coelicolor* M1146 or *R. jostii* RHA1. The latter two expression hosts produced similar intensity of blue colour.

To analyse the role of IndD in the synthesis of indigoidine, *indC* was cloned separately and expressed in *S. coelicolor* M1146, *R. jostii* RHA1 and *E. coli* BL21(DE3). In the case of *E. coli*, the *indC* gene was co-expressed with the *S. clavuligerus* PPTase. With all three expression hosts this experiment resulted again in synthesis of blue pigment. Interestingly, in all cases the colour intensity of the pigment produced by IndC was higher compared to that of the strains carrying IndC(D) (Figure 3).

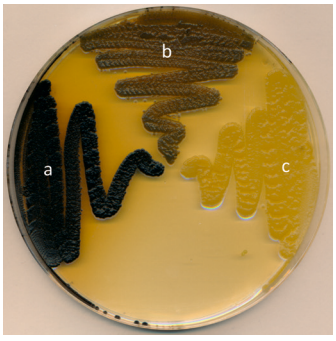


Figure 3. Blue pigment production by *S. coelicolor* M1146 transformants on solid DNA medium supplemented with 325 nM of glutamate incubated at RT for 10 days. a) IndC; b) IndC(D); c) *S. coelicolor* M1146 containing the cosmid 4H02A. Blue pigment production was highest in the strain expressing IndC. In c) the *S. coelicolor* M1146 strain carrying the complete gene cluster with *indC(D)* did not produce any blue pigment.

The blue pigment was purified from *S. coelicolor* M1146/IndC(D) and *S. coelicolor* M1146/IndC. The purified compound was insoluble in water and methanol but soluble in DMSO. The mass and structure of the compound from *S. coelicolor* M1146/IndC(D) was elucidated by a combination of MS and 1D and 2D NMR analyses.

The ^{1}H NMR spectrum of this compound displayed three signals corresponding to the protons of an amino- and an amide function as well as an olefinic substructure (11.30 (broad singlet; NH), 8.18 (singlet; CH), 6.46 (singlet; NH2) ppm). The ^{13}C NMR spectrum showed only five carbon signals further indicating the presence of a symmetrical structure (165.5 (C-2), 160.3 (C-6), 137.0 (C-3), 124.3 (C-5), 107.8 (C-4) ppm.). HMBC couplings of the amino protons with C-4, C-5 and C-6 as well as coupling of the olefinic proton H-4 with C-2, C-3, C-5 and C-6 established the structure of the bipyridyl pigment. This NMR data is in full accordance with published data about the structure of indigoidine (Figure 4) [7].

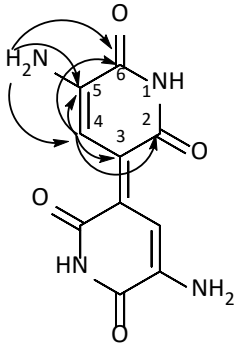


Figure 4. Key HMBC couplings detected in the NMR analysis of the blue pigment.

Coexpression of *S. clavuligerus* IndC and IndD in *R. jostii* RHA1

To further investigate the role of the IndD in the production or modification of indigoidine, a strain was constructed independently expressing the *S. clavuligerus* IndC and the IndD 4-OT-like domain. For this purpose, *indD* was expressed under a thiostrepton-inducible promoter, in the *R. jostii* RHA1/IndC strain. The blue colour of the medium used to grow *R. jostii* RHA1 strains expressing IndC(D), *R. jostii*/IndC and *R. jostii*/IndC + pTip-IndD was evaluated by eye and by

spectrophotometer at 595 nm. *R. jostii*/IndC + pTip-IndD produced blue pigment in amounts comparable to *R. jostii*/IndC, and clearly much higher than *R. jostii*/IndC(D). Separate expression of IndD thus did not restore the phenotype observed for the expression of the IndC(D).

In vitro analysis of the *S. clavuligerus* IndD

With the aim of studying the activity of IndD when separated from IndC(D) and therefore containing an N-terminal free Pro-1 residue, IndD protein (Figures 1, 2) was successfully overexpressed in *E. coli* BL21(DE3) and partially purified in a high yield (Figure 5).

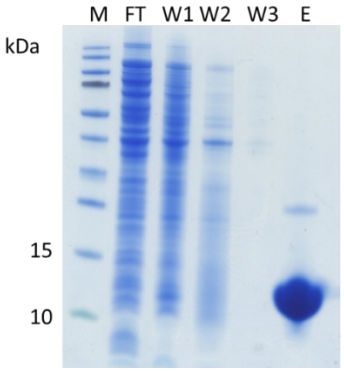


Figure 5. SDS-PAGE analysis of IndD after its expression and His-tag purification using *E. coli* BL21(DE3) as host. The protein is about 11 kDa in size. M: Molecular weight marker proteins. FT: Flow through. W1: Wash 1 (25 mM imidazole in buffer). W2: Wash 2 (50 mM imidazole in buffer). W3: Wash 3 (100 mM imidazole in buffer). E: Elution (250 mM imidazole in buffer).

Analysis by ESI-MS showed that the His-tagged IndD protein has a mass of 9805 ± 2 Da, consistent with the correct processing of the initial methionine, resulting in a protein with N-terminal proline residue, which is necessary for the proper activity of 4-OT enzymes [21]. Next, *in vitro* activity assays were performed. Indigoidine purified from *S. coelicolor* M1146 could not be used as a substrate, as the pigment precipitated in aqueous buffers. The common 4-OT substrates 2-hydroxymuconate and phenylenolpyruvate were tested, but no tautomerase activity was detected for IndD [39, 40]. IndD was active, however, with *trans*-β-nitrostyrene and acetaldehyde as substrates, in a promiscuous C–C bond-

forming Michael-type addition reaction, catalyzed by some of these tautomerase enzymes [40]. We also attempted to purify His-tagged IndC(D) and IndC proteins from *E. coli* BL21 (DE3) by Ni-NTA column chromatography, but the expression level was relatively low and both proteins were lost during the purification procedure.

Discussion

Indigoidine is a dark blue, insoluble pigment that can be synthesized by the single-module NRPS IndC. In the studied genome sequences, the *indC* gene is often found clustered with 5 additional genes suggesting a role for these genes in further modification of indigoidine [12]. *S. clavuligerus,* an industrially relevant bacterium with a high potential for the biosynthesis of diverse secondary metabolites, was reported to contain an indigoidine biosynthetic gene cluster [25] but production of indigoidine by this strain has never been observed, implying that this cluster is cryptic or silent. Interestingly, the *S. clavuligerus indC* gene is fused to an *indD* gene encoding a 4-OT domain. In most other strains containing an indigoidine gene cluster, IndD is encoded by a separate gene. First, we attempted the heterologous expression of a genomic fragment of *S. clavuligerus* containing *indC(D)* and its flanking genes in different *Streptomyces* hosts, but no blue pigment production or phenotypic difference was observed compared to the unmodified strain. One of the flanking genes of *indC* in *S. clavuligerus* is DUF4243. The homologous DUF4243 of *P. luminescens* has been suggested to act as a repressor in indigoidine biosynthesis [12]. Experimental data on the indigoidine cluster from *P. luminescens* heterologously expressed in *E. coli* showed that expression of *indC* alone was sufficient to produce indigoidine. When the complete cluster except for the *indA* gene was expressed in *E. coli*, no indigoidine production was observed. Indigoidine was produced again in this *E. coli* host strain after the further deletion of the DUF4243 or IndD encoding genes from this gene cluster. However, with both constructs a much lower intensity of blue pigment was observed, compared to the expression of IndC alone. Deletion of DUF4243 gene in the parent strain did not induce the production of indigoidine, implying that there is further regulation occurring [12]. Deletion of DUF4243 in the indigoidine biosynthesis cluster of *S. clavuligerus* in future work may provide us with more insights about the regulation of this cluster. Another explanation

for the lack of indigoidine synthesis in *Streptomyces* hosts strains is that the cluster is active but that indigoidine is not the end-product of the biosynthetic pathway. When we expressed the atypical *indC(D)* gene separately, under the control of the strong constitutive promoter *ermE**, the blue pigment was synthesized in all tested *Streptomyces*, *Rhodococcus* and *E. coli* hosts, for the first time showing that the fused IndC(D) protein from *S. clavuligerus* is functional and can synthesize indigoidine. With the *E. coli* host, the pigment was only produced when the *S. clavuligerus* PPTase was expressed together with IndC(D). These results concur with the data from Takahashi et al. [7], using IndC from *Streptomyces lavendulae* ATCC11924 and the PPTase *svp* from *Streptomyces verticillus.* It was previously shown by Yu et al. [26] that IndC from *S. chromofuscus* alone can produce indigoidine. This *indC* gene was afterwards reported to also contain a fused *indD* [16]. All IndC enzymes containing an IndD fusion thus may be capable of catalyzing the production of indigoidine.

To further analyse *indC(D)* of *S. clavuligerus* and the role of IndD in the synthesis of the blue product, *indD* was removed from the *indC* gene and this truncated *indC(D)* was expressed in the same strains mentioned above. Here it is important to note that IndD cannot be functional while fused to IndC: only after deletion of the IndD N-terminal methionine the resulting Pro-1 residue is free (i.e., N-terminal) for correct activity of IndD [21]. A post-translational modification may be needed to separate both enzymes and to allow a proper functioning of both. Compared to IndC(D), blue pigment synthesis from IndC was higher in all three strains. This may be due to better expression or activity levels of IndC than full length IndC(D). But the IndD domain consists of only 74 amino acids out of the 1357 in IndC(D), and it is situated at the C-terminus of this NRPS. It is therefore unlikely that IndD interferes with IndC expression. Perhaps IndD is interfering with the IndC enzyme activity, e.g. by producing an allosteric interference or by obstructing the correct folding of this NRPS, but such a mechanism would seem wasteful and biologically implausible.

Having the tautomerase fused to the NRPS would ensure that the level of expression of both enzymes is largely the same, which could explain why at least 30 bacterial strains have these two enzymes fused. It may also indicate that the tautomerase acts immediately after the formation of the product by IndC or even during its synthesis. In this case, the molecule might be modified before the condensation of the indigoidine monomers, thus preventing the formation of indigoidine. Having both genes fused would ensure the proximity of both enzymes for a faster reaction before the condensation of the molecule. The fusion of these genes might have also been the result of selection for inactivation of IndD by blocking the Pro-1 residue.

Our data shows that the separate *S. clavuligerus* IndD expressed in *E. coli* has a promiscuous enzymatic activity indicating correct folding of this domain (see below). Co-expression of IndC and IndD in *R. jostii* RHA1 was expected to revert the phenotype to that of expression of the complete *indC(D)* gene (synthesis of a reduced amount of blue pigment), but no difference was observed. The lack of activity could be due to a failed removal of the first methionine, thus lowering the activity of IndD. Another possibility is that any downregulatory effect of IndD on IndC activity is exerted via its direct fusion to each other by affecting the expression levels of the fused genes or the correct folding of the NRPS, and/or *in vivo* IndD activity reducing the amount of blue pigment. Unfortunately, no such activity of IndD with the blue pigment could be detected *in vitro*. Further studies of IndD should provide more insights in its regulatory and/or catalytic roles.

To analyse the enzymatic activity of the *S. clavuligerus* IndC(D) and IndC proteins, the enzymes were expressed in *E. coli* BL21 (DE3) but their purification was unsuccessful. IndC(D) and IndC expression levels were poor and these proteins were lost during the different His-Tag purification steps. However, expression and purification of the IndD domain alone was successful. Indigoidine could not be tested as a

substrate for IndD because of the insoluble nature of this compound. However, enzymatic activity was detected in a Michael-type addition reaction using *trans*-β-nitrostyrene and acetaldehyde as the substrates, presumably forming the corresponding γ-nitroaldehyde. This Michael-type addition reaction involves the formation of a nucleophilic enamine intermediate between Pro-1 and acetaldehyde, which reacts with the electrophilic nitroalkene. This reaction has been discovered as a promiscuous activity of 4-OT [40]. This result shows that the fused 4-OT domain of *S. clavuligerus* could have enzymatic activity as well, but further work is needed to identify the natural activity of IndD of *S. clavuligerus*.

The putative indigoidine biosynthesis gene cluster is cryptic in wild type *S. clavuligerus*, and heterologous expression of the gene cluster did not result in production of the blue pigment. Separate expression of the atypical indigoidine synthetase of *S. clavuligerus*, containing an extra C-terminal 4-OT domain, under the control of the *ermE** promoter was successful, and we confirmed indigoidine production. The data also shows that *R. jostii* RHA1 can be used to express *Streptomyces* genes resulting in synthesis of secondary metabolites. *Rhodococcus* strains grow relatively fast and are more easily to manipulate genetically [48] than *Streptomyces*, making them very interesting expression host strains that already find increasing use as industrial production strains [49]. The role of the extra 4-OT domain found in *S. clavuligerus* IndC(D) was also studied. Deletion of this domain resulted in a higher production of indigoidine. *In vitro* enzyme activity assays with IndD showed that the domain has enzymatic activity when separated from this NRPS. However, no activity was observed on typical tautomerase substrates. The biosynthesis pathway and the product(s) synthesized by this gene cluster thus remain to be determined. The study of the synthesis and regulation of easily detectable compounds like indigoidine will help us recognize the essential genes in the different clusters that may need to be activated in order to

trigger the expression of cryptic secondary metabolite gene clusters, leading ultimately to the discovery of new compounds.

Acknowledgments

This research was supported by the Dutch Technology Foundation (STW), which is part of the Netherlands Organization for Scientific Research (NWO) and partly funded by the Ministry of Economic Affairs (STW 10463), and by the University of Groningen. Harshwardhan Poddar and Gerrit J. Poelarends acknowledge funding by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement number 242293. We thank Stefano Donadio (Ktedogen, Italy) for valuable discussions.

References

1. Claessen, H. Über einen indigoblauen Farbstoff erzeugenden *Bacillus* aus Wasser. *Zbl. Bakt. , I. Abt. Orig.* **VII**, 13-17 (1890).

2. Elazari-Volcani, B. On *Pseudomonas indigofera* (Voges) Migula and its pigment. *Arch. Mikrobiol.* **10**, 343-358 (1939).

3. Kuhn, R., Starr, M. P., Kuhn, D. A., Bauer, H. & Knackmuss, H. J. Indigoidine and other bacterial pigments related to 3,3'-bipyridyl. *Arch. Mikrobiol.* **51**, 71-84 (1965).

4. Reverchon, S., Rouanet, C., Expert, D. & Nasser, W. Characterization of indigoidine biosynthetic genes in *Erwinia chrysanthemi* and role of this blue pigment in pathogenicity. *J. Bacteriol.* **184**, 654-665 (2002).

5. Novakova, R., Odnogova, Z., Kutas, P., Feckova, L. & Kormanec, J. Identification and characterization of an indigoidine-like gene for a blue pigment biosynthesis in *Streptomyces aureofaciens* CCM 3239. *Folia Microbiol. (Praha)* **55**, 119-125 (2010).

6. Starr, M. P., Cosens, G. & Knackmuss, H. J. Formation of the blue pigment indigoidine by phytopathogenic *Erwinia*. *Appl. Microbiol.* **14**, 870-872 (1966).

7. Takahashi, H. *et al*. Cloning and characterization of a *Streptomyces* single module type non-ribosomal peptide synthetase catalyzing a blue pigment synthesis. *J. Biol. Chem.* **282**, 9073-9081 (2007).

8. Cude, W. N. *et al*. *Phaeobacter* sp. strain Y4I utilizes two separate cell-to-cell communication systems to regulate production of the antimicrobial indigoidine. *Appl. Environ. Microbiol.* **81**, 1417-1425 (2015).

9. Cude, W. N. *et al*. Production of the antimicrobial secondary metabolite indigoidine contributes to competitive surface colonization by the marine roseobacter *Phaeobacter* sp. strain Y4I. *Appl. Environ. Microbiol.* **78**, 4771-4780 (2012).

10. Thapa, K., Oja, T. & Metsä-Ketelä, M. Molecular evolution of the bacterial pseudouridine-5'-phosphate glycosidase protein family. *FEBS J.* **281**, 4439-4449 (2014).

11. Sánchez, C., Du, L., Edwards, D. J., Toney, M. D. & Shen, B. Cloning and characterization of a phosphopantetheinyl transferase from *Streptomyces verticillus* ATCC15003, the producer of the hybrid peptide-polyketide antitumor drug bleomycin. *Chem. Biol.* **8**, 725-738 (2001).

12. Brachmann, A. O. *et al*. Triggering the production of the cryptic blue pigment indigoidine from *Photorhabdus luminescens*. *J. Biotechnol.* **157**, 96-99 (2012).

13. Kurniawan, Y. N. *et al*. Regulation of production of the blue pigment indigoidine by the pseudo gamma-butyrolactone receptor FarR2 in *Streptomyces lavendulae* FRI-5. *J. Biosci. Bioeng.* (2015).

14. Kurniawan, Y. N., Kitani, S., Maeda, A. & Nihira, T. Differential contributions of two SARP family regulatory genes to indigoidine biosynthesis in *Streptomyces lavendulae* FRI-5. *Appl. Microbiol. Biotechnol.* **98**, 9713-9721 (2014).

15. Burks, E. A. *et al*. Kinetic, crystallographic, and mechanistic characterization of TomN: elucidation of a function for a 4-oxalocrotonate tautomerase homologue in the tomaymycin biosynthetic pathway. *Biochemistry* **50**, 7600-7611 (2011).

16. Olano, C. *et al*. Activation and identification of five clusters for secondary metabolites in *Streptomyces albus* J1074. *Microb. Biotechnol.* **7**, 242-256 (2014).

17. Whitman, C. P. The 4-oxalocrotonate tautomerase family of enzymes: how nature makes new enzymes using a beta-alpha-beta structural motif. *Arch. Biochem. Biophys.* **402**, 1-13 (2002).

18. Zandvoort, E., Geertsema, E. M., Baas, B. J., Quax, W. J. & Poelarends, G. J. Bridging between organocatalysis and biocatalysis: asymmetric addition of acetaldehyde to beta-nitrostyrenes catalyzed by a promiscuous proline-based tautomerase. *Angew. Chem. Int. Ed. Engl.* **51**, 1240-1243 (2012).

19. Subramanya, H. S. *et al*. Enzymatic ketonization of 2-hydroxymuconate: specificity and mechanism investigated by the crystal structures of two isomerases. *Biochemistry* **35**, 792-802 (1996).

20. O'Brien, P. J. & Herschlag, D. Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* **6**, R91-R105 (1999).

21. Poddar, H., Rahimi, M., Geertsema, E. M., Thunnissen, A. M. & Poelarends, G. J. Evidence for the formation of an enamine species during aldol and Michael-type addition reactions promiscuously catalyzed by 4-oxalocrotonate tautomerase. *ChemBioChem* **16**, 738-741 (2015).

22. Baas, B. J., Zandvoort, E., Wasiel, A. A. & Poelarends, G. J. Demethionylation of Pro-1 variants of 4-oxalocrotonate tautomerase in *Escherichia coli* by co-expression with an engineered methionine aminopeptidase. *FEBS Open Bio* **4**, 651-658 (2014).

23. Poelarends, G. J., Veetil, V. P. & Whitman, C. P. The chemical versatility of the beta-alpha-beta fold: catalytic promiscuity and divergent evolution in the tautomerase superfamily. *Cell. Mol. Life Sci.* **65**, 3606-3618 (2008).

24. Rouanet, C. & Nasser, W. The PecM protein of the phytopathogenic bacterium *Erwinia chrysanthemi*, membrane topology and possible involvement in the efflux of the blue pigment indigoidine. *J. Mol. Microbiol. Biotechnol.* **3**, 309-318 (2001).

25. Medema, M. H. *et al*. The sequence of a 1.8-Mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. *Genome Biol. Evol.* **2**, 212-224 (2010).

26. Yu, D., Xu, F., Valiente, J., Wang, S. & Zhan, J. An indigoidine biosynthetic gene cluster from *Streptomyces chromofuscus* ATCC 49982 contains an unusual IndB homologue. *J. Ind. Microbiol. Biotechnol.* **40**, 159-168 (2013).

27. Gómez-Escribano, J. P. & Bibb, M. J. *Streptomyces coelicolor* as an expression host for heterologous gene clusters. *Meth. Enzymol.* **517**, 279-300 (2012).

28. Sambrook, J., Fritsch, E. F. & Maniatis, T. in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).

29. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. in *Practical Streptomyces Genetics* (John Innes Foundation, Norwich, 2000).

30. Woodcock, D. M. *et al*. Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucleic Acids Res.* **17**, 3469-3478 (1989).

31. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113-130 (1986).

32. McLeod, M. P. *et al*. The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 15582-15587 (2006).

33. Menges, R., Muth, G., Wohlleben, W. & Stegmann, E. The ABC transporter Tba of *Amycolatopsis balhimycina* is required for efficient export of the glycopeptide antibiotic balhimycin. *Applied Microbiology and Biotechnology* **77**, 125-134 (2007).

34. Gust, B., Kieser, T. & Chater, K. F. REDIRECT technology: PCR-targeting system in *Streptomyces coelicolor*. *Norwich, UK: The John Innes Foundation.* (2002).

35. Arensk̈otter, M., Baumeister, D., Kalscheuer, R. & Steinbüchel, A. Identification and application of plasmids suitable for transfer of foreign DNA to members of the genus *Gordonia*. *Appl. Environ. Microbiol.* **69**, 4971-4974 (2003).

36. Wolf, P. A critical reappraisal of Waddell's technique for ultraviolet spectrophotometric protein estimation. *Anal. Biochem.* **129**, 145-155 (1983).

37. Nakashima, N. & Tamura, T. Isolation and characterization of a rolling-circle-type plasmid from *Rhodococcus erythropolis* and application of the plasmid to multiple-recombinant-protein expression. *Appl. Environ. Microbiol.* **70**, 5557-5568 (2004).

38. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845-858 (2015).

39. Whitman, C. P., Aird, B. A., Gillespie, W. R. & Stolowich, N. J. Chemical and enzymic ketonization of 2-hydroxymuconate, a conjugated enol. *J. Am. Chem. Soc.* **113**, 3154-3162 (1991).

40. van der Meer, J. Y. *et al*. Using mutability landscapes of a promiscuous tautomerase to guide the engineering of enantioselective Michaelases. *Nat. Commun.* **7**, 10911 (2016).

41. Reverchon, S., Nasser, W. & Robert-Baudouy, J. *pecS*: a locus controlling pectinase, cellulase and blue pigment production in *Erwinia chrysanthemi*. *Mol. Microbiol.* **11**, 1127-1139 (1994).

42. Wang, S. C., Johnson, W. H.,Jr, Czerwinski, R. M., Stamps, S. L. & Whitman, C. P. Kinetic and stereochemical analysis of YwhB, a 4-oxalocrotonate tautomerase homologue in *Bacillus subtilis*: mechanistic implications for the YwhB- and 4-oxalocrotonate tautomerase-catalyzed reactions. *Biochemistry* **46**, 11919-11929 (2007).

43. Harayama, S., Rekik, M., Ngai, K. L. & Ornston, L. N. Physically associated enzymes produce and metabolize 2-hydroxy-2,4-dienoate, a chemically unstable intermediate formed in catechol metabolism via meta cleavage in *Pseudomonas putida*. *J. Bacteriol.* **171**, 6251-6258 (1989).

44. Almrud, J. J. *et al*. The crystal structure of YdcE, a 4-oxalocrotonate tautomerase homologue from *Escherichia coli*, confirms the structural basis for oligomer diversity. *Biochemistry* **41**, 12010-12024 (2002).

45. Doekel, S., Eppelmann, K. & Marahiel, M. A. Heterologous expression of nonribosomal peptide synthetases in *B. subtilis*: construction of a bi-functional *B. subtilis/E. coli* shuttle vector system. *FEMS Microbiol. Lett.* **216**, 185-191 (2002).

46. Doroghazi, J. R. & Metcalf, W. W. Comparative genomics of Actinomycetes with a focus on natural product biosynthetic genes. *BMC Genomics* **14**, 611-2164-14-611 (2013).

47. Xu, F., Gage, D. & Zhan, J. Efficient production of indigoidine in *Escherichia coli*. *J. Ind. Microbiol. Biotechnol.* **42**, 1149-1155 (2015).

48. van der Geize, R. *et al*. A novel method to generate unmarked gene deletions in the intracellular pathogen *Rhodococcus equi* using 5-fluorocytosine conditional lethality. *Nucleic Acids Res.* **36**, e151 (2008).

49. van der Geize, R. & Dijkhuizen, L. Harnessing the catabolic diversity of rhodococci for environmental and biotechnological applications. *Curr. Opin. Microbiol.* **7**, 255-261 (2004).

CHAPTER 3

Identification and characterization of the tunicamycin-like antibiotics (MM 19290) of *Streptomyces clavuligerus*

Ana Ceniceros[1], Mirjan Petrusma[1], Kirstin Scherlach[2], Wigard P. Kloosterman[4,6], Axel Trefzer[4,7], Danae Morales-Ángeles[1], Theo Verboom[3], Christian Hertweck[2], Christina Vandenbroucke-Grauls[3], Roel Bovenberg[4], Ulrike Müller[4], Lubbert Dijkhuizen[1], Eriko Takano[1,5]

1. Microbial Physiology, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Nijenborg 7, 9747 AG, Groningen, The Netherlands
2. Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute Beutenbergstr. 11a, 07745 Jena, Germany
3. Department of Medical Microbiology & Infection Control, VU University Medical Center, Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam
4. DSM Biotechnology Center, DSM Food Specialties B.V., 2613 AX Delft, The Netherlands
5. Current address: Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN United Kingdom
6. Current address: Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands
7. Current address: Life Technologies, GeneArt, Im Gewerbepark B35, 93059 Regensburg, Germany.

Abstract

Streptomyces clavuligerus is a Gram-positive soil bacterium known for the synthesis of clavulanic acid, a β-lactamase inhibitor that is one of its main secondary metabolites. Here we report the construction of a mutant strain that is fully blocked in the synthesis of clavulanic acid, designated *S. clavuligerus* strain Δ7. The mutations introduced are also known to block the synthesis of holomycin. Blocking these major biosynthetic pathways was hypothesized to result in the overexpression and /or activation of other secondary metabolite biosynthetic pathways. Indeed, *S. clavuligerus* strain Δ7 was found to produce bioactive compounds, identified as tunicamycin-like antibiotics (MM 19290). These compounds are produced in trace amounts only by the wild type strain *S. clavuligerus* ATCC 27064. MM 19290 molecules were purified in sufficient amounts from *S. clavuligerus* strain Δ7 to allow their analysis by MS and MS/MS. This resulted in putative identification and characterization of 6 different masses corresponding to derivatives of tunicamycin, corynetoxins and streptovirudines. These results contribute to our knowledge on the diversity of tunicamycin-related compounds synthesized by *S. clavuligerus* ATCC 27064.

Introduction

Members of the genus *Streptomyces* are known to produce many secondary metabolites, small molecules that are not essential for growth but may provide competitive advantages to the organisms that produce them. In many cases, they have important applications for humans, in medicine as antimicrobials or as drug components. About one third of currently used medicinal antimicrobial compounds are produced by members of the genus *Streptomyces* [1]. Secondary metabolites also have applications in food and textile industries as pigments or as sweeteners like stevioside [2] or in agriculture as herbicides [3]. Furthermore, some secondary metabolites are used as precursors in the synthesis of bioplastics [4]. In view of the industrial importance of secondary metabolites, it has become essential to understand how these compounds are produced and how their biosynthesis is regulated. Based on the rapid development of computational methods we have been able to identify many cryptic putative secondary metabolite clusters hidden in the genome sequences of a variety of microorganisms [5]. In most cases, it is unknown why these clusters are silent. The activation of these cryptic clusters is currently attempted aiming to identify novel and bioactive natural compounds. Several techniques have been used for this purpose. In some cases, modifying regulators, activators or repressors, both general or pathway specific, led to the production of a new compound [6]. Another strategy is to force the expression of biosynthetic genes in the targeted cluster [7-10] or to heterologously express these clusters in more efficient or more easily cultivated strains [11-15]. Another approach is to inactivate known pathways to free precursors that subsequently may be used in primary or secondary metabolism [16, 17]. Also, the further elucidation of partially known biosynthetic pathways is important, e.g. to elucidate which precursors are used and thereby understand how these pathways may be competing or interacting with each other.

S. clavuligerus strain ATCC 27064 is a known industrial producer of clavulanic acid and cephamycin C. Clavulanic acid is one of the main

secondary metabolites from *S. clavuligerus*. It is a β-lactamase inhibitor that is co-formulated with amoxicillin under the brand name Augmentin [18]. Cephamycin C is a β-lactam antibiotic with activity against Gram-positive and Gram-negative bacteria [19]. In addition, *S. clavuligerus* produces various other β-lactam compounds (clavams) [20], trace amounts of holomycin, a member of the pyrrothine class of antibiotics [21, 22], and trace amounts of a tunicamycin-like antibiotic (MM 19290), an inhibitor of N-linked glycosylation [21]. The genome of *S. clavuligerus* is comprized of a 6.8 Mb linear chromosome and an additional large linear plasmid of 1.9 Mb [23]. Bioinformatics analyses of this genome revealed the presence of 48 putative biosynthetic gene clusters [23].

This work reports the construction of a mutant strain of *S. clavuligerus*, carrying 7 deletions in the clavulanic acid biosynthesis pathway and thereby blocked in the production of clavulanic acid and clavams (*S. clavuligerus* strain Δ7). As shown by de la Fuente *et al.* [22], a blockage in the beginning of the clavulanic acid synthesis, before the formation of clavaminic acid, is known to impair the production of holomycin, a compound that does not share precursors with clavulanic acid. It is hypothesized that a clavulanic acid precursor, e.g. clavaminic acid or an earlier intermediate, acts as a regulator of holomycin synthesis [22]. *S. clavuligerus* strain Δ7 did not produce cephamycin C, clavams and holomycin. Instead tunicamycin-like molecules were detected. No structural information for (different) *S. clavuligerus* tunicamycin-like molecules MM 19290 was available yet. In this study, we have purified and characterized 6 different tunicamycin-like molecules produced by *S. clavuligerus* strain Δ7.

Materials and methods

Strains, media and growth conditions

Composition of the Tryptone Soya Broth (TSB), R5, and 2YT media used is described in Kieser *et al.* [24]. The 50% CM-3, CLA-SA1 (agar medium) and PAXSA4 (sporulation agar) are described in Koekman and Hans [25]. For antibiotic production, *S. clavuligerus* strain Δ7 was grown in a 20 ml preculture of TSB for 2 days at 25 °C. Then 500 ml of CLA-SA1 medium was inoculated with a 1:100 dilution from this preculture and incubated at 25 °C and 220 rpm for 5 days. Reporter strains listed in Table 1 and a collection of bacterial strains from the Free University Medical Center, both Gram-positive and Gram-negative species, were grown in Luria-Broth medium (Sigma).

Table 1. Bacterial strains used in this study.

Bacterial strain	Description/comments	Reference
***S. clavuligerus* ATCC 27064**	Clavulanic acid producer	ATCC collection
***Streptomyces clavuligerus* Δ7**	*Streptomyces clavuligerus* Δ*ceaS2,* Δ*bls2,* Δ*pah2,* Δ*cas2,* Δ*oat2,* Δ*oppA1*, Δ*claR* Gram-positive, producer strain	This work
***Bacillus subtilis* ATCC 6633**	Gram-positive, reporter strain	ATCC collection
***Escherichia coli* JM101**	Gram-negative, reporter strain	Sambrook and Russell [26]
***Kocuria rhizophila* ATCC 9341**	Gram-positive, reporter strain	ATCC collection
Serratia marcescens	Gram-negative, reporter strain	University of Groningen, soil isolate

Construction of *ceaS2-claR* deletion cassette

Seven genes from the clavulanic acid gene cluster were simultaneously deleted: *ceaS2, bls2, pah2, cas2, oat2, oppA1* and *claR*. A thiostrepton marker was inserted in between 3000 bp flanking regions of the targeted genes. The resulting deletion cassette was fused to pSET152 [24], while removing the integrase function. The deletion cassette was transferred to *S. clavuligerus* ATCC 27064 by conjugation. The following primer sequences (with the restriction site used for cloning shown underlined) were used to amplify the 3000 bp flanking regions: ceaS2-claR Fw-down HindIII-5'aagcttcgagcagactcgtggtgcgg, ceaS2-claR Rv-down XbaI 5'tctagagggaagaccgtctcgtcccg, ceaS2-claR Fw-up NsiI 5'atgcatgtcgccgaggagatacacgg, ceaS2-claR Rv-up DraI 5' tttaaatcgatacacgggacatgagc.

Conjugation protocol in *S. clavuligerus* ATCC 27064

Escherichia coli ET12567 with plasmid pUZ8002 [24] was transformed with the deletion cassette. One ml of an overnight culture of ET12567/pUZ8002 containing the deletion cassette grown on LB containing chloramphenicol (25 µg/ml), kanamycin (50 µg/ml) and apramycin (50 µg/ml) was inoculated into 100 ml fresh 2xYT with apramycin (25 µg/ml) and grown at 37°C to an OD600 of 0.4. Cells were washed once with 2xYT and resuspended in 10 ml of 2xYT. *S. clavuligerus* mycelium was grown for 2 days in liquid TSB medium. Then, 0.5 ml ET12567/pUZ8002 containing the deletion cassette was mixed with 1 ml of *S. clavuligerus* mycelium washed 3 times in 2xYT. The mix was briefly centrifuged and resuspended in 0.5 ml of 2xYT. Subsequently, 100 µl of dilutions 10^{-1} to 10^{-2} were plated on R5 agar without sucrose. Plates were incubated overnight and overlaid with 200 µl Milli Q water containing apramycin (5 mg/ml), nalidixic acid (2.5 mg/ml) and ampicillin (5 mg/ml). To allow for a second cross-over, colonies of the ex-conjugants obtained from the overlaid plates were re-streaked at least 5 times on PAXSA4 agar with thiostrepton selection, with at least once a re-streak from spores.

Colonies that had performed the double crossover were apramycin sensitive and thiostrepton resistant. All isolated potential deletion mutants were checked by PCR for the absence of the *ceaS2* and *oppA1* genes and the presence of the thiostrepton marker gene. This resulted in identification of strain Δ7, containing the thiostrepton marker gene and lacking the targeted clavulanic acid genes.

Clavulanic acid production and HPLC analysis

To test for clavulanic acid production, *S. clavuligerus* ATCC 27064 and *S. clavuligerus* strain Δ7 were cultured in CLA CM3 50% medium for 4 days. Then 80 µl culture supernatant was taken and treated with 20 µl clavulanic acid derivatisation reagent (8.25 g imidazole, 2 ml 5 M HCl, pH 6.8, in a total volume of 40 ml), a procedure adapted from [27]. Amounts of the clavulanic-imidazole derivative were determined by HPLC using a TSP P4000 Pump, an AS 3000 injector, and a UV1000 detector. Chromatographic separations were performed using a Nova-Pak C18 Column, 5 µm, 4.6 mm X 150 mm with the following protocol: 8 min at 100% 0.1 M K_2HPO_4, 6% MeOH, pH 3.2 (H_2SO_4) (A), then a gradient was applied from 100% A to 30% A and 70% MeOH (B) in 4 min and a second gradient from 30% A to 100% A in 2 min. Finally, 100% A was kept for 15 min. The clavulanic-imidazole derivative was detected at 311 nm.

Extraction procedure

S. clavuligerus strain Δ7 cultures were spun down by centrifugation. The supernatant was mixed with EtAc 1:1 (v/v). The EtAc phase was dried down, resuspended in methanol and filtered using a 3 kDa column (Amicon Ultra 0.5 Centrifugal filter devices) according to the manufacturing instructions. A final concentration of 5000x was used for further studies. Each extract was tested for bioactivity before continuing with the purification of metabolites.

Purification of the bioactive compound

The purification process followed in this work is summarized in Figure 1.

