

Aus dem Institut für Insektenbiotechnologie
Professur für Naturstoffforschung mit Schwp. Insektenbiotechnologie
der Justus-Liebig-Universität Gießen

Heterologous expression of bacterial natural product biosynthetic gene clusters



INAUGURAL-DISSERTATION

zur Erlangung des Doktorgrades (Dr. rer. nat.)
im Fachbereich Agrarwissenschaften, Ökotoxologie und
Umweltmanagement der Justus-Liebig-Universität Gießen

vorgelegt von

Zerlina Gabriela Wuisan

aus

Indonesien

Gießen, im September 2021

Mit Genehmigung des Fachbereichs Agrarwissenschaften, Ökotoxikologie und
Umweltmanagement der Justus-Liebig-Universität Gießen

1. Gutachter

Professor Dr. Till F. Schäberle

Institut für Insektenbiotechnologie

Fachbereich Agrarwissenschaften, Ökotoxikologie
und Umweltmanagement

Justus-Liebig-Universität Gießen

2. Gutachterin

Professor Dr. Gabriele M. König

Institut für Pharmazeutische Biologie

Mathematisch-Naturwissenschaftlichen Fakultät

Rheinische Friedrichs-Wilhelms-Universität Bonn

Erklärung gemäß der Promotionsordnung des Fachbereichs 09 vom 07. Juli 2004 § 17 (2)

Ich erkläre:

Ich habe die vorgelegte Dissertation selbständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe.

Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht.

Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten.

Gießen, _____

Zerlina Gabriela Wuisan

Acknowledgements

Herewith I would like to express my gratitude towards everyone who made this work possible and contributed to its accomplishment.

First and foremost, praise to God, the Almighty, for His blessings. He has provided me with more than I could ever have imagined that I could finally finish this journey.

My deepest gratitude is dedicated to Prof. Dr. Till F. Schäberle for his support and guidance since the beginning of this journey until the very end. His hardwork, empathy, and friendliness has shaped a motivating workplace environment that I am very grateful of. This work is only possible to be concluded by his encouragement and constructive advice.

I would like to express my sincere gratitude to Prof. Dr. Gabriele M. König for the opportunity given to me to join the INDOBIO project. I appreciated all help, support and kindness of hers, especially to readily officiate as the second reviewer.

I am highly thankful to Prof. Dr. Heike Wägele for the INDOBIO collaboration and her help my study, especially during the beginning of my study in Bonn.

I would also like to extend my sincere gratitude to the two referees and chairperson of the examination committee.

I am also grateful to my colleagues in the working group of Prof. Schäberle: Nils Böhringer, Riyanti, I Dewa Made Kresna, and Jil-Christine Kramer for the fruitful scientific discussion and for being good friends. I thank Dr. Yang Liu, Dr. Ute Mettal, Lei Wang, and Jean-Marie Pohl for their valuable assistance in mass spectrometry analysis and chemical compound purification; Dr. Katja Fisch and Dr. Luis Linares-Otoya for their support in the molecular biology field. I thank Ms. Alexandra Bender for her help in administrative process, and to all the visiting researchers in Schäberle working group who had enriched our research group with productive collaborations.

Many thanks should also go to the former and current employees of Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Dr. John Heep, Dr. Christoph Hartwig, Dr. Marius Spohn, Dr. Philipp Heise, Dr. Michael Marner, Dr.

Markus Oberpaul, and Regina Zweigert for the great teamwork and generous help throughout my study.

I also appreciate my colleagues from the working group of Prof. König, where I spent the first six months of my PhD research: Dr. Max Crüsemann, Dr. Jamshid Amiri Moghaddam, Dr. Cora Hertzner, and Emilie Goralski.

I thank the Federal Ministry of Education and Research (BMBF) and the German Academic Exchange Service (DAAD) for the financial assistance.

I would like to thank my parents and my sister for believing in me and supporting me throughout my study. I am grateful to my friends in Komsel Giessen for their prayers and moral support.

To those who I might have not mentioned but have supported me in one way or another, I do appreciate the kindness and friendship.

Abstract

Antimicrobial resistance (AMR) has been declared as a global threat by the United Nations (UN). Therefore, it is of an utmost importance to find novel bioactive compounds that could be developed into a drug lead. As the conventional antibiotic discovery pipeline dried up, a synthetic biology-based strategy has become a promising alternative, due to the advances in sequencing technology. The strategy includes genome mining to detect putative biosynthetic gene clusters (BGCs), genetic engineering (e.g. to add promoters) and cloning of the BGCs into expression vectors, and heterologous expression to identify compounds that are produced by the BGCs. This strategy is advantageous to obtain novel compounds from cryptic or silent BGCs, as well as to prove the link between a BGC and a compound of interest.

This study is separated into two main parts: Chapter III: Heterologous expression of non-ribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) BGCs and Chapter IV: Heterologous expression of ribosomally synthesized and post-translationally modified peptide (RiPP) BGCs.

In Chapter III, the genomic DNA of bioactive understudied bacteria from unusual sources, *i.e.* the egg mass of marine slugs and the gut of burying beetles, were analysed *in silico* for putative NRPS and PKS BGCs. The putative BGCs were successfully cloned into expression vectors, and synthetic promoters were inserted. Then, the integrative vectors that carry the respective putative BGCs were introduced into different *Streptomyces* heterologous expression hosts. Although the heterologous expression of the BGCs did not result in the detection of natural products (NPs) that could be linked to the respective BGC, the methods applied and discussed here were proven to be efficient in capturing of large BGCs from Actinobacteria. Moreover, the constructs that were generated can be used as the basis for future studies. The prospective work to improve the heterologous expression of NP BGCs is discussed in section III.4.3.

