

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/138641> holds various files of this Leiden University dissertation.

**Author:** Du, C.

**Title:** Multi-omics studies of the control of growth and antibiotic production of streptomyces

**Issue Date:** 2020-12-09

Multi-omics studies of the control of  
growth and antibiotic production of  
*Streptomyces*

Chao DU (杜超)





# Multi-omics studies of the control of growth and antibiotic production of *Streptomyces*

Proefschrift

ter verkrijging van  
de graad van Doctor aan de Universiteit Leiden,  
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,  
volgens besluit van het College voor Promoties  
te verdedigen op woensdag 9 december 2020  
klokke 10:00 uur

door

Chao DU (杜超)

geboren te Fushun, China

in 1987

## **PROMOTORES**

Prof.dr. G.P. van Wezel

Prof.dr. J.M. Raaijmakers

## **PROMOTIECOMMISSIE**

Prof.dr. A.H. Meijer

Prof.dr. H.P. Spaink

Prof.dr. J.H. de Winde

Dr. A. Wentzel (SINTEF)

Dr. J.J.J. van der Hooft (Wageningen University)

# Table of Contents

## **Chapter 1**

Introduction	2
--------------	---

## **Chapter 2**

Mining for Microbial Gems: Integrating Proteomics in the Postgenomic Natural Product Discovery Pipeline	6
---	---

## **Chapter 3**

Application of systems biology methods for the identification of novel natural products in <i>Streptomyces</i> species	22
--	----

## **Chapter 4**

Unravelling the response of <i>Streptomyces roseifaciens</i> to challenge with small molecules by genome-wide proteomics	40
--	----

## **Chapter 5**

Analysis of the background-reduced antibiotic production host <i>Streptomyces coelicolor</i> M1152 using quantitative proteomics	64
--	----

## **Chapter 6**

A novel nucleoid-associated protein specific to Actinobacteria that binds to GATC sequences	84
---	----

## **Chapter 7**

General discussion and future perspectives	110
Nederlandse Samenvatting	113

<b>References</b>	<b>118</b>
-------------------	------------

## **Appendix I**

Supplementary materials for Chapter 3	144
---------------------------------------	-----

## **Appendix II**

Supplementary materials for Chapter 4	148
---------------------------------------	-----

## **Appendix III**

Supplementary materials for Chapter 5	156
---------------------------------------	-----

## **Appendix IV**

Supplementary materials for Chapter 6	160
---------------------------------------	-----



# Chapter 1

---

## Introduction

Since the discovery of penicillin by Alexander Fleming in 1928 (Fleming, 1929), numerous lives have been saved from bacterial infections by the use of antibiotics in the clinic. However, as Fleming already predicted, antibiotic resistance emerged with the increased usage of antibiotics. At this moment in time, bacterial infections are once again a huge threat to human health (WHO, 2014). Bioactive natural products (NPs) are essential weapons in the battle between human and drug-resistant bacteria, and new drugs are constantly needed. The vast majority of the NPs are produced by microbes and have a wide variety of chemical skeletons, including polyketides synthesized by polyketide synthases (PKS), complex peptides produced by non-ribosomal peptide synthases (NRPS) or ribosomally produced and post-translationally modified peptides (RiPPs), terpenes or aminoglycosides. After the initial boom in antibiotic discovery in the so-called golden age (roughly 1950-1970), there has been a sharp decline in the number of new antibiotics discovered (Cooper & Shlaes, 2011, Baltz, 2008). This is largely due to replication in the search of novel NPs, in other words many of the compounds that are found have already been seen before. It has led to a substantial reduction in the return of investment of high-throughput screening, while at the same time synthetic antibiotics have failed to meet the needs (Kolter & van Wezel, 2016). Hence, there is an urgent need for new strategies to discover novel antimicrobial compounds from nature. To this end, a fundamental, systems-based understanding of the biosynthesis and regulation of antibiotic production is required.

Most antibiotics are secondary metabolites produced by microbes, and they are often produced in low amounts. Actinobacteria are the producers of approximately two-thirds of all known antibiotics and many other medical relevant NPs (Barka *et al.*, 2016, Hopwood, 2007). Actinobacteria have a complex mycelial life cycle that is unique as compared to other bacteria. It starts with the germination of a spore that grows out to form vegetative hyphae via tip extension and branching (Barka *et al.*, 2016, Chater & Losick, 1997). When the environmental situation requires reproduction, for example due to nutrient starvation, actinomycetes start developing areal hyphae, which eventually differentiate into chains of unigenomic spores (Claessen *et al.*, 2014, Flärdh & Buttner, 2009). The production of antibiotics correlates temporally to the developmental growth phase (Bibb, 2005, van der Heul *et al.*, 2018).

The production of antibiotics is often facilitated by a group of proteins encoded by several genes clustered in large gene cluster, which we call a biosynthetic gene cluster (BGC). The BGCs of Actinobacteria can be 100 kb or even larger in size, encoding both core enzymes and those carrying out the decorating steps, as well as resistance and transport genes (Cimermanic *et al.*, 2014). Each type of NP thereby has its own unique features that can be used to find and predict new BGCs specifying similar molecules. Bioinformaticians have made and are still improving

*in-silico* tools that are able to predict the possible BGCs from genome sequences, utilising the information extracted from known BGCs (Blin *et al.*, 2017, Skinnider *et al.*, 2016, Rottig *et al.*, 2011, Agrawal *et al.*, 2017). Harnessing the power of the artificial intelligence boom across every aspect of human life, these tools provide more information about predicted gene clusters and their likely product(s), whereby prediction accuracy is continuously being improved. The development of these *in-silico* tools combined with the rapidly increasing number of available genomes have shown that the ability of Actinobacteria to produce bioactive NPs have been grossly underestimated (Hopwood, 2007, Fedorova *et al.*, 2012).

One way to explore the hidden treasures from Actinobacteria is to try to understand the biosynthetic cell factory as a whole system and tackle the cellular machineries which are blocking the biosynthesis of interesting compounds. It is highly likely that a large part of the NP repository is expressed under specific environmental conditions, responding to interactions with other microbes and higher organisms, as well as to biotic and abiotic stresses (Seipke *et al.*, 2012, van der Meij *et al.*, 2017). At the same time, there are non-traditional NPs like peptidic NPs, which were not a focus in traditional screening methods but may still prove very valuable. Exploring their potential requires a systems-level investigation of the regulation network of the target Actinobacteria, requiring different ‘omics’ technologies. Proteomics technologies, which directly study the global protein levels of metabolic enzymes and other proteins, are becoming increasingly important in these studies. A systematic review of utilizing systems biology, especially proteomics related methods and results, in microbial NP research is provided in **Chapter 2**. In this review, it is discussed how recent advances in genomics, metabolomics and bioinformatics have helped NP discovery, with a particular focus on the application of advanced proteomics, including both biased and unbiased proteomics pipelines.

In **Chapter 3**, natural product proteomining, an important NP research pipelines developed in our laboratory (Gubbens *et al.*, 2014), was put into practice. This chapter includes two studies that both achieved fluctuations in the metabolome of a specific target strain. A workflow combining metabolomics and bioinformatics allowed us to connect a compound to its cognate BGC, namely a prenylated isatin antibiotic in *Streptomyces* sp. MBT28 and C-glycosyl-pyranonaphthoquinones in *Streptomyces roseifaciens* (MBT76). Their corresponding BGCs were confirmed by proteomics and mutagenesis studies, respectively.

In **Chapter 4**, elicitors were added to enforce fluctuation of the secondary metabolome of the gifted NP producer *S. roseifaciens*, and the concomitant changes in the protein levels analysed using high coverage full proteomics studies. We aimed to provide an overview of whole proteomics response patterns of different types of small molecules in *S. roseifaciens*. The elicitors used in this study



should mimic the environmental factors present in the natural habitat of *Streptomyces*. These molecules we used included plant hormones, oligosaccharides and pathogen-related compounds. The results show that the response to these molecules can be categorized into three main clusters, which not necessarily corresponded to the elicited antimicrobial activity. We then focused specifically on the effect of the addition of the plant hormone jasmonic acid via a growth phase-dependent proteomics study. A jasmonic acid response (*jar*) gene cluster was found that related to the antimicrobial activity of *S. roseifaciens*.

A different method in activating the expression of cryptic BGCs is heterologous gene expression. This method has the advantage of allowing highly modulated expression and simplified identification. However, to achieve this goal, an optimised production host suitable for the efficient production of NPs is required. In **Chapter 5**, the background-reduced host *Streptomyces coelicolor* M1152 (Gomez-Escribano & Bibb, 2011) was investigated using quantitative proteomics. The effects caused by the optimization in this strain on growth and production of *S. coelicolor*, including the removal of major native BGCs and introduction of an additional *rpoB*[C1298T] mutation, were analysed by comparing *S. coelicolor* M1152 to its parent *S. coelicolor* M145. This revealed interesting differences that might provide guidance in future strain optimization of *S. coelicolor* and other *Streptomyces* hosts, including a delayed response of members of the PhoP regulon and enhanced production of ectoine biosynthetic enzymes, which relate to stress responses.

Antibiotic production correlates temporally to the onset of development (Bibb, 2005, van der Heul *et al.*, 2018). In order to systematically investigate the regulation of NP production in Actinobacteria, more knowledge on the global control of development is important. In **Chapter 6**, we describe the discovery of a new family of DNA binding proteins, represented by SCO1839. SCO1839 is a nucleoid associated protein (NAP) that belongs to a family of Actinobacteria-specific regulators. The protein is expressed in the later stages of development and is highly abundant in spores. We found that the protein binds specifically to a small DNA motif centred around the palindromic sequence GATC. ChIP-Seq experiments revealed that SCO1839 binds to 1000s of sites on the chromosome and may affect the expression of some 10% of all genes in *S. coelicolor*. A positive feedback loop between SCO1839 and the global antibiotic activator AtrA was also proposed in this study.

A summary of the most important discoveries of this thesis is provided in **Chapter 7**, as well as a brief discussion and future perspectives.

# Mining for Microbial Gems: Integrating Proteomics in the Postgenomic Natural Product Discovery Pipeline

Chao Du, Gilles P. van Wezel

Proteomics (2019) 18: 1700332  
DOI: 10.1002/pmic.201700332

## Abstract

Natural products (NPs) are a major source of compounds for medical, agricultural, and biotechnological industries. Many of these compounds are of microbial origin, and, in particular, from Actinobacteria or filamentous fungi. To successfully identify novel compounds that correlate to a bioactivity of interest, or discover new enzymes with desired functions, systematic multiomics approaches have been developed over the years. Bioinformatics tools harness the rapidly expanding wealth of genome sequence information, revealing previously unsuspected biosynthetic diversity. Varying growth conditions or application of elicitors are applied to activate cryptic biosynthetic gene clusters, and metabolomics provide detailed insights into the NPs they specify. Combining these technologies with proteomics-based approaches to profile the biosynthetic enzymes provides scientists with insights into the full biosynthetic potential of microorganisms. The proteomics approaches include enrichment strategies such as employing activity-based probes designed by chemical biology, as well as unbiased (quantitative) proteomics methods. In this review, the opportunities and challenges in microbial NP research are discussed, and, in particular, the application of proteomics to link biosynthetic enzymes to the molecules they produce, and vice versa.

## A Systematic View of Natural Product Discovery

Natural products (NPs) are metabolites with specialized functions in nature, many of which have agricultural, industrial, or medical applications, such as antibiotics, antifungals, anticancer compounds, herbicides, and immunosuppressants. NPs are found in a wide variety of chemical skeletons, including polyketides synthesized by polyketide synthases (PKS), peptides produced by non-ribosomal peptide synthases (NRPS) or ribosomally produced and post-translationally modified peptides (RiPPs), terpenes, aminoglycosides, or gamma-butyrolactones. The introduction of penicillin in the 1940s showed the importance of NPs to treat infectious diseases, and this has greatly contributed to expanding human life span (Cragg *et al.*, 1997, Newman & Cragg, 2016). However, the exponential increase of antimicrobial resistance means that bacterial infections now once more pose a major threat to human health (WHO, 2014). The high frequency of rediscovery of known molecules, so-called replication, necessitates new approaches to rejuvenate drug screening (Baltz, 2008, Cooper & Shlaes, 2011, Genilloud *et al.*, 2011, Payne *et al.*, 2007). Filamentous fungi and bacteria of the order of Actinomycetales are the major producers of natural products and produce the vast majority of the antibiotics that are used in clinic (Bérdy, 2005, Newman & Cragg, 2007). Some two thirds of all antibiotics are produced by actinomycetes, the majority of which are sourced from members of the genus *Streptomyces*.

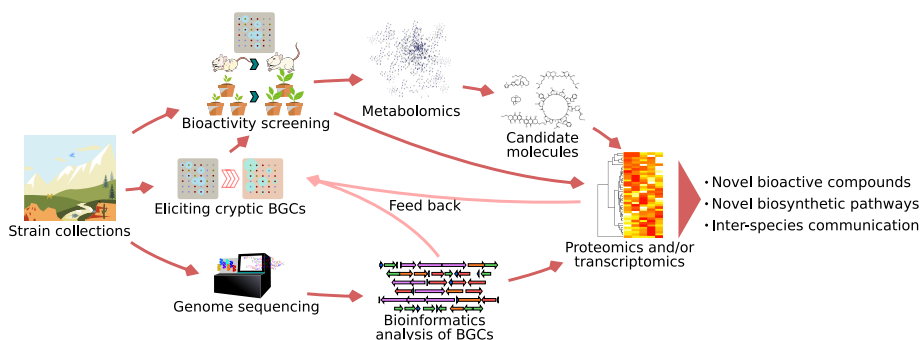
An important issue to solve is how we can identify novel NPs produced by known microorganisms. Considering that they have so far been missed in screening campaigns, it is logical to assume that many of these sought-after molecules are either expressed at a lower level than the ones that have already been identified or differ significantly in their chemical properties. The biosynthetic process is a complex system, wherein each element influences the final product and its production in often subtle ways. One issue is that under routine screening conditions, the full biosynthetic potential of the microorganisms is not visible, as many biosynthetic gene clusters (BGCs) are silent or cryptic under such laboratory conditions (Rutledge & Challis, 2015, Zhu *et al.*, 2014a). Indeed, it is highly likely that a large part of the NP repository is expressed under specific environmental conditions, responding to interactions with other microbes and higher organisms, as well as to biotic and abiotic stresses (Seipke *et al.*, 2012, van der Meij *et al.*, 2017). These issues go hand in hand with the problem of replication, in other words, that known compounds are ubiquitous and continuously rediscovered, thus frustrating efforts to discover new but often minor compounds in the NP pool (Cooper & Shlaes, 2011). At the same time, there are non-traditional NPs like peptidic NPs, which were not a focus in traditional screening methods. Thus, new methods and strategies are required, combining biological insights with new analytical and genomics tools (Kolter & van Wezel, 2016). Still, new NP discovery

needs to harness this NP “dark matter,” as these may have crucial functions and high promise for application (Baltz, 2007).

2 Following the genome sequencing revolution and the concurrent development of genomics technologies, new types of high-throughput analytical methods are emerging rapidly, offering new opportunities for drug screening. These new methods include bioinformatics of large numbers of genomes or metagenomes; whole cell or community-based transcriptomics using next-generation sequencing; proteomics based on mass spectrometry (MS); and metabolomics based on MS or nuclear magnetic resonance (NMR) spectroscopy. Application of these methods in targeted studies on NP biosynthesis gives better insights into the full potential of the producing microorganisms and will thus boost the return on investment of screening efforts. Many of the strategies that are being developed on the basis of these methods treat the biological system as a whole, aiming to find correlations between the sought-after metabolite and changes at the systems level (Figure 1). In this review, we highlight recent advances in proteomics-based technologies in combination with chemical biology, genomics and metabolomics, and their applications to NP research are discussed, with focus on microbial sources.

## Harnessing the Genome Revolution

In the later part of the twentieth century, mutational analysis combined with emerging DNA sequencing technologies allowed scientists to start to map known NPs to their corresponding BGCs, and to elucidate the biosynthetic logic. Actinobacteria are Gram-positive bacteria that are found in soil and aquatic environments, many of which have a mycelial lifestyle. They are prolific producers of NPs, which include some two-thirds of all known antibiotics, as well as many other bioactive NPs (Barka *et al.*, 2016, Bérdy, 2012). David Hopwood was a visionary when he picked *Streptomyces coelicolor* in the 1950s for its ability to produce clearly discernible pigments, as these pigments were later instrumental in the discovery of the first antibiotic BGCs, just as the genome sequence itself later formed the example for the genome sequencing revolution (Hopwood, 2007). The actinorhodin BGC served as an example for type II polyketides, whereby a spontaneous non-producer was complemented genetically to find the first piece of a large biosynthetic jigsaw (Rudd & Hopwood, 1979). A similar strategy was followed to identify the BGC for the type I polyketide undecylprodigiosin (Rudd & Hopwood, 1980). These experiments, for the first time, revealed that the genes for antibiotic production are clustered. Similarly, NRPs are specified by large gene clusters, exemplified by for example, the BGC for calcium dependent antibiotic (CDA) in *S. coelicolor* (Hopwood & Wright, 1983, Hojati *et al.*, 2002).



**Figure 1. Systematic NP research workflow.** Strains collected from different ecological niches are tested for their bioactivities, preferably in a high-throughput way, with eliciting strategies applied to optimize the chance of activating cryptic BGCs. A metabolomics workflow is then applied to find candidate molecules and to dereplicate previously identified compounds. Together with the genome sequence information, quantitative proteomics and/or transcriptomics will help identifying the biosynthetic pathway and/or regulatory network.

Indeed, in bacteria, genes for NPs are typically clustered in large gene clusters of up to 100 kb or even larger in size, with core enzymes and those carrying out the decorating steps, as well as resistance and transport genes. Each type of NP thereby has its own unique features that can be used to find and predict new BGCs specifying similar molecules. As mentioned, the landmark event in this field was the publication of the complete genome sequence of the “antibiotics factory” *S. coelicolor* (Bentley *et al.*, 2002), which cost more than 10 million pounds to complete. Soon other genomes followed suit, uncovering a wealth of yet undiscovered chemical diversity (Cruz-Morales *et al.*, 2013, Ikeda *et al.*, 2003, Ohnishi *et al.*, 2008, Oliynyk *et al.*, 2007). Combining existing knowledge of biosynthetic pathways with the ever-growing wealth of genome sequence information, the possibilities of genomics-based NP mining are seemingly endless. New bioinformatics tools are being developed to help scientists to sieve through the information and prioritize BGCs of interest, based on the computed pattern of existing protein and DNA sequences (Ziemert *et al.*, 2016). Other than mining conserved biosynthetic enzymes with tools like BLAST and HMMer manually, new genome-mining tools including antiSMASH (Blin *et al.*, 2017, Medema *et al.*, 2011), PRISM (Skinnider *et al.*, 2016), and many others have enabled high-throughput genome-mining for different classes of NPs. With all the identified gene clusters, researchers have also created tools that link back from NPs to possible protein domain organization and harness this information to find the responsible BGCs from the databases (Dejong *et al.*, 2016). Genome-mining algorithms like NRPSpredictor2 (Rottig *et al.*, 2011) and RiPPMiner (Agrawal *et al.*, 2017) utilize machine-learning technologies that form the new generation in bioinformatics. Harnessing the power of the artificial intelligence boom across every aspect of

human life, these tools provide more information about predicted gene clusters and their likely product(s), whereby prediction accuracy is continuously improved. Meanwhile, biochemists and bioinformaticists are working closely together to develop efficient dereplication pipelines, aiming at finding specifically new chemical structures. Such efforts include finding “oddly” structured gene clusters, in other words BGCs that either contain rare or unknown biosynthetic genes or have unexpected combinations of known genes. The MIBiG (minimum information about a biosynthetic gene cluster) database is an important new community-based resource that is used for dereplication purposes in genome-mining tools like antiSMASH (Blin *et al.*, 2017). Except connecting BGCs to their chemical potential and environmental diversity, MIBiG is also built to provide guidance in gene cluster engineering (Medema *et al.*, 2015). However, despite all the spectacular developments in the genome-mining technologies, there is still a tremendous amount of work to do in order to identify new molecules at a high frequency and put them into use.

## Eliciting the Expression of Cryptic BGCs

Prior to going into the challenges of how to identify new bioactive compounds, it is important to look into ways to activate the sought-after cryptic antibiotics. After all, the strain collections of large industry have been mined intensively via high-throughput screening, and rescreening these strains is only then feasible if we find conditions where the expression of a significantly large number of compounds that have previously been missed due to low abundance, is elicited. To develop eliciting approaches, we need to understand the underlying regulatory networks that control the expression of the BGCs. Here, we will explain some of the most important principles. For more extensive reviews on the control of antibiotic production, the readers are referred elsewhere (Bibb, 2005, Liu *et al.*, 2013, Sanchez *et al.*, 2010, van Wezel & McDowall, 2011). As mentioned, the genes for NPs are typically clustered. Two levels of control exist, namely by cluster-situated regulators (CSRs) that activate or repress a specific BGC, and by global regulators that often respond to environmental signals and transmit these to a range of BGCs and other metabolic pathways (Liu *et al.*, 2013). To activate a specific BGC of interest, over-expression of the CSR or changing the upstream region with all its regulatory elements by the promoters in the cluster is an effective method. At this moment in time, such approaches are hardly amenable to high-throughput application. However, new strategies are under active development to explore and exploit possibilities of NP-producing microorganisms (Loureiro *et al.*, 2018).

Manipulation of NP BGCs in many bacteria requires interference with global regulatory networks, allowing generic eliciting strategies such as the addition of specific elicitor molecules to the growth media. The first example of a fully elucidated global regulatory cascade toward the onset of antibiotic production is

that controlled by the nutrient sensory GntR-family regulator DasR. DasR connects the pathways for aminosugar metabolism and transport and secondary metabolism (Craig *et al.*, 2012, Rigali *et al.*, 2006, Rigali *et al.*, 2008, Świątek *et al.*, 2012). DasR is a highly pleiotropic regulator, as demonstrated by recent systems biology analyses of chitin- or N-acetylglucosamine-grown cultures of *S. coelicolor* (Nazari *et al.*, 2012, Świątek *et al.*, 2015). DasR directly controls the transcription of genes for actinorhodin, prodiginines, calcium-dependent antibiotic and cryptic polyketide Cpk, as well as siderophores in *S. coelicolor*, as shown by transcript assays and systems-wide DNA binding experiments using ChIP-chip analysis (Świątek *et al.*, 2015). The activity of DasR is modulated by phosphorylated metabolic derivatives of N-acetylglucosamine (GlcNAc), in particular, GlcNAc-6P and glucosamine-6P (GlcN-6P). Addition of GlcNAc to the culture media activates antibiotic production in several actinomycetes, including cryptic antibiotics (Rigali *et al.*, 2008). This technology is now being applied in global screening approaches.

Another well-studied global regulatory cascade revolves around CebR, a cellulose utilization regulator that also regulates the production of the phytotoxin thaxtomine. Thaxtomine is a cellulose biosynthesis inhibitor that is produced by *Streptomyces scabies* (Marushima *et al.*, 2009, Francis *et al.*, 2015, Jourdan *et al.*, 2018). CebR directly controls the thaxtomine BGC and inhibits the expression of the pathway-specific activator gene *txtR*. Cellobiose and cellotriose inhibit the affinity of CebR for DNA, resulting in relieve of CebR repression (Francis *et al.*, 2015). Scanning (cryptic) BGCs for binding sites of known regulators is an excellent approach to find elicitors that activate their expression (Rigali *et al.*, 2018). Identification of CebR or DasR binding sites inside a BGC of interest would logically predict that the addition of cellobiose or GlcNAc, respectively, will derepress the BGC. As soon as more of such “lock and key” combinations have been identified, the arsenal of eliciting strategies and hence screening regimes will rapidly increase. Interestingly, recent data have shown that cross-talk exists between global regulatory networks, such as between DasR and AtrA (Nothaft *et al.*, 2010), adding an additional level of complexity, whereby the specific response at the BGC level is not always easily predicted (Urem *et al.*, 2016).

Co-culturing is another promising way of eliciting NP biosynthesis as it mimics traits of competition in natural environments. Co-culturing of streptomycetes revealed that a large number of species produces antibiotics in response to neighbouring streptomycetes (Abrudan *et al.*, 2015, Ueda *et al.*, 2000). The mechanisms of the interaction between species is often complicated. In the interaction between two *Streptomyces* strains, promomycin, a compound that one of these strains produces, not only acted as an antibiotic, but also elicited antibiotic production of many other *Streptomyces* strains (Amano *et al.*, 2010). In addition to environmental approaches, targeting known cellular components and processes by druggable pharmaceutical compounds is a promising alternative strategy



(Craney *et al.*, 2012, Pimentel-Elardo *et al.*, 2015). Fungal chromatin was successfully targeted via histone deacetylase (HDAC) or DNA methyltransferase inhibitors (Palmer & Keller, 2010). Epigenomic manipulation by HDAC inhibitors influences the expression of a large numbers of genes, including cryptic BGCs (Cichewicz, 2010). Novel aspercryptins (Henke *et al.*, 2016) and other compounds with new structures (Mao *et al.*, 2015) were discovered based on this method, underlining its utility for screening. The concept of using small molecules as elicitors by perturbing the biological system has been extended with high-throughput screening of small molecule libraries and has shown promising results (Craney *et al.*, 2012). For more details on environmental and HT screening approaches to find elicitors, the readers are referred to recent reviews (Okada & Seyedsayamdost, 2017, van der Meij *et al.*, 2017).

## Proteomic-Based Approaches in Natural Product Mining

### Chemical Biology: Activity-Based Probes

An element commonly found in the NP-synthetic machinery is the carrier protein (CP) domain, which acts as an anchor for tethering biosynthetic intermediates in PKS, NRPS, and fatty acid synthase (FAS) systems (Lai *et al.*, 2006). CP domains can be labelled enzymatically by using CoA analogues and PPTase (in vitro) or tagged CoA precursors (in vivo) (La Clair *et al.*, 2004, Meier *et al.*, 2006). Alternatively, activity-based protein profiling (ABPP) may be used as a proteomics strategy to identify enzymatic activity in complex biological samples. Here, the active site is labelled with a covalent reporter probe, often a labelled inhibitor, which can be detected quantitatively either using protein gel electrophoresis or via gel-free methods such as LC-MS/MS and microarray platforms (Evans & Cravatt, 2006). The highly modular characteristics of PKS and NRPS biosynthetic systems makes them an excellent target for ABPP. Specific fluorescently labelled probes have been designed to target the acyltransferase (AT) and thioesterase (TE) domains (Meier *et al.*, 2008). Individual adenylation (A) domains were later targeted in similar manner (Konno *et al.*, 2015). ABPP is suitable for low-cost analysis of NP biosynthetic potential of many strains and samples, eliminating the use of LC-MS/MS in the initial screening steps while retaining the compatibility with online methods (Meier *et al.*, 2008). Activity-based probes can also act as enrichment intermediates to enhance the selectivity and dynamic range of LC-MS/MS analysis (Meier *et al.*, 2009). This enrichment process can be of particular value for the identification of NP biosynthetic machineries that are poorly expressed.

Combination of ABPP with CP-domain enzymatic labelling methods, as well as downstream tandem MS, was used to set up the so-called orthogonal active site identification system (OASIS). In OASIS, CP domains and TE domains are labelled in tandem using the biotinylated probes mentioned before and enriched, followed by offline multidimensional LC-MS/MS analysis (Meier *et al.*, 2009). OASIS was

first used to detect all of the known type I PKS and NRPS systems of *Bacillus subtilis*. The utility of the method in biosynthesis studies was underlined by the good correlation of the activity of surfactin biosynthesis with the phenotypic differences between individual strains (Meier *et al.*, 2009, Meier *et al.*, 2008).

ABPP methods share the limitation that only a small number of domains can be targeted. The limited number of detectable active domains greatly limits the target range of activity-based methods. When a domain is absent, as is the case in many modular systems, the whole system will in principle be eliminated from the data sets (Hertweck, 2009). Also, type II PKS systems are less amenable to such profiling approaches due to the fact that the biosynthetic domains are encoded by different genes, while in type III PKS systems, the KS domains perform all the functions (Shimizu *et al.*, 2017). Another problem that should be taken into account is that in vivo labels often act as inhibitors, which may well affect the target strain or its biosynthetic profile, and hence the results are not unbiased.

Targeted proteomics is applied in the deconvolution of bioactive NPs. Chemical proteomics thereby allows identifying the cellular targets for NPs. In this approach, the bioactive molecule of interest is immobilized and used as a bait for proteins from a target organism. Proteins that have significant affinity for the immobilized NP are then considered as potential targets, and the nature of these proteins can be resolved using proteomics. This will help elucidating the mode of action, and the cellular pathway targeted by the NP. For applications of chemical proteomics in target deconvolution, the readers are referred elsewhere (Rix & Superti-Furga, 2009).

### Direct Proteomic Analysis of Biosynthetic Enzymes

A common characteristic of NRPS and PKS systems is that the biosynthetic machineries are very large, with the synthetases comprising of many domains to form a molecular assembly-line (Fischbach & Walsh, 2006). The growing NP scaffold is covalently tethered to carrier protein (CP) domains by a thioester bond (phosphodiester linkage) to the 4'-phosphopantetheine (PPant) arm, a post-translational modification on a serine residue in the CP active site. This PPant modification can be easily detached from the peptide by infrared multiphoton dissociation (IRMPD) or collision-induced dissociation (CID) techniques used in tandem MS; this offers a good way to detect the presence of CP domains (Meluzzi *et al.*, 2008). Taking advantage of the large size and the unique PPant marker ions, the Proteomic Investigation of Secondary Metabolism (PrISM) method was developed (Bumpus *et al.*, 2009). In PrISM, high-molecular weight proteins of the complex whole-cell mixture are prefractionated by SDS-PAGE, digested and subjected to LC-MS/MS. When PPant diagnostic ions are detected via tandem MS, the corresponding peptide is identified as part of an NP synthase. Based on the

genome sequence, reverse genetics will allow obtaining the gene and hence also the BGC (Bumpus *et al.*, 2009).

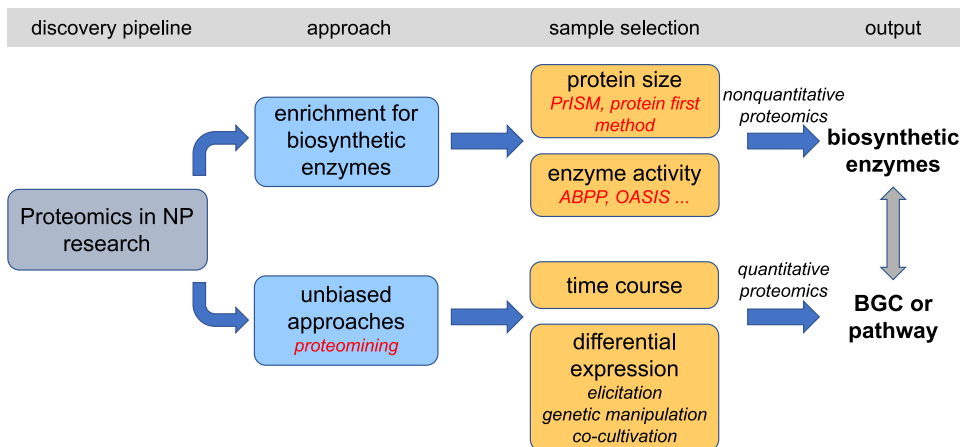
Alternatively, the method can be applied directly to proteins isolated from SDS-PAGE gels without identifiable PPant ions, when the sequence is sufficiently different from that of known biosynthetic proteins. An example is the proteomic analysis of large proteins from *Bacillus* isolate NK2003 extracted from SDS-PAGE gels followed by reverse genetics and comparison to the genome sequence revealing the presence of a new NRPS, which was shown to produce the novel NRP koranimine (Evans *et al.*, 2011). This method was later applied to actinomycetes whose genome sequence had not yet been determined, thereby uncovering 10 NRPS/PKS gene clusters from six strains (Chen *et al.*, 2012). This protein-first method was also used in the discovery of a new class of peptide aldehyde NP called flavopeptins (Chen *et al.*, 2013), thus showing the utility of proteomics-based drug discovery approaches. However, in reality, the costs of genome sequencing have dropped so steeply that most strains that are analysed in detail will generally have been sequenced. The availability of strain-specific databases greatly reduces search time and, at the same time, increases the identifiable peptide number by more than one order of magnitude (Albright *et al.*, 2014). In a case study, the authors identified both known NPs (griseobactin and rakicidin D) and a new siderophore-like compound in *Streptomyces*.

### Unbiased Analysis of the Total Proteome

ABPP and PrISM methods are biased toward the known characters of NP producing enzymes, including a known bioactivity relating to well-conserved domains for ABPP, or in the case of PrISM, on size and the PPant arm. While there is still a great undiscovered biodiversity of enzymes that harbour such characteristics, the bias it generates restricts the target scope to a relatively narrow range. The massive technological advances in genome sequencing and MS-based proteomics make unbiased full proteome investigation much more feasible, offering a great opportunity for NP discovery (Figure 2). The rapidly emerging data-independent and hyper-reaction monitoring MS are pushing this opportunity to a new level by recording the MS/MS spectra for all available MS ions, whereby the raw data will remain available for future implementation (Law & Lim, 2013, Chapman *et al.*, 2014). Also, label-free quantitative proteomics strategies enable the analysis of a nearly unlimited number of samples (Patel *et al.*, 2009).

The first unambiguous proteomics study of biosynthetic enzymes was performed on the proteome of *Myxococcus xanthus*, which was subjected to orthogonal 2D offline HPLC separation, followed by tandem MS in a data-dependent mode (Schley *et al.*, 2006). The study not only successfully identified proteins from four known NRPS and PKS genes, but also identified proteins from several unknown NRPS and PKS genes. Some of these were smaller than 200 kDa and contained no

domains that could be detected using molecular probes. This approach was applied successfully to analyse the NP-producing potential of nitrogen-fixing *Frankia* species (Udwary *et al.*, 2011).



**Figure 2. Integration of proteomics strategies in natural product discovery pipelines.** Integration of proteomics strategies in natural product discovery pipelines. The biased (top) proteomics pipelines are directed at the identification of specific biosynthetic enzymes or bioactive peptides, based on biochemical principles. The unbiased methods (bottom) aim to statistically connect biosynthetic pathways to a bioactivity of interest.

### Natural Product Proteomining: Proteomics to Link NPs to the Biosynthetic Enzymes

Next-generation sequencing combined with whole-cell proteomics offers unprecedented analytical power to detect the biosynthetic machinery of a strain under a specific growth condition. At the same time, as mentioned before, MS- and NMR-based metabolomics allow the efficient profiling of the NPs produced under different growth conditions (Wu *et al.*, 2015b, Gao *et al.*, 2014).

Connection of these data sets will provide guidance for the identification of new NPs and the corresponding biosynthetic enzymes, and for optimization of the production of compounds of interest, as shown for the fungus *Aspergillus fumigatus* (Owens *et al.*, 2014) and for *S. coelicolor* (Gubbens *et al.*, 2014). The latter work describes a quantitative proteomics platform called natural product proteomining developed on the basis of these methods, to readily connect bioactive compounds to their biosynthetic enzymes based on statistical analysis of cultures grown under different conditions. Since the method is based on the combination of unsupervised (statistical) methods to connect a bioactivity of interest to its cognate BGC, prior knowledge of the BGC or the nature of the bioactive molecule is in principle not required (Figure 2). The pipeline entails ensuring significant

fluctuation of a given bioactivity between cultures, for example, by varying growth conditions, adding elicitors, or—in the case of a BGC of interest—by mutation or over-expression of the regulatory gene(s). The cultures are then analysed by NMR- or MS-based metabolomics to find all NPs in the sample, and by proteomics to obtain insights into the gene expression profiles. In this way, all NPs and biosynthetic enzymes that correlate statistically to the bioactivity of interest are identified (Gubbens *et al.*, 2014). Analysis of mutants for the global regulatory genes *dasR* and *rok7B7* in *S. coelicolor* proved the utility of the method, with strong correlation between the level of each of the known NPs and the expression of their cognate BGCs.

Natural product proteomining was applied for NP mining of the soil isolate *Streptomyces* sp. MBT70, which was grown under five different culturing conditions, whereby bioactivity was correlated to the metabolome (using semiquantitative NMR-based metabolomics) and to the protein profiles of its biosynthetic enzymes (using proteomics). This readily identified the BGC and its product, a novel naphthoquinone antibiotic called juglomycin C (Gubbens *et al.*, 2014). For these methods to be successful, a prerequisite is to achieve significant fluctuation in the metabolome. This was achieved via the use of elicitors or changing growth conditions (Gubbens *et al.*, 2014, Gubbens *et al.*, 2017) or via altered expression or mutation of regulatory genes via directed or random mutagenesis (Wu *et al.*, 2017, Wu *et al.*, 2015a).

The major alternative to proteomics-based profiling is the application of transcriptomics. This has the advantage of offering very high resolution in terms of changes in the global expression profiles, as in principle every single gene is identified via DNA microarrays or RNA-seq. One issue hereby is that these methods are more costly than proteomics, and, in particular, that RNA is prone to degradation. With the large numbers of genes whose transcription is altered, it may be difficult to see the key effects of a certain perturbation of the system. Tools like DAVID (Huang *et al.*, 2009) and GSEA-P (Subramanian *et al.*, 2007) address this challenge by providing general classification of classes of genes whose expression has been altered, thus allowing the detection of global changes in the system. Recent developments in transcriptome modelling include tSOT (Kim *et al.*, 2016) and BeReTa (Kim *et al.*, 2017) that aim at finding otherwise unrecognizable regulatory patterns in NP production. Combination with regulon prediction algorithms such as PredictRegulon (Yellaboina *et al.*, 2004) and PREDetector (Hiard *et al.*, 2007) that scan genomes for common regulatory elements within a network, can provide guidance in the design of approaches to specifically alter the transcription of subclasses of genes, such as those for NPs.

## Proteomics-Based Analysis of Peptidic Natural Products

Peptidic NPs (PNPs) are naturally applicable to tandem MS technology, which is widely used in shot-gun proteomics to perform *de novo* peptide sequence elucidation (Mohimani & Pevzner, 2016, Ng *et al.*, 2009). This is not a trivial task, as tandem MS spectra are often complex and ambiguous. Several new programs have been developed with the focus of *de novo* identification of peptides from tandem MS data. By integrating artificial intelligence into the pipeline, programs like DeepNovo have achieved greater peptide identification rates and accuracy as compared to other algorithms (Tran *et al.*, 2017). Yet the *de novo* identification of unknown PNPs remains challenging, in particular, in the light of their high biodiversity and complexity (Duncan *et al.*, 2010). First, PNPs are extensively modified peptide sequences and only a small number of the modification patterns are known, and they are often very hard to predict from the primary sequence. Second, PNPs use numerous non-proteinogenic amino acids rather than the 20 canonical amino acids. Third, except that PNPs often contain disulphide bridges, they may include cyclic, branched-cyclic, or even higher order structures. And last but not least, bioactive PNPs, like all other NPs, may function synergistically, and such information is often lost during fractionation. These factors make MS analysis of PNPs highly challenging, in particular, as most of the compounds are not yet in the databases.

One technology to address this problem is natural product peptidogenomics (NPP), which aims at the rapid characterization of ribosomally produced posttranslational peptides (RiPPs) and non-ribosomally synthesized peptides (NRPs) and their BGCs from sequenced organisms (Kersten *et al.*, 2011). The NPP workflow starts from MS analysis of peptidic compounds in a sample, followed by enrichment of the peptides and their analysis via tandem MS. Searchable sequence tags are then generated, and the mass shifts are considered as possible posttranslational modifications (PTMs). Thus, the complexity of the tandem MS data is greatly reduced, and the peptide tags can be efficiently generated for data mining. Another proteomics-based PNP discovery pipeline is PepSAVI-MS (Kirkpatrick *et al.*, 2017). This pipeline starts with the fractionation of small peptides from biological samples, and then analyses the bioactivity of these fractions in a high-throughput manner. The bioactive fractions are further analysed using LC-MS/MS, allowing the statistical linkage of the bioactivity to specific peptides. The sequence and PTMs of the peptides can then be derived from the LC-MS/MS data. A regression model has been included to account for possible peptides that co-contribute to the combined activity from different fractions.

Known PNPs can be dereplicated by using the excellent databases that are currently available to the community, such as BAGEL (van Heel *et al.*, 2013), MIBiG (Medema *et al.*, 2015), and NORINE (Flissi *et al.*, 2016). Tools developed to identify the novel PNPs include the NRP-dereplication algorithm for cyclic

peptides (Ng *et al.*, 2009), informatics search algorithm for natural products (iSNAP) (Ibrahim *et al.*, 2012), and DEREPLICATOR (Mohimani *et al.*, 2016). The latter provides a high-throughput tool featuring the power of molecular networking techniques applied in metabolomics, thereby greatly increasing the publicly available mass-spectral library of PNPs in the GNPS network. Currently, many projects are ongoing worldwide that focus on identifying PNPs, making use of MS-based technologies. The high sensitivity and automation potential of MS technologies is thereby combined with the possibility of connecting proteomics and metabolomics as an effective method to dereplicate compounds and connect them to their cognate BGC. We therefore expect that many new PNPs (and new PNP families) will be uncovered in the years to come.

## Future Perspectives

After the major successes of the golden era of drug discovery, it has become increasingly difficult to find truly novel bioactive compounds. How do we find the proverbial needles in the haystack? Scientists are turning to microbes from remote and extreme environments or attempting to unlock the potential of the reservoir of cryptic NPs. To avoid wasting time and resources on known molecules and their BGCs, scientists require new tools that allow them to obtain such information as quickly as possible. While initially projects were based on single “omics” technologies, modern-era drug research approaches should follow a multiomics strategy (Figure 1). Taking advantage of the numerous genome sequences available, genomics has become the most advanced “omics” technology. The development of genome-mining tools for NPs has already set a good foundation for other “omics” technologies, in particular, transcriptomics and proteomics. MS-based proteomics pipelines not only provide identification of the biosynthetic machinery, but also provide direct identification on the end product, that is, the PNPs. New generations of mass spectrometers allow switching between proteomics and metabolomics modes, providing the analytic power of both “omics” technologies. This development will foreseeably increase the number of inter-omics studies in NP research. However, low-abundant proteins are still notoriously hard to be identified in proteomics experiments. Furthermore, peptides bearing unknown posttranslational modifications are typically missed as well. Thus, instruments with high resolving power, high capacity, and low detection limits are needed, as well as extensive databases for proteomics NP research. The current development of innovative MS technologies now offers unparalleled resolution.

Scientists are rapidly moving from biased protein enrichment methods, based on size or activity, toward non-enriched, unbiased whole proteome analysis methods. Still, protein enrichment methods greatly simplify the protein samples, which gives some advantage in screening a large number of samples. Also, when dealing with low abundant biosynthetic enzymes or those that are post-translationally modified,

the enrichment process will help in their identification. On the downside, enrichment discards large amounts of information regarding the regulatory network and metabolite flow in a given host. Unbiased proteomics methods allow analysing the biosynthesis process as a complete system involving not only the biosynthetic pathway, but also other cellular or inter-cellular systems including but not limited to primary metabolic pathways, developmental pathways, and regulatory networks. In addition, whenever new knowledge is acquired on NP biosynthesis, scientists may revisit their old datasets and find possible new correlations. This advantage is enhanced by data-independent MS, which allows reanalysing data from raw MS-MS/MS spectra. Another limitation for the traditional isotope labelling quantitation methods is sample limitation due to the limited number of available labels. This issue can be resolved by using label-free quantitative MS. Without labelling steps, high-resolution time series experiments are possible, thereby approaching the scope of transcriptomics. The use of multiple replicates will greatly enhance data reproducibility for, in particular, low-abundance proteins, including important regulatory components in NP biosynthesis. However, all the advantages of recent technology improvements are accompanied by equally large challenges. The computational power to process the exponentially increasing data size is becoming limited. Furthermore, interpretation of the abundance of thousands of proteins demands extensive knowledge of the whole biological system. Like GNPS and MIBiG, scientists are building a dereplication pipeline for proteomics. The ProteomeXchange consortium ([www.proteomexchange.org](http://www.proteomexchange.org)) aims to connect online proteomics databases like PRIDE ([www.ebi.ac.uk/pride](http://www.ebi.ac.uk/pride)) and many others into one universal access point, in order to help researchers to share and explore each other's datasets. We believe that this is a very promising development for NP research.

Multimomics studies are part and parcel of strategies to discern novel structural diversity in the huge chemical space. Thus, for successful dereplication strategies, and to optimally harness the genome sequence and biosynthetic information, scientists need to understand and integrate all types of “next generation” methods. And this will further help scientists to breach barriers toward identifying truly novel drugs that are needed to keep diseases at bay.

## Acknowledgements

The work was supported by grant 14221 from The Netherlands Organization for Scientific research (NWO-TTW).





## Chapter 3

---

# Application of systems biology methods for the identification of novel natural products in *Streptomyces* species

Chao Du, Changsheng Wu, Jacob Gubbens, Koji Ichinose, Young Hae Choi, and Gilles P. van Wezel

Adapted from:

Wu, C., Du, C., Gubbens, J., Choi, Y.H., and van Wezel, G.P. (2015) Metabolomics-driven discovery of a prenylated isatin antibiotic produced by *Streptomyces* species MBT28. J. Nat. Prod. 78: 2355-2363.  
DOI: 10.1021/acs.jnatprod.5b00276

Wu, C., Du, C., Ichinose, K., Choi, Y.H., and van Wezel, G.P. (2017) Discovery of C-Glycosylpyranonaphthoquinones in *Streptomyces* sp. MBT76 by a combined NMR-based metabolomics and bioinformatics workflow. J. Nat. Prod. 80: 269-277.  
DOI: 10.1021/acs.jnatprod.6b00478

## Abstract

Actinomycetes are a major source of antimicrobials, anticancer compounds, and other medically important natural products. Genome mining has revealed an extensive biosynthetic potential of Actinomycetes. Major challenges in the screening of these microorganisms are to activate the expression of cryptic biosynthetic gene clusters and the development of technologies for efficient dereplication of known molecules. Here we applied a systems biology approach consisting of NMR-based metabolomics, proteomics, bioinformatics, and targeted activation of cryptic gene clusters as an effective pipeline for bioactive microbial natural products. We achieved the identification of a previously unidentified isatin-type antibiotic produced by *Streptomyces* sp. MBT28, and novel C-glycosyl-pyranonaphthoquinones in *Streptomyces roseifaciens* MBT76 (NCCB 100637, DSM 106196). Genome mining and proteomics study on production fluctuated mutants helped in 7-prenyl isatin and related isatin type antibiotics characterization. The *qin* gene cluster from *S. roseifaciens* was highlighted by genome mining as a potential producer of novel glycosylated pyranonaphthoquinones. An approach consisting of combined NMR-based metabolomics and regulator over-expression was applied to activate the biosynthetic gene cluster and identify novel C-glycosyl-pyranonaphthoquinones.

## Introduction

Discovery and development of antibiotics to fight bacterial diseases is one of the greatest triumphs of modern medicine. Bacterial infections were once conquered by the use of antibiotics. However, the increasing antimicrobial resistance now again form a major threat to human health (WHO, 2014). In this battle, bioactive natural products (NPs) provide us prolific weapon sources in the battle against drug resistant bacteria. As producers of approximately two-thirds of all known antibiotics and many other medically relevant natural products, actinomycetes are a major source of bioactive NPs (Barka *et al.*, 2016, Hopwood, 2007). Whole genome sequencing revealed that actinomycetes harbour numerous silent and hence likely untapped biosynthetic gene clusters (BGCs) that may not yet be associated with known metabolites. Indeed, even the extensively studied model organism *Streptomyces coelicolor* was shown to possess a far greater potential than anticipated (Bentley *et al.*, 2002). Furthermore, bioinformatics tools developed specifically for mining genome sequences for the identification of biosynthetic gene clusters allow the prediction of the chemical output on the basis of accumulated biosynthetic knowledge (Medema *et al.*, 2015, Fedorova *et al.*, 2012). These developments in genome mining mark the start of a new era of genomics-based drug discovery, with the potential of greatly expanding the chemical space of bioactive natural products. A bottleneck is that many of the biosynthetic pathways uncovered by genome sequencing are in a dormant state under routine laboratory conditions, generally referred to as cryptic or silent gene clusters, and specific approaches to activate their expression are required (Rutledge & Challis, 2015, Zhu *et al.*, 2014a).

One way of activation cryptic BGCs utilizes different culturing conditions, allowing fluctuation of the production of possibly cryptic antibiotics, followed by the metabolic profiling-based identification of the bioactivity of interest. Systems biology methods can then step-in and reveal the genetic source of the bioactive compounds and its production mechanism. One challenge in this research path lies in finding the appropriate chemical triggers or ecological cues to elicit the production of cryptic antibiotics (Zhu *et al.*, 2014a, Yoon & Nodwell, 2014). Strategies that have been employed include heterologous expression (Schümann & Hertweck, 2006), the use of chemical elicitors (Zhu *et al.*, 2014b, Craney *et al.*, 2012), and inducing antibiotic resistance (Ochi *et al.*, 2014, Hosaka *et al.*, 2009). Selection of streptomycin- or rifampicin-resistant mutants, typically caused by mutations in *rpsL* (ribosomal protein S12) or *rpoB* (RNA polymerase  $\beta$ -subunit), respectively, enhances the production of antibiotics (Hosaka *et al.*, 2009, Wang *et al.*, 2008, Tamehiro *et al.*, 2003) and may also lead to the production of novel antibiotics (Fu *et al.*, 2014). Another method to find and activate the production of cryptic antibiotics is through direct genome mining and genetic manipulation of interesting BGCs. In bacteria, biosynthetic genes are often clustered in large gene

clusters, with core biosynthetic enzymes and those patriated in decorating the products. In addition, the controlling regulators as well as enzymes related with the resistance are also encoded in the gene cluster. The number of functional enzymes in BGCs being revealed is increasing fast, which made genome mining tools including antiSMASH (Blin *et al.*, 2017), PRISM (Skinnider *et al.*, 2016) and many others capable of more and more accurate BGC discovery (Du & van Wezel, 2018). New biochemical potentials normally lie in the new combinations of decorating enzymes found by genome mining, studying the new potential in genome mining results becomes a good method in finding BGC of interest. When a potentially new BGC is located, activating its activity becomes essential again. This can be achieved by fluctuating the growth environments, as is discussed above. But now we can activate these cryptic BGCs by direct manipulating the regulators located in and out of the gene cluster. Generally, in actinomycetes, the transcription of genes encoding the biosynthetic machinery for secondary metabolites involves multiple regulatory cascades and networks (van Wezel & McDowall, 2011). The regulatory signals are transmitted through global regulatory networks and ultimately transmitted to the pathway-specific regulatory genes that control the expression of the biosynthetic genes (Bibb, 2005). Molecular biology strategies, including manipulating the pleiotropic regulators and pathway-specific regulators to circumvent the regulatory can possibly trigger the biosynthesis of the corresponding natural products (Rutledge & Challis, 2015, Abdelmohsen *et al.*, 2015).

Once differential expression of a bioactivity is achieved in a producing organism, e.g., by varying culturing conditions or by making mutants, the next challenge is to rapidly establish whether the natural product is sufficiently novel to warrant full structure elucidation. Metabolomics approaches globally identify the low molecular weight metabolites in biological samples. Chemical profiling techniques using LC-MS or NMR followed by multivariate data analysis allow scientists to compare and detect molecules that are differentially synthesized between diverse biological samples, thereby effectively narrowing down the search for the sought-after biomarker and avoiding chemical redundancy in an early stage (Wu *et al.*, 2015b, Krug & Muller, 2014). At the same time, the more readily available bioinformatic tools allows prediction of the metabolite produced by the biosynthetic gene clusters, and this knowledge can be directly related to the  $^1\text{H}$  NMR spectrum by examining the expected chemical shift and/or splitting pattern of typical protons in the predicted molecular motifs. Together with proteomics and/or transcriptomics technologies applied, it is possible to bridge the gap between bioinformatics-driven gene cluster analysis and experimental NPs discovery (Wu *et al.*, 2016a, Du & van Wezel, 2018).

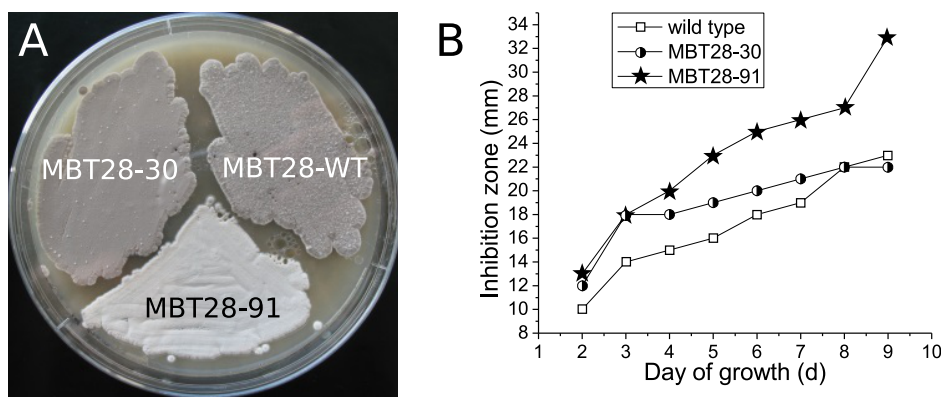
In this chapter, we first describe the use of a proteomics-based genome mining approach in combination with multivariate data analysis correlated bioactivity of

NMR peaks, facilitating the characterization of the previously undescribed antibiotic 7-prenylisatin and its possible BGC. In a workflow of using NMR-based metabolomics and bioinformatics to identify novel pyranonaphtoquinones in *S. roseifaciens*, we made direct changes in the genome to consecutively express the pathway specific activator of a cryptic type II PKS gene cluster (designated *qin*). This resulted in the activation of the production of a family of pyranonaphtoquinones with intriguing chemical architecture.