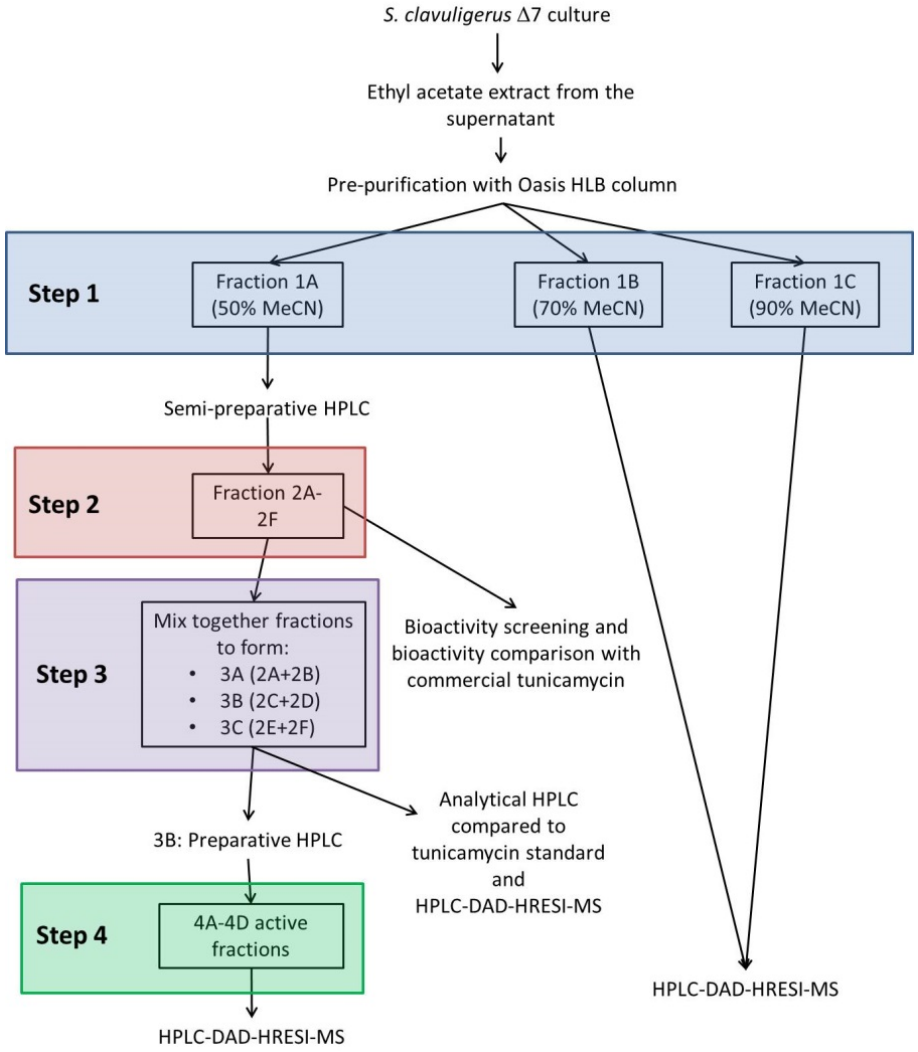


Figure 1. Purification steps performed with supernatants of *S. clavuligerus* strain Δ7 cultures. Three bioactive fractions were obtained from the pre-purification step (Fractions 1A-C). Fraction 1A was further purified through three HPLC steps before it was analyzed by high-performance liquid chromatography diode-array detector-high-resolution electro-spray ionization mass spectrometry (HPLC-DAD-HRESI-MS). Fractions 1B and 1C were directly analyzed by HPLC-DAD-HRESI-MS.

Solid phase extraction

The filtered crude extracts from the supernatants of the liquid cultures were dried and resuspended in Milli Q water. The extracts were loaded on an Oasis HLB column (6 ml). After a wash step with water, the sample was eluted using 10%, 30%, 50%, 70%, 90% and 100% of acetonitrile (MeCN). All the fractions were collected, dried and resuspended in methanol (keeping the original 5000x concentration) and tested in a bioactivity assay (see below).

Analytical HPLC

The bioactive fractions obtained from the Oasis HLB column (50%, 70% and 90% of acetonitrile fractions, fractions 1A-1C, Figure 1) were analyzed by an analytical HPLC to determine whether the activity observed was produced by a single or by multiple compounds. The HPLC equipment consisted of a Jasco Pu-980 Intelligent HPLC Pump and a Jasco LG-980-02 Ternary gradient. Chromatographic separations were performed using a Grace Platinum C18-EPS column (250 X 4.6 mm). Sample injections were performed by a Jasco 851-AS Intelligent sampler. The injection volume was 50 µl. The eluent was monitored at 210 nm by a Jasco UV-975 Intelligent UV/Vis detector. Acetonitrile was used as mobile phase (10% MeCN for 4 min, 10-100% MeCN in 46 min, 100% MeCN for 10 min). A flow of 1 ml/min was used in this purification and fractions of 2 ml were collected dried and resuspended in the initial volume to maintain the 5000x concentration.

Bioactivity tests

Bioactivity-guided purification

To follow the bioactive compounds in the different purification steps, bioactivity was determined by the agar diffusion method. Thus, 50 µl of solution were added to holes (9 mm in diameter) in NA agar plates seeded with 0.5 mL of a bacterial suspension. After incubation at 37 °C for 24 h the halo of the inhibition zone was measured.

Characterization of the bioactivity range of the bioactive fractions 2A-2F obtained from the semiprarative HPLC (see below)

Hand warm LB agar was inoculated with the reporter strain before pouring the plates. Then 5 µl of the samples dissolved in methanol HPLC grade were spotted over the agar and allowed to dry. The plates were incubated overnight at 30 °C and checked for the presence of an inhibition halo. Commercial tunicamycin (Sigma) was used as standard.

HPLC bioactivity-guided fractionation (Fraction 1A, Figure 1)

To identify the peak corresponding to the bioactive compound in the complex mixture of fraction A1, further purification was performed using a semipreparative HPLC consisting on a C12 Jupiter Phenomenex analytical column (250 X 4.6 mm, 4 µm), an Agilent G1311B 1260 Infinity Quaternary LC System connected to an Agilent G1329A (1200 Series) standard autosampler. The detection was performed with the Agilent G1365D (1200 Series) multiple wavelength detector. Acetonitrile containing 1% trifluoroacetic acid (TFA) was used as mobile phase. A gradient of acetonitrile was used for elution (20% of MeCN for 4 min, 20-25% in 2 min, going up to 35% of MeCN, 35-50% MeCN in 3 min, 50% MeCN for 4 min, 50-100% MeCN in 34 min; flow rate: 2.5 ml/min). In this way 5 ml fractions were collected, dried and resuspended in methanol to the initial volume to maintain the concentration. Bioactivity assays (see below) were performed with each fraction to follow the bioactive compound(s). Six bioactive fractions were obtained (fractions 2A-F, Figure 1).

Preparative HPLC

Extra purifications steps were performed using a Shimadzu LC-8a series preparative HPLC with diode-array detector (DAD) (column: Phenomenex Synergi Fusion RP18 250 x 20, gradient mode with MeCN/0.01% TFA (H_2O): 20% MeCN to 83% MeCN in 30 min, then 83% MeCN for 10 min; flow rate 10 mL/min or Nucleosil 100-5 RP18 250 x 10 mm column using a MeCN gradient: (40% MeCN to 100% MeCN in 30 min, then 100% MeCN

for 10 min; flow rate 5 mL/min)). Four bioactive fractions were obtained (4A-4D, Figure 1).

HPLC-DAD-HRESI-MS and MS/MS analysis (Fractions 3A-3C, 4A-4D, 1B and 1C, Figure 1)

HPLC-DAD-HRESI-MS: Exactive Orbitrap High Performance Benchtop LC-MS with an electrospray ion source and an Accela HPLC system (Thermo Fisher Scientific, Bremen). HPLC conditions: C18 column (Betasil C18 3 µm 150 x 2.1 mm) and gradient elution (MeCN/with 0.1 % HCOOH (v/v) 5/95 for 1 min, going up to 98/2 in 15 min, then 98/2 for another 3 min; flow rate 0.2 mL/min; injection volume: 3 µL). MS/MS measurements were performed in the all ion fragmentation mode (HCD 45 eV).

Results

Construction of *S. clavuligerus* strain Δ7

Clavulanic acid is one of the main secondary metabolites of *S. clavuligerus* strain ATCC 27064. We deleted the first 7 genes from the clavulanic acid biosynthetic gene cluster (Figure 2); the resulting mutant was designated *S. clavuligerus* strain Δ7. These deletions blocked the first steps of the clavulanic acid biosynthesis pathway (Figure 2), abolishing the production of both clavulanic acid and holomycin [22].

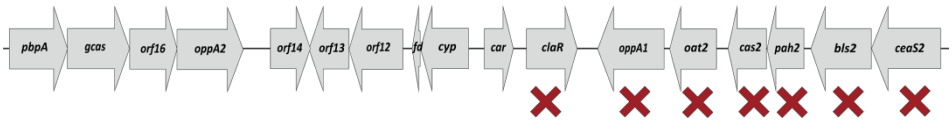


Figure 2. Clavulanic acid biosynthetic gene cluster of *S. clavuligerus*. Disrupted genes are shown with a cross underneath. Genes disrupted and the function of the respective proteins encoded: *ceaS2* ((2 carboxyethyl) arginine synthase), *bls2* (β-lactam synthetase), *pah2* (proclavaminate amidinohydrolase), *cas2 (*clavaminate synthase), *oat2* (ornithine *N*-acetyltransferase), *oppA1* (clavulanate oligopeptide binding protein (transporter)), *claR* (pathway-specific transcriptional activator).

Phenotypical analysis of *S. clavuligerus* strain Δ7

S. clavuligerus strain Δ7 was grown on different solid media to analyse growth, sporulation and antibiotic production. Sporulation of the mutant strain Δ7 was significantly slower than in the wild type strain, taking more than 2 weeks to sporulate in both SFM and CLA-SA1 media compared to approximately 1 week for the wild type strain.

S. clavuligerus strain ATCC 27064 and mutant Δ7 were grown in shake flasks with CM3 50% medium (for 4 days at 30°C). The clavulanic acid produced was analyzed by HPLC in culture supernatants. Under these growth conditions the ATCC 27064 strain produced around 100 mg/ml of clavulanic acid, but its synthesis was completely abolished in strain Δ7 (data not shown)

S. clavuligerus strain ATCC 27064 and strain Δ7 were tested for bioactivity against the reporter strains *B. subtilis* (Gram-positive) and *E. coli* (Gram-negative). Bioactivity (growth inhibition) was observed in all tested media against both reporter strains. Higher bioactivity was observed against *B. subtilis* than against *E. coli* (data not shown)

Antibiotic production by *S. clavuligerus* strain Δ7 was also tested in CLA-SA1 liquid media. After 5 days growth, the supernatant was extracted with ethyl acetate (EtAc) and tested for bioactivity by applying droplets of the extract on LB agar plates containing Gram-positive (*B. subtilis* and *K. rhizophila*) and Gram-negative (*S. marcescens* and *E. coli*) reporter strains. Growth inhibitory activity against all 4 strains was observed with these crude supernatant extracts. Again, highest activity was observed against *B. subtilis* (Supplementary information SI, Figure S1), therefore this reporter strain was chosen to follow the bioactive compounds during subsequent purification steps.

Purification and characterization of bioactive compounds

The bioactive compound(s) produced by *S. clavuligerus* strain Δ7 were subjected to several bioactivity-guided purification steps (Figure 1) (see supplementary information SI and Figure S2). Fractions 3A-C (Step 3, Figure 1) all showed bioactivity with 3B being the most active fraction. Sample 3B was further purified. Obtained bioactive fractions (Step 4, fractions 4A-4D, Figure 1) were analyzed by HPLC-DAD-HRESI-MS. Several compounds were detected with molecular masses ranging from 788 to 858 Da (Table 2). The most prominent detected ion mass, found in fraction 4B, was *m/z* 831 Da $[M+H]^+$ (Figure 3a, Table 2). From the high-resolution MS data, a molecular formula of $C_{38}H_{62}N_4O_{16}$ was deduced. Dereplication with commercially available databases suggested a potential identity as tunicamycin derivative. To corroborate this hypothesis, MS/MS analysis was performed on fraction 4B and on a tunicamycin standard. The molecular composition of all detected fragment ions deduced from HRESI-MS was in agreement with the

proposed structure of tunicamycins and with published data (Figure 3b). Identification of tunicamycins was corroborated by comparison of the HPLC retention times with an authentic tunicamycin reference.

To deduce whether active fractions 3A to 3C all contained tunicamycin-like compounds, they were further analyzed by HPLC-DAD-HRESI-MS. The analysis showed the presence of tunicamycin masses in all fractions, similar to the compounds detected in fraction 4B.

The detected tunicamycin-like compounds were found in fractions derived from fraction 1A (Figure 1). Fraction 1A displayed the highest bioactivity of the three fractions obtained from purification step 1. However, bioactivity was also detected in fractions 1B and 1C. Fractions 1B and 1C were directly analyzed by HPLC-DAD-HRESI-MS, which resulted in the detection of the same tunicamycin-like masses present in fraction 1A. In all analyzed fractions the mass of 830 Da was the most abundant, indicating that this is the predominant derivative produced by *S. clavuligerus* strain Δ7.

a)

b)

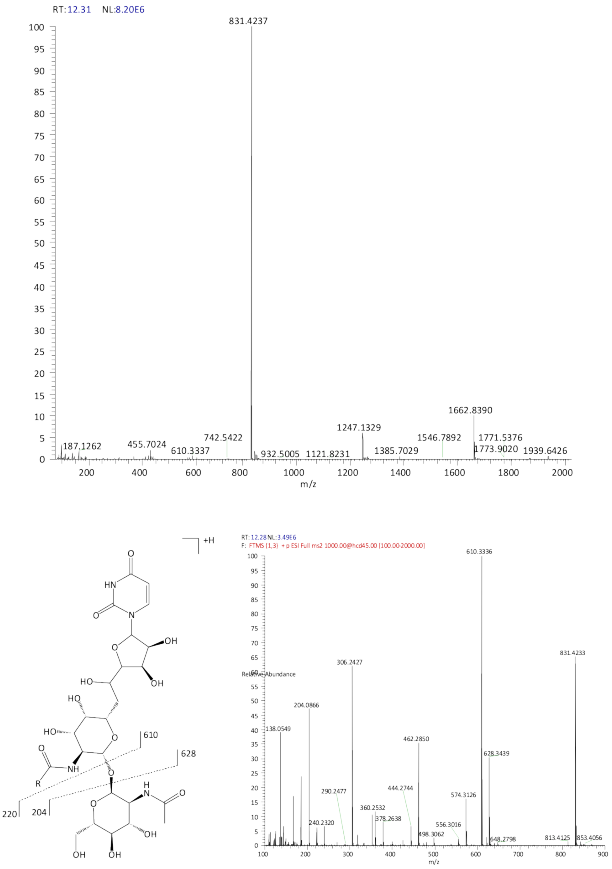


Figure 3. a) Mass spectrum of the main tunicamycin derivative detected in the LC-MS analysis of fractions 4A-4D (Figure 1) with a mass of 830 Da, which could correspond to tunicamycin IV and V or streptovirudin D2. b) Fractionation pattern of the tunicamycin core structure. Fragments of 204, 610 and 628 Da were observed in the MS/MS results obtained for the main detected mass (830 Da) in the samples, confirming that these molecules are tunicamycin derivatives.

S. clavuligerus ATCC 27064 is known to produce trace amounts of the tunicamycin-related antibiotic MM 19290 [21], but no structural information is available yet [28]. In our work with *S. clavuligerus* strain Δ7 we obtained sufficient amounts of MM 19290 to be able to identify for the first time the mass of 6 different molecules that are produced by *S.*

clavuligerus (Table 2), which were then confirmed to contain the tunicamycin core structure by MS/MS (Figure 3).

Table 2. Exact masses (Da) of the 6 putatively identified tunicamycin derivatives detected by LC-MS in the different *S. clavuligerus* bioactive fractions. The molecule with an m/z = 831.42, which corresponding to for instance tunicamycin IV, V or streptovirudin D_2, was the most abundant mass detected. In bold and underlined, molecules present in the tunicamycin standard used for comparison and that were also found in the experimental samples (Sigma Aldrich specifications of the product). Names of the structures were taken from Eckardt [29].

Monoisotopic mass	Detected ion	Calculated sum formula (ion)	Putative identities
830.42	*m/z* [M+H]+= 831.42	$C_{38}H_{63}N_4O_{16}$	Tunicamycin IV or V Streptovirudin D_2
788.37	*m/z* [M+H]+= 789.38	$C_{35}H_{57}N_4O_{16}$	Streptovirudin A_2
802.38	*m/z* [M+H]+= 803.39	$C_{36}H_{59}N_4O_{16}$	Streptovirudin B_{2a}= Tunicamycin I
816.40	*m/z* [M+H]+= 817.41	$C_{37}H_{61}N_4O_{16}$	Tunicamycin III or Tunicamycin II Streptovirudin C_2
844.43	*m/z* [M+H]+= 845.44	$C_{39}H_{65}N_4O_{16}$	Tunicamycin VII or VIII Corynetoxin U16i
858.45	*m/z* [M+H]+= 859.46	$C_{40}H_{67}N_4O_{16}$	Tunicamycin IX or X Corynetoxin U17A or U17i

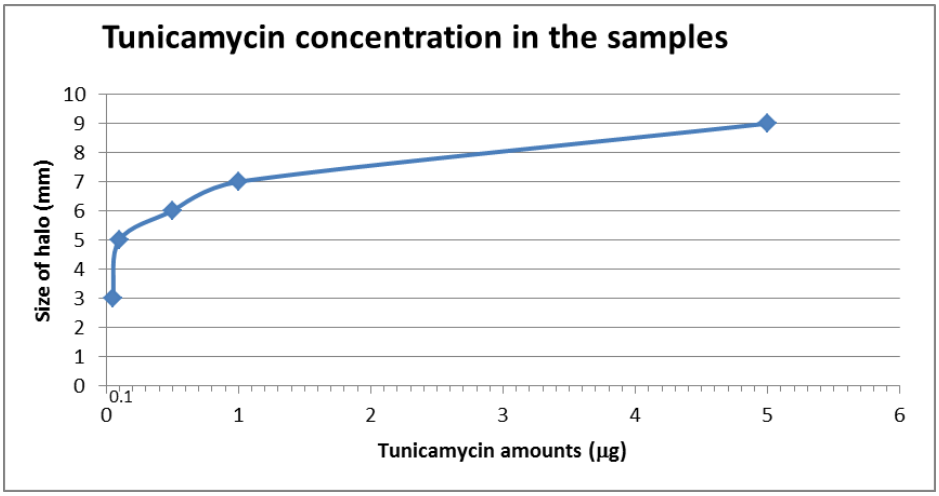
The EtAc extracts were screened by LC-MS for other known *S. clavuligerus* compounds, e.g. cephamycin C, clavams or holomycin. None of these compounds were found. Cephamycin C is a highly polar compound and therefore unlikely to dissolve in EtAc. Clavams share the first 6 steps of their biosynthesis pathway with clavulanic acid, corresponding to the proteins encoded by the *ceaS*2, *bls*2, *pah*2, *cas*2 (2 steps) and *oat*2 genes, which have been deleted in *S. clavuligerus* strain Δ7. The large linear plasmid of this strain carries a paralogous copy of these genes that are known to be activated when the strain is grown in soy media, where also the production of clavulanic acid and clavams is higher [20]. We grew *S. clavuligerus* strain Δ7 in media lacking soy, explaining the absence of clavams.

Characterization of the bioactivity of the purified compounds

In order to analyse the activity range of these compounds, fractions 1B-1C (Step 1, Figure 1) and 2A-2F (Step 2, Figure 1) were tested against different bacterial strains from the Free University Medical Center. No bioactivity was observed for Gram-negative strains (data not shown). Bioactivity was only observed against Gram-positive strains *B. subtilis*, *Streptococcus pyogenes*, *Streptococcus iniae*, *Staphylococcus citreus* and *Staphylococcus saprophyticus*, but not against *Listeria monocytogenes*, *Enterococcus faecium*, or *Lactobacillus casei* (Supplementary information SII Figure S3). Fractions 2C and 2D produced the largest inhibition halo.

Semi-quantification of the concentration of putative tunicamycin derivatives present in the samples

The bioactivity of the fractions 2A-2F (Step 2, Figure 1) was also compared to that of different concentrations of commercial tunicamycin using *B. subtilis*, *S. pyogenes*, *S. citreus*, *L. casei*, *L. monocytogenes*, *E. faecium*, *S. saprophyticus* and *S. iniae* (Supplementary information SII Figure S3). With all the strains tested, the size of the inhibition halo caused by the fractions with highest activity (2C and 2D, Figure 1) was similar to the one observed for 0.1 μg of tunicamycin (Figure 4). Since 5 μl of these samples were used for the assay, it can be calculated that fractions 2C and 2D contain a bioactive concentration equivalent to a tunicamycin concentration of about 20 μg/ml. The total amount of tunicamycin equivalents obtained in purification step 2 was calculated based on this estimate, resulting in approximately 430 μg. The initial amount of tunicamycin equivalents in the culture, however, was higher due to loss of the compound through the different purification steps. The diameter of the inhibition halo produced by fractions 2C and 2D was between 60-66% of the halo observed from the crude extract, which gives an approximation of the purification yield obtained.



Bioactive fractions	2A	2B	2C	2D	2E	2F
Size of halo (mm)	0	4	5	5	3	0

Figure 4. Semi-quantification of tunicamycins synthesized by *S. clavuligerus* strain Δ7. Comparison of the inhibition halo sizes observed with *B. subtilis* using the purified fractions 2A-F and different amounts of commercial tunicamycin (5 µg, 1 µg, 0.5 µg, 0.1 µg and 0.05 µg). The fractions that show the highest bioactivity (2C and 2D) produced a halo of similar size to the one observed when spotting 0.1 µg of commercial tunicamycin, indicating that the concentration in these samples is around 20 µg/ml. This test was also performed using *S. pyogenes, S. citreus, L. casei, L. monocytogenes, E. faecium, S. saprophyticus* and *S. iniae.* Fractions 2C and 2D produced a halo of similar size as 0.1 µg of tunicamycin with all the strains tested (Supplementary information SII Figure S3). For fraction numbers, see Figure 1.

Discussion and conclusions

S. clavuligerus ATCC 27064 is known for the production of clavulanic acid and cephamycin C. In addition to these two main secondary metabolites the strain also produces a few other bioactive compounds, but in lower amounts, or even trace amounts. Analysis of its genome sequence showed that, apart from the known compounds from this strain, it has the potential to produce around 40 extra compounds encoded in cryptic gene clusters [23]. This large potential for secondary metabolite biosynthesis makes the characterization of this strain of high interest.

Tunicamycins are a group of nucleoside antibiotics that were first described by Takatsuki *et al.* [30] as products of *Streptomyces lysosuperificus*. The first structures of tunicamycin molecules were elucidated in 1977 [31]. Tunicamycins consist of an uracil moiety, an *N*-acetylglucosamine moiety, a sugar containing 11 carbons plus a 2-aminodiadose residue, and an *N*-acyl chain which may be of different length. These molecules show bioactivity against Gram-positive bacteria, fungi, yeasts and viruses. Over 15 different tunicamycin molecules that differ in the length of their *N*-acyl chain have been described from different producers [32]. Subsequently, other highly related nucleoside antibiotics called streptovirudines and corynetoxins were described which display structural differences with tunicamycins. Following the classification of these closely related compounds [29], streptovirudins contain shorter fatty acid residues than tunicamycins, and sometimes they contain an anteiso methyl branching in the acyl chain [33]. In the case of series I of streptovirudines they contain dehydrouracil instead of uracil. Classes B_{2a} and C_2 have the same structure as tunicamycin molecules. Corynetoxins, however, contain slightly longer fatty acid chains than tunicamycins. These molecules also contain anteiso methyl branching but only in acyl chains with an odd number of carbon atoms [29, 33, 34]. The reason why tunicamycins are produced in such an array of diverse molecules is unknown. The various molecules, however, need different

concentrations to show the same activity [35]. Other bioactive compounds are also produced in arrays of different derivatives. In these cases, the different molecules also show a different degree of bioactivity [36, 37].

The physiological role of bioactive compounds is often unknown. In some cases, the production of antibiotics is too low to reach a lethal level in the environment, which suggests that its function in the producer strain is different. One hypothesis is that they act as signalling molecules [38]. This could be the case for tunicamycins in *S. clavuligerus* ATCC 27064 where they are produced in trace amounts [21]. Their different configurations may modulate the strength of the response that they regulate.

The tunicamycin gene cluster from *S. clavuligerus* was simultaneously first identified by Wyzsynsky *et al.* [28] and by Karki *et al.* [39] (Figure 5a). The putative biosynthetic pathway predicted by Wyzsynsky and collaborators for tunicamycin is shown in Figure 5b, with the various enzyme steps encoded by the genes in the biosynthetic cluster. No mutagenesis work has been previously done on tunicamycin biosynthesis in *S. clavuligerus*.

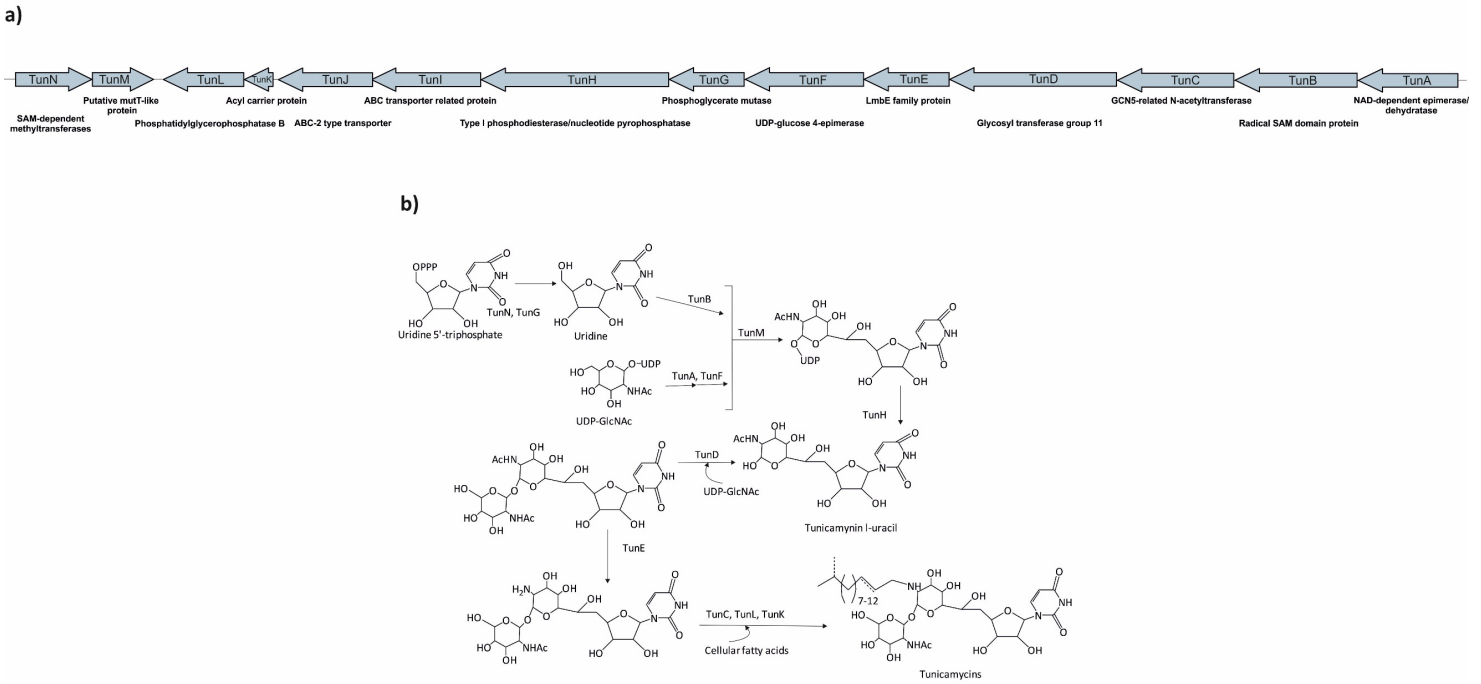


Figure 5. Biosynthetic gene cluster (a) and putative biosynthetic pathway (b) of tunicamycins in *S. clavuligerus*, based on Wyzsynsky *et al.* [28].

To gain further insight in the production of secondary metabolites by *S. clavuligerus* ATCC 27064, we constructed a mutant strain blocked in the biosynthetic pathway of clavulanic acid by deleting seven genes (resulting in *S. clavuligerus* strain Δ7). It is believed that an intermediate of the clavulanic acid biosynthesis pathway activates the production of holomycin [22]. When the pathway is interrupted in the first steps, this intermediate is not formed and holomycin is not produced. *S. clavuligerus* strain Δ7 contains mutations mainly in early steps of the clavulanic acid biosynthesis pathway, but also in *claR.* Single deletion of this regulatory gene led to overproduction of holomycin [22], but we were not able to detect holomycin in the strain Δ7 samples. This corroborates the hypothesis that an early intermediate in the synthesis of clavulanic acid induces the production of holomycin. *S. clavuligerus* strain Δ7 tested positive for production of antimicrobial compounds on solid and liquid media, active against *E. coli* and *B. subtilis*. Bioactive compounds were purified using several HPLC steps (Figure 1). We succeeded in producing and purifying sufficient amounts of these bioactive compounds from *S. clavuligerus* strain Δ7 to be able to structurally identify and characterize 6 different tunicamycin derivatives. Tunicamycins are produced in trace amounts by *S. clavuligerus* ATCC 27064 [21] and until now no structural information was available [28]. The masses of the 6 structurally different tunicamycin-like molecules identified as products of mutant strain Δ7 correspond to the masses of both tunicamycins and streptovirudines [29]. Two of the molecules synthesized by strain Δ7 also have the same exact mass as various corynetoxins (Table 2). The mass 830.42 Da corresponding to the tunicamycin derivative IV, V or streptovirudin derivative D2 was the predominant molecule produced by this strain. Due to the high structural similarities between tunicamycins, streptovirudines and corynetoxins, some of the homologues of tunicamycin detected by LC-MS in our work in fact may represent more than one compound. For a more precise identification of the compounds produced by *S. clavuligerus* strain Δ7, NMR analyses would be required. Besides the tunicamycins-like

compounds, no other bioactive compounds from *S. clavuligerus* were detected in any of the bioactive extracts.

Tunicamycins are compounds with a wide variety of uses. Their capacity to inhibit *N*-glycosylation has many uses in medicine. They lower the infectivity and productivity of different viruses through the modification of the glycosylated proteins of the envelope [40, 41], modify cell surface proteins in bacteria, which in *Streptococcus pneumoniae* affects the glycosylation of adhesins, reducing the capacity of these organisms to adhere to different tissues [42]. Moreover, they can be used to treat tumours by modifying protein glycosylation in tumoural cells which in some cases reduces the invasive properties of the cells and therefore the metastasis [43, 44].

In this study, we were able to purify enough MM 19290 tunicamycin-like antibiotics from mutant strain Δ7 of *S. clavuligerus* strain ATCC 27064 to perform a partial structural analysis that allowed identification of 6 different masses corresponding to tunicamycin-like molecules. The results add to a better understanding of the bioactive potential of the industrially relevant strain *S. clavuligerus*.

Acknowledgments

This research was supported by the Dutch Technology Foundation (STW), which is part of the Netherlands Organization for Scientific Research (NWO) and partly funded by the Ministry of Economic Affairs (STW 10463), and by the University of Groningen. We thank Stefano Donadio (Ktedogen, Italy) for valuable discussions and Andrea Perner (HKI Jena) for technical support.

Supplementary information SI

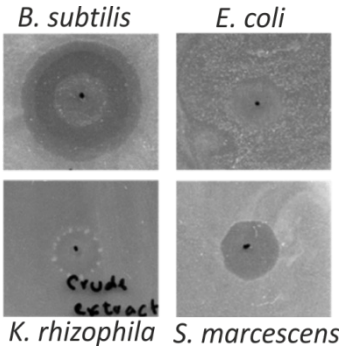


Figure S1. Bioactivity assays performed using the ethyl acetate supernatant extract from *S. clavuligerus* strain Δ7 grown for 5 days in CLA-SA1. Crude extract was tested for bioactivity against *B. subtilis*, *K. rhizophila, E. coli* and *S. marcescens.* Bioactivity was most clear using *B. subtilis* as reporter strain.

Purification steps (Figure 1)

First, the EtAc extract was loaded on a reverse phase SPE-Oasis column. Fractions were eluted using different concentrations of acetonitrile:water and tested for bioactivity against *B. subtilis*. Bioactive fractions eluted at 50% acetonitrile (MeCN) (Fraction 1A), 70% MeCN and 90% MeCN (Fractions 1B and 1C). Highest activity was observed in fraction 1A. Only the aqueous phase was loaded into the column. A water-insoluble residue remained from the EtAc extract and was also tested for bioactivity. A very faint halo was observed, indicating that most of the bioactive compound had dissolved in the aqueous phase.

HPLC analysis was done on fractions 1A to 1C obtained from the SPE-Oasis column (Figure S2). The obtained fractions were tested for bioactivity. From fraction 1A three bioactive fractions were obtained (Figure S2). Bioactive fractions from fractions 1B and 1C eluted at the same acetonitrile concentration (85-95%), suggesting that it is the same compound. Further purification was necessary for fraction 1A due to the high complexity of the sample.

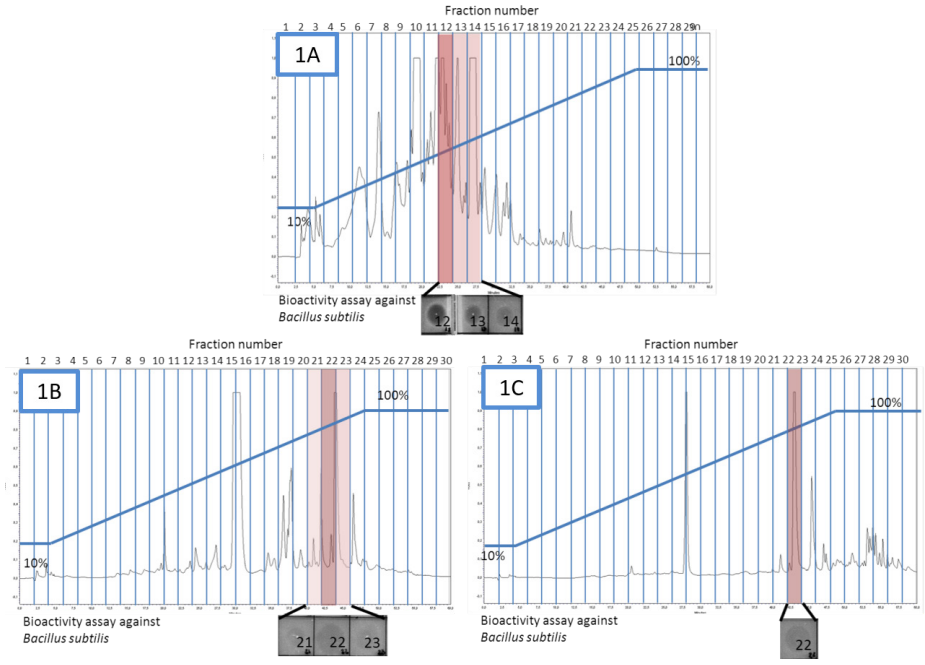


Figure S2. Screening of the different active Oasis HLB fractions by analytical HPLC, equipped with a Grace Platinum C18-EPS column, and bioactivity evaluation of the fractions. a) Fraction 1A (Figure 1) eluted at 50% MeCN from the Oasis HLB column. After the analytical HPLC, fractions 12-15 showed bioactivity. b) Fraction 1B (Figure 1) eluted at 70% MeCN fraction from Oasis HLB column. After the analytical HPLC, bioactivity was observed in fractions 21 to 23. c) Fraction 1C (Figure 1) eluted at 90% MeCN from Oasis HLB column. After the analytical HPLC, bioactivity was observed in fraction 22. The bioactive fractions against *B. subtilis* are shown below each chromatogram. The vertical lines indicate the fractions collected. The red areas indicate the fractions that show bioactivity. The blue line indicates the gradient of acetonitrile during the program.

Fraction 1A was subjected to semipreparative HPLC. The resulting fractions were tested for bioactivity against *B. subtilis* by agar diffusion assay. Six bioactive fractions were obtained from this analysis (Fraction 2A-2F). Highest activity was observed for fractions 2C and 2D. Fractions 2A and 2B, 2C and 2D and 2E and 2F were mixed to form fraction 3A,

fraction 3B and fraction 3C (Figure 1). Fraction 3B was further purified by preparative HPLC.