In Chapter IV, the heterologous expression of a RiPP BGC was aimed to link it to a compound of interest, *i.e.* darobactin A. Additionally, the minimal BGC to produce darobactin A was determined and the heterologous expression system was optimized to increase production yield and decrease fermentation period to ensure compound supply for future animal studies. The heterologous expression system developed in this study has become the basis for the heterologous expression of natural darobactin analogs¹, marine-derived darobactins², and the bioengineering of derivatives libraries³.

In conclusion, the synthetic biology-based approach for the discovery of novel antibiotics is an advancing field. The tools for better BGC prediction are continuously being perfected and with the bioinformatic tools that are now available, putative BGCs could be predicted, unveiling the potential of NP discovery from understudied bacteria. The generation of a defined metagenomic library as the DNA source for capturing target BGCs was proven to increase TAR cloning efficiency up to 92.8%. However, the optimization of the heterologous system is necessary for a successful biosynthesis, especially for the heterologous expression of NP BGCs from understudied bacteria. In contrast, the heterologous expression of BGCs from well-studied bacteria of the genus *Photorhabdus* in the comprehensively studied host *Escherichia coli* has proven the functionality of the darobactin A BGC.

¹ Nils Böhringer, Robert Green, Yang Liu, Ute Metta, Michael Marner, Seyed Majed Modaresi, Roman P. Jakob, Zerlina G. Wuisan, Timm Maier, Akira Inishi, Sebastian Hiller, Kim Lewis and Till F. Schäberle. 2021. **Mutasynthetic production and antimicrobial characterisation of Darobactin analogs.** (under review)

² Nils Böhringer, Jil-Christine Kramer, Leo Padva, Zerlina G. Wuisan, Yang Liu, Michael Marner, and Till F. Schäberle. 2021. **Genome- and Metabolome-Guided Discovery of a Dedicated Darobactin Halogenase Provides Biosynthetic Insights into Marine Bama Inhibitors.** (under review)

³ I Dewa Made Kresna, Zerlina G. Wuisan, ..., and Till F. Schäberle. 2021. **In vivo characterization of DarE flexibility towards DarA and study of structure-activity relation of the derivatives.** (publication in preparation)

Parts of this study have been published in:

Imai, Y., Meyer, K.J., Inishi, A., Favre-Godal, Q., Green, R., Manuse, S., Caboni, M., Mori, M., Niles, S., Ghiglieri, M., Honrao, C., Ma, X., Guo, J.J., Makriyannis, A., Linares-Otoya, L., Böhringer, N., Wuisan, Z.G., Kaur, H., Wu, R., Mateus, A., Typas, A., Savitski, M.M., Espinoza, J.L., O'Rourke, A., Nelson, K.E., Hiller, S., Noinaj, N., Schäberle, T.F., D'Onofrio, A., Lewis, K., 2019. **A new antibiotic selectively kills Gram-negative pathogens.** *Nature* 576, 459–464. <https://doi.org/10.1038/s41586-019-1791-1>

Wuisan, Z.G., Kresna, I.D.M., Böhringer, N., Lewis, K., Schäberle, T.F., 2021. **Optimization of heterologous Darobactin A expression and identification of the minimal biosynthetic gene cluster.** *Metabolic Engineering* 66, 123–136. <https://doi.org/10.1016/j.ymben.2021.04.007>

Table of Contents

| | |
|--|-----------|
| I. GENERAL INTRODUCTION | 1 |
| I.1. ANTIBIOTICS DISCOVERY AND RESISTANCE | 1 |
| I.2. SECONDARY METABOLITES AS SOURCE OF ANTIBIOTICS | 5 |
| I.2.1. <i>Polyketides (PKs)</i> | 6 |
| I.2.2. <i>Non-ribosomal peptides (NRPs)</i> | 10 |
| I.2.3. <i>Ribosomally synthesized and post-translationally modified peptides (RiPPs)</i> | 13 |
| I.3. HETEROLOGOUS EXPRESSION STRATEGY FOR SECONDARY METABOLITE PRODUCTION | 16 |
| I.3.1. <i>Genome mining</i> | 18 |
| I.3.2. <i>Cloning and heterologous expression of BGC</i> | 19 |
| II. SCOPE OF THE STUDY | 23 |
| III. HETEROLOGOUS EXPRESSION OF NRPS AND PKS BGCS | 25 |
| III.1. INTRODUCTION | 25 |
| III.1.1. <i>INDOBIO project</i> | 25 |
| III.1.2. <i>Burying beetle project</i> | 28 |
| III.2. MATERIAL AND METHOD | 32 |
| III.2.1. <i>In silico analysis of putative BGC</i> | 32 |
| III.2.2. <i>Construction of Gordonia library</i> | 32 |
| III.2.3. <i>Colony PCR</i> | 37 |
| III.2.4. <i>Fosmid/plasmid isolation</i> | 38 |
| III.2.5. <i>TAR cloning</i> | 39 |
| III.2.6. <i>Gibson assembly</i> | 42 |
| III.2.7. <i>λ-Red Recombination for promoter insertion</i> | 43 |
| III.2.8. <i>Triparental conjugation</i> | 43 |
| III.2.9. <i>Gordonia cultivation for nocobactin NA dereplication</i> | 44 |
| III.2.10. <i>Heterologous expression and compound detection</i> | 44 |
| III.3. RESULTS | 46 |
| III.3.1. <i>INDOBIO Project</i> | 46 |
| III.3.2. <i>Burying beetle project</i> | 70 |
| III.4. DISCUSSION..... | 78 |
| III.4.1. <i>INDOBIO Project</i> | 78 |
| III.4.2. <i>Burying beetle Project</i> | 82 |
| III.4.3. <i>Prospective work to improve heterologous expression of NP BGC</i> | 84 |
| III.5. SUMMARY | 87 |