## Results and discussion

### Streptomycin resistant mutant of *Streptomyces* sp. MBT28 with elevated bioactivity caused by 7-prenylisatin production

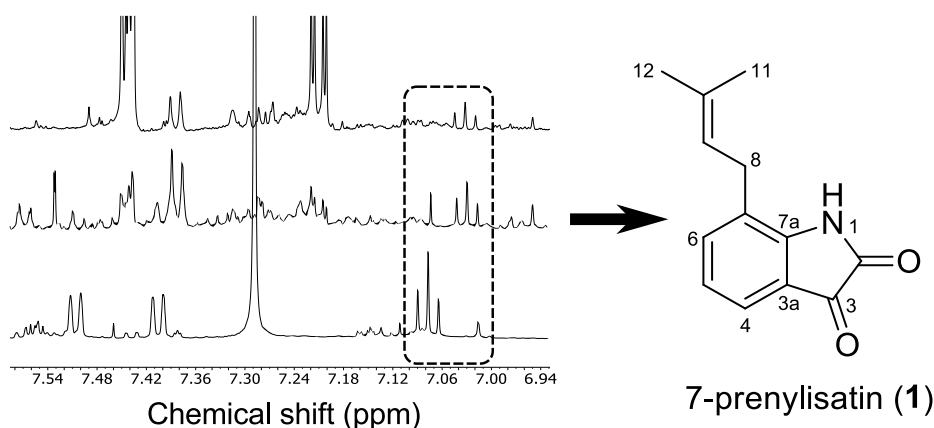
*Streptomyces* sp. MBT28 is a strain of our collection of actinomycetes from remote mountain soils (Zhu *et al.*, 2014b). In this collection, *Streptomyces* sp. MBT28 was identified as a producer of various antibiotics, one of which was identified as borrelidin (Zhu *et al.*, 2014b). However, other secondary metabolites were produced in very low quantities. In an attempt to enhance antibiotic production, two streptomycin resistant ( $\text{Str}^R$ ) mutants are selected from SFM agar plates containing  $10 \mu\text{g}\cdot\text{mL}^{-1}$  streptomycin. One mutant with a white phenotype (*Streptomyces* sp. MBT28-91) shows strong bioactivity and another with a wildtype phenotype (*Streptomyces* sp. MBT28-30) was chosen (Figure 1). They both have less enhanced bioactivity. We checked the streptomycin-resistant hot spot gene *rpsL* and *rsmG* in the two  $\text{Str}^R$  strain by sequencing, only *Streptomyces* sp. MBT28-30 carried one mutation on *rsmG* gene.



**Figure 1. Antibiotic activity of *Streptomyces* sp. MBT28 and its streptomycin-resistant mutants.** Morphological appearance of *Streptomyces* sp. MBT28 (WT) and its  $\text{Str}^R$  mutants *Streptomyces* sp. MBT28-30 and *Streptomyces* sp. MBT28-91 grown on SFM agar for 5 days at  $30^\circ\text{C}$  (A), and time course of inhibition activity against *Bacillus subtilis* (B). For activity test, strains were grown on MM agar supplemented with 0.5% mannitol and 1% glycerol, samples taken at 24 h intervals (X axis) were overlaid with *B. subtilis* to give inhibition zone (Y axis). MBT-28-91 showed much better activity than those of *Streptomyces* sp. MBT28 and *Streptomyces* sp. MBT28-30.

NMR spectroscopy, particularly  $^1\text{H}$  NMR, commonly used for the analysis of industrial natural products, is suited for such analyses as it is fast, reproducible, and benefits from a relatively easy sample preparation (Kim *et al.*, 2010). Multivariate data analysis (MDA) were used to reduce the dimensionality of

multivariate dataset and thus to discriminate samples. For this purpose, *Streptomyces* sp. MBT28, *Streptomyces* sp. MBT28-30, *Streptomyces* sp. MBT28-91 were grown on MM agar plates supplemented with 0.5% (w/v) mannitol and 1% (w/v) glycerol as carbon sources. Ten biological replicates were performed for each sample.  $^1\text{H}$  NMR was done, and the results show distinct difference of antibiotic-overproducing mutant *Streptomyces* sp. MBT28-91. Multivariate data analysis of NMR signals with bioactivity shows one clear signal related with the increased bioactivity in *Streptomyces* sp. MBT28-91. NMR-guided fractionation (Grkovic *et al.*, 2014) tracing the characteristic signal resulted in the identification of 7-prenylisatin (Figure 2). The NMR-based metabolomics and compound identification are described in detail in (Wu *et al.*, 2015a).



**Figure 2. NMR-guided purification of 7-prenylisatin from *Streptomyces* sp. MBT28-91.** Characteristic proton signals  $\delta$  7.03 (t,  $J$  = 7.2 Hz) that correlated to the high bioactivity are boxed, which were used as tracking signals for chromatographic separation (dash boxed). Chemical shift divergence was due to the different deuterated solvents used (methanol- $d_4$  for top and middle rows, and  $\text{CDCl}_3$  for bottom row).

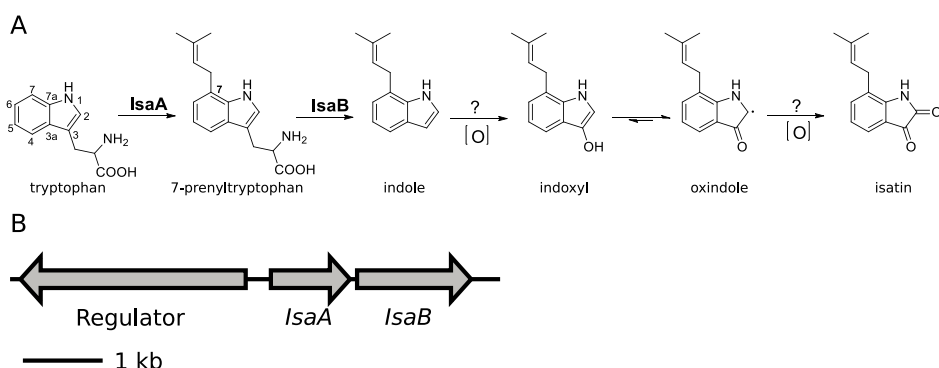
Isatins exhibit a variety of biological activities such as antibacterial, antifungal, and anticancer properties (Vine *et al.*, 2009). However, only a few isatin-type compounds of microbial origin have been identified (Sobolevskaya *et al.*, 2009, Breinholt *et al.*, 1996, Graefe & Radics, 1986). 5-prenylisatin from the fungus *Chaetomium globosum* was reported to have antifungal activity (Breinholt *et al.*, 1996), while 6-prenylisatin was the first example of an isatin derivative discovered in a streptomycete, namely *Streptomyces albus*, which showed antimicrobial activity against Gram-positive bacteria, with an MIC against *Bacillus subtilis* of 20  $\mu\text{g}\cdot\text{mL}^{-1}$  (Graefe & Radics, 1986). Purified 7-prenylisatin was tested for its efficacy against *B. subtilis* using an MTP-based MIC test (See Materials and Methods section). This showed that 7-prenylisatin acted as an antibiotic with an MIC of around 25  $\mu\text{g}\cdot\text{mL}^{-1}$  against *B. subtilis*, which is similar to that of 6-prenylisatin. No



antimicrobial activity was observed against the Gram-negative *Escherichia coli* K12.

### Identification of the biosynthetic gene cluster for 7-prenylisatin

To characterize the 7-prenylisatin biosynthetic pathway, the genome of *Streptomyces* sp. MBT28 was sequenced, and the resulting draft genome sequence was annotated as described before (Girard *et al.*, 2014). The genes responsible for forming the 1*H*-indole-2,3-dione scaffold (isatin) have not previously been identified in bacteria. However, the prenyltransferase (PTase) catalysing the initial prenylation of the indole nucleus has been well characterized. Genes for indole PTases have been found in actinomycetes in two different genomic contexts (Ozaki *et al.*, 2013): Type A in combination with a tryptophanase, and Type B in combination with a flavin-dependent monooxygenase (FMO). For instance, in *Streptomyces* sp. SN-593, IptA (type A) installs a dimethylallyl group on the C-6 position of the tryptophan indole core, while in *Streptomyces coelicolor* A3(2), a Type B PTase prenylates tryptophan at C-5 (Takahashi *et al.*, 2010). Interestingly, the *Streptomyces* sp. MBT28 genome contained both types of gene clusters. However, considering that a tryptophanase is required for cleavage of the side chain (Figure 3A), the type A cluster was the most likely candidate for the bioassembly of 7-prenylisatin.



**Figure 3. Proposed biosynthetic pathway of 7-prenylisatin (A) and its gene cluster (B).**

To validate that 7-prenylisatin is indeed synthesized by a Type A indole PTase, we applied the *natural product proteomining*. This research pipeline makes use of the fact that the expression of biosynthetic proteins correlates closely with the bioactivity of interest. Then quantitative proteomics might possibly show the correlated protein levels and are therefore correlated to changes in the production of the bioactive compound (Gubbens *et al.*, 2014). For this, protein extracts of *Streptomyces* sp. MBT28 and its mutants *Streptomyces* sp. MBT28-30 and

*Streptomyces* sp. MBT28-91 were obtained from cultures grown on MM agar plates, followed by stable isotope dimethyl labelling, and subsequent LC-MS/MS analysis. Relative quantification of protein levels between the three strains were obtained from the result. In total, 1,382 proteins were identified, and the expression levels of 1,038 proteins could be obtained. Based on the intensity-dependent significance-B value (Cox & Mann, 2008), 177 of the quantified proteins were significantly enhanced or reduced ( $p < 0.05$ ) in at least one comparison (Table S1). Among the proteins that were significantly enhanced in strain *Streptomyces* sp. MBT28-91 compare to the wild-type strain, we found the two key enzymes of the Type A indole PTase gene cluster, namely the aromatic prenyltransferase (IsaA) and the tryptophanase (IsaB) (Table 1 and Figure 3B), were all significantly upregulated in *Streptomyces* sp. MBT28-91. While in *Streptomyces* sp. MBT28-30, the two enzymes were upregulated only in a small scale. The change of IsaB is still significant but the fold change is in a much lower scale compared to *Streptomyces* sp. MBT28-91. As *Streptomyces* sp. MBT28-91 has much higher 7-prenylisatin production than *Streptomyces* sp. MBT28-30, and the production of the later strain is slightly higher than wild type, a strong correlation was found between the amount of 7-prenylisatin production and the expression level of the Type A indole PTase gene cluster. In contrast, none of the proteins encoded by the Type B cluster were detected, which indicates a possibility that this cluster was not expressed under these conditions. Although the non-detection of Type B indole PTase gene cluster in proteomics is not a definitive evidence of zero presence, but taken together the bioinformatics analysis, the biosynthetic pathway and the strong correlation on Type A cluster with 7-prenylisatin production, it is suggested that the Type A gene cluster is responsible for the biosynthesis of 7-prenylisatin.

Table 1. Protein level differences (in fold change mutant/wt) of the Type A indole PTase gene cluster as determined by quantitative proteomics.

Protein	MBT28-91/WT (2log)	Significance-B	MBT28-30/WT (2log)	Significance-B	Description*
Isa Regulator	ND <sup>†</sup>	ND	ND	ND	Transcriptional regulator
IsaA	3.0	0.023	0.9	0.116	Aromatic prenyltransferase
IsaB	2.8	0.031	1.6	0.040	Tryptophanase

\* based on BLAST homology searches

<sup>†</sup> Not detected

Thus, a biosynthetic route for 7-prenylisatin was proposed based on the existing literature for related isatin compounds (Figure 3A). Three connected enzymatic reactions are required for the biosynthesis of 7-prenylisatin from tryptophan, namely a prenylation followed by carbon bond cleavage and an oxidation reaction. It is yet unclear which gene(s) are responsible for the oxidation of indole into isatin

backbone, as no such genes were found within or in close proximity of the *isa* gene cluster (Table S1). Three proteins with an oxidation-related PQQ domain were all found significantly up-regulated (with  $p$ -value  $< 0.05$ ) in the overproducing strain as compared to the parental strain. Further analysis is required to establish whether one of these enzymes may execute the oxidation reaction to generate the final isatin moiety.

### Identification and activation of the cryptic type II PKS gene cluster (*qin*) for glycosylated pyranonaphthoquinones

*Streptomyces roseifaciens* (van der Aart *et al.*, 2019), previously named *Streptomyces* sp. MBT76, which originates from the Qinling mountains in China, was identified as a prolific producer of antibiotics, including those with efficacy against multiple Gram-positive and Gram-negative multi-drug resistant pathogens (Zhu *et al.*, 2014b). Further detailed metabolic characterization of the strain identified many natural products often with interesting chemistry, including isocoumarins, prodiginines, acetyltryptamine, and fervenulin, among others (Wu *et al.*, 2016c). To unravel the biosynthetic potentials of *S. roseifaciens*, it was subjected to Illumina/Solexa whole genomic sequencing, and the genome was assembled in 13 contigs, with a total genome size of 8.64 Mb. In total, 7974 coding sequences (CDS) were predicted using the GeneMark algorithm (Lukashin & Borodovsky, 1998). Analysis of the contigs by antiSMASH (Blin *et al.*, 2013) presented a possible 55 putative biosynthetic gene clusters specifying secondary metabolites, 22 of which encoding polyketide synthases (PKS).

One 41 kb biosynthetic gene cluster for a type II PKS, designated *qin* was of particular interest considering its potential to specify pyranonaphthoquinones. These molecules represent a well-studied family of aromatic polyketides with highly complex chemical architecture and pronounced bioactivities (Oja *et al.*, 2015, Metsä-Ketelä *et al.*, 2013), including the representative members actinorhodin (Okamoto *et al.*, 2009), medermycin (Ichinose *et al.*, 2003), and granaticin (Ichinose *et al.*, 1998). In the *qin* gene cluster, besides the central PKS genes that are responsible for the biosynthesis of the pyranonaphthoquinone backbone, the presence of genes for the deoxyaminosugar D-forosamine strongly suggested that the end product should be glycosylated. This genetic organization was similar to the clusters for the synthesis of the glycosylated pyranonaphthoquinone medermycin in *Streptomyces* sp. AM-7161 (*med*) (Ichinose *et al.*, 2003) and granaticin in *Streptomyces violaceoruber* Tü22 (*gra*) (Ichinose *et al.*, 1998). Coclustering of glycosylation-associated biosynthetic genes with those for the aglycones is typical of microbial genomes, which facilitates matching the biosynthetic gene cluster to the corresponding NPs (Kersten *et al.*, 2013). In comparison, *qin*-ORF29 encoding an NADPH-dependent FMN reductase was absent in either the *med* or *gra* biosynthetic gene clusters, and glycosylation with a D-forosamine is unprecedented in the pyranonaphthoquinone

family. This promoted an investigation into the potentially novel product(s) of the *qin* gene cluster. Despite our previous detailed chemical investigations of *S. roseifaciens* (Wu *et al.*, 2016c, Zhu *et al.*, 2014b), the corresponding molecules had not been identified, suggesting that the gene cluster may be cryptic under the many different growth conditions that had been tested.

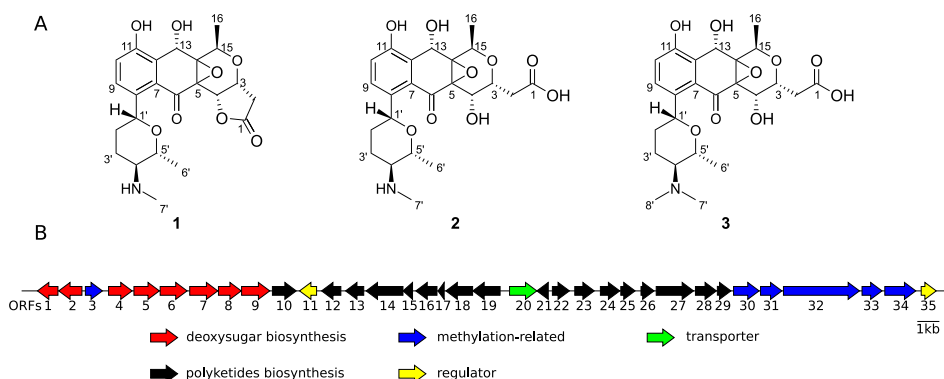
The *qin* gene cluster contains two putative regulatory genes, namely the SARP-family transcriptional regulatory gene *qin*-ORF11, and the TetR-family regulatory gene *qin*-ORF35. Many pathway-specific activators for secondary metabolite production in streptomycetes belong to the SARP family (Bibb, 2005), while TetR-family regulators often act as repressors (Ramos *et al.*, 2005). As an example, in the type II PKS gene cluster *aur1* of *Streptomyces aureofaciens* CCM 3239, which specifies the angucycline-family auricin (Novakova *et al.*, 2005), *aur1P* and *aur1PR3* encode SARP-family activators (Novakova *et al.*, 2011), while the *tetR*-type *aur1R* encodes a negative regulator (Novakova *et al.*, 2010). To activate the *qin* gene cluster, we opted to over-express the likely pathway-specific activator gene *qin*-ORF11. For this, the gene was amplified by PCR and inserted behind the *ermE* promoter (Bierman *et al.*, 1992) in the conjugative and integrative vector pSET152. The resulting plasmid was then introduced to *S. roseifaciens* via conjugation and integrated into the chromosome to create the recombinant derivative *S. roseifaciens*-1.

### Proposed biosynthetic pathway for qinimycins based on bioinformatics analysis

Ex-conjugant *S. roseifaciens*-1 and its parent *S. roseifaciens* were cultivated in parallel in modified NMMP liquid media (Wu *et al.*, 2016c). Extracts of the fermentation were analysed by <sup>1</sup>H NMR profiling (Kim *et al.*, 2010). Some metabolites usually produced by the wild-type strain, such as 1*H*-pyrrole-2-carboxamide, acetyltryptamine, fervenulin and 2-hydroxy-3-methoxy-benzamide (Wu *et al.*, 2016c), was aborted in *S. roseifaciens*-1. While the production of 6,8-dihydroxy-3-methyl-isocoumarin was not affected by ORF11 over-expression. Conversely, the <sup>1</sup>H NMR spectra of *S. roseifaciens*-1 indicates the presence of an α,β-unsaturated ketone aryl moiety typical of pyranonaphthoquinones (Wu *et al.*, 2017). These evidences indicated that indeed the *qin* gene cluster was actively expressed in *S. roseifaciens*-1. In light of the substantial change in the metabolome of *S. roseifaciens* and *S. roseifaciens*-1, we hypothesized that *qin*-ORF11 may control multiple biosynthetic gene clusters for specialized metabolites (Wu *et al.*, 2017).

To identify the metabolic product(s) of the *qin* gene cluster, the crude extract of *S. roseifaciens*-1 was separated by semi-preparative HPLC-UV chromatography. In the <sup>1</sup>H NMR spectroscopy and further UHPLC-UV-ToF-HRMS analysis of the resulting fractions, we confirmed the presence of unknown compounds C<sub>23</sub>H<sub>27</sub>NO<sub>8</sub>

(1), C<sub>23</sub>H<sub>29</sub>NO<sub>9</sub> (2), and C<sub>24</sub>H<sub>31</sub>NO<sub>9</sub> (3) with a relative abundance of 6:2:1 (Figure 4), which we designated as qinimycins.



**Figure 4. Secondary metabolites produced by *S. roseifaciens*-1.** (A) Organization of the qin locus in *S. roseifaciens*. (B) Genes for the minimal PKS (presented in black) are similar to those for biosynthesis of the pyranonaphthoquinone kalafungin, while the eight genes in red encode enzymes for production of deoxyaminosugar D-forosamine.

The qinimycins have intriguing chemical features, representing a new branch within the extensively studied pyranonaphthoquinone family of natural products. A detailed discussion of the biosynthetic pathway and chemical novelties of qinimycins is provided in Wu *et al.* (2017).

**Table 2. Antimicrobial activity of qinimycins.**

Compound	Inhibition zone (mm)			
	<i>Bacillus subtilis</i>	<i>Escherichia coli</i>	<i>Staphylococcus aureus</i>	<i>Pseudomonas aeruginosa</i>
Qinimycins (Fr1)	15	0	20	0
AMP	23	20	10	7
APRA	10	7	7	15
NC	0	0	0	0

For Qinimycins, 25  $\mu$ L of HPLC fragment 1 (Fr1) was spotted as a 2 mg·mL<sup>-1</sup> solution in methanol; For AMP and APRA, 5  $\mu$ L was spotted as a 1 mg·mL<sup>-1</sup> solution in milliQ water. AMP, ampicillin; APRA, apramycin; NC, negative control (methanol)

The bioactivity of qinimycins was tested using agar diffusion assays. Growth inhibition was seen against the Gram-positive bacteria *B. subtilis* 168 and *Staphylococcus aureus* CECT976, but not for Gram-negative *E. coli* JM109 or *Pseudomonas aeruginosa* PAO1, thus establishing bioactivity against Gram-positive but not Gram-negative bacteria (Table 2). However, assessment of the MIC of the mixture of molecules revealed minimal inhibition concentrations (MICs) against *B. subtilis* and *S. aureus* of 50  $\mu$ g·mL<sup>-1</sup> and 100  $\mu$ g·mL<sup>-1</sup>, respectively. These

high values are indicative of very limited bioactivity of the qinimycins under the tested conditions. A biosynthetic model for qinimycins is proposed based on the functional assignments from sequence analysis and the known metabolites that are produced, see detailed discussion in Wu *et al.* (2017).

## Conclusion

A major issue in harnessing the huge potential reservoir of natural products is the increasingly high frequency of re-discovery of known antibiotics used in routine screening (Wang *et al.*, 2008, Tamehiro *et al.*, 2003, Hu & Ochi, 2001, Hosoya *et al.*, 1998). This is known as the Replication Issue. One promising strategy is to ensure (i) fluctuations in the production level of the compound of interest, and (ii) perform metabolomics combined with multivariate data analysis to correlate bioactive molecules to the observed bioactivity, combined with (iii) correlate the changes in the transcriptome or proteome profiles to the bioactivity to identify candidate gene clusters (Wu *et al.*, 2015d). This strategy was successfully applied in elucidating bioactive compounds produced by *Streptomyces* sp. MBT28, resulted in the identification of the antibiotic 7-prenylisatin, which was a near silent antibiotic under routine growth conditions. To activate the production of 7-prenylisatin and elucidate its structure, streptomycin-resistant mutants were selected and subsequently analysed by NMR-based metabolomics (Nicholson & Lindon, 2008) to identify those NMR signals that correlated statistically to the enhanced bioactivity, and natural product proteomining (Gubbens *et al.*, 2014) to identify the corresponding biosynthetic proteins. Another efficient way of solving the Replication Issue and to streamline the discovery of novel molecules from actinomycetes is to combine genome mining, targeted genetic manipulation, and metabolomics analysis. Prior knowledge of NPs that can be expected based on bioinformatics thereby provides information on which biosynthetic cluster is more interesting in terms of discovering new natural products. This information also simplifies and reduces the time consumption of the chromatographic isolation process and structure determination of target compound(s). The identification of qinimycin compounds from *S. roseifaciens* provides proof of concept for this principle. Whereby activation of the cryptic *qin* type II PKS gene cluster by constitutive expression of its pathway-specific activator gene (*qin*-ORF11) followed by NMR-based metabolic profiling, identified novel glycosylated pyranonaphthoquinones. The elucidated biosynthetic pathway for the qinimycins in the genetically tractable *S. roseifaciens* offers new insights into the biosynthesis of this family of natural products.

## Experimental section

### Strains and culturing conditions

*Streptomyces* sp. MBT28 and *Streptomyces roseifaciens* MBT76 (NCCB 100637, DSM 106196) were obtained from Molecular Biotechnology culture collection, IBL, Leiden University. Spores were stored in 20% glycerol and maintained in -20°C. Soy flower agar medium (SFM) (Kieser *et al.*, 2000) was used to generate streptomycin-resistant (Str<sup>R</sup>) mutants, grow over-expression conjugants of *S. roseifaciens*, and for preparing spore suspensions. Standard solid minimal medium (MM) (Kieser *et al.*, 2000) supplemented with both 0.5% (w/v) mannitol and 1% (w/v) glycerol was used for activity tests and to prepare samples for metabolomics and proteomics for *Streptomyces* sp. MBT28 experiments. *Bacillus subtilis* 168 was grown in Luria-Bertani broth (LB) at 37°C. *Escherichia coli* JM109 (Sambrook *et al.*, 1989) was used for routine cloning. *E. coli* ET12567 (MacNeil *et al.*, 1992) containing pUZ8002 (Kieser *et al.*, 2000) was used for introducing non-methylated DNA into *S. roseifaciens* by conjugation. The basal medium for *S. roseifaciens* growth was modified minimal liquid medium NMMP (Kieser *et al.*, 2000) without PEG6000 and containing 1% (w/v) glycerol and 0.5% (w/v) mannitol as the carbon sources, which was further supplemented with 0.8% (w/v) Bacto peptone (Wu *et al.*, 2016c). Tryptone soy broth with 10% (w/v) sucrose (TSBS) was used to grow *S. roseifaciens* mycelia as receptor for conjugation experiments. *Streptomyces* sp. MBT28 Str<sup>R</sup> mutants were selected by plating 10 µl of spores containing 1×10<sup>5</sup> colony forming units (CFU) onto SFM agar plates containing streptomycin (10 µg·mL<sup>-1</sup>) at 30°C. Colonies typically developed after nine days. For chemical analysis of *S. roseifaciens* secondary metabolites, 50 mL modified NMMP was inoculated with 10<sup>6</sup> spores of *S. roseifaciens* in 250 mL Erlenmeyer flasks equipped with a spring. Cultures were incubated at 30°C with constant shaking at 220 rpm for 120 h.

### Genome sequencing, assembly, and annotation

DNA was extracted from *Streptomyces* sp. MBT28 and *S. roseifaciens* as described previously (Kieser *et al.*, 2000). Genome sequencing and annotation was done essentially as described previously (Girard *et al.*, 2014). 100-nt paired-end reads were obtained and the quality of the short reads verified using FastQC (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>). Depending on quality, reads were trimmed to various lengths at both ends. Processed raw reads were subsequently used as input for the Velvet assembly algorithm. The resulting contigs were analysed using the GeneMark hmm algorithm with the *Streptomyces coelicolor* genome as model for ORF finding (Lukashin & Borodovsky, 1998). The genomes were additionally annotated using the RAST server with default options. BLASTP for putative function prediction and HMMER for protein-domain prediction, manually inspected for some and visualized using Artemis. The resulting genomes have been deposited at GenBank. The GenBank WGS project

accession number of *Streptomyces* sp. MBT28 is LARV00000000, for *S. roseifaciens* the accession number is LNBE00000000.

### Antibiotic activity assays

*Streptomyces* sp. MBT28 was grown for six days on MM agar plates and overlaid with LB agar containing 100  $\mu\text{L}$  of exponentially *B. subtilis* cells ( $\text{OD}_{600}$  of 0.4). The activity was initially assessed as the zone of growth inhibition after overnight incubation at 37°C. MIC values against *B. subtilis* were carried out as described before (Zhu *et al.*, 2014b). All MIC determinations were performed in triplicate. Ampicillin was used as positive control, and the solvent chloroform as the negative control.

Antimicrobial activity of qinimycins was determined according to a disc diffusion method as described before (Wu *et al.*, 2015c, Wu *et al.*, 2016b). 25  $\mu\text{L}$  of Fr1 (2  $\text{mg}\cdot\text{mL}^{-1}$  in methanol) was spotted onto paper discs (6 mm diameter) placed on agar plates containing a soft agar overlay with indicator bacteria. Indicator bacteria were *Bacillus subtilis* 168, *E. coli* ASD19, *Staphylococcus aureus* CECT976, or *Pseudomonas aeruginosa* PAO1. Ampicillin and apramycin were used as positive controls, whereby 5  $\mu\text{L}$  was spotted of a 1  $\text{mg}\cdot\text{mL}^{-1}$  solution in miliQ water. The solvent methanol was used as the negative control. After incubation at 37°C for 18 h, growth inhibition zones (in mm) were recorded as antimicrobial activity. The MIC assay against *B. subtilis* and *S. aureus* were carried out in 96-well plate by serial double dilution method, as previously described (Zhu *et al.*, 2014b). All MIC determinations were performed in duplicate.

### Primers and plasmid construction

PCRs were performed in a minicycler (MJ Research, Massachusetts, U.S.), using Phusion polymerase (Stratagene, California, U.S.) as described before (Colson *et al.*, 2007). Sequences of all oligonucleotides used in this study was listed in Table S2. For cloning, *rpsL* was amplified from *Streptomyces* sp. MBT28 genomic DNA using oligo nucleotides *rpsL*\_For and *rpsL*\_Rev; *rsmG* was amplified using oligonucleotides *rsmG*\_For and *rsmG*\_Rev. *qin*-ORF11 was amplified by PCR from *S. roseifaciens* genomic DNA as described before (Colson *et al.*, 2007) using primers: SC06\_0044\_F\_EcoRI and SC06\_0044\_R\_EcoRI\_NdeI. The PCR product was subcloned as an EcoRI fragment and subsequently constructed as an NdeI-HindIII with constitutive *ermE* promoter fragment into pSET152 (Bierman *et al.*, 1992), which integrates at the genome  $\phi\text{C31}$  sites. The final recombinant plasmid pCSW01 (pSET152/*ermE*/*qin*-ORF11) was made and *E. coli* ET12567/pUZ8002 was transformed with it. This transformant was then used to conjugate two-day old mycelia of *S. roseifaciens*. The ex-conjugants were confirmed by PCR using primers SC06\_0044\_SF and SC06\_0044\_SR.



### Metabolomics sampling and preparation

*Streptomyces* sp. MBT28 was grown on MM agar plates supplemented with 0.5% mannitol and 1% glycerol. The agar with mycelia was cut into small pieces and extracted with ethyl acetate by soaking in the solvent overnight at room temperature. Samples were dried under reduced pressure at 40°C and re-dissolved in methanol. The solvent was then evaporated at room temperature under nitrogen gas flow. 50 mL cultures of *S. roseifaciens* or *S. roseifaciens*-1 were harvested by centrifugation at 4000 rpm for 10 min, and the supernatant extracted twice with 20 mL of ethyl acetate (EtOAc). The organic phase was washed with 30 mL of water and subsequently dried with 5 g of anhydrous Na<sub>2</sub>SO<sub>4</sub>. EtOAc was removed under vacuum at 38°C and the residue was dissolved in 2.0 mL of EtOAc. The solvent was then evaporated at room temperature under nitrogen gas. The extracts from both experiments were subsequently dipped into liquid nitrogen and lyophilized using a freeze dryer (Edwards Ltd., Crawley, England). Crude extracts were partitioned between methanol and n-hexane to remove lipids. Ten replicates were performed for *Streptomyces* sp. MBT28 samples and fifteen replicates were performed for *S. roseifaciens* samples.

### Metabolomics, NMR guided separation, compound purification

For methods on NMR-based metabolomics and compound identification, we refer to the original publications by Wu, Du et al. (2015a, 2017)).

### Quantitative proteomics analysis

To prepare samples for proteomics, strains were grown on MM agar plates overlaid with 0.1 µm polycarbonate filter papers prior to inoculation. After six days of incubation at 30 °C, mycelia were scraped from the filters. Protein extraction, in-solution digestion, C18 column dimethyl labelling, strong cation-exchange chromatography (SCX) fractionation, and LC-MS/MS analysis were performed as previously described (Gubbens *et al.*, 2014, Li *et al.*, 2013). Generally, mycelia were washed and then sonicated for disruption of cell wall. After removal of cell debris, protein concentration was measured using a Bradford assay, and 0.167 mg of protein was precipitated for each sample. Proteins were dissolved using RapiGest SF Surfactant (Waters, Massachusetts, U.S.) and treated with iodoacetamide before trypsin digestion. Peptides were dimethyl labelled on C-18 resin and the three samples mixed to obtain 0.5 mg of peptides for fractionation by SCX. 24 fractions were collected and subjected to LC-MS/MS analysis on an LTQ-orbitrap MS (Thermo, Pennsylvania, U.S.) (Li *et al.*, 2013). Data analysis was performed using MaxQuant 1.4.1.2 (Cox & Mann, 2008). Dimethyl labelling was selected as quantification method, carbamidomethylation of cysteine was selected as fixed modification, and oxidation of methionine was set as variable modification (Gubbens *et al.*, 2012). MS/MS spectra were searched against a database of translated coding sequences. Peptide and protein identification FDR were set to 0.01, and a minimum of two quantification events was specified to obtain a protein

quantification. To determine the significance of the found expression ratios, intensity-dependent B-significance values (Cox & Mann, 2008) were calculated.

## **Acknowledgments**

We are grateful to H. Zhu for discussions. The work was supported by a grant from the Chinese Scholarship Council to C.S.W. and by VENI grant 11890 from The Netherlands Technology Foundation STW to J.G.



## Chapter 4

---

# Unravelling the response of *Streptomyces roseifaciens* to challenge with small molecules by genome-wide proteomics

Chao Du, Somayah Elsayed, Marta del Rio, Bogdan I. Florea, Jos Raaijmakers, Gilles P. van Wezel

Manuscript in preparation

## Abstract

Bacteria of the order of the Actinomycetales are the most prominent antibiotic producers. Of these, members of the genus *Streptomyces* produce half of all antibiotics used in the clinic today. To unlock the full potential of these bacteria in terms of natural product biosynthesis, eliciting strategies need to be developed that mimic the environmental conditions, to activate silent biosynthetic gene clusters (BGCs). Here, we analysed the response of the gifted antibiotic producer *Streptomyces roseifaciens* to challenge with small molecules, with the aim to find elicitors of the expression of cryptic BGCs. The bioactivity changes were correlated to the proteome level shifts using quantitative proteomics. Hydroxycoumarin was identified as an efficient elicitor, which may be explained by changes in energy metabolism. Surprisingly, despite their major differences on antimicrobial activity, only minor effects were seen in protein profiles when *S. roseifaciens* was exposed to jasmonic acid, chitosan, benzoic acid, or N-acetylglucosamine. A time series proteomics study on the immediate response of *S. roseifaciens* to jasmonic acid (JA) identified a BGC designated *jar* and a gene encodes an MFS transporter as major determinants for the JA response. The *jar* cluster directly correlated to the antimicrobial activity of *S. roseifaciens*, as shown by mutational analysis.

## Introduction

Streptomycetes are soil-dwelling microorganisms with a complex multicellular life cycle (Claessen *et al.*, 2014). These bacteria are a major source of clinically approved drugs, such as antibiotics and anticancer compounds (Hopwood, 2007, Newman & Cragg, 2016, Barka *et al.*, 2016). As antibiotics are being widely used across the globe, the increasing antibiotic-resistant bacteria are becoming a severe problem. The failure to identify novel bioactive molecules is thereby a huge bottleneck in the battle against resistant microbes. Nowadays, the brute force antibiotic discovery gradually comes to an end as the easy to find compounds have been mostly discovered and synthetic antibiotics fail to meet the needs (Kolter & van Wezel, 2016). Advances in genomics showed that the potential of Actinobacteria as producers of bioactive natural products may have been grossly underestimated (Hopwood, 2007). Scientists now have the task to unleash the full producing capacity of these bacteria.

In bacteria, the biosynthetic genes responsible for the production of specific NPs are typically clustered, thus forming so-called biosynthetic gene clusters (BGCs). Bioinformatics algorithms based on the current understanding of BGCs often gives about 20 to 50 potential BGCs of different kinds from a single *Streptomyces* genome, many of which are silent or cryptic under routine laboratory conditions. One way to activate such silent BGCs is genetic manipulation of the regulatory network that controls the BGC, for example by over-expression of pathway-specific activators (Sanchez *et al.*, 2010, van der Heul *et al.*, 2018). Another method that is often used when prior information is not available, is to change the external and/or internal biological factors by fluctuating the growth environment (Blin *et al.*, 2017, Rutledge & Challis, 2015, Zhu *et al.*, 2014a). These so-called eliciting strategies aim to mimic the environmental conditions in the laboratory, where the activating factors of antibiotic production is often missing (van der Meij *et al.*, 2017).

There is a strong correlation between the production level of a certain NP and the expression of its cognate BGC (Du & van Wezel, 2018). Therefore, once fluctuation of NP production has been achieved, samples from different conditions can be compared using metabolomics. Quantitative proteomics or transcriptomics allow correlation of a specific BGC to its cognate bioactive compound (Gubbens *et al.*, 2012, Wu *et al.*, 2016a). This allows efficient dereplication of known molecules and gene clusters. Proteomics, as a very important method that allows to access the cellular machinery that directly participates in most biological process, has also evolved to have better coverage and accuracy during the years, and is becoming a very attractive tool in NP research (Du & van Wezel, 2018, Schubert *et al.*, 2017).

Since a large part of antibiotics is produced by soil dwelling bacteria of the genus *Streptomyces*, an important pathway of activating the cryptic antibiotic resources is to investigate the interactions between plant and microbe. We know that the ability of producing compounds like jasmonic acid, salicylic acid, and auxin is important for the plants to modulate their surrounding microbiome. This modulation is very important for pathogen resistance of the plants (van der Meij *et al.*, 2017, Xiao *et al.*, 2002, Gopalakrishnan *et al.*, 2011, Carvalhais *et al.*, 2015, Lebeis *et al.*, 2015). Recently, researchers proposed the *cry for help hypothesis*, which entails that plants recruit beneficial Actinobacteria and activate their antibiotic production by releasing stress-related molecules as exudates upon challenge by pests (van der Meij *et al.*, 2017). Recent studies in our laboratory have shown that *Streptomyces roseifaciens* responds to the plant hormone jasmonic acid by enhancing antibiotic production. *S. roseifaciens* is a very gifted strain in terms of NP biosynthesis, which originates from the Qinling mountains in China (van der Aart *et al.*, 2019). The *S. roseifaciens* genome contains 55 putative BGCs for small molecules as predicted by antiSMASH (Blin *et al.*, 2013). Indeed, the strain produces a wide range of NPs, including antibiotics with efficacy against multiple Gram-positive and Gram-negative multi-drug resistant pathogens (Wu *et al.*, 2016c, Zhu *et al.*, 2014b) and other NPs such as isocoumarins, prodiginines, acetyltryptamine, fervenulin and pyranonaphthoquinones (Wu *et al.*, 2016c, Wu *et al.*, 2017).

In this chapter, the response of *S. roseifaciens* to a range of small molecules are shown. A set of 12 small molecules was selected based on their different properties, including plant hormones, oligosaccharides and pathogen-related compounds. Quantitative proteomics was performed to obtain insights into the changes in the protein profiles, and correlate these to metabolic information of the NPs produced. The response to the 12 compounds could be clustered in three main groups, of which hydroxycoumarin stood out as a single group. The response to jasmonic acid (JA) was worked out more, and a 4 h time-course experiment following induction by JA revealed a JA-repressed gene cluster (*jar* cluster) and a JA-activated MFS-type transporter.

## Results and Discussion

### Small molecules selected for proteomics study of the response of *S. roseifaciens*

In order to explore the potential of *S. roseifaciens* as a natural product producer, a set of 12 different small molecules was selected to analyse their potential as elicitors of cryptic BGCs (Table 1). One major group of small molecules tested were plant hormone compounds, including jasmonic acid (JA), salicylic acid (SA), indol-3-acetic acid (IAA) and its synthetic analogue naphthaleneacetic acid (NAA). Of these, JA elicits the production of antibiotics in *S. roseifaciens* (van der Meij et al. unpublished data). The plant hormones IAA and NAA are auxins that induce plant cell elongation and cell division (Hopkins, 1999). Some *Streptomyces* species are also able to produce auxins, which might have a plant growth-promoting effect (Sadeghi et al., 2012, Viaene et al., 2016). A second group of small molecules consisted of hydroxycoumarin (HyC), N-acetylglucosamine (GlcNAc), cellobiose (CB), chitosan (CS), cinnamic acid (CA), ferulic acid (FA), benzoic acid (BA), and butyrate (BR). These compounds are related with cell components of other organisms or metabolites of those components. Of these, GlcNAc was previously shown to activate silent BGCs in *Streptomyces*, via the inactivation of the global antibiotic repressor DasR (Craig et al., 2012, Rigali et al., 2008, van Wezel et al., 2009). CS, which consists of both GlcNAc and GlcN units, is an incomplete deacetylation product of chitin. It can induce defence-related responses in plants (Hadwiger, 2013). CB is a hydrolysate of the highly abundant natural polymer cellulose, and can be efficiently used by *Streptomyces* (Colson et al., 2008, Liao et al., 2014). CA is a precursor of a large number of plant substances. FA and BA are derivatives of CA. All three molecules are metabolized by *Streptomyces* species (Sutherland et al., 1981, Sutherland et al., 1983).

### Antimicrobial activity can be both induced and suppressed by the selected small molecules

To assess the actual effects of these selected small molecules, antimicrobial tests were done. *S. roseifaciens* spores were grown as spots on MM with the different compounds present in the medium, followed by overlay with *E. coli* strain ASD19 (Figure 1). This revealed major differences, whereby GlcNAc, CB, BR and CS decreased the total antibiotic activity of *S. roseifaciens*, while most other compounds elicited the bioactivity, except SA which had no noticeable effect. Interestingly, the compounds with the most distinct effect, namely CS, GlcNAc, JA and BA, clustered together in the proteomic response. As discussed later in this Chapter, in fact we did not see a clear correlation between the full proteome-level response and the antimicrobial activities.

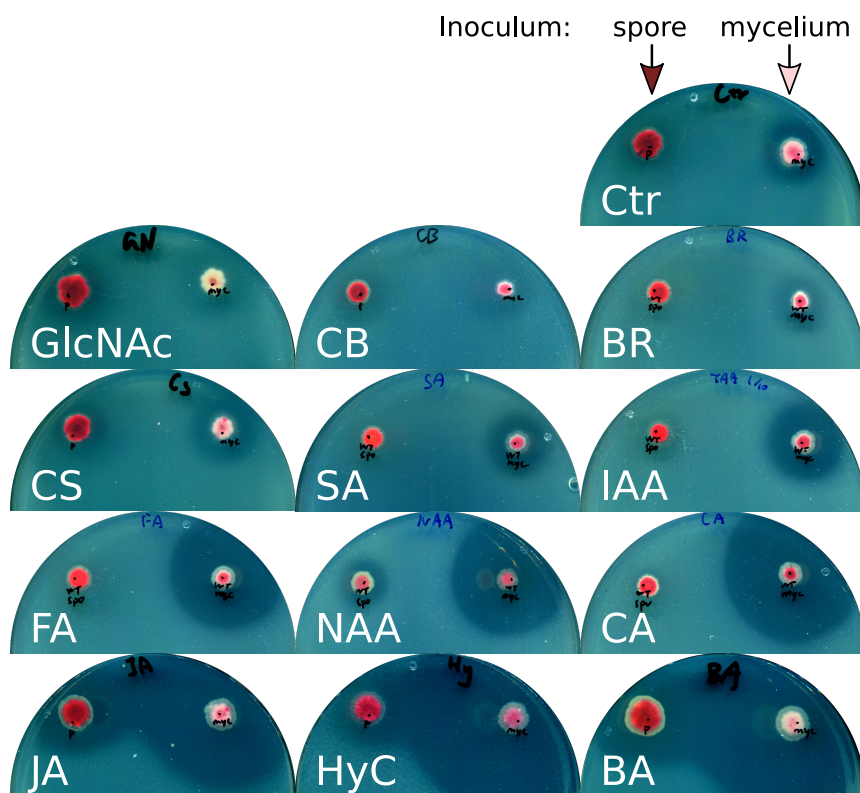


Table 1. The small molecules used as putative elicitors in this study.

Abbreviation	Compound name	Notes
JA	Jasmonic acid	Plant hormone of jasmonate class, involved in the stress response and known as transducers of plant secondary metabolite production (Farmer et al., 2003).
SA	Salicylic acid	Plant hormone involved in response to stress and development (Jin et al., 2017, Ahmad et al., 2019).
IAA	Indole-3-acetic acid	Plant hormone of auxin class, can induce plant cell elongation and cell division (Hopkins, 1999).
NAA	Naphthalene acetic acid	Synthetic plant hormone in the auxin family (Hopkins, 1999).
HyC	Hydroxycoumarin	Fungal metabolite relating to plant cell damage (Bocks, 1967, Brown, 1986, Liu et al., 2009), mimicking fungal presence.
GlcNAc	N-acetylglucosamine	Involved in nutrient signalling of <i>Streptomyces</i> , related to regulation of antibiotics production (Rigali et al., 2008).
CB	Cellobiose	Disaccharide from the hydrolysis of cellulose, induces expression of cellulases in many fungi (Ilmén et al., 1997).
CS	Chitosan	Chitin deacetylation product.
CA	Cinnamic acid	Precursor for the synthesis of a huge number of plant substances, has antimicrobial activity (Ferreira & Teixeira, 2003).
FA	Ferulic acid	Product of cinnamic acid metabolism, abundant in plant cell wall, can be released by wounding. Induces <i>vir</i> genes of <i>Agrobacterium tumefaciens</i> (Heath et al., 1995, Kalogeraki et al., 1999).
BA	Benzoic acid	Product of cinnamic acid metabolism, rich in many crops, food preservative (del Olmo et al., 2017).
BR	Butyrate	End product of bacterial anaerobic fermentation, also a building block of NPs produced by <i>Streptomyces</i> (Omura et al., 1983, Reynolds et al., 1988).

### Proteome level responses to small molecules

To obtain biomass for proteomics experiments, *S. roseifaciens* was pre-cultured on MM agar plates covered with a cellophane disc for 48 h at 30°C. After this, the cellophane disk with mycelial biomass was moved to a fresh MM agar plate containing one of the small molecules at a concentration of 500  $\mu$ M or same amount of solvent as control. Then the mycelia were grown for another 48 h and harvested for protein extraction and proteomic measurement. Triplicate experiments were performed for each of the 13 growth conditions (12 small molecules + control), giving a total of 39 samples.

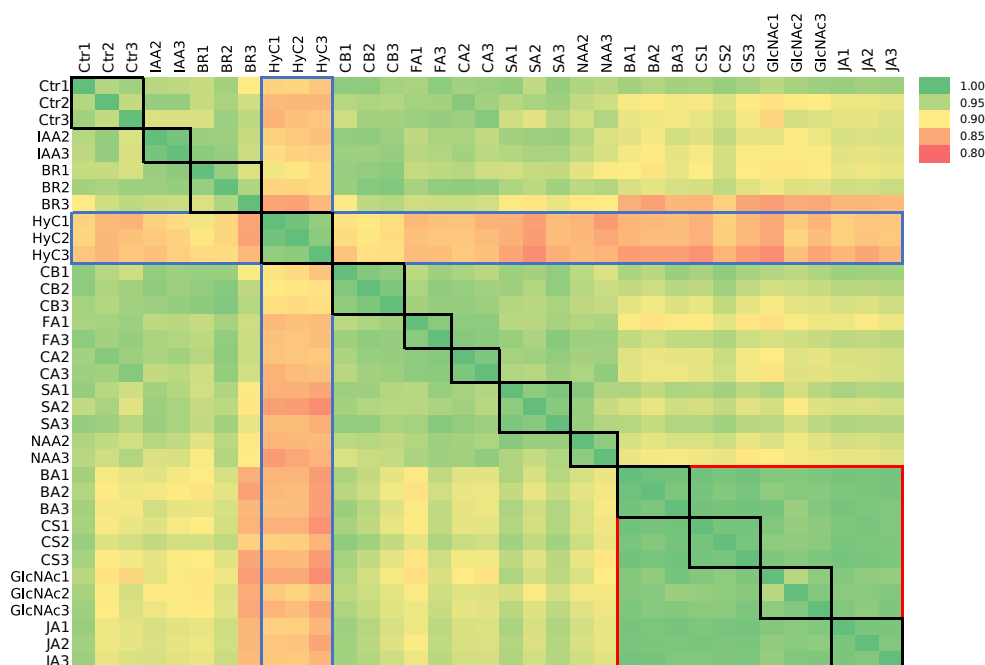


**Figure 1. Effect of small molecules on the antimicrobial activity of *S. roseifaciens*.**

Spots of spore (left) and mycelium (right) were grown on MM agar supplemented with 1% glycerol and 0.5% mannitol. Small molecules were added to 500  $\mu$ M in the media before inoculation. *E. coli* ASD19 was used as indicator strain, and overlay was done on the fifth day after inoculation. Pictures were taken 16 h after the overlay. Ctr for control.

Pearson's correlation coefficients were calculated to show the differential protein levels between different samples, while at the same time allowing general assessment of the quality of the data. A cut-off score of 0.95 was maintained for the correlation within triplicates. As a result, IAA replicate #1, FA replicate #2, CA replicate #1 and NAA replicate #1 were removed from the data for their low correlation with the other two replicates. The data normalization and quantification were done again using the rest of samples, ensuring best data quality. Subsequently, single quantifications among replicates were removed. In total, 1,897 proteins could be quantified in at least one condition. The number of quantified proteins in each sample ranged from 1,507 to 1,739, with an average of 1,638. Of the quantified proteins, 1,242 were shared by all 13 conditions (two or more identifications among replicates). The matrix of correlation coefficients revealed that hydroxycoumarin (HyC) samples had the most distinct proteome

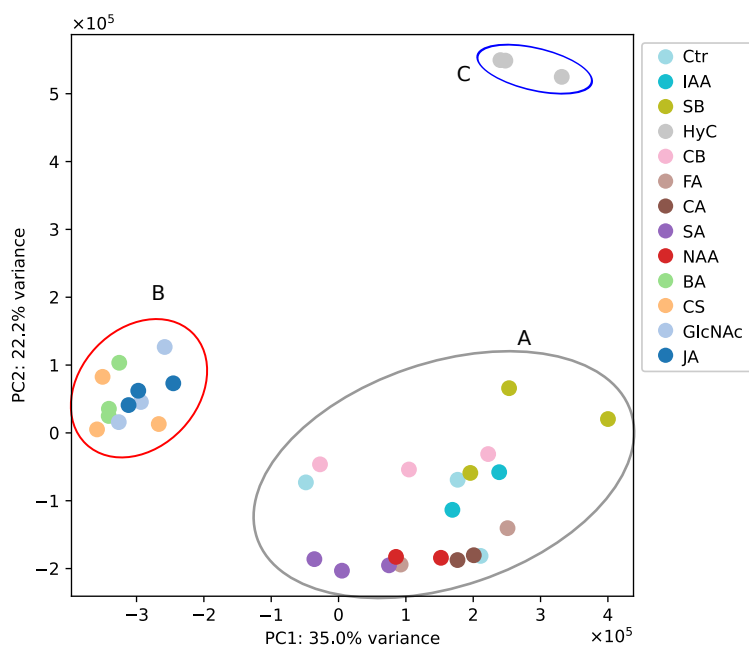
profile as compared to samples obtained from growth on other small molecules (Figure 2 blue box), with a correlation coefficient  $< 0.9$  for the comparison. The correlation matrix also showed that protein profiles were very similar for samples obtained from MM with either JA, GlcNAc, CS, or BA, with a correlation coefficient  $> 0.95$  (Figure 2 red box). Principal component analysis (PCA) was done, and three clusters were found amongst all conditions (Figure 3). Samples treated with auxins (IAA, NAA), SA, CA and its derivative FA, CB, or BR clustered together with the control, which indicates that these compounds did not trigger a global change in gene expression in *S. roseifaciens*. As suggested by the high correlation, samples obtained from mycelia grown on JA, GlcNAc, chitosan or BA formed another closely related cluster, well separated from the control and related samples. Samples obtained from MM with hydroxycoumarin (HyC) were remote from these two large clusters. This indicates that HyC elicits a distinct systematic response as compared to the other compounds and the control.



**Figure 2. Correlations between protein profiles in response to small molecules.**

Replicates are indicated using black borders (three replicates analysed, samples that failed to comply with statistical threshold not shown). Hydroxycoumarin-treated samples are indicated with a blue box. The four compounds giving highly similar responses are highlighted with a red box. Ctr for control.