Supplementary information SII

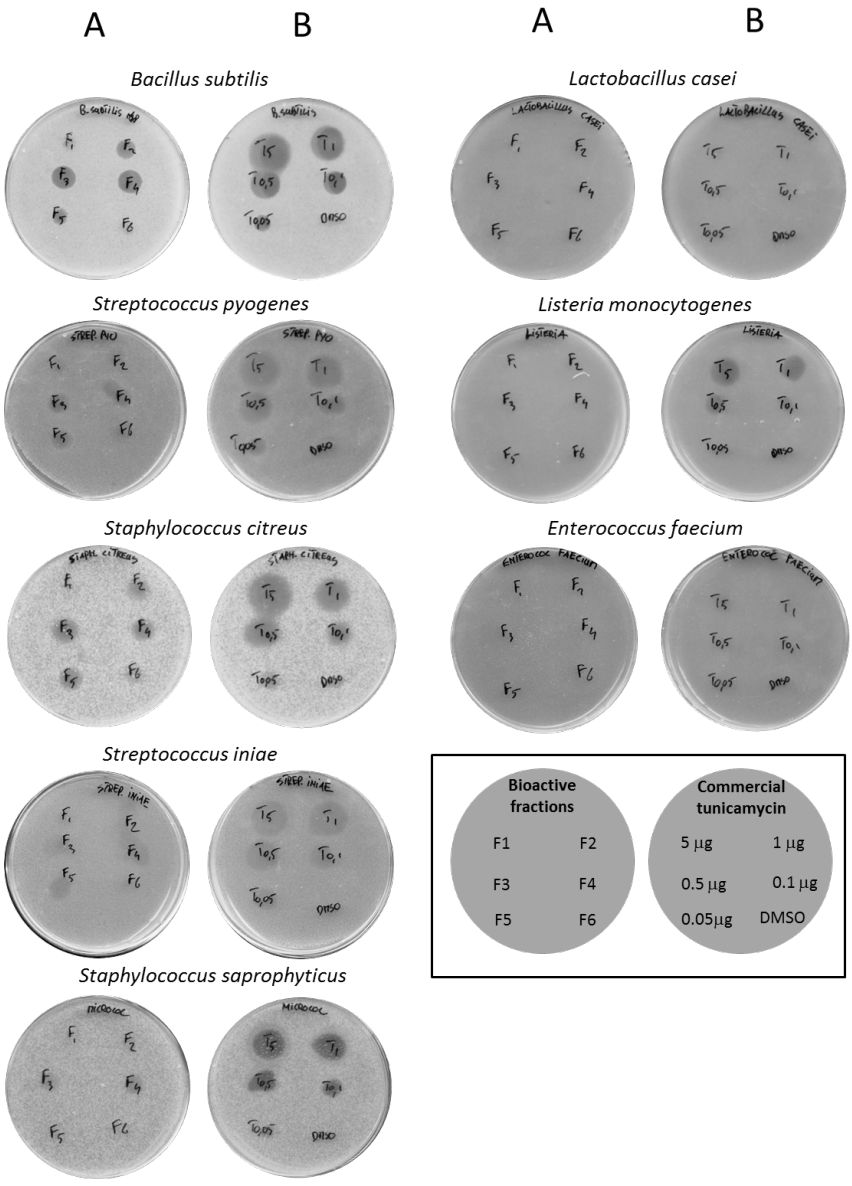


Figure S3. Comparison of the bioactivity observed for the purified fractions (column A) and different amounts of commercial tunicamycin (5 µg, 1 µg, 0.5 µg, 0.1 µg and 0.05 µg; column B). The reporter strains used were *B. subtilis, Streptococcus pyogenes, S. citreus, S. iniae, S. saprophyticus, L. casei, L. monocytogenes* and *E. faecium.* 2A-2F: fractions 2A to 2F (Figure 1).

References

1. Chater, K. F. *Streptomyces* inside-out: a new perspective on the bacteria that provide us with antibiotics. *Philos. Trans. R. Soc. Lond. , B, Biol. Sci.* **361**, 761-768 (2006).

2. Pal, P. K. *et al*. Crop-ecology and nutritional variability influence growth and secondary metabolites of *Stevia rebaudiana Bertoni*. *BMC Plant Biol.* **15**, 67-015-0457-x (2015).

3. Sica, V. P. *et al*. Optimizing production and evaluating biosynthesis in situ of a herbicidal compound, mevalocidin, from Coniolariella sp. *J. Ind. Microbiol. Biotechnol.* **43**, 1149-1157 (2016).

4. Mooney, B. P. The second green revolution? Production of plant-based biodegradable plastics. *Biochem J.* **418**, 219-232 (2009).

5. Weber, T. *et al*. antiSMASH 3.0–a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43**, W237-43 (2015).

6. Gottelt, M., Kol, S., Gómez-Escribano, J. P., Bibb, M. & Takano, E. Deletion of a regulatory gene within the *cpk* gene cluster reveals novel antibacterial activity in *Streptomyces coelicolor* A3(2). *Microbiology* **156**, 2343-2353 (2010).

7. Olano, C. *et al*. Activation and identification of five clusters for secondary metabolites in *Streptomyces albus* J1074. *Microb. Biotechnol.* **7**, 242-256 (2014).

8. Hesketh, A., Chen, W. J., Ryding, J., Chang, S. & Bibb, M. The global role of ppGpp synthesis in morphological differentiation and antibiotic production in *Streptomyces coelicolor* A3(2). *Genome Biol.* **8**, R161 (2007).

9. Gao, C., Hindra, Mulder, D., Yin, C. & Elliot, M. A. Crp is a global regulator of antibiotic production in *Streptomyces*. *MBio* **3**, 10.1128/mBio.00407-12 (2012).

10. Świątek, M. A. *et al*. The ROK family regulator Rok7B7 pleiotropically affects xylose utilization, carbon catabolite repression, and antibiotic production in *Streptomyces coelicolor*. *J. Bacteriol.* **195**, 1236-1248 (2013).

11. Komatsu, M. *et al*. Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites. *ACS Synth. Biol.* **2**, 384-396 (2013).

12. Gómez-Escribano, J. P. & Bibb, M. J. *Streptomyces coelicolor* as an expression host for heterologous gene clusters. *Meth. Enzymol.* **517**, 279-300 (2012).

13. Izawa, M., Kawasaki, T. & Hayakawa, Y. Cloning and heterologous expression of the thioviridamide biosynthesis gene cluster from *Streptomyces olivoviridis*. *Appl. Environ. Microbiol.* **79**, 7110-7113 (2013).

14. Baltz, R. H. *Streptomyces* and *Saccharopolyspora* hosts for heterologous expression of secondary metabolite gene clusters. *J. Ind. Microbiol. Biotechnol.* **37**, 759-772 (2010).

15. Komatsu, M., Uchiyama, T., Omura, S., Cane, D. E. & Ikeda, H. Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2646-2651 (2010).

16. Olano, C., Lombo, F., Méndez, C. & Salas, J. A. Improving production of bioactive secondary metabolites in Actinomycetes by metabolic engineering. *Metab. Eng.* **10**, 281-292 (2008).

17. Ochi, K. & Hosaka, T. New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters. *Appl. Microbiol. Biotechnol.* **97**, 87-98 (2013).

18. Leigh, D. A., Bradnock, K. & Marriner, J. M. Augmentin (amoxycillin and clavulanic acid) therapy in complicated infections due to beta-lactamase producing bacteria. *J. Antimicrob. Chemother.* **7**, 229-236 (1981).

19. Jacks, T. M., Schleim, K. D., Judith, F. R. & Miller, B. M. Cephamycin C treatment of induced enterotoxigenic colibacillosis (scours) in calves and piglets. *Antimicrob. Agents Chemother.* **18**, 397-402 (1980).

20. Zelyas, N. J., Cai, H., Kwong, T. & Jensen, S. E. Alanylclavam biosynthetic genes are clustered together with one group of clavulanic acid biosynthetic genes in *Streptomyces clavuligerus*. *J. Bacteriol.* **190**, 7957-7965 (2008).

21. Kenig, M. & Reading, C. Holomycin and an antibiotic (MM 19290) related to tunicamycin, metabolites of *Streptomyces clavuligerus*. *J. Antibiot. (Tokyo)* **32**, 549-554 (1979).

22. de la Fuente, A., Lorenzana, L. M., Martín, J. F. & Liras, P. Mutants of *Streptomyces clavuligerus* with disruptions in different genes for clavulanic acid biosynthesis produce large amounts of holomycin: possible cross-regulation of two unrelated secondary metabolic pathways. *J. Bacteriol.* **184**, 6559-6565 (2002).

23. Medema, M. H. *et al*. The sequence of a 1.8-Mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. *Genome Biol. Evol.* **2**, 212-224 (2010).

24. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. in *Practical Streptomyces Genetics* (John Innes Foundation, Norwich, 2000).

25. Koekman, B. & Hans, M. in *Engineering Complex Phenotypes in Industrial Strains* (ed Patnaik, R.) pp. 169-183 (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2012).

26. Sambrook, J. & Russell, D. W. in *Molecular Cloning: A Laboratory Manual.* (Cold Spring Harbor Laboratory, New York, 2001).

27. Foulstone, M. & Reading, C. Assay of amoxicillin and clavulanic acid, the components of Augmentin, in biological fluids with high-performance liquid chromatography. *Antimicrob. Agents Chemother.* **22**, 753-762 (1982).

28. Wyszynski, F. J., Hesketh, A. R., Bibb, M. J. & Davis, B. G. Dissecting tunicamycin biosynthesis by genome mining: cloning and heterologous expression of a minimal gene cluster. *Chem. Sci.* **1**, 581-589 (2010).

29. Eckardt, K. Tunicamycins, streptovirudins, and corynetoxins, a special subclass of nucleoside antibiotics. *J. Nat. Prod.* **46**, 544-550 (1983).

30. Takatsuki, A., Arima, K. & Tamura, G. Tunicamycin, a new antibiotic. I. Isolation and characterization of tunicamycin. *J. Antibiot. (Tokyo)* **24**, 215-223 (1971).

31. Takatsuki, A. *et al*. The structure of tunicamycin. *Agric. Biol. Chem.* **41**, 2307-2309 (1977).

32. Tsvetanova, B. C. & Price, N. P. Liquid chromatography-electrospray mass spectrometry of tunicamycin-type antibiotics. *Anal. Biochem.* **289**, 147-156 (2001).

33. Eckardt, K. *et al*. Streptovirudins-new antibiotics with antiviral effect. *Z. Allg. Mikrobiol.* **13**, 625-627 (1973).

34. Keenan, R. W., Hamill, R. L., Occolowitz, J. L. & Elbein, A. D. Biological activities of isolated tunicamycin and streptovirudin fractions. *Biochemistry* **20**, 2968-2973 (1981).

35. Duksin, D. & Mahoney, W. C. Relationship of the structure and biological activity of the natural homologues of tunicamycin. *J. Biol. Chem.* **257**, 3105-3109 (1982).

36. Hung, P. P., Marks, C. L. & TardrewP.L. The biosynthesis and metabolism of erythromycins by *Streptomyces erythreus*. *J. Biol. Chem.* **240**, 1322-1326 (1965).

37. Maget-Dana, R., Lakey, J. H. & Ptak, M. A comparative monomolecular film study of antibiotic A21978C homologues of various lipid chain length. *Biochim. Biophys. Acta* **962**, 201-207 (1988).

38. Raaijmakers, J. M. & Mazzola, M. Diversity and natural functions of antibiotics produced by beneficial and plant pathogenic bacteria. *Annu. Rev. Phytopathol.* **50**, 403-424 (2012).

39. Karki, S., Kwon, S. & Kwon, H. Cloning of tunicamycin biosynthetic gene cluster from *Streptomyces chartreusis* NRRL 3882. *J. Korean Soc. Appl. Biol. Chem.* **54**, 136-140 (2010).

40. Hutt-Fletcher, L. M., Balachandran, N. & LeBlanc, P. A. Modification of Epstein-Barr virus replication by tunicamycin. *J. Virol.* **57**, 117-123 (1986).

41. Klenk, H. & Schwarz, R. T. Viral glycoprotein metabolism as a target for antiviral substances. *Antiviral Res.* **2**, 177-190 (1982).

42. Pulverer, G., Beuth, J., Ko, H. L., Sölter, J. & Uhlenbruck, G. Modification of glycosylation by tunicamycin treatment inhibits lectin-mediated adhesion of *Streptococcus pneumoniae* to various tissues. *Zentralbl Bakteriol Mikrobiol Hyg A.* **266**, 137-144 (1987).

43. Pulverer, G. *et al*. Glycoprotein modifications of sarcoma L-1 tumor cells by tunicamycin, swainsonine, bromoconduritol or 1-desoxynojirimycin treatment inhibits their metastatic lung colonization in Balb/c-mice. *J. Cancer Res. Clin. Oncol.* **114**, 217-220 (1988).

44. Zhang, Z. *et al*. Modification of glycosylation mediates the invasive properties of murine hepatocarcinoma cell lines to lymph nodes. *PLoS One* **8**, e65218 (2013).

CHAPTER 4

Genome-based exploration of the specialized metabolic capacities of the genus *Rhodococcus*

Ana Ceniceros[1], Lubbert Dijkhuizen[1], Mirjan Petrusma[1], Marnix Medema[2]

1. Microbial Physiology, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands
2. Bioinformatics Group, Wageningen University, Droevendaalsesteeg 1, 6708 PB, Wageningen, The Netherlands

Abstract

Bacteria of the genus *Rhodococcus* are well known for their ability to degrade a large range of organic compounds. Some rhodococci are free-living, saprophytic bacteria; others are animal and plant pathogens. Recently, several studies have shown that their genomes encode putative pathways for the synthesis of a large number of specialized metabolites that are likely to be involved in microbe-microbe and host-microbe interactions. To systematically explore the specialized metabolic potential of this genus, we here performed a comprehensive analysis of the biosynthetic coding capacity across publicly available rhododoccal genomes, and compared these with those of several *Mycobacterium* strains as well as that of *Amycolicicoccus subflavus*. The results show that most predicted biosynthetic gene cluster families in these strains are clade-specific and lack any homology with gene clusters encoding the production of known natural products. Interestingly, many of these clusters appear to encode the biosynthesis of lipopeptides, which may play key roles in the diverse environments were rhodococci thrive, by acting as biosurfactants, pathogenicity factors or antimicrobials. We also identified several gene cluster families that are universally shared among all three genera, which therefore may have a more 'primary' role in their physiology. Inactivation of these clusters by mutagenesis may help to generate weaker strains that can be used as live vaccines. The genus *Rhodococcus* thus provides an interesting target for natural product discovery, in view of its large and mostly uncharacterized biosynthetic repertoire, its relatively fast growth and the availability of effective genetic tools for its genomic modification.

Introduction

Specialized metabolites, also known as secondary metabolites, are small molecules that are not essential for growth and reproduction of the producer organism but give them a survival advantage. One example is the production of antibiotics, which inhibit the growth of surrounding organisms competing for the same resources. Specialized metabolites are applied in human society in various ways [1-4] and comprise diverse classes of chemicals, including polyketides, peptides (produced either ribosomally or nonribosomally), saccharides, terpenes and alkaloids [5].

The bacterial genus that has been most extensively studied for its capacity to produce bioactive compounds is *Streptomyces*. Streptomycetes are the source of most of the natural antibiotics that are used in modern medicine [6]. Antibiotics were a revolution in medicine, being the cure for many—until then—deadly illnesses such as the plague, leprosy, tuberculosis or syphilis. Unfortunately, many pathogenic bacteria have developed resistance to antibiotics, in some cases even to all antibiotics currently available [7]. Several large-scale efforts are under way to find new antibiotic compounds that can be used to fight these strains. However, in many cases, these efforts suffer from frequent rediscovery of compounds previously identified from other strains [8]. The development of bioinformatics tools to analyze the growing amount of available bacterial genomic sequence information has shown that the number of biosynthetic gene clusters (BGCs) that may encode pathways capable of producing specialized metabolites is much greater than initially thought, even in strains that were already known for their specialized metabolite repertoires [9, 10], but also in many other strains from a wide range of taxonomic groups [11, 12].

One of the actinobacterial genera that has received relatively little attention from the natural products research community is *Rhodococcus*. Rhodococci are actinomycetes that contain mycolic acids in their cell

walls; they are closely related to the genus *Mycobacterium*, host of hazardous pathogenic strains such as *Mycobacterium tuberculosis* and *Mycobacterium leprae*. Two *Rhodococcus* species—*Rhodococcus equi* and *Rhodococcus fascians*—are animal and plant pathogens, respectively. Traditionally, *Rhodococcus* strains have been studied for their capacity to degrade complex organic compounds and many of them have been isolated from chemically contaminated environments [13, 14]. A recent study has shown that rhodococci not only have a vast specialized catabolic repertoire, but also a large specialized anabolic repertoire: the genomes of four *Rhodococcus* strains were shown to harbour a vast number of different BGCs, including a strikingly high number of nonribosomal peptide synthetase (NRPS)-encoding BGCs compared to other actinobacteria [12]. Among the specialized metabolites previously described in *Rhodococcus*, there are several siderophores: the hydroxamate-type siderophores rhequichelin, heterobactin, rhodochelin [15-17], and the catecholate-type siderophore rhequibactin [18]. Additionally, multiple *Rhodococcus* strains have also been reported to produce antibiotics [19]. Four (groups of) rhododoccal natural products with antimicrobial activity have been described in literature; lariantin peptide antibiotics with anti-mycobacterial activity [20], the polyketide aurachin RE from *Rhodococcus erythropolis* JCM 6824 (which has a structure similar to that of aurachin C from the Gram-negative organism *Stigmatella aurantiaca* [21, 22]), a group of peptide antifungals named rhodopeptins [23] and the recently described humimycins [24]. The gene cluster responsible for the synthesis of rhodopeptins has not yet been identified.

Here, we performed an extensive genomic analysis of the biosynthetic potential of twenty *Rhodococcus* strains with complete genome sequences available. In view of the close phylogenetic distance that *Rhodococcus* has with *Mycobacterium*, we also have analyzed several *Mycobacterium* strains, four free-living strains and three obligate pathogens. Also the only available complete genome sequence from the newly discovered genus *Amycolicicoccus* [25], which is a taxonomic

intermediate between *Rhodococcus* and *Mycobacterium*, was included, adding to a total of 28 strains. Based on a computational reconstruction of gene cluster families (GCFs) in these strains, we found several BGCs shared between all species that may play key roles in survival and could therefore be studied as potential drug targets to combat pathogenic strains, as well as clade-specific clusters that have a high probability of synthesizing novel natural products not previously described in other strains. Particularly, a striking variety of putative lipopeptide BGCs was observed. Most of the NRPSs in *Rhodococcus* strains contain an N-terminal condensation domain that belongs to the C-starter subfamily, which is known to acylate the first residue of the NRP; additionally, we found various different CoA-ligases encoded in NRPS BGCs that may be involved in lipidation. Altogether, we provide a comprehensive overview of the genomic basis of *Rhodococcus* specialized metabolic diversity and show that *Rhodococcus* is a promising and thus far underexplored target genus for natural product discovery, in view of the large number of unknown clusters present in their genomes and the availability of techniques for genetic manipulation of this genus [26].

Materials and methods

AntiSMASH analysis

Genome sequences were obtained from NCBI (http://www.ncbi.nlm.nih.gov) and loaded into antiSMASH, which was run with default settings plus inclusion of the ClusterFinder algorithm (http://antiSMASH.secondarymetabolites.org)

16S rRNA Phylogenetic tree

16S ribosomal RNA sequences were obtained from NCBI. In the case of species that contain more than one 16S rRNA genes, only one of them was included in the phylogenetic analysis since all of them appeared together in the tree. The analysis was performed which was performed using MEGA 6.0[27] (http://www.megasoftware.net/). The alignment was performed using MUSCLE with default parameters. The tree was generated using the Neighbour-Joining method and the bootstrap test with 1000 repetitions.

BGC similarity networks and gene cluster family reconstruction

A network containing 2363 nodes was generated using BiG-SCAPE (https://git.wageningenur.nl/yeong001/BGC_networks) optimized for a good separation of NRPS clusters. The parameters used were Jaccard index (0.2), similarity of domain order measured by the goodman-kruskal γ index (0.05) and the domain duplication similarity, weighted by sequence identity (0.75). Six network versions were produced with different cut-offs of 0.6, 0.65, 0.7, 0.75, 0.8 and 0.85. The lower the cut-off, the fewer connections are kept between the clusters. A cut-off of 0.75 was used for further analysis.

Analysis of the network

Groups of gene clusters with more than 7 nodes were curated manually by MultiGeneBlast (http://multigeneblast.sourceforge.net/) to confirm that they indeed constitute a GCF. Gene clusters of known molecular family but unknown final product (NRPs, PKs, RiPPs…) were considered as

the same cluster when the identity of the main biosynthetic enzymes was 60% or higher inside the same genus and 50% from different genera. If the main biosynthetic enzymes could not be identified, the organization predicted in the network was considered correct. Known and described GCFs were also curated manually with MultiGeneBlast and added or removed when required.

Gain/loss diagrams

The gain/loss diagram was created using Count [28]. A total of 112 GFCs (GCFs shared with more than 7 strains and all predicted NRPSs) were used in the study. Dollo parsimony and Wagner parsimony (gain penalty: 1) were used for the analysis.

Heat map generation

The Python Module Seaborn was used to generate the heat map. Pairwise distances were calculated using the Euclidean method, the hierarchical clustering of the gene clusters was performed with the Complete method. Strains were ordered according to their appearance in the 16S-rRNA based phylogenetic tree. In cases of strains containing more than one 16S rRNA genes, they were checked for consistent signal and the first occurring 16S rRNA in each strain chromosome was used to generate the phylogenetic tree from the heat map and the three *R. opacus* strains which have 44 to 45 BGCs.

Results and discussion

Rhodococcus has great potential for specialized metabolite production

In order to establish a phylogenetic framework to understand biosynthetic diversity in rhodococci and their relatives, a 16S rRNA phylogenetic analysis was first performed with all 28 selected strains and two *Streptomyces* species as outgroups: the model organism *Streptomyces coelicolor* and the industrial clavulanic acid producer *Streptomyces clavuligerus,* both of which are well-known specialized metabolite producers (Figure 1). The tree shows the previously described paraphyletic nature of the *Rhodococcus* genus [29, 30].

Figure 1. Neighbour-joining 16S rRNA phylogenetic tree of all 30 strains studied. A total of 1000 bootstrap replicates were performed in this analysis; bootstrap values are given in percentages.

A computational analysis (using antiSMASH + ClusterFinder [11, 31, 32]) of their biosynthetic capacity showed a considerable number of BGCs in all 28 strains (Figure 2). The strain with the highest number of gene clusters (128 BGCs, out of which 32 belong to known families and 96 are putative ClusterFinder-predicted BGCs) is *Rhodococcus opacus* R7, which also has the largest genome with 10.1 Mb; it is followed by *Rhodococcus jostii* RHA1 and the other *R. opacus* strains. The strain with the fewest gene clusters (*M. leprae*, with 15 BGCs) has the smallest genome with 3.3 Mb (Figure 2a). Both in *Mycobacterium* and *Rhodococcus*, pathogenic strains generally have a smaller genome and a limited array of BGCs [33]. Obligate pathogens live in much more stable conditions than soil bacteria and therefore they do not need to adapt to sudden environmental changes, which reduces the need of different biosynthetic and/or catabolic pathways and frequently results in genome minimization [34]. They still need a minimum arsenal of molecules to compete for resources with the host they infect. For instance, *M. tuberculosis* is known to need siderophores to capture iron, which is in most cases in low availability since the host organism sequesters it for its own use [16, 18, 35]. These siderophores are therefore targets to treat mycobacterial infections.

Interestingly, the percentage of ClusterFinder-predicted putative clusters that are not assignable to a known type of molecule is about 75% of the predicted BGCs in *Rhodococcus*, compared to the 64% and 46% for *S. coelicolor* and *S. clavuligerus* respectively (Figure 2b). These putative BGCs contain diverse types of enzymes, different between each cluster. The absolute numbers of these clusters are generally highest in rhodococcal genomes, which indicates that they encode the largest variety of thus far unknown molecules.

These findings corroborate the study of Doroghazi and collaborators[12], who recently have studied a smaller number of rhodococcal and mycobacterial genomes, and noted that they contain a large number of putative BGCs in their genomes and that the NRPS/PKS proportion in

rhodococci is higher compared to that in other actinomycetes. Our results show that these observations extend throughout the *Rhodococcus* genus (Figure 2). Indeed, most of the *Rhodococcus* strains contain a range of NRPS-encoding gene clusters but only one or two Type I PKS-encoding gene clusters, and they do not contain any Type II or Type III PKS clusters, except for the two strains of *Rhodococcus aethrivorans* and *Rhodococcus* sp. AD45 that contain one or two Type III PKSs.

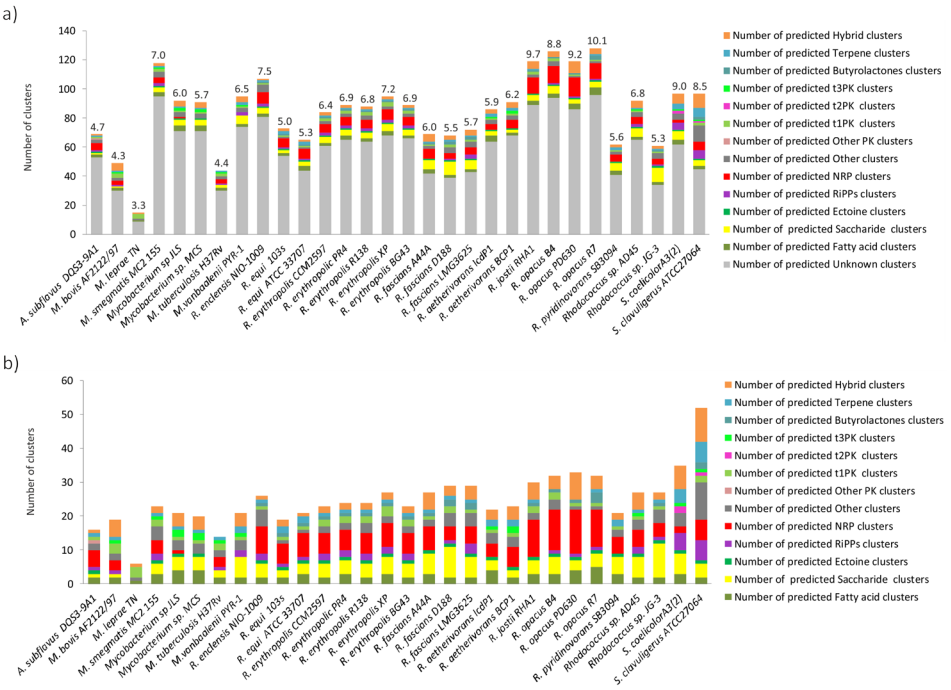


Figure 2. Number and families of BGCs detected in all 28 *Rhodococcus* and *Mycobacterium* strains, compared to the well-known specialized metabolite producers *S. coelicolor* and *S. clavuligerus*. a) Total number of biosynthetic gene clusters predicted by antiSMASH, including ClusterFinder. On top of each bar, numbers represent the size of each genome in Mb. b) BGCs predicted by antiSMASH to code for known families of biosynthetic pathways. It can be observed that the proportion of NRPS BGCs in *Rhodococcus* compared to other types of molecules is much higher, especially in *R. jostii* RHA1 and *R. opacus* strains. (T3PK: Type III polyketide. T2PK: Type II. polyketide, T1PK=Type I polyketide, PK: Polyketide. NRP: nonribosomal peptide. RiPP: ribosomally synthetized and posttranslational modified peptide)

Intriguingly, the genome of *Rhodococcus pyridinovorans* SB3094, which is the strain with the smallest genome of all free-living rhodococci studied

in this work (5.24 Mb), shows a genomic duplication of a 366-kb region of its chromosome covering four of its BGCs. The genus *Rhodococcus* is known to contain a great redundancy of genes but duplication of complete gene clusters has not been reported before [36]. Notably, in some industrial strains, copy number variation of BGCs has been shown to lead to increased specialized metabolite production [37]; a similar high-producing phenotype may have driven this evolutionary event and its fixation in a population. The combination of a small genome and duplication of 7% of it indicates that this strain is highly specialized to its environment [33, 38]. *R. pyridinovorans* SB3094 was isolated from oil fields, thus this strain's specialized metabolism may have adapted to optimized degradation of fatty acids and increased survival in hydrophobic environments. In fact, the only NRPS cluster duplicated (NRPS-17) has a C-starter domain, which is known to acylate the first residue of the NRP [39], indicating that it probably encodes the biosynthesis of a lipopeptide that could potentially serve as biosurfactant. Surfactants are compounds that decrease the surface tension of two fluids and improve the availability of hydrophobic compounds such as oil. Also, *R. pyridinovorans* is the only strain along with *Rhodococcus sp.* AD45 to lack a butyrolactone BGC; one potential reason for this may be that the solubility properties of butyrolactones limits their use in quorum sensing in this environment. This cluster is known to be very conserved in *Rhodococcus* and is thought to play an important role in this genus [12]; in streptomycetes, the quorum-sensing γ-butyrolactone molecules are known to be involved in the regulation of their specialized metabolism.

One aim of this analysis was to identify biosynthetic pathways that may be essential to rhodococcal and mycobacterial metabolism and would thus present possible drug targets to combat pathogenic strains; a second aim was to identify gene clusters that are only present in one or a few strains, which may encode biosynthetic pathways for the production of novel bioactive compounds. For this purpose, we used the BiG-SCAPE software (Navarro-Muñoz, Yeong, Medema et al., in preparation) to

construct a sequence similarity network that categorizes the different gene clusters into separate groups, thus providing a powerful visualization of shared/non-shared BGCs in all the studied strains (Figure 3).

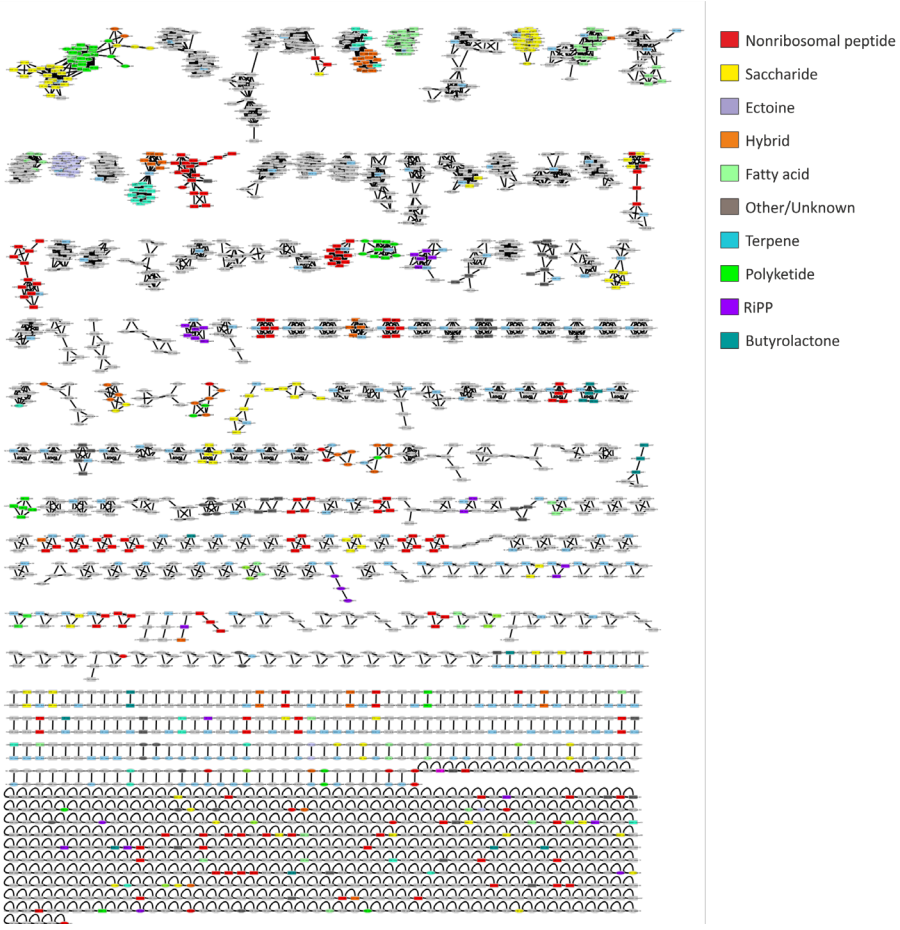


Figure 3. Sequence similarity network relating the gene clusters detected by antiSMASH from all strains. Each cluster is represented by a square in the case of *Rhodococcus* or as an ellipse in the case of *Mycobacterium* and an octagon in the case of *Amycolicicoccus subflavus*. The different shapes are colour-coded for different BGC families with a colour scheme consistently applied in other figures as well. The upper part of the network represents the group of GCFs shared across multiple strains, while the lower part contains the clusters that are only present in one strain.

In order to more fully understand the evolutionary histories of all studied strains that have led to the currently observed BGC repertoires

represented in the network (Figure 3), we used ancestral state reconstruction with Count [28] to identify the most parsimonious BGC gain/loss events (Figure S1). For that analysis, we used a total of 114 different BGCs: all families shared between more than 7 strains and all NRPS BGCs detected in this work through the network (Table S1). Figure S1 shows which GCFs are conserved through evolution and which ones are not. Twenty-four GCFs are jointly present in the genomes of different strains of *Mycobacterium* and *Amycolicicoccus* (which form a monophyletic clade), as well as in *Rhodococcus*. Altogether, all *Rhodococcus* clades share 36 GCFs, the previous 24 plus 12 more GCFs that are only present in all *Rhodococcus* strains. Notably, it can be observed that the branches leading to *R. jostii* RHA1 and the three strains from *R. opacus* show many GCF gain events (Figure S1), which indicates that ecological specialization of these strains involved acquisition of several biosynthetic pathways through horizontal gene transfer.

Shared GCFs may have essential functions and therefore offer possible targets to combat pathogenic strains

A detailed analysis was performed on the GCFs shared among more than seven strains, which amount to a total of 37 GCFs. We reasoned that biosynthetic pathways strongly conserved between *Mycobacterium* and *Rhodococcus* may offer possible drug targets in pathogenic strains from both genera, since their conservation suggests that they are important for survival. A heat map representation of the presence/absence patterns of these 37 GCFs in each strain was performed to analyze the data (Figure 4).

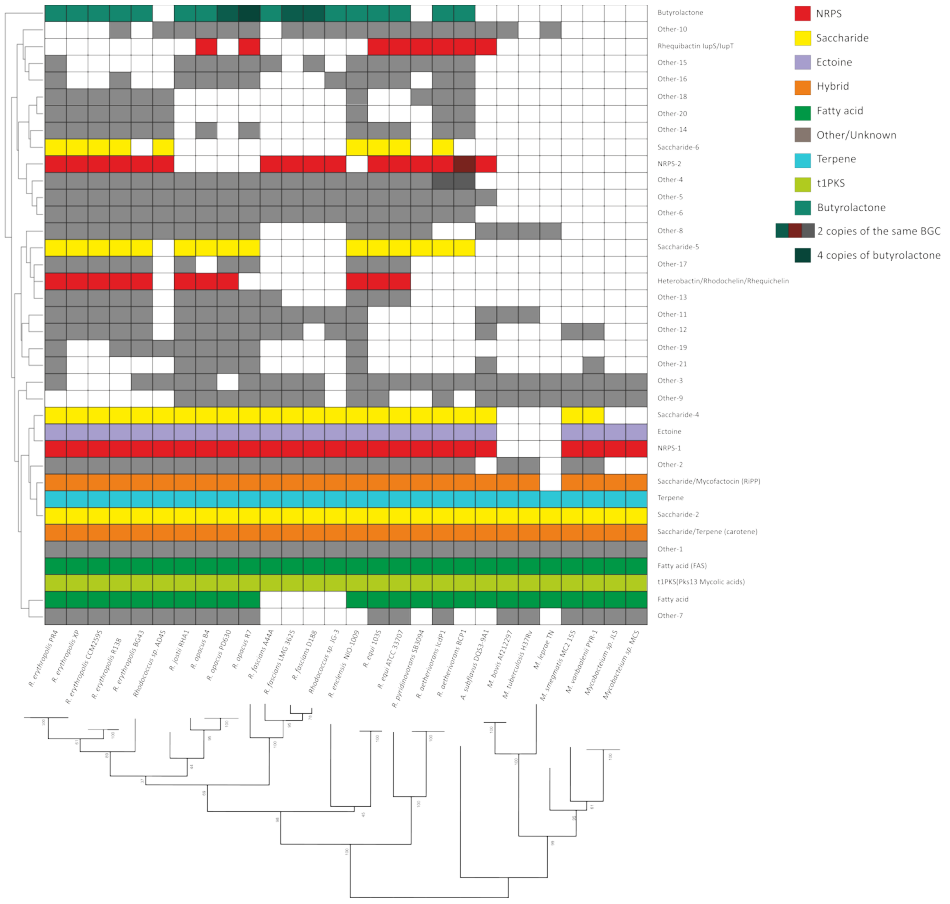


Figure 4. Heat-map representing the presence/absence patterns of GCFs shared between more than seven strains. The strains are displayed horizontally in phylogenetic order. Vertically, they were ordered using hierarchical clustering (see Methods for details). Six clusters, which belong to various families, are shared between all species and genera. Darker areas represent more than one copy of the cluster.

The vast majority of these GCFs have no experimentally characterized members. Five GCFs have members that have been previously described in at least one strain: 1) the Type I PKS cluster that contains the gene *pks13* catalyzes the last condensation step of mycolic acid biosynthesis [40], which is present in all studied strains. This cluster is grouped together in the network with a saccharide cluster which may also be related to cell wall biosynthesis: it encodes arabinogalactan biosynthesis enzymes, family 2 glycosyltransferases and O-antigen transporters; 2) the NRPS

BGC encoding the biosynthesis of the siderophore heterobactin, present only in 11 *Rhodococcus* strains [41]; 3) the carotenoid BGC [42], not detected in *Mycobacterium* strains; 4) the mycofactocin BGC, encoding the biosynthesis of a ribosomally synthesized and post-translationally modified peptide of unknown function that was initially discovered by bioinformatics analysis [43, 44], and is only absent in *M. leprae*; 5) the butyrolactone gene cluster, detected by the presence of an *afsA* homologue, encoding the main biosynthetic enzyme of the γ-butyrolactone signalling molecules known to be involved in the regulation of the secondary metabolism in *Streptomyces* [45], which are present in all *Rhodococcus* strains except for *Rhodococcus* sp. AD45 and *R. pyridinovorans* SB3094. Five of the 37 GCFs are present in all strains from the three genera studied; each of them is predicted to encode a biosynthetic pathway for molecules belonging to a different family. One of them is the already mentioned **Type 1 PKS *pks13***, involved in mycolic acid biosynthesis. All these organisms are known to possess mycolic acids in their cell walls, including *A. subflavus,* but these mycolic acids vary in their complexity [40, 46].