| | | |
|-------------|--|------------|
| IV. | HETEROLOGOUS EXPRESSION OF RIPP BGC | 89 |
| IV.1. | INTRODUCTION | 89 |
| IV.2. | MATERIAL AND METHODS | 93 |
| IV.2.1. | <i>Plasmid and strain construction</i> | <i>93</i> |
| IV.2.2. | <i>Heterologous expression for functional proof of DAR BGC</i> | <i>96</i> |
| IV.2.3. | <i>Qualitative analysis of different promoters</i> | <i>97</i> |
| IV.2.4. | <i>Generation of E. coli mutant strains as heterologous hosts.....</i> | <i>98</i> |
| IV.2.5. | <i>Heterologous host strains, media and maintenance.....</i> | <i>99</i> |
| IV.2.6. | <i>Heterologous expression for optimization of DAR production</i> | <i>100</i> |
| IV.2.7. | <i>UPLC-HRMS analysis</i> | <i>101</i> |
| IV.2.8. | <i>Determination of E. coli survival rate in the presence of DAR</i> | <i>102</i> |
| IV.2.9. | <i>Expression of DarA.....</i> | <i>102</i> |
| IV.2.10. | <i>Trypsin digestion</i> | <i>103</i> |
| IV.2.11. | <i>Quantification of darA and darE expression level.....</i> | <i>104</i> |
| IV.3. | RESULTS | 106 |
| IV.3.1. | <i>Functional prove of the DAR BGC in E. coli.....</i> | <i>107</i> |
| IV.3.2. | <i>Selection of the heterologous producer strain.....</i> | <i>108</i> |
| IV.3.3. | <i>Generation of an E. coli strain with increased DAR resistance level</i> | <i>110</i> |
| IV.3.4. | <i>Comparison of different DAR BGCs.....</i> | <i>112</i> |
| IV.3.5. | <i>Influence of the transporter-encoding genes darBCD</i> | <i>114</i> |
| IV.3.6. | <i>A dedicated peptidase is not part of the BGC</i> | <i>115</i> |
| IV.3.7. | <i>Integration of an additional darA copy.....</i> | <i>116</i> |
| IV.3.8. | <i>Integration of native darA upstream regions</i> | <i>117</i> |
| IV.3.9. | <i>Quantification of darA and darE transcription.....</i> | <i>118</i> |
| IV.4. | DISCUSSION..... | 120 |
| IV.5. | SUMMARY | 126 |
| V. | CONCLUSION | 127 |
| VI. | REFERENCES..... | 128 |
| VII. | APPENDIX..... | 149 |

I. General Introduction

I.1. Antibiotics discovery and resistance

The discovery of penicillin by Alexander Fleming in 1928 marks the start of antibiotic era (Fleming, 1929). Thereafter, the systematic screening that was developed by Selman Waksman initiated the “Golden Era of Antibiotic Discovery” between 1940 to 1960. The Waksman antibiotic discovery platform had a similar setup to the serendipitous finding by Fleming and was based on the inhibition zone of soil-derived actinomycetes against susceptible microorganisms on agar plates (Schatz *et al.*, 1944). The method led to the discovery of streptomycin, the first antibiotic against tuberculosis. Since then, new classes of antibiotics were discovered and new antibiotics were introduced to the healthcare system.

Antibiotics have significantly increased human life expectancy. In 1900, prior to the antibiotic era, human life expectancy was only 47.3 years, with infectious diseases, *e.g.* pneumoniae, tuberculosis, influenza, and diarrhea, representing the most common cause of death. In 1970, directly after the Golden Era of Antibiotics, the life expectancy in the United States had increased to 70.8 years, and up to 78.7 years in 2018. The most common cause of death had also shifted to non-infectious diseases, such as heart diseases, cancer, cerebrovascular diseases, diabetes, and respiratory diseases (“Mortality and Cause of Death, 1900 v. 2010,” 2014; “NVSS - Mortality - Historical Data,” 2019).

However, the antibiotic discovery pipeline used to discover novel antibiotics in the “Golden Era of Antibiotic Discovery” was drained after 20 years. Rediscovery of known compounds occurred more often and the discovery identification of new antibiotic classes became increasingly difficult. After the Golden Era, only two new clinically relevant antibiotic classes were discovered, lipopeptides (daptomycin) in 1986 and diarylquinolines (bedaquiline) in 1997 (**Figure I.1**) (Lewis, 2013).

Antibiotic resistance as a threat to antibiotic therapy has appeared early, even before the introduction of antibiotics for clinical use. From bacterial growth inhibitory assay, Fleming noticed a coli-typhoid bacterial group that was not affected by penicillin (Fleming, 1929). A further study on *Bacillus coli* (*Escherichia coli*) identified an enzyme that inactivates penicillin (Abraham and Chain, 1940). Shortly after penicillin was introduced to treat staphylococcal infection, penicillin-resistant

Staphylococcus aureus strains were observed from patients that received penicillin therapy. It was also observed *in vitro* that prolonged exposure to penicillin with increasing concentration could render resistance (Rammelkamp and Maxon, 1942). Since then, antibiotic resistance has appeared almost in parallel to the discovery of antibiotics (**Figure I.1**).