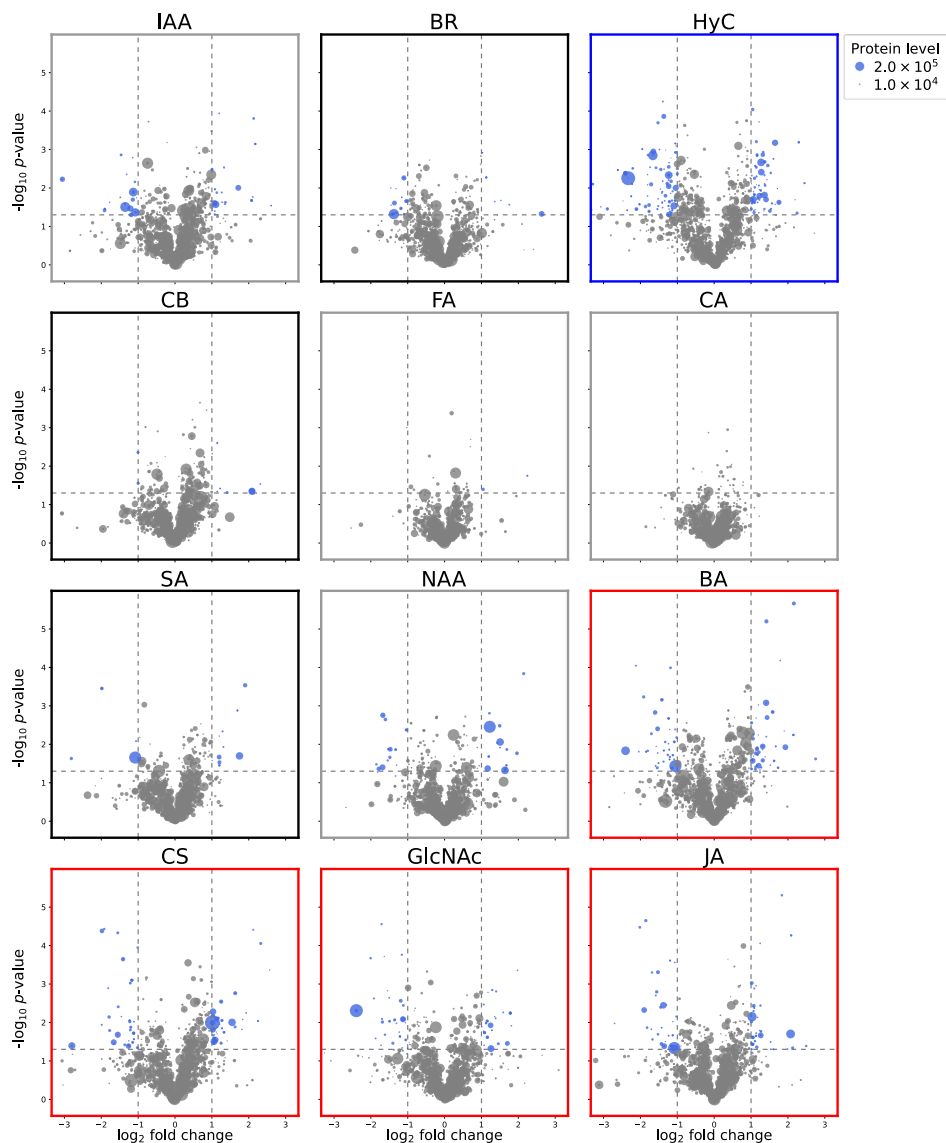
However, the response to the small molecules in terms of the clustering at the proteome level did not correlate well with the antimicrobial activity. In fact, GlcNAc and CS were among the four compounds that suppressed the antimicrobial activity of *S. roseifaciens*, while JA and BA had the most eliciting effect. This indicates that the antimicrobial activity induced in JA and BA did not relate to the general changes at the proteomics level. Instead, this may be due to the changes in a few key enzymes, small differences in the growth conditions between experiments, or the effects of elicitors on growth.



**Figure 3. Principle component analysis (PCA) of the proteomics data.** Each dot represents one single sample; treatments of different compounds are colour coded. Circles (A, B, C) are drawn indicating clustered samples that are similar to each other. Ctr for control.

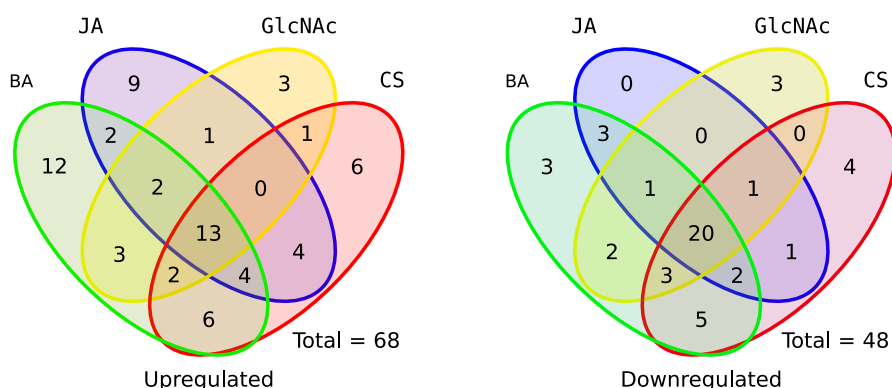
### Differentially expressed proteins (DEPs) induced by small molecules

Proteins that differed at least two-fold in abundance with the control and a  $p$ -value  $< 0.05$  in T-test were considered as differentially expressed proteins (DEPs). The DEPs are shown in the volcano plots as blue circles with sizes varying depending on their level (Figure 4). All DEP numbers from different treatments are listed in Table 2 and in Table S1. Consistent with the correlation analysis and PCA results, HyC samples had the most up- and down-regulated proteins, followed by PCA cluster B (BA, CS, JA, GlcNAc), while those for PCA cluster A showed only few significant changes to the control.



**Figure 4. Volcano plots of DEPs following addition of different compounds.** The significantly changed (fold change  $\geq 2$ ,  $p$ -value  $< 0.05$ ) proteins are indicated by blue dots, other proteins are shown as grey dots. The size of each dot represents the higher average protein level of control and test condition. HyC treatment is indicated by blue bordered plot. Red bordered plots indicate the four conditions that clustered closely (Figure 3). Grey bordered plots (IAA, NAA, FA and CA) are the treatments with one sample removed from the analysis, which led to higher  $p$ -values (lower  $-\log_{10} p$ -value). The blue border plot corresponds to HyC samples, red border plots to the four clustered molecules (BA, CS, JA, and GlcNAc), as shown in Figure 3.

In order to get an overview of the function of DEPs, gene ontology (GO) enrichment analysis was performed for the DEPs of all treatments. For each treatment, three sets of proteins were tested for enrichment the GO aspect “biological process”, they were up-regulated, down-regulated and combined DEPs. All over-represented GO classes which were the lowest in the GO hierarchy were listed in Table 3. For the HyC treated samples, proteins related to “ATP synthesis coupled proton transport” were significantly overrepresented (corrected  $p$ -value < 0.05) in the extracted DEPs, whereby three of the five proteins belonging to this category were significantly up-regulated. This suggests that energy metabolism was affected by HyC, which may explain the fundamental changes in the overall proteomic shift shown in the correlation and PCA graph (Figure 2, Figure 3). Alpha-L-glutamate ligase SC03\_2012 was among the most up-regulated DEPs following HyC treatment ( $\log_2$  fold change = 2.45). This is a RimK-family ATP hydrolase that likely plays a role in the biosynthesis of small molecules (Zhao *et al.*, 2013). The two paralogues of pyruvate dehydrogenase E1 component (SC01\_0222 and SC04\_0475) were down-regulated when grown on HyC, with  $\log_2$  fold changes of -3.29 and -0.76, respectively. The down-regulation of the key metabolic enzyme pyruvate dehydrogenase suggests changes in central metabolism.



**Figure 5. Venn diagram showing DEPs shared between four conditions that induced similar responses in *S. roseifaciens*.** About 20% and 40% proteins were commonly up- and down-regulated, respectively, indicating a larger heterogeneity in the up-regulated proteins than down-regulated proteins.

Within the cluster of samples treated with BA, JA, CS, and GlcNAc, 13 proteins were up-regulated, and 20 proteins down-regulated under all conditions (Figure 5, Table 4). The most highly up-regulated DEPs shared between the four conditions include 4-aminobutyrate aminotransferase (SC04\_2868; fold change 1.81-2.33), transaconitate 2-methyltransferase (SC04\_3021; fold change 1.59-2.12), and glucosamine-fructose-6-phosphate aminotransferase (GlmS) (SC04\_0181; fold

change 1.54-1.78). 4-aminobutyrate aminotransferase catalyses transamination between primary amines and  $\alpha$ -keto acids; trans-aconitate 2-methyltransferase metabolises trans-aconitate, a toxic metabolite derived from the TCA cycle (Cai *et al.*, 2001), GlmS catalyses the conversion of fructose-6-phosphate to glucosamine-6-phosphate (GlcN-6P). GlcN-6P is a central metabolite that stands at the crossroads of aminosugar metabolism, glycolysis, nitrogen metabolism and cell-wall synthesis in *Streptomyces* (Urem *et al.*, 2016). For GlcNAc and Chitosan treated samples, increased GlcN-6P levels were expected, but this does not explain the up-regulation of GlmS. Among the most strongly down-regulated proteins common to all four conditions, there were two MarR-type regulatory proteins, SC04\_2720 (fold change -3.72 to -2.70) and SC05\_0108 (foldchange -1.22 to -1.90). MarR family regulators typically control stress-related regulons, including resistance to multiple antibiotics, organic solvents and oxidative stress agents (Alekhshun *et al.*, 2001, Alekhshun & Levy, 1999). Although the effects of the clustered small molecules were not the same, the proteome level similarities still provide valuable information for further studies.

Table 2. Number of DEPs in samples treated with different small molecules.

Small molecules	Abbreviation	Up-regulated	Down-regulated	Total DEPs
Hydroxycoumaric acid	HyC	67	92	159
Benzoic acid	BA	44	39	83
Chitosan	CS	36	36	72
Jasmonic acid	JA	35	28	63
N-acetylglucosamine	GlcNAc	25	30	55
Indole-3-acetic acid	IAA	23	18	41
Naphthaleneacetic acid	NAA	18	15	33
Butyrate	BR	8	15	23
Salicylic acid	SA	12	4	16
Cellobiose	CB	7	2	9
Ferulic acid	FA	2	1	3
Cinnamic acid	CA	0	0	0

Table 3. Overrepresented gene ontology (GO) biological process terms in hydroxycoumarin treated samples.

DEP sets	GO ID	Class name	p-value	Test set	Reference set	Proteins in test set
Up	0015986	ATP synthesis coupled proton transport	$6.48 \times 10^{-3}$	3/31	5/891	SC04_0848 SC04_2604 SC04_2607
All	0015986	ATP synthesis coupled proton transport	$3.79 \times 10^{-2}$	3/61	5/891	SC04_0848 SC04_2604 SC04_2607

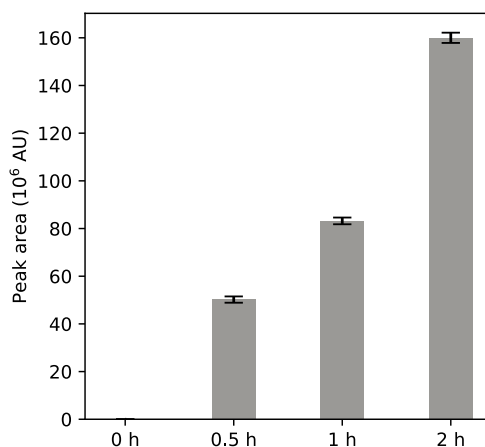
Table 4. Common DEPs for the PCA cluster B treatments

ID	Log <sub>2</sub> fold change					Annotation
	BA	CS	GlcNAc	JA	Mean	
sc04_2868	2.16	2.33	1.81	2.13	2.11	4-aminobutyrate aminotransferase
sc04_3021	2.12	2.12	1.59	2.07	1.98	Trans-aconitate 2-methyltransferase
sc04_2231	1.79	1.84	1.98	1.61	1.8	Phosphoenolpyruvate carboxykinase [GTP]
sc04_3260	1.93	1.82	1.7	1.65	1.77	50S ribosomal protein L16
sc04_0181	1.67	1.63	1.78	1.54	1.66	L-glutamine-D-fructose-6-phosphate amidotransferase
sc03_0574	1.41	1.33	1.26	1.26	1.31	MMPL domain protein
sc04_2879	1.44	1.17	1.29	1.07	1.24	Proline-tRNA ligase
sc02_0728	1.31	1.27	1.28	1.11	1.24	Alkaline d-peptidase
sc04_2093	1.21	1.1	1.36	1.26	1.23	Uncharacterized protein
sc07_0017	1.2	1.14	1.18	1.25	1.19	Delta-aminolaevulinic acid dehydratase
sc04_0048	1.16	1.05	1.16	1.01	1.1	DNA-binding protein HU 1
sc05_0086	1.04	1.12	1.07	1.14	1.09	Zinc protease
sc03_2044	1.17	1.06	1.08	1.05	1.09	Uncharacterized protein
sc04_1365	-1.24	-1.11	-1.19	-1.03	-1.14	Soj family protein
sc04_0799	-1.18	-1.17	-1.16	-1.09	-1.15	DUF88 domain-containing protein
sc04_0177	-1.2	-1.11	-1.21	-1.12	-1.16	Oligoribonuclease
sc04_1335	-1.33	-1.21	-1.3	-1.12	-1.24	Carbamoyl-phosphate synthase large chain
sc06_0096	-1.14	-1.36	-1.15	-1.43	-1.27	Non-specific serine/threonine protein kinase
sc03_0077	-1.6	-1.41	-1.13	-1.2	-1.33	Putative serine/threonine protein kinase
sc04_2018	-1.53	-1.25	-1.4	-1.33	-1.38	Enoyl-CoA hydratase/isomerase
sc05_0108	-1.22	-1.23	-1.23	-1.9	-1.39	MarR-family transcriptional regulator
sc07_0025	-1.57	-1.55	-1.24	-1.38	-1.44	Putative oxidoreductase
sc06_0206	-1.7	-1.45	-1.5	-1.45	-1.53	ATP-dependent chaperone ClpB
sc04_2325	-1.53	-1.54	-1.55	-1.58	-1.55	Oligopeptide/dipeptide ABC transporter ATPase subunit
sc03_1237	-1.84	-1.55	-1.49	-1.42	-1.58	LclR family transcriptional regulator
sc03_1622	-1.42	-1.21	-2.01	-1.69	-1.58	Cold-shock DNA-binding domain protein
sc04_1597	-1.91	-1.81	-1.64	-1.52	-1.72	Putative cysteinyl-tRNA synthetase
sc08_0001	-1.67	-1.79	-1.92	-1.56	-1.73	Phosphoenolpyruvate-dependent sugar phosphotransferase
sc04_0830	-1.8	-1.77	-1.69	-1.7	-1.74	Glutamate synthase subunit beta
sc06_0083	-1.84	-1.98	-1.71	-1.85	-1.85	Phenylacetic acid degradation protein PaaN
sc04_2933	-2.04	-1.99	-1.7	-1.82	-1.89	Acyl-CoA synthetase AMP-forming /AMP-acid ligase II-like protein
sc07_0022	-2.12	-1.92	-2.12	-2.02	-2.04	Putative secreted serine-rich protein
sc04_2720	-3.72	-3.44	-3.71	-2.7	-3.39	Regulatory protein MarR



### ***S. roseifaciens* efficiently converts JA to JA-Gln in liquid cultures**

It was shown previously that plant hormone JA is able to elicit antibiotic production in *S. roseifaciens*, and that *S. roseifaciens* aminoacylates JA to form jasmonoyl-glutamine (JA-Gln) (jasmonoyl-glutamine; van der Meij, 2020). To investigate the effect of JA on *S. roseifaciens*, and to identify possible enzymes involved in the conversion of JA to JA-Gln, an induction experiment was carried out. A 500 mL pre-culture was set up, and sampling started after 16 h ( $t=0$ , early exponential phase). After three samples were taken, the culture was evenly divided over 30 smaller culture-flasks with or without JA (15 flasks each). For both experimental and control group, mycelia of three cultures were harvested after 15 min, 30 min, 1 h, 2 h, and 4 h. For each sample, a small part was kept for proteomics, the rest was extracted for metabolomic analysis, to investigate the conversion of JA-Gln.



**Figure 6. Relative intensity of jasmonoyl-glutamine at different time points.**

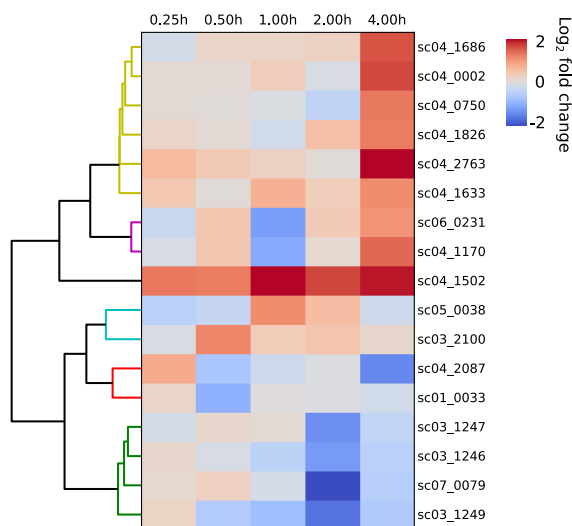
Mycelium was grown in TSBS, a pre-culture was grown for 16 h before JA treatment ( $t = 0$  h). The time after JA treatment was shown on the x axis. Samples were taken in parallel with proteomics samples. The same amount of extract was injected into the LC-MS system. Error bars indicate the standard errors of three replicates.

The formation of JA-Gln was previously identified by both MS and NMR in the extracts of *S. roseifaciens* grown in the presence of JA for 5 days (van der Meij, 2020). JA-Gln was detected using LC-MS by its retention time, together with its MS and MS<sup>2</sup> spectra. In the MS spectra, the  $[M + H]^+$  ion of the compound was observed at  $m/z$  339.1925. As for the MS<sup>2</sup> spectra, the most characteristic fragments observed were the  $[M + H]^+$  ions of Gln ( $m/z$  147.08) and JA-H<sub>2</sub>O ( $m/z$  193.12) (Figure S1), confirming this compound is indeed JA-Gln. The peak area of JA-Gln was extracted for semi-quantification. The results show that JA-Gln was

clearly detected 30 min after induction, increasing 1- and 2-fold at 1 h and 2 h after induction, respectively (Figure 6). Unfortunately, JA does not ionize well in the positive mode, so it was not detected in the LC-MS analysis performed.

### Effect of JA on the proteome of *S. roseifaciens*

Pearson's correlation coefficients were calculated for proteomics data and a cut-off score of 0.97 was maintained for the correlation within triplicates. As a result, control sample at 1 h after induction was removed. The data normalization and quantification were done again using the rest of samples, ensuring best data quality. Subsequently, single quantifications amongst replicates were removed. In total, 2,137 proteins were quantified in at least one replicate. For each sample, the number of quantified proteins ranged from 1,981 to 2,058. New correlation coefficients were calculated between samples (Figure S2). The 4 h samples stood out the most in comparison to the other samples. No clear difference was observed from the correlations between JA and control samples indicating a clear shift in the proteome of *S. roseifaciens* after 2 h. Comparing JA treated samples with control samples at the same time points, only 17 DEPs conformed to the criterion of two-fold difference in protein levels and a  $p$ -value  $< 0.05$  in T-test (Figure 7, Table S2).



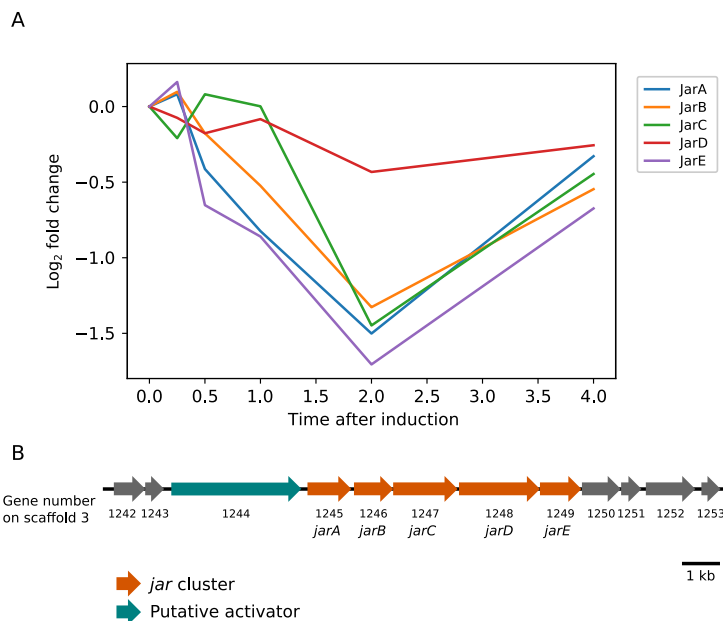
**Figure 7. Heatmap showing log<sub>2</sub> fold change of protein level comparing samples from JA-treated cultures and control cultures.** Only proteins significantly changed (log<sub>2</sub> ratio  $\geq 1$ ,  $p$ -value  $< 0.1$ ) are shown and clustered based on the hierarchical distance of the profile.

### Jar enzymes are suppressed by JA

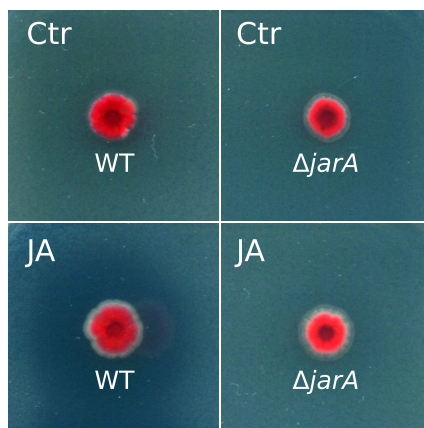
Among the 17 DEPs, SC03\_1246, SC03\_1247, SC03\_1249, and SC07\_0079 formed a cluster that were down-regulated 2 h after induction in the JA-treated

samples (Figure 7). Proteins numbered from SC03\_1245 to SC03\_1249 all showed a similar pattern (Figure 8A), suggesting they are encoded by a single gene cluster. *In-silico* operon analysis suggested that the genes SC03\_1242 to SC03\_1253 most likely form a single operon, which we designated as the JA-responsive (*jar*) cluster (Figure 8B). SC03\_1245 to SC03\_1249 were nominated JarA-E. Interestingly, JarA, JarB and JarC share high homology with the paerucumarin biosynthetic enzymes PvcA, PvcB, and PvcC (BLAST positives 63%, 64%, 78%; e-value  $9 \times 10^{-100}$ ,  $3 \times 10^{-96}$ , 0, respectively) from *Pseudomonas aeruginosa* (Clarke-Pearson & Brady, 2008). Paerucumarin is an isocyanide functionalized coumarin that affects biofilm development by *Pseudomonas aeruginosa*, possibly through modulating the intracellular signalling pathways including the extracellular ion binding pathway (Qaisar *et al.*, 2016). The biosynthesis of paerucumarin requires PvcABCD, with PvcC and PvcD functioning as two-component flavin adenine dinucleotide-dependent monooxygenases (Clarke-Pearson & Brady, 2008). In *S. roseifaciens*, an orthologue of *pvcD* is missing. In *E. coli*, however, the orthologue of PvcC/JarC suffices for oxidation of its substrate. We therefore propose that JarABC may catalyse the formation of a paerucumarin-like compound, whereby the biosynthetic enzymes are suppressed in response to JA. The *jar* cluster also encodes a putative OmpR/PhoB type transcriptional regulator (SC03\_1244), similar to the global regulator of antibiotics AfsR in *S. coelicolor* (Blast identity 34%). Like AfsR, SC03\_1244 has a DNA binding domain and a BTAD domain which possibly mediates its interaction with other regulators (Yeats *et al.*, 2003, Floriano & Bibb, 1996, Horinouchi, 2003). It is yet unclear if this regulator controls the expression of the *jar* cluster.

To analyse the possible role of the *jar* cluster in the response to JA, a *jarA* knock-out mutant was created using CRISPR-Cas9 by removing the +65 to +913 region relative to the translational start of *jarA*. Agar diffusion assays with *E. coli* ASD19 as the indicator strain revealed that *jarA* mutants had lost the ability for its antimicrobial activity to be induced by JA (Figure 9). This suggests that indeed the *jar* operon plays a role in this process. One hypothesis is that the metabolite produced by the Jar proteins may regulate antibiotic production by *S. roseifaciens* in a concentration-dependent manner, with an inverse correlation between the metabolite concentration and antibiotic activity. Further work is needed to identify the nature of the metabolite produced from the *jar* gene cluster and elucidate the signal transduction pathway from JA sensing to the elicitation of antibiotic production.



**Figure 8. Jar proteins suppressed by JA and gene organization of the jar gene cluster in *S. roseifaciens*.** A) Log<sub>2</sub> fold change of the average protein levels between jasmonic acid treatment and control sample. B) Genes of the *jar* cluster (coloured genes) and its surrounding genes on the genome.



**Figure 9. Antimicrobial assays for *jarA* knock-out mutant of *S. roseifaciens*.** Spores of *S. roseifaciens* were spotted on MM agar media supplemented with 1% glycerol and 0.5% mannitol and grown for 5 days before overlay with the indicator strain *E. coli* ASD19. Pictures were taken 16 h after overlay.

### JA can promote the level of an MFS transporter

Among the 17 DEPs, SC04\_1502 was the only protein that was up-regulated at all time points from 15 min to 4 h. SC04\_1502 is a putative major facilitator superfamily (MFS) transporter with 14 transmembrane domains. Functional analysis of SC04\_1502 showed that this protein is a tetracycline-type resistance protein/drug resistance transporter and a likely transporter of small molecules (InterPro entry IPR001411). MFS transporters have varying substrate specificity. High specificity proteins include the tetracenomycin exporter TcmA of *Streptomyces glaucescens* (Martín *et al.*, 2005) and the riboflavin exporter RfnT in *Rhizobiaceae* (Vitreschak *et al.*, 2002); lower specificity proteins include the multiple drug resistance efflux pump EmrB in *E. coli* (Lomovskaya & Lewis, 1992) and the siderophore iron transporter Str1 in *Schizosaccharomyces pombe* (Pelletier *et al.*, 2003).

To test if the transporter may be involved in the transport of JA or JA-Gln, a SC04\_1502 knock-out mutant was made by replacing the +372 to +1628 region (relative to the translational start site) with the apramycin resistance cassette, followed by deletion of the cassette using Cre recombinase so as to avoid polar effects. Subsequently, the in-frame deletion mutant was tested for its response to JA. While the antimicrobial activity of the mutant was decreased as compared to the parental strain, the bioactivity was still inducible by JA (Figure S3). However, this does not necessarily rule out the possibility of the involvement of SC04\_1502 in transporting JA and/or JA-Gln. For example, its function may be compensated by other enzymes such as the highly abundant SC04\_3061 (Figure S4) or by diffusion. Further work is needed to elucidate the function of the MFS family transporter SC04\_1502 and its role in the response of JA.

### Conclusion

In this chapter, the response of *S. roseifaciens* to different candidate elicitors of natural product biosynthesis was compared using quantitative proteomics. Selected candidates include plant hormones and sugars derived from natural polysaccharides. Of these, hydroxycoumarin induced a response that is distinct from that of the 12 other growth conditions. This may be caused by changes in energy flow indicated by the changes in ATP generating enzymes. Jasmonic acid, N-acetylglucosamine, chitosan and benzoic acid induced a surprisingly similar response, the basis for which remains unknown. Comparison of the proteomic changes revealed the up-regulation of a few metabolic enzymes. GlmS, which connects glycolysis to GlcNAc metabolism was among the most strongly up-regulated enzymes in all samples within this group. The proteomics data also show two MarR-family regulators that were down-regulated.

We paid special attention to the plant hormone JA, as previous data obtained in our laboratory showed that many Actinobacteria respond to JA by enhancing their antibiotic production (Anne van der Meij, PhD thesis). In this research, we found one BGC that was down-regulated in *S. roseifaciens* when JA was added and was therefore designated *jar* to highlight its responsiveness to JA. The *jar* cluster shares homology with the BGC for paerucumarin production in *P. aeruginosa*. A *jarA* mutant was made, which indeed showed altered antibiotic production and hence its possible involvement in the JA response. In fact, *S. roseifaciens* had lost its ability to respond to JA, whereby the induction of antibiotic production by JA had been lost in the *jarA* null mutant. A second candidate element of the JA signalling pathway that we identified was an MFS-family transporter, whose level was significantly promoted by JA. The precise roles of these proteins in JA sensing and the resulting change in antibiotic production awaits further investigation.

## Material and methods

### Strains and culturing conditions

*S. roseifaciens*, previously *Streptomyces* sp. MBT76 (van der Aart *et al.*, 2019), was used in this study. *E. coli* ASD19 (Avalos *et al.*, 2018) was used as an indicator strain in antimicrobial activity assays. *E. coli* JM109 was used for routine cloning. Methylation deficient strain *E. coli* ET12567 containing driver plasmid pUZ8002 (Paget *et al.*, 1999) was used in conjugation experiments for introducing DNA into *Streptomyces*.

For solid-grown cultures, spores collected from 14 days MYM culture of *S. roseifaciens* were used. MM agar (Kieser *et al.*, 2000) supplemented with 0.5% mannitol and 1% glycerol was used for proteomics culture preparation. Approximately  $10^7$  spores were inoculated on each plate covered with cellophane and grown at 30°C for 48 h. Then the cellophane disks were carefully moved to a new plate containing desired concentration of small molecules (Table S5) and grown for another 48 h. Mycelium was then collected using a small spatula from the cellophane and transferred to a 2 mL EP tube with two 4 mm diameter ion beads. The samples were snap-frozen in liquid nitrogen and stored in -80°C for protein extraction.

*S. roseifaciens* mycelial stocks were made by inoculating 500  $\mu$ L initiator mycelium in 250 mL TSBS medium in a 2 L Erlenmeyer flask equipped with a steel spring, this culture was grown at 30°C with constant shaking at 200 rpm. After 2 days of growth, mycelium was collected by centrifugation and resuspended in 20% glycerol and aliquots was made and stored at -80°C. 1.5 mL *S. roseifaciens* mycelium stock was inoculated into 500 mL pre-warmed TSBS medium in a 3 L Erlenmeyer flask equipped with a spring and culture in a 30°C shaking incubator with a constant shaking speed of 200 rpm. After 16 h growth, 3  $\times$  15 mL culture was collected as sample time 0, the rest of cultures was divided per 15 mL to new 100 mL flasks with JA or with equal amount of solvent (ethanol). The final concentration of JA was 0.01% (w/v). At each of the following time point: 15 min, 30 min, 1 h, 2 h and 4 h, samples were collected in triplicate. For each sample, 2 mL was pelleted for proteomics sample preparation, the remaining volume was used for metabolomics analysis.

### Knockout mutants of *S. roseifaciens*

Oligonucleotides used in this study was listed in Table S3. PCR was preformed using Pfu DNA polymerase using standard protocol (Colson *et al.*, 2007). All plasmid and constructs used for generating knockout mutants are summarized in Table S4. The constructs generated in this study were verified by Sanger sequencing performed in BaseClear (Leiden, The Netherlands). *E. coli* strains were grown in Luria broth at 37°C supplemented with the appropriate antibiotics

(ampicillin, apramycin, kanamycin and/or chloramphenicol at 100, 50, 25 and 25  $\mu\text{g}\cdot\text{mL}^{-1}$ , respectively) depending on the vector used (Table S4).

Knockout of *jarA* was preformed using pCRISPR-Cas9 technology (Tong *et al.*, 2015a). A spacer sequence was designed at the middle of *jarA*, with the sequence of ACAGGTCCTCCAGCATGAAG. The design of sgRNA guides Cas9 to make a double strand brake at around nt position +610 relative to the *jarA* translational start site. After spacer integration, the PCR product of *jarA* flanking regions for homology-directed repair designed to knock out the +65 to +913 regions (relative to the *jarA* translational start site) were inserted and a construct made using Gibson assembly. The resulting knockout construct was designated pGWS1460. This construct was introduced into *S. roseifaciens* through conjugation. Loss of the pGWS1460 was confirmed by the apramycin sensitivity of the exconjugants, knockout of *jarA* was further confirmed by PCR.

Knockout of the MFS family protein SC04.1502 was preformed using pWHM3-oriT system which was a conjugatable version of pWHM3 (Vara *et al.*, 1989). The knockout strategy was essentially the same as described previously (Zhang *et al.*, 2018), conjugation was used to introduce construct to *S. roseifaciens*. Basically, around 1500 bp of upstream and downstream region of SC04.1502 was amplified by PCR from *S. roseifaciens* genome. The upstream region was thereby cloned as *HindIII-XbaI* fragment, and the downstream region was cloned as *XbaI-EcoRI* fragment. The apramycin resistance cassette *aac(3)IV* with flanking *loxP* sites was digested from construct made in previous study (Zhang *et al.*, 2018). These three fragments were ligated into *EcoRI-HindIII*-digested pWHM3-oriT. The presence of the *loxP* recognition sites allows the efficient removal of the apramycin resistance cassette following the introduction of a plasmid pUWL-Cre expressing the Cre recombinase (Fedoryshyn *et al.*, 2008, Khodakaramian *et al.*, 2006). Knock-out construct pGWS1461 was created for the deletion of nucleotide positions +372 to +1628 of SC04\_1502, where +1 refers to the translation start site of SC04\_1502. Introducing this construct to *S. roseifaciens* followed by losing this construct resulted in replacing the respective region with apramycin resistance cassette, which was subsequently removed using Cre expressing construct pUWLCre. The confirmation of knockout and losing the apramycin resistance cassette was confirmed by PCR followed by Sanger sequencing of the product.

### Metabolomics analysis

For each sample collected from liquid grown cultures, about 0.5 g of Diaion® HP20 (Resindion, Binasco, Italy) was added and the tube was then placed on a rotator at 4°C overnight. The HP20 beads were then collected on glass wool, washed with water, and then soaked in methanol overnight at room temperature. The methanol was evaporated under vacuum at 40°C and the residue was dissolved in methanol again at a fixed concentration of 1  $\text{mg}\cdot\text{mL}^{-1}$  for analysis.



LC-DAD-HRESIMS spectra were obtained using a Waters Acquity UPLC system, equipped with Waters Acquity PDA, and coupled to a Thermo Instruments MS system (LTQ Orbitrap XL). The UPLC system was run using Acquity UPLC HSS T3 C<sub>18</sub> column (1.8  $\mu$ m, 100 Å, 2.1  $\times$  100 mm). Solvent A was 0.1% formic acid, 95% H<sub>2</sub>O and 5% acetonitrile. Solvent B was 0.1% formic acid, 95% ACN and 5% H<sub>2</sub>O. The gradient used was 2% B for 0.5 min, 2-40% for 5.5 min, 40-100% for 2 min, and 100% for 3 min. The flow rate was 0.5 mL·min<sup>-1</sup>. The MS conditions used were capillary voltage 5 V, capillary temperature 300 °C, auxiliary gas flow rate 5 arbitrary units, sheath gas flow rate 50 arbitrary units, spray voltage 3.5 kV, mass range 100-2000 *m/z*, FT resolution 30000. The MS spectra were acquired in the positive mode.

### Sample preparation for proteomics

For mycelium collected from plates, the 2 mL EP tubes containing mycelium and metal beads were put into an adaptor and disrupted using TissueLyser II (Qiagen, Venlo, The Netherlands) for 2 cycles of 30 s at 30 Hz. The samples were kept frozen during this process by cooling using liquid nitrogen before and in between cycles. 0.6 mL disrupting buffer (4% SDS, 0.06 M DTT, 100 mM Tris-HCl pH 7.6, 50 mM EDTA) is then added. Samples were vortexed to allow mixing with disrupting buffer and restore to room temperature. Cell debris was then removed by centrifugation at 16,000  $\times$  g for 10 minutes at 20°C, 50  $\mu$ L supernatant was taken for protein precipitation.

For liquid grown cultures, samples were centrifuged at 5,000  $\times$  g for 10 min, cell pellets were resuspended in 1 mL disrupting buffer and transferred to 1.5 mL screw-cap tube with about 0.5 g 0.1 mm glass beads. All samples were disrupted using FastPrep FP120 (Thermo, Pennsylvania, U.S.) at speed 6.0 (m·s<sup>-1</sup>) for 5 times 30 s disruption in a 4°C cold room, with 5 minutes in between runs to cool down the samples. Samples after disruption were centrifuged at 16,000  $\times$  g for 10 minutes at 20°C to remove cell debris. 50  $\mu$ L supernatant was taken for protein precipitation.

Protein extracts from both liquid and plate grown cultures were precipitated using chloroform-methanol method described by (Wessel & Flugge, 1984). Briefly, 200  $\mu$ L methanol was added to 50  $\mu$ L sample and mixed, then 100  $\mu$ L chloroform was added and mixed. 150  $\mu$ L water was then added and vortexed for 30 s. Phase separation was achieved by centrifugation at 16,000 g for 5 minutes. After removing the water phase, 150  $\mu$ L methanol was added to the remainder followed by vortex and centrifugation at 16,000 g for 5 minutes to precipitate and pellet proteins. Supernatant was then removed, and the protein pellets were then dried in a vacuum centrifuge.

Protein pellets were resuspended in 50  $\mu$ L of 0.05 % RapiGest SF solution (in 50 mM ammonium bicarbonate) and incubated at 95°C for 5 minutes to dissolve as

much protein as possible. 14  $\mu\text{L}$  was taken for BCA assay to determine protein concentration. 5 mM DTT was added in the rest of samples and incubate at 60°C for 30 minutes. Iodoacetamide was then added to 21.6 mM, and samples were incubated at room temperature in dark. Trypsin was then added at an enzyme to substrate ratio of 1:100 allowing overnight digestion at 37°C. Samples were then acidified to degrade RapiGest by adding 5% TFA to 0.5% and incubated at 37°C for 60 minutes, followed by centrifugation at 16,000 g for 10 min. The supernatant was desalted using STAGE-tip microcolumns as described previously (Gubbens *et al.*, 2014, Rappsilber *et al.*, 2007). Briefly, 8  $\mu\text{g}$  of acidified peptides of each sample was loaded on STAGE-tip, washed twice with 0.5% formic acid solution, eluted with elution solution (80% acetonitrile, 0.5% formic acid). Acetonitrile was then evaporated in a SpeedVac (Thermo, Pennsylvania, U.S.). Final peptide concentration was adjusted to 100  $\text{ng}\cdot\mu\text{L}^{-1}$  using sample solution (3% acetonitrile, 0.5% formic acid) before analysis.

### UPLC and mass spectrometry measurement and data processing

200 ng (2  $\mu\text{L}$ ) digested peptide was injected and analysed by reversed-phase liquid chromatography on a nanoAcquity UPLC system (Waters, Massachusetts, U.S.) equipped with HSS-T3 C18 1.8  $\mu\text{m}$ , 75  $\mu\text{m}$  X 250 mm column (Waters, Massachusetts, U.S.). A gradient from 1% to 40% acetonitrile in 110 min (ending with a brief regeneration step to 90% for 3 min) was applied. [Glu<sup>1</sup>]-fibrinopeptide B was used as lock mass compound and sampled every 30 s. Online MS/MS analysis was done using Synapt G2-Si HDMS mass spectrometer (Waters, Massachusetts, U.S.) with an UDMSE method set up as described in (Distler *et al.*, 2014)

Raw data from all samples were first analysed using the vendor software ProteinLynx Global SERVER (PLGS) version 3.0.3. Generally, mass spectrum data were generated using an MS<sup>E</sup> processing parameter with charge 2 lock mass 785.8426, and default energy thresholds. For protein identification, default workflow parameters except an additional acetyl in N-terminal variable modification were used. Reference protein database was made from predicted 7974 coding sequences described in (Wu *et al.*, 2016c). The resulted dataset was imported to ISOQuant (version 1.8, Distler *et al.*, 2014) for label-free quantification. ISOQuant processing includes spectrum alignment, data normalization at spectrum level, peptide quantification by top3 method and subsequent protein quantification. Default high identification parameters were used in the quantification process.

### Acknowledgments

We would like to acknowledge Bogdan I. Florea of Leiden University, Leiden, The Netherlands, for running and for monitoring the proteome measurements, and the

bio-organic synthesis group at Leiden University for providing the opportunity to use their instrumentation.

# Analysis of the background-reduced antibiotic production host *Streptomyces coelicolor* M1152 using quantitative proteomics

Chao Du<sup>1</sup>, Dino van Dissel<sup>2</sup>, Alexander Wentzel<sup>2</sup>, Gilles P. van Wezel<sup>1</sup>

1. Microbial Biotechnology, Institute of Biology, Leiden University, Leiden, The Netherlands

2. Department of Biotechnology and Nanomedicine, SINTEF Industry, Trondheim, Norway

Sulheim, S., Kumelj, T., van Dissel, D., Salehzadeh-Yazdi, A., Du, C., van Wezel, G.P., Nieselt, K., Almaas, E., Wentzel, A., and Kerkhoven, E.J. (2020) Enzyme-constrained models and omics analysis of *Streptomyces coelicolor* reveal metabolic changes that enhance heterologous production. iScience: 101525.

DOI: 10.1016/j.isci.2020.101525

## Abstract

Streptomycetes are major producers of bioactive natural products. Genome mining has uncovered a wealth of biosynthetic gene clusters (BGCs) for bioactive molecules, yet many of the clusters are silent under laboratory conditions. An important strategy to explore these potentials is to express BGCs in an optimised heterologous production host, but the effects of the genomic changes on growth and production are not well understood. Here, using quantitative proteomics, we have analysed the differences in global protein profiles between *Streptomyces coelicolor* M145 and its derivative strain M1152, which is a widely used platform for heterologous gene expression. M1152 has four major natural BGCs removed, and in addition a point mutation in *rpoB*. Our data show that the response to phosphate depletion is deregulated in M1152, with delayed up-regulation of several members of the PhoP regulon. Level of proteins belonging to the GlnR regulon was reduced in M1152. GlnR and PhoP are connected regulons that globally control natural product biosynthesis. Biosynthetic proteins for the osmoprotectant hydroxyectoine were over-represented in strain M1152, while the level of  $\gamma$ -butyrolactone (GBL) synthase ScbA and its GBL-responsive regulator ScbR was strongly induced. Taken together, this work reveals that M1152 has undergone significant changes in the global regulatory networks that control antibiotic production. This provides novel insights into the consequences of strain-optimization approaches and will provide guidance for future efforts in this important field of research.

## Introduction

Natural products are the primary source of antibiotics, antitumor compounds, food preservatives, and many more (Newman & Cragg, 2016, Hopwood, 2007). Human society has benefited enormously from the introduction of new antibiotics, allowing the treatment of infections with microbial pathogens. Recently, however, the dramatic decline of new antibiotics discovery combined with a rapid rise of multidrug resistant infections impose huge new threats to human health (Kolter & van Wezel, 2016, Cooper & Shlaes, 2011, Baltz, 2008). This antibiotic crisis had led to the urgent need of novel antibiotics (Baltz, 2007, Zhu *et al.*, 2014a). Most antibiotics are produced by members of the Actinobacteria, and the majority of those by streptomycetes (Barka *et al.*, 2016). Since the first complete genome of the model organism *S. coelicolor* was published, it is becoming increasingly clear that *Streptomyces* genomes harbour a very large repository of biosynthetic gene clusters (BGCs) that might potentially produce new compounds (Bentley *et al.*, 2002, Medema & Fischbach, 2015, Ziemert *et al.*, 2016).

One major effort in natural product discovery is to find, study, and manipulate the biosynthetic gene clusters (BGCs) of the producer organisms. This includes engineering the expression of biosynthetic enzymes and increasing the precursor and intermediates pool, thereby optimizing the production of the natural products which are already produced by native strains (Komatsu *et al.*, 2010, Myronovskiy & Luzhetskyy, 2019). However, the activation of the cryptic biosynthetic potential remains difficult. This is among others due to the fact that we do not yet understand the ecological conditions under which the biosynthetic gene clusters are activated in the natural habitat (van der Meij *et al.*, 2018). To overcome these limitations, expression in heterologous production hosts is becoming indispensable (Nepal & Wang, 2019, Gomez-Escribano & Bibb, 2011). One advantage of heterologous expression is that we can choose genetically well understood strains that can be cultured robustly and are genetically tractable. One of the most important organisms for heterologous expression is *Streptomyces*. Heterologous expression in *Streptomyces* is now becoming a popular tool for novel antibiotics discovery and production (Myronovskiy & Luzhetskyy, 2019, Ahmed *et al.*, 2020, Gomez-Escribano & Bibb, 2011). Heterologous expression can also be used to determine the boundaries of the BGC (Komatsu *et al.*, 2013, Liu *et al.*, 2018), to create artificial natural products by combining different BGCs (Park *et al.*, 2011), or for characterizing the functions of individual genes (Waldman *et al.*, 2015).

*Streptomyces coelicolor*, as the most studied *Streptomyces* model organism (Bentley *et al.*, 2002, Hoskisson & van Wezel, 2019), is a producer of many different natural products, including five antibiotics, namely actinorhodin (ACT), undecylprodigiosin (RED), calcium-dependent antibiotic (CDA), coelimycin (CPK) and the plasmid-encoded methylenomycin (MMY). Antibiotic production is tightly

controlled, and correlates to the onset of development (Bibb, 2005, van der Heul *et al.*, 2018). Nutrient availability is thereby an important factor, and coordinated lysis of the vegetative or substrate mycelium precedes the onset of morphological and chemical differentiation (Chater & Losick, 1997, Manteca *et al.*, 2005, Tenconi *et al.*, 2018). Upon nutrient limitation, *S. coelicolor* responds with a flux change of energy and intermediate metabolites which initiate secondary metabolism (van Keulen & Dyson, 2014). This is a tightly controlled process that resulted in major metabolic change from primary to secondary metabolism, so this process is often referred as metabolic switch. The regulation of the metabolic switch involves both global and pathway-specific regulators that form complex regulatory networks (Rigali *et al.*, 2008, Romero *et al.*, 2014, Urem *et al.*, 2016). Depletion of inorganic phosphate (Pi) is one of the most studied inducer of major metabolic switch (Martín, 2004, Nieselt *et al.*, 2010). The primary response to Pi depletion is mediated through the two-component system PhoR-PhoP (Sola-Landa *et al.*, 2003, Sola-Landa *et al.*, 2005). PhoP acts as a global regulator of various metabolomic processes, including secondary metabolite biosynthesis (Martín *et al.*, 2011, Sola-Landa *et al.*, 2003, Santos-Beneit *et al.*, 2009). Another important global regulator that related with nutrient depletion is GlnR, which controls nitrogen uptake and metabolism related genes (Fink *et al.*, 2002, Tiffert *et al.*, 2008, Amin *et al.*, 2012, Wang & Zhao, 2009). Interestingly, there is significant cross-talk between the P- and N-responses in *Streptomyces*, due to partially overlapping regulons of GlnR and PhoP (Sola-Landa *et al.*, 2012, Santos-Beneit *et al.*, 2012). Novel systems biology approaches of multi-dimensional, multi-omics global analyses are required to help us understand the regulatory networks in a more comprehensive way (Hwang *et al.*, 2014, Rokem *et al.*, 2007). Another gene that has great regulatory impact on antibiotic production in *Streptomyces* is *rpoB*. It encodes an RNA polymerase  $\beta$ -subunit known to respond to stringent response elicitor guanosine 5'-diphosphate 3'-diphosphate (ppGpp), which causes rapid cessation of RNA synthesis and other cellular reactions (Cashel, 1996, Xu *et al.*, 2002). It has been suggested that the *rpoB*[C1298T] mutation functions by mimicking the activated form in promoting the onset of secondary metabolism in *Streptomyces* (Hu *et al.*, 2002).

In order to explore the full potential of *S. coelicolor* as a superhost for the production of antibiotics or other valuable natural product, we here analysed the background-reduced host *S. coelicolor* M1152 with its parent, *S. coelicolor* M145. As an important production host (Gomez-Escribano & Bibb, 2014, Nepal & Wang, 2019), *S. coelicolor* M1152 was generated by removal of four major BGCs, namely the polyketides ACT, RED and CPK, and the non-ribosomal peptide antibiotic CDA (Gomez-Escribano & Bibb, 2011). In addition, mutations were introduced in *rpoB*[C1298T] which boosts antibiotic production (Hu *et al.*, 2002).

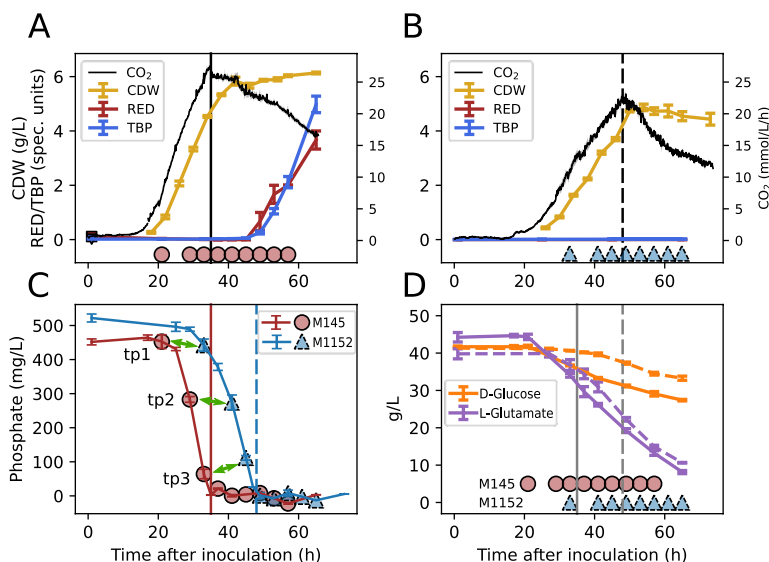
To obtain more insights into how the metabolic switch redirects primary and secondary metabolism, a proteomics study was done as part of the systematic rebuilding of Actinomycetes for natural product formation (SYSTERACT) project using a multi-Omics approach, and applying the optimised cultivation system published previously (Wentzel *et al.*, 2012a). Here, we compared the protein profiles of *S. coelicolor* M1152 and its parent M145 using samples from batch fermenters obtained from the SYSTERACT project. For this, a new advanced proteomics pipeline was applied, featuring label-free and TOP3 quantification techniques. The differences between the two strains and the impact of the new insights on the use of heterologous production strains are discussed.



## Results and Discussion

### Comparison of the growth of background-reduced expression host *S. coelicolor* M1152 with its parent M145

We aimed to investigate the changes in global protein profiles in *S. coelicolor* M1152 as compared to its parent *S. coelicolor* M145. *S. coelicolor* M145 and M1152 were cultivated in bioreactors in SSBM-P media. SSBM was designed with two carbon sources, D-glucose and L-glutamate, to ensure sufficient biomass before the onset of secondary metabolism. Use of L-glutamate as the sole nitrogen source leads to a well-defined transition phase, which allows antibiotic production after sufficient biomass accumulation (Wentzel *et al.*, 2012a). This established system was tested for high-resolution time-course analyses on the levels of transcriptomics (Nieselt *et al.*, 2010, Alam *et al.*, 2010), proteomics (Thomas *et al.*, 2012), and metabolomics (Wentzel *et al.*, 2012b).

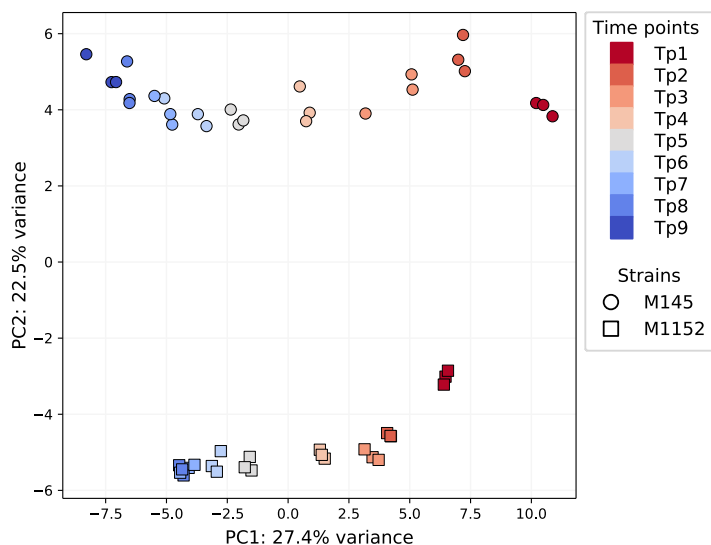


**Figure 1. Batch cultivation data of *S. coelicolor* M145 and M1152.** A) M145 and B) M1152 were grown on SMMB-P medium in benchtop bioreactors in triplicate. Time points of collecting proteomics and transcriptomics samples are indicated by purple triangle on the x axis line. CDW, cell dry weight; RED, undecylprodigiosin; TBP, total blue pigments / actinorhodins. Error bars for CDW, RED, TBP and greyed area for CO<sub>2</sub> generation represent the standard error of three biological replicates. C) Phosphate (Pi) concentration measured during the same batch culture. The first 3 sampling time points (tp1 to tp3) are indicated by green arrows. D) Concentration of the two carbon sources during fermentation of the two strains. Circle and triangles in all subplots indicate the time points of collecting proteomics and transcriptomics samples for *S. coelicolor* M145 and M1152, respectively. Vertical lines indicate the time when phosphate in the medium has been depleted for respective fermentation experiments.

Samples were taken in 4 h intervals during exponential growth, around the metabolic switch and during the stationary phase (when natural product formation takes place) (Figure 1). Growth of M145 reached stationary phase at around 42 h with a cell dry weight (CDW) of 5.8 g·L<sup>-1</sup>, while for M1152, the stationary phase was reached at around 50 h with a CDW of only 4.7 g·L<sup>-1</sup>. This alteration in the growth characteristics is consistent with the literature (Gomez-Escribano & Bibb, 2011). The online analysis showed maximum CO<sub>2</sub> production after 35 h for M145 and after 47 h for M1152 (Figure 1A, B). This is consistent with the timing of phosphate depletion in both strains (Figure 1C). Shortly after phosphate depletion, both ACT and RED were produced by *S. coelicolor* M145, indicating that the metabolic switch had induced antibiotic production (Figure 1A). D-glucose and L-glutamate were consumed concomitantly during fermentation, with glutamate consumed slightly faster than glucose (Figure 1D). Both remained in excess throughout the cultivation. The consumption of glucose and glutamate by M1152 were all reduced, with glucose consumption reduced more.

### **Proteomics analysis reveals major differences between *S. coelicolor* M145 and M1152**

To investigate the changes in more detail, samples from the fermentations with M145 and M1152 were compared by proteomics. 27 mycelial samples were collected from *S. coelicolor* M145 corresponding to triplicate samples from 9 time points, and 24 mycelial samples were collected from *S. coelicolor* M1152 corresponding to triplicate samples from 8 time points (Figure 1). In total, 3,278 proteins were identified combining all samples from *S. coelicolor* M145 and M1152 fermentations. After the removal of proteins with low quantities (see Materials and Methods section for details on the criteria), 1,879 proteins were reliably quantified in at least one time point (Table S1). These same samples were also analysed by RNA-seq (Sulheim *et al.*, 2020), allowing comparison of proteome and transcriptome data.