The GCF called **Terpene,** which is also shared by all strains, includes a lycopene cyclase as well as genes encoding the enzymes SufD and NifU, known to be involved in the iron-sulphur cluster biosynthesis. Iron-sulphur clusters are known to be cofactors of different proteins [47]. This cluster also contains the genes encoding RipA and RipB, known to be essential for cell division in some *Mycobacterium* species, although only one of these is necessary for the cells to survive in *M. smegmatis* [48].

The third universally shared GCF is **Saccharide-2** family; BGCs that are members of this family contain genes for the synthesis of menaquinone synthesis, which is also known as vitamin K2 and, among other functions, plays a role in the respiratory electron transport chains in bacteria. It also is known to be an important factor in the latent phase of infection in *M. tuberculosis* [49].

Two other GCFs are shared between all species, for each of which the function is less certain. One of them is predicted to contain fatty acid BGCs encoding a **fatty acid synthase (FAS)**; the other one (**Other-1**) contains BGCs that may be involved in heme biosynthesis, containing genes encoding a hydroxymethylbilane synthase, a glutamyl-tRNA reductase for the synthesis of heme (HemC and HemA respectively), a possible redox-sensing transcriptional repressor, a putative phosphoserine phosphatase, a probable UDP-glucose 4-epimerase, a predicted excisionase, a pyrroline-5-carboxylate reductase and several proteins of unknown function. Similarly, to the four known universally shared clusters, the latter two unknown clusters are most likely important for the survival of rhodococci and mycobacteria. Characterizing the function and products of these clusters may lead to the identification of novel targets to develop vaccines against pathogenic strains such as *M. tuberculosis, Mycobacterium bovis, M. leprae*, *R. fascians* or *R. equi*.

Another GCF present in all strains encodes predicted saccharide-terpene hybrid clusters (indicated as **Saccharide/terpene** from now on). However, a closer analysis shows that it is also predicted in some strains as an "unknown" cluster, as just the terpene cluster or as just the saccharide cluster, which indicates that they may represent two different clusters that are close together on the chromosome in some species and are therefore (likely incorrectly) predicted by antiSMASH to constitute a hybrid cluster. Consequently, the strains that have only one of these clusters were also grouped in the same set, even if they are in fact not the same cluster. In the end, this therefore does not constitute a bona fide GCF with members in all strains. The terpene cluster (which was not detected in *Mycobacterium*) contains the genes encoding the enzymes for carotene biosynthesis, while the saccharide cluster is known to be involved in cell wall synthesis in *M. tuberculosis* [50]. Given their close proximity on the genome, it cannot be excluded that the enzyme products of these two BGCs act together in some species to produce glycosylated carotenoids.

Among the partially shared GCFs, it is worth mentioning that the ectoine GCF and the NRPS-1 GCF are present in all strains except for pathogenic mycobacteria. **Ectoine** is an osmolyte that is produced in high salt conditions as osmoprotectant [51]. This molecule thus, does not seem essential for these three pathogenic *Mycobacterium* strains. The **NRPS-1** BGCs contain one NRPS, a reductase, a putative aldehyde dehydrogenase, a probable aromatic ring dioxygenase, a putative gamma-glutamyltransferase, a probable multidrug resistance transporter from the MFS family, a putative GntR transcriptional regulator and an aminopeptidase. The presence of the transporter-encoding gene indicates that this gene cluster may encode the biosynthetic pathway of a bioactive molecule for which self-resistance is needed. Interestingly, the **NRPS-1** cluster is not present in pathogenic *Mycobacterium* strains. A BLAST search with the amino acid sequences of this multidrug resistance transporter gene in the genomes of *M. tuberculosis*, *M. bovis* and *M. leprae* did not show any hit with significant homology, which indicates that their genomes do not retain this resistance mechanism and that these strains may be sensitive to the product of this BGC. Three other clusters of unknown function (Other-4, Other-5 and Other-6) were detected in all *Rhodococcus* strains. Other-5 is also present in *A. subflavus*. **Other-4** contains a homologue of a cutinase enzyme which in phytopathogenic organisms is involved in the infection process by degrading the plant cell-wall [52, 53]. These cutinase enzymes were also predicted in *M. tuberculosis* and are thought to be involved in providing substrates to form mycolic acids, or in pathogenicity [53]. Unfortunately, the function of the **Other-4** cluster could not be predicted in more detail. **Other-5** includes a protein with 92% identity to the IdeR global iron-dependent regulator described in *M. tuberculosis*, which is also homologous to the diphtheria toxin repressor DtxR from *Corynebacterium diphteriae* [54]. IdeR was also found in *R. equi* and *R. erythropolis* [55]. In pathogenic strains, it is known to regulate different virulence factors which are activated when the bacterial strain enters a

host cell, where iron levels are scarce due to iron sequestering enzymes from the host as transferrin [54]. This cluster also includes an enzyme with a PAC2 (proteasome assembly chaperone) domain involved in the formation of the proteasome, which is essential for pathogenicity in *M. tuberculosis* [56]. A protein from the superfamily II of RNA and DNA helicases, a UDP-galactose-4-epimerase, two hydrolases and two hypothetical proteins are also encoded in this gene cluster. However, the products of the **Other-5** clusters remain unknown. Also for **Other-6** we could not identify a function. It encodes two multidrug transporters, which indicates that this cluster may be producing a bioactive compound. It also contains genes coding for three transcriptional regulators, a putative esterase, two dehydrogenases, a glutamate decarboxylase, a glutathione S-transferase, an acetyl transferase, a possible glycolate oxidase FAD-linked subunit, a possible enoyl-CoA hydratase, a monooxygenase and four hypothetical proteins. **Saccharide-5** is also present in all *Rhodococcus* strains, except for *Rhodococcus* sp. AD45, both *R. fascians* strains and *Rhodococcus sp*. JG-3. No putative function was deduced for this cluster. It contains genes coding for a DNA-binding helix-turn-helix protein, an isocitrate lyase, a 3-hydroxybutyryl-CoA dehydrogenase, a putative 5-methyltetrahydropteroyltriglutamate-homocysteine S-ethyltransferase, a cellulase, a GroES-like protein and a hypothetical protein.

Figure 4 shows that *Rhodococcus sp*. AD45 lacks nine clusters that are present in all its close relatives: **Saccharide-5, Other-11, 12, 13, 17, 19, 21,** the **Butyrolactone** and the **Heterobactin** gene clusters. This strain was isolated from fresh water sediments and is able to use isoprene as sole source of carbon. However, the size of its genome is not much smaller than its closest relatives, the *R. erythropolis* strains. Indeed, this is the third *Rhodococcus* strain for which the majority of its BGCs belong to species-specific GCFs as it will be discussed later in this work, which suggests that this strain has adapted to a very specific environment by losing some otherwise conserved gene clusters, and by gaining others,

probably through horizontal gene transfer from other strains in its environment.

R. pyridinovorans is a free-living *Rhodococcus* strain with a small genome, and about 366 kb of its genome is duplicated. When this 366 kb region was searched using BLAST, it showed that *Rhodococcus* sp. P52 also contains a region with 98% identity and 85% coverage of this genomic fragment. This strain was also isolated from oil fields, which suggests that the enzymes encoded in this region are important for survival in this harsh environment. Part of these 366 kb are also present in the genomes of *R. aetherivorans* and *Rhodococcus* sp. WB1 with a query coverage of 46-47% and an identity of 85%. The fact that this strain thus has a minimized genome, while still having four clusters duplicated, suggests a strong adaptation to its environment [33, 38], probably at least partially facilitated by increased biosynthesis of the products of these four clusters. For none of these clusters, members of their parent GCFs have been experimentally characterized, and none of them are shared with more than three strains. Two of them are predicted as saccharide gene clusters; a third cluster is of an unknown biosynthetic family. The fourth cluster is an NRPS cluster (**NRPS-17**), which is only shared with *R. equi* strains; interestingly, it contains a C-starter domain, indicating that it is a lipopeptide which might be acting as a biosurfactant that can benefit the strain in its hydrophobic environment (see further discussion below).

NRPS clusters are the most dominant family in rhodococci and may encode interesting novel compounds

NRPS clusters are highly represented in *Rhodococcus* genomes. NRPSs can synthesize a great variety of peptides; more than 500 different precursors have been identified that can be used by NRPSs [57], thus creating a highly-varied array of compounds [57]. Each precursor amino acid added is specified by the adenylation domain of an NRPS module. Apart from this diversity of precursors, the peptide can be modified after it is released from the NRPS by other tailoring enzymes that produce

significant changes in the structure. In total, 79 distinctive NRPS GCFs were found across all 28 strains. Only one of them is shared with all strains, except for the pathogenic *Mycobacterium* strains; the product of this shared cluster is not known. Most of the NRPS clusters are present only in one strain or only in a small group of related strains as is the case of the *R. erythropolis* clade, a clade comprising *R. opacus* strains and *R. jostii*, and the *R. fascians* clade (Figure 6). In the case of *R. fascians*, it is possible that these clusters are involved in pathogenicity, as is believed to be the case for **NRPS-31** which was described in *R. fascians* D188 [52]. Mutagenesis and expression studies performed with this gene cluster revealed that it plays a role in pathogenicity but is not essential. This cluster is located on the plasmid of this strain, but the final product of the cluster and its physiological role are still unknown.

A few NRPS gene clusters that have been described in different strains of *Rhodococcus* represent siderophores in all cases, reflecting the importance of iron uptake for these bacteria. The importance of iron in rhodococcal physiology is also corroborated by the different gene clusters detected containing genes encoding the biosynthesis of heme groups, porphyrin and iron-sulphur cofactors. The gene cluster involved in the synthesis of one important iron-scavenging molecule, the hydroxamate-type siderophore rhodochelin, has been described in *R. jostii* RHA1 [15]. This cluster was identified in most *Rhodococcus* strains. The corresponding gene cluster from *R. erythropolis* PR4 has been identified previously as the gene cluster responsible for synthesis of the siderophore heterobactin [17] and in *R. equi* 103S as responsible for the synthesis of rhequichelin [16]. The three molecules are chemically different but closely related. The encoded protein sequences and gene order in the clusters are very similar, and therefore they are grouped together in our network. Some of the genes present in each cluster vary between species and as described in Miranda-Casaluengo et al. [16], three out of four adenylation domains from the NRPSs are conserved between species with the only differences being two active sites in the second adenylation domain.

These differences are probably responsible for the structural variations between heterobactin and rhodochelin (Figure 5) as well as the differences with the predicted structure of rhequichelin [16].

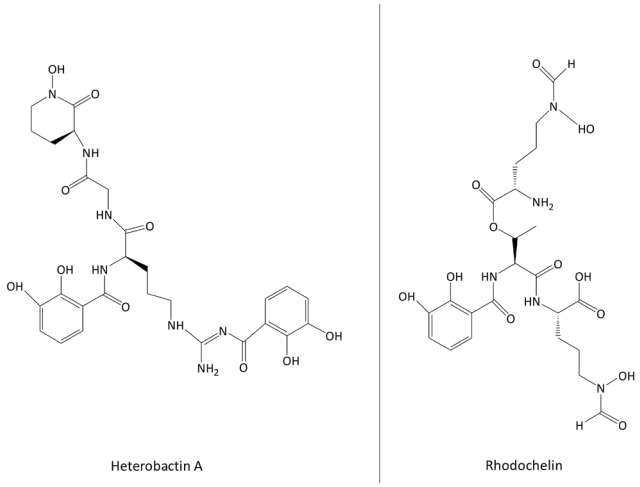


Figure 5. Structure of the siderophores heterobactin A and rhodochelin. Adapted from Bosello et al. [15, 17].

In case of rhodochelin, further genes are involved in its synthesis, located outside this gene cluster [15]. This NRPS gene cluster thus is a good example of how very similar NRPS clusters can be responsible for the production of different molecules. Another group of NRPS-synthesized siderophores described from *Rhodococcus* is constituted by the catecholate-type siderophore rhequibactins synthesized by the rhequibactin BGC [18, 58]. This siderophore was described in *R. equi* 103S and is synthesized by two different NRPS, IupS and IupT; it is thought to be used only in saprophytic growth, since its deletion did not affect the pathogenicity but prevents the growth of the strain as a free-living organism. The *iupU* gene was also described in this strain by Miranda-Casaluengo et al. [18] and was believed to be related to rhequibactin biosynthesis. The molecular product of this pathway was predicted to be a non-soluble siderophore believed to work similar to mycobactin, which is cell wall-bound and therefore not diffusible [18]. Recent work, describing the *Rhodococcus* antibiotics humimycins, identified the *iupU* gene as encoding the NRPS responsible

for the synthesis of humimycin A [24]. Humimycins were synthesized based on the predicted product of the NRPS encoded by IupU and by another NRPS present in the *R. erythropolis* genome (encoding the biosynthesis of a variant, humimycin B). Both molecules showed potent activity against methicillin resistant *Staphylococcus aureus* (MRSA), by targeting lipid II flippases. The NRPSs contain a C-starter domain, which indicated that the first residue of the NRP is acylated. Therefore, a β-hydroxymyristic acid was added to the N-terminal residue of the compounds, but no further modification was done in the product. It is not unlikely that the actual natural products are further modified by other enzymes encoded in these BGCs and therefore, their natural function may be different from that of an antibiotic. It remains possible that they function as a siderophore, as suspected previously regarding IupU. While Miranda-Casaluengo et al. [18] show that the expression of *iupU* is not controlled by iron, as is normally the case for siderophores and its deletion does affect growth of *R. equi* 103s in low iron conditions when growing as a free-living organism. Chu et al. [24] screened for the presence of humimycins in the culture broth of *Rhodococcus* strains but they could not find any compound with a similar structure, and concluded that it may be a silent gene cluster. While the IupU pathway indeed maybe silent in *R. equi* ATCC 33707 and *R. erythropolis* SK121, the expression of *iupU* was shown to be constitutive and high in *R. equi* 103s [18]. If the BGC indeed does encode the biosynthesis of a non-diffusible siderophore, it would not be possible to find it in the culture broth; rather, it would most likely be attached to the cell envelopes of the producing organisms. Of course, it remains possible that IupU does synthesize a real humimycin-like antibiotic: for example, one or more intermediates of the synthesis could be affecting the expression of rhequibactin, which would provide an alternative explanation for the low-iron phenotype of the knockout. Interestingly, yet another NRPS GCF, **NRPS-5**, which is present in all *R. erythropolis* strains, has a NRPS with the same domain architecture as the humimycin NRPSs, but with different predicted substrates in modules 1, 4, 6 and 7. This cluster also contains a

long chain fatty acid CoA-ligase-encoding gene. Further experimental studies will be needed to verify the natural physiological roles of each of these intriguing nonribosomal peptides.

Two mycobacterial GCFs encoding siderophore-producing NRPSs were detected in our analysis, which are responsible for the production of the hydroxamate-type siderophore mycobactin and exochelin[35], respectively. The second set of genes necessary for mycobactin biosynthesis, located on a different locus, were only detected in *M. tuberculosis* [59]. Mycobactin is known to be essential for *M. tuberculosis* pathogenicity. It is hydrophobic and is localized in the cell wall, and is thought to work together with the siderophore carboxymycobactin, which is a soluble siderophore. It is believed that carboxymycobactin transfers the iron to mycobactin which is then reduced from Fe^{3+} to Fe^{2+} and transported into the cytosol [35]. The transfer is thought to be mediated by the iron-dependent membrane protein HupB [35]. The mycobactin BGC was detected in every *Mycobacterium* strain except for *M. leprae* TN. Exochelin has been described in *M. smegmatis*. In the genome annotations that we used, the NRPS is only one ORF instead of two. FxbB and FxbC are fused and the BGC containing this fused NRPS was only detected in *M. smegmatis*.

The products of the other NRPS GCFs (74 in total) remain unknown, which suggests a great potential for identifying novel nonribosomal peptides. The clusters **NRPS-36, NRPS-30, NRPS-29** are only present in the facultative horse pathogen *R. equi*. This strain has also been described as a human opportunistic pathogen with a pathogenicity mechanism similar to that of *M. tuberculosis,* mainly infecting alveolar macrophages; this has attracted large interest towards the strain [60]. The products of these three NRPS clusters have not been identified to date. They may give the producer an advantage in its environment and also may have a role in pathogenicity. The same applies for the NRPSs only present in *R. fascians* species. Mutational inactivation of these gene clusters, followed by tests

on infective abilities and survival of these mutants in host cells, should clarify whether they can be used to develop drugs that may target any of these enzymes, or even a vaccine.

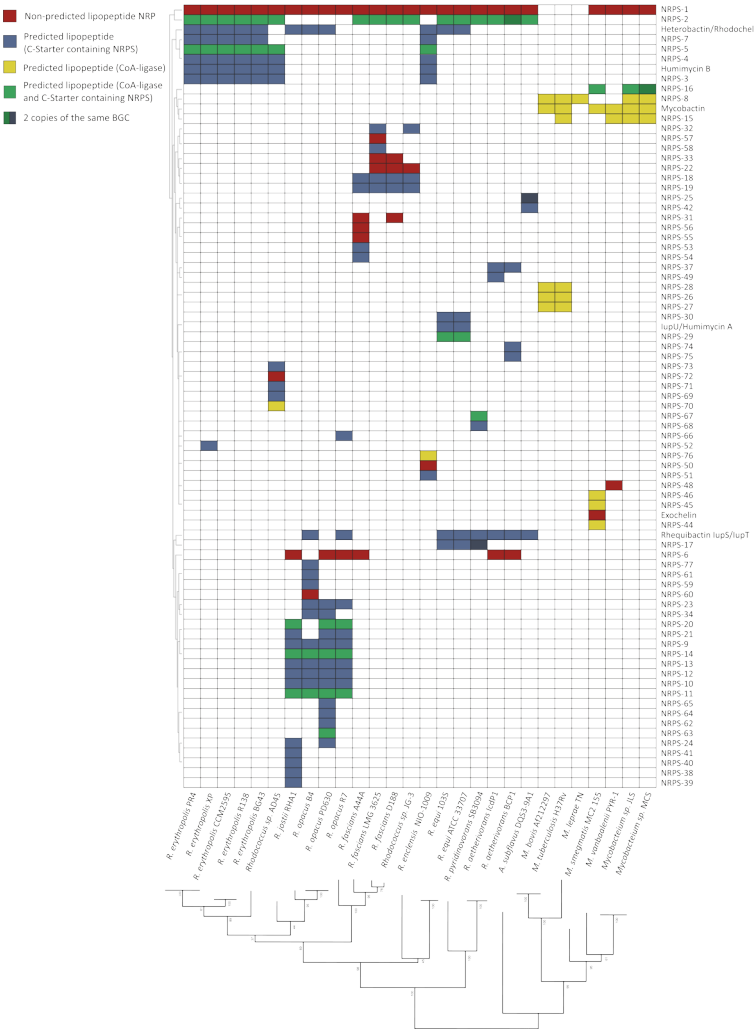


Figure 6. Heat-map showing all NRPS clusters and their presence/absence pattern in the different strains. White indicates absence of the cluster. Blue indicates presence of a gene cluster containing a C-starter domain in at least one of the NRPSs, which indicates that it encodes the biosynthetic pathway for a lipopeptide. Yellow indicates presence of a NRPS cluster containing an acyl CoA-ligase/synthetase, also indicative of the final product being a lipopeptide. Green indicates presence of NRPS clusters containing both one or more acyl CoA ligases/synthetases and a C-starter domain in one of the NRPS. Red indicates presence of NRPS clusters not predicted to synthesize lipopeptides.

Putative lipopeptides

We observed that several of these NRPS gene clusters contain individual fatty acid CoA ligases/synthetases as well as C-starter domains [39], suggesting that the final product maybe a lipopeptide. *Rhodococcus* strains seem to have preference for lipopeptides formed by a C-starter domain or the combination of C-starter domain and an acyl CoA-ligase/synthetase, while *Mycobacterium* strains have a bigger proportion of clusters with a CoA-ligase and without C-starter domains (Figure 6). Lipopeptides include antibiotics, such as daptomycin, the last major antibiotic that has been commercialized [61]. However, lipopeptides have other functions besides antibiotics: some function as surfactants, others display haemolytic activity [62] and still others play a role in establishing infection and/or biofilm formation [63]. They are formed from a cyclic oligopeptide, non-ribosomally synthesized, to which an acyl chain is attached [64]. These compounds are known to have different antimicrobial activity and toxicity depending on the length of the acyl chain. The acyl chain can be attached to the oligopeptide by different methods [65]: through a stand-alone acyl-carrier protein (ACP) and fatty acid ligase (AL), as is the case of daptomycin, by a hybrid NRPS/PKS enzyme containing an ACP and AL domain as is the case for mycosubtilin, or by a specialized C-starter domain in the NRPS, as is the case for surfactin. In the case of the calcium-dependent antibiotic from *Streptomyces* species, the fatty acid is synthesized in a specific pathway (Fab enzymes encoded in the CDA gene cluster and enzymes from the primary metabolism), and is then attached to a stand-alone ACP that directly transfers the lipid to the condensation domain of the NRPS where it is attached to the peptide [66]. Of the 79 distinct NRPS GCFs found our analysis, 69 show hallmarks of encoding the biosynthesis of lipopeptides. Also, all rhodococcal genomes studied encode putative lipopeptide BGCs, regardless of their specific ecological diversity. Still, the wide variety of bioactivities known to be associated with lipopeptides explains why they are likely to be important in various niches: they may aid in hydrocarbon degradation in oil field-dwelling

rhodococci through surfactant-mediated dispersion, solubilization, or emulsification of hydrophobic substrates [67], they may aid the infection process in pathogenic strains or function as antibiotics in saprophytic ones. To zoom in on the biosynthetic diversity of lipopeptides encoded in rhodococcal genomes, we performed a detailed analysis of the fatty acid CoA-ligases and synthetases encoded in NRPS clusters. Out of the 79 GCFs, 19 encoded distinct CoA-ligases that may be involved in lipopeptide biosynthesis[13, 68]. Notably, such ligases may also be encoded outside the cluster: for example, the enzymes in charge of the transfer of the acyl chain to the peptide part of mycobactin are in different loci as the peptide synthetases [59]. A phylogenetic study of these enzymes, comparing them to previously described fatty acid CoA-ligases involved in the synthesis of characterized lipopeptides (Figure S2), highlights their diversity.

Given the fact that such a wide variety of rhododoccal lipopeptide BGCs exists, we hypothesize that they have adapted to specific ecological sub-functions during evolution. If this indeed is the case, we predict that a dynamic evolution of lipopeptide BGC repertoires has occurred. In order to study this dynamic evolution, we performed ancestral state reconstruction of the 67 clusters using the software Count [28] to identify GCF gain/loss events across the evolutionary history of the *Rhodococcus* genus (Figure 7). Indeed, the vast majority of putative lipopeptide GCFs showed a taxon-specific distribution. None are conserved throughout *Mycobacterium* and *Rhodococcus*, but NRPS-2 and rhequibactin IupS/IupT are present in several strains of *Rhodococcus* and in *A. subflavus* (Figures 4 and 7). The clade with the largest number of predicted lipopeptides is the one containing *R. jostii* RHA1 and the three *R. opacus* strains. With 17 putative lipopeptide BGCs, *R. opacus* P630 contains the largest number, followed by *R. jostii* RHA1 with 15 putative lipopeptide BGCs. The other *Rhodococcus* clades harbour between three and eight putative lipopeptide BGCs per genome. The *R. jostii* RHA1 and *R. opacus* PD630 and B4 strains have been studied for their ability to accumulate carbon in triacyclglycerols (TAGs) when growing in nitrogen

limited media and using different carbon sources [36, 69, 70]. Alvarez et al. [69] have shown a higher triacyclglycerol accumulation when using gluconic acid as the sole carbon source and an even higher production in the case of strain PD630 when grown in olive oil [36, 69]. Lipopeptide surfactants may facilitate degradation of the hydrophobic compounds present in their growth media, allowing their import into the TAG biosynthesis pathway. Interestingly, *M. tuberculosis* has 5 putative lipopeptide GCFs (**NRPS-15, NRPS-26, NRPS-27, NRPS-28 and NRPS-8**), in addition to mycobactin, making it the *Mycobacterium* strain with the largest number of lipopeptides. **NRPS-26** and **NRPS-28** are not canonical NRPS clusters. They do not possess a modular NRPS enzyme but stand-alone domains usually found in modular enzymes, as described in Wang *et al*. [71]. Lipopeptides are known to induce a strong immune response and many of them have remained uncharacterized [72, 73]. The products of these putative lipopeptide BGCs may allow development of vaccines against *M. tuberculosis* by inducing a response of the host immune system.

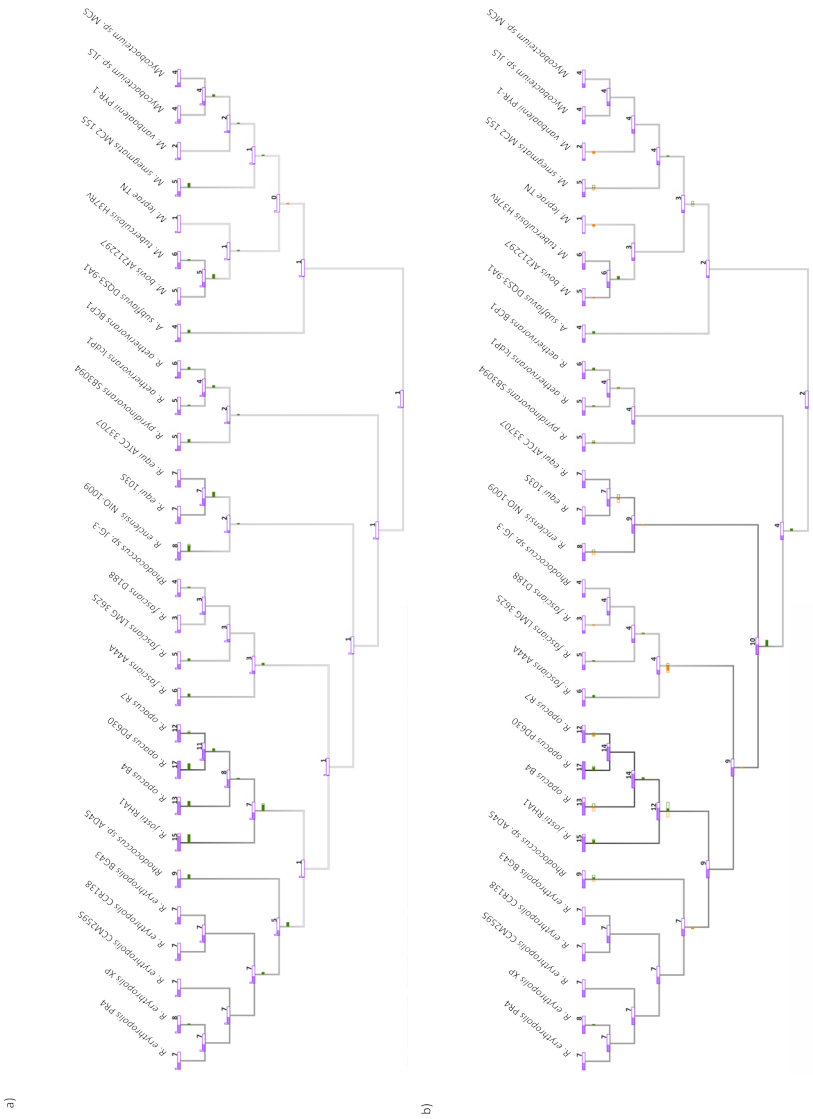


Figure 7. Gain/loss events of the putative lipopeptides in all species. Horizontal lines under each node indicate the number of gained or lost clusters. Green to the right gain, yellow to the left loss. The purple bar on each node represents the proportion of clusters present in each strain from the total 19 clusters analyzed. a) Wagner parsimony with a gain penalty of 1. b) Dollo parsimony

Many rhodococci and mycobacteria contain strain-specific GCFs

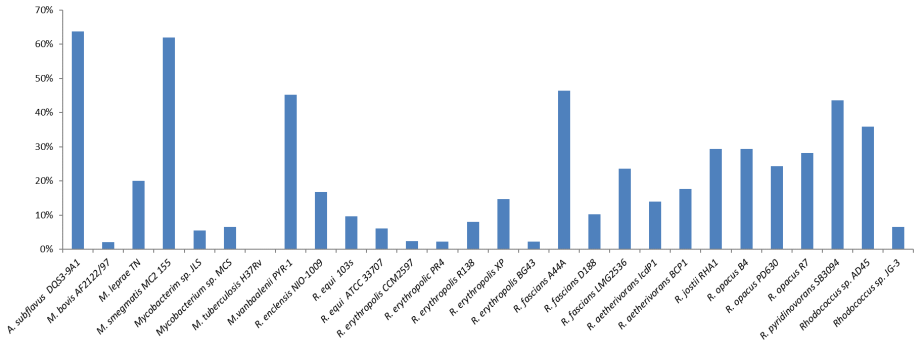


Figure 8. Percentage of strain-specific gene clusters from all detected GCFs present in each strain. *M. tuberculosis* is the only one that shares all its BGCs with other species, while *M. smegmatis* and *A. subflavus* only share 38.14% and 36.23% of their BGCs with the other strains studied. The number of gene clusters not shared by *R. fascians* A44A (46.38%) compared to its close relatives *R. fascians* LMG2536 (23.61%) and *R. fascians* D188 (10.29) is also striking but explained by their presence in different clades in the phylogenetic tree (Figure 1), also observed in [74].

Every strain except for *M. tuberculosis* has strain-specific GCFs (Figure 8). *M. tuberculosis* shares all its GFCs with at least one other strain, in many cases with *M. bovis*. The *M. smegmatis* and *A. subflavus* strains display the highest number of strain-specific GCFs (about 65%) followed by *M. vanbaalenii* and *R. fascians* A44. A. *A. subflavus* is the only strain of its genus studied, therefore it was expected that it would have many specific GCFs. On the contrary, this was not expected for the *R. fascians* strains. *R. fascians* A44A has almost 50% of species-specific clusters, while only 10% and 23% of the GFCs from *R. fascians* D188 and *R. fascians* LMG 2536, respectively, are species-specific. Creason et al. previously indicated that *R. fascians* A44A is relatively distantly related to the other *R. fascians* strains [74]. Possibly, these differences are also responsible for host range specialization among members of this phytopathogenic species as well as differences in the symptoms caused.

Another interesting observation is that most GCFs present on plasmids are unique to the species they are found in. This is true in the cases of *R. jostii* RHA1, *R. erythropolis* R138, all three strains of *R. opacus* and *R.*

pyridinovorans SB3094. Plasmids are mobile and therefore the GCFs encoded on them might be expected to be present in more than one species. To find out whether these clusters are shared with other strains, which also may indicate which strains share the same habitat, *R. jostii* RHA1 was studied in more detail. Fifteen out of the 35 non-shared clusters were found to be present on the plasmids of *R. jostii* RHA1. This species has a total of 18 clusters on its 3 plasmids, thus most of them are not shared with the rest of the species in this study, not even with its closest relative *R. opacus*. The number of strain-specific clusters present on the plasmids of *R. jostii* RHA1 is comparable to that in *R. opacus* R7, which has 11 strain-specific clusters on its plasmids from 37 non-shared of a total of 128 predicted gene clusters. However, it is much higher than the number of strain-specific clusters present on the plasmids of their close relatives *R. opacus* PD680 and *R. opacus* B4: only three strain-specific clusters are present on the plasmids of *R. opacus* PD680 and five in *R. opacus* B4. Due to the previously mentioned special nature of the *R. pyridinovorans* SB3094 genome, which is small and has a 366 kb duplication, we also analyzed the two BGCs predicted on its plasmid. Homologs of the enzyme-coding genes in both these clusters were also found in other rhodococci isolated from oil-contaminated soil, soil, waste water, and even in *Rhodococcus gordonia*, which has been isolated from clinical material and phenol-contaminated soil. This suggests that indeed these BGCs may confer specific traits to the strain that allow it to thrive in such conditions. Experimental studies are needed to further characterize these clusters.

Activation of cryptic BGCs may allow discovery of new bioactive compounds

As presented in this work, most BGCs identified in the different strains studied here are completely unknown. Various methods have been developed to induce the expression of such BGCs as the ones described in Chapter 1 in this thesis. Some of these approaches are the deletion of known biosynthetic pathways to make precursors available for other

routes as we performed in Chapter 3, the co-cultivation of two or more strains, the manipulation of regulatory systems as described in Chapter 5 or the heterologous expression in other strains as we show in Chapter 2. In some cases, introduction of synthetic promoters may be needed to enforce BGC expression in the heterologous host [75]. These and other synthetic biology techniques can be used for the activation of cryptic metabolic routes; the final goal of synthetic biology in this context is to be able to design molecules by combining different regulatory elements and biosynthetic genes (building blocks) [76, 77]. As such, it has the potential to allow for the refactoring as well as further engineering of cryptic gene clusters such as those studied in this paper. Powered by such technologies, the secondary metabolism from *Rhodococcus* can be used to identify targets to fight related pathogenic strains as well as to identify and engineer novel bioactive natural products.

Acknowledgements:

AC was financially supported by the University of Groningen, and by the Dutch Technology Foundation (STW), which is part or the Netherlands organization for Scientific Research (NWO) and partly funded by the Ministry of Economic Affairs (STW 10463). MHM is supported by VENI grant 863.15.002 from The Netherlands Organization for Scientific Research (NWO).

Supplementary information

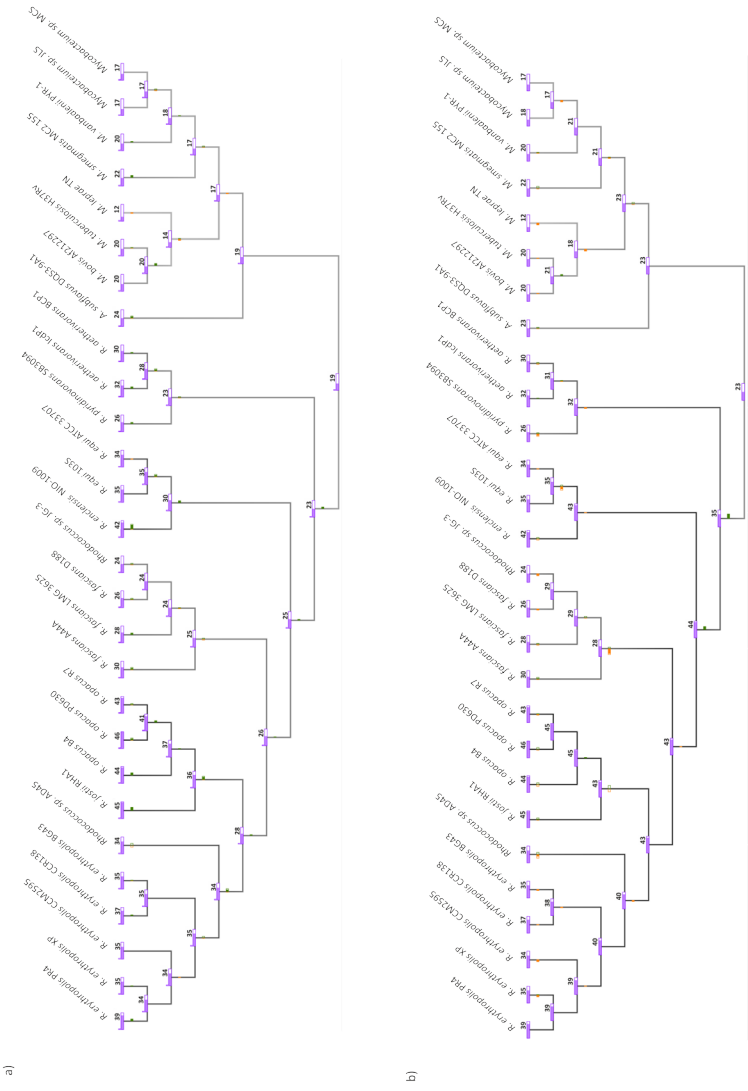


Figure S1. Gain/loss diagram. Clusters shared with more than 7 strains and all NRPS were analyzed for presence or absence in each strain and represented in the phylogenetic tree. Twenty-three GFCs are present in *Rhodococcus, Amycolicicoccus and Mycobacterium* strains (lower part of the dendogram). Horizontal lines under each node indicate the number of gained or lost clusters. Green to the right gain, yellow to the left loss. The purple bar on each node represents the proportion of clusters present in each strain from the total 113 clusters analyzed. a) Analysis made by Wagner parsimony with gain penalty=1. b) analysis made by Dollo parsimony.

Figure S2. Phylogenetic analysis of the acyl CoA-ligases/synthetases detected in this study using described acyl CoA-ligases/synthetases as references. The Neighbour-joining method was used with 1000 bootstrap replicates. Bootstrap values are given as percentage. FAAL: Fatty acid AMP Ligases, FACS: Fatty acids Acyl-CoA Synthetases, FACL: Fatty acid Acyl-CoA ligases FATP: Fatty acid transport protein, MACS: medium-chain Acyl-CoA synthetase. BACL: Bile acid CoA ligases.