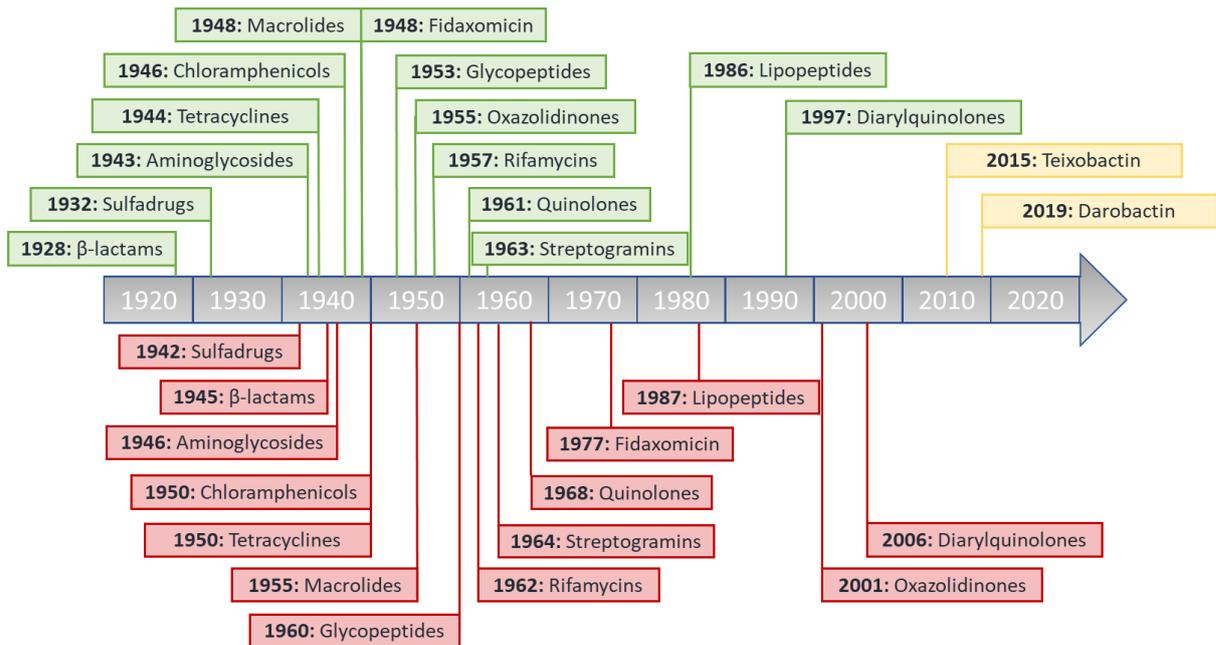


Figure I.1: Timeline of antibiotic discovery and resistance. This timeline shows the discovery year of clinically useful antibiotic classes (in green), and the year when the first resistance was reported (in red). The time frame of 1940 – 1960 marks the “Golden Era of Antibiotic Discovery”, when most of the clinically relevant antibiotic classes were discovered. Thereafter, the discovery pipeline faltered, and only two new classes were reported after the “Golden Era”. Resistance towards the antibiotics was also observed almost in parallel to the antibiotic discovery. Two new antibiotics with novel mode-of-action (in yellow) might mark the discovery of new antibiotic classes.

The natural resistance mechanisms had been available in the bacterial community even before the clinical use of antibiotics (Barlow and Hall, 2002). The resistance mechanisms are usually used by the bacteria to defend themselves from the compounds that they produced and to survive in the microbial community. They could also acquire the resistance through horizontal gene transfer or by genetic mutations (Reygaert, 2018). Moreover, the anthropogenic activity, e.g. unregulated use of antibiotics, inappropriate prescription, and extensive use in agriculture, was

believed to significantly contribute to the emergence of the antibiotic resistance crisis (Ventola, 2015).

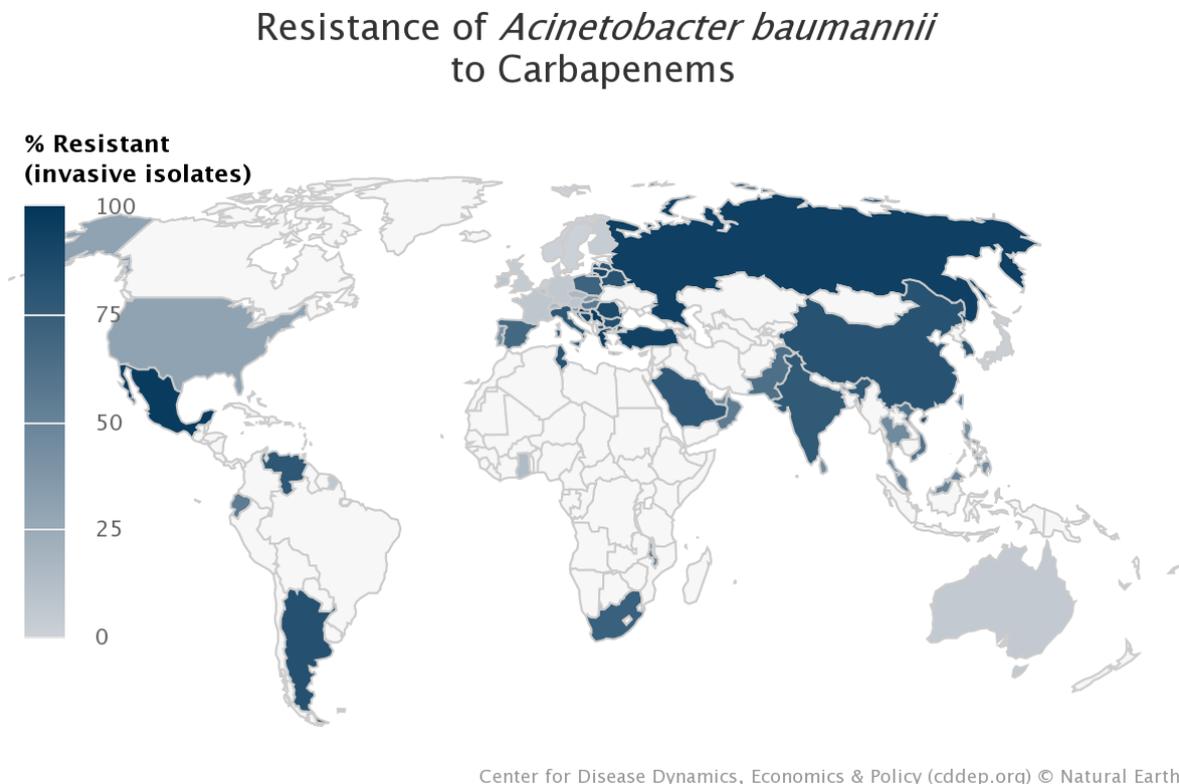


Figure I.2: Resistance map of *Acinetobacter baumannii* to carbapenems.