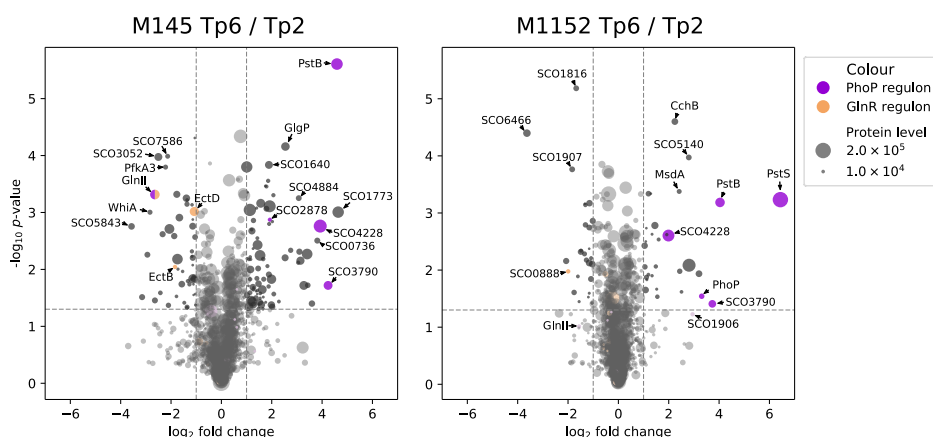
**Figure 2. Principal component analysis (PCA) of proteomics data from all samples.** The variance stabilizing transformed (VST) was used. The proteins from genes deleted in M1152 were excluded. Aligned sampling time points are coloured coded from red to blue, different strains are distinguished by the shape of data point.

As *S. coelicolor* M1152 grows slower than its parent M145 (Figure 1), data were aligned to represent a comparative growth phase. The metabolic switch occurred between the third and fourth sampling points for both strains (Figure 1C). This allowed sequential alignment of the samples from the two strains, named tp(n). To understand the overall differences between different samples, variance stabilizing transformation (VST) of the normalized TOP3 quantification data was done to reduce the systematical variance especially for the low abundant proteins (Love *et al.*, 2014). This normalization was recently reviewed to be excellent for label-free proteomics data (Välikangas *et al.*, 2016). General assessments of data transformation are shown in Figures S1 and S2. To investigate the coherence of the collected protein concentrations between the different samples a principal component analysis (PCA) performed on the transformed dataset (Figure 2). All proteins encoded by the four BGCs that had been deleted from M1152 were omitted from the PCA analysis. The PCA plot shows a clear time course migration of samples on principal component 1 (PC1) between samples from the same strain. The continuity of samples from one strain indicates a good stability of our data. Moreover, there was a very clear separation between *S. coelicolor* M145 and M1152 samples on principal component 2 (PC2), which represents 22.5% of the variance (Figure 2), indicating that the genome modifications in M1152 led to a profound overall protein level alteration. However, like for M145, M1152 samples also shifted to one side of PC1 in a time-dependent manner. To investigate which

proteins caused the general shift in the protein profiles, protein levels were compared between *S. coelicolor* M1152 and M145 for the aligned time points, whereby the raw quantification value with fold change  $\geq 2$  or  $\leq 0.5$  and with T-test  $p$ -value  $< 0.05$  were considered differentially expressed proteins (DEPs). In total, there were 408 DEPs in at least one time point comparing the two strains, representing 21.7% of quantifiable proteome in this experiment (Table S2).

### Response of the PhoP regulon to Pi depletion is delayed in strain M1152

Since the metabolic switch was induced by phosphate depletion (Figure 1), we investigated the change in protein profiles during phosphate depletion. The same criteria for DEP were used comparing selected timepoints before (tp2) and after (tp6) Pi depletion within each strain. The distribution of variance in these samples was comparable (Figure S3). In total, 108 DEPs were found in M145 and 47 in M1152 (Table S3). The reduced number of total DEPs in the BGC-deleted background suggests that secondary metabolism may act as a modulator of the phosphate response in *S. coelicolor*. Indeed, many PhoP regulon proteins were up-regulated after Pi depletion (Figure 3, purple dots). Especially for strain M1152, the PhoP regulon proteins were dominant (Fisher's exact test  $p$ -value  $< 0.001$ ) among the DEPs.



**Figure 3. Volcano plot showing comparisons between representative time points before (tp2) and after (tp6) phosphate depletion.** The time points were chosen to represent different developmental state. Proteins known to be regulated by PhoP and GlnR are coloured purple and orange, respectively. The protein names of outstanding points are marked with an arrow. The higher average protein level of the test and control samples is transformed to the size of each dot. Proteins not significantly changed ( $p$ -value  $\geq 0.05$  or  $0.5 < \text{fold change} < 2$ ) are greyed out as transparent dots. Dashed lines showing the boundaries of significance threshold.

To investigate further how PhoP regulon proteins respond to Pi depletion and the differences of the response between two strains, a time-course protein level curve was made for all known PhoP regulon proteins that were quantified (Figure 4). At the transcriptome level, all genes belonging to the PhoP regulon were shown to be up-regulated immediately after Pi depletion in both strains (Sulheim *et al.*, 2020). This is consistent with previous microarray data based on the same system (Nieselt *et al.*, 2010). As expected, also many proteins that belong to the PhoP regulon were up-regulated directly after Pi depletion. However, some proteins were unchanged or even down-regulated after Pi depletion (Table 1). This delay in the regulation of protein levels may be explained by the effect of post-translational control (Vogel & Marcotte, 2012).

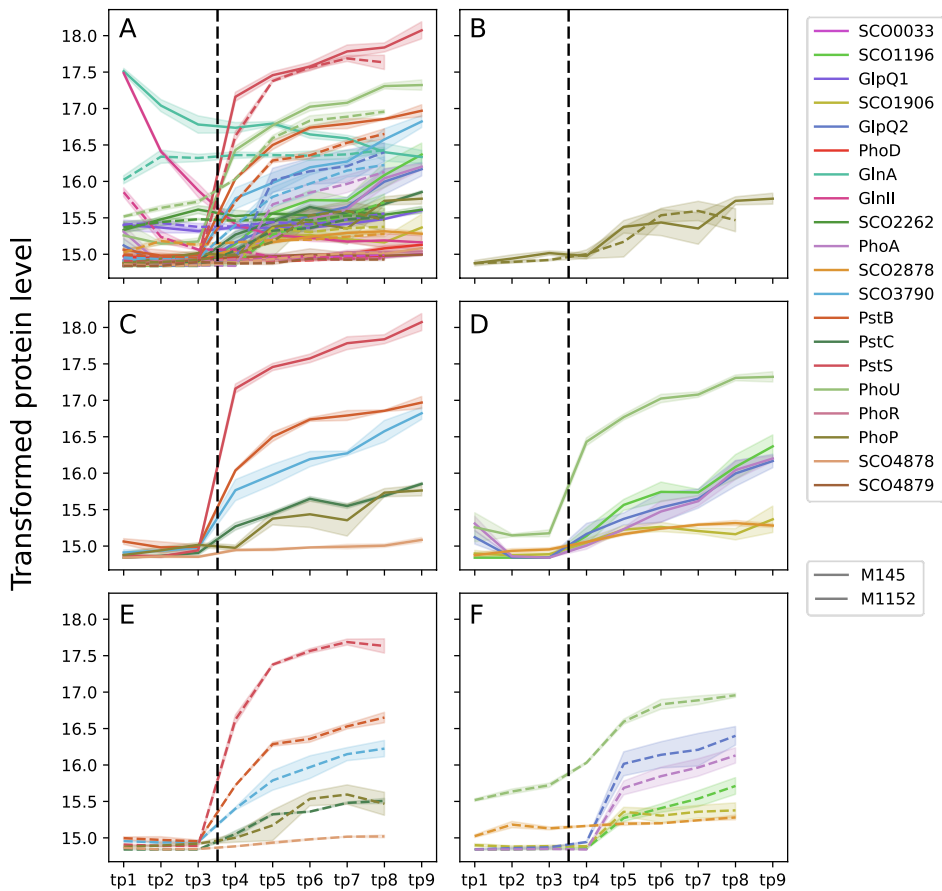
PhoP itself (Figure 4C) showed a delayed response in both strains, whereby the response to Pi depletion was delayed by one sampling interval (4 h). The pattern of protein levels of the PhoP regulon can be categorized into three subcategories (Table 1), namely those that were up-regulated in both M145 and M1152 at the time of Pi depletion (group A), those that showed delayed up-regulation in M1152 (group B), and those that did not show any up-regulation (group C). Group A includes proteins that were up-regulated in both *S. coelicolor* M145 and M1152. This group includes the phosphate transport proteins PstB, PstC, PstS, phosphate scavenging protein SCO3790, and SCO4878 that relates to phosphate storage (Martín *et al.*, 2012) (Figure 4B). The second category consists of proteins that were up-regulated in strain M145 immediately upon Pi depletion but showed delayed or reduced response in M1152 (Figure 4D). Four of these proteins are related to phosphate scavenging. The delayed up-regulation of several proteins of the PhoP regulon and of PhoP itself, in the phase immediately following the Pi depletion, may be one of the factors that causes slow growth. Perhaps this difference may contribute to the slower growth of *S. coelicolor* M1152.

Table 1. Response pattern of PhoP regulon proteins\*

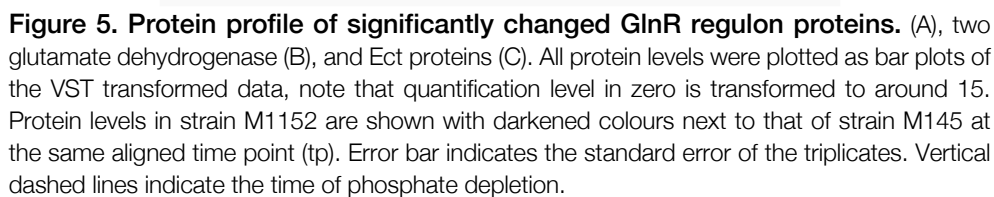
	Protein ID (name)	Response pattern	Functions related with phosphate depletion	Annotation
PhoP	SCO4230 (PhoP)	Delayed up-regulation (both)	Signal transduction	Pho response regulator
	SCO3790	Both up-regulated at Pi depletion	Phosphate scavenging	Putative phosphatase
	SCO4139 (PstB)	Both up-regulated at Pi depletion	Phosphate transport	Phosphate ABC transport system ATP-binding protein
Group A	SCO4141 (PstC)	Both up-regulated at Pi depletion	Phosphate transport	Phosphate ABC transport system permease protein
	SCO4142 (PstS)	Both up-regulated at Pi depletion	Phosphate transport	Secreted phosphate-binding protein
	SCO4878	Both up-regulated at Pi depletion	Phosphate storage	Putative glycosyltransferase
Group B	SCO2878	Up-regulated in M145, but no change in M1152	Unknown	Hypothetical protein
	SCO1196	Delayed up-regulation in M1152	Phosphate scavenging	Putative secreted protein, PLC- like phosphodiesterase motif
	SCO1906	Delayed up-regulation in M1152	Phosphate scavenging	Putative secreted phosphatase
	SCO1968 (GlpQ2)	Delayed up-regulation in M1152	Phosphate scavenging	Secreted glycerophosphodiester phosphodiesterase
	SCO2286 (PhoA)	Delayed up-regulation in M1152	Phosphate scavenging	Secreted alkaline phosphatase
	SCO4228 (PhoU)	Delayed up-regulation in M1152	Signal transduction	Putative modulator of the Pho system
	SCO4159 (GlnR)	Detected only in aligned time point 1	Signal transduction	Transcriptional regulatory protein
	SCO4879	Up-regulated but with fluctuation	Phosphate storage	Conserved hypothetical protein
	SCO2210 (GlnII)	Down-regulated	Cross regulation with GlnR regulon	Glutamine synthetase
	SCO0033	No response	Unknown	Putative secreted neuraminidase
Group C	SCO1565 (GlpQ1)	No response	Phosphate scavenging	Secreted glycerophosphodiester phosphodiesterase
	SCO2068 (PhoD)	No response	Phosphate scavenging	Secreted phospholipase D
	SCO2198 (GlnA)	No response	Cross regulation with GlnR regulon	Glutamine synthetase I
	SCO2262	No response	Unknown	Putative oxidoreductase
	SCO4229 (PhoR)	No response	Signal transduction	Pho histidine kinase
	SCO4261	No response (ND <sup>†</sup> in M1152)	Unknown	Putative response regulator
	SCO5583 (AmtB)	No response (ND in M1152)	Cross regulation with GlnR regulon	Ammonium transporter

\* PhoP regulon gene list adapted from (Allenby *et al.*, 2012), with extension of the proteins belonging to the same regulon: SCO4139 and SCO4141 from SCO4142 regulon, SCO4230 from SCO4228 regulon. Only proteins identified in the proteomics experiment are listed.

<sup>†</sup> Not detected



**Figure 4. Level of detected PhoP regulon proteins in aligned time points.** A), all detected PhoP regulon proteins, including some GlnR regulon proteins that proved to be co-regulated by PhoP. All other subplots are showing a subset of proteins in figure A). Protein levels in *S. coelicolor* M145 are shown in solid lines, strain M1152 are shown in dashed lines. B) PhoP. C) and E) PhoP regulon proteins regulated similarly in both strains (group A in Table 1). D) and F), PhoP regulon proteins differently regulated in the two strains (Group B in Table 1). For all plots, variance stabilizing transformed (VST) data was used for better demonstration of low abundant proteins (Love *et al.*, 2014). Black vertical dashed lines indicating the time of phosphate depletion. Shaded area indicating the standard error of triplicates, colour coding is shared across all subplots. PhoP regulon was adapted from (Allenby *et al.*, 2012), see Table 1 for detailed annotation.





### Levels of the GlnR regulon proteins are reduced in *S. coelicolor* M1152

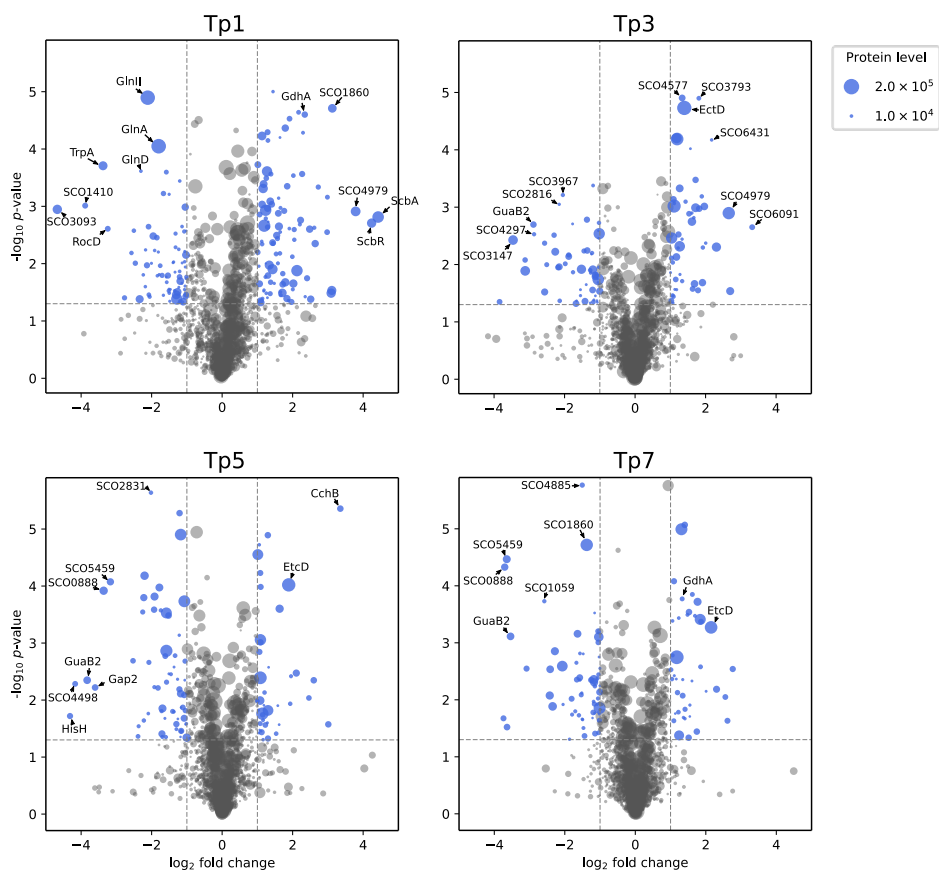
According to the offline measurement of L-glutamate (Figure 1), nitrogen was not depleted during the course of the fermentation. However, the cross-talk between the PhoP and GlnR regulons (Sola-Landa *et al.*, 2012, Santos-Beneit *et al.*, 2012) prompted investigation of the GlnR regulon during Pi depletion. 14 out of 21 reported GlnR regulon proteins were identified by our proteomics analysis. GlnII, GlnA, AmtB, GlnK, GlnD, EctD and EctB were down-regulated only in the parental strain M145 during fermentation (Figure 3 for GlnII, EctD and EctB, Figure 5A for GlnA, AmtB). None of these proteins was changed in M1152 after Pi depletion. Interestingly, GlnA showed up-regulation at tp2 in M1152, while it was down-regulated in M145. Previous studies have shown that *glnA* and *glnII* are differentially regulated and probably related to the control of development (Fink *et al.*, 2002, Reuther & Wohlleben, 2007). The different level of GlnA and other GlnR regulon proteins in strain M1152 indicates that their growth phase-dependent regulation has changed in M1152 as compared to M145. The *amtB-glnK-glnD* operon is regulated by both PhoP and GlnR (Rodríguez-García *et al.*, 2009, Wang *et al.*, 2012), but the levels of these proteins was unchanged upon Pi depletion.

## 5

Interestingly, the level of GlnK was significantly lower at all time points in M1152 as compared to M145 (Figure 5A, Table S2). GlnK is one of the nitrogen-sensory proteins that acts as a pleiotropic regulator for secondary metabolism and morphological differentiation (Waldvogel *et al.*, 2011, Perez-Redondo *et al.*, 2012). It is important to realise that the reduced level of GlnK in M1152 may have a major effect on the expression of BGCs that are regulated by this protein. GlnII and GlnK directly relate to nitrogen assimilation (Tiffert *et al.*, 2008). Their down-regulation in M1152 is an indication of reduced nitrogen assimilation. Conversely, the level of glutamate dehydrogenase (GdhA) was higher in M1152 at all time points as compared to M145. The protein profile changes along the time series could be separated by Pi depletion (Figure 5B). The second glutamate dehydrogenase encoded by the *S. coelicolor* genome, SCO2999, showed a five-fold increase from tp1 to tp6 in M145, corresponding to the onset of antibiotic production. In M1152 the increase was less pronounced, although SCO2999 levels were higher at early time points than in M145. Interestingly, the ammonia excretion was not significantly different between the two strains, which has been shown in the metabolism simulations build upon the same proteomics data (Sulheim *et al.*, 2020). This stable ammonia excretion can be explained by the observed balance between SCO2999 and GdhA. Combining these observations, we propose that the levels of the two glutamate dehydrogenases is inversely regulated in *S. coelicolor* and may compensate each other.

Another important GlnR-controlled process is that of the biosynthesis of ectoine and hydroxyectoine (Shao *et al.*, 2015, Bursy *et al.*, 2008). Ectoine compounds act as protectants against osmotic and heat stress (Bursy *et al.*, 2008), and the ectoine

BGC is also a likely target of control by GlnK (Waldvogel *et al.*, 2011). All Ect proteins (EctA to EctD) were detected at a higher level in M1152 than in M145 (Figure 5C, Table S2). The up-regulation of the biosynthesis of osmoprotectant compounds on stress resistance of M1152 needs to be investigated.



**Figure 6. Volcano plots showing comparisons between *S. coelicolor* M1152 and M145 at representative aligned time points.** The proteins from genes deleted in M1152 were excluded. Dashed lines showing the boundaries of significance threshold ( $p$ -value 0.05,  $\log_2$  fold change  $\pm 1$ ). Protein names of outstanding points or those have been discussed in the text were marked with an arrow. The proteins are coloured blue as they met the significance threshold. The higher average protein level of the test and control samples is transformed to the size of each dot.

### ScbR and ScbA are up-regulated in *S. coelicolor* M1152

To get a full view of the differences caused by the genome modifications in *S. coelicolor* M1152, the aligned corresponding samples were compared between

M1152 and M145 (Figure 6, Table S2). The most pronounced protein level differences were seen at tp1, corresponding to the time just before phosphate depletion, with 171 DEPs. The number of DEPs then dropped to values between 126 (time point 2) and 80 (time point 6). Interestingly, at tp1, ScbR and ScbA were among the most up-regulated proteins (Figure 6). These two proteins play a major role in  $\gamma$ -butyrolactone (GBL) mediated auto-regulation of antibiotic production and development (Takano *et al.*, 2001). ScbA synthesizes the GBL Scb1, while ScbR binds Scb1 and acts as a ScbA-responsive activator (Hsiao *et al.*, 2007). GBL-activated ScbR is a repressor of many developmental pathways including antibiotic production and development (Li *et al.*, 2015). The up-regulation of ScbR and ScbA is consistent with previous study showing that in AlaMM (minimal medium supplemented with alanine as nitrogen source), *S. coelicolor* M1152 produces more GBLs as compared to its parent strain M145 (Sidda *et al.*, 2016). It is known that ScbA is also controlled by ScbR2 (Li *et al.*, 2015), a gene that is lacking in M1152 because it is part of the *cpk* cluster that was deleted from its genome. Thus, the over-presence of ScbA is a likely result of *scbR2* deletion, which might relieve the suppression of *scbA* (Gomez-Escribano & Bibb, 2011, Li *et al.*, 2015). Further quantitative studies are needed to dissect the actual mechanism of ScbA, ScbR, ScbR2 and possible other participants. However, the up-regulation of ScbR should definitely be taken into account when expressing heterologous BGCs that are subject to control (repression) by ScbR.

## Conclusions

In this work, the protein profiles in background-reduced host *S. coelicolor* M1152 - which lacks four major BGCs and has a mutation in *rpoB* - were compared to those of its parent *S. coelicolor* M145 in samples obtained from batch fermentations. We found that upon Pi depletion, part of the PhoP regulon was up-regulated, while the rest remained unchanged. Interestingly, half of the up-regulated proteins showed delayed response in M1152, which might result in reduced response in Pi depletion, and thus affects growth. Worth noticing is also the major differences shown for proteins of the GlnR regulon. Most of the GlnR regulon proteins were not affected by the Pi depletion induced by the metabolic switch. However, we noticed that some of these proteins showed significant differences between M1152 and M145. In particular, the down-regulation at early time points was not seen in M1152, indicating changes in growth phase-dependent control. We also observed significant reduction in the level of GlnK in M1152 as compared to that in M145, which may be an additional factor that may be relevant for heterologous expression of BGCs. The strong induction in ectoine biosynthesis in strain M1152 may reflect of elevated osmotic stress. Finally, the ScbR/ScbA regulatory system was highly up-regulated in M1152, likely as a result of the absence of *scbR2*, which alters the global control of BGC expression.

In summary, we have systematically compared the global protein profiles in *S. coelicolor* M145 and its derivative M1152 in a well-defined fermentation system. This provides comprehensive insights into the changes in growth and protein profiles that are caused by the genomic modifications in *S. coelicolor* M1152. Our proteomics data, connected to the extant metabolic and transcriptome data, provide detailed insights into the effects of rational strain design of *S. coelicolor* aimed at obtaining optimised platforms for the expression of heterologous BGCs. At the same time, these studies will also form useful reference datasets for future 'omics studies on the model organism *S. coelicolor*.

## Material and methods

### Strains, cultivation conditions, sampling procedures, and analyses of media components

Experiments were performed using *S. coelicolor* A3(2) M145 (Kieser *et al.*, 2000) and its derivative M1152 (Gomez-Escribano & Bibb, 2011). M1152 lacks the four major biosynthetic gene clusters for actinorhodin (ACT), undecylprodigiosin (RED), coelimycin (CPK), and calcium-dependent antibiotic (CDA), and carrying the pleiotropic, previously described antibiotic production enhancing mutation *rpoB*[C1298T] (Hu & Ochi, 2001). Both strains were kindly provided by Mervyn Bibb (John Innes Centre, Norwich, UK).

TriPLICATE cultivations of the strains were performed based on germinated spore inocula in 1.8 L phosphate-limited medium SSBM-P, applying all routines of the optimised submerged batch fermentation strategy for *S. coelicolor* established and described before (Wentzel *et al.*, 2012a). Spores were obtained by cultivation on soy flour-mannitol (SFM) agar plates (Kieser *et al.*, 2000), harvested and suspended in 20% (v/v) glycerol, following by storage at -80°C.  $10^9$  CFU of spores were germinated and inoculated into 50 mL  $2 \times$  YT medium (van Wezel *et al.*, 2006). An even dispersion of the germinated spores was achieved by vortex mixing (30 s), ensuring comparable inocula among biological replicas. Each fermenter (1.8 L starting volume culture medium in 3 L Applikon fermenters) was inoculated with 4.5 mL germinated spore suspension (corresponding to  $9 \times 10^8$  CFU). Phosphate-limited medium SSBM-P (Nieselt *et al.*, 2010) was used, containing Na-glutamate, 55.2 g·L<sup>-1</sup>; D-glucose, 40 g·L<sup>-1</sup>; MgSO<sub>4</sub>, 2.0 mM; phosphate, 4.6 mM; supplemented minimal medium trace element solution SMM-TE (Claessen *et al.*, 2003), 8 mL·L<sup>-1</sup> and TMS1, 5.6 mL·L<sup>-1</sup>. TMS1 consisted of FeSO<sub>4</sub> · 7H<sub>2</sub>O, 5 g·L<sup>-1</sup>; CuSO<sub>4</sub> · 5H<sub>2</sub>O, 390 mg·L<sup>-1</sup>; ZnSO<sub>4</sub> · 7H<sub>2</sub>O, 440 mg·L<sup>-1</sup>; MnSO<sub>4</sub> · H<sub>2</sub>O, 150 mg·L<sup>-1</sup>; Na<sub>2</sub>MoO<sub>4</sub> · 2H<sub>2</sub>O, 10 mg·L<sup>-1</sup>; CoCl<sub>2</sub> · 6H<sub>2</sub>O, 20 mg·L<sup>-1</sup>, and HCl, 50 mL·L<sup>-1</sup>. Clerol FBA 622 fermentation defoamer (Diamond Shamrock, Pennsylvania, U.S.) was added to the growth medium before inoculation. Throughout fermentations, pH 7.0 was maintained constant by automatic addition of 2 M HCl. Dissolved oxygen levels were maintained at a minimum of 50% by automatic adjustment of the stirrer speed (minimal agitation 325 rpm). The aeration rate was constant 0.5 L·(L × min)<sup>-1</sup> sterile air. Dissolved oxygen, agitation speed and carbon dioxide evolution rate were measured and logged on-line, while samples for the determination of cell dry weight, levels of growth medium components and secondary metabolites concentrations, as well as for transcriptome and proteome analysis were withdrawn throughout the fermentation trials. For proteome analysis, 5 mL samples were taken at 21, 29, 33, 37, 41, 45, 49, 53 hours after inoculation of strain M145, and at 33, 41, 45, 49, 53, 57, 61, 65 hours after inoculation of strain M1152. Samples of bacteria culture were centrifuged (3200 × g, 5 min, 4°C), and the resulting cell pellets frozen rapidly at -80°C until further processing.

## Proteomics and data processing

### *Sample preparation and NanoUPLC-MS analysis*

Samples were prepared essentially as described before (Gubbens *et al.*, 2012). Briefly, mycelia for proteome analysis were resuspended in 100  $\mu$ L lysis buffer (4% SDS, 100 mM Tris-HCl pH 7.6, 50 mM EDTA) and disrupted by sonication. Total protein was precipitated using the chloroform-methanol method (Wessel & Flugge, 1984), and the proteins dissolved in 0.1% RapiGest SF surfactant (Waters, Massachusetts, U.S.) at 95°C. The protein concentration was measured at this stage using BCA method. Protein samples were then reduced by adding 5 mM DTT and incubate at 60°C for 30 min, followed by thiol group protection with 21.6 mM iodoacetamide incubation at room temperature in dark for 30 min. Then 0.1  $\mu$ g trypsin (recombinant, proteomics grade, Roche, Bavaria, Germany) per 10  $\mu$ g protein was added, and samples were digested at 37°C overnight. After digestion, trifluoroacetic acid was added to 0.5% and samples were incubated at 37°C for 30 min followed by centrifugation to degrade and remove RapiGest SF. Peptide solution containing 8  $\mu$ g peptide was then cleaned and desalted using STAGE-Tipping (Rappsilber *et al.*, 2007). Briefly, 8  $\mu$ g of peptide was loaded on a conditioned StageTip with 2 pieces of 1 mm diameter SDB-XC plug (Empore, Minnesota, U.S.), washed twice with 0.5% formic acid solution, and eluted with elution solution (80% acetonitrile, 0.5% formic acid). Acetonitrile was then evaporated in a SpeedVac. Final peptide concentration was adjusted to 40 ng· $\mu$ L<sup>-1</sup> using sample solution (3% acetonitrile, 0.5% formic acid) for analysis.

200 ng digested peptide was injected and analysed by reversed-phase liquid chromatography on a nanoAcquity UPLC system (Waters, Massachusetts, U.S.) equipped with HSS-T3 C18 1.8  $\mu$ m, 75  $\mu$ m  $\times$  250 mm column (Waters, Massachusetts, U.S.). A gradient from 1% to 40% acetonitrile in 110 min (ending with a brief regeneration step to 90% for 3 min) was applied. [Glu<sup>1</sup>]-fibrinopeptide B was used as lock mass compound and sampled every 30 s. Online MS/MS analysis was done using Synapt G2-Si HDMS mass spectrometer (Waters, Massachusetts, U.S.) with an UDMS<sup>E</sup> method set up as described in (Distler *et al.*, 2014).

### *Quantification and statistical analysis*

Raw data from all samples were first analysed using the vender software ProteinLynx Global SERVER (PLGS) version 3.0.3. Generally, mass spectrum data were generated using an MS<sup>E</sup> processing parameter with charge 2 lock mass 785.8426, and default energy thresholds. For protein identification, default workflow parameters except an additional acetyl in N-terminal variable modification were used. Reference protein database was downloaded from GenBank with the accession number NC\_003888.3. The resulted dataset was imported to ISOQuant version 1.8 (Distler *et al.*, 2014) for label-free quantification. Default high identification parameters were used in the quantification process.

TOP3 quantification was filtered to remove identifications meet these two criteria: 1. identified in lower than 70% of samples of each strain and 2. sum of TOP3 value less than  $1 \times 10^5$ . Cleaned quantification data was further subjected to DESeq2 package version 1.22.2 (Love *et al.*, 2014) and PCA was conducted after variance stabilizing transformation (vst) of normalized data.

#### *Proteome data availability*

The proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol *et al.*, 2018) partner repository with the dataset identifier PXD013178 and 10.6019/PXD013178.

## **Acknowledgments**

We would like to acknowledge Bogdan I. Florea of Leiden University, Leiden, The Netherlands, for running and for monitoring the proteome measurements, and the bio-organic synthesis group at Leiden University for providing the opportunity to use their instrumentation. This study was conducted in the frame of ERA-net for Applied Systems Biology (ERA-SysAPP) project SYSTERACT and the project INBioPharm of the Centre for Digital Live Norway

# A novel nucleoid-associated protein specific to Actinobacteria that binds to GATC sequences

Chao Du, Joost Willemse, Victor J. Carrion, and Gilles P. van Wezel

Manuscript in preparation



## Abstract

Nucleoid-associated proteins (NAPs) are known to affect the development of *Streptomyces*, which is a direct regulating factor in antibiotic production. SCO1839 is a small DNA binding protein which was previously suggested as a novel NAP in *Streptomyces coelicolor*, but no function or binding site was attributed to the protein. Here, we show that SCO1839 is an NAP that plays a role in the control of morphogenesis and antibiotic production in *S. coelicolor*. SCO1839 did not show significant matches against the Pfam database and thus likely represents a novel protein family. Chromatin immunoprecipitation sequencing (ChIP-Seq) showed that SCO1839 has a very large number of binding sites, with one or more sites in more than 2500 distinct genomic locations. ChIP-Seq analysis showed that SCO1839 binds to a short palindromic sequence (GATC), with preferable two A/T bases on one side. Interestingly, by far the largest difference in enrichment between 25 h and 48 h was seen for SCO1311, which encodes a tRNA editing enzyme, and SCOt32, which specifies the tRNA that recognises the rare leucine codon CUA. In both cases, enrichment was much higher in 25 h than in 48 h samples. Disruption of SCO1839 resulted in changes of development speed, colony morphology, and spore shape. The effect of SCO1839 on antibiotic production may in part be caused by the interaction between SCO1839 and the global transcriptional regulator AtrA, which activates antibiotic production. In conclusion, our results highlight the role of SCO1839 as DNA binding protein, we also proved that this protein is involved in the morphological development and antibiotic production of *S. coelicolor*.

## Introduction

Streptomycetes are filamentous soil bacteria with a complex life cycle, which are well known for their ability to produce various kinds of antibiotics. Streptomycetes are a major source of clinical drugs (Barka *et al.*, 2016, Hopwood, 2007, Bérdy, 2005). The life cycle of *Streptomyces* starts with the germination of a spore that grows out to form vegetative hyphae. Via tip extension and branching, this results in the formation of a dense mycelial network (Barka *et al.*, 2016, Chater & Losick, 1997). When the environmental situation requires sporulation, for example due to nutrient starvation, streptomycetes start their reproductive growth phase by developing aerial hyphae, which eventually differentiate into chains of unigenomic spores (Claessen *et al.*, 2014, Flärdh & Buttner, 2009). The production of antibiotics temporally correlates to the developmental growth phase (Bibb, 2005, van der Heul *et al.*, 2018). The complexity of the underlying regulatory networks is underlined by the fact that the *Streptomyces coelicolor* genome encodes some 900 regulatory proteins, of which only a fraction has been functionally characterized (Bentley *et al.*, 2002). Many of these are involved in the control of development and antibiotic production, such as the *bld* and *whi* genes that are responsible for the control of aerial hyphae formation and sporulation, respectively, and global regulatory genes such as *afsR*, *dasR* and *atrA* that pleiotropically control antibiotic production (van der Heul *et al.*, 2018).

The control of chromosome structure is also an important regulating factor of gene expression. In bacteria, the organization of chromosome structure is mediated by a diverse group of proteins referred to collectively as nucleoid-associated proteins (NAPs) (Dillon & Dorman, 2010). These are generally small DNA binding proteins involved in processes such as controlling gene expression, nucleoid structure or repair. Well-known NAPs in *Streptomyces* include Lsr2, HupA, HupS, and sIHF. Lsr2 binds non-specifically to AT-rich sequences and can globally repress gene expression (Gehrke *et al.*, 2019). HupA and HupS are homologs of HU (for histone-like factor U) proteins, which are differentially regulated depending on the developmental growth phase (Salerno *et al.*, 2009). sIHF is one of the basic architectural element conserved in many actinobacteria and is able to influence the regulation of secondary metabolism and cell development (Yang *et al.*, 2012). IHF binds a well conserved nucleotide sequence, while HU binds to random DNA sequences (Swinger & Rice, 2004). A detailed proteomic survey identified 24 proteins with NAP-like properties, namely the known Lsr2, HupA, HupS and sIHF and 20 yet unidentified proteins (Bradshaw *et al.*, 2013). Although the functions of many NAPs are still not clear, but there are certainly some have pervasive influences on the transcriptome, such as BldC (Bush *et al.*, 2019, Dorman *et al.*, 2020). One of the putative NAPs was SCO1839, a small protein that was previously identified via pull-down assays as associated with the promoter region of *ssgR* (Kim *et al.*, 2015). SsgR is the transcriptional activator of the cell division

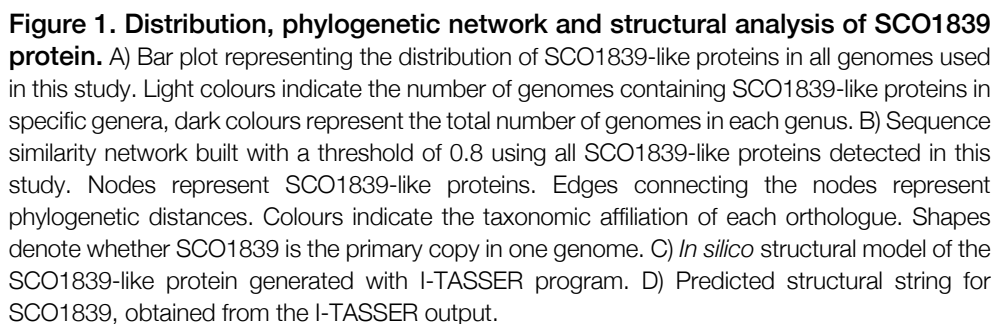
activator gene *ssgA* (Traag *et al.*, 2004). SsgA and its paralogue SsgB are both required for sporulation (Keijser *et al.*, 2003, van Wezel *et al.*, 2000a), and together coordinate the onset of sporulation-specific cell division in *Streptomyces*, whereby SsgB directly recruits the cell division scaffold protein FtsZ to the future sites of septation (Willemse *et al.*, 2011). The association of SCO1839 with the *ssgR* promoter thus implies a possible involvement in the control of sporulation. In this study, we show that SCO1839 represents a novel family of small DNA binding proteins, which plays a role in the regulation of *Streptomyces* development and antibiotic production. SCO1839 is specific to the Actinobacteria, with an HTH DNA binding motif containing three helices. Deletion of SCO1839 resulted in a significant change in colony phenotype. Chromatin immunoprecipitation coupled with massively parallel DNA sequencing (ChIP-Seq) showed the genome binding sites and a specific binding motif centred around the palindromic sequence GATC.

## Results and discussion

### SCO1839 is a small NAP specific for Actinobacteria

SCO1839 is a small protein of 73 amino acids (7.6 kDa) and a predicted isoelectric point (pI) of 10.53, indicative of an alkaline protein. A Pfam sequence search (El-Gebali *et al.*, 2018) did not yield any significant matches to known protein families, strongly suggesting that SCO1839 is the first member of a new protein family. Previously, an affinity capture assay using the promoter region of the cell division regulatory gene *ssgR* as a bait pulled down SCO1839 (Kim *et al.*, 2015). SCO1839 was also identified as one of the most abundant possible NAP (nucleoid-associated protein) by Bradshaw and colleagues (Bradshaw *et al.*, 2013).

To obtain more insights into the distribution and phylogeny of SCO1839, a conserved Hidden Markov Model (HMM) domain was constructed using all SCO1839-like proteins from *Streptomyces* species. Consequently, an HMM search against all available bacteria full genomes in the database was performed. No hits were found outside the order Actinomycetales, strongly suggesting that SCO1839 is an Actinobacteria-specific protein (Figure 1A and B). Eight main clusters of similar groups of homologs were found. The largest cluster mainly consists SCO1839 homologs from *Streptomyces*, *Amycolatopsis*, *Pseudonocardia*, *Frankia*, and *Actinomadura*. Other major clusters including a cluster of *Nocardia* and *Rhodococcus*, *Micromonospora* and *Salinispora*, *Geodermatophilus* and *Blastococcus*. *Rhodococcus* itself forms two separate clusters. Interestingly, 24 *Actinomadura* species and 38 *Streptomyces* species contain two paralogues of SCO1839, which divided into two additional clusters. For most genera, more than 90% of the species contains at least one copy of SCO1839 like proteins. *Rhodococcus* stands out as only 28.7% (92 out of 321) of the genomes containing a SCO1839-family protein (Figure 1A), and these proteins divide into three distinct clusters. This could be due to the fact that it is the only nonsporulating genus among all listed genera. The fact that the nonsporulating genus *Rhodococcus* stands out clearly from the sporulating genera, in combination with the low frequency of occurrence, suggests that perhaps SCO1839 has diversified in the non-sporulating *Rhodococcus* and also that SCO1839 may primarily be sporulation-specific.



*In silico* structural modelling of SCO1839 was performed using the I-TASSER server (Yang *et al.*, 2015), revealing a putative single DNA binding helix-turn-helix (HTH) motif, in the form of a tri-helical structure (Figure 1C and D). No homology was seen to any other known transcriptional regulator family from bacteria. The regions flanking the residues belonging to the HTH motif are very short. Nine other protein structures found in PDB (Protein data bank, rcsb.org, Berman *et al.*, 2000) share structural analogy and contained DNA binding motifs that structurally resemble SCO1839. These proteins come from a wide range of organisms and functions, including DNA helicase from archaeon bacteria *Pyrococcus furosus* (PDB structure ID: 2ZJ8), *E. coli* (2VA8, 2P6R), human (5AGA); ribosome protein from yeast (5MRC), cell division cycle protein from human (2DIN), terminator binding protein from yeast (5EYB), regulator from *Staphylococcus aureus* (2R0Q), and a tRNA synthetase from *Archaeoglobus fulgidus* (2ZTG). However, none of the analogous proteins were as small as SCO1839 and neither contained only one DNA binding motif. To the best of our knowledge, no other bacterial protein with similar structure has been reported before. Therefore, we propose that the SCO1839-like proteins form a new family of bacterial DNA binding proteins.

### **Deletion of SCO1839 accelerates sporulation of *S. coelicolor***

To obtain insights into the possible role of SCO1839 in the life cycle of *S. coelicolor*, a knock-out mutant was generated using a strategy published previously (Świątek *et al.*, 2012), whereby the +1 to +207 region of the gene was replaced by the apramycin resistance cassette flanked by *loxP* sites, followed by deletion of the cassette using Cre recombinase so as to avoid polar effects. To genetically complement the mutant and see if the wild-type phenotype would be restored, the -565/+228 region of SCO1839 was amplified from the *S. coelicolor* chromosome and cloned into pHJL401, a shuttle vector that is useful for genetic complementation due to its low copy number in streptomycetes (van Wezel *et al.*, 2000b). To also analyse the effect of enhanced expression of SCO1839, a second strain was constructed using CRISPR-Cas9 wherein the native promoter of SCO1839 (-157 to +4, start codon switched to ATG) was replaced by the strong constitutive *ermE* promoter (Bibb *et al.*, 1985). For details see the Materials and Methods section.

When grown on MM agar plates with mannitol as the carbon source, colonies of the SCO1839 mutant appeared darker in colour and more irregularly shaped as compared with the parental strain *S. coelicolor* M145 (Figure 2A, 6). However, on soya flour mannitol (SFM) agar plates, the mutant colonies showed a smoother surface, devoid of the ridges that are typical of wild-type colonies (Figure 2A, 8). Interestingly, the SCO1839 mutant failed to produce blue-pigmented antibiotic actinorhodin on SFM agar when colonies were grown at higher densities (Figure 2B, right). The SCO1839 mutant phenotype was restored in the complemented strain on minimum medium and SFM medium when growing as single colonies

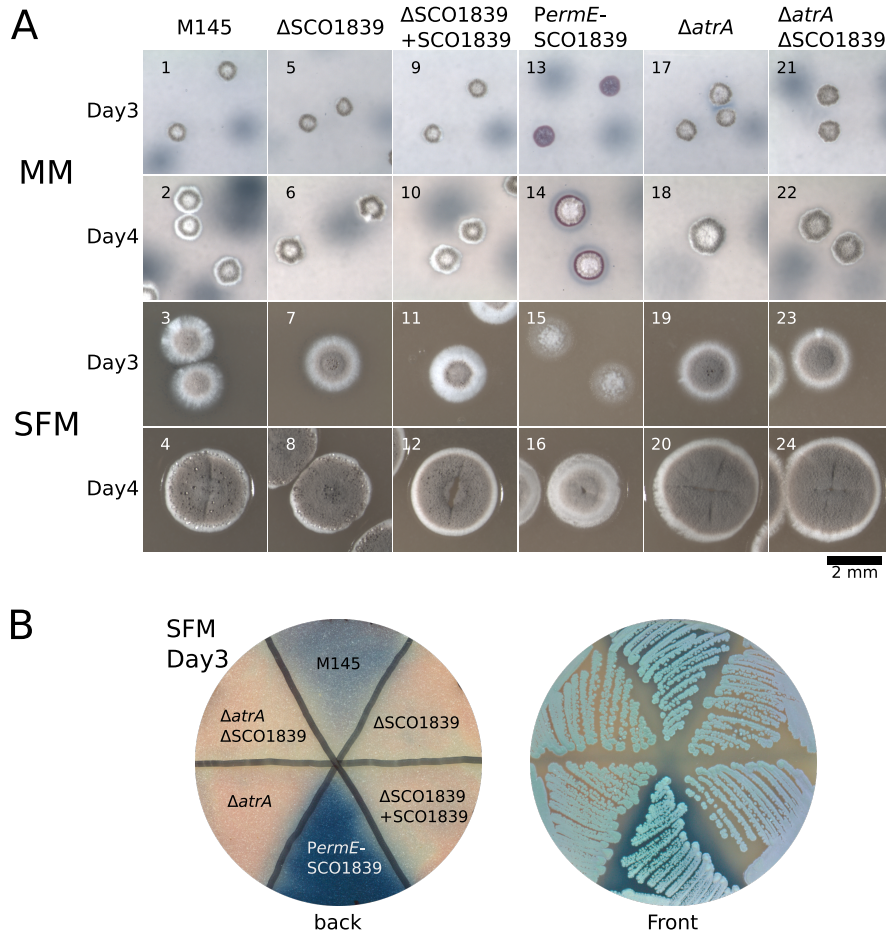
(Figure 2A, 9-12). However, at higher density, actinorhodin production on SFM was not restored (Figure 2B, right). The strain expressing SCO1839 from *ermE* promoter produced large amounts of undecylprodigiosin on MM agar (Figure 2A, 13, 14) and showed a reduced growth on SFM agar (Figure 2A, 15, 16).

In order to more precisely monitor the differences in the timing of development between SCO1839 mutants and their parent, time-lapse imaging was performed on confluent mycelial lawns. This method allows monitoring multiple morphological characteristics and in particular quantify differences in timing of the individual developmental stages, based on pigmentation of the mycelia. When aerial hyphae are formed, intensity increases due to the white pigmentation, while brightness decreases again when grey-pigmented spores are produced (see Material and Methods section for details). Deletion of SCO1839 resulted in acceleration of development by 2-5 h (Figure 3A). 54 h after inoculation, the light intensity of the SCO1839 mutant again increased (Figure 3A). This may be due to premature germination and renewed growth. The complemented mutant had a far less pronounced acceleration of its life cycle, confirming that this phenomenon was primarily due to the deletion of SCO1839. Conversely, the enhanced and constitutive expression of SCO1839 delayed sporulation by approximately 18 h (Figure 3A and B). Taken together, this strongly suggests that the expression of SCO1839 correlates to the timing of sporulation.

## 6

Table 1. Statistics of spore length to width ratio

Strain	n	Median	M145 median	Mann-Whitney <i>U</i> test of difference		Variance	M145 variance	Levene's test of equal variance	
				<i>U</i>	<i>p</i> -value			<i>W</i>	<i>p</i> -value
ΔSCO1839	332	1.69		$1.2 \times 10^5$	$9.3 \times 10^{-16}$	0.145		49.51	$3.93 \times 10^{-12}$
ΔSCO1839 + SCO1839	556	1.50		$1.6 \times 10^5$	0.40	0.099		13.49	$2.52 \times 10^{-4}$
pErmE- SCO1839	441	1.60	1.49 (n = 560)	$1.5 \times 10^5$	$1.4 \times 10^{-7}$	0.125	0.069 (n = 560)	34.89	$4.78 \times 10^{-9}$
ΔSCO1839 ΔAtrA	565	1.52		$1.6 \times 10^5$	0.20	0.114		13.01	$3.23 \times 10^{-4}$
ΔAtrA	561	1.48		$1.5 \times 10^5$	0.09	0.076		2.01	0.157

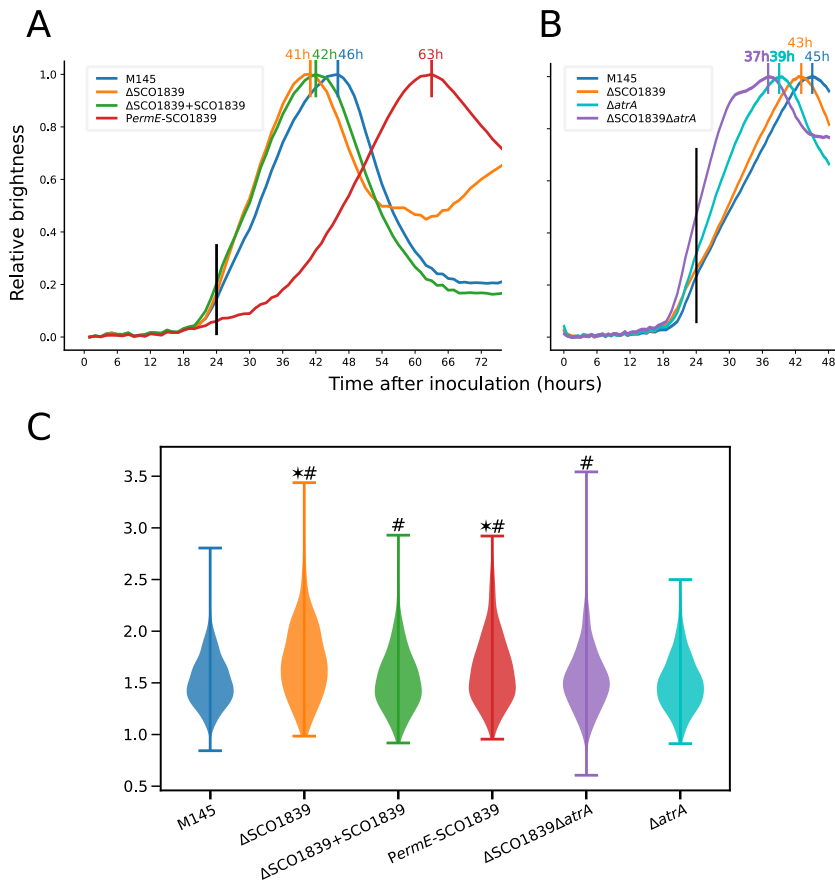


**Figure 2. Colony morphology and pigment production of *S. coelicolor* M145 and mutants.** A) *S. coelicolor* M145 and its mutants  $\Delta$ SCO1839,  $\Delta$ *atrA* and  $\Delta$ *atrA* $\Delta$ SCO1839, as well as the M145 + *PermE*-SCO1839 and  $\Delta$ SCO1839 + *PermE*-SCO1839. Strains were grown as single colonies on MM agar supplemented with mannitol or on SFM agar at 30°C during 3 and 4 days. All pictures were taken at the same scale. B) *S. coelicolor* M145 and derivatives compared on SFM medium. Note the enhanced Act production by M145 + *PermE*-SCO1839.

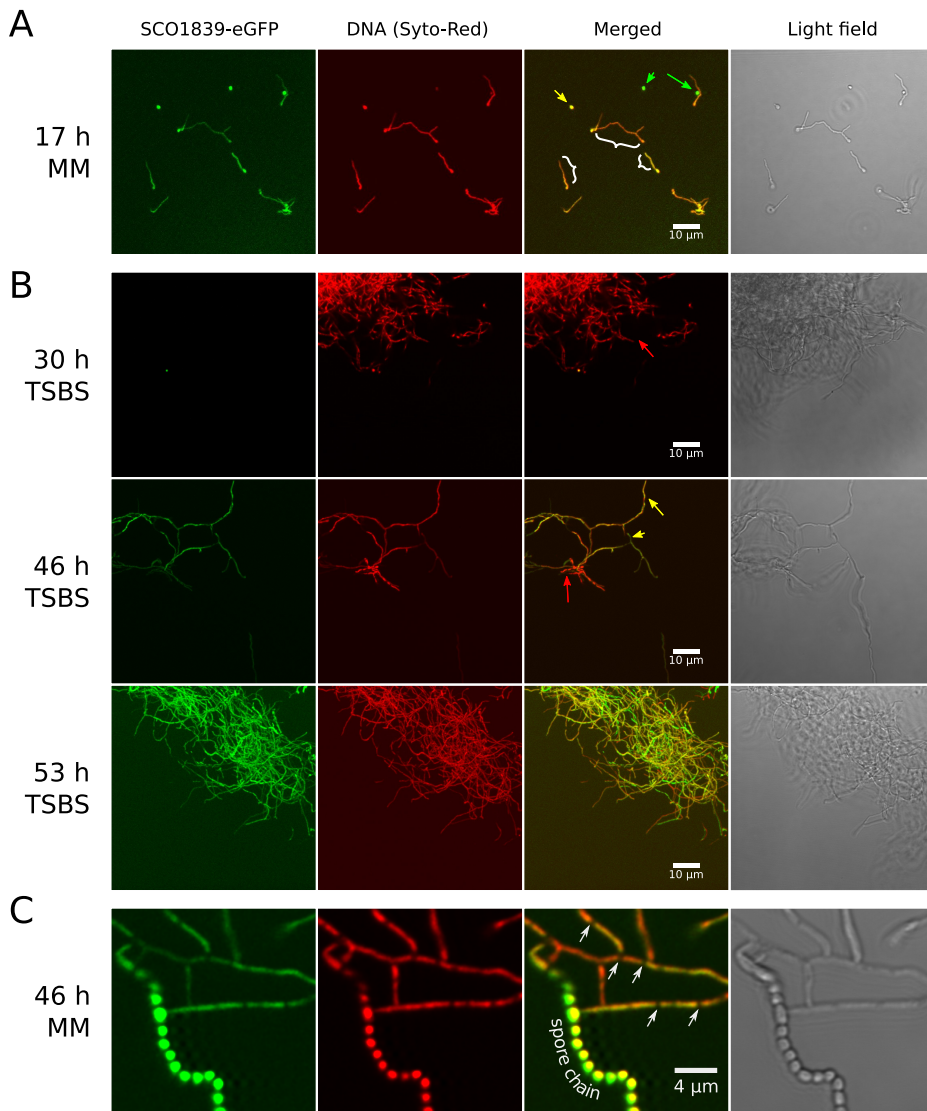
Closer examination of the spores by scanning electron microscopy (SEM) revealed that spores were significantly enlarged in both the deletion mutant and the strain with enhanced expression of SCO1839 (Mann-Whitney U test, p-value < 0.01, Table 1, Figure 3 C). The complemented strain produced spores with a normal size. All engineered strains ( $\Delta$ SCO1839,  $\Delta$ SCO1839+SCO1839, *PermE*-SCO1839), including the complementation mutant, had a variance in the distribution of spore length to width ratio that differed significantly from that of the parent strain (Levene's test, p-value < 0.01), with the SCO1839 deletion mutant showing the



largest variance. Complementation of the mutant reduced the large variance to a level closest to the wild-type. Thus, deletion of *SCO1839* alters the timing of sporulation and also leads to significant alteration of the spore morphology.