Table S1 GCF name and its corresponding number in antiSMASH results from *R. jostii* unless stated**.**

GCF	Cluster number from antiSMASH results in *R. jostii* unless stated
NRPS-1	64
NRPS-2	*R. equi* ATCC33707 60
Requibactin IupT/IupS	*R. equi* ATCC33707 26
Heterobactin/Rhodochelin/Requichelin	71
NRPS-3	*R. erythropolis* R138 43
NRPS-4	*R. erythropolis* R138 39
Humimycins	*R. erythropolis* R138 44
NRPS-5 (similar to humimycin gene cluster)	*R. erythropolis* R138 46
Mycobactin	*M. tuberculosis* 32 Mycobactin
NRPS-6	93
NRPS-7	*R. erythropolis* R138 47
NRPS-8	*M. tuberculosis* 18,
NRPS-9	2
NRPS-10	33
NRPS-11	34
NRPS-12	35
NRPS-13	75
NRPS-14	91
NRPS-15	*M. tuberculosis* 37
NRPS-16	*M. smegmatis* 11
NRPS-17	*R. equi* ATCC33707 52
NRPS-18	*R. fascians LMG 3625* 32
NRPS-19	*R. fascians LMG 3625* 7,
NRPS-20	4
NRPS-21	36
NRPS-22	*R. fascians LMG 3625* 10
NRPS-23	*R. opacus PD630* 112
NRPS-24	1
NRPS-25	*A. subflavus* 64 and 68
NRPS-26	*M. tuberculosis* 26
NRPS-27	*M. tuberculosis* 50
NRPS-28	*M. tuberculosis* 1
NRPS-29	*R. equi* ATCC33707 49
NRPS-30	*R. equi* ATCC33707 7
NRPS-31	*R. fascians* A44A 69
NRPS-32	*R. fascians LMG 3625* 54
NRPS-33	*R. fascians* LMG3625 65
NRPS-34	*R. opacus PD630 27*
Rhequibactin IupU	*R. equi* ATTCC33707 11
NRPS-37	*R. aetherivorans* 48
NRPS-38	3

NRPS-39	82
NRPS-40	94
NRPS-41	106
NRPS-42	*A. subflavus* 57
Exochelin	*M. smegmatis* 1
NRPS-44	*M. smegmatis* 67
NRPS-45	*M. smegmatis* 86
NRPS-46	*M. smegmatis* 89
NRPS-48	*M. vanbaalenii* PYR-1 82
NRPS-49	*R. aetherivorans* 65
NRPS-50	*R. enclensis* 103
NRPS-51	*R. enclensis* 73
NRPS-52	*R. erythropolis* XP 71
NRPS-53	*R. fascians* A44A 49
NRPS-54	*R. fascians* A44A 63
NRPS-55	*R. fascians* A44A 65
NRPS-56	*R. fascians* A44A 66
NRPS-57	*R. fascians LMG 3625* 3
NRPS-58	*R. fascians LMG 3625* 38
NRPS-59	*R. opacus* B4 4
NRPS-60	*R. opacus* B4 42
NRPS-61	*R. opacus* B4 97
NRPS-62	*R. opacus* PD630 28
NRPS-63	*R. opacus* PD630 36
NRPS-64	*R. opacus* PD630 41
NRPS-65	*R. opacus* PD630 46
NRPS-66	*R. opacus* R7 34
NRPS-67	*R. pyridinovorans* 17
NRPS-68	*R. pyridinovorans* 7
NRPS-69	*Rhodococcus sp*. AD45 10
NRPS-70	*Rhodococcus sp*. AD45 52
NRPS-71	*Rhodococcus sp*. AD45 6
NRPS-72	*Rhodococcus sp*. AD45 88
NRPS-73	*Rhodococcus sp*. AD45 9
NRPS-74	*Rhodococcus sp* BCP1 57
NRPS-75	*Rhodococcus sp* BCP1 86
NRPS-76	*R. enclensis* 97
NRPS-77	*R. opacus* B4 2 plasmid
Saccharide-1 -Terpene(carotene)	11 Carotene not in Mycobacterium
Fatty acid or synthase (FAS)	17
Saccharide-2	27
Other-1	29
Terpene	101
Mycolic acid t1PKS and synthesis of arabinogalactan	60 and 61
Saccharide-3- (Mycofactocin RiPP)	*R. erythropolis* R138 28
Ectoine	14

Other-2	97 and 98
Fatty acid	77
Saccharide-4	89
other-3	12
Other-4	54
Other-5	96
Other-6	90
Other-7	9 and 10
Other-8	25 and 26
Other-9	*R. aetherivorans* cluster 54
Other-10	62
Other-11	38
Other-12	16
Saccharide-5	*R. equi* ATCC33707 56
Other-13	92
Other-14	*R. equi* ATCC33707 30
Other-15	19
Other-16	30
Other-17	68
Other-18	*R. erythropolis* R138 64
Other-19	84
Saccharide-6	*R. erythropolis* CCM2595 55
Other-20	*R. erythropolis* R138 9
Other-21	74
Butyro+B2:B38lactone	69

References

1. Garneau-Tsodikova, S., Dorrestein, P. C., Kelleher, N. L. & Walsh, C. T. Protein assembly line components in prodigiosin biosynthesis: characterization of PigA,G,H,I,J. *J. Am. Chem. Soc.* **128**, 12600-12601 (2006).

2. Kumar, A., Vishwakarma, H. S., Singh, J., Dwivedi, S. & Kumar, M. Microbial pigments: production and their applications in various industries. *IJPCBS* **5**, 203-212 (2015).

3. Pal, P. K. *et al*. Crop-ecology and nutritional variability influence growth and secondary metabolites of *Stevia rebaudiana Bertoni*. *BMC Plant Biol.* **15**, 67-015-0457-x (2015).

4. Mousa, W. K. & Raizada, M. N. The diversity of anti-microbial secondary metabolites produced by fungal endophytes: an interdisciplinary perspective. *Front. Microbiol.* **4**, 65 (2013).

5. O'Connor, S. E. Engineering of secondary metabolism. *Annu. Rev. Genet.* **49**, 71-94 (2015).

6. Chater, K. F. *Streptomyces* inside-out: a new perspective on the bacteria that provide us with antibiotics. *Philos. Trans. R. Soc. Lond. , B, Biol. Sci.* **361**, 761-768 (2006).

7. Parida, S. K. *et al*. Totally drug-resistant tuberculosis and adjunct therapies. *J. Intern. Med.* **277**, 388-405 (2015).

8. Wietz, M., Månsson, M., Vynne, N. G. & Gram, L. in *Marine Microbiology: Bioactive Compounds and Biotechnological Applications* (ed Kim, S. K.) (Wiley-VCH, Weinheim, 2013).

9. Medema, M. H. *et al*. The sequence of a 1.8-Mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. *Genome Biol. Evol.* **2**, 212-224 (2010).

10. Bentley, S. D. *et al*. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141-147 (2002).

11. Cimermancic, P. *et al*. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412-421 (2014).

12. Doroghazi, J. R. & Metcalf, W. W. Comparative genomics of Actinomycetes with a focus on natural product biosynthetic genes. *BMC Genomics* **14**, 611-2164-14-611 (2013).

13. McLeod, M. P. *et al*. The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 15582-15587 (2006).

14. Di Gennaro, P., Rescalli, E., Galli, E., Sello, G. & Bestetti, G. Characterization of *Rhodococcus opacus* R7, a strain able to degrade naphthalene and o-xylene isolated from a polycyclic aromatic hydrocarbon-contaminated soil. *Res. Microbiol.* **152**, 641-651 (2001).

15. Bosello, M., Robbel, L., Linne, U., Xie, X. & Marahiel, M. A. Biosynthesis of the siderophore rhodochelin requires the coordinated expression of three independent gene clusters in *Rhodococcus jostii* RHA1. *J. Am. Chem. Soc.* **133**, 4587-4595 (2011).

16. Miranda-Casoluengo, R. *et al*. The hydroxamate siderophore rhequichelin is required for virulence of the pathogenic actinomycete *Rhodococcus equi*. *Infect. Immun.* **80**, 4106-4114 (2012).

17. Bosello, M. *et al*. Structural characterization of the heterobactin siderophores from *Rhodococcus erythropolis* PR4 and elucidation of their biosynthetic machinery. *J. Nat. Prod.* **76**, 2282-2290 (2013).

18. Miranda-CasoLuengo, R., Prescott, J. F., Vázquez-Boland, J. A. & Meijer, W. G. The intracellular pathogen *Rhodococcus equi* produces a catecholate siderophore required for saprophytic growth. *J. Bacteriol.* **190**, 1631-1637 (2008).

19. Borisova, R. B. Isolation of a *Rhodococcus* soil bacterium that produces a strong antibacterial compound. *Electronic Theses and Dissertations. East Tennessee State University* **Paper 1388** (2011).

20. Iwatsuki, M. *et al*. Lariatins, antimycobacterial peptides produced by *Rhodococcus sp*. K01-B0171, have a lasso structure. *J. Am. Chem. Soc.* **128**, 7486-7491 (2006).

21. Kunze, B., Höfle, G. & Reichenbach, H. The aurachins, new quinoline antibiotics from myxobacteria: production, physico-chemical and biological properties. *J. Antibiot. (Tokyo)* **40**, 258-265 (1987).

22. Kitagawa, W. & Tamura, T. A quinoline antibiotic from *Rhodococcus erythropolis* JCM 6824. *J. Antibiot. (Tokyo)* **61**, 680-682 (2008).

23. Chiba, H., Agematu, H., Sakai, K., Dobashi, K. & Yoshioka, T. Rhodopeptins, novel cyclic tetrapeptides with antifungal activities from *Rhodococcus* sp. III. Synthetic study of rhodopeptins. *J. Antibiot. (Tokyo)* **52**, 710-720 (1999).

24. Chu, J. *et al*. Discovery of MRSA active antibiotics using primary sequence from the human microbiome. *Nat. Chem. Biol.* **12**, 1004-1006 (2016).

25. Wang, Y. N. *et al*. *Amycolicicoccus subflavus* gen. nov., sp. nov., an Actinomycete isolated from a saline soil contaminated by crude oil. *Int. J. Syst. Evol. Microbiol.* **60**, 638-643 (2010).

26. van der Geize, R. *et al*. A novel method to generate unmarked gene deletions in the intracellular pathogen *Rhodococcus equi* using 5-fluorocytosine conditional lethality. *Nucleic Acids Res.* **36**, e151 (2008).

27. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725-2729 (2013).

28. Csurös, M. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* **26**, 1910-1912 (2010).

29. Hamid, M. E. *et al*. *Nocardia africana* sp. nov., a new pathogen isolated from patients with pulmonary infections. *J. Clin. Microbiol.* **39**, 625-630 (2001).

30. Rainey, F. A., Burghardt, J., Kroppenstedt, R. M., Klatte, S. & Stackebrandt, E. Phylogenetic analysis of the genera *Rhodococcus* and *Nocardia* and evidence for the evolutionary origin of the genus *Nocardia* from within the radiation of *Rhodococcus* species. *Microbiology* **141**, 523-528 (1995).

31. Weber, T. *et al*. antiSMASH 3.0–a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43**, W237-43 (2015).

32. Medema, M. H. *et al*. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39**, W339-46 (2011).

33. Batut, B., Knibbe, C., Marais, G. & Daubin, V. Reductive genome evolution at both ends of the bacterial population size spectrum. *Nat. Rev. Microbiol.* **12**, 841-850 (2014).

34. Moran, N. A. Microbial minimalism: genome reduction in bacterial pathogens. *Cell* **108**, 583-586 (2002).

35. Sritharan, M. Iron homeostasis in *Mycobacterium tuberculosis*: mechanistic insights into siderophore-mediated iron uptake. *J. Bacteriol.* **198**, 2399-2409 (2016).

36. Hernández, M. A. *et al*. Biosynthesis of storage compounds by *Rhodococcus jostii* RHA1 and global identification of genes involved in their metabolism. *BMC Genomics* **9**, 600-2164-9-600 (2008).

37. Weber, S. S., Polli, F., Boer, R., Bovenberg, R. A. & Driessen, A. J. Increased penicillin production in *Penicillium chrysogenum* production strains via balanced overexpression of isopenicillin N acyltransferase. *Appl. Environ. Microbiol.* **78**, 7107-7113 (2012).

38. Giovannoni, S. J., Cameron Thrash, J. & Temperton, B. Implications of streamlining theory for microbial ecology. *ISME J.* **8**, 1553-1565 (2014).

39. Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D. H. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.* **7**, 78 (2007).

40. Marrakchi, H., Lanéelle, M. A. & Daffé, M. Mycolic acids: structures, biosynthesis, and beyond. *Chem. Biol.* **21**, 67-85 (2014).

41. Carran, C. J., Jordan, M., Drechsel, H., Schmid, D. G. & Winkelmann, G. Heterobactins: A new class of siderophores from *Rhodococcus erythropolis* IGTS8 containing both hydroxamate and catecholate donor groups. *Biometals* **14**, 119-125 (2001).

42. Choi, S. K., Harada, H., Matsuda, S. & Misawa, N. Characterization of two beta-carotene ketolases, CrtO and CrtW, by complementation analysis in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **75**, 1335-1341 (2007).

43. Haft, D. H. Bioinformatic evidence for a widely distributed, ribosomally produced electron carrier precursor, its maturation proteins, and its nicotinoprotein redox partners. *BMC Genomics* **12**, 21-2164-12-21 (2011).

44. Khaliullin, B. *et al*. Mycofactocin biosynthesis: modification of the peptide MftA by the radical S-adenosylmethionine protein MftC. *FEBS Lett.* (2016).

45. Kato, J. Y., Funa, N., Watanabe, H., Ohnishi, Y. & Horinouchi, S. Biosynthesis of gamma-butyrolactone autoregulators that switch on secondary metabolism and morphological development in *Streptomyces*. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 2378-2383 (2007).

46. Lanéelle, M. A. *et al*. A novel mycolic acid species defines two novel genera of the *Actinobacteria, Hoyosella* and *Amycolicicoccus*. *Microbiology* **158**, 843-855 (2012).

47. Bandyopadhyay, S., Chandramouli, K. & Johnson, M. K. Iron-sulfur cluster biosynthesis. *Biochem. Soc. Trans.* **36**, 1112-1119 (2008).

48. Martinelli, D. J. & Pavelka, M. S.,Jr. The RipA and RipB peptidoglycan endopeptidases are individually nonessential to *Mycobacterium smegmatis*. *J. Bacteriol.* **198**, 1464-1475 (2016).

49. Dhiman, R. K. *et al*. Menaquinone synthesis is critical for maintaining mycobacterial viability during exponential growth and recovery from non-replicating persistence. *Mol. Microbiol.* **72**, 85-97 (2009).

50. Moran, A. P. in *Microbial Glycobiology: Structures, Relevance and Applications* (eds Holst, O., Brennan, P. J. & Itzstein, M.) 1036 (Elsevier, USA, 2009).

51. Bursy, J. *et al*. Synthesis and uptake of the compatible solutes ectoine and 5-hydroxyectoine by *Streptomyces coelicolor* A3(2) in response to salt and heat stresses. *Appl. Environ. Microbiol.* **74**, 7286-7296 (2008).

52. Francis, I. *et al*. pFiD188, the linear virulence plasmid of *Rhodococcus fascians* D188. *Mol. Plant Microbe Interact.* **25**, 637-647 (2012).

53. Schué, M. *et al*. Two cutinase-like proteins secreted by *Mycobacterium tuberculosis* show very different lipolytic activities reflecting their physiological function. *FASEB J.* **24**, 1893-1903 (2010).

54. Rodriguez, G. M., Voskuil, M. I., Gold, B., Schoolnik, G. K. & Smith, I. *ideR*, an essential gene in *Mycobacterium tuberculosis*: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. *Infect. Immun.* **70**, 3371-3381 (2002).

55. Boland, C. A. & Meijer, W. G. The iron dependent regulatory protein IdeR (DtxR) of *Rhodococcus equi*. *FEMS Microbiol. Lett.* **191**, 1-5 (2000).

56. Jastrab, J. B. & Darwin, K. H. Bacterial Proteasomes. *Annu. Rev. Microbiol.* **69**, 109-127 (2015).

57. Strieker, M., Tanović, A. & Marahiel, M. A. Nonribosomal peptide synthetases: structures and dynamics. *Curr. Opin. Struct. Biol.* **20**, 234-240 (2010).

58. Miranda-Casoluengo, R. *et al*. The iron-regulated *iupABC* operon is required for saprophytic growth of the intracellular pathogen *Rhodococcus equi* at low iron concentrations. *J. Bacteriol.* **187**, 3438-3444 (2005).

59. Vergnolle, O., Xu, H. & Blanchard, J. S. Mechanism and regulation of mycobactin fatty acyl-AMP ligase FadD33. *J. Biol. Chem.* **288**, 28116-28125 (2013).

60. van der Geize, R., Grommen, A. W., Hessels, G. I., Jacobs, A. A. & Dijkhuizen, L. The steroid catabolic pathway of the intracellular pathogen *Rhodococcus equi* is important for pathogenesis and a target for vaccine development. *PLoS Pathog.* **7**, e1002181 (2011).

61. Silver, L. L. Challenges of antibacterial discovery. *Clin. Microbiol. Rev.* **24**, 71-109 (2011).

62. Aranda, F. J., Teruel, J. A. & Ortiz, A. Further aspects on the hemolytic activity of the antibiotic lipopeptide iturin A. *Biochim. Biophys. Acta* **1713**, 51-56 (2005).

63. Raaijmakers, J. M., De Bruijn, I., Nybroe, O. & Ongena, M. Natural functions of lipopeptides from *Bacillus* and *Pseudomonas*: more than surfactants and antibiotics. *FEMS Microbiol. Rev.* **34**, 1037-1062 (2010).

64. Debono, M. *et al*. A21978C, a complex of new acidic peptide antibiotics: isolation, chemistry, and mass spectral structure elucidation. *J. Antibiot. (Tokyo)* **40**, 761-777 (1987).

65. Chooi, Y. H. & Tang, Y. Adding the lipo to lipopeptides: do more with less. *Chem. Biol.* **17**, 791-793 (2010).

66. Kraas, F. I., Giessen, T. W. & Marahiel, M. A. Exploring the mechanism of lipid transfer during biosynthesis of the acidic lipopeptide antibiotic CDA. *FEBS Lett.* **586**, 283-288 (2012).

67. Noordman, W. H., Wachter, J. H., de Boer, G. J. & Janssen, D. B. The enhancement by surfactants of hexadecane degradation by *Pseudomonas aeruginosa* varies with substrate availability. *J. Biotechnol.* **94**, 195-212 (2002).

68. van der Geize, R. & Dijkhuizen, L. Harnessing the catabolic diversity of Rhodococci for environmental and biotechnological applications. *Curr. Opin. Microbiol.* **7**, 255-261 (2004).

69. Alvarez, H. M. *et al*. Identification of phenyldecanoic acid as a constituent of triacylglycerols and wax ester produced by *Rhodococcus opacus* PD630. *Microbiology* **148**, 1407-1412 (2002).

70. Castro, A. R., Rocha, I., Alves, M. M. & Pereira, M. A. *Rhodococcus opacus* B4: a promising bacterium for production of biofuels and biobased chemicals. *AMB Express* **6**, 35-016-0207-y. Epub 2016 May 14 (2016).

71. Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L. & Sivonen, K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 9259-9264 (2014).

72. Bastian, M., Braun, T., Bruns, H., Röllinghoff, M. & Stenger, S. Mycobacterial lipopeptides elicit CD4+ CTLs in *Mycobacterium tuberculosis*-infected humans. *J. Immunol.* **180**, 3436-3446 (2008).

73. Seshadri, C., Turner, M. T., Lewinsohn, D. M., Moody, D. B. & Van Rhijn, I. Lipoproteins are major targets of the polyclonal human T cell response to *Mycobacterium tuberculosis*. *J. Immunol.* **190**, 278-284 (2013).

74. Creason, A. L., Davis, E. W.,2nd, Putnam, M. L., Vandeputte, O. M. & Chang, J. H. Use of whole genome sequences to develop a molecular phylogenetic framework for *Rhodococcus fascians* and the *Rhodococcus* genus. *Front. Plant. Sci.* **5**, 406 (2014).

75. Myronovskyi, M. & Luzhetskyy, A. Native and engineered promoters in natural product discovery. *Nat. Prod. Rep.* **33**, 1006-1019 (2016).

76. King, J. R., Edgar, S., Qiao, K. & Stephanopoulos, G. Accessing nature's diversity through metabolic engineering and synthetic biology. *F1000Res* **5**, 10.12688/f1000research.7311.1. eCollection 2016 (2016).

77. Smanski, M. J. *et al*. Synthetic biology to access and expand nature's chemical diversity. *Nat. Rev. Microbiol.* **14**, 135-149 (2016).

CHAPTER 5

Molecular characterization of a *Rhodococcus jostii* RHA1 γ-butyrolactone-like signalling molecule and its main biosynthesis gene *gblA*

Ana Ceniceros, Lubbert Dijkhuizen and Mirjan Petrusma

Microbial Physiology, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

Abstract

Rhodococcus are most well-known as catabolic powerhouses. *Rhodococcus* genome sequence analysis recently has revealed a surprisingly large (and unexplored) potential for the production of secondary metabolites as well. Also, putative γ-butyrolactone gene clusters have been identified in some rhodococci. These signalling molecules regulate secondary metabolism in *Streptomyces*. This work provides evidence for synthesis of a γ-butyrolactone-like molecule by rhodococci (RJB), the first report in non-*Streptomyces* species. The *Rhodococcus jostii* RHA1 RJB molecule was detected by a reporter system based on the γ-butyrolactone receptor protein (ScbR) of *Streptomyces coelicolor*. This RJB is structurally identical to 6-dehydro SCB2, predicted precursor of *S. coelicolor* γ-butyrolactone SCB2. The *R. jostii* RHA1 key RJB biosynthesis gene was identified (*gblA*): Deletion mutagenesis of *gblA* resulted in complete loss of RJB synthesis whereas higher RJB levels were detected when *gblA* was overexpressed. Interaction of the RJB molecule with the *Streptomyces* γ-butyrolactone receptor system may indicate that communication occurs between these two Actinomycete genera in their natural habitat. Furthermore, RJB may provide a highly relevant tool for the awaking of cryptic secondary metabolic gene clusters in rhodococci. This study provides preliminary evidence that *R. jostii* RHA1 indeed synthesizes diffusible molecules with antimicrobial activity.

Introduction

Rhodococcus is a genus of aerobic, acid resistant, non-sporulating, Gram-positive soil bacteria (family Nocardiaceae, order Actinomycetales) which contain mycolic acids in their cell walls[1]. This genus is well-known for its catabolic versatility [2-5], but little is known about its secondary metabolism. Computational analysis has shown that this genus has a great potential for synthesis of secondary metabolites [4, 6-9, Chapter 4]. Analysis of several Actinomycete genomes, including 4 strains of the genus *Rhodococcus*, *R. jostii* RHA1*, R. equi* 103S*, R. opacus* B4 and *R. erythropolis* PR4, uncovered a relatively high percentage of genes encoding nonribosomal peptide synthetases (NRPS) in rhodococci. Also, conserved γ-butyrolactone biosynthesis gene clusters were identified in these rhodococci [8]. The physiological roles of γ-butyrolactones have been extensively studied in members of the genus *Streptomyces* only, although their putative biosynthetic genes also appear to be present in other Actinomycete genera [10]. These signalling molecules are known to participate in the regulation of secondary metabolism and to induce a range of physiological responses [11-14]. γ-Butyrolactones bind to one or more receptor proteins which belong to the TetR family of transcriptional regulators. Several of these receptor proteins have been described in the genus *Streptomyces*, involved in controlling expression of secondary metabolite gene clusters. In most cases, they act as repressors, and their binding to DNA thus results in blocking the expression of the target genes. Binding of γ-butyrolactone ligands induces a change in receptor protein conformation, releasing the repression they were exerting.

Three main groups of γ-butyrolactones have been described to date, classified according to their structures: A-factor type, which contains a keto group in the carbon 6 of the molecule; IM-2, with a hydroxyl group in the same carbon in the R configuration; and VB type, also with a hydroxyl group in the carbon 6 but in an S configuration [15-19] (Fig. 1).

Within each group there is further diversity depending on the structure of the acyl chain connected to carbon 6.

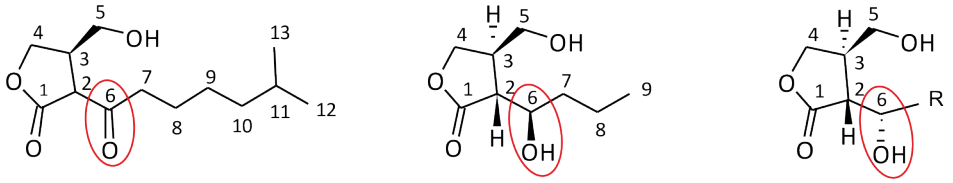


Figure 1. Illustration of the three γ-butyrolactone structural types described to date, differing in group and conformation at the C6 position. A-factor was first described in *Streptomyces griseus*[14]; IM-2 molecules were first identified in *Streptomyces lavendulae*[18]; VB molecules were first reported in *Streptomyces venezuelae*[16]. Adapted from Martin-Sanchez[20]

The enzymes involved in the biosynthesis of γ-butyrolactones have been described and partially characterized in several *Streptomyces* strains. They have been named according to the species that employ them or the compound that is regulated by these molecules. For unification purposes, we have renamed these genes so that they can be applied to any butanolide system in any strain. GblA (**g**amma-**b**utyro**l**actone biosynthesis **A**) catalyzes the first step of the biosynthesis by condensing a glycerol derivative with a fatty acid derivative (Fig. 2). This enzyme was first described in *S. griseus* where it was named AfsA (Fig. 2). After this step two different pathways have been predicted. Pathway A is believed to be catalyzed by non-specific enzymes. Pathway B includes a reductase GblC (**g**amma-**b**utyro**l**actone biosynthesis **C**), named BprA in *S. griseus* (Fig. 2). GblC is predicted to reduce the double bond in carbons 3 and 2 in the lactone ring (conversion of compound 4 into 5). In some species, there is also a short chain dehydrogenase (GblD, **g**amma-**b**utyro**l**actone biosynthesis **D**) that reduces the keto group in carbon 6 of 6-dehydro-γ-butyrolactone forms (compound 6 to compound 7 in Fig. 2).

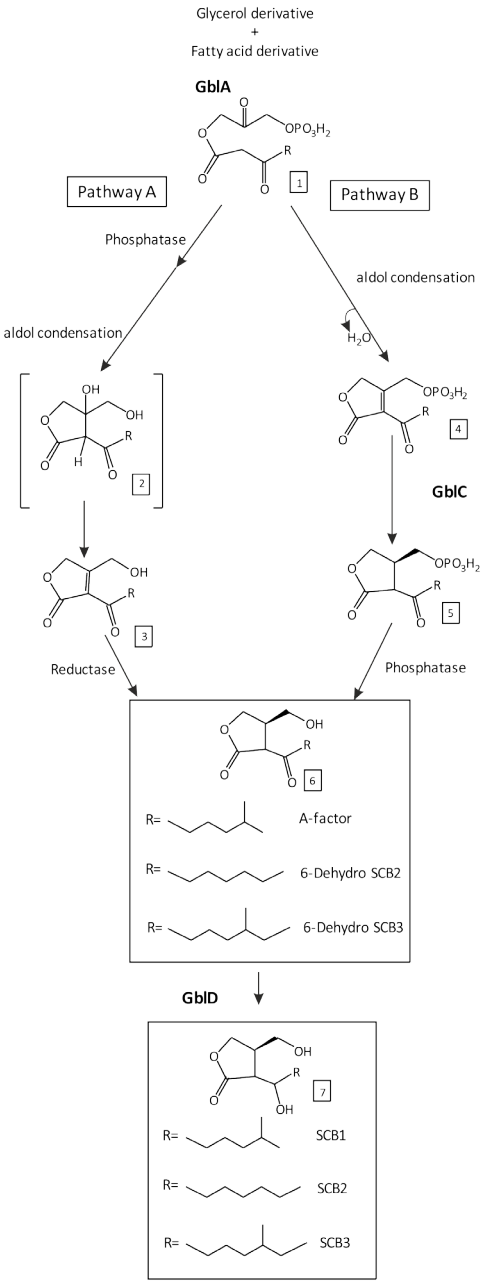


Figure 2. Predicted biosynthetic pathway of γ-butyrolactones in *Streptomyces* species (adapted from [14, 20, 21]) The generic names given to the enzymes in this work are shown at the (putative) steps that they catalyze. A-factor is the γ-butyrolactone from *S. griseus*. SCBs (**S**treptomyces **C**oelicolor **B**utyrolactones), γ-butyrolactones described in *Streptomyces coelicolor*.

The butanolide system has only been studied in *Streptomyces*; it has not been explored for interspecies communication although there is evidence that other genera also have this system [10]. In this work, we identified a γ-butyrolactone-like molecule in *R. jostii* RHA1, the first time that such a molecule has been identified in the genus *Rhodococcus*. A *R. jostii* RHA1 *gblA* deletion mutant did not induce the growth of strain LW16/pTE134, strain that can only grow on kanamycin in presence of γ-butyrolactones (table 1) indicating that the *gblA* gene is essential for RJB (***R****hodococcus* ***j****ostii* **b**utyrolactone) synthesis. Moreover, an overproduction of this RJB was observed when *gblA* was overexpressed in *R. jostii* RHA1. LC-MS analysis of extracts from these *R. jostii* RHA1 strains indicated that the RJB molecule synthesized has the same structure as 6-dehydro SCB2, a stereoisomer of A-factor, the known γ-butyrolactone from *S. griseus* and a predicted precursor of SCB2 (Figs. 1, 2), a known γ-butyrolactone from *S. coelicolor* [20].

Materials and methods

Strains and growth conditions

All strains used in this study are described in Table 1 and the media used in Supplementary Table S1 online. *Rhodococcus* strains were grown at 30°C. Luria-Bertani agar from Sigma-Aldrich was used as a standard medium. Appropriate antibiotics were used at the following concentrations: apramycin 50 μg/ml, kanamycin 200 μg/ml. For γ-butyrolactone extraction, *Rhodococcus* strains were grown in modified SMM solid (SMMS) medium [22] (Supplementary Table S1 online) at 30°C.

Table 1. Microbial strains used in this work

Strain	Details	Reference
Rhodococcus jostii RHA1	Wild type	[4]
RHA1-OE	*R. jostii* RHA1 +pRM4-*gblA*	This work
RHA1-Δ*gblA*	*R. jostii* RHA1Δ*gblA*	This work
RHA1-C	*R. jostii* RHA1Δ*gblA*+pRM4-*gblA*	This work
S. coelicolor LW16/pTE134	γ-Butyrolactone reporter strain. *S. coelicolor* M145 Δ*scbA*Δ*scbR* (LW16) containing pTE134 (*scbR* and Km resistance gene under the control of a γ-butyrolactone inducible promoter (*cpkOp*)	[23]
S. coelicolor M145	Wild type strain of *S. coelicolor*	[24]
E. coli DH5α™	Cloning strain. F– Φ80*lacZ*ΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1	Invitrogen
Aspergillus niger N402	Bioactivity reporter strain. Wild type.	[25]
Micrococcus luteus ATCC9341/ *Kocuria rhizophila*	Bioactivity reporter strain. Wild type.	
Bacillus subtilis ATCC6633	Bioactivity reporter strain. Wild type.	
E. coli JM101	Bioactivity reporter strain F´ traD36 proA+B+ lacIq Δ(lacZ)M15/ Δ(lac-proAB) glnV thi	NEB
Mycobacterium smegmatis MC2 155	Bioactivity reporter strain. Wild type.	

Phenotypic characterization of the different strains

R. jostii RHA1 WT, RHA1-Δ*gblA* and RHA1-OE strains were grown on different solid media to check for phenotypic differences. Growth, pigmentation and antibiotic activity against *A. niger, M. smegmatis, E. coli, B. subtilis* and *K. rhizophila* on solid media were monitored by direct observation. Cell shapes were studied under a Zeiss Axioskop 2 phase contrast microscope. The medium components used are described in Supplementary Table S1 online. The media used for antibiotic production tests were Trypton Soya Agar (TSA), Difco Nutrient Agar (DNA), Luria Broth Agar (LBA), Starch Casein Agar (SCA), Minimum Salt Medium (MSM) nitrogen deficient and complimented with casamino acids, Supplemented Minimum Medium Solid (SMMS) and low pH SMMS: in this case pH was not adjusted after mixing components. *R. jostii* RHA1 strains were grown for 4 days before plating the reporter strains (Table 1) next to the *Rhodococcus* colonies. Growth of the reporter strains was followed in time and scored after 4 days.

Bioinformatic analysis

The following complete genome sequences of rhodococci were obtained from GenBank (http://www.ncbi.nlm.nih.gov/genbank/): *Rhodococcus aetherivorans* IcdP1, *Rhodococcus equi* 103S, *Rhodococcus equi* ATCC 33707, *Rhodococcus erythropolis* CCM2595, *Rhodococcus erythropolis* PR4, *Rhodococcus erythropolis* R138, *Rhodococcus opacus* B4, *Rhodococcus opacus* PD630, *Rhodococcus opacus* R7, *Rhodococcus pyridinivorans* SB3094, *Rhodococcus* sp. AD45, *Rhodococcus* sp. B7740, *Rhodococcus aetherivorans* BCP1. AntiSMASH [26] was used for the detection of γ-butyrolactone biosynthetic gene clusters in these genomes using the ClusterFinder algorithm.

Deletion mutagenesis

Unmarked gene deletion mutagenesis [27] was used to delete *gblA*. Primers were designed to amplify fragments of 1.5 Kb upstream and downstream of *gblA*, including the first and last ~200 bp from *gblA* to ensure that the

surrounding genes were not affected by the deletion. Both fragments were cloned in pK18mobSacB (ATCC® 87097™) between EcoRI and PstI, producing the deletion construct pK18mobSacB-Δ*gblA.* The deletion construct was transformed into *R. jostii* RHA1 by electrotransformation. The strain was checked by PCR with primers outside the 1.5 Kb homologous regions and by sequencing of the resulting product.

gblA complementation and overexpression

Primers ScbA-jostii-NdeI-F and ScbA-jostii-BamHI-R were used to amplify the *R. jostii* RHA1 *gblA* gene (Table 2) and clone it into pRM4 [28] under the control of the strong constitutive promoter *ermE**(PMR4-*gblA*). This construct was introduced into *R. jostii* RHA1 wild type strain and RHA1-Δ*gblA* by electrotransformation obtaining the strains RHA1-OE and RHA1-C, respectively. Strains were checked by PCR with primers annealing in PMR4 at both sides of the insert and by sequencing of the resulting products.

Table 2. Primers used in this work

Primer	Sequence	Amplicon target
ScbA-jostii-NdeI-F	GCGATACATATGGCGCAAAT TTCCCGGCCGAT	*R. jostii gblA* gene
ScbA-jostii-BamHI-R	CGCTATGGATCCCTAGCGAG CGCATGCGCTCA	*R. jostii gblA* gene
afsA del GTG FW XbaI	GATTATCTAGAGAAGACCTC GGCCACGGATTG	Upstream region of *R. jostii gblA* gene for deletion
afsA del GTG Rv EcoRI	TACTTGAATTCGGGCTTTCGT GAACGACCTC	Upstream region of *R. jostii gblA* gene for deletion
afsA del UAG Rv XbaI	GATTATCTAGAGACGAGCGA GCCACGATCC	Downstream region of *R. jostii gblA* gene for deletion
afsA del UAG Fw PstI	GTCAACTGCAGGCCGGGCGA GATCGTTCAC	Downstream region of *R. jostii gblA* gene for deletion

Transformation of *Rhodococcus* strains

A modified electrotransformation protocol from Arenskotter *et al*. [29] was used to introduce the different constructs described in this work into

R. jostii RHA1. Strains were grown in 50 ml LB containing 1% w/v of glycine in a 250 ml Erlenmeyer flask at 30°C and 220 rpm to an OD600 of 0.8-1. Cells were washed twice with 15 ml of chilled deionized water and concentrated to 2.5 ml 10% glycerol and aliquoted in 400 μl. Subsequently, 100 ng to 1 μg of DNA was added to each 400 μl and the sample kept on ice for at least 10 min. Cells were pulsed with a Biorad Xcell gene pulser at 1.75 kV, 50 μF and 200 Ω (field strength of 8.75 kV cm^{-1}). Ice cold LB was added immediately after the pulse and the cell samples were allowed to recover for 4 h at 30°C and 220 rpm. Subsequently the cells were plated on selective media.

γ-Butyrolactone extraction

Extraction of γ-butyrolactones was performed following the procedure described in Hsiao *et al*. [23]. *R. jostii* RHA1 WT and mutant strains derived were grown on modified SMMS [22] (Table S1). Per strain, 40 standard (90 mm diameter) petri dishes were used. After 4 days of growth at 30 °C, when the strains started to produce carotenes, indication of an active secondary metabolism, the agar of each plate was cut into pieces and extracted as described in Hsiao *et al*. [23]. Extracts were dried at 30 °C in a rotary evaporator, and then resuspended in 160 μl of methanol per 40 petri dishes.