Note: Data for different countries are collected with different methodologies from 2014 to 2018. White area shows no data available. (“Resistance Map: Antibiotic Resistance,” 2021)

In 2016, the UN declared antimicrobial resistance (AMR) as a “fundamental threat” to global health, development, and security (WHO News Release, 2016), for example the resistance of *Acinetobacter baumannii* to carbapenems that can be observed across countries in Asia, America, Australia, Europe, and Africa (**Figure I.2**). Antimicrobial-resistant strains were responsible for more than 700,000 deaths globally (*Review on Antimicrobial Resistance: Tackling drug-resistant infections globally*, 2014). The treatment of common bacterial infections as well as life-threatening disease like tuberculosis and malaria became challenging due to AMR. A priority pathogens list, for which new antibiotics are urgently needed, was published by the World Health Organization (WHO News Release, 2017). The list includes *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, *Enterobacteriaceae*

in critical priority, and *Enterococcus faecium*, *Staphylococcus aureus*, *Helicobacter pylori*, *Campylobacter* spp., *Salmonellae*, *Neisseria gonorrhoeae* in high priority. The research and development of new antibiotics are now expected to be directed towards these pathogens.

The antibiotics discovered during the “Golden Era” were mostly isolated from soil-derived actinomycetes, detected through inhibition assays according to Waksman’s method. In order to find new antibiotics, new methods should be developed and new bioresources should be explored (Lewis, 2013). The discovery of teixobactin and darobactin are great examples, where potentially novel classes of antibiotics that target the priority pathogens set by WHO were discovered by integration of the iChip technology and exploration of nematode gut symbionts (Imai *et al.*, 2019; Ling *et al.*, 2015). The iChip technology allowed the isolation of formerly uncultured and unknown bacteria, *e.g.* *Eleftheria terrae*, which then led to the isolation of teixobactin, a peptide antibiotic without detectable resistance biosynthesized by non-ribosomal peptide synthetase pathway (Ling *et al.*, 2015). Meanwhile, the gut microbiome of nematodes, an under-explored source during the “Golden Age”, was screened applying the traditional Waksman’s method, with slight alteration that concentrated extracts were prepared from the bacterial cultures. This cumulated in the identification of darobactin, a selective Gram-negative antibiotic biosynthesized by post-ribosomal peptide synthesis pathway (Imai *et al.*, 2019).

I.2. Secondary metabolites as source of antibiotics

Secondary metabolites are compounds that are not necessary for the growth and reproduction of bacterial cultures, but are required for survival in the bacterial natural environment (Demain and Fang, 2000). They have diverse functions, such as (i) defence mechanism against other organisms, either their predator, prey, or competitor, (ii) metal chelating agent, (iii) signal for their symbionts, (iv) sexual hormones, (v) effectors for differentiation. These secondary metabolites are of great value for human health, especially the bioactive secondary metabolites, which function as antibiotics, antitumor, antiviral, and antifungal agents (Bérdy, 2005).

Most of the clinically relevant antibiotics that were discovered in the “Golden Era” are actinomycetes’ secondary metabolites, *i.e.* β -lactams, aminoglycosides, chloramphenicols, macrolides, tetracyclines, rifamycins, glycopeptides, streptogramins, lipopeptides, and fidaxomicin (de Crécy-Lagard *et al.*, 1997; Ehrlich *et al.*, 1947; Griffith, 1981; Hochlowski *et al.*, 1987; Jukes, 1985; Liras and Demain, 2009, p. 16; Margalith and Beretta, 1960; McGUIRE *et al.*, 1952; Raja *et al.*, 2003; Schatz *et al.*, 1944). The above-mentioned antibiotic classes are biosynthesized by two metabolic pathways: (i) the polyketide synthase (PKS)- and (ii) the non-ribosomal peptide synthetase (NRPS)-dependent pathway, except for the aminoglycosides, which are produced by the saccharides biosynthetic pathway. In addition to these ones, further pathways can be employed to biosynthesize secondary metabolites with bioactivity, *e.g.* ribosomally synthesized and post-translationally modified peptides (RiPPs), alkaloids, and terpenes (Cushnie *et al.*, 2014; Mahizan *et al.*, 2019; Montalbán-López *et al.*, 2021).

The genes that are involved in the biosynthesis of secondary metabolites are often grouped together and collectively termed as biosynthetic gene cluster (BGC) (Medema *et al.*, 2015). The MIBiG (Minimum Information about Biosynthetic Gene Cluster) Repository contains curated biosynthetic gene clusters of known function from bacteria (1670 entries), fungi (249 entries), plants (19 entries), and unknown source (Kautsar *et al.*, 2020). The database currently contains 1923 entries, 799 of which are PKS, 605 NRPS, 258 RiPP, 173 saccharide, 130 terpene, and 253 other clusters (<https://mibig.secondarymetabolites.org/stats>, accessed 15 June 2021). The following sections will discuss the three most common pathways for the

biosynthesis of secondary metabolites: (i) polyketides (PKs), (ii) non-ribosomal peptides (NRPs), and (iii) ribosomally synthesized and post-translationally modified peptides (RiPPs).