**Figure 3. Growth and spore shape analysis of *SCO1839* mutants.** A) and B) scanner measurements of confluent plate brightness of *S. coelicolor* M145 and mutants ( $\Delta$ SCO1839,  $\Delta$ SCO1839+SCO1839, *PermE*-SCO1839,  $\Delta$ atrA and  $\Delta$ atrA $\Delta$ SCO1839) grown as confluent lawns. The X-axis represents the time after inoculation and the Y-axis the normalized brightness measured from time lapse scan pictures. Mutants shown as different colours, curve peaks are marked on top. Black vertical lines illustrate 24 h time point. C) Violin plot showing the distributions of the ratio of spore length and width as measured using scanning electron microscope images. Each strain was grown on MM agar for 5 days before imaging. \* with significantly ( $p < 0.01$ ) different length to width ratio compared to *S. coelicolor* M145 (Mann-Whitney U test), # with significantly ( $p < 0.01$ ) different variance compared to *S. coelicolor* M145 (Levene's test).



**Figure 4. Confocal micrographs showing the expression of SCO19839-eGFP in different developmental stages and its co-localization with genomic DNA.** A) Germinating spores on MM agar supplemented with mannitol. Dormant spores are highlighted using green arrows, swelled spores by yellow arrows, and germ tubes with white curly brackets. B) Mycelia grown in liquid TSBS media. SCO19839 expressing mycelium is indicated by yellow arrows, not expressing mycelium is indicated by red arrows. C) Mycelium grown on MM agar. SCO19839-eGFP and DNA were both absent at some location indicated by arrows.

## **SCO1839 expression after early development stage and is highly abundant in spores**

To determine the location and timing of SCO1839 expression, a mutant strain that expresses SCO1839-eGFP fusion protein was constructed. For this, the gene for eGFP was fused to the C-terminal of SCO1839 using CRISPR-Cas9. Confocal images with DNA staining with Syto-Red showed that SCO1839-eGFP protein was strongly expressed in spores (Figure 4A). The fluorescence disappeared once the vegetative hyphae extended after germination, indicating that SCO1839 is not expressed during early stages of growth. In liquid TSBS, SCO1839-eGFP was only expressed approximately 45 h after inoculation and became stronger thereafter (Figure 4B). Transcriptional analysis using quantitative PCR (qPCR) confirmed a two-fold increase in SCO1839 expression at 48 h as compared to 24 h (Figure 7C). The qPCR analysis supported the results of the time series transcriptomics data available in the NCBI GEO database (Barrett *et al.*, 2012). Specifically, we compared our results with available data that include early developmental stages and that had been performed in different media (Gehrke *et al.*, 2019, Castro-Melchor *et al.*, 2010, Nieselt *et al.*, 2010). These authors showed that transcription of SCO1839 is low at early developmental stages, and increases as development progresses, in a medium-independent manner (Gehrke *et al.*, 2019, Castro-Melchor *et al.*, 2010, Nieselt *et al.*, 2010). This indicates that the timing of SCO1839 depends on the growth conditions and the developmental stage.

## 6

## **SCO1839 has thousands of DNA binding sites and specifically recognizes a small motif with GATC as core sequence**

To obtain insights into the genome-wide DNA binding capacity of SCO1839, ChIP-Seq analysis was performed after 25 h (vegetative growth) and 48 h (sporulation). This should reveal all binding sites of SCO1839 on the *S. coelicolor* chromosome. For this purpose, the original copy of SCO1839 on the genome was fused with a sequence encoding a triple FLAG tag in its C-terminal using CRISPR-Cas9 (see Materials and Methods section). The strain had a phenotype that was highly similar to that of the parent (Figure S1). The ChIP-Seq results showed a wide distribution of SCO1839 binding sites (Figure 5). Half of the binding sites that showed a high fold enrichment (50% of the sites at 25 and 38 h) colocalized with low G+C content regions. In total, 2825 and 2919 binding regions were identified for the vegetative and sporulation growing stage samples, respectively. Interestingly, there was a very high overlap (> 90%, 2402) between the SCO1839 binding regions in both stages. This not only shows that the binding sites of SCO1839 is largely growth phase-independent, but also that the experiments were highly reproducible. This is highlighted by the very similar binding patterns in Figure 5A.

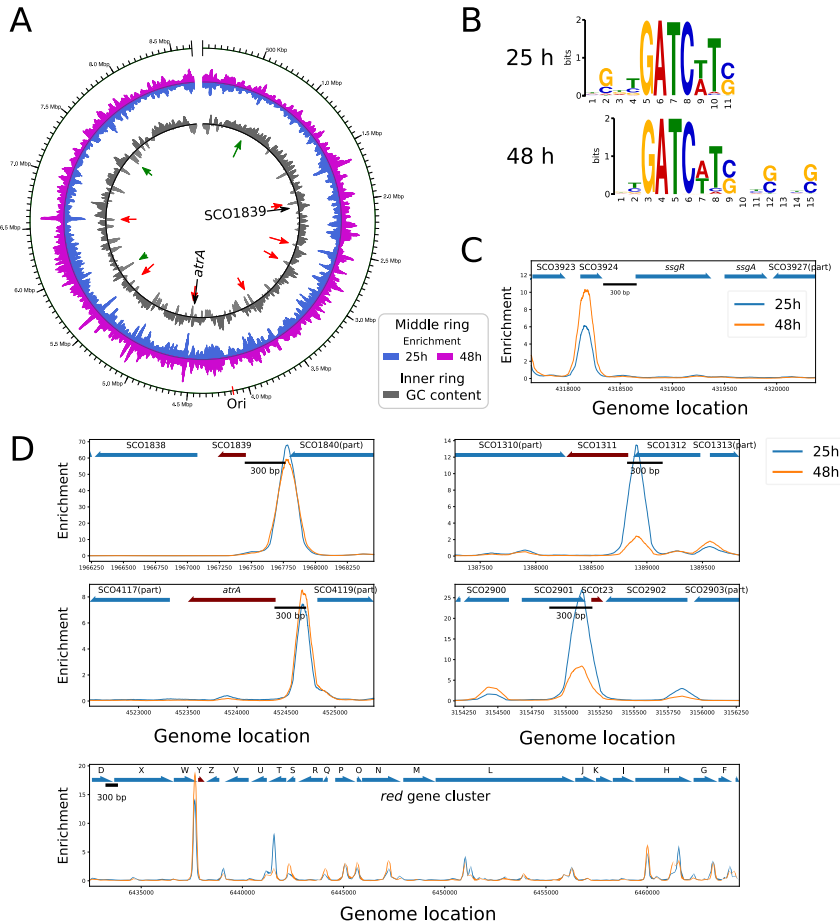
To obtain the consensus sequence required by SCO1839 binding, the binding regions from ChIP-Seq results were extracted and modelled using MACS2 and

MEME-ChIP (Zhang *et al.*, 2008, Machanick & Bailey, 2011). As a result, GATC(A/T)T was found to be the most significant binding motif (Figure 5B). The most conserved binding core was GATC, which is a palindrome known as recognition site for DNA methylation (Barras & Marinus, 1989). The predicted motifs also showed G/C preference on the flanking region separated by two base pairs gaps (Figure 5B). It is important to note that virtually all binding regions that were identified as significant (> 99.8%) contained a GATC motif, and most (88.2% for 25 h, 84.0% for 48 h) contained the consensus sequence GATC(A/T)T. The *S. coelicolor* genome contains in total 6,500 GATC(A/T)T sequences, 64% and 68% were detected as binding regions in the 25 h and 48 h samples, respectively. Note that many binding sites have more than one copy of the motif. The *S. coelicolor* genome contains 43,535 GATC sequences, of which only 24% (25 h) and 30% (48 h) are in the identified SCO1839 binding regions. This indicates that the minimal binding motif is not the only factor that determines *in vivo* binding of SCO1839. Other determining factors might include yet undiscovered elements in the flanking regions, secondary genome structure, and/or the binding of other proteins to the same region.

The affinity of SCO1839 to the DNA binding motif was further tested *in vitro* using electrophoretic mobility shift assays (EMSA). The results showed that SCO1839 could bind to the DNA with one GATC motif with random flanking sequence, but the additional nucleotide (AT)T on one side increased the affinity of SCO1839 by around two-fold (Figure 6B). Furthermore, when four GATC motifs were present in the short 50 bp DNA fragment, simultaneous binding of three SCO1839 proteins was observed. Perhaps the two GATC sequences in close proximity were bound by a single SCO1839 protein due to steric hindrance. Taken together, SCO1839 showed good binding to a motif centred around GATC, and this is probably the only requirement for SCO1839 to bind *in vitro*.

The GATC motif is the target sequence for deoxyadenosine (DAM) methylase, which is essential for DNA mismatch repair in *E. coli* (Barras & Marinus, 1989). Another typical DNA methylation is on the second deoxycytosine of the sequence CCTGG (DCM) (May & Hattman, 1975). *S. coelicolor* lacks both methylation system and degrades methylated exogenous DNA (Flett *et al.*, 1997, Kieser & Hopwood, 1991). Which proteins are involved in the recognition and restriction system in *Streptomyces* remains unknown (Liu *et al.*, 2010, González-Cerón *et al.*, 2009). The fact that SCO1839 binds to GATC sequences could be correlated with the restriction system in *S. coelicolor*. Therefore, the transformation efficiency was tested for methylated DNA and nonmethylated DNA to  $\Delta$ SCO1839 strain. However, the result indicates the barrier still exists (data not shown). Next, we tested the affinity difference of SCO1839 to GATC and GA<sup>m</sup>TC in electrophoretic mobility shift assay (EMSA). The results showed an approximately two-fold decreased affinity of SCO1839 for methylated DNA (Figure 6D). As DNA is not

methyated in *S. coelicolor* (Flett *et al.*, 1997), the biological function of this difference if affinity cannot yet be explained. However, it is logical to assume that methylation causes some steric hindrance, which negatively affects the binding of SCO1839.



**Figure 5. SCO1839 protein binding sites analysis.** A) Genome-wide distribution of SCO1839 protein binding sites along the *S. coelicolor* genome. The outer ring shows the genome location; the middle ring shows the local average fold enrichment from ChIP-Seq analysis, 48 h sample oriented outwards, 25 h sample oriented inwards; the inner ring shows the local average G+C content. The G+C percentage above median is plotted outwards, below median inwards. Bin size for local averaging was 20,000 bp. SCO1839 and *atrA* have been indicated by black arrows. B) SCO1839 DNA binding motifs predicted by MEME-ChIP program. C) Enrichment level (compared to the corresponding total DNA control sample) around *ssgR*. D) Enrichment level of genome regions around genes highlighted in Table 2, with target genes coloured red. The *red* gene cluster is shown as a whole, with significant binding only upstream of *redY*. Note that the y-axes of the plots are not at the same scale.



relative to the translation start site (Figure 5C). This is in accordance with the previous observation that SCO1839 binds to the *ssgR* promoter region (Kim *et al.*, 2015).

Table 2. Genes strongly bound by SCO1839 at their promoter regions

Gene ID	Strand	Product	Binding width		Binding summit position**		Fold enrichment	
			25h	48h	25h	48h	25h	48h
SCO1311*	-	hypothetical protein with tRNA edit domain	271	ND <sup>†</sup>	-76	-72 <sup>†</sup>	13.41	2.42 <sup>†</sup>
SCO1839*	-	transcriptional regulator	547	530	-326	-324	67.75	58.18
SCO2077	-	DivIVA	251	300	-90	-91	17.19	13.48
SCO2081	-	YlmD	271	378	-348	-348	10.44	17.18
SCO2084	-	MurG (UDPdiphospho-muramoylpentapeptide beta-N-acetylglucosaminyltransferase)	280	313	-280	-273	13.22	13.41
SCO2086	-	MurD (UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase)	482	471	-90	-100	9.63	10.62
SCO2088	-	MurF (UDP-N-acetylmuramoylalanyl-D-glutamyl-2, 6-diaminopimelate- D-alanyl-alanyl ligase)	227	285	-5	-10	7.37	14.21
SCO3213	-	TrpG (anthranilate synthase component II)	247	299	-240	-236	7.85	10.69
SCO3224	-	ABC transporter ATP-binding protein	349	367	+33	+48	14.23	15.90
SCO3225	+	AbsA1 (two component sensor kinase)	<i>ab</i>	<i>ab</i>	-168	-183	<i>ab</i>	<i>ab</i>
SCO4118*	-	AtrA (TetR family transcriptional regulator)	228	255	-284	-289	7.42	8.23
SCO4702	+	RplC (50S ribosomal protein L3)	459	511	-34	-37	12.19	22.16
SCO4707	+	RplV (50S ribosomal protein L22)	290	388	-198	-196	15.80	25.17
SCO4713	+	RplX (50S ribosomal protein L24)	372	406	-7	-5	58.66	44.69
SCO4714	+	RplE (50S ribosomal protein L5)	<i>ab</i>	<i>ab</i>	-330	-328	<i>ab</i>	<i>ab</i>
SCO4717	+	RplF (50S ribosomal protein L6)	372	418	-324	-326	25.78	23.66
SCO4718	+	RplR (50S ribosomal protein L18)	252	641	+46	+45	10.27	19.35
SCO4721	+	RplO (50S ribosomal protein L15)	380	432	+22	+21	34.82	29.13
SCO4726	+	RpmJ (50S ribosomal protein L36)	310	761	-3	+21	6.69	15.32
SCO4727	+	RpsM (30S ribosomal protein S13)	<i>ab</i>	<i>ab</i>	-305	-282	<i>ab</i>	<i>ab</i>
SCO5880*	+	RedY (RedY protein)	286	347	-168	-161	14.59	18.85
SCOt02	-	tRNA Val (anticodon CAC)	509	559	-136	-138	56.47	55.65
SCOt17	+	tRNA Gly (anticodon UCC)	295	348	-325	-320	20.40	28.92
SCOt23*	+	tRNA Leu (anticodon UAG)	338	280	-78	-90	25.89	8.10
SCOt49	+	tRNA Thr (anticodon GGU)	261	294	-170	-168	15.21	13.95
SCOt50	+	tRNA Met (anticodon CAU)	<i>ab</i>	<i>ab</i>	-289	-287	<i>ab</i>	<i>ab</i>

\* Shown in Figure 5C

\*\* Relative to gene start site (+1)

<sup>†</sup> Peak not detected, local summit position and corresponding fold enrichment are shown

*ab* as above, shared binding region as the gene above

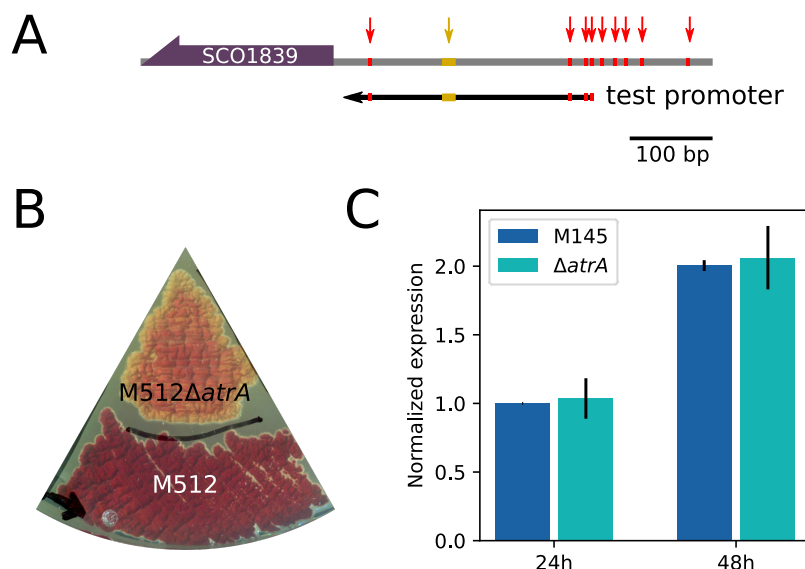
Interestingly, SCO1839 did not only bind to the promoter regions of five tRNA genes, but also to the promoter region of SCO1311, which has a tRNA-editing domain responsible for hydrolysing mis-acylated tRNA (coverage 68%, Pfam: PF04073). Importantly, the promoter region of SCO1311 and SCOt32 showed the largest difference in SCO1839 binding between 25 h and 48 h samples, with enrichment at 25 h more than three times higher than at 48 h (Figure 5D, Table 2). SCOt23 specifies the leucyl-tRNA with anticodon UAG, which is required for the translation of the very rare leucine codon CUA. The rarest codon in *S. coelicolor* is another leucine codon, namely UUA. The tRNA recognizing the UUA codon is specified by *bldA*, and the corresponding TTA codon occurs specifically in many genes involved in development and antibiotic production, making those genes *bldA* dependent (Leskiw *et al.*, 1991, Li *et al.*, 2007, Lawlor *et al.*, 1987, Chater & Chandra, 2008). The CTA leucine codon is also very rare, representing only 0.31% of all leucine codons in *S. coelicolor*. It will be interesting to see if SCOt23 also plays a role in the control of developmental gene expression, and what the role is of SCO1839 in the control of SCOt23 transcription.

Considering that over-expression of SCO1839 enhances undecylprodigiosin (Red) production on MM agar (Figure 2A) it is worth noticing that SCO1839 binds to the promoter region of *redY*, which is part of the BGC for Red (Figure 5D). However, since *redY* is not considered as a key factor in Red production (Cho *et al.*, 2008), we do not want to speculate too much on this topic. The connection (if any) between SCO1839 and Red biosynthesis needs to be investigated further.

### Interaction between the global regulators SCO1839 and AtrA

An interesting hit in the list of SCO1839 binding sites was *atrA* (Figure 7C). We previously experimentally validated that both SCO1839 and AtrA bind to the upstream region of the cell division activator gene *ssgR*, suggesting competition or cooperativity (Kim *et al.*, 2015). Furthermore, AtrA is required as transcriptional activator of the actinorhodin BGC (Uguru *et al.*, 2005), and control of *atrA* may therefore explain at least in part the altered antibiotic production in the SCO1839 mutant (Figure 2). To analyse the possible relationship between the two regulators further, we created *atrA* mutants for *S. coelicolor* M145 and its SCO1839 null mutant. For this, the -41 to +46 region relative to the *atrA* translational start site was removed using pCRISPR-Cas9. We then followed growth of the strains using time lapse imaging (Figure 3B). The *atrA* mutant and *atrA*/SCO1839 double mutant showed accelerated sporulation by 6 and 8 h, respectively, while the SCO1839 mutant developed only slightly earlier (around 2 h). The accelerated growth was also reflected by the larger colony sizes (Figure 2A, 20, 24).





**Figure 7. SCO1839 promoter activity assay.** A) Graphic representation of the promoter region of SCO1839. Red arrows point at the GATC positions; yellow arrow points at the predicted AtrA binding site. B) Promoter probing experiment showing *S. coelicolor* M512ΔatrA (top) and *S. coelicolor* M512 (bottom) harbouring a plasmid wherein *redD* is expressed from the SCO1839 promoter region (redness of the mycelium is a measure for the promoter activity). Patches streaked on R5 agar were grown for 4 days at 30°C. C) Normalized qPCR result showing SCO1839 expression in *S. coelicolor* M145 and its ΔatrA mutant on samples taken 24 h and 48 h after inoculation MM agar plates covered with a cellophane. Error bars represent the standard error of the mean derived from two biological replicates. All expression values of SCO1839 were normalized against the expression in M145 at 24 h.

Interestingly, in the SCO1839 promoter region we found an AtrA binding site using PREDetector (Tocquin *et al.*, 2016) with a high score of 13.3. This score is the same as for the confirmed AtrA binding site upstream of *actII-ORF4*, and this therefore suggests direct control of SCO1839 by AtrA. To test the possible transcriptional control of SCO1839 by AtrA, promoter probing experiments were performed based on the *redD* reporter system (van Wezel *et al.*, 2000b). For this, we used the -297 to -10 region relative to the SCO1839 translation start site (Figure 7A). The promoter-probe system is based on the *redD* gene, which activates the red-pigmented antibiotic undecylprodigiosin. When a promoter sequence is inserted in front of the promoterless *redD*, the gene will be transcribed, and the level of red pigment that is produced can be used as a measure of promoter activity. As hosts we used *S. coelicolor* M512 (which is *S. coelicolor* M145 ΔactII-ORF4 ΔredD and hence non-pigmented) and its *atrA* mutant. The results clearly show that the

SCO1839 promoter is less actively transcribed in the *atrA* mutant (Figure 7B). We therefore conclude that AtrA *trans*-activates SCO1839 transcription. To further validate this, we used quantitative reverse transcription PCR (q-RT-PCR) to analyse the transcription of SCO1839 in *S. coelicolor* M145 and its *atrA* mutant. Surprisingly, the expression of SCO1839 was slightly higher in the mutant at both 24 h and 48 h time points, which conflicts with the promoter probing results (Figure 7C). Thus, expression of SCO1839 likely depends on more than the tested region and involves multiple factors.

## Conclusions

In this study, we have shown that SCO1839 is a highly pleiotropic DNA binding protein that plays a role in the (timing of) development and antibiotic production of *S. coelicolor*. Strains in which SCO1839 had been deleted or over-expressed showed accelerated and delayed sporulation, respectively, suggesting that SCO1839 plays a role in the timing (and in particular delay) of development. Localization experiments of SCO1839 showed high concentrations of protein especially inside spores. In both *in-vivo* and *in-vitro* experiments revealed that SCO1839 binds specifically to GATC DNA motif with a preference of (A/T)T on one side. In addition, we have shown that the methylation of adenine in the GATC sequence reduced affinity of SCO1839 for its binding sites. SCO1839 binds to the promoter region of *atrA*, which correlates to the activation of actinorhodin production. In return, AtrA activates SCO1839 expression, which results in an interesting two-way feedback loop between these two important proteins.

## Materials and methods

### Bioinformatics

The SCO1839 from *S. coelicolor* was used as query in HMMER web server (Potter *et al.*, 2018) to obtain all SCO1839-like proteins from the database, resulting in 727 hits. Sequences with an E value < 0.01 (684 sequences) were selected to generate a Hidden Markov model (HMM) profile using HMMER suit (v. 3.1b2, Eddy, 2011). This profile was used to search against a custom database containing 146,856 genomes with all available bacteria genomes (access date Feb. 9, 2019). Hits with E value  $\leq 5.5 \times 10^{-9}$  (2,317 sequences) were aligned to the generated HMM profile using the hmalign tool from HMMER suit. Using the alignment, a network was built calculating the pairwise distance between all the detected SCO1839 proteins and the threshold for clustering was settled at 0.8. Network visualizations were constructed using Cytoscape (v. 3.7.1, Shannon *et al.*, 2003).

Four sequences of known AtrA binding motif (Table S2) was used as input for PREDetector (v. 3.1, Tocquin *et al.*, 2016) to find possible AtrA binding motif in *S. coelicolor* genome.

### Strains and growth conditions

All strains used in this study are listed in Table S3. *Escherichia coli* strain JM109 was used for routine cloning, *E. coli* ET12567 (MacNeil *et al.*, 1992) for preparing non-methylated DNA, ET12567 containing driver plasmid pUZ8002 (Paget *et al.*, 1999) was used in conjugation experiments for introducing DNA to *Streptomyces*. *E. coli* strains were grown in Luria broth at 37°C supplemented with the appropriate antibiotics (ampicillin, apramycin, kanamycin and/or chloramphenicol at 100, 50, 25 and 25  $\mu\text{g}\cdot\text{mL}^{-1}$ , respectively) depending on the vector used. *S. coelicolor* A3(2) M145 was the parent for all mutants. For promoter probing experiments, *S. coelicolor* M512 (M145  $\Delta\text{redD}$   $\Delta\text{actII-ORF4}$  (Floriano & Bibb, 1996)) was used. *Streptomyces* strains were grown on soya flour medium (SFM) for conjugation, SFM or MM agar medium supplemented with 0.5% mannitol for phenotype characterization, and R5 agar plates for protoplast regeneration and promoter probing. Solid cultures were grown in a 30°C incubator unless described specifically. For liquid cultures, approximately  $10^6$  spores were inoculated in 100 mL Erlenmeyer flask equipped with steel spring, containing 15 mL TSBS (tryptone soya broth sucrose) medium (Kieser *et al.*, 2000), and grown at 30°C with constant shaking at 180 rpm. Antibiotics used for screening *Streptomyces* transformants were apramycin and thiostrepton (20 and 10  $\mu\text{g}\cdot\text{mL}^{-1}$ , respectively).

### Plasmids and Constructs

Primers used for PCR and short double strand DNA fragment are listed in Table S4. PCR was preformed using Pfu DNA polymerase using standard protocol as described previously (Colson *et al.*, 2007). All plasmids and constructs described

in this study are summarized in Table S5. The constructs generated in this study were verified by sequencing performed in BaseClear (Leiden, The Netherlands).

The SCO1839 knock-out strategy was based on the unstable multi-copy vector pWHM3 (Vara *et al.*, 1989) as described previously (Zhang *et al.*, 2018). Roughly 1.5 Kb of upstream and downstream region of SCO1839 were amplified by PCR from M145 genome. The upstream region (relative to SCO1839 direction) was thereby cloned as a *Bam*HI-*Xba*I fragment, and the downstream region was cloned as a *Xba*I-*Hind*III fragment. The apramycin resistance cassette *aac*(3)IV with flanking *loxP* sites was digested from construct pGWS728 made in previous study using *Xba*I (Zhang *et al.*, 2018). And these three fragments were ligated into *Hind*III-*Bam*HI-digested pWHM3. The presence of the *loxP* recognition sites allows the efficient removal of the apramycin resistance cassette following the introduction of a plasmid pUWL-Cre expressing the Cre recombinase (Fedoryshyn *et al.*, 2008, Khodakaramian *et al.*, 2006). Knock-out construct pGWS1255 was created for the deletion of nucleotide positions -565 to +228 of SCO1839, where +1 refers to the translation start site. Introducing this construct to *S. coelicolor* A3(2) M145 followed by losing this construct resulted in replacing the respective region with apramycin resistance cassette, which was subsequently removed using Cre expressing construct pUWL-Cre. For complementation of the SCO1839 deletion mutant, the nucleotide positions -565 to +228 relative SCO1839 translation start site, containing entire coding region of SCO1839 (with stop codon) and its promoting region was cloned as a *Hind*III-*Bam*HI fragment and inserted into the multiple cloning site of low copy number plasmid pHJL401. This resulted in the complementation construct pGWS1260. This construct was then transformed to SCO1839 null mutant, resulting in strain GAD014.

For FLAG-tagging and enhanced green fluorescent protein (eGFP) fusion expression of SCO1839, the 3×FLAG or eGFP sequences were fused to the end of original copy of SCO1839 on genome using codon optimised CRISPR-Cas9 system (Cobb *et al.*, 2015). The spacer sequence located at the end of SCO1839 was the same for both 3×FLAG tag and eGFP knock-in. The sgRNA with this spacer guides the Cas9 enzyme to make a double strand break after the stop codon. This spacer sequence was inserted into the pCRISPomyces-2 plasmid as described by (Cobb *et al.*, 2015). Templates for homology-directed repair (HDR) were made and inserted to spacer containing pCRISPomyces-2 briefly as follows. For the 3×FLAG tag knock-in, about 1 kb region of SCO1839 and its upstream region were amplified by PCR from the *S. coelicolor* genome, including an additional connecting sequence replacing the stop codon of SCO1839. Additionally, around 1 kb of downstream region was PCR-amplified from the genome, whereby a 3×FLAG sequence was added upstream to ensure in-frame expression of the 3×FLAG tag. The two fragments were connected by overlap extension PCR (Figure S5). *Xba*I sites were added in both ends, and the fragment was then inserted into an *Xba*I site

in the spacer of pCRISPomyces-2, resulting in a SCO1839-3×FLAG knock-in construct pGWS1298. For the eGFP knock-in, the DNA sequence of eGFP was cloned as a *Bam*HI-*Hind*III fragment from pGWS526 (Zhang *et al.*, 2016). SCO1839 and its upstream region were cloned as an *Xba*I-*Bam*HI fragment, while the downstream region was cloned as *Hind*III-*Xba*I fragment (Figure S2). These three fragments were directly ligated into *Xba*I-digested pCRISPomyces-2, resulting in eGFP knock-in construct pGWS1299.

Mutagenesis was done according to (Cobb *et al.*, 2015). After conjugation of pGWS1298 and pGWS1299 to *S. coelicolor* A3(2) M145, ex-conjugants were patched onto SFM agar plates containing apramycin (50 µg·mL<sup>-1</sup>), then positive ex-conjugants were plated onto fresh SFM agar plates and grown at 37°C. Spores were then collected and checked for loss of the construct. Apramycin sensitive strains were incubated for spores and the genomes were checked by PCR for the desired recombination events. A successful 3×FLAG tag knock-in strain was identified and designated GAD043, eGFP knock-in strain was designated GAD099.

For over-expression of SCO1839, the *ermE* promoter was cloned to replace part of the original promoter region of SCO1839 using the same CRISPR-Cas9 system. As spacer sequence we used the promoter region of SCO1839 and this was inserted into pCRISPomyces-2 (Cobb *et al.*, 2015). This guides Cas9 to make double-strand breaks at around nt position -89 relative to the translation start site of SCO1839. To assemble the template for HDR, approximately 1 kb upstream and downstream region of SCO1839 promoter were cloned as *Xba*I-*Sec*I and *Nde*I-*Xba*I fragments, respectively. The fragments were designed removing -157 to +4 region of SCO1839. *PermE* was digested from pHM10a (Motamedi *et al.*, 1995) as a *Sec*I-*Nde*I fragment. These three fragments were then inserted into pCRISPomyces-2, producing pGWS1295. Following the same procedure as above, strain GAD039 was obtained with expresses SCO1839 from the *ermE* promoter.

To produce protein for EMSA experiments, SCO1839-His<sub>6</sub> was expressed in *E. coli* strain BL21 CodonPlus (DE3)-RIPL (Invitrogen, Massachusetts, U.S.). For this, SCO1839 and its downstream region were cloned as an *Nde*I-*Eco*RI fragment and ligated into pET28a for the expression of 6×His tag at N-terminal of SCO1839 in *E. coli*. To generate methylated and non-methylated DNA for electrophoretic mobility shift assay (EMSA), part of promoter region of SCO1839 and a random region of SCO1839 non-binding region were cloned as 324 bp and 361 bp *Eco*RI-*Hind*III fragments, respectively, and inserted into the multiple cloning site of pUC19 (Sambrook *et al.*, 1989). Resulting in EMSA testing DNA baring constructs pGWS1300 for SCO1839 binding region and pGWS1451 for non-binding region.

Knock-out of *atrA* was achieved using another CRISPR-Cas9 system that was described by (Tong *et al.*, 2015a). The spacer sequence was designed at the

beginning of *atrA* to guide Cas9 create a double-strand break at around nt position +4 of *atrA*. The template was assembled by PCR amplification of the upstream and downstream regions of *atrA*, so as to remove -41 to +46 position relative to the translation start site of *atrA*, with overhangs for Gibson assembly (Gibson *et al.*, 2009). The two fragments together with *StuI* digested spacer containing pCRISPR-Cas9 construct was assembled using Gibson assembly kit (NEB, Massachusetts, U.S.) and the resulting construct was designated pGWS1452, which was further used for the knock-out of *atrA* in *S. coelicolor* M145, SCO1839 null mutant GAD003, and *S. coelicolor* M512. The knock-out procedure was the same as using pCRISPMycos-2. The obtained M145  $\Delta$ *atrA* strain was designated GAD075,  $\Delta$ SCO1839 $\Delta$ *atrA* strain was designated GAD078, M512  $\Delta$ *atrA* strain was designated GAD093.

For testing the promoter efficiency in *S. coelicolor* strain M512 and its *atrA* null mutant, the promoter region of SCO1839 (-297 to -10) was cloned as a *Bam*HI-*Sac*I fragment and ligated into *Bam*HI-*Sac*I digested pIJ2587 (van Wezel *et al.*, 2000b). The resulting construct pGWS1454 was then transformed into *S. coelicolor* M512 and *S. coelicolor* M512  $\Delta$ *atrA*.

### Time-lapse confluent plate morphology monitoring

Approximately  $10^7$  spores were plated on MM agar supplemented with mannitol. The plates were then placed upside down in Perfection V370 scanner (Epson, Nagano, Japan) located inside 30°C incubator. A scanning picture was taken every hour, and images were processed using custom python script to get the brightness value of the plate. Specifically, the pictures were first converted to grey scale. To avoid the interference of mycelium grown in the edge which might show different developmental properties, 70% the diameter of the plate from the centre was selected as the region of interest (ROI). The average grey value of all the pixels within ROI was used as the brightness of the mycelium lawn. The measured values from one plate were then normalized to range 0 to 1.

### Scanning electron microscopy (SEM)

Mycelia were grown on MM agar supplemented with mannitol and grown for 5 days. Sample preparation and imaging was done as described before (Piette *et al.*, 2005), using JSM-7600F scanning electron microscope (JEOL, Tokyo, Japan). For each strain, 5 images with  $7,500\times$  magnification were taken in randomly selected spore-rich areas. The length and width of spores in each picture were measured using ImageJ version 1.52p strictly according to a randomized file list, in order to minimize human error. Only spores which are approximately parallel to the focal plane were measured.

## Confocal imaging

Sterile coverslips were inserted at an angle of 45° into minimum medium agar plates, spores of SCO1839-eGFP fusion protein expressing strain was carefully inoculated at the intersection angle, and incubate at 30°C. Before imaging, a drop of water containing 5  $\mu$ M cell-permeant Syto-Red nucleic acid dye (Invitrogen, Massachusetts, U.S.) and stained for 5 min at room temperature. For imaging liquid grown cultures, 5  $\mu$ L of TSBS culture were dropped on a glass slide containing a 5  $\mu$ L water containing 10  $\mu$ M Syto-Red and stained for 5 min at room temperature. Imaging was performed with a Zeiss LSM system as described before (Willemse & van Wezel, 2009). eGFP fluorescent were excited using 488 nm laser and monitored with a 505-530 nm band-pass filter, the DNA Syto-Red red fluorescent was excited using 633 nm laser and monitored using 635 nm secondary dichroic beam splitter. All images were processed in ImageJ version 1.52p.

## DNA-protein cross-linking and chromatin immunoprecipitation

10<sup>8</sup> spores of strain GAD043 were plated on MM medium covered with cellophane. After 25 h or 48 h growth, cellophane disks were soaked up-side-down in PBS solution containing 1% formaldehyde for 20 min for DNA-protein crosslinking. Ten plates were collected for 25 h samples; four plates were collected for 48 h samples. Then the disks were moved to PBS solution containing 0.5 M glycine for 5 min to stop crosslinking reaction. The mycelium was then collected, washed in PBS and resuspended in 0.5 mL lysis buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 15 mg·mL<sup>-1</sup> lysozyme, 1× protease inhibitor (Roche, Bavaria, Germany) and incubated at 37°C for 20 min. After incubation, 0.5 mL IP buffer (100 mM Tris-HCl pH 8.0, 250 mM NaCl, 0.8% v/v Triton-X-100) was added to the sample, and chromosomal DNA sheared to 100 to 500 bp using Bioruptor Plus water bath sonication system (Diagenode, Liège, Belgium). After centrifugation to remove cell debris, lysates were incubated with 40  $\mu$ L Anti-FLAG M2 affinity gel (cat A2220, Sigma-Aldrich, St. Louis, U.S.) suspension according to the manufacturer's instructions and incubated at 4°C overnight. After centrifugation and washing, the pellets and 50  $\mu$ L of untreated total extracts (controls) were incubated in 100  $\mu$ L IP elution buffer (50 mM Tris-HCl pH7.5, 10 mM EDTA, 1% m/v SDS) at 65°C overnight to reverse cross-links. Beads were then removed by centrifugation before DNA extraction with phenol-chloroform. The DNA sample was then extracted with chloroform and the water layer was further purified using DNA Clean & Concentrator kit (Zymo Research, California, U.S.). The samples were then sent for next generation sequencing using BGI-Seq platform (BGI, Hong Kong, China)

## ChIP-Seq data analysis

Clean reads received from sequencing contractor were aligned to the *S. coelicolor* M145 genome with RefSeq accession number NC\_003888.3 using bowtie2 (version 2.3.4, Langmead & Salzberg, 2012). Resulted SAM files are sorted using

SAMtools (version 1.9, Li *et al.*, 2009) producing BAM files. MACS2 (version 2.1.2, Zhang *et al.*, 2008) was then used for binding peak prediction and peak modelling by comparing the chromatin immunoprecipitated DNA sample with the corresponding whole genome sample. The models for both samples are shown in Figure S4. The enrichment data used in Figure 5 was calculated for each nucleotide using MACS2 'bdgcmp' command with '-m FE' switch. The peak summit positions including sub peak positions of each predicted binding region were then extracted, and the  $\pm 150$  bp region of each summit was extracted from genome sequence using python script dependent on the Biopython module (version 1.70, Cock *et al.*, 2009). Extracted sequences were subjected to MEME-ChIP (version 5.0.2, Machanick & Bailey, 2011), which is suitable for large sequence sets, for binding motif prediction.

The enrichment data of two samples was averaged separately in a moving bin of 20,000 bp and plotted at opposite directions as the middle ring of the circular genome diagram. The G+C content was calculated using the same moving bin and centred at the middle of maximum and minimum value before plotting as the inner ring on the plot (Figure 5A). For determining the overlap of low G+C content regions and high enrichment regions, the genome was divided into 1,000 bp long sections, the G+C content and average enrichment levels were calculated. The sections with G+C content below the first quartile was considered low in G+C, and those with average enrichment level above the third quartile were considered high in enrichment. To find genes possibly regulated by SCO1839, the locations of promoter regions (-350 to +50) of all genes were extract from the GenBank file containing annotations and checked for overlap with  $\pm 150$  bp location of the summit of SCO1839 binding peaks. This was done using python script dependent on module Biopython and pybedtools (version 0.8, Dale *et al.*, 2011) and external BEDTools (version 2.27, Quinlan, 2014).

### Electrophoretic mobility shift assay (EMSA)

Construct pGWS1286 expressing SCO1839-His<sub>6</sub> was introduced into *E. coli* BL21 CodonPlus (DE3)-RIPL and expressed and purified as described before (Mahr *et al.*, 2000). Purified protein was dialyzed over-night at 4°C against EMSA buffer (10 mM Tris-HCl, pH 7.9, 0.1 mM EDTA, 50 mM KCl). 50 bp double strand DNA was generated by gradual cooling of reverse complemented single strand oligonucleotides in T4 DNA ligase buffer (NEB, Massachusetts, U.S.) from 95°C to 12°C in 45 min. For testing the methylation effect on SCO1839 binding, long DNA fragments of the SCO1839 promoter region were cloned into pUC19, forming pGWS1300. The non-binding control fragment was randomly selected from the genome that showed no enrichment in ChIP-Seq experiment, this fragment was then cloned into pUC19, to give pGWS1451. pGWS1300 was then transformed to both DAM methylation effective *E. coli* strain JM109 and DAM deficient *E. coli* strain ET12567, while pGWS1451 for negative control was transformed to strain



ET12567 only. The target fragments were then digested from extracted constructs and blunt-ends created using DNA polymerase I Klenow fragment (NEB, Massachusetts, U.S.) and dNTPs. The *in-vitro* DNA-protein interaction tests were done in EMSA buffer in a total reaction volume of 10  $\mu$ L and incubated at 30°C for 15 min. The reactions were then loaded on 5% polyacrylamide gels and separated by electrophoresis. The gel was briefly staining with ethidium bromide and imaged in a Gel Doc imaging system (BioRad, California, U.S.).

### RNA extraction and quantitative reverse transcription PCR

*S. coelicolor* strains were inoculated on minimal medium agar plates covered with cellophane disks with a final inoculum of  $10^7$  CFU per plate. Biomass from two (for 24 h) or one (48 h) agar plate was collected in 15 mL tube. The RNA was extracted using modified Kirby mix (Kieser *et al.*, 2000). The total nucleic acid was precipitated using 1 volume isopropanol and 0.1 volume of 3 M sodium acetate and then dissolved followed by purification using RNeasy columns (Qiagen, Venlo, The Netherlands) according to vendor's instructions, and the remainder of the DNA was removed by DNase I (NEB, Massachusetts, U.S.) treatment followed by phenol chloroform extraction. The complete removal of DNA was confirmed by PCR. cDNA was synthesized using iScript cDNA synthesis kit (BioRad, California, U.S.) according to vendor's instructions. Quantitative PCR was done using iTaq Universal SYBR green qPCR kit (BioRad, California, U.S.). The PCR program was set as following: 95°C 30 s; 40 cycles of 95°C 10 s, 60°C 30 s, plate read; melt curve from 65°C to 95°C with 5 s per 0.5°C increment. House-keeping genes *gyrA* (SCO3873) and *rpmE3* (SCO5359) that are known to show constitutive expression were used as inner controls. The data was analysed using CFX Manager software (version 3.1, BioRad, California, U.S.) using  $\Delta\Delta C_q$  method, which is an implementation of the method described by (Vandesompele *et al.*, 2002).

### Acknowledgments

We are grateful to Prof.dr. K.J. McDowall and to Prof.dr. R.T. Dame for stimulating discussions.

## Chapter 7

---

General discussion and future perspectives

Actinobacteria are well known for their complex life cycle and the ability to produce bioactive natural products (Barka *et al.*, 2016, Hopwood, 2007, Bérdy, 2005). Due to the extensive screening of these bacteria in the 20th century, the frequency of finding truly novel antibiotics has decreased dramatically. Therefore, new systematic approaches should deliver the goods in terms of finding the novel drugs needed to counteract the problems associated with the emerging antibiotic resistance. Given that conventional NP screening methods identify primarily those compounds that are produced at higher concentrations under laboratory conditions, new methods are required to uncover the bioactive compounds that are specified by cryptic BGCs (Baltz, 2007). From a metabolomics perspective, this means that we need to improve the efficiency in dereplication (Wu *et al.*, 2015b, Krug & Muller, 2014). To allow us to identify the NPs produced from BGCs that are cryptic under laboratory conditions, we need to discover ways to activate them. Mutagenesis, changing culturing conditions and mimicking environmental factors are three main methods to achieve undirected activation of BGCs. The mutagenesis method is exemplified in **Chapter 3**, where streptomycin-resistant mutants were generated, and strains with enhanced or reduced antibiotic activity were selected for analysis. The sought-after BGC was then identified by correlating the production of the bioactive metabolite using LC-MS with the expression of BGCs as seen by proteomics. Alternatively, multiple small molecules mimicking environmental factors may be applied to activate antibiotic production, such as shown for the promising NP producer *S. roseifaciens* in **Chapter 4**. This study revealed eliciting strategies that may be applied in a systematic way.

## 7

Systems biology is a key element in NP research because it allows us to understand the biological system as a whole, which takes advantage of the huge technological advances that were missing in the past (Butcher *et al.*, 2004). The recent advances in 'omics technologies, especially transcriptomics, metabolomics, and proteomics, has made systems-level investigation more achievable. Of these technologies, proteomics is important for the study of NP biosynthesis as it allows scientists to determine the expression levels of the metabolic enzymes (Du & van Wezel, 2018). As shown in **Chapter 4**, statistical analysis revealed that jasmonic acid, N-acetylglucosamine, chitosan and benzoic acid induced a surprisingly similar response at whole proteome level in *S. roseifaciens*. However, this did not lead to a similar response in the antimicrobial activity, indicating that this activity might relate to the changes in a few key enzymes, small differences in the growth conditions between experiments, or the effects of elicitors on growth rate or morphology. Second, gene ontology enrichment analysis gives finer detail of the samples. The results showed that hydroxycoumarin, which is a good elicitor of antimicrobial activity, probably affects energy metabolism in *S. roseifaciens*. Finally, high-resolution proteomics studies on the short-term responses to JA identified a gene cluster (*jar*) that relates to the antimicrobial activity of *S.*

*roseifaciens*, though it does not directly correlate to the biosynthesis of the bioactive metabolites.

Expressing BGCs to obtain more insights into the metabolites they specify is not always straight forward, especially for those from non-culturable organisms or identified in metagenomic studies. In order to access this part of the NP repository, heterologous expression of BGCs in an optimised host is an important method (Nepal & Wang, 2019, Gomez-Escribano & Bibb, 2011). It has been shown to be effective in finding novel antibiotics (Du *et al.*, 2013), determining the boundaries of the BGC (Komatsu *et al.*, 2013, Liu *et al.*, 2018), creating artificial natural products by combining different BGCs (Park *et al.*, 2011), and characterizing functions of individual genes (Waldman *et al.*, 2015). However, choosing a suitable host for targeted expression of BGCs remains difficult. This fact, together with other technical difficulties such as transfer of large gene fragments and the requirement of biosynthetic repertoire from the natural producer that may be lacking in the new host, slows down the application of heterologous production platforms. Furthermore, what is the basis to decide which BGC to prioritise, if we do not know what the cognate NP is? Gene synthesis is still relatively expensive and hence high-throughput gene synthesis is not commonplace. In attempts to solve these issues, an background-reduced strain *S. coelicolor* M1152 was made to maximise production and at the same time reduce the complexity of extracts (Gomez-Escribano & Bibb, 2011). However, the optimization of this strain is mainly based on experience, while systems level details about the effect of the optimization process was missing so far. Therefore, we performed a time-series multi-omics study in a fully defined fermentation system and analysed the samples via high resolution proteomics (**Chapter 5**). Significant differences were observed in the global protein profiles between the background-reduced strain and its wild-type parent, including many members of the regulon of the global transcription factors PhoP, GlnR, and ScbR. We also found upregulation of ectoine biosynthetic enzymes, which may reflect increased osmotic stress in *S. coelicolor* M1152. These findings will be beneficial for further rational design of this strain or provide guidance to the optimization of other promising host strains.

The first fully sequenced model organism in the major antibiotic producing phylum, Actinobacteria, is *S. coelicolor* (Bentley *et al.*, 2002). Its ability to produce many different antibiotics, in particular PKS and NRPS, suggests it is a logical choice as host for the heterologous expression of BGCs. *S. coelicolor* has been the model organism for research on streptomycetes since the 1960s (Hoskisson & van Wezel, 2019) and a large proportion of our understanding of the regulatory networks that control secondary metabolism is obtained from this species (Rigali *et al.*, 2018, Urem *et al.*, 2016). This regulatory network is also closely correlated with the development of *Streptomyces* (van Wezel & McDowall, 2011).

An interesting protein is SCO1839, which is member of a yet unstudied protein family that is unique to Actinobacteria. Previous studies on developmental genes of *S. coelicolor* suggested that SCO1839 is involved in the regulation of sporulation (Kim *et al.*, 2015). **Chapter 6** provides a detailed analysis on the DNA binding properties of SCO1839. We found that SCO1839 is a likely nucleoid associated protein (NAP) that binds to a very small DNA sequence centred around the palindrome GATC. As expected from the short motif, the binding site of SCO1839 is widespread on the chromosome. Besides SCO1839, many other small DNA binding proteins are encoded by the *S. coelicolor* genome that control development, as exemplified by BldC (Hunt *et al.*, 2005). Its mode of DNA binding appears to be distinct from that of other regulators, involving asymmetric head-to-tail oligomerization on direct repeats that results in dramatic DNA distortion (Schumacher *et al.*, 2018, Bush *et al.*, 2019). Unlike SCO1839, no significant binding motif has so far been proposed for BldC. However, just like the DNA binding sites of BldC vary significantly in length, the lengths of the SCO1839 target sequences also vary, as shown by the ChIP-Seq data. Thus, we need to elucidate the DNA-binding model for SCO1839 and compare it with that of BldC and other actinobacterial regulators. *Streptomyces* have many different NAPs whose functions have not yet been fully deciphered. These proteins not only provide structure to the nucleoid but also perform important roles in the regulation of various cellular processes. The study of SCO1839 and its specific binding motif has provided a new route to decode function of NAPs in *Streptomyces*. It also emphasised the importance of NAPs in participating development processes of *Streptomyces*, which are major regulatory factors in secondary metabolism.

## 7

In conclusion, this thesis describes the application of advanced proteomics and other 'omics technologies in NP research in *Streptomyces*, in particular as part of systems-wide studies. This includes the proteomining technology to connect BGCs to a bioactivity of interest, studies of proteome-level responses of *Streptomyces* to small molecules mimicking environmental signals, and analysis of the consequences of strain design approaches for the optimization of *S. coelicolor* as heterologous host for NP production. Additionally, a new family of small DNA binding proteins that might function in epigenetic regulation of antibiotic production was described, combining molecular biology and bioinformatics analysis. Nowadays, new high-throughput methods, and especially 'omics techniques, generate magnitudes more data than ever before. The challenge thereby lies in combining multi-dimensional datasets in such a way that we can answer major biological questions. I hope that this thesis has made a contribution to fulfil this daunting task. Systems-wide approaches are key to obtaining a more comprehensive understanding of the regulatory networks that control NP biosynthesis and the complex biology of *Streptomyces*.

## Nederlandse samenvatting

Actinobacteriën staan bekend om hun complexe levenscyclus en het vermogen om bioactieve natuurlijke producten (NP) te produceren (Hopwood, 2007, Fedorova *et al.*, 2012). Door de uitgebreide screening van deze bacteriën in de 20e eeuw is de frequentie van het vinden van echt nieuwe antibiotica dramatisch afgenomen. Nieuwe systematische benaderingen zouden daarom nieuwe geneesmiddelen moeten opleveren die nodig zijn om de problemen van de opkomende antibioticaresistentie tegen te gaan. Aangezien conventionele NP-screeningsmethoden voornamelijk die stoffen identificeren die onder laboratoriumomstandigheden in hogere concentraties worden geproduceerd, zijn nieuwe methoden nodig om de bioactieve stoffen te ontdekken die worden gespecificeerd door cryptische biosynthetische genclusters (BGCs) (Baltz, 2007). Vanuit metabolomics-perspectief betekent dit dat we de efficiëntie van dereplicatie moeten verbeteren (Wu *et al.*, 2015b, Krug & Muller, 2014). Om ons in staat te stellen de NPs te identificeren die zijn geproduceerd uit BGCs die onder laboratoriumomstandigheden cryptisch zijn, moeten we manieren vinden om ze te activeren. Mutagenese, veranderende groeiomstandigheden en het nabootsen van omgevingsfactoren zijn drie belangrijke methoden om algemene activering van BGCs te bereiken. De mutagenesemethode wordt geïllustreerd in **Hoofdstuk 3**, waar streptomycine-resistente mutanten werden gegenereerd en stammen met verhoogde of verminderde antibiotica-activiteit werden geselecteerd voor verdere analyse. Het verantwoordelijke BGC werd vervolgens geïdentificeerd door de productie van de bioactieve metaboliet met LC-MS te correleren aan de expressie van BGCs welke door proteoom-analyse is verkregen. Een alternatief om de antibioticaproductie te activeren is het nabootsen van omgevingsfactoren door meerdere kleine moleculen, zoals aangetoond voor de veelbelovende NP-producent *S. roseifaciens* in **Hoofdstuk 4**. Deze studie bracht strategieën aan het licht die op een systematische manier kunnen worden toegepast.

Systeembioologie is een essentieel onderdeel van NP-onderzoek, omdat het ons in staat stelt het biologische systeem als geheel te begrijpen. Dit onderzoeksveld profiteert nu van de enorme technologische vooruitgang die in het verleden ontbrak (Butcher *et al.*, 2004). De recente vooruitgang in omics-technologieën, met name transcriptomics, metabolomics en proteomics, heeft dit type onderzoek beter haalbaar gemaakt. Onder deze technologieën is proteomics belangrijk voor de studie van NP-biosynthese, omdat het wetenschappers in staat stelt de expressieniveaus van de metabole enzymen te bepalen (Du & van Wezel, 2018). Zoals aangetoond in **Hoofdstuk 4** bleek uit statistische analyse dat jasmonzuur, N-acetylglucosamine, chitosan en benzeencarbonzuur een verrassend vergelijkbaar effect op proteoom-niveau in *S. roseifaciens* veroorzaakten. Dit leidde echter niet tot een vergelijkbaar effect in de antimicrobiële activiteit, wat kan betekenen dat dit mogelijk verband houdt met de veranderingen in enkele belangrijke enzymen,

kleine verschillen in de groeiomstandigheden tussen experimenten of de effecten van elicitors op groeisnelheid of morfologie. Ten tweede brengt verrijkingsanalyse met genontologie de resolutie terug naar het tussenliggende niveau. De resultaten toonden aan dat hydroxycoumarine, dat een goede antimicrobiële activiteit opwekt, waarschijnlijk het energiemetabolisme van *S. roseifaciens* beïnvloedt. Door het gebruik van hoge-resolutie proteomics voor het onderzoeken van de kortetermijnreactie op jasmonzuur werd ten slotte een gencluster (*jar*) geïdentificeerd. Dit gencluster is gerelateerd aan de antimicrobiële activiteit van *S. roseifaciens*, maar correleert niet direct aan de biosynthese van de biologische activiteit.