Kanamycin assay

The Km assays performed in this study were done following the protocol from Hsiao *et al*. [23]. From the extract of each strain, 60 μl were concentrated to 6 μl and spotted onto a DNA plate (Supplementary Table S1 online) containing 4.5 μg ml^{-1} of kanamycin and uniformly inoculated with *S. coelicolor* strain LW16/pTE134. As positive control, 6 μl of a stock solution of 1.5 mg ml^{-1} of chemically synthesized 6-dehydro SCB2 was used. Dried ethyl acetate resuspended in methanol was used as negative control. Results were reproducible with 2-3 biological replicates for each strain.

Liquid chromatography-Mass spectrometry analysis

For identification of *R. jostii* RHA1 γ-butyrolactone-like molecules, HPLC-MS analysis was performed using an Accella1250™ HPLC system coupled with the benchtop ESI-MS Orbitrap Exactive™ (Thermo Fisher Scientific, San Jose, CA). A Reversed Phase C18 (Shim Pack Shimadzu XR-ODS 3x75 mm) column was used and a gradient from 2% to 95% of acetonitrile:water (0.1% Formic Acid) as follows: 2 min 2% acetonitrile, 2-10 min gradient to 95% acetonitrile, 1 min 95% acetonitrile. To separate further the peaks from A-factor and 6-dehydro SCB2, a gradient from 2% to 80% acetonitrile was applied to the separation: 2 min 2% acetonitrile, 2-25 min in 2-80% acetonitrile, 1 min 80% acetonitrile. Data was analyzed using Xcalibur software from Thermo Scientific. LC-MS analysis was performed with 2-4 biological replicates per strain.

Synthesis of γ-butyrolactone standards

The γ-butyrolactones used in this study were chemically synthesized as described by Martin-Sanchez [20].

Analysis of the interactions between *Streptomyces coelicolor* and *R. jostii* RHA1

R. jostii RHA1 wild type, strains RHA1-OE and RHA1-Δ*gblA* were plated onto modified minimum salt media (MSM) containing casamino acids instead of NH_4NO_3 (see Supplementary Table S1 online). After 4 days, *S. coelicolor* M145 was plated next to the patches of the *R. jostii* strains. Following a further incubation for 18 h, the *S. coelicolor* growth stage and production of coloured antibiotics was checked every 2 h; after 24 h of incubation these parameters were checked once a day for a week. Results were reproducible in 5 independent experiments.

Results

γ-Butyrolactone gene clusters in rhodococci

Analysis of the predicted γ-butyrolactone gene cluster of *R. jostii* RHA1 [8, 10] revealed the presence of a *gblA* gene (RHA1_RS22510) (Fig. 3a), encoding for GblA, the putative first enzyme in the biosynthetic pathway of γ-butyrolactones. It contains two AfsA repeats, the predicted active sites of GblA enzymes. $GblA_{jostii}$ has 37-41% AA identity with (partially) characterized homologues of *S. venezuelae* (JadW1), *S. coelicolor* (ScbA) and *S. griseus* (AfsA) (Fig. 3a). The homologues of these three *Streptomyces* species have 43-65% AA identity between each other. The γ-butyrolactone receptor protein (GblR) from *R. jostii* RHA1, annotated as a TetR regulator, has 34-36% AA identity with the corresponding proteins in the three *Streptomyces* strains; the homologues of the three *Streptomyces* species have 37-56% AA identity between each other (Fig. 3a). The γ-butyrolactone gene cluster of *R. jostii* RHA1 also includes a GblE enzyme with a NAD-epimerase/dehydratase predicted function. A (partially) characterized homologue is JadW2 of *S. venezuelae* (35% AA identity), shown to be essential for the synthesis of γ-butyrolactones [30]. However, it is not clear in which step of the biosynthesis pathway this enzyme acts. BLAST searches were also performed with GblD of *Streptomyces coelicolor,* a short chain dehydrogenase known to contribute to synthesis of γ-butyrolactones in some *Streptomyces* species, e.g. in *S. coelicolor* [20] and *S. venezuelae* [31]. This yielded a large number of homologues with 30%-40% AA identity to the query, mostly spread throughout the *R. jostii* RHA1 genome, and with a few of them located in the proximity of the *gbl* gene cluster. Also, two homologues of GblC were found in the *R. jostii* RHA1 genome, with ~ 35% AA identity to the *S. coelicolor* GblC. These genes are not located in close proximity to the *gbl* gene cluster.

All studied *Rhodococcus* strains possess a predicted γ-butyrolactone gene cluster, except for *R. pyridinovorans*. *R. opacus* PD630 and *R. opacus* R7

contain multiple gene clusters (Fig. 3b) in their genomes. *R. opacus* R7 contains γ-butyrolactone clusters on 2 different plasmids while in all other cases the clusters are located on the chromosome only. Most γ-butyrolactone gene clusters have a similar organization, with the *gblR* gene flanked by *gblA* and *gblE* but divergently oriented. The *gblE* gene however is not always present (Fig. 3b).

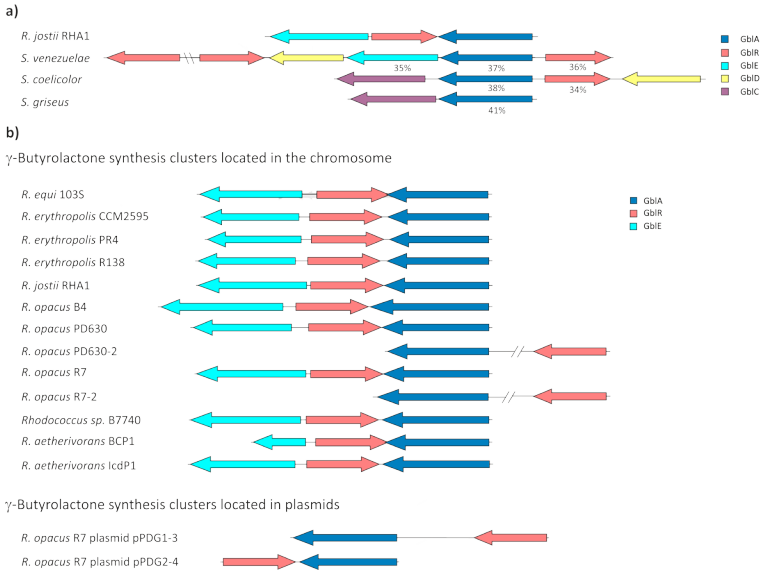


Figure 3. Predicted γ-butyrolactone gene clusters in different strains. a) Organization of the predicted γ-butyrolactone gene cluster of *Rhodococcus jostii* RHA1 compared to that of the known clusters in different *Streptomyces* strains. AA identity of the *R. jostii* RHA1 enzymes to the corresponding enzymes encoded by the genes in each strain is stated below each gene. b) Comparison of the organization of the different γ-butyrolactone gene clusters of rhodococci. A cluster is present in virtually all studied rhodococci, with a highly similar organization. Only *R. opacus* R7 contains γ-butyrolactone clusters on its plasmids. GblA: γ-butyrolactone first biosynthetic enzyme. GblR: γ-butyrolactone receptor protein. GblE: γ-butyrolactone biosynthetic enzyme E, predicted to be a NAD-epimerase/dehydratase. GblD: γ-butyrolactone biosynthetic enzyme D, short chain dehydrogenase. GblC: γ-butyrolactone biosynthetic enzyme C, reductase.

To analyse whether *R. jostii* RHA1 is producing any γ-butyrolactone-like molecules, a detection assay developed for *S. coelicolor* was performed with ethyl acetate extracts of *R. jostii* RHA1 agar plates [23]. This test is based on release of the repression by the γ-butyrolactone receptor ScbR

of transcription of a Km resistance gene in the *S. coelicolor* LW16/pTE134 indicator strain. When plating this reporter strain on solid media with Km, it will only be able to grow if γ-butyrolactone molecules are present that are able to bind to the ScbR receptor protein and thereby allow transcription of the Km resistance gene. The *R. jostii* RHA1 extracts obtained, as described in the methods section, indeed reproducibly induced the growth of the LW16/pTE134 reporter strain (Fig. 4a). *R. jostii* RHA1 thus indeed synthesizes γ-butyrolactone-like molecules (RJB). This Km bioassay is known to be very specific, since changes in the γ-butyrolactone aliphatic side chain are known to significantly affect the affinity of ScbR for these molecules [19]. Generic γ-butyrolactones also are not able to trigger this system [32]. *R. jostii* RHA1 apparently synthesizes one or more RJB molecules that are able to bind to the *S. coelicolor* γ-butyrolactone receptor protein ScbR, enabling growth of the reporter strain. In view of the high specificity of this assay it is likely that these *R. jostii* RHA1 RJB molecules structurally are most similar to SCBs, the γ-butyrolactone molecules of *S. coelicolor* [17, 19].

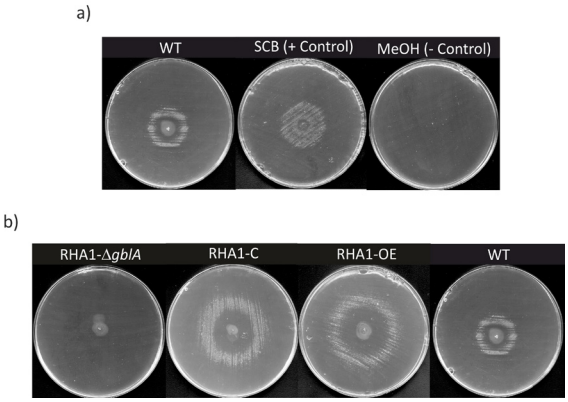


Figure 4. Detection of γ-butyrolactone-like molecules in *R. jostii* RHA1 ethyl acetate agar culture extracts by Km bioassay. When γ-butyrolactones are present in the sample the reporter strain *S. coelicolor* LW16/pTE134 forms a halo of growth around the sample application area in the centre of kanamycin agar plates. Size of the halo of growth is indicative for the concentration of the diffusible γ-butyrolactones in the sample[23]. a) *R. jostii* RHA1 ethyl acetate extract (left), positive control with chemically synthesized 6-dehydro SCB2 (middle), negative control (solvent) (right). b)

R. jostii-Δ*gblA* (left), complemented strain (RHA1-C), *gblA* overexpression strain (RHA1-OE) and wild type strain (right).

To further analyse the γ-butyrolactone system of *R. jostii* RHA1 the putative γ-butyrolactone biosynthesis gene *gblA* was deleted from the genome using unmarked gene deletion mutagenesis [27]. Ethyl acetate extracts of the deletion strain RHA1-Δ*gblA* were made and tested for RJB synthesis. In contrast to the wild type *R. jostii* RHA1, extracts of RHA1-Δ*gblA* did not induce the growth of the LW16/pTE134 reporter strain in the Km bioassay, indicating the absence of γ-butyrolactones (Fig. 4b). No other phenotypical differences were observed for this strain compared to the WT strain. The RHA1-Δ*gblA* deletion strain was fully complemented in RJB production by reintroducing the wild type *gblA* gene (strain RHA1-C) (Fig. 4b). The complementation was performed by inserting PMR4 containing $gblA_{jostii}$ under the control of a constitutive strong promoter (*ermE**) (PMR4-*gblA*), which resulted in a higher yield of RJB than the wild type strain (bigger halo of growth, see Fig. 4b). The role of GblA in RJB production was further studied by construction of the overexpression strain RHA1-OE, introducing PMR4-*gblA* into wild type *R. jostii* RHA1. Next, RJB synthesis by the RHA1-OE strain was analyzed. Also in this case the Km bioassay showed a bigger halo of growth of the reporter strain compared to wild type *R. jostii* RHA1 extracts, which indicates a higher RJB concentration in the RHA1-OE sample due to a further diffusion from the application point. Thus, overexpression of *gblA* results in an enhanced RJB production (Fig. 4b).

Characterization of the γ-butyrolactone-like molecules synthesized by *R. jostii* RHA1

The extracts of wild type *R. jostii* RHA1, RHA1-Δ*gblA*, RHA1-OE and RHA1-C were analyzed for RJBs by liquid chromatography coupled to a mass spectrometer (LC-MS). The structures of RJBs are apparently rather similar to those of the *S. coelicolor* γ-butyrolactones since they bind to ScbR, as evident from the Km bioassay (see above). Therefore, the LC-MS

data was analyzed searching for metabolites with masses similar to those of known *S. coelicolor* γ-butyrolactones. A peak eluting at 7.70 min with a mass of m/z 241.1441 amu $[M-H]^-$ was detected in the extracts from *R. jostii* RHA1 wild type strain, RHA1-OE and RHA1-C, but it was missing in RHA1-Δ*gblA*. The *gblA* gene thus is essential for its synthesis, indicating that this peak indeed represents a *R. jostii* RHA1 RJB molecule (Fig. 5a). The mass of the detected *R. jostii* RHA1 RJB molecule corresponds to the A-factor signalling molecule of *S. griseus* and also to the intermediate compound 6-dehydro SCB2 of *S. coelicolor* [20] (Fig. 2, 5a). Synthetic standards of both molecules were also analyzed on LC-MS (Fig. 5a). The extracts of the different *Rhodococcus* strains yielded peaks with the same retention time as 6-dehydro SCB2. The RHA1-OE and RHA1-C peaks had a higher intensity than in the *R. jostii* RHA1 wild type strain, corresponding to the Km bioassay results, showing a bigger halo than seen with the wild type strain (Fig. 4b). To analyse whether the molecule detected in the *R. jostii* RHA1 extracts is similar to 6-dehydro SCB2 or to A-factor, the extract from RHA1-OE was spiked with the synthetic standards of these compounds and run in the LC-MS with a longer gradient (Fig. 5b). As a control, a mixture of both standards (A-factor and 6-dehydro SCB2) was also run in the same conditions. The mixture of both standards and the extract spiked with the standard of A-factor showed two different peaks at 14.86 min and 15.17 min, corresponding to A-factor and 6-dehydro SCB2, respectively. When the extract was spiked with 6-dehydro SCB2 the peaks completely overlapped, confirming that the single RJB detected is structurally identical to 6-dehydro SCB2 (Fig. 5b). The *R. jostii* RHA1 samples were also compared to the available chemically synthesized standards of *S. coelicolor* γ-butyrolactones [20], but we were not able to find any other known γ-butyrolactone-like molecules in *R. jostii* RHA1 extracts. When the samples were screened for masses between 187 and 350, a mass range that includes all described γ-butyrolactones, two peaks had a higher intensity in the RHA1-OE and RHA1-C strains than in the *R. jostii* RHA1 wild type; these two peaks were not visible in the RHA1-Δ*gblA*

deletion strain (see Fig. S1 online). One peak eluted at 7.08 min and showed three different masses, m/z 211.0972 amu $[M-H]^-$, m/z 279.1369 amu $[M-H]^-$ and m/z 289.1658 amu $[M-H]^-$. Another peak eluted at 7.88 min that corresponds to a mass of m/z 255.1236 amu $[M-H]^-$. None of these masses correspond to known γ-butyrolactones, including the recently described in Sidda *et al.* [33].

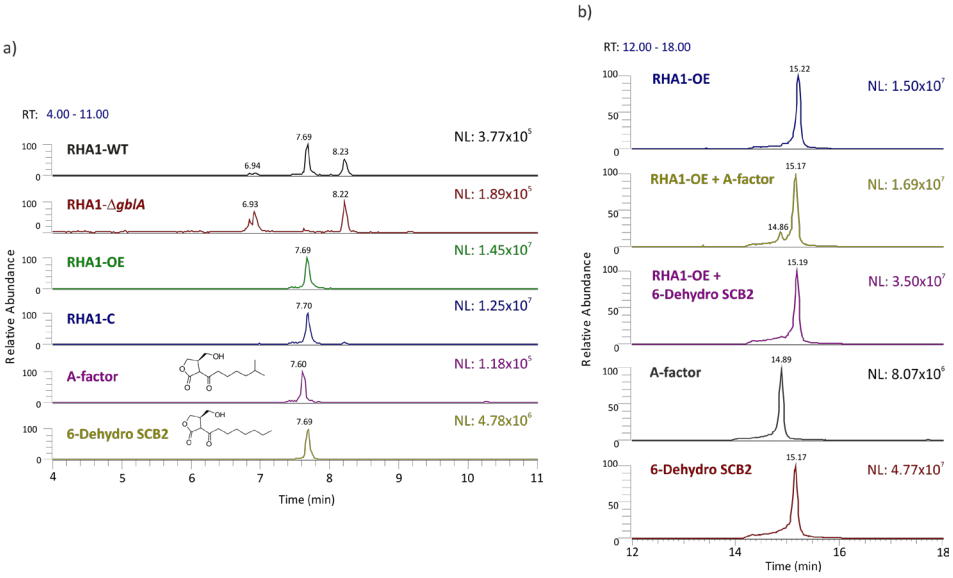


Figure 5. LC-MS analysis of ethyl acetate extracts of the various *R. jostii* RHA1 strains grown for 4 days on SMMS. a) A peak eluting at 7.70 min with a mass of mass m/z 241.1441 amu $[M-H]^-$ was detected in all samples except in RHA1-Δ*gblA.* RHA1-C, the complemented strain, and RHA1-OE, the *gblA* overexpression strain, both showed a higher intensity of this peak. This mass corresponds to the described γ-butyrolactone from *S. griseus* (A-factor) and the stereoisomer 6-dehydro SCB2, known to be an intermediate in the synthesis of the γ-butyrolactone SCB2 in *S. coelicolor*. The standard of A-factor showed a peak eluting at 7.60 min while the standard of 6-dehydro SCB2 eluted at 7.69 min. b) Extracts of RHA1-OE spiked with standards of A-factor or 6-dehydro SCB2 using a longer gradient to separate both peaks further. The spiked extract of *R. jostii* RHA1 with both standards confirmed that the molecule synthesized by *R. jostii* RHA1 has the same retention time and mass as 6-dehydro SCB2. (NL: Normalization Level; RT: Retention Time)

Phenotypical characterization of constructed *R. jostii gblA* strains

The γ-butyrolactone system is known to regulate secondary metabolite synthesis, morphogenesis or both, in streptomycetes [14, 34]. *R. jostii* RHA1 has almost 120 putative secondary metabolite biosynthetic gene clusters

in its genome and most of them are uncharacterized[4]. Wild type *R. jostii* RHA1 was screened for secondary metabolite production during growth on different agar media. Production of bioactive compounds was tested with various indicator strains, two Gram-positive strains (*Kocuria rhizophila* and *Bacillus subtilis*), one acid-resistant Gram-positive strain (*Mycobacterium smegmatis*), a Gram-negative strain (*Escherichia coli*) and a fungal species (*Aspergillus niger*). These reporter strains were plated next to the *Rhodococcus* colonies. *R. jostii* RHA1 WT exerted clear inhibition of growth towards *K. rhizophila* on SCA medium and *M. smegmatis* in low pH SMMS (Fig. 6). Inhibition of sporulation of *A. niger* was observed on LBA, TSA and DNA agar media (Fig. 6).

The RHA1-Δ*gblA* deletion strain and the RHA1-OE overexpression strain were also analyzed for phenotypic changes in different growth media. No difference in bioactivity was observed between *R. jostii* RHA1 wild type and derived strains in any medium or with any reporter strain tested. The growth rates and colony shapes of all constructed strains were also analyzed on all tested solid media, and their cell shapes in liquid LB medium, however, no differences were observed compared to the wild type strain (data not shown).

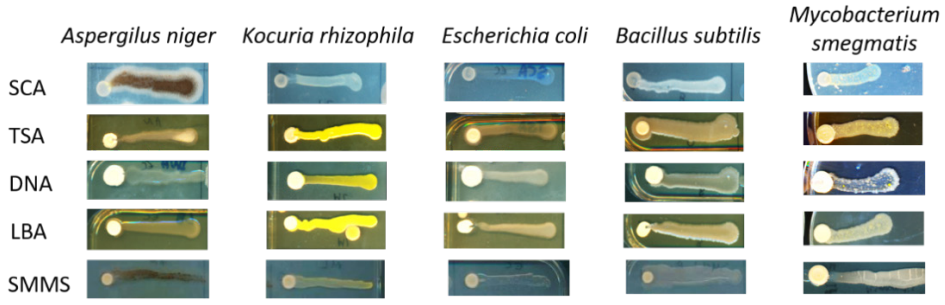


Figure 6. Bioactivity tests with *R. jostii* RHA1 spotted directly from glycerol stocks onto SCA, TSA, DNA, LBA and SMMS (pH 5.5) agar plates using *A. niger*, *K. rhizophila*, *E. coli*, *B. subtilis* and *M. smegmatis* as reporter strains. After 4 days of incubation, the reporter strains were applied on a horizontal line towards the *R. jostii* RHA1 patch. Inhibition of *A. niger* sporulation was observed on TSA, DNA and LBA (note loss of black pigment from the conidia). Growth inhibition of *K. rhizophila* and *M. smegmatis* was observed on SCA and SMMS, respectively (especially visible close to the *R. jostii* RHA1 patch at the left).

Interaction between *R. jostii* RHA1 and *S. coelicolor* M145

The Km bioassay performed with extracts from *R. jostii* RHA1 showed that the *R. jostii* RHA1 RJB interacts with the γ-butyrolactone receptor protein ScbR of *S. coelicolor* (see above). These different genera thus may be capable of interspecies communication. To study a possible interaction between *R. jostii* RHA1 and *S. coelicolor* M145, both strains were inoculated next to each other on agar plates. γ-Butyrolactones diffuse into the agar and therefore an exchange of signalling molecules between species is possible. *R. jostii* RHA1 was allowed to grow for 4 days, using carotene production as indication that secondary metabolism was active. Subsequently, *S. coelicolor* M145 was plated next to it. As a control, *S. coelicolor* M145 and *R. jostii* RHA1 also were plated separately on the same agar media. After a further 24 h, *S. coelicolor* M145 developed aerial mycelium with its characteristic white pigmentation when grown next to the RHA1-OE overexpression strain, but not when growing next to the RHA1-WT and the RHA1-Δ*gblA* strains (Fig. 7).

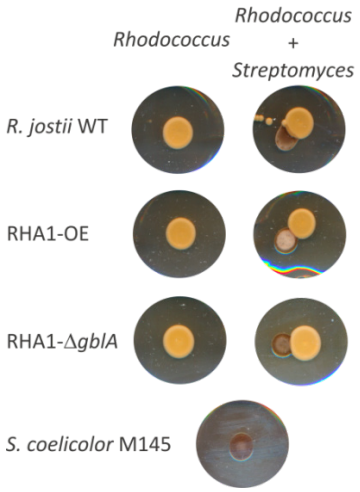


Figure 7. Interactions observed between *R. jostii* RHA1 WT, RHA1-OE and RHA1-Δ*gblA* with *S. coelicolor* M145 on MSM agar containing casamino acids. *Rhodococcus* strains were spotted directly from glycerol stocks and grown for 4 days before spores of *S. coelicolor* M145 were plated next to the patches of the *Rhodococcus* strains. On the left, the different strains from *R. jostii* RHA1 were grown separately. On the right, *S. coelicolor* M145 was grown next to the different *R. jostii* RHA1 strains. At the bottom, *S. coelicolor* M145 grown on its own. After 24 h incubation, *S. coelicolor* M145 started to develop a white colour characteristic for the formation of aerial mycelia when grown next to RHA1-OE, but not when grown on its own or with *R. jostii* RHA1 WT or RHA1-Δ*gblA*.

Effects of γ-butyrolactones on development of S*. coelicolor* have not been observed before. This difference in sporulation (Fig. 7) may be due to the *R. jostii* RHA1 RJB alone, or caused by unknown RHA1-OE compounds accumulating in response to the enhanced synthesis of RJB. To test this further, we added chemically synthesized 6-dehydro SCB2 to a confluent lawn of *S. coelicolor* M145 on MSM agar with casamino acids as nitrogen source. No growth difference was observed in the presence or absence of 6-dehydro SCB2 and the control (methanol) indicating that *S. coelicolor* M145 differentiation was not affected by γ-butyrolactones, but the production of its pigmented antibiotics was observed earlier than in the methanol control (data not shown).

Discussion

Rhodococci are Gram-positive soil bacteria known for the great variety of catabolic pathways which are encoded in their relatively large chromosomes, 9.6 Mb in case of *R. jostii* RHA1 [4]. Recent genome analyses showed that rhodococci also contain a large number of uncharacterized putative secondary metabolite gene clusters [4, 8]. Here we show that γ-butyrolactone gene clusters are not only present in the genomes of *R. jostii* RHA1*, R. equi* 103S*, R. opacus* B4 and *R. erythropolis* PR4 [8], but occur even more widespread in rhodococci (Fig. 3). Analysis of the γ-butyrolactone gene cluster in *R. jostii* RHA1 predicted the presence of genes encoding various homologues of enzymes known to be involved in the γ-butyrolactone biosynthesis pathway in *Streptomyces* [14], namely GblA, GblE and GblR. *R. jostii* RHA1 thus may employ a biosynthetic pathway similar to the one described in *S. griseus* (Fig. 2), also lacking the GblD enzyme. However, the GblE enzyme encoded in *R. jostii* RHA1 is not present in *S. griseus*. Instead this GblE enzyme is homologous to the γ-butyrolactone enzyme JadW2 from *S. venezuelae* which is known to be essential for the production of γ-butyrolactones in this strain [30] but it is not known which step it catalyzes. BLAST searches with GblD of *S. coelicolor* yielded a large number of dehydrogenases with 30%-40% AA identity to the query, spread throughout the *R. jostii* RHA1 genome. Also, two homologues of GblC were found in the *R. jostii* RHA1 genome, with ~ 35% AA identity to the *S. coelicolor* GblC. Homologues of these enzymes with higher identity than the ones found in *R. jostii* RHA1 are also present in the genomes of different *Streptomyces* strains. These enzymes have never been reported to be part of the γ-butyrolactone biosynthesis pathways in these streptomycetes. Clearly, simple sequence analysis is not sufficient to predict the involvement of *R. jostii* RHA1 genes homologues in the synthesis of RJB. Deletion mutagenesis of these putative *gblC*, *gblD* and *gblE* genes in *R. jostii* RHA1 followed by LC-MS analysis of cell extracts of these mutant strains, searching for

intermediates accumulating, may serve to elucidate the biosynthetic pathway in this strain. Since this pathway has not been completely elucidated in *Streptomyces*, other not yet identified pathway specific enzymes may also be involved in the synthesis of these signalling molecules. We carefully checked whether other genes surrounding the predicted γ-butyrolactone gene clusters (Fig. 3) in the different *Rhodococcus* strains are conserved, but this was not the case. Additional γ-butyrolactone biosynthesis genes thus remain to be identified in *R. jostii* RHA1.

Our data provides the first evidence of γ-butyrolactone synthesis in the genus *Rhodococcus*. The binding of exogenous molecules to γ-butyrolactone receptor proteins from *Streptomyces* previously has been observed for extracts of the cultures of other non-*Streptomyces* species [34, 35]. The latter study also suggested that *Amycolatopsis mediterrani* and *Micromonospora echinospora* produce an IM-2 type molecule and *Actinoplanes teichomyceticus* a VB type of γ-butyrolactone (Fig. 1). This conclusion was based on the efficiency by which these molecules bind to the *S. virginiae* and *S. lavendulae* γ-butyrolactone receptor proteins respectively, but no structural analysis was performed. We used LC-MS analysis to further compare the compounds in *R. jostii* RHA1 extracts with different chemically synthesized standards of known γ-butyrolactones. The *R. jostii* RHA1 RJB was identified as 6-dehydro SCB2 (Fig. 5). 6-Dehydro SCB2 is an isomer of the γ-butyrolactone described in *S. griseus* (A-factor) and is a predicted precursor of one of the described γ-butyrolactones in *S. coelicolor* (SCB2) [20]. In *S. coelicolor,* a GblD enzyme reduces the keto group in carbon 6 to a hydroxyl group (Fig. 2). We did not find a *gblD* homologue in the *R. jostii* RHA1 gene cluster (Fig. 3a), which corresponds to the observation that it is producing a 6-dehydro form of the molecule. When the samples were screened by LC-MS for a mass range that includes all known γ-butyrolactones, two peaks were observed in RHA1-OE and RHA1-C that were not present in the deletion strain and were in a lower intensity in the WT strain. The masses

corresponding to these peaks did not match to those of any γ-butyrolactone described to date. These molecules could be γ-butyrolactones with novel structures, or totally different compounds, e.g. products of a biosynthesis pathway regulated by RJB. The detected mass of m/z 255.1236 amu $[M-H]^-$ differs in m/z 14 from 6-dehydro SCB2, which would correspond to the loss of two hydrogen atoms and the gain of an oxygen atom. In further work we will attempt the isolation of sufficient amounts of these molecules for NMR analysis to elucidate their structures.

Unmarked deletion mutagenesis of the *gblA* gene in *R. jostii* RHA1 abolished RJB synthesis. Various homologues of this gene in *Streptomyces* species are known to be essential for biosynthesis of γ-butyrolactone molecules, catalyzing the first step of the biosynthesis, the condensation of a glycerol derivative with a fatty acid derivative (Fig. 2). Both Km bioassays (Fig. 4) and LC-MS analysis of extracts of the various *R. jostii* RHA1 (mutant) strains (Fig. 5) confirmed that *gblA* is essential for RJB synthesis. γ-Butyrolactones are known to regulate secondary metabolism and morphogenesis in the genus *Streptomyces* [14, 34]. *R. jostii* RHA1 contains a large number of putative secondary metabolite clusters that are mostly uncharacterized. RJB may be involved in control of the expression of one or more of these clusters. Although a few *Rhodococcus* antimicrobials are known this genus has remained largely unexplored for production of secondary metabolites [36-39]. In this work, we detected bioactivity from *R. jostii* RHA1 against *K. rhizophila, Aspergillus niger* and *Mycobacterium smegmatis*. Lariatins, cyclic peptides that have bioactivity against *Mycobacterium* species, were found in *R. jostii* K01-B0171 [37], but the enzymes involved in the synthesis of these compounds are not encoded in the genome from *R. jostii* RHA1. We have not been able to find a phenotypical difference between the *R. jostii* RHA1 WT, RHA1-Δ*gblA* and RHA1-OE strains, therefore further experiments are needed to analyse any regulatory effects of the γ-butyrolactone system. Mutagenesis analysis of GblR may help identify any *R. jostii* RHA1 genes

that are regulated by its RJB. Various systems may be controlled by RJB in rhodococci, analogous to the situation in the genus *Streptomyces*. In some species of *Streptomyces* γ-butyrolactones are known to be involved in morphogenesis and sporulation, as is the case in *S. griseus.* Deletion of AfsA in *S. griseus* blocked its sporulation and streptomycin production [34, 40]. The RJB in *R. jostii* RHA1 may be controlling the synthesis of one or more secondary metabolites that have remained unidentified, or it may be directly or indirectly influencing the primary metabolism in this strain. The γ-butyrolactone system is known to be present in several Actinomycete genera [19, 34, 35]. Actinomycetes are soil bacteria that live in a rich community of microorganisms. The γ-butyrolactone system may have developed as a way to communicate between different species. To test whether such interspecies communication occurs between *R. jostii* RHA1 and *S. coelicolor*, we plated these strains next to each other. *S. coelicolor* sporulation clearly was accelerated when growing next to RHA1-OE compared to *S. coelicolor* growing alone or next to the RHA1 WT and RHA1-Δ*gblA* strains. These results thus indicate that the *R. jostii* RHA1 RJB γ-butyrolactone affects morphological differentiation in *S. coelicolor.* This effect is known in other *Streptomyces* species [14, 41] but has never been described before in *S. coelicolor.* The addition of 6-dehydro SCB2 to a confluent lawn of *S. coelicolor* however did not induce sporulation which indicated that the phenotypical difference observed is not a direct effect of this RJB.

This work reports synthesis of a γ-butyrolactone-like molecule by *R. jostii* RHA1, a non-*Streptomyces* strain. This RJB molecule is structurally very similar to the γ-butyrolactones described in *Streptomyces* and interacts with the *S. coelicolor* butanolide system. In future work, we aim to elucidate the physiological roles of these signalling molecules in *Rhodococcus* metabolism, with specific interest in possible regulatory effects on representatives of the many secondary metabolite biosynthetic gene clusters in this genus. We have shown that *R. jostii* RHA1 produces compounds with antibiotic activity, with at least one of

them active against *M. smegmatis* and therefore potentially also against the fast-emerging multidrug resistant *Mycobacterium tuberculosis*. Activation of cryptic secondary metabolite clusters in rhodococci may potentially unlock the biosynthesis of novel compounds that are of interest to the pharmaceutical industry.

Acknowledgements:

AC was financially supported by the University of Groningen, and by the Dutch Technology Foundation (STW), which is part or the Netherlands organization for Scientific Research (NWO) and partly funded by the Ministry of Economic Affairs (STW 10463). We thank Subramaniyan Mannathan and Adriaan J. Minnaard of the Bio-Organic Chemistry department of the University of Groningen for providing the chemically synthesized SCBs, and Dr. Lara Martin-Sanchez for fruitful discussions.

References

1. Gürtler, V., Mayall, B. C. & Seviour, R. Can whole genome analysis refine the taxonomy of the genus *Rhodococcus*? *FEMS Microbiol. Rev.* **28**, 377-403 (2004).

2. Curragh, H. *et al*. Haloalkane degradation and assimilation by *Rhodococcus rhodochrous* NCIMB 13064. *Microbiology* **140 (Pt 6)**, 1433-1442 (1994).

3. van der Geize, R. & Dijkhuizen, L. Harnessing the catabolic diversity of Rhodococci for environmental and biotechnological applications. *Curr. Opin. Microbiol.* **7**, 255-261 (2004).

4. McLeod, M. P. *et al*. The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 15582-15587 (2006).

5. Taylor, C. R. *et al*. Isolation of bacterial strains able to metabolize lignin from screening of environmental samples. *J. Appl. Microbiol.* **113**, 521-530 (2012).

6. Letek, M. *et al*. The genome of a pathogenic *Rhodococcus*: cooptive virulence underpinned by key gene acquisitions. *PLoS Genet.* **6**, e1001145 (2010).

7. Hjerde, E. *et al*. Draft genome sequence of the Actinomycete *Rhodococcus* sp. strain AW25M09, isolated from the Hadsel Fjord, Northern Norway. *Genome Announc* **1**, e0005513-13 (2013).

8. Doroghazi, J. R. & Metcalf, W. W. Comparative genomics of Actinomycetes with a focus on natural product biosynthetic genes. *BMC Genomics* **14**, 611-2164-14-611 (2013).

9. Orro, A. *et al*. Genome and phenotype microarray analyses of *Rhodococcus sp.* BCP1 and *Rhodococcus opacus* R7: genetic determinants and metabolic abilities with environmental relevance. *PLoS One* **10**, e0139467 (2015).

10. Polkade, A. V., Mantri, S. S., Patwekar, U. J. & Jangid, K. Quorum sensing: an under-explored phenomenon in the phylum. *Front. Microbiol.* **7**, 131 (2016).

11. Takano, E., Chakraburtty, R., Nihira, T., Yamada, Y. & Bibb, M. J. A complex role for the gamma-butyrolactone SCB1 in regulating antibiotic production in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **41**, 1015-1028 (2001).

12. Bibb, M. J. Regulation of secondary metabolism in Streptomycetes. *Curr. Opin. Microbiol.* **8**, 208-215 (2005).

13. Li, X. *et al*. ScbR- and ScbR2-mediated signal transduction networks coordinate complex physiological responses in *Streptomyces coelicolor*. *Sci. Rep.* **5**, 14831 (2015).

14. Kato, J. Y., Funa, N., Watanabe, H., Ohnishi, Y. & Horinouchi, S. Biosynthesis of gamma-butyrolactone autoregulators that switch on secondary metabolism and morphological development in *Streptomyces*. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 2378-2383 (2007).

15. Khokhlov, A. S. in *Frontiers of Bioorganic Chemistry and Molecular Biology* (ed Ananchenko, S. N.) 201-210 (Pergamon, 1980).

16. Yamada, Y., Sugamura, K., Kondo, K., Yanagimoto, M. & Okada, H. The structure of inducing factors for virginiamycin production in *Streptomyces virginiae*. *J. Antibiot. (Tokyo)* **40**, 496-504 (1987).

17. Takano, E. *et al*. Purification and structural determination of SCB1, a gamma-butyrolactone that elicits antibiotic production in *Streptomyces coelicolor* A3(2). *J. Biol. Chem.* **275**, 11010-11016 (2000).

18. Sato, K., Nihira, T., Sakuda, S., Yanagimoto, M. & Yamada, Y. Isolation and structure of a new butyrolactone autoregulator from *Streptomyces sp*. FRI-5. *J. Ferment. Bioeng.* **68**, 170-173 (1989).

19. Hsiao, N. H. *et al*. Analysis of two additional signaling molecules in *Streptomyces coelicolor* and the development of a butyrolactone-specific reporter system. *Chem. Biol.* **16**, 951-960 (2009).