I.2.1. Polyketides (PKs)

Polyketide natural products are a large family of secondary metabolites that are biosynthesized from the decarboxylative condensation of acyl-CoA precursors. The polyketide chain elongation is catalysed by polyketide synthases (PKSs), which are multifunctional enzymes or multienzyme complexes (Shen, 2003). There are three types of PKSs, type I and type II – based on the similarity to architectural organization of fatty acid biosynthetic systems – and type III (**Table I.1**).

Table I.1 Distinctive characteristics of different types of PKS (Weissman, 2009)

| PKS Type | Operation mode | Substrate activation mode | Typical Product |
|----------|----------------|---------------------------|----------------------|
| I | Modular | ACP | Reduced |
| I | Iterative | ACP | Aromatic and Reduced |
| II | Iterative | ACP | Aromatic |
| III | Iterative | CoA | Aromatic |

Type I PKS

Type I PKSs are huge multidomain enzymes that serve different functions. Type I PKSs can be subdivided into two different operation mode, modular and iterative. The modular type I PKSs, which are more commonly found in bacteria, have linearly arranged domains. A group of domains that are used for the incorporation of one building block is called ‘module’, and the polyketide chain elongation was achieved following assembly line of the modules (Donadio *et al.*, 1991). The iterative type I PKSs, which are more dominant in fungi, extend the polyketide chain by a set of iteratively functioning domains (**Figure I.3**).

The domains in type I PKS are: acyltransferase (AT) that selects the acyl-CoA extender unit, the ketosynthase (KS) that catalyses a Claisen-type decarboxylative condensation for polyketide chain elongation, the optional processing enzymes like dehydratase (DH), enoylreductase (ER), ketoreductase (KR) that determine the degree of β -keto reduction, the acyl carrier protein (ACP) that activates the acyl-CoA monomer and transfers the growing polyketide chain to the downstream module, and thioesterase (TE) that releases the polyketide chain from the assembly line. (Keatinge-Clay, 2012). The nascent polyketide chain will then be processed with tailoring enzymes to create the final product as exemplified by erythromycins biosynthesis. Erythromycins biosynthesis is often used as the typical model of type I PKS, by which the 6-deoxyerythronolide B polyketide core is produced and then modified by tailoring enzymes, e.g. TDP-mycarose glycosyltransferase, TDP-desosamine glycosyltransferase, C-12 hydroxylase, (O)-methyltransferase, to yield erythromycins (Staunton and Wilkinson, 1997; Zhang *et al.*, 2010).

Recently, the term 'modules' was proposed to be redefined, so that the assembly line for modular type I PKS begins with AT, optionally followed by DH, ER, KR, then the acyl carrier protein ACP and ketosynthase KS instead of the previous order of KS+AT+DH+ER+KR+ACP (Keatinge-Clay, 2017; Zhang *et al.*, 2017).

Type II PKS

Type II PKSs are multienzyme complexes, which typically can be found in bacteria and produce aromatic compounds like actinorhodin, tetracenomycin C, and tetracycline. The heterologous expression of a large DNA fragment (>30 kb) from actinorhodin-producing *Streptomyces coelicolor* into non-producing mutants of *S. coelicolor* and *S. parvulus* led to the production of actinorhodin and the first elucidation of PKS biosynthetic genes (Malpartida and Hopwood, 1984). The genes encoded for a multienzyme complex with homology to enzymes involved in type II fatty acid synthase (FAS), e.g. KS, ACP, and KR. Due to the similar architectural arrangement towards the type II FAS, this PKS system with monofunctional enzymes is classified as type II PKS (Smith and Tsai, 2007). Another distinctive characteristics of type II PKSs is the iterative use of the discrete enzymes, as demonstrated by the tetracenomycin C biosynthesis (Motamedi and Hutchinson, 1987).

A type II PKS consists of two KS subunits with high sequence homology, KS_{α} and KS_{β} , and malonyl-CoA:ACP transferase (MCAT) for the elongation of the polyketide chain. KS_{α} catalyses Claisen-type condensation of the polyketide building blocks. KS_{β} was formerly described as chain length factors (CLF) that determines the length of the carbon chain. Furthermore, the replacement of the active site cysteine by glutamine in KS_{β} indicates the decarboxylative activity of KS_{β} , which is similar to the KS_Q domain in the modular type I PKS, and suggests KS_{β} to function as chain initiation factor (Bisang *et al.*, 1999; Hertweck *et al.*, 2007). After the decarboxylation of malonyl-ACP by KS_{β} , KS_{α} forms a C-C bond between the activated acetyl unit and the malonyl building block. The polyketide chain grows via the iteratively acting enzymes, until it reaches a certain chain length (usually either 16, 20, or 24). Then, additional enzyme subunits modify the polyketide chain according to their function (ketoreductase (KR) for β -keto reduction, cyclase (CYC) and aromatase (ARO) for cyclization/aromatization of the chain) (Staunton and Weissman, 2001; Wang *et al.*, 2020).

Type III PKS

The type III PKS is the simplest type of PKSs. Without the involvement of ACP to tether the growing polyketide chain, acyl-CoAs are directly used as building blocks. The homodimeric ketosynthase works repetitively to catalyse the decarboxylative condensation of the monomers, and produces linear polyketide intermediates. These intermediates are then cyclized by Claisen condensation as exemplified by chalcone synthase, aldol condensation as exemplified by stilbene synthase, or lactonization as exemplified by acridone synthase (Lim *et al.*, 2016).

Type III PKSs were primarily found in plants. 1,3,6,8-tetrahydroxynaphthalene (THN) was the first compound that was reported to be produced by a type III PKS from bacteria (Funa *et al.*, 1999). Thereafter, more compounds were reported to be biosynthesized via a type III PKS pathway, such as germicidin A from *Streptomyces viridochromogenes* and *Streptomyces coelicolor*, alkylresorcinol and alkylpyrone from *Azotobacter vinelandii*, and 2,4-Diacetylphloroglucinol from *Pseudomonas* sp. (Katsuyama and Ohnishi, 2012).