Het tot uiting brengen van cryptische BGCs om meer inzicht te krijgen in de metabolieten die ze specificeren, is niet altijd ongecompliceerd, vooral niet voor die van niet-kweekbare micro-organismen of voor die geïdentificeerd zijn in metagenoom studies. Om toegang te krijgen tot dit deel van de NP-repository is heterologe expressie van BGCs in een geoptimaliseerde host een belangrijke methode (Nepal & Wang, 2019, Gomez-Escribano & Bibb, 2011). Deze methode is effectief gebleken bij het vinden van nieuwe antibiotica (Du *et al.*, 2013), het bepalen van de grenzen van BGCs (Komatsu *et al.*, 2013, Liu *et al.*, 2018), het creëren van kunstmatige natuurlijke producten door verschillende BGCs te combineren (Park *et al.*, 2011) en voor het karakteriseren van functies van individuele genen (Waldman *et al.*, 2015). Het blijft echter moeilijk om een geschikte gastheer te kiezen voor specifieke BGCs. Dit feit, samen met andere technische problemen, zoals het overdragen van grote genfragmenten en de vereiste van biosynthetische repertoire van de natuurlijke producent dat mogelijk ontbreekt in de nieuwe gastheer, vertraagt de toepassing van heterologe productieplatforms. Wat is bovendien de basis om te beslissen welke BGC prioriteit te geven, als we niet weten wat de verwante NP is? Gensynthese is nog steeds relatief duur, en daarom is grootschalige synthese van genen niet gebruikelijk. Om deze problemen op te lossen werd een geoptimaliseerde stam, *S. coelicolor* M1152, gemaakt om de productie te maximaliseren en tegelijkertijd de complexiteit van het maken van extracten te verminderen (Gomez-Escribano & Bibb, 2011). Het optimaliseren van deze stam is echter voornamelijk gebaseerd op ervaring, en het was onbekend wat het effect van het optimalisatieproces op het systeemniveau is. Daarom hebben we een multi-omics studie uitgevoerd over een tijdreeks in een volledig gedefinieerd fermentatiesysteem en vervolgens de monsters geanalyseerd via proteomics met een hoge resolutie (**Hoofdstuk 5**).

Significante verschillen werden waargenomen in de globale proteïneprofielen tussen de stam met gereduceerde achtergrond en zijn wildtype ouder, waaronder veel leden van de regulon van de globale transcriptiefactoren PhoP, GlnR en ScbR. We vonden ook positieve regulatie van biosynthetische enzymen van ectoïne, welke een verhoogde osmotische stress in *S. coelicolor* M1152 kunnen aangeven.

Deze bevindingen bevorderen een verder rationeel ontwerp van deze stam of een vormen een leidraad voor de optimalisatie van andere veelbelovende gastheerstammen.

Het eerste volledig gesequentieerde modelorganisme in het phylum van meest belangrijke antibioticumproducenten, Actinobacteriën, is *S. coelicolor* (Bentley *et al.*, 2002). Het vermogen om veel verschillende antibiotica te produceren, in het bijzonder PKS en NRPS, suggereert dat het een logische keuze is als gastheer voor de heterologe expressie van BGCs. *S. coelicolor* is sinds 1960s het modelorganisme voor onderzoek naar streptomyceten (Rigali *et al.*, 2018, Urem *et al.*, 2016). Dit regelgevende netwerk hangt ook nauw samen met de ontwikkeling van *Streptomyces* (van Wezel & McDowall, 2011).

Een interessant proteïne is SCO1839, dat lid is van een nog niet bestudeerde proteïnefamilie die uniek is voor Actinobacteriën. Eerdere studies naar genen betrokken bij de ontwikkeling van *S. coelicolor* hebben voorgesteld dat SCO1839 mogelijk deelneemt aan het reguleren van sporulatie (Kim *et al.*, 2015). **Hoofdstuk 6** geeft een gedetailleerde analyse van de DNA-bindingseigenschappen van SCO1839. We ontdekten dat SCO1839 een waarschijnlijk nucleoïde-geassocieerd proteïne (NAP) is dat zich bindt aan een zeer kleine DNA-sequentie rond het palindroom GATC. Zoals verwacht van het korte motief, is de bindingslocatie van SCO1839 wijdverspreid op het chromosoom. Naast SCO1839 worden veel andere kleine DNA-bindende proteïnen gecodeerd door het *S. coelicolor*-genoom dat de ontwikkeling controleert, zoals geïllustreerd door BldC (Hunt *et al.*, 2005). De wijze van DNA-binding lijkt te verschillen van die van andere regulatoren, welke functioneren via asymmetrische kop-staart-oligomerisatie op directe DNA-herhalingssequenties, wat resulteert in dramatische DNA-ervorming (Schumacher *et al.*, 2018, Bush *et al.*, 2019). In tegenstelling tot SCO1839 is er tot nu toe geen significant bindingsmotief voorgesteld voor BldC. Aangezien de bindingsplaatsen op het DNA van BldC echter aanzienlijk in lengte variëren, is er ook variatie in de lengtes van de SCO1839-bindingssequenties, zoals blijkt uit de ChIP-Seq-gegevens. Het is dus nodig om het DNA-bindings model voor SCO1839 te ophelderen en te vergelijken met dat van BldC en andere actinobacterie regulatoren.

*Streptomyces* hebben veel verschillende NAPs waarvan de functies nog niet volledig zijn ontcijferd. Deze proteïnen geven niet alleen structuur aan de nucleoïde, maar spelen ook een belangrijke rol bij de regulering van verschillende cellulaire processen. De studie van SCO1839 en zijn specifieke bindingsmotief heeft een nieuwe route opgeleverd om de functie van NAPs in *Streptomyces* te decoderen. Het benadrukte ook het belang van NAPs in deelnemende ontwikkelingsprocessen van *Streptomyces*, die belangrijke regulerende factoren zijn in het secundaire metabolisme.



Concluderend, dit proefschrift beschrijft de toepassing van geavanceerde proteomics en andere 'omics-technologieën in NP-onderzoek in *Streptomyces*, in het bijzonder als onderdeel van systeembrede studies. Dit omvat de proteomining-technologie om BGCs te verbinden met een bioactiviteit die van belang is, studies van de gevolgen op proteoom-niveau van *Streptomyces* op kleine moleculen die omgevingssignalen nabootsen, en analyse van de gevolgen van stam ontwerpbenaderingen voor de optimaliseren van *S. coelicolor* als heterologe gastheer voor productie van de NP. Bovendien werd een nieuwe familie van kleine DNA-bindende proteïnen beschreven die zouden kunnen functioneren in de epigenetische regulatie van de antibiotica-productie, waarbij de analyse van moleculaire biologie en bio-informatica worden gecombineerd. Tegenwoordig genereren nieuwe high-throughput-methoden, en in het bijzonder 'omics-technieken, veel meer gegevens dan ooit tevoren. De uitdaging ligt daarbij in het zo combineren van multidimensionale datasets dat we grote biologische vragen kunnen beantwoorden. Ik hoop dat dit proefschrift een bijdrage heeft geleverd aan het vervullen van deze lastige taak. Systeembrede benaderingen zijn essentieel voor het verkrijgen van een meer uitgebreid begrip van de regulerende netwerken die NP-biosynthese en de complexe biologie van *Streptomyces* controleren.

## References

---

- Abdelmohsen, U.R., Grkovic, T., Balasubramanian, S., Kamel, M.S., Quinn, R.J., and Hentschel, U. (2015) Elicitation of secondary metabolism in Actinomycetes. *Biotechnol. Adv.* **33**: 798-811.
- Abrudan, M.I., Smakman, F., Grimbergen, A.J., Westhoff, S., Miller, E.L., van Wezel, G.P., and Rozen, D.E. (2015) Socially mediated induction and suppression of antibiosis during bacterial coexistence. *Proc. Natl. Acad. Sci. USA* **112**: 11054-11059.
- Agrawal, P., Khater, S., Gupta, M., Sain, N., and Mohanty, D. (2017) RiPPMiner: A bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links. *Nucleic Acids Res.* **45**: W80-W88.
- Ahmad, F., Singh, A., and Kamal, A., (2019) Chapter 23 – Salicylic acid-mediated defense mechanisms to abiotic stress tolerance. In: Plant Signaling Molecules. M.I.R. Khan, P.S. Reddy, A. Ferrante & N.A. Khan (eds). Woodhead Publishing, pp. 355-369.
- Ahmed, Y., Rebets, Y., Estévez, M.R., Zapp, J., Myronovskiy, M., and Luzhetskyy, A. (2020) Engineering of *Streptomyces lividans* for heterologous expression of secondary metabolite gene clusters. *Microb. Cell Fact.* **19**: 5.
- Alam, M.T., Merlo, M.E., Hodgson, D.A., Wellington, E.M.H., Takano, E., Breitling, R., and The, S.C. (2010) Metabolic modeling and analysis of the metabolic switch in *Streptomyces coelicolor*. *BMC Genom.* **11**: 202.
- Albright, J.C., Goering, A.W., Doroghazi, J.R., Metcalf, W.W., and Kelleher, N.L. (2014) Strain-specific proteogenomics accelerates the discovery of natural products via their biosynthetic pathways. *J. Ind. Microbiol. Biotechnol.* **41**: 451-459.
- Alekshun, M.N., and Levy, S.B. (1999) The *mar* regulon: Multiple resistance to antibiotics and other toxic chemicals. *Trends Microbiol.* **7**: 410-413.
- Alekshun, M.N., Levy, S.B., Mealy, T.R., Seaton, B.A., and Head, J.F. (2001) The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 Å resolution. *Nat. Struct. Biol.* **8**: 710-714.
- Allenby, N.E.E., Laing, E., Bucca, G., Kierzek, A.M., and Smith, C.P. (2012) Diverse control of metabolism and other cellular processes in *Streptomyces coelicolor* by the PhoP transcription factor: Genome-wide identification of in vivo targets. *Nucleic Acids Res.* **40**: 9543-9556.
- Amano, S., Morota, T., Kano, Y.K., Narita, H., Hashidzume, T., Yamamoto, S., Mizutani, K., Sakuda, S., Furihata, K., Takano-Shiratori, H., Takano, H., Beppu, T., and Ueda, K. (2010) Promomycin, a polyether promoting antibiotic production in *Streptomyces* spp. *J. Antibiot.* **63**: 486-491.
- Amin, R., Reuther, J., Bera, A., Wohlleben, W., and Mast, Y. (2012) A novel GlnR target gene, *nnar*, is involved in nitrate/nitrite assimilation in *Streptomyces coelicolor*. *Microbiology* **158**: 1172-1182.
- Avalos, M., Boetzer, M., Pirovano, W., Arenas, N.E., Douthwaite, S., and van Wezel, G.P. (2018) Complete genome sequence of *Escherichia coli* AS19, an antibiotic-sensitive variant of *E. coli* strain B REL606. *Genome Announc.* **6**: e00385-00318.
- Baltz, R.H. (2007) Antimicrobials from Actinomycetes: back to the future. *Microbe* **2**: 125-131.
- Baltz, R.H. (2008) Renaissance in antibacterial discovery from Actinomycetes. *Curr. Opin. Pharmacol.* **8**: 557-563.

- Barka, E.A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H.-P., Clément, C., Ouhdouch, Y., and van Wezel, G.P. (2016) Taxonomy, physiology, and natural products of Actinobacteria. *Microbiol. Mol. Biol. Rev.* **80**: 1-43.
- Barras, F., and Marinus, M.G. (1989) The great GATC: DNA methylation in *E. coli*. *Trends Genet.* **5**: 139-143.
- Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C.L., Serova, N., Davis, S., and Soboleva, A. (2012) NCBI GEO: Archive for functional genomics data sets – update. *Nucleic Acids Res.* **41**: D991-D995.
- Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J., and Hopwood, D.A. (2002) Complete genome sequence of the model Actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**: 141-147.
- Bérdy, J. (2005) Bioactive microbial metabolites. *J. Antibiot.* **58**: 1-26.
- Bérdy, J. (2012) Thoughts and facts about antibiotics: Where we are now and where we are heading. *J. Antibiot.* **65**: 385-395.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000) The protein data bank. *Nucleic Acids Res.* **28**: 235-242.
- Bibb, M.J. (2005) Regulation of secondary metabolism in streptomycetes. *Curr. Opin. Microbiol.* **8**: 208-215.
- Bibb, M.J., Janssen, G.R., and Ward, J.M. (1985) Cloning and analysis of the promoter region of the erythromycin resistance gene (*ermE*) of *Streptomyces erythraeus*. *Gene* **38**: 215-226.
- Bierman, M., Logan, R., O'Brien, K., Seno, E.T., Nagaraja Rao, R., and Schoner, B.E. (1992) Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene* **116**: 43-49.
- Blin, K., Medema, M.H., Kazempour, D., Fischbach, M.A., Breitling, R., Takano, E., and Weber, T. (2013) antiSMASH 2.0 – a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.* **41**: 1-9.
- Blin, K., Wolf, T., Chevrete, M.G., Lu, X.W., Schwalen, C.J., Kautsar, S.A., Duran, H.G.S., Santos, E.L.C.D.L., Kim, H.U., Nave, M., Dickschat, J.S., Mitchell, D.A., Shelest, E., Breitling, R., Takano, E., Lee, S.Y., Weber, T., and Medema, M.H. (2017) antiSMASH 4.0 – improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45**: 36-41.
- Bocks, S.M. (1967) Fungal metabolism – I.: The transformations of coumarin, o-coumaric acid and *trans*-cinnamic acid by *Aspergillus niger*. *Phytochemistry* **6**: 127-130.
- Bradshaw, E., Saalbach, G., and McArthur, M. (2013) Proteomic survey of the *Streptomyces coelicolor* nucleoid. *J. Proteom.* **83**: 37-46.

- Breinholt, J., Demuth, H., Heide, M., Jensen, G.W., Moller, I.L., Nielsen, R.I., Olsen, C.E., and Rosendahl, C.N. (1996) Prenisatin (5-(3-Methyl-2-butenyl)-indole-2,3-dione): An antifungal isatin derivative from *Chaetomium globosum*. *Acta Chem. Scand.* **50**: 443-445.
- Brown, S.A., (1986) Biochemistry of plant coumarins. In: The Shikimic Acid Pathway. E.E. Conn (ed). Boston, MA: Springer US, pp. 287-316.
- Bumpus, S.B., Evans, B.S., Thomas, P.M., Ntai, I., and Kelleher, N.L. (2009) A proteomics approach to discovering natural products and their biosynthetic pathways. *Nat. Biotechnol.* **27**: 951-956.
- Bursy, J., Kuhlmann, A.U., Pittelkow, M., Hartmann, H., Jebbar, M., Pierik, A.J., and Bremer, E. (2008) 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.
- Bush, M.J., Chandra, G., Al-Bassam, M.M., Findlay, K.C., and Buttner, M.J. (2019) BldC delays entry into development to produce a sustained period of vegetative growth in *Streptomyces venezuelae*. *mBio* **10**: e02812-02818.
- Butcher, E.C., Berg, E.L., and Kunkel, E.J. (2004) Systems biology in drug discovery. *Nat. Biotechnol.* **22**: 1253-1259.
- Cai, H., Strouse, J., Dumlao, D., Jung, M.E., and Clarke, S. (2001) Distinct reactions catalyzed by bacterial and yeast trans-aconitate methyltransferases. *Biochemistry* **40**: 2210-2219.
- Carvalhais, L.C., Dennis, P.G., Badri, D.V., Kidd, B.N., Vivanco, J.M., and Schenk, P.M. (2015) Linking jasmonic acid signaling, root exudates, and rhizosphere microbiomes. *Mol. Plant. Microbe Interact.* **28**: 1049-1058.
- Cashel, M., (1996) The stringent response. In: *Escherichia coli* and *Salmonella*: cellular and molecular biology. pp. 1458-1496.
- Castro-Melchor, M., Charaniya, S., Karypis, G., Takano, E., and Hu, W.-S. (2010) Genome-wide inference of regulatory networks in *Streptomyces coelicolor*. *BMC Genom.* **11**: 578.
- Chapman, J.D., Goodlett, D.R., and Masselon, C.D. (2014) Multiplexed and data-independent tandem mass spectrometry for global proteome profiling. *Mass Spectrom. Rev.* **33**: 452-470.
- Chater, K.F., and Chandra, G. (2008) The use of the rare UUA codon to define "Expression Space" for genes involved in secondary metabolism, development and environmental adaptation in *Streptomyces*. *J. Microbiol.* **46**: 1-11.
- Chater, K.F., and Losick, R., (1997) Mycelial life style of *Streptomyces coelicolor* A3(2) and its relatives. In: Bacteria as multicellular organisms. J.A. Shapiro & M. Dworkin (eds). New York: Oxford University Press, pp. 149-182.
- Chen, Y., McClure, R.A., Zheng, Y., Thomson, R.J., and Kelleher, N.L. (2013) Proteomics guided discovery of flavopeptins: Anti-proliferative aldehydes synthesized by a reductase domain-containing non-ribosomal peptide synthetase. *J. Am. Chem. Soc.* **135**: 10449-10456.
- Chen, Y., Ntai, I., Ju, K.S., Unger, M., Zamdborg, L., Robinson, S.J., Doroghazi, J.R., Labeda, D.P., Metcalf, W.W., and Kelleher, N.L. (2012) A proteomic survey of nonribosomal peptide and polyketide biosynthesis in Actinobacteria. *J. Proteome Res.* **11**: 85-94.

- Cho, H.J., Kim, K.-J., Kim, M.H., and Kang, B.S. (2008) Structural insight of the role of the *Hahella chejuensis* HapK protein in prodigiosin biosynthesis. *Proteins: Structure, Function, and Bioinformatics* **70**: 257-262.
- Cichewicz, R.H. (2010) Epigenome manipulation as a pathway to new natural product scaffolds and their congeners. *Nat. Prod. Rep.* **27**: 11-22.
- Cimermancic, P., Medema, Marnix H., Claesen, J., Kurita, K., Wieland Brown, Laura C., Mavrommatis, K., Pati, A., Godfrey, Paul A., Koehrsen, M., Clardy, J., Birren, Bruce W., Takano, E., Sali, A., Linington, Roger G., and Fischbach, Michael A. (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**: 412-421.
- Claessen, D., Rink, R., de Jong, W., Siebring, J., de Vreugd, P., Boersma, F.G.H., Dijkhuizen, L., and Wösten, H.A.B. (2003) A novel class of secreted hydrophobic proteins is involved in aerial hyphae formation in *Streptomyces coelicolor* by forming amyloid-like fibrils. *Genes Dev.* **17**: 1714-1726.
- Claessen, D., Rozen, D.E., Kuipers, O.P., Sogaard-Andersen, L., and van Wezel, G.P. (2014) Bacterial solutions to multicellularity: A tale of biofilms, filaments and fruiting bodies. *Nat. Rev. Microbiol.* **12**: 115-124.
- Clarke-Pearson, M.F., and Brady, S.F. (2008) Paerucumarin, a new metabolite produced by the *pvc* gene cluster from *Pseudomonas aeruginosa*. *J. Bacteriol.* **190**: 6927-6930.
- Cobb, R.E., Wang, Y., and Zhao, H. (2015) High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS Synth. Biol.* **4**: 723-728.
- Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., and de Hoon, M.J.L. (2009) Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**: 1422-1423.
- Colson, S., Stephan, J., Hertrich, T., Saito, A., van Wezel, G.P., Titgemeyer, F., and Rigali, S. (2007) Conserved *cis*-acting elements upstream of genes composing the chitinolytic system of streptomycetes are DasR-responsive elements. *J. Mol. Microbiol. Biotechnol.* **12**: 60-66.
- Colson, S., van Wezel, G.P., Craig, M., Noens, E.E., Nothaft, H., Mommaas, A.M., Titgemeyer, F., Joris, B., and Rigali, S. (2008) The chitobiose-binding protein, DasA, acts as a link between chitin utilization and morphogenesis in *Streptomyces coelicolor*. *Microbiology* **154**: 373-382.
- Cooper, M.A., and Shlaes, D. (2011) Fix the antibiotics pipeline. *Nature* **472**: 32-32.
- Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**: 1367-1372.
- Cragg, G.M., Newman, D.J., and Snader, K.M. (1997) Natural products in drug discovery and development. *J. Nat. Prod.* **60**: 52-60.
- Craig, M., Lambert, S., Jourdan, S., Tenconi, E., Colson, S., Maciejewska, M., Ongena, M., Martin, J.F., van Wezel, G., and Rigali, S. (2012) Unsuspected control of siderophore production by N-acetylglucosamine in *Streptomyces*. *Environ. Microbiol. Rep.* **4**: 512-521.

- Craney, A., Ozimok, C., Pimentel-Elardo, S.M., Capretta, A., and Nodwell, J.R. (2012) Chemical perturbation of secondary metabolism demonstrates important links to primary metabolism. *Chem. Biol.* **19**: 1020-1027.
- Cruz-Morales, P., Vijgenboom, E., Iruegas-Bocardo, F., Girard, G., Yáñez-Guerra, L.A., Ramos-Aboites, H.E., Pernodet, J.-L., Anné, J., van Wezel, G.P., and Barona-Gómez, F. (2013) The genome sequence of *Streptomyces lividans* 66 reveals a novel tRNA-dependent peptide biosynthetic system within a metal-related genomic island. *Genome Biol. Evol.* **5**: 1165-1175.
- Dale, R.K., Pedersen, B.S., and Quinlan, A.R. (2011) Pybedtools: A flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics* **27**: 3423-3424.
- Dejong, C.A., Chen, G.M., Li, H., Johnston, C.W., Edwards, M.R., Rees, P.N., Skinnider, M.A., Webster, A.L.H., and Magarvey, N.A. (2016) Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching. *Nat. Chem. Biol.* **12**: 1007.
- del Olmo, A., Calzada, J., and Nuñez, M. (2017) Benzoic acid and its derivatives as naturally occurring compounds in foods and as additives: Uses, exposure, and controversy. *Crit. Rev. Food Sci. Nutr.* **57**: 3084-3103.
- Dillon, S.C., and Dorman, C.J. (2010) Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.* **8**: 185-195.
- Distler, U., Kuharev, J., Navarro, P., Levin, Y., Schild, H., and Tenzer, S. (2014) Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nat. Methods* **11**: 167-170.
- Dorman, C.J., Schumacher, M.A., Bush, M.J., Brennan, R.G., and Buttner, M.J. (2020) When is a transcription factor a NAP? *Curr. Opin. Microbiol.* **55**: 26-33.
- Du, C., and van Wezel, G.P. (2018) Mining for microbial gems: Integrating proteomics in the postgenomic natural product discovery pipeline. *Proteomics* **18**: 1700332.
- Du, D., Zhu, Y., Wei, J., Tian, Y., Niu, G., and Tan, H. (2013) Improvement of gougierotin and nikkomycin production by engineering their biosynthetic gene clusters. *Appl. Microbiol. Biotechnol.* **97**: 6383-6396.
- Duncan, M.W., Aebersold, R., and Caprioli, R.M. (2010) The pros and cons of peptide-centric proteomics. *Nat. Biotechnol.* **28**: 659-664.
- Eddy, S.R. (2011) Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**: e1002195.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., Sonnhammer, E.L.L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S.C.E., and Finn, R.D. (2018) The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**: D427-D432.
- Evans, B.S., Ntai, I., Chen, Y., Robinson, S.J., and Kelleher, N.L. (2011) Proteomics-based discovery of koranimine, a cyclic imine natural product. *J. Am. Chem. Soc.* **133**: 7316-7319.
- Evans, M.J., and Cravatt, B.F. (2006) Mechanism-based profiling of enzyme families. *Chem. Rev.* **106**: 3279-3301.
- Farmer, E.E., Alméras, E., and Krishnamurthy, V. (2003) Jasmonates and related oxylipins in plant responses to pathogenesis and herbivory. *Curr. Opin. Plant Biol.* **6**: 372-378.

- Fedorova, N.D., Moktali, V., and Medema, M.H., (2012) Bioinformatics approaches and software for detection of secondary metabolic gene clusters. In: *Fungal secondary metabolism*. Springer, pp. 23-45.
- Fedoryshyn, M., Welle, E., Bechthold, A., and Luzhetskyy, A. (2008) Functional expression of the Cre recombinase in Actinomycetes. *Appl. Microbiol. Biotechnol.* **78**: 1065-1070.
- Ferreira, R.M.B., and Teixeira, A.R.N., (2003) Amino acids - Metabolism. In: *Encyclopedia of Food Sciences and Nutrition* (Second Edition). B. Caballero (ed). Oxford: Academic Press, pp. 197-206.
- Fink, D., Weißschuh, N., Reuther, J., Wohlleben, W., and Engels, A. (2002) Two transcriptional regulators GlnR and GlnRII are involved in regulation of nitrogen metabolism in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **46**: 331-347.
- Fischbach, M.A., and Walsh, C.T. (2006) Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **106**: 3468-3496.
- Flärdh, K., and Buttner, M.J. (2009) *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat. Rev. Microbiol.* **7**: 36-49.
- Fleming, A. (1929) On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*. *Br. J. Exp. Pathol.* **10**: 226-236.
- Flett, F., Mersinias, V., and Smith, C.P. (1997) High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting Streptomyces. *FEMS Microbiol. Lett.* **155**: 223-229.
- Flissi, A., Dufresne, Y., Michalik, J., Tonon, L., Janot, S., Noé, L., Jacques, P., Leclère, V., and Pupin, M. (2016) Norine, the knowledgebase dedicated to non-ribosomal peptides, is now open to crowdsourcing. *Nucleic Acids Res.* **44**: D1113-D1118.
- Floriano, B., and Bibb, M. (1996) *afsR* is a pleiotropic but conditionally required regulatory gene for antibiotic production in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **21**: 385-396.
- Francis, I.M., Jourdan, S., Fanara, S., Loria, R., and Rigali, S. (2015) The cellobiose sensor CebR is the gatekeeper of *Streptomyces scabies* pathogenicity. *mBio* **6**: e02018-02014.
- Fu, P., Jamison, M., La, S., and MacMillan, J.B. (2014) Inducamides A–C, chlorinated alkaloids from an RNA polymerase mutant strain of *Streptomyces* sp. *Org. Lett.* **16**: 5656-5659.
- Gao, W., Sun, H.X., Xiao, H., Cui, G., Hillwig, M.L., Jackson, A., Wang, X., Shen, Y., Zhao, N., Zhang, L., Wang, X.J., Peters, R.J., and Huang, L. (2014) Combining metabolomics and transcriptomics to characterize tanshinone biosynthesis in *Salvia miltiorrhiza*. *BMC Genom.* **15**: 73.
- Gehrke, E.J., Zhang, X., Pimentel-Elardo, S.M., Johnson, A.R., Rees, C.A., Jones, S.E., Hindra, Gehrke, S.S., Turvey, S., Boursalie, S., Hill, J.E., Carlson, E.E., Nodwell, J.R., and Elliot, M.A. (2019) Silencing cryptic specialized metabolism in *Streptomyces* by the nucleoid-associated protein Lsr2. *eLife* **8**: e47691.
- Genilloud, O., González, I., Salazar, O., Martín, J., Tormo, J.R., and Vicente, F. (2011) Current approaches to exploit Actinomycetes as a source of novel natural products. *J. Ind. Microbiol. Biotechnol.* **38**: 375-389.



- Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison Iii, C.A., and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**: 343.
- Girard, G., Willemse, J., Zhu, H., Claessen, D., Bukarasam, K., Goodfellow, M., and van Wezel, G.P. (2014) Analysis of novel *kitasatosporae* reveals significant evolutionary changes in conserved developmental genes between *Kitasatospora* and *Streptomyces*. *Antonie van Leeuwenhoek* **106**: 365-380.
- Gomez-Escribano, J.P., and Bibb, M.J. (2011) Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechnol.* **4**: 207-215.
- Gomez-Escribano, J.P., and Bibb, M.J. (2014) Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: From genome mining to manipulation of biosynthetic pathways. *J. Ind. Microbiol. Biotechnol.* **41**: 425-431.
- González-Cerón, G., Miranda-Olivares, O.J., and Servín-González, L. (2009) Characterization of the methyl-specific restriction system of *Streptomyces coelicolor* A3(2) and of the role played by laterally acquired nucleases. *FEMS Microbiol. Lett.* **301**: 35-43.
- Gopalakrishnan, S., Pande, S., Sharma, M., Humayun, P., Kiran, B.K., Sandeep, D., Vidya, M.S., Deepthi, K., and Rupela, O. (2011) Evaluation of Actinomycete isolates obtained from herbal vermicompost for the biological control of *Fusarium* wilt of chickpea. *Crop Protect.* **30**: 1070-1078.
- Graefe, U., and Radics, L. (1986) Isolation and structure elucidation of 6-(3'-methylbuten-2'-yl)isatin, an unusual metabolite from *Streptomyces albus*. *J. Antibiot.* **39**: 162-163.
- Grkovic, T., Pouwer, R.H., Vial, M.-L., Gambini, L., Noël, A., Hooper, J.N.a., Wood, S.a., Mellick, G.D., and Quinn, R.J. (2014) NMR fingerprints of the drug-like Natural-Product space identify irochotazine A: A chemical probe to study parkinson's disease. *Angew. Chem. Int. Ed. Engl.* **53**: 6070-6074.
- Gubbens, J., Janus, M., Florea, B.I., Overkleeft, H.S., and van Wezel, G.P. (2012) Identification of glucose kinase-dependent and -independent pathways for carbon control of primary metabolism, development and antibiotic production in *Streptomyces coelicolor* by quantitative proteomics. *Mol. Microbiol.* **86**: 1490-1507.
- Gubbens, J., Wu, C., Zhu, H., Filippov, D.V., Florea, B.I., Rigali, S., Overkleeft, H.S., and van Wezel, G.P. (2017) Intertwined precursor supply during biosynthesis of the catecholate-hydroxamate siderophores qinichelins in *Streptomyces* sp. MBT76. *ACS Chem. Biol.*: 2756-2766.
- Gubbens, J., Zhu, H., Girard, G., Song, L., Florea, Bogdan I., Aston, P., Ichinose, K., Filippov, Dmitri V., Choi, Young H., Overkleeft, Herman S., Challis, Gregory L., and van Wezel, Gilles P. (2014) Natural product proteomining, a quantitative proteomics platform, allows rapid discovery of biosynthetic gene clusters for different classes of natural products. *Chem. Biol.* **21**: 707-718.
- Hadwiger, L.A. (2013) Multiple effects of chitosan on plant systems: solid science or hype. *Plant Sci.* **208**: 42-49.

- Heath, J.D., Nester, E.W., and Charles, T.C., (1995) Ti plasmid and chromosomally encoded two-component systems important in plant cell transformation by *Agrobacterium* species. In: Two-component signal transduction. American Society of Microbiology, pp. 367-385.
- Henke, M.T., Soukup, A.A., Goering, A.W., McClure, R.A., Thomson, R.J., Keller, N.P., and Kelleher, N.L. (2016) New aspercryptins, lipopeptide natural products, revealed by HDAC inhibition in *Aspergillus nidulans*. *ACS Chem. Biol.* **11**: 2117-2123.
- Hertweck, C. (2009) The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed. Engl.* **48**: 4688-4716.
- Hiard, S., Marée, R., Colson, S., Hoskisson, P.A., Titgemeyer, F., van Wezel, G.P., Joris, B., Wehenkel, L., and Sébastien, R. (2007) PREDetector: A new tool to identify regulatory elements in bacterial genomes. *Biochem. Biophys. Res. Commun.* **357**: 861-864.
- Hirano, S., Tanaka, K., Ohnishi, Y., and Horinouchi, S. (2008) Conditionally positive effect of the TetR-family transcriptional regulator AtrA on streptomycin production by *Streptomyces griseus*. *Microbiology* **154**: 905-914.
- Hojati, Z., Milne, C., Harvey, B., Gordon, L., Borg, M., Flett, F., Wilkinson, B., Sidebottom, P.J., Rudd, B.A.M., Hayes, M.A., Smith, C.P., and Micklefield, J. (2002) Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. *Chem. Biol.* **9**: 1175-1187.
- Hopkins, W.G., (1999) *Introduction to plant physiology*. John Wiley & Sons, Ltd.
- Hopwood, D.A., (2007) *Streptomyces in nature and medicine: the antibiotic makers*. Oxford University Press.
- Hopwood, D.A., and Wright, H.M. (1983) CDA is a new chromosomally-determined antibiotic from *Streptomyces coelicolor* A3(2). *Microbiology* **129**: 3575-3579.
- Horinouchi, S. (2003) AfsR as an integrator of signals that are sensed by multiple serine/threonine kinases in *Streptomyces coelicolor* A3(2). *J. Ind. Microbiol. Biotechnol.* **30**: 462-467.
- Hosaka, T., Ohnishi-Kameyama, M., Muramatsu, H., Murakami, K., Tsurumi, Y., Kodani, S., Yoshida, M., Fujie, A., and Ochi, K. (2009) Antibacterial discovery in Actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. *Nat. Biotechnol.* **27**: 462.
- Hoskisson, P.A., and van Wezel, G.P. (2019) *Streptomyces coelicolor*. *Trends Microbiol.* **27**: 468-469.
- Hosoya, Y., Okamoto, S., Muramatsu, H., and Ochi, K. (1998) Acquisition of certain streptomycin-resistant (str) mutations enhances antibiotic production in bacteria. *Antimicrob. Agents Chemother.* **42**: 2041-2047.
- Hsiao, N.-H., Söding, J., Linke, D., Lange, C., Hertweck, C., Wohlleben, W., and Takano, E. (2007) ScbA from *Streptomyces coelicolor* A3(2) has homology to fatty acid synthases and is able to synthesize  $\gamma$ -butyrolactones. *Microbiology* **153**: 1394-1404.
- Hu, H., and Ochi, K. (2001) Novel approach for improving the productivity of antibiotic-producing strains by inducing combined resistant mutations. *Appl. Environ. Microbiol.* **67**: 1885-1892.

- Hu, H., Zhang, Q., and Ochi, K. (2002) Activation of antibiotic biosynthesis by specified mutations in the *rpoB* gene (encoding the rna polymerase  $\beta$  subunit) of *Streptomyces lividans*. *J. Bacteriol.* **184**: 3984.
- Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**: 44-57.
- Hunt, A.C., Servin-Gonzalez, L., Kelemen, G.H., and Buttner, M.J. (2005) The *bldC* developmental locus of *Streptomyces coelicolor* encodes a member of a family of small DNA-binding proteins related to the DNA-binding domains of the MerR family. *J. Bacteriol.* **187**: 716-728.
- Hwang, K.-S., Kim, H.U., Charusanti, P., Palsson, B.Ø., and Lee, S.Y. (2014) Systems biology and biotechnology of *Streptomyces* species for the production of secondary metabolites. *Biotechnol. Adv.* **32**: 255-268.
- Ibrahim, A., Yang, L., Johnston, C., Liu, X., Ma, B., and Magarvey, N.A. (2012) Dereplicating nonribosomal peptides using an informatic search algorithm for natural products (iSNAP) discovery. *Proc. Natl. Acad. Sci. USA* **109**: 19196-19201.
- Ichinose, K., Bedford, D.J., Tornus, D., Bechthold, A., Bibb, M.J., Peter Revill, W., Floss, H.G., and Hopwood, D.A. (1998) The granaticin biosynthetic gene cluster of *Streptomyces violaceoruber* Tü22: Sequence analysis and expression in a heterologous host. *Chem. Biol.* **5**: 647-659.
- Ichinose, K., Ozawa, M., Itou, K., Kunieda, K., and Ebizuka, Y. (2003) Cloning, sequencing and heterologous expression of the medermycin biosynthetic gene cluster of *Streptomyces* sp. AM-7161: Towards comparative analysis of the benzoisochromanquinone gene clusters. *Microbiology* **149**: 1633-1645.
- Ikedo, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., and Omura, S. (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat. Biotechnol.* **21**: 526-531.
- Ilmén, M., Saloheimo, A., Onnela, M.L., and Penttilä, M.E. (1997) Regulation of cellulase gene expression in the filamentous fungus *Trichoderma reesei*. *Appl. Environ. Microbiol.* **63**: 1298.
- Jin, J.B., Cai, B., and Zhou, J.-M. (2017) Salicylic acid. In: Hormone Metabolism and Signaling in Plants. J. Li, C. Li & S.M. Smith (eds). Academic Press, pp. 273-289.
- Jourdan, S., Francis, I.M., Deflandre, B., Tenconi, E., Riley, J., Planckaert, S., Tocquin, P., Martinet, L., Devreese, B., Loria, R., and Rigali, S. (2018) Contribution of the  $\beta$ -glucosidase BglC to the onset of the pathogenic lifestyle of *Streptomyces scabies*. *Mol. Plant Pathol.* **19**: 1480-1490.
- Kalogeraki, V.S., Zhu, J., Eberhard, A., Madsen, E.L., and Winans, S.C. (1999) The phenolic *vir* gene inducer ferulic acid is O-demethylated by the VirH2 protein of an *Agrobacterium tumefaciens* Ti plasmid. *Mol. Microbiol.* **34**: 512-522.
- Keijsers, B.J., Noens, E.E., Kraal, B., Koerten, H.K., and van Wezel, G.P. (2003) The *Streptomyces coelicolor* *ssgB* gene is required for early stages of sporulation. *FEMS Microbiol. Lett.* **225**: 59-67.
- Kersten, R.D., Yang, Y.L., Xu, Y., Cimerancic, P., Nam, S.J., Fenical, W., Fischbach, M.A., Moore, B.S., and Dorrestein, P.C. (2011) A mass spectrometry-guided

- genome mining approach for natural product peptidogenomics. *Nat. Chem. Biol.* **7**: 794-802.
- Kersten, R.D., Ziemert, N., Gonzalez, D.J., Duggan, B.M., Nizet, V., Dorrestein, P.C., and Moore, B.S. (2013) Glycogenomics as a mass spectrometry-guided genome-mining method for microbial glycosylated molecules. *Proc. Natl. Acad. Sci. USA* **110**: E4407-E4416.
- Khodakaramian, G., Lissenden, S., Gust, B., Moir, L., Hoskisson, P.A., Chater, K.F., and Smith, M.C.M. (2006) Expression of Cre recombinase during transient phage infection permits efficient marker removal in *Streptomyces*. *Nucleic Acids Res.* **34**: e20-e20.
- Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A., (2000) *Practical Streptomyces genetics*. The John Innes Foundation, Norwich, United Kingdom.
- Kieser, T., and Hopwood, D.A., (1991) Genetic manipulation of *Streptomyces*: integrating vectors and gene replacement. In: *Methods Enzymol.*: Academic Press, pp. 430-458.
- Kim, H.K., Choi, Y.H., and Verpoorte, R. (2010) NMR-based metabolomic analysis of plants. *Nat. Protoc.* **5**: 536-549.
- Kim, M., Sun, G., Lee, D.Y., and Kim, B.G. (2017) BeReTa: A systematic method for identifying target transcriptional regulators to enhance microbial production of chemicals. *Bioinformatics* **33**: 87-94.
- Kim, M., Yi, J.S., Lakshmanan, M., Lee, D.-Y., and Kim, B.-G. (2016) Transcriptomics-based strain optimization tool for designing secondary metabolite overproducing strains of *Streptomyces coelicolor*. *Biotechnol. Bioeng.* **113**: 651-660.
- Kim, S., Traag, B., Hasan, A., McDowall, K., Kim, B.-G., and van Wezel, G.P. (2015) Transcriptional analysis of the cell division-related *ssg* genes in *Streptomyces coelicolor* reveals direct control of *ssgR* by AtrA. *Antonie van Leeuwenhoek* **108**: 201-213.
- Kirkpatrick, C.L., Broberg, C.A., McCool, E.N., Lee, W.J., Chao, A., McConnell, E.W., Pritchard, D.A., Hebert, M., Fleeman, R., Adams, J., Jamil, A., Madera, L., Stromstedt, A.A., Goransson, U., Liu, Y., Hoskin, D.W., Shaw, L.N., and Hicks, L.M. (2017) The "PepSAVI-MS" pipeline for natural product bioactive peptide discovery. *Anal. Chem.* **89**: 1194-1201.
- Kolter, R., and van Wezel, G.P. (2016) Goodbye to brute force in antibiotic discovery? *Nat. Microbiol.* **1**: 15020.
- Komatsu, M., Komatsu, K., Koiwai, H., Yamada, Y., Kozono, I., Izumikawa, M., Hashimoto, J., Takagi, M., Omura, S., Shin-ya, K., Cane, D.E., and Ikeda, H. (2013) Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites. *ACS Synth. Biol.* **2**: 384-396.
- Komatsu, M., Uchiyama, T., Omura, S., Cane, D.E., and Ikeda, H. (2010) Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. USA* **107**: 2646-2651.
- Konno, S., Ishikawa, F., Suzuki, T., Dohmae, N., Burkart, M.D., and Kakeya, H. (2015) Active site-directed proteomic probes for adenylation domains in nonribosomal peptide synthetases. *Chem. Commun. (Camb.)* **51**: 2262-2265.

- Krug, D., and Muller, R. (2014) Secondary metabolomics: The impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products. *Nat. Prod. Rep.* **31**: 768-783.
- La Clair, J.J., Foley, T.L., Schegg, T.R., Regan, C.M., and Burkart, M.D. (2004) Manipulation of carrier proteins in antibiotic biosynthesis. *Chem. Biol.* **11**: 195-201.
- Lai, J.R., Koglin, A., and Walsh, C.T. (2006) Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis. *Biochemistry* **45**: 14869-14879.
- Langmead, B., and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**: 357.
- Larson, J., and Hershberger, C. (1986) The minimal replicon of a streptomycete plasmid produces an ultrahigh level of plasmid DNA. *Plasmid* **15**: 199-209.
- Law, K.P., and Lim, Y.P. (2013) Recent advances in mass spectrometry: data independent analysis and hyper reaction monitoring. *Expert Rev. Proteomics* **10**: 551-566.
- Lawlor, E.J., Baylis, H.A., and Chater, K.F. (1987) Pleiotropic morphological and antibiotic deficiencies result from mutations in a gene encoding a tRNA-like product in *Streptomyces coelicolor* A3(2). *Genes Dev.* **1**: 1305-1310.
- Lebeis, S.L., Paredes, S.H., Lundberg, D.S., Breakfield, N., Gehring, J., McDonald, M., Malfatti, S., Glavina del Rio, T., Jones, C.D., Tringe, S.G., and Dangl, J.L. (2015) Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* **349**: 860.
- Leskiw, B.K., Lawlor, E.J., Fernandez-Abalos, J.M., and Chater, K.F. (1991) TTA codons in some genes prevent their expression in a class of developmental, antibiotic-negative, *Streptomyces* mutants. *Proc. Natl. Acad. Sci. USA* **88**: 2461-2465.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078-2079.
- Li, N., Kuo, C.-L., Paniagua, G., van den Elst, H., Verdoes, M., Willems, L.I., van der Linden, W.A., Ruben, M., van Genderen, E., Gubbens, J., van Wezel, G.P., Overkleeft, H.S., and Florea, B.I. (2013) Relative quantification of proteasome activity by activity-based protein profiling and LC-MS/MS. *Nat. Protoc.* **8**: 1155-1168.
- Li, W., Wu, J., Tao, W., Zhao, C., Wang, Y., He, X., Chandra, G., Zhou, X., Deng, Z., Chater, K.F., and Tao, M. (2007) A genetic and bioinformatic analysis of *Streptomyces coelicolor* genes containing TTA codons, possible targets for regulation by a developmentally significant tRNA. *FEMS Microbiol. Lett.* **266**: 20-28.
- Li, X., Wang, J., Li, S.S., Ji, J.J., Wang, W.S., and Yang, K.Q. (2015) ScbR- and ScbR2-mediated signal transduction networks coordinate complex physiological responses in *Streptomyces coelicolor*. *Sci. Rep.* **5**.
- Liao, C., Rigali, S., Cassani, C.L., Marcellin, E., Nielsen, L.K., and Ye, B.-C. (2014) Control of chitin and N-acetylglucosamine utilization in *Saccharopolyspora erythraea*. *Microbiology* **160**: 1914-1928.

- Liu, B., Raeth, T., Beuerle, T., and Beerhues, L. (2009) A novel 4-hydroxycoumarin biosynthetic pathway. *Plant Mol. Biol.* **72**: 17.
- Liu, G., Chater, K.F., Chandra, G., Niu, G., and Tan, H. (2013) Molecular regulation of antibiotic biosynthesis in *Streptomyces*. *Microbiol. Mol. Biol. Rev.* **77**: 112-143.
- Liu, G., Ou, H.-Y., Wang, T., Li, L., Tan, H., Zhou, X., Rajakumar, K., Deng, Z., and He, X. (2010) Cleavage of phosphorothioated DNA and methylated DNA by the type IV restriction endonuclease ScoMcrA. *PLoS Genet.* **6**: e1001253.
- Liu, X., Liu, D., Xu, M., Tao, M., Bai, L., Deng, Z., Pfeifer, B.A., and Jiang, M. (2018) Reconstitution of kinamycin biosynthesis within the heterologous host *Streptomyces albus* J1074. *J. Nat. Prod.* **81**: 72-77.
- Lomovskaya, O., and Lewis, K. (1992) *emr*, an *Escherichia coli* locus for multidrug resistance. *Proc. Natl. Acad. Sci. USA* **89**: 8938.
- Loureiro, C., Medema, M.H., van der Oost, J., and Sipkema, D. (2018) Exploration and exploitation of the environment for novel specialized metabolites. *Curr. Opin. Biotechnol.* **50**: 206-213.
- Love, M.I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**.
- Lukashin, A.V., and Borodovsky, M. (1998) GeneMark.hmm: New solutions for gene finding. *Nucleic Acids Res.* **26**: 1107-1115.
- Machanick, P., and Bailey, T.L. (2011) MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics* **27**: 1696-1697.
- MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H., and MacNeil, T. (1992) Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111**: 61-68.
- Mahr, K., van Wezel, G.P., Svensson, C., Krengel, U., Bibb, M.J., and Titgemeyer, F. (2000) Glucose kinase of *Streptomyces coelicolor* A3(2): Large-scale purification and biochemical analysis. *Antonie van Leeuwenhoek* **78**: 253-261.
- Manteca, Á., Fernández, M., and Sánchez, J. (2005) A death round affecting a young compartmentalized mycelium precedes aerial mycelium dismantling in confluent surface cultures of *Streptomyces antibioticus*. *Microbiology* **151**: 3689-3697.
- Mao, X.M., Xu, W., Li, D., Yin, W.B., Chooi, Y.H., Li, Y.Q., Tang, Y., and Hu, Y. (2015) Epigenetic genome mining of an endophytic fungus leads to the pleiotropic biosynthesis of natural products. *Angew. Chem. Int. Ed. Engl.* **54**: 7592-7596.
- Martín, J.F. (2004) Phosphate control of the biosynthesis of antibiotics and other secondary metabolites is mediated by the PhoR-PhoP system: An unfinished story. *J. Bacteriol.* **186**: 5197.
- Martín, J.F., Casqueiro, J., and Liras, P. (2005) Secretion systems for secondary metabolites: How producer cells send out messages of intercellular communication. *Curr. Opin. Microbiol.* **8**: 282-293.
- Martín, J.F., Santos-Beneit, F., Rodríguez-García, A., Sola-Landa, A., Smith, M.C.M., Ellingsen, T.E., Nieselt, K., Burroughs, N.J., and Wellington, E.M.H. (2012) Transcriptomic studies of phosphate control of primary and secondary metabolism in *Streptomyces coelicolor*. *Appl. Microbiol. Biotechnol.* **95**: 61-75.
- Martín, J.F., Sola-Landa, A., Santos-Beneit, F., Fernández-Martínez, L.T., Prieto, C., and Rodríguez-García, A. (2011) Cross-talk of global nutritional regulators in the

- control of primary and secondary metabolism in *Streptomyces*. *Microb. Biotechnol.* **4**: 165-174.
- Marushima, K., Ohnishi, Y., and Horinouchi, S. (2009) CebR as a master regulator for cellulose/cellooligosaccharide catabolism affects morphological development in *Streptomyces griseus*. *J. Bacteriol.* **191**: 5930-5940.
- May, M.S., and Hattman, S. (1975) Analysis of bacteriophage deoxyribonucleic acid sequences methylated by host- and R-factor-controlled enzymes. *J. Bacteriol.* **123**: 768-770.
- Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M.A., Weber, T., Takano, E., and Breitling, R. (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39**: W339-W346.
- Medema, M.H., and Fischbach, M.A. (2015) Computational approaches to natural product discovery. *Nat. Chem. Biol.* **11**: 639-648.
- Medema, M.H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J.B., Blin, K., de Bruijn, I., Chooi, Y.H., Claesen, J., Coates, R.C., Cruz-Morales, P., Duddela, S., D sterhus, S., Edwards, D.J., Fewer, D.P., Garg, N., Geiger, C., Gomez-Escribano, J.P., Greule, A., Hadjithomas, M., Haines, A.S., Helfrich, E.J.N., Hillwig, M.L., Ishida, K., Jones, A.C., Jones, C.S., Jungmann, K., Kegler, C., Kim, H.U., K tter, P., Krug, D., Masschelein, J., Melnik, A.V., Mantovani, S.M., Monroe, E.A., Moore, M., Moss, N., N tzmann, H.-W., Pan, G., Pati, A., Petras, D., Reen, F.J., Rosconi, F., Rui, Z., Tian, Z., Tobias, N.J., Tsunematsu, Y., Wiemann, P., Wyckoff, E., Yan, X., Yim, G., Yu, F., Xie, Y., Aigle, B., Apel, A.K., Balibar, C.J., Balskus, E.P., Barona-G mez, F., Bechthold, A., Bode, H.B., Borriss, R., Brady, S.F., Brakhage, A.A., Caffrey, P., Cheng, Y.-Q., Clardy, J., Cox, R.J., De Mot, R., Donadio, S., Donia, M.S., van der Donk, W.A., Dorrestein, P.C., Doyle, S., Driessen, A.J.M., Ehling-Schulz, M., Entian, K.-D., Fischbach, M.A., Gerwick, L., Gerwick, W.H., Gross, H., Gust, B., Hertweck, C., H fte, M., Jensen, S.E., Ju, J., Katz, L., Kaysser, L., Klassen, J.L., Keller, N.P., Kormanec, J., Kuipers, O.P., Kuzuyama, T., Kyrpides, N.C., Kwon, H.-J., Lautru, S., Lavigne, R., Lee, C.Y., Linqun, B., Liu, X., Liu, W., *et al.* (2015) Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* **11**: 625.
- Meier, J.L., Mercer, A.C., and Burkart, M.D. (2008) Fluorescent profiling of modular biosynthetic enzymes by complementary metabolic and activity based probes. *J. Am. Chem. Soc.* **130**: 5443-5445.
- Meier, J.L., Mercer, A.C., Rivera, H., and Burkart, M.D. (2006) Synthesis and evaluation of bioorthogonal pantetheine analogues for in vivo protein modification. *J. Am. Chem. Soc.* **128**: 12174-12184.
- Meier, J.L., Niessen, S., Hoover, H.S., Foley, T.L., Cravatt, B.F., and Burkart, M.D. (2009) An orthogonal active site identification system (OASIS) for proteomic profiling of natural product biosynthesis. *ACS Chem. Biol.* **4**: 948-957.
- Meluzzi, D., Zheng, W.H., Hensler, M., Nizet, V., and Dorrestein, P.C. (2008) Top-down mass spectrometry on low-resolution instruments: Characterization of phosphopantetheinylated carrier domains in polyketide and non-ribosomal biosynthetic pathways. *Bioorg. Med. Chem. Lett.* **18**: 3107-3111.