20. Martín-Sánchez, L. in *Quorum sensing in Streptomyces coelicolor: Regulation of the SCB signalling system that controls the synthesis of antibiotics* (PhD thesis, University of Groningen, 2016) http://hdl.handle.net/11370/2d24931d-c9c8-4ccc-832a-03f3286bce66.

21. Sakuda, S., Higashi, A., Tanaka, S., Nihira, T. & Yamada, Y. Biosynthesis of virginiae butanolide A, a butyrolactone autoregulator from *Streptomyces*. *J. Chem. Soc. Perkin Trans. I* **114**, 663-668 (1992).

22. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. in *Practical Streptomyces Genetics* (John Innes Foundation, Norwich, 2000).

23. Hsiao, N. H., Gottelt, M. & Takano, E. Chapter 6. Regulation of antibiotic production by bacterial hormones. *Methods Enzymol.* **458**, 143-157 (2009).

24. Bentley, S. D. *et al*. Complete genome sequence of the model Actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141-147 (2002).

25. Bos, C. J. *et al*. Genetic analysis and the construction of master strains for assignment of genes to six linkage groups in *Aspergillus niger*. *Curr. Genet.* **14**, 437-443 (1988).

26. Medema, M. H. *et al*. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39**, W339-46 (2011).

27. van der Geize, R., Hessels, G. I., van Gerwen, R., van der Meijden, P. & Dijkhuizen, L. Unmarked gene deletion mutagenesis of *kstD*, encoding 3-ketosteroid Delta1-dehydrogenase, in *Rhodococcus erythropolis* SQ1 using sacB as counter-selectable marker. *FEMS Microbiol. Lett.* **205**, 197-202 (2001).

28. Menges, R., Muth, G., Wohlleben, W. & Stegmann, E. The ABC transporter Tba of *Amycolatopsis balhimycina* is required for efficient export of the glycopeptide antibiotic balhimycin. *Applied Microbiology and Biotechnology* **77**, 125-134 (2007).

29. Arenskötter, M., Baumeister, D., Kalscheuer, R. & Steinbüchel, A. Identification and application of plasmids suitable for transfer of foreign DNA to members of the genus *Gordonia*. *Appl. Environ. Microbiol.* **69**, 4971-4974 (2003).

30. Lee, Y. J., Kitani, S., Kinoshita, H. & Nihira, T. Identification by gene deletion analysis of *barS2*, a gene involved in the biosynthesis of gamma-butyrolactone autoregulator in *Streptomyces virginiae*. *Arch. Microbiol.* **189**, 367-374 (2008).

31. Shikura, N., Yamamura, J. & Nihira, T. *barS1*, a gene for biosynthesis of a gamma-butyrolactone autoregulator, a microbial signaling molecule eliciting antibiotic production in *Streptomyces* species. *J. Bacteriol.* **184**, 5151-5157 (2002).

32. Bhukya, H., Bhujbalrao, R., Bitra, A. & Anand, R. Structural and functional basis of transcriptional regulation by TetR family protein CprB from *S. coelicolor* A3(2). *Nucleic Acids Res.* **42**, 10122-10133 (2014).

33. Sidda, J. D. *et al*. Overproduction and identification of butyrolactones SCB1-8 in the antibiotic production superhost *Streptomyces* M1152. *Org. Biomol. Chem.* **14**, 6390-6393 (2016).

34. Khokhlov, A. S. in *Microbial Autoregulators* (Harwood Academic Publishers, 1991).

35. Choi, S. U., Lee, C. K., Hwang, Y. I., Kinosita, H. & Nihira, T. Gamma-butyrolactone autoregulators and receptor proteins in non- *Streptomyces* Actinomycetes producing commercially important secondary metabolites. *Arch. Microbiol.* **180**, 303-307 (2003).

36. Chiba, H., Agematu, H., Sakai, K., Dobashi, K. & Yoshioka, T. Rhodopeptins, novel cyclic tetrapeptides with antifungal activities from *Rhodococcus sp*. III. Synthetic study of rhodopeptins. *J. Antibiot. (Tokyo)* **52**, 710-720 (1999).

37. Iwatsuki, M. *et al*. Lariatins, antimycobacterial peptides produced by *Rhodococcus* sp. K01-B0171, have a lasso structure. *J. Am. Chem. Soc.* **128**, 7486-7491 (2006).

38. Kitagawa, W. & Tamura, T. A quinoline antibiotic from *Rhodococcus erythropolis* JCM 6824. *J. Antibiot. (Tokyo)* **61**, 680-682 (2008).

39. Borisova, R. B. in *Isolation of a Rhodococcus soil bacterium that produces a strong antibacterial compound* (East Tennessee State University, 2011) http://dc.etsu.edu/etd/1388.

40. Horinouchi, S. & Beppu, T. Hormonal control by A-factor of morphological development and secondary metabolism in *Streptomyces*. *Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci.* **83**, 277-295 (2007).

41. Wang, L. & Vining, L. C. Control of growth, secondary metabolism and sporulation in *Streptomyces venezuelae* ISP5230 by*jadW* 1, a member of the *afsA* family of γ-butyrolactone regulatory genes. *Microbiology* **149**, 1991-2004 (2003).

CHAPTER 6.1

Summary and concluding remarks

Natural products

Natural products are compounds synthesized by living beings that have a great variety of functions for the source organism and for human use. One example is that of secondary metabolites. These are defined as molecules that are not essential for the producing organism but that may provide an advantage for their survival. Pigments, chelators and bioactive molecules (e.g. antibiotics) are some of the most important groups of molecules that are part of secondary metabolism [1-6]. Actinomycetes are the main bacterial source of secondary metabolites, with *Streptomyces* as one of the most important producers. Most of the antibiotics that are now being used in medicine have been isolated from *Streptomyces* [7]. The fast development of resistance to antibiotics in pathogenic bacteria is a well-known problem that has been extensively described and that needs to be addressed [8, 9]. From 1962 to 2000 not a single new antibiotic class has been introduced for treatment of infectious diseases. The last class commercialized is that of lipopeptides in the early 2000s, but this chemical class was already known from 1952 [10]. The discovery of new classes of molecules to fight the growing number of resistant microorganisms is therefore currently a priority. Great efforts are being made in further exploring the secondary metabolite richness that nature provides.

Activation of putative biosynthesis gene clusters (BGCs)

Different tools and strategies have been developed to identify gene clusters that may encode pathways for the biosynthesis of novel antibiotic compounds in the growing number of available microbial genomes. With the ongoing development of bioinformatic tools and genome sequencing it is becoming evident that the number of microbial natural products currently known is just the tip of the iceberg of the real repertoire available. This was also observed for the genome sequences of

streptomycetes, bacteria traditionally used for the production of secondary metabolites, showing the potential to produce many unknown compounds with putative antimicrobial properties [11, 12]. Also, strains of genera not well-known for the production of secondary metabolites, e.g. *Rhodococcus* or *Mycobacterium* [13], harbour a great number of BGCs in their genomes. Most of these BGCs have remained cryptic (their products are unknown) and different techniques have been developed aiming to activate them, including heterologous expression of BGCs, redirecting precursors by inactivating known biosynthetic pathways, manipulating regulatory systems or building improved biosynthesis pathways by synthetic biology approaches [14-17] (Chapter 1).

In the work reported in this PhD thesis we explored different techniques to activate BGCs in *Streptomyces clavuligerus,* a well-known producer of antibiotic compounds (Chapter 2 and 3). We also performed a deep bioinformatic analysis of the surprisingly large repertoire of cryptic secondary gene clusters found in the genome sequences of rhodococci and Mycobacteria (Chapter 4). This included the γ-butyrolactone signalling system which is well-known for its regulatory role in streptomycetes secondary metabolism [18]. This cluster had been previously predicted in four strains of rhodococci [13] but we show that it extends throughout the genus. We characterized it in more detail in *Rhodococcus jostii* RHA1 (Chapter 5).

Heterologous production of indigoidine

S. clavuligerus is known for the synthesis of clavulanic acid, a β-lactamase inhibitor that is co-formulated with amoxicillin to produce the well-known drug Augmentin [19, 20]. Clavulanic acid is one of the main secondary metabolites produced by *S. clavuligerus*. Its biosynthesis pathway is partially shared with the one of clavam antibiotics, such as alanyl-clavam [21]. The wild type strain is also known to produce trace amounts of

holomycin and tunicamycin-like antibiotics MM 19290 [22]. Genome analysis showed that this strain contains an impressive number of cryptic BGCs [12]. We targeted the cryptic BGC for the biosynthesis of the blue pigment indigoidine. This pigment is produced by diverse bacteria and it is thought to have anti-oxidative and antimicrobial properties [23-25]. It is however not known whether indigoidine is the final product of the biosynthesis pathway and the function of most genes in this BCG are unknown. The cryptic gene cluster of *S. clavuligerus* encodes a homologue of the nonribosomal peptide synthetase needed for the production of the blue pigment indigoidine (IndC) [26]. Interestingly, IndC of *S. clavuligerus* has an extra domain representing a 4-oxalocrotonate tautomerase (4-OT)-like enzyme, of unknown function. This tautomerase-like enzyme is usually encoded by a separate gene in the indigoidine biosynthesis gene cluster (*indD*). In Chapter 2 we show that such a fusion of putative *indC* and *indD* genes occurs more wide spread, and is present in at least 30 different bacterial strains. Heterologous expression of the *S. clavuligerus indC(D)* gene plus flanking genes in several *Streptomyces* host strains did not result in blue pigment production. Only expression of *indC(D)* alone, controlled by a strong promoter, resulted in production of indigoidine in *Streptomyces coelicolor, R. jostii*, and *Escherichia coli*, showing for the first time that IndC(D) from *S. clavuligerus* indeed encodes an indigoidine synthetase. We also show that separate expression of the *S. clavuligerus* IndC, lacking the tautomerase domain, yields more indigoidine than expression of the complete *indC(D)* gene. To further study the activity of the tautomerase domain in IndC(D), a truncated *S. clavuligerus* gene, encoding only the IndD domain, was successfully expressed in *E. coli* and purified as an active enzyme, catalyzing a promiscuous Michael-like addition reaction but none of the canonical reactions described for 4-OT tautomerases. The physiological function of this tautomerase-like enzyme is still unknown

but its fusion to IndC may ensure the same expression levels, guaranteeing that both enzymes are in close proximity for a possible fast reaction before the indigoidine monomers are condensed to form indigoidine, or even as a way to select for an inactive tautomerase. Chapter 2 shows that secondary metabolites from *S. clavuligerus* can be produced in different heterologous hosts, including *Rhodococcus*, an actinomycete with faster growth than *Streptomyces*. A range of molecular tools are available for rhodococci making these strains highly interesting as host for heterologous expression (and engineering) of *Streptomyces* secondary metabolite gene clusters.

Six derivatives of antibiotics MM 19290 were isolated and partially characterized from *S. clavuligerus* strain Δ7

Aiming to activate or upregulate secondary metabolite biosynthesis pathways in *S. clavuligerus* ATCC 27064, we constructed a mutant strain that is fully blocked in the synthesis of its main secondary metabolite clavulanic acid, *S. clavuligerus* strain Δ7. The mutations introduced are known to block also the synthesis of holomycin [27]. *S. clavuligerus* strain Δ7 was found to produce bioactive compounds. After extraction and several purification steps, we were able to analyse the bioactive compounds by mass spectrometry (MS) and MS/MS. This resulted in the putative identification and characterization of 6 different compounds with masses equivalent to those of different derivatives of tunicamycin, and the structurally related corynetoxins and streptovirudines [28]. Tunicamycin-like compounds named MM 19290 are known to be produced only in trace amounts by the wild type strain *S. clavuligerus* ATCC 27064 [22, 29] and no structural information was available yet. Different derivatives of tunicamycin show different activity [30]. Tunicamycins are able to inhibit *N*-glycosylation which makes them highly toxic, but they are used in various ways in research and medicine [31-34]. New derivatives of these antibiotics

may show different activities that could be further exploited. The results presented in this work contribute to our knowledge on the diversity of tunicamycin-related compounds synthesized by *S. clavuligerus* ATCC 27064 (Chapter 3).

The great potential for secondary metabolite production of *Rhodococcus*

Rhodococci are well-known for their capacity to degrade a wide range of (aromatic) compounds [35-38] but hardly explored for their ability to produce secondary metabolites. Recently it has become evident that rhodococci encode a large number of unexplored putative BGCs [13]. The genus *Rhodococcus* is closely related to that of *Mycobacterium*, which comprises feared pathogens such as *Mycobacterium tuberculosis* and *Mycobacterium leprae*. To further explore the specialized metabolic potential of *Rhodococcus*, we performed a computational analysis of 20 available rhodococcal genomes, and compared these with those of several *Mycobacterium* strains as well as that of *Amycolicicoccus subflavus.* Most of the predicted BGCs from these strains lack any homology with gene clusters responsible for synthesis of known natural products. We identified five gene clusters that are shared among all strains of these three genera studied. Their analysis led to identification of PKS13 as the last condensation step in mycolic acid synthesis; this is an essential biosynthetic activity for these genera and already used as target to combat pathogenic *Mycobacterium* strains [39]. Essential enzymes involved in cell wall synthesis, in iron-sulphur cluster biosynthesis or vitamin K12 synthesis, were also found in these shared clusters. We were not able to assign a function to two of the universally shared gene clusters but their presence in the genomes of all these strains suggests that they likely also have an essential role. These shared clusters thus may be responsible for the production of interesting novel compounds and / or

provide targets to combat the pathogenic strains. Mutational inactivation of these clusters may generate weaker strains that can be used as live vaccines.

We also studied the predicted gene clusters that only are present in closely related strains (clade-specific) and therefore appear in one branch of the phylogenetic tree, and the clusters that are only present in single strains. Strain-specific clusters are interesting targets for secondary metabolite production, since there is less risk that the final product is known already from a different species; in many studies this has been problematic and many compounds have been re-discovered [40]. By studying the conservation patterns of the predicted BGCs across the different species, we hypothesized about their biological roles and prioritized them for a future functional characterization. We also observed that from all rhodococci studied, *R. jostii* RHA1 and three strains from *Rhodococcus opacus* contain the highest number of BGCs in their genomes, even higher than the predicted number in the well-known secondary metabolite producers *S. coelicolor* or *S. clavuligerus*, making them highly interesting strains to study in more detail (Chapter 4).

Identification and characterization of the γ-butyrolactone system from *R. jostii* RHA1

In view of the great number of cryptic secondary metabolite gene clusters present in rhodococci, it is highly relevant to gain insight into their secondary metabolism regulation. Our further analysis focussed on *R. jostii* RHA1, one of the most interesting *Rhodococcus* strains. One of the most conserved BGCs in rhodococci is a homologue of the extensively studied quorum-sensing system for γ-butyrolactones, known to participate in the activation of secondary metabolism in the *Streptomyces* genus [18, 41-43]. The data shows that *R. jostii* RHA1 synthesises γ-butyrolactones, the first report of the production of these molecules

outside the *Streptomyces* genus. Extracted γ-butyrolactones from *R. jostii* RHA1 (RJBs) were shown to be able to interact with receptor protein ScbR of *S. coelicolor* using a previously developed reporter system [44]. This indicates that inter-genera communication occurs via these RJB molecules. LC-MS analysis of the same extracts using synthetic standards of the *S. coelicolor* γ-butyrolactones [45] identified an RJB molecule with identical mass and elution time as 6-dehydro SCB2, the predicted precursor of *S. coelicolor* γ-butyrolactone SCB2. We furthermore identified the *R. jostii* RHA1 key RJB biosynthesis gene *gblA*. Following deletion of the *gblA* gene, no RJB production could be detected whereas higher RJB levels were synthesized when *gblA* was overexpressed. We were also able to detect for the first time the synthesis of diffusible molecules with antimicrobial activity from *R. jostii* RHA1 (Chapter 5). The only known secondary metabolite BGCs from this strain are involved in the formation of siderophores. The detected antimicrobial molecules thus may be novel and further increase our interest in this strain. Knowing the relevant role that γ-butyrolactones have in *Streptomyces* secondary metabolism [46], RJBs could be used as tool for the awakening of cryptic secondary metabolic gene clusters in rhodococci.

The work in this thesis provides new insights in the activation of cryptic secondary metabolite gene clusters in *S. clavuligerus* and *R. jostii* RHA1 and shows in detail the potential that the genus *Rhodococcus* has to produce novel natural products. Both *Streptomyces* and *Rhodococcus* are highly interesting genera to explore for novel secondary metabolites, native ones or following expression of foreign clusters. They encode a surprisingly high number of cryptic BGCs in their genomes [11-13] and techniques for DNA manipulation are available for both genera [47, 48]. *Streptomyces* has historically been used for this purpose and has been extensively studied for the production of bioactive compounds [7].

Furthermore, great effort has been made to unravel the regulatory networks in *Streptomyces* species resulting in a better understanding of gene expression that will surely be key to the current efforts being dedicated to activate cryptic gene clusters [41, 49-52]. *Rhodococcus* however shows a great potential to produce secondary metabolites that are unexplored [13]. The genus is extensively studied for its great catabolic potential [37]; and molecular tools are available to perform genetic manipulation [48]. The presence of the common γ-butyrolactone system between both genera indicates that this system is a promising tool for activation of the secondary metabolism in both genera. In order to improve our chances of discovering novel antimicrobials that can be used against resistant bacterial strains, we need to apply a combination of different approaches. First, the use of bioinformatics tools to detect putative BGCs in newly sequenced genomes, serving to choose those strains that may be the source of interesting new compounds and to select the most promising cryptic gene clusters for further study (Chapter 4). The next step is to select a suitable host strain for activation of these gene clusters, which also will define which of the different strategies will be followed to induce the production of these compounds (Chapters 1 - 3). The analysis of the regulatory systems of donor and hosts strains is also vital to understand why these gene clusters are not active and how their expression can be triggered (Chapter 5). This thesis therefore gives insight in all the necessary steps needed to attempt and perform the activation of cryptic BGCs.

HOOFDSTUK 6.2

Samenvatting en conclusies

Natuurlijke producten

Natuurlijke producten zijn verbindingen die gesynthetiseerd worden door levende organismen en die een grote variatie aan functies hebben voor het producerende organisme en voor de mens. Secundaire metabolieten zijn een voorbeeld van natuurlijke producten. Dit zijn moleculen die niet essentieel zijn voor het producerende organisme maar die wel zijn overleving kunnen bevorderen. Pigmenten, chelatoren en bioactieve moleculen (bv antibiotica) zijn belangrijke groepen van secundaire metabolieten [1-6]. Actinomyceten zijn de belangrijkste bacteriële bron van secundaire metabolieten, en leden van het genus *Streptomyces* zijn de belangrijkste producenten. De meeste antibiotica die momenteel medisch toegepast worden zijn dan ook geisoleerd uit *Streptomyces* stammen [7]. De snelle ontwikkeling van antibioticaresistentie in ziekteverwekkende (pathogene) bacteriën is een bekend probleem dat in detail beschreven is en met urgentie bestreden moet worden [8, 9]. In de periode 1962 tot 2000 is er geen enkele nieuwe klasse van antibiotica geïntroduceerd voor de behandeling van bacteriële infecties. De meest recente klasse die commerciële toepassing gevonden heeft is die van de lipopeptiden, en wel in 2000, maar deze antibiotica waren al bekend vanaf 1952 [10]. De ontdekking van geheel nieuwe klassen van moleculen waarmee de groeiende aantallen resistente microorganismen bestreden kunnen worden is daarom nu een grote prioriteit. Momenteel wordt dan ook op relatief grote schaal onderzoek verricht naar de natuurlijke rijkdom aan secundaire metabolieten.

Activatie van potentiële biosynthetische genclusters (BGCs)

In de afgelopen jaren zijn verschillende nieuwe technieken en strategiën ontwikkeld voor de identificatie van genclusters die coderen voor biosyntheseroutes voor nieuwe antibiotica in de steeds grotere aantallen microbiële genoomsequenties die nu beschikbaar komen. Door deze

nieuwe ontwikkelingen in de bioinformatica en genoomsequentieanalyse wordt het steeds duidelijker dat het aantal bekende microbiële natuurlijke producten nog maar een puntje van de ijsberg is. Ook bij de analyse van de genoomsequenties van Streptomyceten, bacteriën die traditioneel gebruikt worden voor de productie van secundaire metabolieten, werd een grote capaciteit waargenomen voor de synthese van vele onbekende verbindingen, potentieel met antimicrobiële eigenschappen [11, 12]. Ook stammen van bacteriegeslachten die niet bekend staan als producenten van secundaire metabolieten, bijvoorbeeld *Rhodococcus* en *Mycobacterium* [13], hebben een relatief groot aantal BGCs in hun genomen. De meeste van deze BGCs zijn cryptisch, wat wil zeggen dat ze niet 'actief' zijn en dat dus het product niet gevormd wordt. Verschillende technieken zijn ontwikkeld om deze clusters te activeren, zoals heterologe expressie (in andere bacteriën), stimuleren van de aanvoer van precursor moleculen door inactivatie van bekende biosynthetische routes, engineering van controle systemen of het construeren van nieuwe en/of verbeterde biosyntheseroutes (synthetische biologie) [14-17] (Hoofdstuk 1).

Dit proefschrift beschrijft de resultaten van onderzoek waarbij verschillende technieken gebruikt zijn om BGCs te activeren in *Streptomyces clavuligerus,* een bekende producent van antibiotica (Hoofdstukken 2 en 3). Ook is er een gedetailleerde bioinformatica analyse uitgevoerd van het verrassend grote aantal cryptische secundaire genclusters die in de genoomsequenties van Rhodococci en Mycobacteria gevonden worden (Hoofdstuk 4). Het γ-butyrolacton signaalsysteem, bekend om zijn regulerende rol bij de synthese van secundaire metabolieten in Streptomycetes [18], bleek ook in deze geslachten voor te komen. De aanwezigheid van dit cluster is eerder voorspeld in 4 *Rhodococcus* stammen [13] maar dit proefschrift laat zien dat het breed verspreid aanwezig is in dit bacteriegeslacht. Een meer gedetailleerde

karakterisatie van het *Rhodococcus jostii* RHA1 γ-butyrolacton systeem wordt hier beschreven (Chapter 5).

Heterologe productie van indigoidine

S. clavuligerus is het meest bekend als producent van clavulaanzuur, een β-lactamase remmer die in combinatie met amoxicillin gebruikt wordt in Augmentin, een geneesmiddel dat bacteriële infecties bestrijdt [19, 20]. Clavulaanzuur is een van de belangrijkste secundaire metabolieten van *S. clavuligerus*. Zijn biosyntheseroute is gedeeltelijk gelijk aan die voor clavam antibiotica, zoals alanyl-clavam [21]. De wild type (natuurlijke) stam produceert ook spoortjes holomycin, en de tunicamycine-achtige antibiotica MM 19290 [22]. Uit genoomanalyse is gebleken dat deze stam verder ook nog codeert voor een indrukwekkend aantal cryptische BGCs [12]. De heterologe expressie van een van deze cryptische BGCs, coderend voor de biosynthese van het blauwe pigment indigoidine, wordt beschreven in Hoofdstuk 2. Dit pigment wordt door verschillende bacteriën geproduceerd en heeft waarschijnlijk anti-oxidatieve en antimicrobiële eigenschappen [23-25]. Het is echter niet bekend of indigoidine het eindproduct van deze biosyntheseroute is; ook zijn de functies van de meeste genen in dit BCG nog onbekend. De cryptische gencluster in *S. clavuligerus* codeert voor een homoloog van de nonribosomale peptide synthetase die nodig is voor de productie van het blauwe pigment indigoidine (IndC) [26]. Ook interessant is dat IndC van *S. clavuligerus* een extra domein heeft, een 4-oxalocrotonate tautomerase (4-OT)-achtig enzym, met een onbekende functie. Dit tautomerase-achtige enzym wordt normal gecodeerd door een afzonderlijk (eigen) gen in de indigoidine biosynthese gencluster (*indD*). Hoofdstuk 2 toont aan dat een dergelijke fusie van *indC* en *indD* genen breder verspreid voorkomt, en wel in tenminste 30 verschillende bacteriestammen. Heterologe expressie van *indC(D)* en de flankerende genen uit *S.*

clavuligerus in verschillende andere *Streptomyces* stammen resulteerde niet in synthese van een blauw pigment. Expressie van alleen *indC(D)*, onder controle van een sterke promoter, resulteerde wel in synthese van indigoidine, zowel in *Streptomyces coelicolor, R. jostii*, als in *Escherichia coli.* Dit is het eerste directe experimentele bewijs dat IndC(D) van *S. clavuligerus* inderdaad codeert voor een indigoidine synthetase. Expressie van alleen het *S. clavuligerus* IndC, dus zonder het tautomerase domein, resulteerde zelfs in synthese van meer indigoidine dan expressie van het complete *indC(D)* gen. Ook het tautomerase domein in IndC(D) werd in meer detail bestudeerd. Daartoe werd een *S. clavuligerus* gen dat alleen codeert voor het losse IndD domein tot expressie gebracht in *E. coli* en gezuiverd als een actief enzym. Analyse toonde aan dit IndD een promiscue Michael-achtige additiereactie katalyseert maar geen van de canonieke (gebruikelijke) reacties die beschreven zijn voor 4-OT tautomerases. De fysiologische functie van dit tautomerase-achtige enzym is nog onbekend. Fusie van IndC en IndD zorgt mogelijkerwijs voor het handhaven van vergelijkbare enzymniveaus, en garandeert dat beide enzyme fysiek dichtbij elkaar zijn. Wellicht wordt hiermee een snelle (onbekende) reactie mogelijk voordat de indigoidine monomeren condenseren tot indigoidine. Momenteel valt ook nog niet uit te sluiten dat vanwege de fusie dit tautomerase enzym inactief is.

Hoofdstuk 2 toont aan dat secundaire metabolieten van *S. clavuligerus* geproduceerd kunnen worden in verschillende heterologe gastheerstammen, waaronder *Rhodococcus*, een actinomyceet die duidelijk sneller groeit dan *Streptomyces*. Momenteel beschikken wij al over goede moleculaire technieken voor metabole engineering van Rhodococci. Mede hierdoor zijn Rhodococci zeer interessant als alternatieve gastheren voor de heterologe expressie (en engineering) van genclusters voor secundaire metabolieten uit *Streptomyces.*

Isolatie en karakterisatie van zes varianten van de MM 19290 tunicamycine antibiotica uit *S. clavuligerus* stam Δ7

In dit onderzoek is een mutant van *S. clavuligerus* ATCC 27064 geconstrueerd (stam Δ7) die volledig geblokkeerd is in de synthese van clavulaanzuur, de belangrijkste secundaire metaboliet geproduceerd door deze bacterie. Ook de synthese van holomycine is geblokkeerd in deze mutant [27]. Vervolgens is onderzocht of in mutant stam Δ7 nu andere biosyntheseroutes voor secundaire metabolieten geactiveerd zijn, of dat hun expressie verbeterd is. Uit analyse bleek dat *S. clavuligerus* stam Δ7 inderdaad nog andere bioactieve verbindingen maakt. Na hun extractie en verschillende zuiveringsstappen zijn deze bioactieve verbindingen met massa spectrometrie (MS) en MS/MS technieken verder geanalyseerd. Dit resulteerde in de identificatie en karakterisatie van 6 verbindingen met massawaarden overeenkomstig met die van verschillende varianten van tunicamycine, en de structureel verwante corynetoxine en streptovirudine verbindingen [28]. Tunicamycine-achtige verbindingen (bekend als MM 19290) worden in spoortjes geproduceerd door de wild type stam *S. clavuligerus* ATCC 27064 [22, 29] maar informatie over hun structuren was nog niet beschikbaar. De verschillende tunicamycine-achtige moleculen variëren in hun activiteit [30]. Zo zijn tunicamycines in staat om cellulaire *N*-glycosyleringsactiviteiten te remmen, waardoor ze zeer toxisch zijn, en ze worden voor verschillende medische en onderzoeksdoelen gebruikt [31-34]. Het is mogelijk dat de nieuwe varianten van deze antibiotica nieuwe (bio)activiteiten hebben die wellicht ook toegepast kunnen worden. De resultaten die gepresenteerd worden in dit hoofdstuk dragen sterk bij aan onze kennis over de diversiteit aan tunicamycine-achtige verbindingen die gesynthetiseerd worden door *S. clavuligerus* ATCC 27064 (Hoofdstuk 3).

Het geslacht *Rhodococcus* heeft een groot potentieel voor productie van secundaire metabolieten

Rhodococci staan bekend om hun vermogen om een grote variatie aan (aromatische) verbindingen af te breken [35-38] maar zijn nog nauwelijks onderzocht op hun vermogen om secundaire metabolieten te synthetiseren. Recent is duidelijk geworden dat de genomen van *Rhodococcus* stammen coderen voor een groot aantal onbekende biosynthetische gen clusters (BGCs) [13]. Het geslacht *Rhodococcus* is nauw verwant aan dat van *Mycobacterium*, met gevreesde pathogene bacteriën zoals *Mycobacterium tuberculosis* en *Mycobacterium leprae*. Teneinde het biosynthetische potentiaal van *Rhodococcus* nader te bestuderen zijn de 20 beschikbare genoomsequenties van Rhodococci geanalyseerd en vergeleken met die van verschillende *Mycobacterium* stammen, en die van *Amycolicicoccus subflavus.* De meeste BGCs van deze stammen blijken geen enkele overeenkomst te hebben met genclusters betrokken bij de synthese van bekende natuurlijke producten. Een 5-tal genclusters bleken in alle onderzochte stammen van deze 3 bacteriegeslachten aanwezig te zijn. Hun analyse resulteerde in identificatie van PKS13 als de laatste condensatiestap in mycolzuursynthese. Mycolzuren vormen het hoofdbestanddeel van de celwand van deze bacteriën. PKS13 katalyseert daarom een essentiële biosynthesestap in deze 3 geslachten. Door deze stap te remmen zouden pathogene *Mycobacterium* stammen bestreden kunnen worden [39]. De gevonden gemeenschappelijke genclusters bleken verder te coderen voor essentiële enzymen betrokken bij celwandsynthese, synthese van ijzer-zwavel clusters of vitamine K12 synthese. Voor 2 van deze gemeenschappelijke genclusters kon geen functie gevonden worden; hun aanwezigheid in de genomen van al deze stammen doet vermoeden dat zij ook coderen voor essentiële functies. Deze gemeenschappelijke genclusters kunnen dus verantwoordelijk zijn voor de synthese van

interessante nieuwe verbindingen, en zijn wellicht ook nieuwe doelen voor bestrijding van pathogene stammen. Inactivatie van deze genclusters door mutagenese zal mogelijkerwijs resulteren in zwakkere stammen die gebruikt kunnen worden als levend vaccin.

Ook werden de genclusters bestudeerd die alleen aanwezig zijn in nauwverwante stammen, en de clusters die specifiek in slechts 1 stam aanwezig zijn. Stam-specifieke clusters zijn interessant omdat de kans dan groter is dat de secundaire metaboliet die geproduceerd wordt nog geheel nieuw is. Herontdekking van al eerder geïdentificeerde metabolieten komt in dit onderzoeksveld helaas vaak voor [40]. Na analyse van de conservering van de BGCs in de verschillende bacteriegeslachten en soorten zijn hypotheses opgesteld voor hun mogelijke biologische rol; tevens zijn ze qua relevantie gerangschikt voor hun toekomstige functionele karakterisatie. Van alle bestudeerde Rhodococci bleken *R. jostii* RHA1 en 3 *Rhodococcus opacus* stammen het grootste aantal BGCs in hun genomen te hebben, zelfs meer dan in de zeer bekende producenten van secundaire metabolieten, *S. coelicolor* en *S. clavuligerus.* Deze *Rhodococcus* stammen zijn dus zeer interessant om in meer detail te bestuderen (Hoofdstuk 4).

Identificatie en karakterisatie van het γ-butyrolacton systeem in *R. jostii* RHA1

Gezien het grote aantal cryptische genclusters voor secundaire metabolieten in Rhodococci is het ook zeer relevant om de regulering van genexpressie en metabolietsynthese op te helderen. Daartoe werd een van de meest geconserveerde BGCs in Rhodococci in meer detail bestudeerd, namelijk het quorum-sensing systeem voor γ-butyrolacton synthese in *R. jostii* RHA1. Quorum-sensing is een mechanisme waarbij meerdere bacteriën hun genexpressie kunnen afstemmen op de populatiedichtheid aan de hand van signaalmoleculen. Het γ-

butyrolactonsysteem is eerder in detail bestudeerd in *S. coelicolor*, en is betrokken bij de activatie van routes voor synthese van secundaire metabolieten [18, 41-43]. De resultaten laten zien dat *R. jostii* RHA1 inderdaad γ-butyrolactonmoleculen synthetiseert, de eerste keer dat dit gerapporteerd wordt buiten het geslacht *Streptomyces*. Door gebruik te maken van een eerder ontwikkeld detectiesysteem [44] kon aangetoond worden dat de uit *R. jostii* RHA1 geextraheerde γ-butyrolactonmoleculen (RJBs) zelfs in staat zijn om een interactie aan te gaan met het receptoreiwit ScbR van *S. coelicolor.* Het is dus mogelijk dat deze RJB moleculen een rol spelen in de communicatie tussen deze 2 geslachten. LC-MS analyse van deze extracten, gebruikmakend van eerder gesynthetiseerde en gekarakteriseerde *S. coelicolor* γ-butyrolactonmoleculen [45], resulteerde in identificatie van een RJB molecuul met de massa en het elutiepatroon van 6-dehydro SCB2, de precursor van het *S. coelicolor* γ-butyrolacton SCB2 molecuul. Ook werd het *gblA* gen geidentificeerd als het belangrijkste gen voor RJB synthese in *R. jostii* RHA1. Deletie van het *gblA* gen in *R. jostii* RHA1 resulteerde in compleet verlies van RJB synthese; overexpressie van het *gblA* gen resulteerde in hogere RJB niveaus. In dit onderzoek namen we ook voor het eerst waar dat *R. jostii* RHA1 moleculen met antimicrobiële activiteit synthetiseert (Hoofdstuk 5). De enige tot nu toe bekende *R. jostii* RHA1 secundaire metabolieten coderen voor sideroforen (ijzerbindende moleculen). Deze *R. jostii* RHA1 moleculen met antimicrobiële activiteit zijn dus wellicht geheel nieuw en daarmee interessante doelen voor verder onderzoek. Gezien de belangrijke rol van γ-butyrolactonmoleculen in *Streptomyces* [46] is het denkbaar dat RJBs gebruikt kunnen worden om de expressie van cryptische genclusters voor secundaire metabolieten in Rhodococci te activeren.

De resultaten van het onderzoek beschreven in dit proefschrift vergroten onze kennis over cryptische genclusters voor secundaire metabolieten in *S. clavuligerus* and *R. jostii* RHA1. Ook is duidelijk geworden dat stammen van het geslacht *Rhodococcus* potentieel veel verschillende en nieuwe natuurlijke producten kunnen synthetiseren. *Streptomyces* en *Rhodococcus* stammen zijn zeer interessant voor verder onderzoek naar nieuwe secundaire metabolieten, als bron van nieuwe BGCs, of als gastheren voor expressie van heterologe genclusters. De genomen van beiden coderen voor een verrassend groot aantal cryptische BGCs [11-13], en technieken voor hun genetische manipulatie zijn beschikbaar [47, 48]. Streptomyceten worden al langer bestudeerd voor dit doeleinde en er is al een schat aan informatie over hun bioactieve verbindingen [7]. Ook zijn er veel studies verricht naar de werking van regulatienetwerken in *Streptomyces* soorten, resulterend in betere inzichten in de regulering van genexpressie, hetgeen van grote waarde is voor de pogingen die momenteel gaande zijn om cryptische genclusters te activeren [41, 49-52]. *Rhodococcus* soorten beschikken echter ook over een groot potentieel voor productie van (vaak geheel nieuwe) secundaire metabolieten [13]. Diverse stammen van het geslacht *Rhodococcus* zijn intensief bestudeerd voor hun grote catabole vermogen [37] en moleculaire technieken zijn beschikbaar voor hun genetische manipulatie [48]. Gezien de aanwezigheid van een vergelijkbaar γ-butyrolactonsysteem in beide geslachten zou dit systeem over en weer gebruikt kunnen worden voor activatie van hun secundaire metabolisme. Idealiter wordt een combinatie van verschillende technieken gebruikt om onze kansen te verbeteren op het ontdekken van nieuwe antimicrobiële stoffen die actief zijn tegen resistente pathogene bacteriën. Ten eerste, het gebruik van bioinformaticatechnieken om BGCs te ontdekken in nieuwe genoomsequenties. Op basis hiervan kunnen die stammen worden geselecteerd voor nader onderzoek die beschikken over cryptische

genclusters die potentieel coderen voor interessante nieuwe verbindingen (Hoofdstuk 4). In de volgende stap kan een geschikte gastheerstam geselecteerd worden voor activatie van deze genclusters, gebruikmakend van de verschillende strategiën voor inductie van de synthese van deze verbindingen (Hoofdstukken 1 - 3). Ook is het essentieel dat de betrokken regulatiesystemen in donor en gastheerstammen opgehelderd worden, om te begrijpen waarom deze genclusters vaak niet actief zijn, en hoe hun expressie aangezet kan worden (Hoofdstuk 5). Dit proefschrift beschrijft in detail de verschillende stappen die genomen kunnen worden om expressie van cryptische BGCs te activeren.