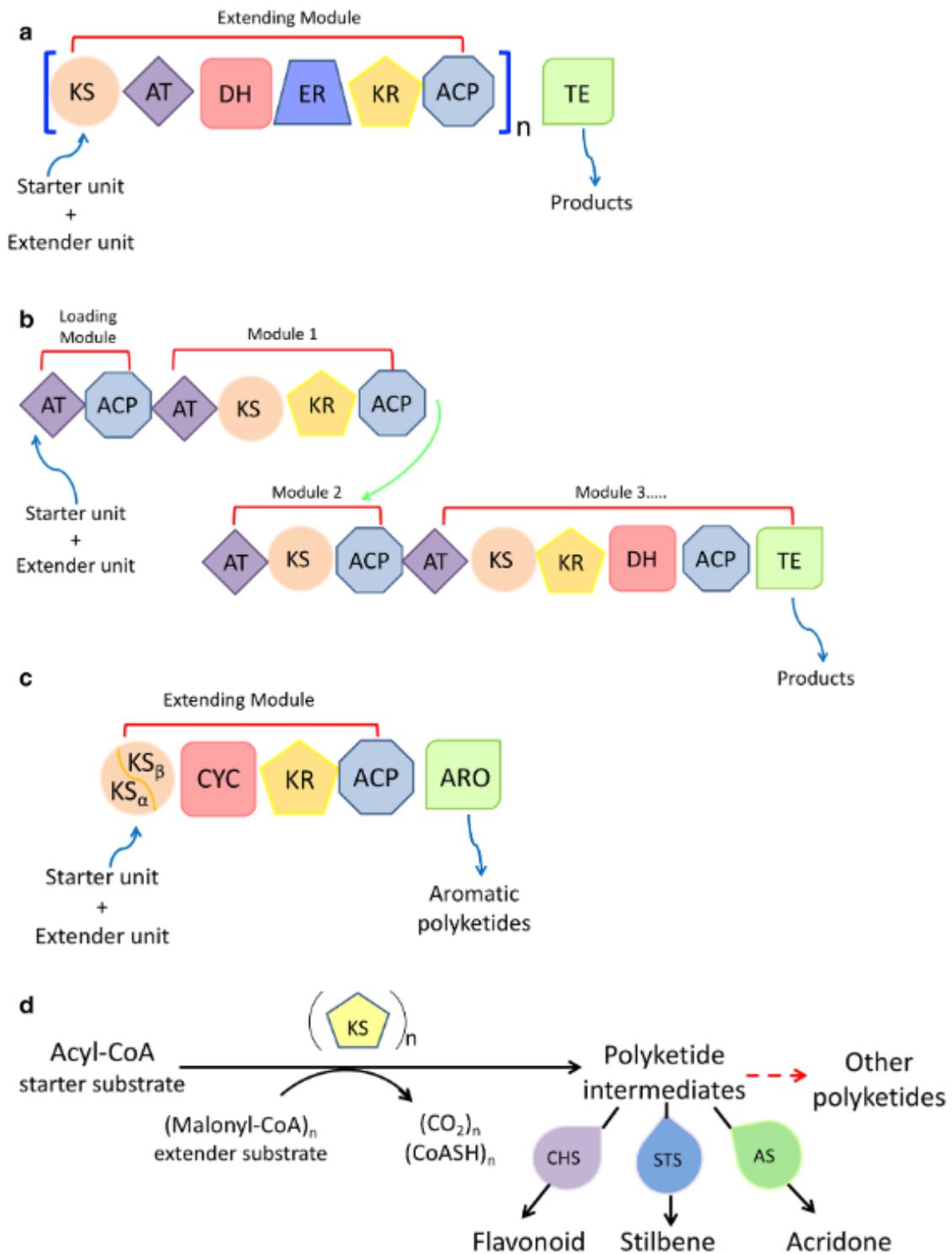


Figure I.3: Schematic overview of different PKS types. Type I PKSs are divided to two different operation modes: (a) iterative and (b) modular. Type II PKSs (c) are multi-enzyme complex that work iteratively to typically produce aromatic polyketides. Type III PKSs (d) are homodimeric enzymes with single active sites that works iteratively. KS: ketosynthase, AT: acyltransferase, DH: dehydratase, ER: enoylreductase, KR: ketoreductase, ACP: acyl carrier protein, TE: thioesterase, CYC: cyclase, ARO: aromatase, CHS: chalcone synthase, STS: stilbene synthase, AS: acridone synthase (Wang *et al.*, 2020).

I.2.2. Non-ribosomal peptides (NRPs)

Non-ribosomal peptides are a family of peptidic secondary metabolites that are synthesized by non-ribosomal peptide synthetases (NRPSs) instead of ribosomes. NRPSs are involved in the biosynthesis of clinically relevant antibiotics, such as β -lactams, glycopeptides, streptogramins, chloramphenicols, and lipopeptides. Similar to PKS biosynthetic pathway, NRPS pathways also exploit a modular enzymatic concept that follows the so-called 'collinearity rule', where each module incorporates an amino acid to a growing peptide chain (Felnagle *et al.*, 2008; Süssmuth and Mainz, 2017).

The NRPS modules consists of three core domains, additional modification domains, and a thioesterase domain (abbreviated as TE or Te) in the last module of the NRPS pathway. The core domains are an adenylation (A) domain, which selectively activates amino acids as the NRP building blocks; a thiolation (T) or peptidyl carrier protein (PCP) domain, which thiolates the aminoacyl-AMP to yield aminoacylthioester; and a condensation (C) domain, which catalyses the formation of a peptide bond between the amino acid monomers (**Figure I.4**). The core domains form the minimal NRPS modules. The Te domain usually exists at the C-terminus of the NRPS assembly line to release the nascent peptide from the NRPS (Miller and Gulick, 2016; Süssmuth and Mainz, 2017).