- Metsä-Ketelä, M., Oja, T., Taguchi, T., Okamoto, S., and Ichinose, K. (2013) Biosynthesis of pyranonaphthoquinone polyketides reveals diverse strategies for enzymatic carbon-carbon bond formation. *Curr. Opin. Chem. Biol.* **17**: 562-570.
- Mohimani, H., Gurevich, A., Mikheenko, A., Garg, N., Nothias, L.-F., Ninomiya, A., Takada, K., Dorrestein, P.C., and Pevzner, P.A. (2016) Dereplication of peptidic natural products through database search of mass spectra. *Nat. Chem. Biol.* **13**: 30.
- Mohimani, H., and Pevzner, P.A. (2016) Dereplication, sequencing and identification of peptidic natural products: From genome mining to peptidogenomics to spectral networks. *Nat. Prod. Rep.* **33**: 73-86.
- Motamedi, H., Shafiee, A., and Cai, S.-J. (1995) Integrative vectors for heterologous gene expression in *Streptomyces* spp. *Gene* **160**: 25-31.
- Myronovskyi, M., and Luzhetskyy, A. (2019) Heterologous production of small molecules in the optimized *Streptomyces* hosts. *Nat. Prod. Rep.* **36**: 1281-1294.
- Nazari, B., Kobayashi, M., Saito, A., Hassaninasab, A., Miyashita, K., and Fujii, T. (2012) Chitin-induced gene expression involved in secondary metabolic pathways in *Streptomyces coelicolor* A3(2) grown in soil. *Appl. Environ. Microbiol.* **79**: 707-713.
- Nepal, K.K., and Wang, G. (2019) Streptomycetes: Surrogate hosts for the genetic manipulation of biosynthetic gene clusters and production of natural products. *Biotechnol. Adv.* **37**: 1-20.
- Newman, D.J., and Cragg, G.M. (2007) Natural products as sources of new drugs over the last 25 years. *J. Nat. Prod.* **70**: 461-477.
- Newman, D.J., and Cragg, G.M. (2016) Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* **79**: 629-661.
- Ng, J., Bandeira, N., Liu, W.T., Ghassemian, M., Simmons, T.L., Gerwick, W.H., Lington, R., Dorrestein, P.C., and Pevzner, P.A. (2009) Dereplication and de novo sequencing of nonribosomal peptides. *Nat. Methods* **6**: 596-599.
- Nicholson, J., and Lindon, J. (2008) Systems biology: metabonomics. *Nature* **455**: 1054-1056.
- Nieselt, K., Battke, F., Herbig, A., Bruheim, P., Wentzel, A., Jakobsen, O.M., Sletta, H., Alam, M.T., Merlo, M.E., Moore, J., Omara, W.A., Morrissey, E.R., Juarez-Hermosillo, M.A., Rodriguez-Garcia, A., Nentwich, M., Thomas, L., Iqbal, M., Legaie, R., Gaze, W.H., Challis, G.L., Jansen, R.C., Dijkhuizen, L., Rand, D.A., Wild, D.L., Bonin, M., Reuther, J., Wohlleben, W., Smith, M.C., Burroughs, N.J., Martin, J.F., Hodgson, D.A., Takano, E., Breitling, R., Ellingsen, T.E., and Wellington, E.M. (2010) The dynamic architecture of the metabolic switch in *Streptomyces coelicolor*. *BMC Genom.* **11**: 10.
- Nothhaft, H., Rigali, S., Boomsma, B., Świątek, M., McDowall, K.J., Van Wezel, G.P., and Titgemeyer, F. (2010) The permease gene nagE2 is the key to N-acetylglucosamine sensing and utilization in *Streptomyces coelicolor* and is subject to multi-level control. *Mol. Microbiol.* **75**: 1133-1144.
- Novakova, R., Homerova, D., Feckova, L., and Kormanec, J. (2005) Characterization of a regulatory gene essential for the production of the angucycline-like polyketide antibiotic auricin in *Streptomyces aureofaciens* CCM 3239. *Microbiology* **151**: 2693-2706.



- Novakova, R., Kutas, P., Feckova, L., and Kormanec, J. (2010) The role of the TetR-family transcriptional regulator Aur1R in negative regulation of the auricin gene cluster in *Streptomyces aureofaciens* CCM 3239. *Microbiology* **156**: 2374-2383.
- Novakova, R., Rehakova, A., Kutas, P., Feckova, L., and Kormanec, J. (2011) The role of two SARP family transcriptional regulators in regulation of the auricin gene cluster in *Streptomyces aureofaciens* CCM 3239. *Microbiology* **157**: 1629-1639.
- Ochi, K., Tanaka, Y., and Tojo, S. (2014) Activating the expression of bacterial cryptic genes by *rpoB* mutations in RNA polymerase or by rare earth elements. *J. Ind. Microbiol. Biotechnol.* **41**: 403-414.
- Ohnishi, Y., Ishikawa, J., Hara, H., Suzuki, H., Ikenoya, M., Ikeda, H., Yamashita, A., Hattori, M., and Horinouchi, S. (2008) Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J. Bacteriol.* **190**: 4050-4060.
- Oja, T., Galindo, P.S.M., Taguchi, T., Manner, S., Vuorela, P.M., Ichinose, K., Metsä-Ketelä, M., and Fallarero, A. (2015) Effective antibiofilm polyketides against *Staphylococcus aureus* from the pyranonaphthoquinone biosynthetic pathways of *Streptomyces* species. *Antimicrob. Agents Chemother.* **59**: 6046-6052.
- Okada, B.K., and Seyedsayamdost, M.R. (2017) Antibiotic dialogues: Induction of silent biosynthetic gene clusters by exogenous small molecules. *FEMS Microbiol. Rev.* **41**: 19-33.
- Okamoto, S., Taguchi, T., Ochi, K., and Ichinose, K. (2009) Biosynthesis of actinorhodin and related antibiotics: Discovery of alternative routes for quinone formation encoded in the *act* gene cluster. *Chem. Biol.* **16**: 226-236.
- Oliynyk, M., Samborsky, M., Lester, J.B., Mironenko, T., Scott, N., Dickens, S., Haydock, S.F., and Leadlay, P.F. (2007) Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. *Nat. Biotechnol.* **25**: 447-453.
- Omura, S., Tsuzuki, K., Tanaka, Y., Sakakibara, H., Aizawa, M., and Lukacs, G. (1983) Valine as a precursor of n-butyrate unit in the biosynthesis of macrolide aglycone. *J. Antibiot.* **36**: 614-616.
- Owens, R.A., Hammel, S., Sheridan, K.J., Jones, G.W., and Doyle, S. (2014) A proteomic approach to investigating gene cluster expression and secondary metabolite functionality in *Aspergillus fumigatus*. *PLoS One* **9**: e106942.
- Ozaki, T., Nishiyama, M., and Kuzuyama, T. (2013) Novel tryptophan metabolism by a potential gene cluster that is widely distributed among Actinomycetes. *J. Biol. Chem.* **288**: 9946-9956.
- Paget, M.S.B., Chamberlin, L., Atrih, A., Foster, S.J., and Buttner, M.J. (1999) Evidence that the extracytoplasmic function sigma factor  $\sigma^E$  is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **181**: 204-211.
- Palmer, J.M., and Keller, N.P. (2010) Secondary metabolism in fungi: Does chromosomal location matter? *Curr. Opin. Microbiol.* **13**: 431-436.
- Park, J.W., Park, S.R., Nepal, K.K., Han, A.R., Ban, Y.H., Yoo, Y.J., Kim, E.J., Kim, E.M., Kim, D., Sohng, J.K., and Yoon, Y.J. (2011) Discovery of parallel pathways of kanamycin biosynthesis allows antibiotic manipulation. *Nat. Chem. Biol.* **7**: 843-852.

- Patel, V.J., Thalassinou, K., Slade, S.E., Connolly, J.B., Crombie, A., Murrell, J.C., and Scrivens, J.H. (2009) A comparison of labeling and label-free mass spectrometry-based proteomics approaches. *J. Proteome Res.* **8**: 3752-3759.
- Payne, D.J., Gwynn, M.N., Holmes, D.J., and Pompliano, D.L. (2007) Drugs for bad bugs: Confronting the challenges of antibacterial discovery. *Nat. Rev. Drug Discovery* **6**: 29-40.
- Pelletier, B., Beaudoin, J., Philpott, C.C., and Labbé, S. (2003) Fep1 represses expression of the fission yeast *Schizosaccharomyces pombe* siderophore-iron transport system. *Nucleic Acids Res.* **31**: 4332-4344.
- Perez-Redondo, R., Rodriguez-Garcia, A., Botas, A., Santamarta, I., Martin, J.F., and Liras, P. (2012) ArgR of *Streptomyces coelicolor* as a versatile regulator. *PLoS One* **7**.
- Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, S., Tiwary, S., Cox, J., Audain, E., Walzer, M., Jarnuczak, A.F., Ternent, T., Brazma, A., and Vizcaíno, J.A. (2018) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**: D442-D450.
- Piette, A., Derouaux, A., Gerkens, P., Noens, E.E.E., Mazzucchelli, G., Vion, S., Koerten, H.K., Titgemeyer, F., De Pauw, E., Leprince, P., van Wezel, G.P., Galleni, M., and Rigali, S. (2005) From dormant to germinating spores of *Streptomyces coelicolor* A3(2): New perspectives from the *crp* null mutant. *J. Proteome Res.* **4**: 1699-1708.
- Pimentel-Elardo, S.M., Sorensen, D., Ho, L., Ziko, M., Bueler, S.A., Lu, S., Tao, J., Moser, A., Lee, R., Agard, D., Fairn, G., Rubinstein, J.L., Shoichet, B.K., and Nodwell, J.R. (2015) Activity-independent discovery of secondary metabolites using chemical elicitation and cheminformatic inference. *ACS Chem. Biol.* **10**: 2616-2623.
- Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R., and Finn, R.D. (2018) HMMER web server: 2018 update. *Nucleic Acids Res.* **46**: W200-W204.
- Qaisar, U., Kruczek, C.J., Azeem, M., Javaid, N., Colmer-Hamood, J.A., and Hamood, A.N. (2016) The *Pseudomonas aeruginosa* extracellular secondary metabolite, Paerucumarin, chelates iron and is not localized to extracellular membrane vesicles. *J. Microbiol.* **54**: 573-581.
- Quinlan, A.R. (2014) BEDTools: The swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinformatics* **47**: 11.12.01-11.12.34.
- Ramos, J.L., Martínez-Bueno, M., Molina-Henares, A.J., Terán, W., Watanabe, K., Zhang, X., Gallegos, M.T., Brennan, R., and Tobes, R. (2005) The TetR family of transcriptional repressors. *Microbiol. Mol. Biol. Rev.* **69**: 326-356.
- Rappsilber, J., Mann, M., and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**: 1896-1906.
- Reuther, J., and Wohlleben, W. (2007) Nitrogen metabolism in *Streptomyces coelicolor*: Transcriptional and post-translational regulation. *J. Mol. Microbiol. Biotechnol.* **12**: 139-146.

- Reynolds, K.A., Ohagan, D., Gani, D., and Robinson, J.A. (1988) Butyrate metabolism in *Streptomyces* - characterization of an intramolecular vicinal interchange rearrangement linking isobutyrate and butyrate in *Streptomyces cinnamonensis*. *J. Chem. Soc. [Perkin 1]*: 3195-3207.
- Rigali, S., Anderssen, S., Naômé, A., and van Wezel, G.P. (2018) Cracking the regulatory code of biosynthetic gene clusters as a strategy for natural product discovery. *Biochem. Pharmacol.*: 24-34.
- Rigali, S., Nothaft, H., Noens, E.E.E., Schlicht, M., Colson, S., Müller, M., Joris, B., Koerten, H.K., Hopwood, D.A., Titgemeyer, F., and Van Wezel, G.P. (2006) The sugar phosphotransferase system of *Streptomyces coelicolor* is regulated by the GntR-family regulator DasR and links N-acetylglucosamine metabolism to the control of development. *Mol. Microbiol.* **61**: 1237-1251.
- Rigali, S., Titgemeyer, F., Barends, S., Mulder, S., Thomae, A.W., Hopwood, D.A., and van Wezel, G.P. (2008) Feast or famine: The global regulator DasR links nutrient stress to antibiotic production by *Streptomyces*. *EMBO Rep.* **9**: 670-675.
- Rix, U., and Superti-Furga, G. (2009) Target profiling of small molecules by chemical proteomics. *Nat. Chem. Biol.* **5**: 616-624.
- Rodríguez-García, A., Sola-Landa, A., Apel, K., Santos-Beneit, F., and Martín, J.F. (2009) Phosphate control over nitrogen metabolism in *Streptomyces coelicolor*: Direct and indirect negative control of *glnR*, *glnA*, *glnII* and *amtB* expression by the response regulator PhoP. *Nucleic Acids Res.* **37**: 3230-3242.
- Rokem, J.S., Lantz, A.E., and Nielsen, J. (2007) Systems biology of antibiotic production by microorganisms. *Nat. Prod. Rep.* **24**: 1262-1287.
- Romero, D.A., Hasan, A.H., Lin, Y.-f., Kime, L., Ruiz-Larrabeiti, O., Urem, M., Bucca, G., Mamanova, L., Laing, E.E., van Wezel, G.P., Smith, C.P., Kaberdin, V.R., and McDowall, K.J. (2014) A comparison of key aspects of gene regulation in *Streptomyces coelicolor* and *Escherichia coli* using nucleotide-resolution transcription maps produced in parallel by global and differential RNA sequencing. *Mol. Microbiol.* **94**: 963-987.
- Rottig, M., Medema, M.H., Blin, K., Weber, T., Rausch, C., and Kohlbacher, O. (2011) NRPSpredictor2 – a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **39**: W362-W367.
- Rudd, B.A., and Hopwood, D.A. (1979) Genetics of actinorhodin biosynthesis by *Streptomyces coelicolor* A3(2). *J. Gen. Microbiol.* **114**: 35-43.
- Rudd, B.A., and Hopwood, D.A. (1980) A pigmented mycelial antibiotic in *Streptomyces coelicolor*: Control by a chromosomal gene cluster. *J. Gen. Microbiol.* **119**: 333-340.
- Rutledge, P.J., and Challis, G.L. (2015) Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **13**: 509-523.
- Sadeghi, A., Karimi, E., Dahaji, P.A., Javid, M.G., Dalvand, Y., and Askari, H. (2012) Plant growth promoting activity of an auxin and siderophore producing isolate of *Streptomyces* under saline soil conditions. *World J. Microbiol. Biotechnol.* **28**: 1503-1509.
- Salerno, P., Larsson, J., Bucca, G., Laing, E., Smith, C.P., and Flärdh, K. (2009) One of the two genes encoding nucleoid-associated HU proteins in *Streptomyces*

- coelicolor* is developmentally regulated and specifically involved in spore maturation. *J. Bacteriol.* **191**: 6489-6500.
- Sambrook, J., Fritsch, E.F., and Maniatis, T., (1989) *Molecular cloning: a laboratory manual*. Cold spring harbor laboratory press, New York.
- Sanchez, S., Chavez, A., Forero, A., Garcia-Huante, Y., Romero, A., Sanchez, M., Rocha, D., Sanchez, B., Avalos, M., Guzman-Trampe, S., Rodriguez-Sanoja, R., Langley, E., and Ruiz, B. (2010) Carbon source regulation of antibiotic production. *J. Antibiot.* **63**: 442-459.
- Santos-Beneit, F., Rodríguez-García, A., and Martin, J.F. (2012) Overlapping binding of PhoP and AfsR to the promoter region of *glnR* in *Streptomyces coelicolor*. *Microbiol. Res.* **167**: 532-535.
- Santos-Beneit, F., Rodríguez-García, A., Sola-Landa, A., and Martín, J.F. (2009) Cross-talk between two global regulators in *Streptomyces*: PhoP and AfsR interact in the control of *afsS*, *pstS* and *phoRP* transcription. *Mol. Microbiol.* **72**: 53-68.
- Schley, C., Altmeyer, M.O., Swart, R., Müller, R., and Huber, C.G. (2006) Proteome analysis of *Myxococcus xanthus* by off-line two-dimensional chromatographic separation using monolithic poly-(styrene-divinylbenzene) columns combined with ion-trap tandem mass spectrometry. *J. Proteome Res.* **5**: 2760-2768.
- Schubert, O.T., Röst, H.L., Collins, B.C., Rosenberger, G., and Aebersold, R. (2017) Quantitative proteomics: Challenges and opportunities in basic and applied research. *Nat. Protoc.* **12**: 1289.
- Schumacher, M.A., den Hengst, C.D., Bush, M.J., Le, T.B.K., Tran, N.T., Chandra, G., Zeng, W., Travis, B., Brennan, R.G., and Buttner, M.J. (2018) The MerR-like protein BldC binds DNA direct repeats as cooperative multimers to regulate *Streptomyces* development. *Nat. Commun.* **9**: 1139.
- Schümann, J., and Hertweck, C. (2006) Advances in cloning, functional analysis and heterologous expression of fungal polyketide synthase genes. *J. Biotechnol.* **124**: 690-703.
- Seipke, R.F., Kaltenpoth, M., and Hutchings, M.I. (2012) *Streptomyces* as symbionts: An emerging and widespread theme? *FEMS Microbiol. Rev.* **36**: 862-876.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**: 2498-2504.
- Shao, Z., Deng, W., Li, S., He, J., Ren, S., Huang, W., Lu, Y., Zhao, G., Cai, Z., and Wang, J. (2015) GlnR-mediated regulation of *ectABCD* transcription expands the role of the *glnR* regulon to osmotic stress management. *J. Bacteriol.* **197**: 3041.
- Shimizu, Y., Ogata, H., and Goto, S. (2017) Type III polyketide synthases: Functional classification and phylogenomics. *ChemBioChem* **18**: 50-65.
- Sidda, J.D., Poon, V., Song, L., Wang, W., Yang, K., and Corre, C. (2016) Overproduction and identification of butyrolactones SCB1-8 in the antibiotic production superhost *Streptomyces* M1152. *Org. Biomol. Chem.* **14**: 6390-6393.
- Skinninger, M.A., Johnston, C.W., Edgar, R.E., Dejong, C.A., Merwin, N.J., Rees, P.N., and Magarvey, N.A. (2016) Genomic charting of ribosomally synthesized natural product chemical space facilitates targeted mining. *Proc. Natl. Acad. Sci. USA* **113**: E6343-E6351.

- Sobolevskaya, M.P., Denisenko, V.A., Fotso, S., Laach, H., Menzorova, N.I., Sibirtsev, Y.T., and Kuznetsova, T.A. (2009) Biologically active metabolites of the actinobacterium *Streptomyces* sp. GW 33/1593. *Russ. Chem. Bull.* **57**: 665-668.
- Sola-Landa, A., Moura, R.S., and Martín, J.F. (2003) The two-component PhoR-PhoP system controls both primary metabolism and secondary metabolite biosynthesis in *Streptomyces lividans*. *Proc. Natl. Acad. Sci. USA* **100**: 6133.
- Sola-Landa, A., Rodríguez-García, A., Amin, R., Wohlleben, W., and Martín, J.F. (2012) Competition between the GlnR and PhoP regulators for the *glnA* and *amtB* promoters in *Streptomyces coelicolor*. *Nucleic Acids Res.* **41**: 1767-1782.
- Sola-Landa, A., Rodríguez-García, A., Franco-Domínguez, E., and Martín, J.F. (2005) Binding of PhoP to promoters of phosphate-regulated genes in *Streptomyces coelicolor*: Identification of PHO boxes. *Mol. Microbiol.* **56**: 1373-1385.
- Subramanian, A., Kuehn, H., Gould, J., Tamayo, P., and Mesirov, J.P. (2007) GSEA-P: A desktop application for gene set enrichment analysis. *Bioinformatics* **23**: 3251-3253.
- Sulheim, S., Kumelj, T., van Dissel, D., Salehzadeh-Yazdi, A., Du, C., van Wezel, G.P., Nieselt, K., Almaas, E., Wentzel, A., and Kerkhoven, E.J. (2020) Enzyme-constrained models and omics analysis of *Streptomyces coelicolor* reveal metabolic changes that enhance heterologous production. *iScience*: 101525.
- Sutherland, J.B., Crawford, D.L., and Pometto, A.L. (1981) Catabolism of substituted benzoic acids by *Streptomyces* species. *Appl. Environ. Microbiol.* **41**: 442.
- Sutherland, J.B., Crawford, D.L., and Pometto, A.L. (1983) Metabolism of cinnamic, p-coumaric, and ferulic acids by *Streptomyces setonii*. *Can. J. Microbiol.* **29**: 1253-1257.
- Świątek, M.A., Bucca, G., Laing, E., Gubbens, J., Titgemeyer, F., Smith, C.P., Rigali, S., and van Wezel, G.P. (2015) Genome-wide analysis of in vivo binding of the master regulator DasR in *Streptomyces coelicolor* identifies novel non-canonical targets. *PLoS One* **10**: e0122479.
- Świątek, M.A., Tenconi, E., Rigali, S., and van Wezel, G.P. (2012) Functional analysis of the N-acetylglucosamine metabolic genes of *Streptomyces coelicolor* and role in the control of development and antibiotic production. *J. Bacteriol.* **194**: 1136-1144.
- Swinger, K.K., and Rice, P.A. (2004) IHF and HU: flexible architects of bent DNA. *Curr. Opin. Struct. Biol.* **14**: 28-35.
- Takahashi, S., Takagi, H., Toyoda, A., Uramoto, M., Nogawa, T., Ueki, M., Sakaki, Y., and Osada, H. (2010) Biochemical characterization of a novel indole prenyltransferase from *Streptomyces* sp. SN-593. *J. Bacteriol.* **192**: 2839-2851.
- Takano, E., Chakraborty, R., Nihira, T., Yamada, Y., and Bibb, M.J. (2001) A complex role for the  $\gamma$ -butyrolactone SCB1 in regulating antibiotic production in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **41**: 1015-1028.
- Tamehiro, N., Hosaka, T., Xu, J., Hu, H., Otake, N., and Ochi, K. (2003) Innovative approach for improvement of an antibiotic-overproducing industrial strain of *Streptomyces albus*. *Appl. Environ. Microbiol.* **69**: 6412-6417.
- Tenconi, E., Traxler, M.F., Hoebreck, C., van Wezel, G.P., and Rigali, S. (2018) Production of prodiginines is part of a programmed cell death process in *Streptomyces coelicolor*. *Front. Microbiol.* **9**.

- Thomas, L., Hodgson, D.A., Wentzel, A., Nieselt, K., Ellingsen, T.E., Moore, J., Morrissey, E.R., Legaie, R., Wohlleben, W., Rodríguez-García, A., Martín, J.F., Burroughs, N.J., Wellington, E.M.H., and Smith, M.C.M. (2012) Metabolic switches and adaptations deduced from the proteomes of *Streptomyces coelicolor* wild type and *phoP* mutant grown in batch culture. *Mol. Cell. Proteomics* **11**: M111.013797.
- Tiffert, Y., Supra, P., Wurm, R., Wohlleben, W., Wagner, R., and Reuther, J. (2008) The *Streptomyces coelicolor* GlnR regulon: Identification of new GlnR targets and evidence for a central role of GlnR in nitrogen metabolism in Actinomycetes. *Mol. Microbiol.* **67**: 861-880.
- Tocquin, P., Naome, A., Jourdan, S., Anderssen, S., Hiard, S., van Wezel, G.P., Hanikenne, M., Baurain, D., and Rigali, S. (2016) PREDetector 2.0: Online and enhanced version of the prokaryotic regulatory elements detector tool. *bioRxiv*.
- Tong, Y., Charusanti, P., Zhang, L., Weber, T., and Lee, S.Y. (2015a) CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS Synth. Biol.*: 1020-1029.
- Tong, Y., Charusanti, P., Zhang, L., Weber, T., and Lee, S.Y. (2015b) CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. *ACS Synthetic Biology*.
- Traag, B.A., Kelemen, G.H., and Van Wezel, G.P. (2004) Transcription of the sporulation gene *ssgA* is activated by the IclR-type regulator SsgR in a whi-independent manner in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **53**: 985-1000.
- Tran, N.H., Zhang, X., Xin, L., Shan, B., and Li, M. (2017) De novo peptide sequencing by deep learning. *Proc. Natl. Acad. Sci. USA* **114**: 8247-8252.
- Udwary, D.W., Gontang, E.A., Jones, A.C., Jones, C.S., Schultz, A.W., Winter, J.M., Yang, J.Y., Beauchemin, N., Capson, T.L., Clark, B.R., Esquenazi, E., Eustáquio, A.S., Freel, K., Gerwick, L., Gerwick, W.H., Gonzalez, D., Liu, W.-T., Malloy, K.L., Maloney, K.N., Nett, M., Nunnery, J.K., Penn, K., Prieto-Davo, A., Simmons, T.L., Weitz, S., Wilson, M.C., Tisa, L.S., Dorrestein, P.C., and Moore, B.S. (2011) Significant natural product biosynthetic potential of actinorhizal symbionts of the genus *Frankia*, as revealed by comparative genomic and proteomic analyses. *Appl. Environ. Microbiol.* **77**: 3617-3625.
- Ueda, K., Kawai, S., Ogawa, H.-O., Kiyama, A., Kubota, T., Kawanobe, H., and Beppu, T. (2000) Wide distribution of interspecific stimulatory events on antibiotic production and sporulation among *Streptomyces* species. *J. Antibiot.* **53**: 979-982.
- Uguru, G.C., Stephens, K.E., Stead, J.A., Towle, J.E., Baumberg, S., and McDowall, K.J. (2005) Transcriptional activation of the pathway-specific regulator of the actinorhodin biosynthetic genes in *Streptomyces coelicolor*. *Mol. Microbiol.* **58**: 131-150.
- Urem, M., Świątek, M.A., Rigali, S., and van Wezel, G.P. (2016) Intertwining nutrient-sensory networks and the control of antibiotic production in *Streptomyces*. *Mol. Microbiol.*: 183-195.
- Välikangas, T., Suomi, T., and Elo, L.L. (2016) A systematic evaluation of normalization methods in quantitative label-free proteomics. *Brief. Bioinform.* **19**: 1-11.
- van der Aart, L.T., Nouioui, I., Kloosterman, A., Igual, J.-M., Willemse, J., Goodfellow, M., and van Wezel, G.P. (2019) Polyphasic classification of the gifted natural

- product producer *Streptomyces roseifaciens* sp. nov. *Int. J. Syst. Evol. Microbiol.* **69**: 899-908.
- van der Heul, H.U., Bilyk, B.L., McDowall, K.J., Seipke, R.F., and van Wezel, G.P. (2018) Regulation of antibiotic production in Actinobacteria: New perspectives from the post-genomic era. *Nat. Prod. Rep.* **35**: 575-604.
- van der Meij, A., (2020) Impact of plant hormones on growth and development of Actinobacteria. In: Institute of Biology Leiden. Leiden: Leiden University, pp.
- van der Meij, A., Willemse, J., Schneijderberg, M.A., Geurts, R., Raaijmakers, J.M., and van Wezel, G.P. (2018) Inter- and intracellular colonization of Arabidopsis roots by endophytic actinobacteria and the impact of plant hormones on their antimicrobial activity. *Antonie van Leeuwenhoek* **111**: 679-690.
- van der Meij, A., Worsley, S.F., Hutchings, M.I., and van Wezel, G.P. (2017) Chemical ecology of antibiotic production by Actinomycetes. *FEMS Microbiol. Rev.* **41**: 392-416.
- van Heel, A.J., de Jong, A., Montalban-Lopez, M., Kok, J., and Kuipers, O.P. (2013) BAGEL3: Automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res.* **41**: W448-W453.
- van Keulen, G., and Dyson, P.J., (2014) Chapter Six - Production of specialized metabolites by *Streptomyces coelicolor* A3(2). In: Adv. Appl. Microbiol. S. Sariaslani & G.M. Gadd (eds). Academic Press, pp. 217-266.
- van Wezel, G.P., Krabben, P., Traag, B.A., Keijser, B.J.F., Kerste, R., Vijgenboom, E., Heijnen, J.J., and Kraal, B. (2006) Unlocking *Streptomyces* spp. for use as sustainable industrial production platforms by morphological engineering. *Appl. Environ. Microbiol.* **72**: 5283-5288.
- van Wezel, G.P., and McDowall, K.J. (2011) The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances. *Nat. Prod. Rep.* **28**: 1311-1333.
- van Wezel, G.P., McKenzie, N.L., and Nodwell, J.R., (2009) Chapter 5. Applying the genetics of secondary metabolism in model Actinomycetes to the discovery of new antibiotics. In: Methods Enzymol.: Academic Press, pp. 117-141.
- van Wezel, G.P., van der Meulen, J., Kawamoto, S., Luiten, R.G., Koerten, H.K., and Kraal, B. (2000a) *ssgA* is essential for sporulation of *Streptomyces coelicolor* A3(2) and affects hyphal development by stimulating septum formation. *J. Bacteriol.* **182**: 5653-5662.
- van Wezel, G.P., White, J., Hoogvliet, G., and Bibb, M.J. (2000b) Application of *redD*, the transcriptional activator gene of the undecylprodigiosin biosynthetic pathway, as a reporter for transcriptional activity in *Streptomyces coelicolor* A3 (2) and *Streptomyces lividans*. *J. Mol. Microbiol. Biotechnol.* **2**: 551-556.
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**: research0034.0031.
- Vara, J., Lewandowska-Skarbek, M., Wang, Y.G., Donadio, S., and Hutchinson, C.R. (1989) Cloning of genes governing the deoxysugar portion of the erythromycin

- biosynthesis pathway in *Saccharopolyspora erythraea* (*Streptomyces erythreus*). *J. Bacteriol.* **171**: 5872-5881.
- Viaene, T., Langendries, S., Beirinckx, S., Maes, M., and Goormachtig, S. (2016) *Streptomyces* as a plant's best friend? *FEMS Microbiol. Ecol.* **92**: 1-9.
- Vine, K.L., Matesic, L., Locke, J.M., Ranson, M., and Skropeta, D. (2009) Cytotoxic and anticancer activities of isatin and its derivatives: a comprehensive review from 2000-2008. *Anticancer Agents Med. Chem.* **9**: 397-414.
- Vitreschak, A.G., Rodionov, D.A., Mironov, A.A., and Gelfand, M.S. (2002) Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. *Nucleic Acids Res.* **30**: 3141-3151.
- Vogel, C., and Marcotte, E.M. (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**: 227.
- Waldman, A.J., Pechersky, Y., Wang, P., Wang, J.X., and Balskus, E.P. (2015) The cremeomycin biosynthetic gene cluster encodes a pathway for diazo formation. *ChemBioChem* **16**: 2172-2175.
- Waldvogel, E., Herbig, A., Battke, F., Amin, R., Nentwich, M., Nieselt, K., Ellingsen, T.E., Wentzel, A., Hodgson, D.A., Wohlleben, W., and Mast, Y. (2011) The PII protein GlnK is a pleiotropic regulator for morphological differentiation and secondary metabolism in *Streptomyces coelicolor*. *Appl. Microbiol. Biotechnol.* **92**: 1219-1236.
- Wang, G., Hosaka, T., and Ochi, K. (2008) Dramatic activation of antibiotic production in *Streptomyces coelicolor* by cumulative drug resistance mutations. *Appl. Environ. Microbiol.* **74**: 2834-2840.
- Wang, J., and Zhao, G.-P. (2009) GlnR positively regulates *nasA* transcription in *Streptomyces coelicolor*. *Biochem. Biophys. Res. Commun.* **386**: 77-81.
- Wang, Y., Cen, X.-F., Zhao, G.-P., and Wang, J. (2012) Characterization of a new GlnR binding box in the promoter of *amtB* in *Streptomyces coelicolor* inferred a PhoP/GlnR competitive binding mechanism for transcriptional regulation of *amtB*. *J. Bacteriol.* **194**: 5237.
- Wentzel, A., Bruheim, P., Øverby, A., Jakobsen, Ø.M., Sletta, H., Omara, W.A.M., Hodgson, D.A., and Ellingsen, T.E. (2012a) Optimized submerged batch fermentation strategy for systems scale studies of metabolic switching in *Streptomyces coelicolor* A3(2). *BMC Syst. Biol.* **6**: 59.
- Wentzel, A., Sletta, H., STREAM Consortium, Ellingsen, T.E., and Bruheim, P. (2012b) Intracellular metabolite pool changes in response to nutrient depletion induced metabolic switching in *Streptomyces coelicolor*. *Metabolites* **2**: 178-194.
- Wessel, D., and Flugge, U.I. (1984) A method for the quantitative recovery of protein in dilute-solution in the presence of detergents and lipids. *Anal. Biochem.* **138**: 141-143.
- WHO, (2014) *Antimicrobial resistance: Global report on surveillance*. World Health Organization.
- Willemse, J., Borst, J.W., de Waal, E., Bisseling, T., and van Wezel, G.P. (2011) Positive control of cell division: FtsZ is recruited by SsgB during sporulation of *Streptomyces*. *Genes Dev.* **25**: 89-99.



- Willemse, J., and van Wezel, G.P. (2009) Imaging of *Streptomyces coelicolor* A3(2) with reduced autofluorescence reveals a novel stage of FtsZ localization. *PLoS One* **4**: e4242.
- Wu, C., Choi, Y.H., and van Wezel, G.P. (2016a) Metabolic profiling as a tool for prioritizing antimicrobial compounds. *J. Ind. Microbiol. Biotechnol.* **43**: 299-312.
- Wu, C., Du, C., Gubbens, J., Choi, Y.H., and van Wezel, G.P. (2015a) Metabolomics-driven discovery of a prenylated isatin antibiotic produced by *Streptomyces* species MBT28. *J. Nat. Prod.* **78**: 2355-2363.
- Wu, C., Du, C., Ichinose, K., Choi, Y.H., and van Wezel, G.P. (2017) Discovery of C-Glycosylpyranonaphthoquinones in *Streptomyces* sp. MBT76 by a combined NMR-based metabolomics and bioinformatics workflow. *J. Nat. Prod.* **80**: 269-277.
- Wu, C., Kim, H.K., van Wezel, G.P., and Choi, Y.H. (2015b) Metabolomics in the natural products field – a gateway to novel antibiotics. *Drug Discov. Today Technol.* **13**: 11-17.
- Wu, C., Medema, M.H., Läkamp, R.M., Zhang, L., Dorrestein, P.C., Choi, Y.H., and van Wezel, G.P. (2016b) Leucanicidin and endophenazines result from methyl-rhamnosylation by the same tailoring enzymes in *Kitasatospora* sp. MBT66. *ACS Chem. Biol.* **11**: 478-490.
- Wu, C., Van Wezel, G.P., and Hae Choi, Y. (2015c) Identification of novel endophenazine antibiotics produced by *Kitasatospora* sp. MBT66. *J. Antibiot.* **68**: 445-452.
- Wu, C., Zhu, H., van Wezel, G.P., and Choi, Y.H. (2016c) Metabolomics-guided analysis of isocoumarin production by *Streptomyces* species MBT76 and biotransformation of flavonoids and phenylpropanoids. *Metabolomics* **12**: 90.
- Wu, C.S., Kim, H.K., van Wezel, G.P., and Choi, Y.K. (2015d) Metabolomics in the natural products field – a gateway to novel antibiotics. *Drug Discov. Today Technol.* **13**: 11-17.
- Xiao, K., Kinkel, L.L., and Samac, D.A. (2002) Biological control of phytophthora root rots on alfalfa and soybean with *Streptomyces*. *Biol. Control* **23**: 285-295.
- Xu, J., Tozawa, Y., Lai, C., Hayashi, H., and Ochi, K. (2002) A rifampicin resistance mutation in the *rpoB* gene confers ppGpp-independent antibiotic production in *Streptomyces coelicolor* A3(2). *Mol. Genet. Genomics* **268**: 179-189.
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015) The I-TASSER Suite: Protein structure and function prediction. *Nat. Methods* **12**: 7-8.
- Yang, Y.-H., Song, E., Willemse, J., Park, S.-H., Kim, W.-S., Kim, E.-j., Lee, B.-R., Kim, J.-N., van Wezel, G.P., and Kim, B.-G. (2012) A novel function of *Streptomyces* integration host factor (slHF) in the control of antibiotic production and sporulation in *Streptomyces coelicolor*. *Antonie van Leeuwenhoek* **101**: 479-492.
- Yeats, C., Bentley, S., and Bateman, A. (2003) New knowledge from old: In silico discovery of novel protein domains in *Streptomyces coelicolor*. *BMC Microbiol.* **3**: 3.
- Yellaboina, S., Seshadri, J., Kumar, M.S., and Ranjan, A. (2004) PredictRegulon: A web server for the prediction of the regulatory protein binding sites and operons in prokaryote genomes. *Nucleic Acids Res.* **32**: W318-W320.

- Yoon, V., and Nodwell, J.R. (2014) Activating secondary metabolism with stress and chemicals. *J. Ind. Microbiol. Biotechnol.* **41**: 415-424.
- Zhang, L., Willemse, J., Claessen, D., and van Wezel, G.P. (2016) SepG coordinates sporulation-specific cell division and nucleoid organization in *Streptomyces coelicolor*. *Open Biol.* **6**: 150164.
- Zhang, L., Willemse, J., Hoskisson, P.A., and van Wezel, G.P. (2018) Sporulation-specific cell division defects in *ylmE* mutants of *Streptomyces coelicolor* are rescued by additional deletion of *ylmD*. *Sci. Rep.* **8**: 7328.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**: R137.
- Zhao, G., Jin, Z., Wang, Y., Allewell, N.M., Tuchman, M., and Shi, D. (2013) Structure and function of *Escherichia coli* RimK, an ATP-grasp fold, l-glutamyl ligase enzyme. *Proteins: Structure, Function, and Bioinformatics* **81**: 1847-1854.
- Zhu, H., Sandiford, S.K., and van Wezel, G.P. (2014a) Triggers and cues that activate antibiotic production by Actinomycetes. *J. Ind. Microbiol. Biotechnol.* **41**: 371-386.
- Zhu, H., Swierstra, J., Wu, C., Girard, G., Choi, Y.H., van Wamel, W., Sandiford, S.K., and van Wezel, G.P. (2014b) Eliciting antibiotics active against the ESKAPE pathogens in a collection of Actinomycetes isolated from mountain soils. *Microbiology* **160**: 1714-1725.
- Ziemert, N., Alanjary, M., and Weber, T. (2016) The evolution of genome mining in microbes - a review. *Nat. Prod. Rep.* **33**: 988-1005.



## Supplementary materials for Chapter 3

### Application of systems biology methods for the identification of novel natural products in *Streptomyces* species

Chao Du, Changsheng Wu, Jacob Gubbens, Koji Ichinose, Young Hae Choi, and Gilles P. van Wezel

Table S1 Significantly up or down regulated (with  $p$ -value less than 0.05) quantified proteins in at least one comparison

Protein IDs	MBT28- 30/WT (2log)	$p$ -value	MBT28- 91/WT (2log)	$p$ -value	MBT28- 91/30 (2log)	$p$ -value	predicted function
NODE_001_gene_001	0.2	0.604	3.1	0.020	2.8	0.008	1 4-alpha-glucan branching enzyme GlgB 2
NODE_001_gene_012	-1.9	0.174	-3.1	0.033	-1.4	0.184	6-phosphofructokinase 2
NODE_001_gene_037	-2.4	0.075	-3.3	0.025	-1.0	0.384	Methylmalonyl-CoA epimerase
NODE_005_gene_007	1.1	0.077	-1.6	0.577	-2.6	0.010	Alanine--tRNA ligase
NODE_009_gene_004	-0.3	0.994	-2.3	0.150	-2.2	0.045	Secreted protease
NODE_012_gene_017	-1.5	0.479	-3.5	0.021	-2.3	0.029	ATP-dependent Clp protease proteolytic subunit 1
NODE_016_gene_001	3.6	0.000	4.3	0.002	0.6	0.576	Secreted protein
NODE_016_gene_019	-1.2	0.658	-3.8	0.010	-2.5	0.014	Ribose-phosphate pyrophosphokinase
NODE_020_gene_023	0.2	0.581	2.8	0.030	2.4	0.026	Secreted protein
NODE_020_gene_034	0.8	0.263	2.8	0.035	2.0	0.064	Flp pilus assembly protein CpaB
NODE_021_gene_028	-2.6	0.068	-3.9	0.007	-1.2	0.272	Cytochrome oxidase subunit I
NODE_021_gene_047	2.4	0.004	3.4	0.012	0.9	0.388	Secreted protein
NODE_021_gene_048	2.5	0.003	4.7	0.001	2.7	0.012	Nuclear export factor GLE1
NODE_025_gene_009	-2.9	0.039	-1.9	0.400	1.0	0.268	Putative PadR-like family transcriptional regulator
NODE_025_gene_015	-1.0	0.532	-3.2	0.029	-2.4	0.029	Ectoine/hydroxyectoine ABC transporter permease EhuC
NODE_025_gene_017	1.3	0.097	3.7	0.006	2.6	0.017	Methyltransferase
NODE_025_gene_022	1.7	0.037	0.6	0.467	-0.9	0.439	Nucleotide-binding protein
NODE_025_gene_023	1.7	0.033	4.5	0.001	2.5	0.021	Peptidoglycan-binding domain 1 protein
NODE_026_gene_007	1.9	0.009	3.7	0.007	1.7	0.072	Secreted protein
NODE_027_gene_001	2.3	0.003	3.4	0.013	1.6	0.088	Conserved hypothetical secreted protein
NODE_029_gene_014	3.2	0.000	2.5	0.051	-0.9	0.411	Secreted protein
NODE_032_gene_004	-1.9	0.174	-3.3	0.025	-1.2	0.254	Membrane protein
NODE_032_gene_009	0.9	0.205	-2.5	0.117	-3.1	0.004	Starvation-induced DNA protecting protein
NODE_033_gene_057	-2.6	0.048	-1.9	0.266	0.7	0.522	Cysteine--tRNA ligase
NODE_035_gene_011	-2.5	0.062	-3.4	0.018	-1.0	0.368	Glycerol-3-phosphate dehydrogenase
NODE_035_gene_012	-2.4	0.073	0.5	0.507	3.2	0.003	Glycerol kinase 1
NODE_035_gene_015	-0.3	0.707	3.0	0.024	2.6	0.007	Methionine synthase
NODE_035_gene_017	3.4	0.000	3.0	0.025	-0.7	0.532	Lipoprotein oligopeptide binding protein
NODE_036_gene_009	0.1	0.429	2.2	0.073	2.1	0.030	Oxidoreductase
NODE_042_gene_025	-1.6	0.262	-3.6	0.013	-1.8	0.089	Nucleotide sugar-1-phosphate transferase
NODE_047_gene_002	-3.8	0.003	-4.4	0.002	-0.5	0.644	Alkylhydroperoxidase like protein AhpD family
NODE_047_gene_007	-3.0	0.022	-2.9	0.056	-0.1	0.958	Alkylhydroperoxidase like protein AhpD family
NODE_048_gene_013	0.9	0.111	3.5	0.011	2.2	0.028	Carboxypeptidase
NODE_048_gene_015	3.9	0.000	3.0	0.025	-0.8	0.493	Metallopeptidase
NODE_051_gene_011	-2.1	0.133	0.2	0.628	2.2	0.039	Monooxygenase
NODE_057_gene_030	-3.3	0.015	-2.5	0.169	0.5	0.585	Putative oxidoreductase
NODE_057_gene_031	-2.7	0.040	-1.6	0.388	1.7	0.105	Putative 6-phospho-3-hexuloisomerase
NODE_057_gene_032	-2.8	0.031	-0.4	0.923	2.3	0.029	Putative triosephosphate isomerase
NODE_058_gene_031	-3.6	0.005	-1.5	0.410	0.8	0.465	Alkylhydroperoxidase AhpD family core domain-containing protein
NODE_062_gene_005	0.0	0.471	-2.4	0.207	-2.3	0.026	DNA-binding protein
NODE_063_gene_013	-3.2	0.013	-2.8	0.064	0.9	0.393	Glutamine synthetase 2
NODE_063_gene_031	2.2	0.007	2.0	0.102	-0.4	0.758	Maltose-binding protein
NODE_064_gene_033	2.5	0.002	1.5	0.202	-1.2	0.286	Secreted protein
NODE_065_gene_019	-2.7	0.040	-2.7	0.078	0.3	0.745	Carboxylesterase
NODE_065_gene_022	1.0	0.163	4.0	0.003	3.1	0.004	Cold shock domain-containing protein CspD
NODE_066_gene_007	-4.0	0.002	-4.0	0.005	-0.2	0.872	Secreted protein
NODE_069_gene_017	4.2	0.000	2.9	0.027	-1.1	0.323	Sugar transporter sugar binding protein
NODE_070_gene_012	2.9	0.000	4.1	0.004	1.0	0.271	Oxidoreductase
NODE_071_gene_024	0.2	0.347	-2.4	0.197	-2.7	0.009	Integral membrane protein
NODE_072_gene_025	1.1	0.134	-0.8	0.798	-2.2	0.043	Succinate-semialdehyde dehydrogenase
NODE_075_gene_003	-0.1	0.555	2.0	0.092	2.7	0.007	Putative methyltransferase
NODE_078_gene_010	-2.6	0.049	-3.8	0.008	-1.0	0.376	Putative 3-oxoacyl-ACP synthase II
NODE_079_gene_007	1.0	0.177	-1.8	0.318	-2.3	0.035	Chloride peroxidase
NODE_079_gene_008	0.7	0.290	-1.5	0.444	-2.4	0.029	Alpha/beta hydrolase
NODE_082_gene_003	2.2	0.004	0.5	0.380	-1.3	0.211	Uncharacterized protein
NODE_089_gene_031	-0.5	0.822	2.7	0.038	3.0	0.002	SPFH domain/Band 7 family protein
NODE_090_gene_004	1.8	0.024	0.9	0.369	-1.1	0.333	Integral membrane protein
NODE_091_gene_005	1.2	0.057	-2.2	0.257	-3.6	0.000	Hypothetical cytosolic protein
NODE_092_gene_010	-0.1	0.526	1.9	0.095	1.9	0.046	Redoxin
NODE_092_gene_011	-1.3	0.587	-3.4	0.024	-2.1	0.042	3-phosphoserine phosphatase
NODE_099_gene_033	-0.8	0.665	2.3	0.066	3.1	0.004	NDP-4-keto-6-deoxy-L-hexose 2 3-reductase
NODE_102_gene_017	2.3	0.003	2.1	0.076	-0.1	0.956	ABC-type cobalt transport system
NODE_110_gene_015	-0.4	0.918	2.1	0.096	2.2	0.039	Pseudouridine-5'-phosphate glycosidase

NODE_110_gene_016	0.4	0.453	2.0	0.107	2.3	0.032	Dioxygenase
NODE_118_gene_013	1.9	0.011	4.0	0.005	2.0	0.041	Predicted protein
NODE_119_gene_005	0.6	0.347	2.6	0.045	1.8	0.085	Putative secreted extracellular small neutral protease
NODE_126_gene_032	1.0	0.179	-1.4	0.492	-2.3	0.030	Putative calcium binding protein
NODE_131_gene_004	1.5	0.056	2.5	0.050	1.0	0.360	lipoprotein
NODE_131_gene_035	-1.9	0.260	-3.2	0.044	-1.2	0.265	Uncharacterized protein
NODE_136_gene_010	-0.9	0.880	2.5	0.047	3.5	0.000	AdpA
NODE_136_gene_016	0.1	0.397	2.1	0.076	1.9	0.045	B-N-acetylhexosaminidase
NODE_137_gene_015	-2.6	0.048	-1.9	0.258	0.9	0.392	Flavoprotein reductase
NODE_139_gene_026	-1.1	0.726	0.8	0.309	2.5	0.011	Putative uncharacterized protein
NODE_143_gene_007	-1.4	0.322	-3.3	0.025	-1.5	0.176	Succinyl-CoA ligase [ADP-forming] subunit beta-1
NODE_143_gene_008	-4.3	0.001	-3.4	0.028	0.5	0.587	Succinyl-CoA ligase [ADP-forming] subunit alpha
NODE_144_gene_030	2.9	0.000	3.6	0.009	0.5	0.541	Iron transport lipoprotein
NODE_146_gene_013	0.3	0.329	-2.9	0.076	-2.6	0.011	Uncharacterized protein
NODE_146_gene_033	1.6	0.041	-1.2	0.590	-2.6	0.014	Dehydrogenase
NODE_152_gene_033	2.1	0.005	4.4	0.002	2.2	0.023	Amino acid transport integral membrane protein
NODE_153_gene_002	2.7	0.001	3.3	0.014	0.4	0.595	Solute-binding lipoprotein
NODE_154_gene_010	2.1	0.005	3.3	0.015	1.3	0.165	Putative secreted protein
NODE_154_gene_060	NaN	NaN	2.8	0.030	2.8	0.004	Aminotransferase
NODE_154_gene_064	-0.3	0.710	1.9	0.104	2.3	0.018	Oxidoreductase
NODE_155_gene_014	4.6	0.000	5.8	0.000	1.2	0.244	Substrate binding protein
NODE_155_gene_032	1.0	0.161	-1.6	0.372	-2.7	0.011	Mesaconyl-CoA hydratase
NODE_155_gene_040	0.8	0.138	-1.7	0.530	-2.4	0.022	Secreted peptidase
NODE_155_gene_052	-1.2	0.627	0.8	0.299	2.0	0.041	Uncharacterized protein
NODE_159_gene_015	3.8	0.000	4.5	0.002	0.6	0.496	Putative membrane protein
NODE_168_gene_002	0.9	0.190	-1.3	0.538	-2.2	0.045	6-pyruvoyl tetrahydropterin synthase
NODE_171_gene_005	-1.9	0.162	-4.0	0.004	-2.4	0.029	Sugar phosphotransferase
NODE_173_gene_018	1.8	0.022	0.8	0.408	-1.1	0.312	Peptidoglycan-binding domain 1 protein
NODE_175_gene_006	1.0	0.180	-1.5	0.432	-2.3	0.032	Putative lysozyme
NODE_180_gene_024	0.9	0.116	3.0	0.023	2.4	0.016	Aromatic prenilyltransferase DMATS type
NODE_180_gene_025	1.6	0.040	2.8	0.031	1.1	0.286	Aromatic amino acid beta-eliminating lyase/threonine aldolase
NODE_188_gene_004	2.4	0.004	-2.4	0.137	-4.7	0.000	Iron sulfur protein (Secreted protein)
NODE_191_gene_023	0.1	0.680	2.0	0.103	2.5	0.018	3-oxoacyl-[acyl-carrier-protein] synthase 2
NODE_196_gene_002	0.6	0.330	3.5	0.010	2.6	0.015	Secreted protein
NODE_198_gene_010	-2.9	0.039	-2.7	0.129	0.5	0.587	Aminoacylase
NODE_207_gene_013	0.9	0.103	3.5	0.011	2.4	0.015	Neutral zinc metalloprotease
NODE_207_gene_015	-1.7	0.226	1.5	0.191	3.5	0.001	Phosphorylase
NODE_207_gene_016	0.6	0.323	3.1	0.020	2.3	0.030	Alpha-amylase
NODE_211_gene_006	3.9	0.000	2.2	0.072	-1.9	0.070	ABC transport system integral membrane protein BldKC
NODE_212_gene_021	-0.1	0.897	-4.5	0.001	-4.3	0.000	Putative signal transduction protein with EFhand domain
NODE_226_gene_007	-1.5	0.434	-3.7	0.012	-2.2	0.030	Exporter
NODE_227_gene_010	1.8	0.028	1.5	0.186	-0.2	0.869	Proteinase
NODE_230_gene_006	-3.0	0.028	-2.2	0.258	0.4	0.637	GCN5-related N-acetyltransferase
NODE_231_gene_019	-2.9	0.030	-3.9	0.006	-0.9	0.393	Peptide transporter
NODE_231_gene_043	-2.7	0.044	-3.1	0.035	-0.7	0.523	FAD-dependent pyridine nucleotide-disulfide oxidoreductase
NODE_244_gene_025	-1.9	0.260	-3.3	0.033	-1.2	0.266	Phosphodiesterase
NODE_245_gene_011	-2.2	0.143	-3.2	0.044	-0.9	0.391	Mycothioli conjugate amidase Mca
NODE_248_gene_003	0.9	0.209	-1.5	0.441	-2.5	0.021	Peptide transport system secreted peptide-binding protein
NODE_249_gene_004	-0.3	1.000	-3.5	0.015	-3.0	0.005	Urease subunit alpha 1
NODE_250_gene_002	-1.3	0.558	1.3	0.188	2.6	0.009	Putative transmembrane transport protein
NODE_250_gene_012	1.4	0.034	2.1	0.076	0.7	0.429	Cationic amino acid transporter
NODE_250_gene_016	-2.2	0.166	-3.2	0.045	-1.1	0.312	Thioredoxin
NODE_251_gene_002	-1.4	0.503	0.8	0.312	2.1	0.033	Sigma 54 modulation protein/SSU ribosomal protein S30P
NODE_259_gene_010	1.7	0.035	1.3	0.242	-0.5	0.659	Subtilisin-like protease
NODE_267_gene_023	3.5	0.000	3.6	0.009	0.1	0.872	Secreted alkaline phosphatase
NODE_271_gene_008	-3.3	0.012	-2.1	0.208	0.1	0.886	UDP-glucose 4-epimerase
NODE_272_gene_009	-0.5	0.854	1.3	0.197	2.0	0.043	WD40 repeat-containing protein
NODE_279_gene_009	1.2	0.113	-1.6	0.378	-2.9	0.007	Putative lipoprotein
NODE_279_gene_082	1.9	0.017	-0.6	0.919	-2.6	0.017	Glyceraldehyde 3-phosphate dehydrogenase
NODE_279_gene_107	0.5	0.236	3.3	0.015	2.7	0.005	Putative cysteine desulphurases SufS
NODE_281_gene_006	-0.1	0.825	-2.6	0.094	-2.6	0.014	3-hydroxyacyl-CoA dehydrogenase PaaC
NODE_282_gene_002	-2.7	0.041	-2.6	0.096	0.1	0.881	Phosphocarrier protein HPr
NODE_286_gene_026	0.7	0.302	-1.7	0.322	-2.4	0.025	Catalase
NODE_288_gene_011	0.7	0.169	-1.6	0.580	-2.1	0.045	Uncharacterized protein
NODE_299_gene_002	-2.8	0.049	-1.4	0.744	1.5	0.127	Bifunctional uroporphyrinogen-III synthetase/response regulator domain protein
NODE_299_gene_014	1.9	0.011	2.2	0.073	0.4	0.616	Sugar transporter sugar-binding protein

NODE_300_gene_002	-1.3	0.389	1.0	0.333	2.3	0.031	Adenylosuccinate synthetase
NODE_309_gene_008	-2.1	0.177	-3.1	0.049	-0.7	0.545	Orotate phosphoribosyltransferase
NODE_316_gene_021	1.3	0.050	-0.6	0.821	-1.4	0.172	Secreted hydrolase
NODE_320_gene_035	0.0	0.502	2.1	0.075	2.2	0.025	3-ketosteroid-delta-1-dehydrogenase (Fragment)
NODE_321_gene_026	-3.2	0.013	-2.7	0.081	0.5	0.647	Phosphoribosylaminoimidazole-succinocarboxamide synthase
NODE_330_gene_015	-3.0	0.022	-3.4	0.019	-0.3	0.790	Peptidase M48
NODE_337_gene_001	-2.5	0.093	-3.8	0.009	-0.3	0.830	Putative oxygenase
NODE_340_gene_005	-0.8	0.937	0.9	0.272	1.9	0.049	Pyruvate phosphate dikinase
NODE_346_gene_006	-0.6	0.899	1.1	0.238	2.0	0.038	Translation initiation factor IF-3
NODE_347_gene_004	4.1	0.000	4.4	0.001	0.5	0.642	Lipoprotein
NODE_359_gene_003	-0.9	0.576	-3.0	0.044	-2.0	0.066	Putative secreted protein
NODE_363_gene_001	1.8	0.025	5.3	0.000	4.2	0.000	Putative iron transport lipoprotein
NODE_366_gene_016	4.0	0.000	3.7	0.007	-0.2	0.882	Glutamate binding protein
NODE_366_gene_028	-0.4	0.744	-3.1	0.056	-2.5	0.013	Predicted protein
NODE_366_gene_037	3.4	0.000	3.6	0.010	0.3	0.714	Secreted protein
NODE_395_gene_033	-1.8	0.311	-3.4	0.025	-1.6	0.131	HesB/YadR/YfhF family protein
NODE_395_gene_039	-2.2	0.098	-3.8	0.008	-1.6	0.145	Uncharacterized protein
NODE_399_gene_018	0.8	0.133	-1.7	0.554	-2.1	0.038	Uncharacterized protein
NODE_410_gene_001	1.6	0.041	2.8	0.031	1.1	0.282	Secreted protein
NODE_416_gene_006	1.7	0.030	1.3	0.243	-0.3	0.822	Germacradienol/geosmin synthase
NODE_431_gene_010	0.3	0.545	2.8	0.030	2.3	0.032	Uncharacterized protein
NODE_434_gene_008	-2.0	0.210	-3.9	0.008	-1.2	0.244	Putative polyketide cyclase
NODE_451_gene_001	-1.1	0.474	0.9	0.361	2.4	0.026	NB-ARC domain-containing protein
NODE_453_gene_003	-2.7	0.038	-2.8	0.060	0.4	0.705	Methylmalonyl-CoA mutase
NODE_461_gene_007	1.7	0.038	-1.6	0.397	-3.4	0.002	Protease
NODE_463_gene_011	2.0	0.008	3.0	0.025	1.3	0.173	ABC-type Fe3+-siderophore transporter ATP-binding protein
NODE_463_gene_013	4.1	0.000	5.9	0.000	2.1	0.054	ABC-type Fe3+-siderophore transporter substrate-binding protein
NODE_464_gene_004	0.4	0.282	2.6	0.039	2.1	0.032	Cobalt transport integral membrane protein
NODE_464_gene_005	1.8	0.024	2.5	0.055	1.2	0.248	Putative ABC transporter ATP-binding protein SCO5958
NODE_487_gene_009	1.4	0.042	4.7	0.001	3.5	0.000	Muramoyl-pentapeptide carboxypeptidase
NODE_487_gene_101	3.6	0.000	4.7	0.001	1.6	0.135	Lipoprotein
NODE_487_gene_102	3.7	0.000	5.1	0.001	2.0	0.041	Solute-binding lipoprotein
NODE_487_gene_103	5.3	0.000	6.5	0.000	1.1	0.282	Secreted protein
NODE_491_gene_002	1.8	0.013	1.4	0.168	-0.3	0.854	Secreted alkaline phosphatase
NODE_510_gene_026	-0.3	0.664	-2.1	0.330	-2.1	0.041	Membrane protein
NODE_514_gene_002	-2.2	0.144	-3.8	0.010	-1.6	0.127	Secreted penicillin binding protein
NODE_526_gene_007	-2.8	0.048	-3.3	0.032	-0.3	0.865	Lipoprotein
NODE_528_gene_003	0.8	0.135	3.2	0.018	2.5	0.010	Carrier protein membrane protein (Fragment)
NODE_532_gene_001	1.3	0.096	2.9	0.028	1.9	0.074	Alpha-1 4-glucan:maltose-1-phosphate maltosyltransferase 1
NODE_534_gene_001	-1.7	0.342	1.3	0.196	2.9	0.003	Alpha-1 4-glucan:maltose-1-phosphate maltosyltransferase 1
NODE_585_gene_004	-0.9	0.607	1.6	0.172	2.4	0.023	DUF364 domain-containing protein
NODE_633_gene_002	0.7	0.284	2.7	0.037	1.7	0.116	Nitrogen regulatory protein P-II
NODE_644_gene_008	2.0	0.008	-1.4	0.749	-3.1	0.002	hydrolase
NODE_659_gene_009	-0.6	0.763	2.2	0.080	2.9	0.008	Amidase
NODE_679_gene_001	-1.9	0.169	-3.9	0.006	-2.1	0.052	Phage tail sheath protein
NODE_746_gene_022	0.9	0.202	-2.5	0.113	-3.4	0.002	Lipoprotein
NODE_746_gene_044	0.8	0.263	-1.3	0.543	-2.1	0.050	Putative glyoxalase/bleomycin resistance family protein
NODE_746_gene_047	0.5	0.230	-2.7	0.119	-3.1	0.003	Phytanoyl-CoA dioxygenase

Table S2. Oligo nucleotides used in this study

Name	Sequence
<i>rpsL</i> _For	CGGCACACAGAAACCGGAGAAG
<i>rpsL</i> _Rev	GTCGATGATGACCGGGCGCTTCG
<i>rsmG</i> _For	TGACGAATTCGTCAGCCTGATAGTTCTGGTGG
<i>rsmG</i> _Rev	TCAGTAAGCTTGCCGTGCAGCAGTGAGCGACG
SC06_0044_F_EcoRI	CGATGAATTCGTCGCGCTTCGTGTG
SC06_0044_R_EcoRI_NdeI	CGATGAATTCGCACCCGGTACAGGAGTGTGTCATATGCGATTCAACCTCATC

### Supplementary materials for Chapter 4

#### Unravelling the response of *Streptomyces roseifaciens* to challenge with small molecules by genome-wide proteomics

Chao Du, Somayah Elsayed, Marta del Rio, Bogdan I. Florea, Jos Raaijmakers, Gilles P. van Wezel



Table S1 Differentially expressed proteins (DEPs) for all small molecule treatments

ID	predicted function	Ctr Mean	HyC			...
			Mean	log2 Fold change	log10 p-Value	...
sc01_0047	Uncharacterized protein	5942.667	18992.33	1.676235	0.743147	...
sc01_0194	transferase	14004.33	9288.667	-0.59233	1.59673	...
sc01_0208	Putative Uncharacterized 50.6 kDa protein in the 5' region of gyrA and gyrB	10565.67	<b>51839.67</b>	<b>2.294673</b>	<b>3.189379</b>	...
sc01_0222	Pyruvate dehydrogenase E1 component	28681.67	<b>2925.333</b>	<b>-3.29346</b>	<b>2.101675</b>	...
sc01_0228	Methyltransferase domain-containing protein	3432.667	<b>8844.333</b>	<b>1.365424</b>	<b>1.816476</b>	...
sc01_0243	ERCC4 domain protein	52936.67	<b>7105.333</b>	<b>-2.89729</b>	<b>1.455557</b>	...
sc01_0253	Glyoxalase	10574.33	<b>27181</b>	<b>1.362032</b>	<b>1.453402</b>	...
sc02_0182	Putative transcriptional regulatory protein	3132	<b>6385.333</b>	<b>1.027678</b>	<b>1.603741</b>	...
...	...	...	...	...	...	...