CAPÍTULO 6.3

Resumen y conclusiones

Productos naturales

Los productos naturales son compuestos producidos por organismos vivos. Poseen una gran variedad de funciones tanto para el organismo productor como para el ser humano. Los metabolitos secundarios son productos naturales no esenciales para el organismo que los produce, pero les confiere una ventaja para su supervivencia. Pigmentos, quelantes y moléculas bioactivas como son los antibióticos son algunos de los grupos de moléculas más importantes que forman parte del metabolismo secundario [1-6]. Los Actinomicetos son la principal fuente bacteriana de metabolitos secundarios, siendo el género *Streptomyces* uno de los más importantes productores. La mayor parte de los antibióticos usados actualmente en medicina provienen de este género [7]. Actualmente nos enfrentamos al gran problema del rápido desarrollo de resistencia a antibióticos de las bacterias patógenas [8, 9]. A este problema contribuye la disminución de antibióticos introducidos como fármacos. Entre 1962 y 2000 no se introdujo ninguna nueva clase de antibióticos para el tratamiento de enfermedades infecciosas. La última clase nueva de antibióticos comercializada fueron los lipopéptidos a principio de los 2000, pero esta clase química se conocía desde 1952 [10]. El descubrimiento de nuevas clases de antibióticos que sean activas contra cepas multi-resistentes es, por lo tanto, una prioridad en la investigación. Actualmente se están desarrollando grandes esfuerzos para explorar más allá la riqueza de metabolitos secundarios que la naturaleza nos ofrece.

Activación de posibles agrupaciones génicas de rutas biosintéticas (Biosynthetic Gene Clusters, BGCs)

Se han desarrollado diferentes herramientas y estrategias para identificar agrupaciones génicas (o clústeres) que codifican las enzimas necesarias para la biosíntesis de nuevos compuestos en la cada vez mayor cantidad de genomas microbianos secuenciados disponibles. Gracias al continuo

desarrollo de herramientas bioinformáticas y de secuenciación, es cada vez más evidente que los productos naturales microbianos que hoy en día se conocen son sólo una pequeña fracción del repertorio realmente existente. Esto también se observó en streptomycetos, bacterias tradicionalmente usadas para la producción de metabolitos secundarios [11, 12]. Igualmente, se ha detectado una sorprendente cantidad de posibles BGCs en el genoma de cepas de géneros que no son conocidos por la producción de metabolitos secundarios, como, por ejemplo, los géneros *Rhodococcus* y *Mycobacterium* [13]. La mayoría de los BGCs son crípticos (su producto es desconocido). Se han perfeccionado diferentes técnicas con el objetivo de activarlos, incluida la expresión heteróloga de BGCs, la inactivación de rutas biosintéticas conocidas para re-direccionar precursores a otras rutas, la manipulación de sistemas de regulación o la construcción de rutas biosintéticas mejoradas mediante biología sintética [14-17] (Capítulo 1).

En esta tesis doctoral hemos explorado diferentes técnicas para activar BGCs en *Streptomyces clavuligerus,* un destacado productor de antibióticos (Capítulos 2 y 3). Así mismo, realizamos un profundo análisis bioinformático del gran repertorio de BGCs de metabolitos secundarios crípticos presentes en los genomas de *Rhodococcus* y *Mycobacterium* (Capítulo 4). Este análisis incluye el sistema de señalización de las γ-butirolactonas que es conocido por su papel regulador en el metabolismo secundario de Streptomycetos [18]. Esta agrupación génica había sido previamente detectada en el genoma de cuatro cepas de *Rhodococcus* [13]. En esta tesis mostramos que este clúster está ampliamente distribuido en este género. Además, caracterizamos el clúster de las γ-butirolactonas en mayor detalle en la cepa *Rhodococcus jostii* RHA1 (Capítulo 5).

Producción heteróloga de indigoidina

S. clavuligerus es conocido por la síntesis de ácido clavulánico, un inhibidor de la enzima β-lactamasa que inhibe la acción de los antibióticos β-lactámicos como la penicilina. El ácido clavulánico está formulado junto con amoxicilina en el conocido medicamento Augmentin [19, 20] y es uno de los principales metabolitos secundarios producidos por *S. clavuligerus*. Su ruta biosintética está parcialmente compartida con la responsable de la síntesis de varios compuestos con estructura clavama como, por ejemplo, alanilclavam [21]. La cepa silvestre puede producir también cantidades traza del antibiótico holomicina y de compuestos antibióticos relacionados con la tunicamicina, llamados MM 19290 [22]. El análisis de la secuencia se su genoma, demostró que esta cepa tiene un gran número de BGCs crípticas [12]. En nuestro trabajo nos centramos en el estudio de la agrupación génica responsable de la síntesis del pigmento azul indigoidina, que es críptica en *S. clavuligerus*. Este pigmento lo producen diversos tipos de bacterias y se cree que tiene propiedades antioxidantes y antimicrobianas [23-25]. Sin embargo, se desconoce si el pigmento es el producto final de la ruta biosintética, ya que la función de la mayoría de los genes del clúster es desconocida. En la agrupación génica presente en *S. clavuligerus* se encuentra codificada la sintetasa de péptido no ribosomal (NRPS) necesaria para la síntesis de indigoidina (IndC) [26], con la singularidad de incluir un dominio extra similar a una 4-oxalocrotonato tautomerasa (4-OT), de función desconocida. Esta enzima tipo 4-OT es normalmente un gen a parte en las agrupaciones génicas de indigoidina (*indD*). En el Capítulo 2 mostramos que la fusión de los genes *indC* e *indD* está presente en, al menos, 30 cepas bacterianas distintas. La expresión heteróloga del gen *indC(D)* de *S. clavuligerus* junto con los genes flanqueantes en varias cepas de *Streptomyces* no resultó en la producción del pigmento azul. Sin embargo, la expresión única del gen *indC(D)*, controlado por un promotor fuerte, resultó en la producción de indigoidina en *Streptomyces coelicolor, R. jostii*, y *Escherichia coli*, mostrando por primera vez que

IndC(D) de *S. clavuligerus* realmente codifica una sintetasa de indigoidina. Así mismo, mostramos que la producción de indigoidina es mayor al expresar IndC de *S. clavuligerus*, separada del domino tipo tautomerasa. Para hacer un estudio más exhaustivo de este dominio, expresamos IndD en *E. coli* y lo purificamos, resultando en una enzima activa capaz de catalizar una reacción similar a la adición de Michael, promiscua para las 4-OT tautomerasas. Sin embargo, no se detectó actividad en ninguna de las reacciones canónicas descritas para las 4-OT tautomerasas. La función fisiológica de esta enzima es todavía desconocida. Su fusión con IndC podría ser beneficiosa para asegurar los mismos niveles de expresión, garantizar la cercanía de las dos enzimas para una posible rápida reacción precedente a la dimerización espontánea de los monómeros de indigoidina para formar indigoidina, o incluso una forma de seleccionar una tautomerasa inactiva. El Capítulo 2 muestra qué diferentes huéspedes heterólogos pueden ser usados para la producción de metabolitos secundarios de *S. clavuligerus* incluido *Rhodococcus*, un actinomiceto de crecimiento más rápido que *Streptomyces*. Hay disponibles una gran variedad de herramientas moleculares para modificar *Rhodococcus,* aportando, por lo tanto, un gran interés a este género como huéspedes heterólogos para la expresión e ingeniería de agrupaciones génicas de metabolitos secundarios de *Streptomyces*.

Aislamiento y caracterización parcial de seis derivados de los antibióticos MM 19290 de *S. clavuligerus* cepa Δ7

Con el objetivo de activar o regular positivamente las rutas de biosíntesis de metabolitos secundarios en *S. clavuligerus* ATCC 27064, construimos una cepa mutante con un bloqueo total de la ruta biosintética de ácido clavulánico (*S. clavuligerus* cepa Δ7). Las mutaciones introducidas también bloquean la síntesis del antibiótico holomicina [27]. Pudimos detectar los compuestos bioactivos producidos por *S. clavuligerus* cepa

Δ7. Tras su extracción y varios pasos de purificación, pudimos analizar los compuestos bioactivos por espectrometría de masas (MS) y espectrometría de masas en tándem (MS/MS). Este análisis nos llevó a identificar seis compuestos con masas equivalentes a diferentes derivados de la tunicamicina y/o a los compuestos estructuralmente relacionados corynetoxinas y streptovirudinas [28]. Anteriormente se había descrito que los compuestos tipo tunicamicinas denominados MM 19290 son producidos en cantidades traza por la cepa salvaje de *S. clavuligerus* ATCC 27064 [22, 29], pero no había información estructural aun disponible. Diferentes derivados de tunicamicinas tienen distinta actividad [30]. Las tunicamicinas son capaces de inhibir la *N*-glicosilación, lo que las hace altamente tóxicas, pero tienen variados usos en investigación y medicina [31-34]. Nuevos derivados de estos antibióticos podrían mostrar diferentes actividades que podrían ser explotadas. Los resultados presentados en esta investigación contribuyen a nuestro conocimiento de la diversidad de antibióticos tipo tunicamicina sintetizados por *S. clavuligerus* ATCC 27064 (Capítulo 3).

El gran potencial de producción de metabolitos secundarios de *Rhodococcus*

Rhodococcus son conocidos por su habilidad de degradar una gran variedad de compuestos aromáticos [35-38] pero prácticamente no se ha explorado su capacidad para producir metabolitos secundarios. Recientemente se ha hecho evidente que los genomas de *Rhodococcus* codifican un gran número de BGCs de posibles metabolitos secundarios [13]. El género *Rhodococcus* está estrechamente relacionado con el género *Mycobacterium*, el cual incluye patógenos peligrosos como el causante de la tuberculosis (*Mycobacterium tuberculosis*) y el causante de la lepra (*Mycobacterium leprae*). Para hacer un estudio más exhaustivo del potencial metabólico especializado de *Rhodococcus*,

hicimos un análisis computacional de 20 genomas de *Rhodococcus* disponibles públicamente y lo comparamos con varias cepas de *Mycobacterium* y con *Amycolicicoccus subflavus.* La mayoría de los BGCs predichos en estas cepas no tienen homología con agrupaciones génicas responsables de la síntesis de compuestos conocidos. Identificamos cinco clústeres compartidos por todas las especies estudiadas de los tres géneros distintos. Su análisis llevó a la identificación de PKS13, enzima que cataliza el último paso de condensación en la síntesis de ácido micólico. Esta es una ruta esencial para estos géneros y actualmente se usa como diana para combatir a cepas patógenas de *Mycobacterium* [39]. Además, encontramos incluidas en las agrupaciones génicas universalmente compartidas enzimas esenciales involucradas en la síntesis de la pared celular, en la biosíntesis del clúster hierro-azufre o en la síntesis de la vitamina K12. No fue posible asignar una función a dos de los clústeres universalmente compartidos, pero su presencia en todos los genomas estudiados sugiere que también poseen una función esencial. Los clústeres compartidos por lo tanto podrían ser responsables de la síntesis de nuevos compuestos y/o proveer de dianas para lidiar con las cepas patógenas. La inactivación de estos clústeres podría generar cepas más débiles que podrían ser usadas como vacunas.

Así mismo, analizamos agrupaciones génicas que están únicamente presentes en cepas filogenéticamente cercanas y que, por lo tanto, aparecen en la misma rama del árbol filogenético (clado) y los clústeres presentes solo en una cepa. Estas últimas podrían ser interesantes fuentes de metabolitos secundarios novedosos, ya que hay menos riesgo de que el producto final sea uno ya conocido producido por otras especies, cómo ha sido el caso en anteriores estudios [40]. A través del estudio del patrón de conservación de los BGCs detectados en las diferentes especies, hemos podido deducir su posible papel biológico y

priorizarlos para una futura caracterización funcional. También pudimos observar que, entre las cepas de *Rhodococcus* estudiadas, *R. jostii* RHA1 y las tres cepas de *Rhodococcus opacus* poseen el mayor número de BGCs en sus genomas, más alto incluso que el número de BGCs detectados en cepas destacadas por la producción de metabolitos secundarios como son *S. coelicolor* o *S. clavuligerus*. Estas cepas de *Rhodococcus* son, por lo tanto, altamente interesantes para ser estudiadas en más detalle (Capítulo 4).

Identificación y caracterización del sistema de γ-butirolactona de *R. jostii* RHA1

En vista del gran número de metabolitos secundarios crípticos presentes en los genomas de *Rhodococcus*, es muy importante adquirir conocimiento sobre la regulación de su metabolismo secundario. Una de las agrupaciones génicas más conservadas en *Rhodococcus* es la responsable de la síntesis de γ-butirolactonas, un sistema de "quorum-sensing" relacionado con la regulación del metabolismo secundario en el género *Streptomyces* [18, 41-43]. Nuestros resultados muestran que *R. jostii* RHA1 sintetiza γ-butirolactonas, la primera prueba de síntesis de estas moléculas fuera del género *Streptomyces*. Mediante el uso de un sistema de reporte previamente desarrollado [44] observamos cómo las γ-butirolactonas extraídas de *R. jostii* RHA1 (RJBs) tienen la capacidad de interactuar con el receptor ScbR de *S. coelicolor*. Este resultado indica que la comunicación entre estos géneros mediante RJBs podría ser posible. Estos mismos extractos fueron analizados mediante cromatografía líquida acoplada a espectrometría de masas (LC-MS), usando estándares sintéticos de γ-butirolactonas de *S. coelicolor* [45]. El resultado de este análisis fue la identificación de una molécula (RJB) con masa y tiempo de elución idénticos a 6-deshidro SCB2, el precursor predicho de la γ-butirolactona de *S. coelicolor* SCB2. Además, identificamos el gen

principal de la síntesis de RJB, *gblA*. Tras la eliminación de este gen, no pudimos detectar la producción de RJBs, mientras que su sobreexpresión llevo a una mayor producción de esta molécula. Así mismo, pudimos detectar por primera vez la producción de compuestos con actividad microbiana producidos por *R. jostii* RHA1 (Capítulo 5). Los únicos metabolitos secundarios descritos en esta cepa son sideróforos. Estos compuestos antimicrobianos detectados podrían ser novedosos, haciendo así aun mayor el interés por el estudio de esta cepa. Conociendo el importante papel de las γ-butirolactonas en el metabolismo secundario de *Streptomyces* [46], RJBs podrían ser de utilidad como herramienta para la activación de rutas del metabolismo secundario crípticas en el género *Rhodococcus*.

El trabajo de investigación de esta tesis proporciona nuevos conocimientos sobre la activación de rutas crípticas del metabolismo secundario en S. *clavuligerus* y *R. jostii* RHA1 y muestra en detalle el potencial que posee el género *Rhodococcus* para la producción de productos naturales. Tanto el género *Streptomyces* como el de *Rhodococcus* son de gran interés para la búsqueda de nuevos metabolitos secundarios nativos o para ser usados como hospedadores de agrupaciones génicas de otras especies. Sus genomas codifican un gran número de rutas crípticas [11-13] y hay disponibles técnicas de manipulación genética para ambos géneros [47, 48]. *Streptomyces* ha sido históricamente utilizado como productor de metabolitos secundarios y ha sido extensamente estudiado para la producción de moléculas bioactivas [7]. Se han hecho grandes esfuerzos para entender las redes regulatorias en diferentes especies de *Streptomyces,* resultando en un mejor entendimiento de la expresión génica en este género que será de gran utilidad para activar BGCs crípticos [41, 49-52]. *Rhodococcus,* sin embargo, apenas ha sido explorado con este fin. Este género ha sido extensamente

estudiado por su gran potencial catabólico [37] y debido a ello, se han desarrollado herramientas de manipulación genética que pueden ahora ser utilizadas para el estudio de su metabolismo secundario [48]. La presencia de un sistema de γ-butirolactona común entre *Streptomyces* y *Rhodococcus* indica que podría ser una herramienta prometedora para estudiar el metabolismo secundario de ambos géneros. Con el objetivo de mejorar nuestras probabilidades de encontrar nuevos compuestos con actividad frente a patógenos multi-resistentes, necesitamos usar una combinación de distintos enfoques. El uso de métodos bioinformáticos facilita la detección y priorización de los BGCs a estudiar (Capítulo 4). El siguiente paso es la selección del hospedador donde estas agrupaciones génicas serán activadas, lo cual definirá también qué estrategias se seguirán para inducir la producción de los compuestos (Capítulos 1-3). El análisis de los sistemas de regulación, tanto de la cepa donante como de la cepa huésped, es vital para entender por qué estos clústeres no son activos y cómo inducir su expresión (Capítulo 5). Esta tesis doctoral, por lo tanto, proporciona conocimiento sobre todos los pasos necesarios para la activación de rutas crípticas.

References

1. Cimermancic, P. *et al*. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412-421 (2014).

2. Butler, M. S. & Buss, A. D. Natural products–the future scaffolds for novel antibiotics? *Biochem. Pharmacol.* **71**, 919-929 (2006).

3. Crotti, S. *et al*. Mass spectrometry in the pharmacokinetic studies of anticancer natural products. *Mass Spectrom Rev.* (2015).

4. Kumar, A., Vishwakarma, H. S., Singh, J., Dwivedi, S. & Kumar, M. Microbial pigments: production and their applications in various industries. *IJPCBS* **5**, 203-212 (2015).

5. Li, K., Chen, W. H. & Bruner, S. D. Microbial siderophore-based iron assimilation and therapeutic applications. *Biometals* **29**, 377-388 (2016).

6. Breitling, R., Ceniceros, A., Jankevics, A. & Takano, E. Metabolomics for secondary metabolite research. *Metabolites* **3**, 1076-1083 (2013).

7. Chater, K. F. *Streptomyces* inside-out: a new perspective on the bacteria that provide us with antibiotics. *Philos. Trans. R. Soc. Lond. , B, Biol. Sci.* **361**, 761-768 (2006).

8. Nguyen, L. Antibiotic resistance mechanisms in *M. tuberculosis*: an update. *Arch. Toxicol.* **90**, 1585-1604 (2016).

9. Yosef, I., Manor, M. & Qimron, U. Counteracting selection for antibiotic-resistant bacteria. *Bacteriophage* **6**, e1096996-Mar (2016).

10. Silver, L. L. Challenges of antibacterial discovery. *Clin. Microbiol. Rev.* **24**, 71-109 (2011).

11. Bentley, S. D. *et al*. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141-147 (2002).

12. Medema, M. H. *et al*. The sequence of a 1.8-Mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. *Genome Biol. Evol.* **2**, 212-224 (2010).

13. Doroghazi, J. R. & Metcalf, W. W. Comparative genomics of Actinomycetes with a focus on natural product biosynthetic genes. *BMC Genomics* **14**, 611-2164-14-611 (2013).

14. Gómez-Escribano, J. P. & Bibb, M. J. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechnol.* **4**, 207-215 (2011).

15. Lincke, T., Behnken, S., Ishida, K., Roth, M. & Hertweck, C. Closthioamide: an unprecedented polythioamide antibiotic from the strictly anaerobic bacterium *Clostridium cellulolyticum*. *Angew. Chem. Int. Ed. Engl.* **49**, 2011-2013 (2010).

16. Olano, C. *et al*. Activation and identification of five clusters for secondary metabolites in *Streptomyces albus* J1074. *Microb. Biotechnol.* **7**, 242-256 (2014).

17. Breitling, R. & Takano, E. Synthetic biology of natural products. *Cold Spring Harb. Perspect. Biol.* **8**, 10.1101/cshperspect.a023994 (2016).

18. Kato, J. Y., Funa, N., Watanabe, H., Ohnishi, Y. & Horinouchi, S. Biosynthesis of gamma-butyrolactone autoregulators that switch on secondary metabolism and morphological development in *Streptomyces*. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 2378-2383 (2007).

19. Koekman, B. & Hans, M. in *Engineering Complex Phenotypes in Industrial Strains* (ed Patnaik, R.) pp. 169-183 (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2012).

20. Leigh, D. A., Bradnock, K. & Marriner, J. M. Augmentin (amoxycillin and clavulanic acid) therapy in complicated infections due to beta-lactamase producing bacteria. *J. Antimicrob. Chemother.* **7**, 229-236 (1981).

21. Zelyas, N. J., Cai, H., Kwong, T. & Jensen, S. E. Alanylclavam biosynthetic genes are clustered together with one group of clavulanic acid biosynthetic genes in *Streptomyces clavuligerus*. *J. Bacteriol.* **190**, 7957-7965 (2008).

22. Kenig, M. & Reading, C. Holomycin and an antibiotic (MM 19290) related to tunicamycin, metabolites of *Streptomyces clavuligerus*. *J. Antibiot. (Tokyo)* **32**, 549-554 (1979).

23. Reverchon, S., Rouanet, C., Expert, D. & Nasser, W. Characterization of indigoidine biosynthetic genes in *Erwinia chrysanthemi* and role of this blue pigment in pathogenicity. *J. Bacteriol.* **184**, 654-665 (2002).

24. Cude, W. N. *et al*. Production of the antimicrobial secondary metabolite indigoidine contributes to competitive surface colonization by the marine roseobacter *Phaeobacter* sp. strain Y4I. *Appl. Environ. Microbiol.* **78**, 4771-4780 (2012).

25. Cude, W. N. *et al*. *Phaeobacter* sp. strain Y4I utilizes two separate cell-to-cell communication systems to regulate production of the antimicrobial indigoidine. *Appl. Environ. Microbiol.* **81**, 1417-1425 (2015).

26. Takahashi, H. *et al*. Cloning and characterization of a *Streptomyces* single module type non-ribosomal peptide synthetase catalyzing a blue pigment synthesis. *J. Biol. Chem.* **282**, 9073-9081 (2007).

27. de la Fuente, A., Lorenzana, L. M., Martín, J. F. & Liras, P. Mutants of *Streptomyces clavuligerus* with disruptions in different genes for clavulanic acid biosynthesis produce large amounts of holomycin: possible cross-regulation of two unrelated secondary metabolic pathways. *J. Bacteriol.* **184**, 6559-6565 (2002).

28. Eckardt, K. Tunicamycins, streptovirudins, and corynetoxins, a special subclass of nucleoside antibiotics. *J. Nat. Prod.* **46**, 544-550 (1983).

29. Wyszynski, F. J., Hesketh, A. R., Bibb, M. J. & Davis, B. G. Dissecting tunicamycin biosynthesis by genome mining: cloning and heterologous expression of a minimal gene cluster. *Chem. Sci.* **1**, 581-589 (2010).

30. Duksin, D. & Mahoney, W. C. Relationship of the structure and biological activity of the natural homologues of tunicamycin. *J. Biol. Chem.* **257**, 3105-3109 (1982).

31. Klenk, H. & Schwarz, R. T. Viral glycoprotein metabolism as a target for antiviral substances. *Antiviral Res.* **2**, 177-190 (1982).

32. Hutt-Fletcher, L. M., Balachandran, N. & LeBlanc, P. A. Modification of Epstein-Barr virus replication by tunicamycin. *J. Virol.* **57**, 117-123 (1986).

33. Pulverer, G., Beuth, J., Ko, H. L., Sölter, J. & Uhlenbruck, G. Modification of glycosylation by tunicamycin treatment inhibits lectin-mediated adhesion of *Streptococcus pneumoniae* to various tissues. *Zentralbl Bakteriol Mikrobiol Hyg A.* **266**, 137-144 (1987).

34. Zhang, Z. *et al*. Modification of glycosylation mediates the invasive properties of murine hepatocarcinoma cell lines to lymph nodes. *PLoS One* **8**, e65218 (2013).

35. Curragh, H. *et al*. Haloalkane degradation and assimilation by *Rhodococcus rhodochrous* NCIMB 13064. *Microbiology* **140 (Pt 6)**, 1433-1442 (1994).

36. van der Geize, R. & Dijkhuizen, L. Harnessing the catabolic diversity of Rhodococci for environmental and biotechnological applications. *Curr. Opin. Microbiol.* **7**, 255-261 (2004).

37. McLeod, M. P. *et al*. The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 15582-15587 (2006).

38. Taylor, C. R. *et al*. Isolation of bacterial strains able to metabolize lignin from screening of environmental samples. *J. Appl. Microbiol.* **113**, 521-530 (2012).

39. Rayasam, G. V. MmpL3 a potential new target for development of novel anti-tuberculosis drugs. *Expert Opin. Ther. Targets* **18**, 247-256 (2014).

40. Wietz, M., Månsson, M., Vynne, N. G. & Gram, L. in *Marine Microbiology: Bioactive Compounds and Biotechnological Applications* (ed Kim, S. K.) (Wiley-VCH, Weinheim, 2013).

41. Takano, E., Chakraburtty, R., Nihira, T., Yamada, Y. & Bibb, M. J. A complex role for the gamma-butyrolactone SCB1 in regulating antibiotic production in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **41**, 1015-1028 (2001).

42. Bibb, M. J. Regulation of secondary metabolism in Streptomycetes. *Curr. Opin. Microbiol.* **8**, 208-215 (2005).

43. Li, X. *et al*. ScbR- and ScbR2-mediated signal transduction networks coordinate complex physiological responses in *Streptomyces coelicolor*. *Sci. Rep.* **5**, 14831 (2015).

44. Hsiao, N. H. *et al*. Analysis of two additional signaling molecules in *Streptomyces coelicolor* and the development of a butyrolactone-specific reporter system. *Chem. Biol.* **16**, 951-960 (2009).

45. Martín-Sánchez, L. in *Quorum sensing in Streptomyces coelicolor: Regulation of the SCB signalling system that controls the synthesis of antibiotics* (PhD thesis, University of Groningen, 2016) http://hdl.handle.net/11370/2d24931d-c9c8-4ccc-832a-03f3286bce66.

46. Horinouchi, S. & Beppu, T. Hormonal control by A-factor of morphological development and secondary metabolism in *Streptomyces*. *Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci.* **83**, 277-295 (2007).

47. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. in *Practical Streptomyces Genetics* (John Innes Foundation, Norwich, 2000).

48. van der Geize, R. *et al*. A novel method to generate unmarked gene deletions in the intracellular pathogen *Rhodococcus equi* using 5-fluorocytosine conditional lethality. *Nucleic Acids Res.* **36**, e151 (2008).

49. Reverchon, S., Nasser, W. & Robert-Baudouy, J. *pecS*: a locus controlling pectinase, cellulase and blue pigment production in *Erwinia chrysanthemi*. *Mol. Microbiol.* **11**, 1127-1139 (1994).

50. Bibb, M. J., Janssen, G. R. & Ward, J. M. Cloning and analysis of the promoter region of the erythromycin resistance gene (*ermE*) of *Streptomyces erythraeus*. *Gene* **38**, 215-226 (1985).

51. Gao, C., Hindra, Mulder, D., Yin, C. & Elliot, M. A. Crp is a global regulator of antibiotic production in *Streptomyces*. *MBio* **3**, 10.1128/mBio.00407-12 (2012).

52. Gottelt, M., Kol, S., Gomez-Escribano, J. P., Bibb, M. & Takano, E. Deletion of a regulatory gene within the *cpk* gene cluster reveals novel antibacterial activity in *Streptomyces coelicolor* A3(2). *Microbiology* **156**, 2343-2353 (2010).

ACKNOWLEDGEMENTS

DANK JE WEL

THANK YOU VERY MUCH

MUCHAS GRACIAS

DANKE SCHÖN

非常感谢

ESKERRIK ASKO

“Everything with a beginning, has an end”. Even PhD thesis!! And as final part of this work, I would like to thank all the people that have helped me one way or another to get here. I could not have done it without you.

First of all, I would like to thank my promotors, Mirjan and Lubbert. The great support I have received from you was essential for finishing the thesis, but also to become a better scientist and evolve as a person. Beiden hartstikke bedankt!

Lubbert, a simple thank you does not feel enough to express my gratitude for all the help I have received from you these last years. I will always be grateful for your help through one of the most difficult periods in my life and for giving me the opportunity of developing myself and finishing this thesis. Thank you also for translating the summary of this thesis to Dutch for me! I am happy to have you as my promotor. I wish you all the best in your retirement and I hope you and Dinie can visit all those places you want to travel to and enjoy. Thank you very much for everything!

Mirjan, where to start… I met you as the Russian PhD student sitting next to me in the bench that spoked surprisingly fluent Dutch and that was running around working very hard. I am not sure how long it took until someone told me you are actually Dutch. Afterwards you became a postdoc in the Strepto group and we developed a friendship that I know we will keep through the years. You were always there supporting me even when you were going through some difficult times yourself. You took the initiative to become my supervisor/co-promotor. Without you this thesis would not have been possible. You gave me a very strong personal and professional support, you were there cheering up, listening to all the frustrations, problems and successes. I am very honored to be able to say that you are my co-promotor, my supervisor and, most importantly, my friend. I wish you all the best in the future!

I also want to thank my paranymphs, Lara and Laura, with whom I have 3-4 whatsapp groups just with the three of us. And even if we all separate now, remember that the world is no longer so big and that one of the whatsapp groups should keep working. Siempre habéis estado ahí para escucharme y apoyarme en todos los momentos de dificultad y también para celebrar los buenos momentos.

Laruska, te conocí durante mi erasmus siendo la chica española sentada enfrente de mi pipeteando qRTs durante horas y siempre dispuesta a ayudar. Durante la tesis fuiste un apoyo indispensable para mí. Sin él no estaríamos hoy aquí. Muchas gracias por haber ido abriéndome camino y creando precedente, haciendo mi camino mucho más fácil. Y por supuesto gracias por esas cenas y cervezas en los momentos en los que necesitaba una distracción o hablar y también en aquellos en los que había que celebrar, los conciertos, los festivales, las fiestas, los viajes… Por ayudarme a ultimar detalles de la tesis y del proceso para acabarla. Y por elegirme como paraninfa. ¡Fue un honor serlo! No sé qué habría hecho sin ti. Me alegro muchísimo de haber coincidido contigo estos años y haber podido conocerte. Ahora empiezas una nueva etapa en Wageningen y con Joost (¡por fin!!), donde sé que serás feliz y te va a ir genial. Muchísimas gracias por todo.

Lauris, no sabes lo mucho que me alegro de que vinieras a Groningen. Viniste a un par de semanas más tarde de mi comienzo de tesis. Desde entonces has ido y venidoun total de 4 veces (por ahora). Y en esta última en el momento perfecto para ser mi paraninfa. Mientras estabas en Madrid estábamos igualmente en contacto y nos veíamos en mi paso por Madrid camino a Pamplona. Tu apoyo directo u online ha sido también indispensable para mí. Siempre ayudando y dando consejos sobre todo tipo de temas, animando, riéndonos, sobretodo riéndonos como empieces tú jeje. Fue un honor ser testigo de tu boda con Guti y de haber podido compartir un tiempo con los dos viviendo en la misma ciudad,

yendo de cenas y cervezas. Y en la última vez que habéis venido ya con Nico hemos podido pasar tardes de cena disfrutando de vuestra compañía y de Nico, que es un niño guapísimo y encantador. Os deseo lo mejor a los tres y ya nos veremos en un sitio o en otro para que pueda malcríalo y darle juguetes.

I want also to thank you Mark, for also being there for me in both good and bad moments, for all our travels and our dinners and beers. Thank you very much for being there supporting me and helping me going through certain difficult times but also for always being available to celebrate. Thank you also for your really tasty cooking and your home-made beer or for building my computer or the Game of Thrones saga. That I will finish (eventually). And all the recipes with which I have developed my Dutch for cooking (voeg, roer, kook, snijd...). And don't worry, wherever I end up, I can always be your ibéricos provider. Thank you very much for everything!!!!

Thanks to all of those friends that have been there at some point in Groningen and that have made chapters in my time here. Joost, Marcos, Alok, Marta, Tim, Ilse, Vero, Santi, Geralt, HaJö, Joana, Javi, Guti, Nico. It was a pleasure to have you around. Joana, mil gracias por tus sabios consejos durante estos años. Javi, thank you very much for the great design of the cover. I like it a lot. Y gracias a los dos por las cenas y quedadas que hemos hecho juntos.

Vero, it was great to have you around, we miss you in Groningen. El tiempo que estuviste aquí fue una etapa genial para mí. Aprendí muchas cosas de ti, y me ayudaste mucho en el trabajo y personalmente. Y también a maldecir en argentino y otro vocabulario como la gran y útil palabra "boludear". Me ha encantado poder haberte visto aun cuando ya te habías ido de Groningen. Pensaba que viviendo en Argentina no iba a

poder verte, pero no ha sido así. Te deseo lo mejor para ti y tu familia y espero que podamos seguir viéndonos. Y dile ¡Hala! a Lari de mi parte!

Leticia, thank you very much to you and your maridito Peter to keep me as fit as possible during this period. Thank you for the dinners and for all the nice wine bottles you brought me as exchange for taking care of Eli and Copi, which I really enjoyed a lot! ¡Y muchas gracias por ayudarme con todo el proceso de finalización de la tesis! We arrived to Groningen at the same time, and we will probably leave also at the same time! I wish you both the best in the future, wherever you end up! And I hope we can meet from time to time!

Geralt, thank you very much for all the nights out and concerts in the city. We still should join for a big festival! Thank you also for taking me out in the last months of my writing period. It helped keeping me sane. And of course, thank you for your help and coffees at work.

HaJö, thank you very much for being there, for your jokes, for your help, for all those beers and dinners, even when they were slightly late for you. Thank you for cooking for us in so many occasions and thank you for all those hugs that we taught you to appreciate. I am sure you have a bright future out there! After all, as you say, everything will be fine! Danke schön!

Martuski, thank you also for the festivals (even when the type of music was not always really your favorite), the parties (remember Mojito night?) for our travels to Brussels, Vienna, Copenhagen… You are finishing one month and 3 days after me! I wish you all the best and a happy life with Tim in Gent! ¡Muchas gracias por todo!

Joost! I am very happy that you and Lara have found each other. I wish you two the best in your future together!

Ilse, thanks for those crazy nights and parties! I wish you the best on you PhD in Leeuwarden and in your future!

Marcos, Alok, Bruno, Marta and later Tim and together with Lara, we have been in 3 festivals where I could see some of my favorite music groups and where I had a lot of fun with you. I hope we can keep this yearly tradition in the future!

HaJö, Cecile, Mirjan and Lara, we teamed up in a very important point of my thesis to try to improve the situation we all were or had went through. Thanks to that, we were successful in finding help and support in the university which lead to a new and brighter step in our careers. Thank you all for being there at that moment and in many others.

I would also like to thank everyone in the Microbial Physiology group, past and present members. Andriy, Stef, Dennis, Elena, Marco, Davide, Wouter, Evelien, Vincent, Rivca, Vero, Cecile, Mirjan, Laura, Lara, Mark, Pieter, Alicia, Geralt, Sander, Marnix, Thai, Sebastian, Ahmed, Hien, Huifang, Xiangfeng, Markus, Tim, Bai, Scott, Sahar, Robert, Hans… I enjoy a lot my time in the group with all of you. Special thanks to Xiangfeng, who started at the same time as I did. We shared a lot of frustrations and problems in the beginning. You also chose me as your paranymph, which I was honored to do. I wish you all the best now back in China and I hope we can keep in touch. And Huifang, thank you for translating “thank you very much” into Chinese 非常感谢

Thanks to some very important members in the group, our technicians. Pieter, thank you very much for your advices, for listening, for your fast-problem-solving abilities. Also for keeping us company late in the evenings, especially in Haren, when we were almost the only group left in that creepy dark building with people already living in it. Cecile, thank you for being always so happy and getting smiles out of people. Thank

you for worrying about us and for offering help constantly. Geralt, you are in other sections, but thank you also for all the help at work. Thank you all for your hard work that keeps everything running!

And how to forget my office mates. Markus, thank you for all the conversations we had in the office, I will miss them. You will soon be also finishing your thesis and I hope you will enjoy whatever time you have left in Groningen and whatever comes later. Bai, thank you also for those conversations where I learned a lot about how China works. It was a lot of fun to have you around! And of course, Tim! Thank you also for making the office life more entertaining!

I also want to thank my students, Danae, Peter, Jacco and Reinoud. I had a great experience with all of you. I learned a lot and I hope you did too. I wish you the best in the future.

Bea, Manon and later Anmara, thank you very much for your very needed help. Without you I would have gotten stuck at so many points. You are indispensable for all of us. Thank you very much for being there.

I would like to also thank my friends from Spain which are each in a different country; Cova, Fernando, Marisa, Natalia, Oihane y Maria. Gracias por las charlas y por los ánimos. Espero que podamos seguir viéndonos en el futuro.

And the last but absolutely not the least, I want to thank my family. A mis tíos y primos, abuelos, muchas gracias por vuestro apoyo y ánimo todos estos años. Y sobre todo a vosotros, papá, mamá y Juan, muchas gracias por vuestro apoyo y por vuestra paciencia. No habría podido llegar aquí sin vosotros. Siempre me habéis animado a hacer lo que yo quiero y me habéis apoyado en todo. Espero algún día poder devolvéroslo. Os quiero mucho.

Thank you all so much!!!