**Bold** numbers indicating the protein in that specific treatment passed the threshold for DEPs ( $p$ -value < 0.05, fold change  $\geq 2$  or  $\leq -2$ )

Complete table can be accessed via:

[https://osf.io/hx9vc/?view\\_only=fdb5749fe6c49e8a6665e29ee2b6cb0](https://osf.io/hx9vc/?view_only=fdb5749fe6c49e8a6665e29ee2b6cb0)

Table S2. DEPs<sup>†</sup> induced by short-term JA treatment

Protein ID	0 h			15 min			30 min			1 h			2 h			4 h			Log <sub>2</sub> ratios of JA/Control								
	Control		Mean	Control		Mean	Control		Mean	Control		Mean	Control		Mean	Control		Mean	Control		Mean	SEM	JA				
	SEM**	SEM	SEM	SEM**	SEM	SEM	SEM**	SEM	SEM	SEM**	SEM	SEM	SEM**	SEM	SEM	SEM**	SEM	SEM	SEM**	SEM	SEM	SEM	JA				
sc01_0003	171.73	18.41	130.50	10.95	143.14	32.71	235.09	21.68	114.50	2.76	180.46	13.77	174.29	22.10	187.25	37.58	172.51	11.96	240.22	60.22	200.53	14.25	0.13	-1.04	-0.05	-0.12	-0.26
sc03_1246	814.23	41.74	754.06	35.54	806.78	68.77	967.86	38.72	856.13	5.85	1681.0	270.38	1169.0	56.73	6645.9	1239.0	2649.0	364.60	5927.2	199.25	4058.8	189.29	0.10	-0.18	-0.52	-1.33	-0.55
sc03_1247	235.94	34.89	245.00	23.42	211.93	16.71	158.31	43.87	167.48	25.81	189.04	34.06	189.24	78.33	861.62	143.00	315.85	37.29	908.33	137.78	666.82	80.93	-0.21	0.08	0.00	-1.45	-0.45
sc03_1249	129.32	13.64	102.77	8.65	115.03	4.20	177.19	10.03	112.76	5.61	366.57	60.46	201.99	26.79	3561.0	824.60	1091.8	171.33	3787.5	88.17	2374.1	133.60	0.16	-0.65	-0.86	-1.71	-0.67
sc03_2100	500.37	158.39	545.66	282.49	492.02	299.45	105.20	32.83	225.20	27.32	201.93	38.33	246.89	143.51	124.81	29.63	164.69	58.50	111.68	48.88	117.98	51.65	-0.15	1.10	0.29	0.40	0.08
sc04_0002	319.88	68.39	301.95	125.37	306.08	104.41	205.99	24.51	207.12	80.48	239.13	78.37	290.35	89.24	233.63	12.53	206.34	94.98	104.11	47.02	313.34	63.97	0.02	0.01	0.28	-0.18	1.59
sc04_0750	396.84	15.64	401.30	16.82	408.36	26.00	410.10	19.62	402.35	20.99	388.51	10.25	354.26	34.80	287.18	26.24	206.13	25.78	107.33	32.13	248.03	43.24	0.03	-0.03	-0.13	-0.48	1.21
sc04_1170	394.22	46.27	425.35	133.40	377.12	165.35	361.41	78.41	466.88	117.53	658.66	24.38	301.28	56.60	318.30	84.40	331.86	94.43	110.65	37.54	282.80	169.83	-0.17	0.37	-1.13	0.06	1.35
sc04_1502	96.83	10.97	88.91	28.93	207.91	26.07	114.09	16.79	259.42	30.44	56.95	19.36	215.97	12.85	102.05	31.08	308.75	7.84	76.14	4.39	275.32	15.28	1.23	1.19	1.92	1.60	1.85
sc04_1633	323.46	27.50	300.17	33.55	384.81	75.34	269.77	52.56	266.76	14.40	162.78	41.38	261.29	66.32	195.44	51.06	235.02	49.42	144.89	23.07	299.26	45.04	0.36	-0.02	0.68	0.27	1.05
sc04_1686	78.77	6.66	83.35	19.95	71.16	35.30	72.62	6.94	80.64	6.91	66.46	15.07	74.07	9.64	56.04	4.28	64.07	24.25	11.97	3.26	33.90	4.77	-0.23	0.15	0.16	0.19	1.50
sc04_1826	350.02	52.46	447.19	127.62	493.15	128.74	316.14	84.27	318.42	103.59	460.82	53.67	390.43	96.70	263.19	90.38	364.63	113.34	224.35	64.68	510.72	96.56	0.14	0.01	-0.24	0.47	1.19
sc04_2087	104.53	20.67	44.41	19.40	74.01	42.92	119.96	27.04	71.47	19.81	182.20	61.89	145.37	35.15	124.57	33.11	114.39	47.34	221.71	42.75	77.86	25.69	0.74	-0.75	-0.33	-0.12	-1.51
sc04_2763	106.39	10.70	101.77	27.63	150.29	37.40	98.12	21.00	124.19	22.43	111.66	4.11	127.48	3.06	88.21	22.37	86.78	30.58	42.08	5.93	158.01	29.02	0.56	0.34	0.19	-0.02	1.91
sc05_0038	234.86	20.28	234.63	47.23	161.05	25.86	225.17	40.44	168.51	20.40	122.24	35.87	251.23	29.70	164.85	43.53	235.10	34.19	161.65	26.11	131.90	12.09	-0.54	-0.42	1.04	0.51	-0.29
sc06_0231	396.04	76.39	451.81	256.57	353.99	222.27	393.23	33.27	510.93	201.85	637.06	103.70	262.52	100.55	342.28	188.36	423.92	154.72	157.03	36.62	307.99	183.65	-0.35	0.38	-1.28	0.31	0.97
sc07_0079	39.60	8.61	30.99	15.59	32.40	12.54	45.13	23.22	52.11	14.86	54.75	38.49	47.64	24.58	433.02	69.70	102.58	34.91	364.71	25.66	240.48	50.47	0.06	0.21	-0.20	-2.08	-0.60

<sup>†</sup>  $p$ -value < 0.1, fold change  $\geq 2$ 

\* Jasmonic acid treatment

\*\* Standard error of mean

Table S3. Oligonucleotides used in this study

Purpose	No.	Name	Sequence 5'-3'
<i>jarA</i> knockout	1	jarA <sub>SpCRko</sub>	CATGCCATGGACAGGTCCTCCAGCATGAAGGTTTTAGAGCTAGAAATAGC
	2	jarA <sub>Up_GF</sub>	TCGTGGAAGGCACTAGAAGGGATGCCAACCTCGTACAGATCG
	3	jarA <sub>Up_GR</sub>	GTAGTGGCTGGGGATTCCCGTCAGGACCTCCCGGATGACTTC
	4	jarA <sub>Down_GF</sub>	CGGGAATCCCCAGCCACTAC
	5	jarA <sub>Down_GR</sub>	GGTCGATCCCCGCATATAGGCGGGCGGGTTGATGATGAC
<i>jarA</i> knockout genotype check	6	03.1245_KOCK_F	GGCCGTGAGCCACCTCACC
	7	03.1245_KOCK_R	CCGGCTCCAGCGCCTGTATC
SC04_1502 knockout	8	04.1502up_F	AGGAAGCTTTCCTTGCCCTACCTTTACCC
	9	04.1502up_R	AGGTCTAGAATGAGCGTGCCGGTGTTCTG
	10	04.1502down_F	ACATCTAGAGCAGGGCGGGCTGACAGAC
	11	04.1502down_R	ACAGAATTCATGGGCATGGCCCATCGG
SC04_1502 knockout genotype check	12	04.1502_KOCK_F	AGTGGGCGGAGAGCTACCAG
	13	04.1502_KOCK_R	GCCACACTCGCACCGCAAAG

Table S4. Plasmids and constructs used in this study

Plasmid and construct	Description	reference/vendor reference
pWHM3-oriT	<i>E. coli</i> / <i>Streptomyces</i> shuttle vector, high copy number and unstable in <i>Streptomyces</i>	Modified from Vara <i>et al.</i> (1989)
pUWL-Cre	<i>E. coli</i> / <i>Streptomyces</i> shuttle vector expressing the Cre recombinase in <i>Streptomyces</i>	Fedoryshyn <i>et al.</i> (2008)
pCRISPR-Cas9	<i>E. coli</i> / <i>Streptomyces</i> shuttle vector, harbouring codon optimised <i>cas9</i> , designed for easy inserting spacer sequences. Recombination template is designed to be inserted through in-vitro assembly to <i>StuI</i> site	Tong <i>et al.</i> (2015b)
pGWS1460	pCRISPR-Cas9 harbouring spacer and HDR region to knockout <i>jarA</i>	This study
pGWS1461	pWHM3-oriT harboring apramycin resistant cassette flanked by <i>loxP</i> sites and flanking regions of SC04_1502 to knockout SC04_1502	This study

Table S5 – CAS number and supplier of the small molecules used in this study

Abbreviations	Compound name	CAS number	Provider, catalog numbers, notes
JA	Jasmonic acid	77026-92-7	Cayman Chemical, 88300
SA	Salicylic acid	69-72-7	Alfa Aesar, A12253
IAA	Indole-3-acetic acid	6505-45-9	Sigma-Aldrich, I5148
NAA	Naphthalene acetic acid	86-87-3	Duchefa, N0903
HyC	Hydroxycoumarine	93-35-6	Alfa Aesar, L04082, 7-Hydroxycoumarine
GlcNAc	N-acetylglucosamine	7512-17-6	Sigma-Aldrich, A3286
CB	Cellobiose	528-50-7	Sigma-Aldrich, C7252
CS	Chitosan	9012-76-4	Sigma-Aldrich, C3646, $\geq 75\%$ deacetylated
CA	Cinnamic acid	621-82-9	Honeywell, 63185H
FA	Ferulic acid	537-98-4	Sigma-Aldrich, W518301
BA	Benzoic acid	65-85-0	Supelco, 1.00136.0100
BR	Sodium butyrate	156-54-7	Sigma-Aldrich, 303410

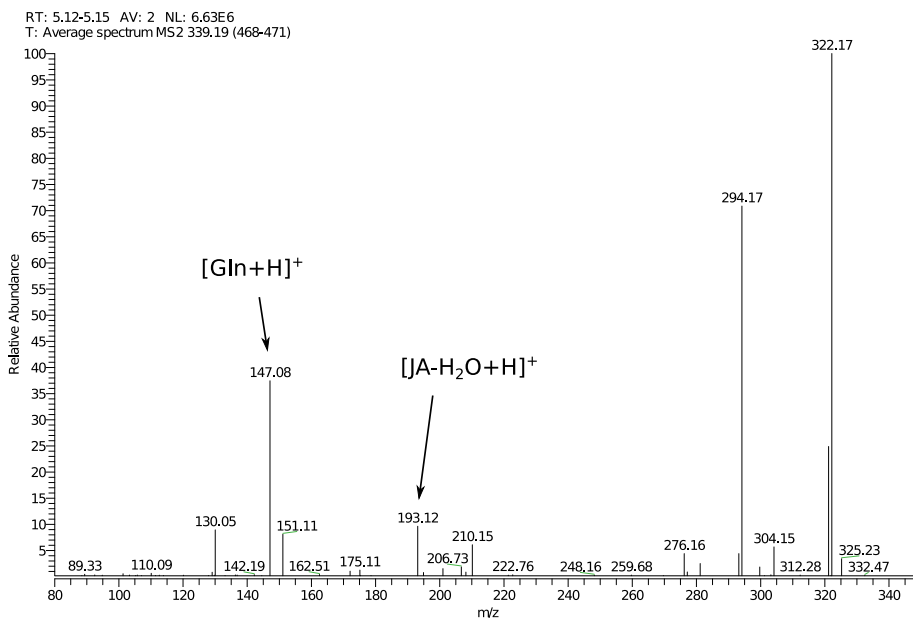
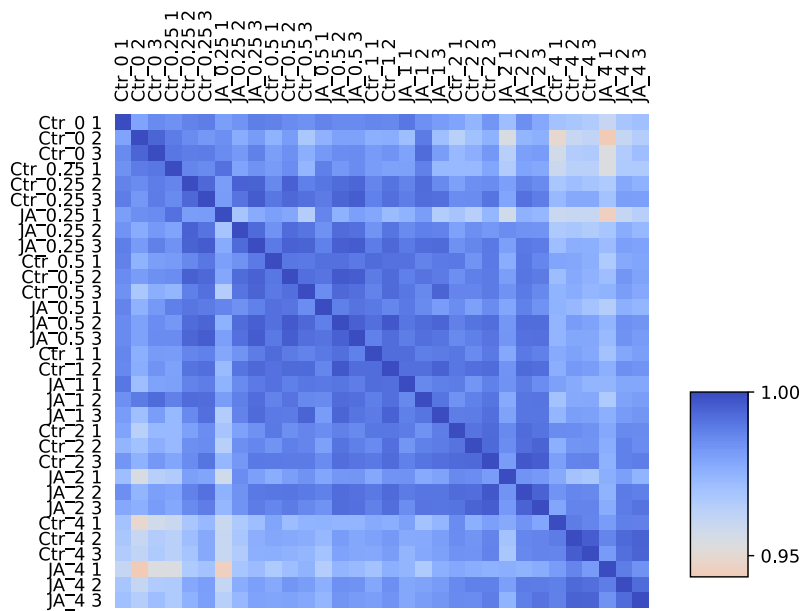
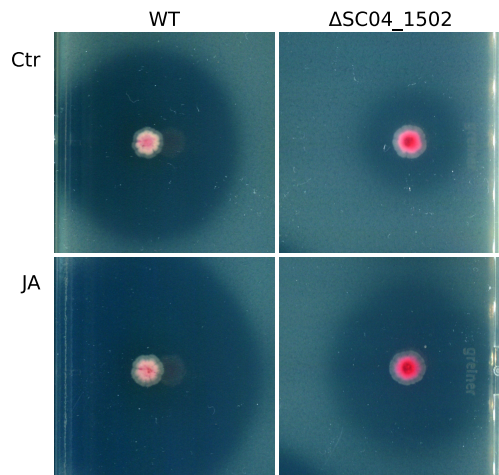


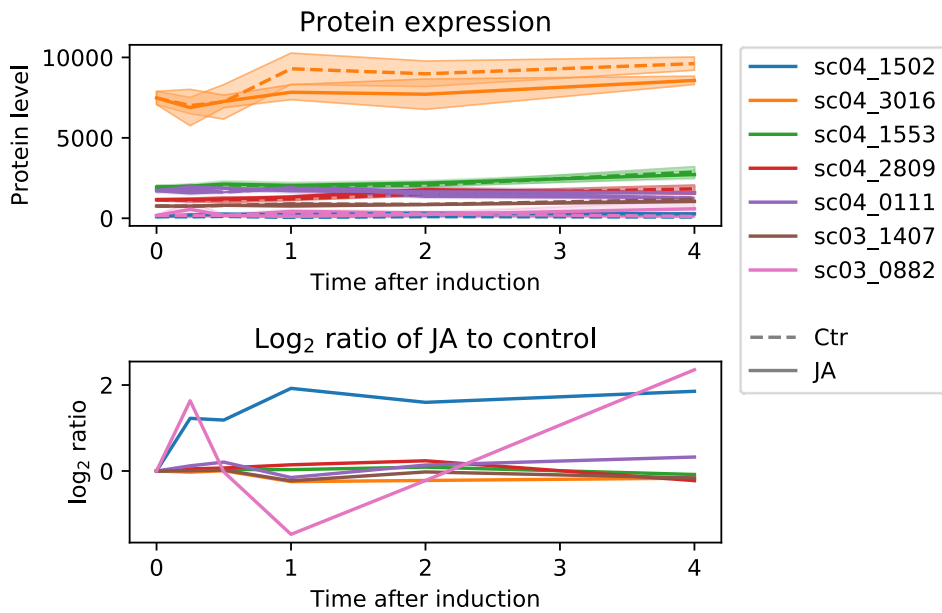
Figure S1. MS2 spectrum of jasmonoyl-glutamine (JA-Gln). Main fragments are annotated by arrows.



**Figure S2. Correlation plot of proteomics data of all samples in short-term JA response experiment.** Ctr: control sample; JA, jasmonic acid treatment; the numbers directly after underline indicating hours after jasmonic acid induction; the last number indicates the replicate number.



**Figure S3. Antimicrobial activity test of the  $\Delta$ SC04\_1502 strain compare with wildtype strain under JA treatment.** Spores of *S. roseifaciens* were grown for 5 days on MM medium supplemented with mannitol. *E. coli* ASD19 was overlaid on top as indicator.



**Figure S4. Abundance pattern of all detected MFS family proteins found in *S. roseifaciens* genome.** Note that SCO03\_0882 (pink line) did not pass the threshold as DEP.

A

### Supplementary materials for Chapter 5

Analysis of the background-reduced  
antibiotic production host *Streptomyces*  
*coelicolor* M1152 using quantitative  
proteomics

Chao Du, Dino van Dissel, Alexander Wentzel, Gilles P. van Wezel



Table S1. Protein levels in each sample

Strain	M145	M145	M145	...
Fermenter	F516	F516	F516	...
Aligned time point	1	2	3	...
Time after inoculation (h)	21	29	33	...
Total quantified	1732	1836	1873	...
SCO0006	1573	ND	ND	...
SCO0018	2086	2013	2045	...
SCO0020	ND	1513	ND	...
SCO0033	3356	2353	3924	...
SCO0070	ND	2638	3283	...
SCO0072	10821	7661	5601	...
SCO0075	2456	2637	1328	...
...	...	...	...	...

ISOQuant output using Top3 quantification method

ND - None detected

Complete table can be accessed via:

[https://osf.io/e352b/?view\\_only=ac6b5126cb5e4523b56bd55231365a51](https://osf.io/e352b/?view_only=ac6b5126cb5e4523b56bd55231365a51)

Table S2. DEPs for aligned time comparison between M1152 and M145

Aligned time point	1		2		3		...
Time after inoculation (h, M1152)	33		41		45		...
Time after inoculation (h, M145)	21		29		33		...
Number of DEPs	171		126		105		...
Protein ID (name)	log <sub>2</sub> fold change	p-value	log <sub>2</sub> fold change	p-value	log <sub>2</sub> fold change	p-value	...
SCO2198 (GlnA)	-1.800	0.000	-0.842	0.010	-0.613	0.105	...
SCO2210 (GlnII)	-2.111	0.000	-2.575	0.002	-2.565	0.030	...
SCO2286 (PhoA)	NA	NA	NA	NA	NA	NA	...
SCO2878	NA	NA	1.688	0.024	1.221	0.005	...
SCO4141 (PstC)	NA	NA	NA	NA	NA	NA	...
SCO4159 (GlnR)	-1.137	0.002	NA	NA	NA	NA	...
SCO4228	0.816	0.285	1.672	0.001	1.716	0.000	...
SCO0888	-1.266	0.031	NA	NA	-3.117	0.013	...
SCO1865 (EctB)	-0.656	0.170	0.685	0.014	0.472	0.138	...
SCO1867 (EctD)	0.558	0.000	1.195	0.002	1.396	0.000	...
SCO4683 (GdhA)	2.344	0.000	1.130	0.009	0.570	0.073	...
SCO5584 (GlnK)	-2.339	0.042	-3.136	0.062	-2.298	0.003	...
SCO5585 (GlnD)	-2.304	0.000	NA	NA	NA	NA	...
...	...	...	...	...	...	...	...

Showing only comparisons between aligned time points

p-values are calculated from student's T-test

None-significant ( $p \geq 0.05$ ,  $-1 < \log_2 \text{fold change} < 1$ ) are greyed out

NA - No data

Complete table can be accessed via:

[https://osf.io/e352b/?view\\_only=ac6b5126cb5e4523b56bd55231365a51](https://osf.io/e352b/?view_only=ac6b5126cb5e4523b56bd55231365a51)

Table S3. DEPs before and after phosphate depletion

Strain tp6/tp2	M145 45/29 h		M1152 57/41 h		Belong to protein group		
Protein ID (name)	log <sub>2</sub> fold change	p-value	log <sub>2</sub> fold change	p-value	glycolysis	TCA	PPP
SCO0617	3.14	0.01	0.85	0.08	-	-	Yes
SCO2627	1.28	0.04	-0.37	0.24	-	-	Yes
SCO4979	3.41	0.01	0.15	0.26	Yes	Yes	-
SCO0169	-1.73	0.01	-0.72	0.38	-	-	-
SCO0186(CrtI)	1.87	0.02	0.70	0.17	-	-	-
SCO0203	-2.07	0.00	-2.35	0.06	-	-	-
SCO0392	3.43	0.02	0.89	0.00	-	-	-
SCO0393	NA	NA	1.13	0.01	-	-	-
SCO0395	3.28	0.02	0.63	0.02	-	-	-
SCO0498(CchB)	-1.16	0.01	2.24	0.00	-	-	-
...	...	...	...	...	...	...	...

Tp2 and tp6 was chosen as representative aligned time points before and after phosphate depletion, respectively.

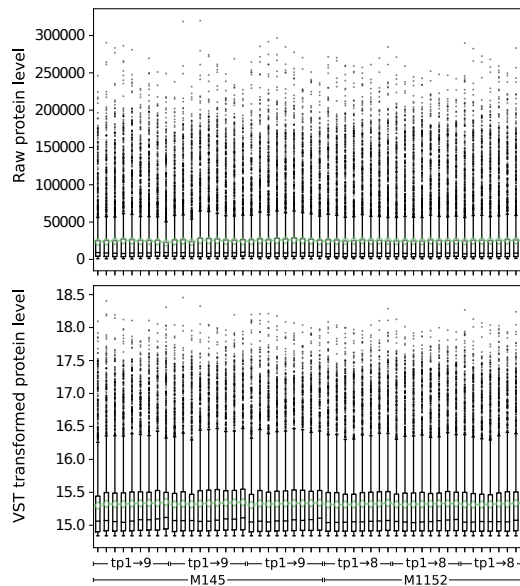
p-values are calculated from student's T-test

None-significant ( $p \geq 0.05$ ,  $-1 < \log_2 \text{fold change} < 1$ ) are greyed out

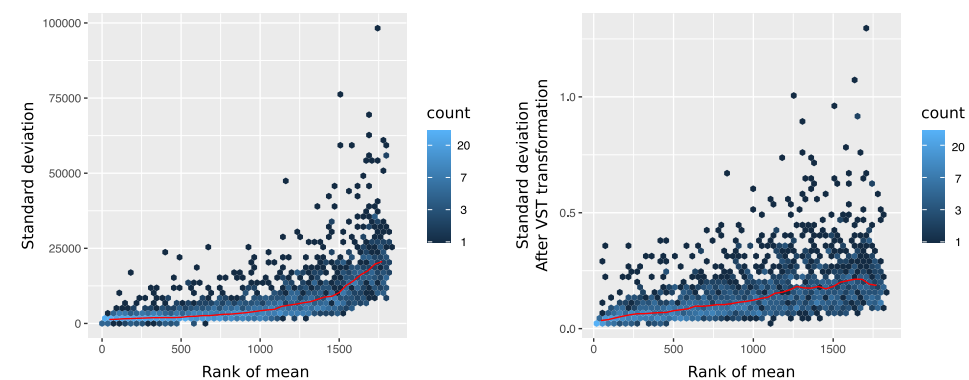
NA - No data

Complete table can be accessed via:

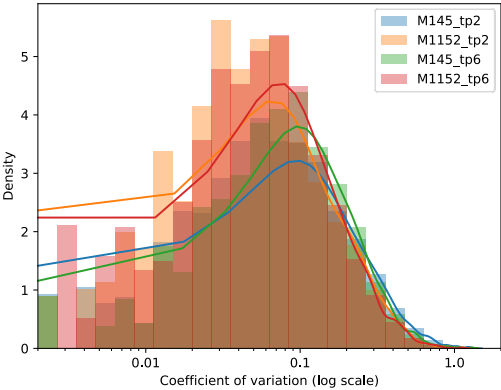
[https://osf.io/e352b/?view\\_only=ac6b5126cb5e4523b56bd55231365a51](https://osf.io/e352b/?view_only=ac6b5126cb5e4523b56bd55231365a51)



**Figure S1. Boxplot of protein levels in each sample before and after VST data transformation.** Outliers are indicated by black dots, the overall means of all proteins in each sample are indicated by green diamonds.



**Figure S2. Mean – standard deviation plot of protein levels in each sample before and after VST data transformation.** The red line depicts the running median estimator (window-width 10%).



**Figure S3. Distribution plot of coefficient of variation (ratio of SD to mean) of each quantified protein.**

### Supplementary materials for Chapter 6

# A novel nucleoid-associated protein specific to Actinobacteria that binds to GATC sequences

Chao Du, Joost Willemse, Victor J. Carrion, and Gilles P. van Wezel

Table S1. SCO1839 binding regions that overlap with promoter regions

Gene	Product	Translation start site on genome	Strand	25h peak fold enrichment	48h peak fold enrichment	25h peak summit position	25h peak width	48h peak summit position	48h peak width	...
SCO0002	ttrA (hypothetical protein)	1251	+	24.53	18.63	-5	304	+3	306	...
SCO0003	DNA- binding protein	3868	+	7.93	6.83	-167	234	-167	255	...
SCO0008	hypothetical protein	11360	-	10.32	4.37	-93	361	-139	285	...
SCO0019	hypothetical protein	19771	-	4.17	5.97	-148	260	-136	533	...
SCO0020	transposase	19973	+	4.17	5.97	-54	260	-66	533	...
SCO0021	hypothetical protein	21616	+	4.91	7.22	+3	205	+5	266	...
SCO0026	hypothetical protein	23315	+	3.68	6.33	-9	172	-6	228	...
SCO0044	Oxidore- ductase	36674	-	5.23	4.32	-27	214	-27	230	...
...	...	...	...	...	...	...	...	...	...	...

-350 to +50 relative to translation start site was considered as promoter region

"Peak" in this table refers to "binding region"

An "overlap" represents > 50% of the peak region overlaps with the promoter region

Complete table can be accessed via:

[https://osf.io/8ds2m/?view\\_only=8710562f6e4043da8d2bfc4a31dab7d8](https://osf.io/8ds2m/?view_only=8710562f6e4043da8d2bfc4a31dab7d8)

Table S2. AtrA binding motif used to predict the binding site in PREDetector

Name	Sequence	Ref
nagE2_AtrAMotif	GGAATCACGGGTTCC	Nothaft <i>et al.</i> (2010)
actII-ORF4_AtrAMotif	GGAATGCCAGATTCT	Uguru <i>et al.</i> (2005)
predicted	GGAACCACCGGTTCC	Hiard <i>et al.</i> (2007)
strR_AtrAMotif	GGAGGGGGCCGTTCC	Hirano <i>et al.</i> (2008)

Table S3. Bacterial strains used in this study

Strain	Genotype/description	Reference/vendor reference
E. coli JM109	See reference	Sambrook <i>et al.</i> (1989)
E. coli ET12567	See reference	MacNeil <i>et al.</i> (1992)
E. coli ET12567/pUZ8002	See reference	Flett <i>et al.</i> (1997)
E. coli BL21 CodonPlus (DE3)-RIPL	See reference	Agilent 230280
S. coelicolor A3(2) M145	See reference	Kieser <i>et al.</i> (2000)
S. coelicolor M512	See reference	van Wezel <i>et al.</i> (2000b)
GAD003	M145ΔSCO1839	This study
GAD014	M145ΔSCO1839 + pGWS1260	This study
GAD039	Part of SCO1839 promoter replaced by PerME	This study
GAD043	3×FLAG fused to SCO1839 C-term	This study
GAD099	eGFP fused to SCO1839 C-term	This study
GAD075	M145ΔatrA	This study
GAD078	M145ΔatrAΔSCO1839	This study
GAD105	M512 + pGWS1454	This study
GAD093	M512ΔatrA	This study
GAD107	M512ΔatrA + pGWS1454	This study

Table S4. Plasmids and constructs used in this study

Plasmid and construct	Description	reference/vendor ID
pWHM3	<i>E. coli</i> / <i>Streptomyces</i> shuttle vector, high copy number and unstable in <i>Streptomyces</i>	Vara <i>et al.</i> (1989)
pUWL-Cre	<i>E. coli</i> / <i>Streptomyces</i> shuttle vector expressing the Cre recombinase in <i>Streptomyces</i>	Fedoryshyn <i>et al.</i> (2008)
pHJL401	<i>E. coli</i> / <i>Streptomyces</i> shuttle vector, 5-10 copies per chromosome in <i>Streptomyces</i>	Larson and Hershberger (1986)
pHM10a	<i>E. coli</i> / <i>Streptomyces</i> shuttle vector, designed for gene over-expression using consecutive promoter <i>PermE</i>	Motamedi <i>et al.</i> (1995)
pCRISPomyces-2	<i>E. coli</i> / <i>Streptomyces</i> shuttle vector, harbouring codon optimised <i>cas9</i> , designed for easy inserting spacer sequences. Recombination template is designed to be inserted at <i>Xba</i> I site.	Cobb <i>et al.</i> (2015)
pCRISPR-Cas9	<i>E. coli</i> / <i>Streptomyces</i> shuttle vector, harbouring codon optimised <i>cas9</i> , designed for easy inserting spacer sequences. Recombination template is designed to be inserted through in-vitro assembly to <i>Stu</i> I site	Tong <i>et al.</i> (2015a)
pIJ2587	<i>E. coli</i> / <i>Streptomyces</i> shuttle vector, harbouring <i>redD</i> with multiple cloning sites in front for inserting external promoter sequence	van Wezel <i>et al.</i> (2000b)
pGWS526	pHJL401 expressing SsgA-eGFP	Zhang <i>et al.</i> (2016)
pGWS728	Construct harbouring <i>aac(3)/IV</i>	Zhang <i>et al.</i> (2018)
pET28a	<i>E. coli</i> vector, designed to build His-tag fusion protein expression construct	Novagene 69864-3
pUC19	<i>E. coli</i> vector with multi-copy origin of replication	NEB N3041
pGWS1255	pWHM3 containing flanking regions of SCO1839 with apramycin resistance cassette with <i>loxP</i> sites inserted as <i>Xba</i> I fragment between flanking regions	This study
pGWS1260	pHJL401 harbouring SCO1839 and its own promoter region	This study
pGWS1298	pCRISPomyces-2 with spacer sequence from near the end of SCO1839, containing recombination template for 3×FLAG tag knock-in	This study
pGWS1299	pCRISPomyces-2 with spacer sequence from near the end of SCO1839, containing recombination template for eGFP knock-in	This study
pGWS1295	pCRISPomyces-2 with spacer sequence from near the beginning of SCO1839, containing recombination template for <i>PermE</i> knock-in	This study
pGWS1286	pET28a with SCO1839 coding sequence built-in, for 6His-SCO1839 fusion protein expression	This study
pGWS1300	pUC19 harbouring partial SCO1839 promoter region for EMSA experiment	This study
pGWS1451	pUC19 harbouring random SCO1839 non-binding region for EMSA experiment	This study
pGWS1452	pCRISPR-Cas9 with spacer sequence from near the beginning of <i>atrA</i> , containing recombination template for <i>atrA</i> knock-out	This study
pGWS1454	pIJ2587 with the promoter region of SCO1839	This study



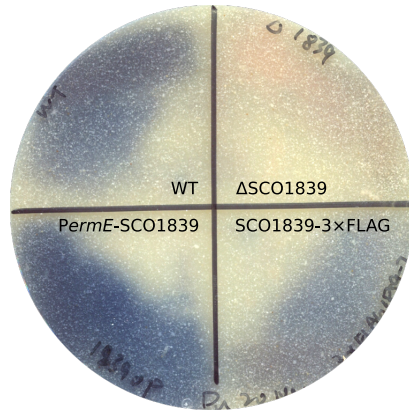
Table S5. Primers used in this study

Purpose	No. Name	Sequence 5'-3'	Note
SCO1839 knock-out and complementation	1 SCO1839FU_F	CGATGATCCGACGCGGACTCGATCATCTG	No.1 changed to XbaI
	2 SCO1839FU_R	CGATTCTAGACCGTGCTCTCTCATGGGAAG	
	3 SCO1839FD_F	CGATTCTAGAGCCACTTCGGCCTGACGTCC	
	4 SCO1839FD_R	CGATAAGCIIICGGCCTGCTCGACGAAAC	
	5 SCO1839CM_F	ACTGAAAGCIIIGGACGTCAGGCCGAAGTG	
	6 SCO1839CM_R	ACTCGGATCCGGTCTCTCGGCCAACGTG	
SCO1839 over-expression	7 1839_Ukp_F	CCGATCTAGACGACGCCGACTCGATCATCTG	No.4 changed to XbaI Spacer assembly Spacer assembly, also used in eGFP knock-in
	8 1839_Ukp_R	CAGTGAGCTCTCCGCCGAACGAGTTTCTCC	
	9 SCO1839OP_F	CAGTCATATGCCCGAGACTCTGAAGAGGG	
	10 SCO1839FD_R_XbaI	CGATTCTAGACCGGCCTGCTCGACGAAAC	
	11 Sp_1839U_F	ACGCTCGGCAAGGCTGATGAGACA	
	12 Sp_1839U_R	AAACTGTCTCATCAGCCTTGCCGA	
3xFLAG tag knock-in	13 1839sp_down_F	ACGCGCCACTTCGGCCTGACGTCC	BamHI locates inside PCR product
	14 1839sp_down_R	AAACGGACGTCAGGCCGGAAGTGGC	
	15 1839flagFL_UR	AGTCGCGCTCGTGGTCTTGTAGTCGGCCGAAGTGGCTTCTTTCG GACTACAAAGGACCACGACGGCGACTACAAAGGACCACGACATCGACTACAA GGACGACGACGACAAGTGACGTCCGGCGGCCCGCCGGCTCCGAAAGTG ACGGTGGCCACCCCGGT	
	16 3xFLAG+1839ending	GCGGCCCTTTACGGTTCCTGGCCTCTCGTAGTCCCGCGCTCTGG	
	17 1839flagCR_DR	CGATTCTAGATGGTCTCTCGGCCAACGTG	
	18 1839flagCR_UF_xbaI	CGATTCTAGACCTCTGTAGTCCCGCGCTCTGG	
eGFP knock-in	19 1839flagCR_DR_xbaI	CGATGGATCCGCCGGAAGTGGCCTTCTTGC	BamHI locates inside PCR product
	20 1839GFPF_UR_BamHI	ATCGAAAGCIIIGTCCGAAAGTGACGGTGGC	
	21 1839GFPF_DF_HindIII	TGGACGTGCCGGGACTTCCTG	
	22 eGFP(ftsZ)_F	ATCGAAAGCIIIGGCCGCTTTACTTTGTACAGCTC	
	23 eGFP_R_HindIII	CTCGACGCCGTCCGGAAGCAGATGATCCGTTCCGGCTGGAGCGCCTCGA	
	24 p1839_50Ssingle_strong	TCGAGGGGCTCCAGCCGGAACGGATCATCTGCTTCCGGACGGCGTCTCGAG	
50 bp EMSA fragments	25 p1839_50Ssingle_strong_R		

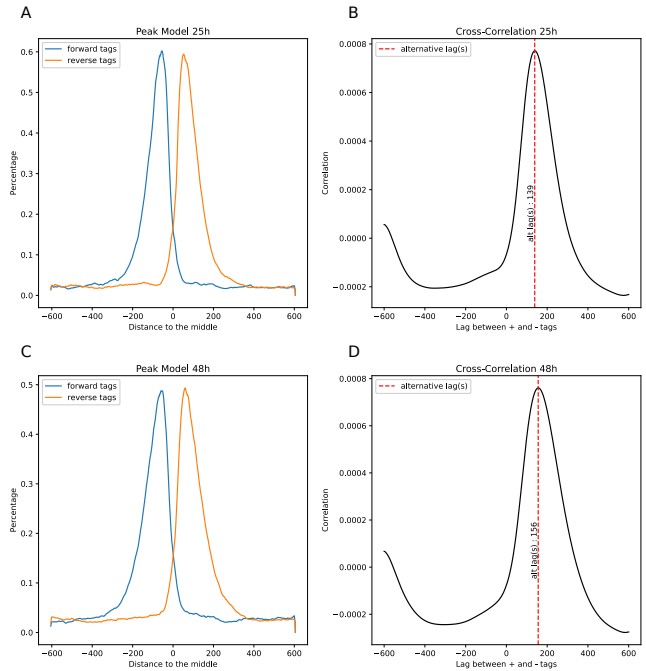
26	p1839_50Single_weak	TCGATGTAGGACACACCTCTTGTGATGAGGAGATCTCACACAGATGGCGGAT	
27	p1839_50Single_weak_R	ATCGCCATCTGTGTGAGATCTCCTCATCAAGAGGTGTGTCTTACATCGA	
28	p1839_50Quadruple	GATGGCGATCACGGATCGGCCGAATGATCCATACCCAGTGATCATCCA	
29	p1839_50Quadruple_R	TGGATGATCCACTGGTTATGGATCATTCGGCCGATCCGTGATCCGCCATC	
30	npio_B50	GCAATTCGTAGGACGGAAGTGCACAGGTGCCCTCAAAATCGGCCCTATAG	Negative control
31	npio_B50_R	CTATAGGGCCGCAATTTGAGGCACCCCTGCGCACITCCGTCTTACGAAATTGC	EcoRI site inside PCR product, the fragment was digested using EcoRI and HindIII for inserting into pUC19
32	p1839_ip_F	CCGCCGAACGAGTTTCTCC	
33	p1839_ip_R_HindIII	GCATAAGCTTCCTCGGCCAACGTGCTG	Negative control
34	noip_A_F_EcoRI	GCATGAATTCGCCGCCGTCGTTGGTGTGTC	
35	noip_A_R_HindIII	GCATAAGCTTGCCAGCGAGGCCGCTTC	
36	pSCO1839_F_BamHI	CGATGGATCCCTCATATGGGAAGTGCCTCTG	
37	pSCO1839_R_SacI	ACGTGAGCTCTCCGTGATCCGCCCATCTGTG	
38	SCO1839_QF	GAGCATGCGGTGCACAAAG	
39	SCO1839_QR	GCGGCAGACCTGAAGAAGAAG	
40	SCO3873_qF	GCGTGGTGGACACGAAGAAG	Inner control
41	SCO3873_qR	GCACCAAGACCCGACGACTAC	
42	SCO5359_qF	TACGTCGAGACGCAGGTACG	Inner control
43	SCO5359_qR	CTGCTTGCCCGGTAGAAC	
44	AtrA_CrSP	CATGCCAATGGTGCATGTTCCAGGATTCTCATGTTTTAGAGCTAGAAATAGC	Spacer sequence in red
45	atrA_UgR	TCGTGGAAGGCACIAGAAAGGACGCCTCGACGAGGATCCAC	
46	atrA_UgF	CGTACTCATCGCACCCGCCGCCGCCATCGAAAGCCGGAG	
47	atrA_DgR	CCGGCGGTGCGATGAGTACG	
48	atrA_DgF	GGTCGATCCCCGCATATAGGACGTACCCGGGACCGCTCTC	

Underlined characters indicating restriction sites or overhangs for assembly.

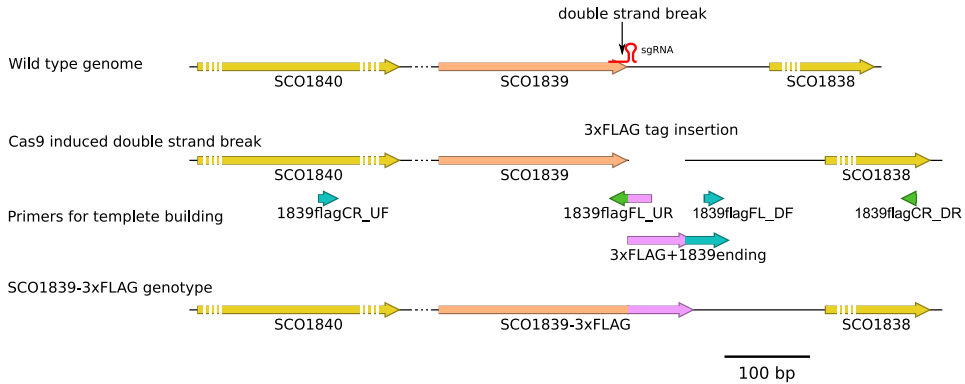
GGATCC, BamHI; GAATTC, EcoRI; AGGCGT, StuI; TCTAGA, XbaI; GAGCTC, SmaI; CATATG, NdeI; AAGCTT, HindIII



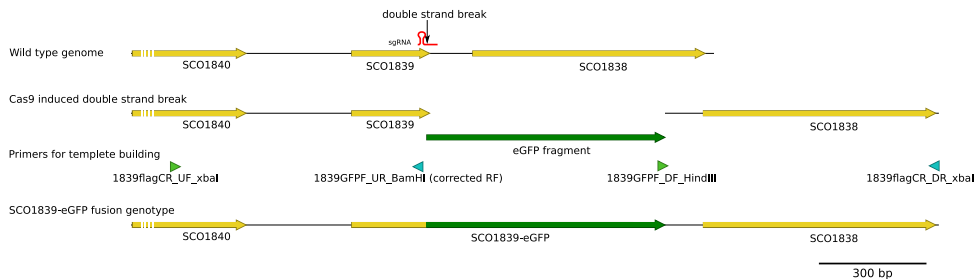
**Figure S1. FLAG-tagged SCO1839 does not affect SCO1839 function.** A strain expressing SCO1839-3×FLAG (lower right) produces a similar amount of blue-pigmented actinorhodin (Act) as the parental strain (upper left). Compare the SCO1839 deletion strain (upper right), which produces reduced amounts of Act, and a strain over-expressing SCO1839 (lower left) that shows enhanced Act production.



**Figure S2. MACS model of 25 h and 48 h ChIP-Seq data.** A and B, 25 h ChIP model, C and D, 48 h ChIP model. A and C, 5' ends of strand-separated tags (reads) from a random sample of 1,000 model peaks, aligned by the centre of their Watson (forward) and Crick (reverse) peaks. B and D, combined correlation peak from both strands, showing the distance (lag) of enriched tags to the centre of predicted binding region.



**Figure S3. Schematic of construct to allow 3xFLAG tag integration at the C-terminal end of SCO1839.** Spacer sequence locates at the end of SCO1839, the sgRNA with this spacer guides the Cas9 protein to make a double-strand break after the stop codon. Templates for homology-directed repair (HDR) were made by cloning SCO1839 and its upstream region from genome with additional connecting sequence, thus replacing the stop codon for connecting with 3xFLAG sequence. The downstream region was PCR-amplified from the genome proceeded by the sequence for a full 3xFLAG sequence. Then these two fragments were connected by overlap extension PCR.



**Figure S4. Schematic of construct to allow eGFP integration at the C-terminal end of SCO1839.** The spacer sequence is fused behind SCO1839, allowing the sgRNA with this spacer to guide Cas9 protein to make a double-strand break after the stop codon. Templates for homology-directed repair (HDR) were made by connecting the SCO1839 region, the eGFP fragment, and the downstream region of SCO1839. The sequence for eGFP was cloned from pGWS526 expressing FtsZ-eGFP as described in Zhang *et al.* (2016).

## Curriculum vitae

Chao Du was born on the 15<sup>th</sup> of August 1987 in Dalian, Liaoning province, China. In the year 2006, he was admitted by Southwest University in Chongqing municipality, China, and obtained Bachelor of Science on the 21<sup>th</sup> of June 2010. He then performed his research as a master student under the supervision of Prof. dr. Junrong Liang at Diatom Laboratory, Xiamen University, Xiamen, Fujian province, China. On 30<sup>th</sup> of September 2013, he finished his research on diatom growth and development and obtained his master's degree. Then, he participated in phytoplankton investigation cruises in the coastal area of Fujian province, China. On the 15<sup>th</sup> of September 2014, he joined Leiden University as a PhD student under the supervision of Prof. dr. Gilles P. van Wezel. Here in Leiden, he performed studies on the development and antibiotics production in *Streptomyces*, the results are presented in this thesis. Currently, he is performing his post-doctoral research in the laboratory of Prof. dr. Gilles P. van Wezel.

## List of publications

**Du, C.**, and van Wezel, G.P. (2018) Mining for microbial gems: Integrating proteomics in the postgenomic natural product discovery pipeline. *Proteomics* 18: 1700332.

Zhang, Z., **Du, C.**, de Barsy, F., Liem, M., Liakopoulos, A., van Wezel, G.P., Choi, Y.H., Claessen, D., and Rozen, D.E. (2020) Antibiotic production in *Streptomyces* is organized by a division of labor through terminal genomic differentiation. *Science Advances* 6: eaay5781.

Wu, C., **Du, C.**, Ichinose, K., Choi, Y.H., and van Wezel, G.P. (2017) Discovery of C-Glycosylpyranonaphthoquinones in *Streptomyces* sp. MBT76 by a combined NMR-based metabolomics and bioinformatics workflow. *J. Nat. Prod.* 80: 269-277.

Wu, C., **Du, C.**, Gubbens, J., Choi, Y.H., and van Wezel, G.P. (2015) Metabolomics-driven discovery of a prenylated isatin antibiotic produced by *Streptomyces* species MBT28. *J. Nat. Prod.* 78: 2355-2363.

Sulheim, S., Kumelj, T., van Dissel, D., Salehzadeh-Yazdi, A., **Du, C.**, van Wezel, G.P., Nieselt, K., Almaas, E., Wentzel, A., and Kerkhoven, E.J. (2020) Enzyme-constrained models and omics analysis of *Streptomyces coelicolor* reveal metabolic changes that enhance heterologous production. *iScience*: 101525.

Xiao, X., Elsayed, S.S., Wu, C., van der Heul, H.U., Metsä-Ketelä, M., **Du, C.**, Prota, A.E., Chen, C., Liu, W., Guo, R., Abrahams, J.P., and van Wezel, G.P. (2020) Functional and structural insights into a novel promiscuous ketoreductase of the lugdunomycin biosynthetic pathway. *ACS Chem. Biol.*

**Du, C.**, Liang, J., Chen, D., Xu, B., Zhuo, W., Gao, Y., Chen, C., Bowler, C., and Zhang, W. (2014) iTRAQ-based proteomic analysis of the metabolism mechanism associated with silicon response in the marine diatom *Thalassiosira pseudonana*. *J. Proteome Res.* 13: 720-734.