The additional domains that are optionally integrated in the NRPS assembly line are: epimerization (E) domain that catalyses the stereochemical transformation of L-amino acids to D-amino acids; methylation (M) domain located at the C-terminus of the A domain, which in most cases catalyses N-methylation and in rare cases catalyses O-, S-, and C-methylation; NADP(H)-dependent reductase (R) domain that substitutes a Te domain to release the NRPS product, thereby yielding a peptide with an aldehyde or an alcohol moiety at its C-terminus; N-formylation (F) domain that occurs rarely in NRPSs; cyclisation (Cy) domain, which is homologous to a C domain, however with additional heterocyclisation function of cysteine, serine, or threonine. Furthermore, the thiazoline ring that was formed by the Cy domain-catalysed heterocyclisation of cysteine can be oxidized to thiazole by an oxidation (Ox) domain (Miller and Gulick, 2016; Süssmuth and Mainz, 2017).

General Introduction

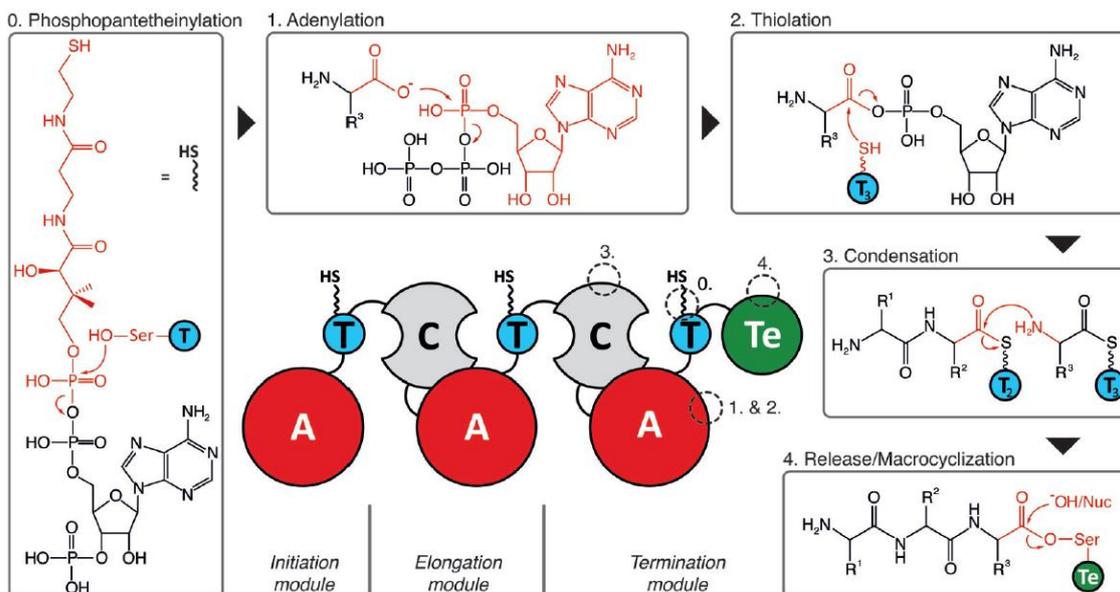


Figure I.4: Mechanism of NRPS biosynthetic pathway. NRPS modules are organized into initiation, elongation, and termination modules. C-A-T domains comprise the minimal NRPS modules. 0: Phosphopantetheinylation of the apo-T domain is catalysed by the enzyme phosphopantetheinyl transferase (PPTase). 1: Adenylation (A) domain uses ATP to activate specific amino acid as the NRP building block and load it to the downstream T domain. 2: Thiolation (T) domain tethers the activated amino acid via a thioester bond of the 4'-phosphopantetheine (Ppant) prosthetic group. 3: Condensation (C) domain catalyses the formation of a peptide bond between the activated amino acid and the growing peptide chain from upstream module. 4: Release/Macrocyclization of the nascent oligopeptide by the thioesterase (Te) domain. Nuc = nucleophile. (Süssmuth and Mainz, 2017)

NRPSs are classified into three types: Type A (linear NRPS), Type B (iterative NRPS), and Type C (Non-linear NRPS). In type A NRPSs, each module works only once to incorporate one amino acid into the growing peptide chain. The type A NRPS is commonly exploited by nature for the biosynthesis of clinically relevant antibiotics, like penicillin, daptomycin, vancomycin, and the suggestively new class of antibiotics, *i.e.* teixobactin. Type B NRPSs repetitively use allocated module(s) to produce compounds with molecular symmetry, as exemplified in the biosynthesis of echinomycin and gramicidin S. Type C NRPSs reuse one domain instead of the whole module to produce an NRP (Felngale *et al.*, 2008; Süssmuth and Mainz, 2017). This mechanism can be observed in the biosynthesis of mannopeptimycin, a glycopeptide with activity against MRSA. In its biosynthesis one A domain is used twice to supply β -hydroxyenduracididine-AMP to be tethered by the T domain in its

own module and the downstream module, the latter lacking an A domain (Magarvey *et al.*, 2006)

The aforementioned biosynthetic pathways can be combined to PKS/NRPS hybrids, in which PKS and NRPS modules produce natural products using short carboxylic acids and amino acids as precursor. Some examples of clinically important compounds that are known to be produced by this biosynthetic pathway are the anticancer agents bleomycin and epothilone; the immunosuppressant rapamycin and FK506; as well as antimicrobial compounds, *e.g.* pristinamycin and bacillaene (Rath *et al.*, 2010). The genome mining study by Wang *et al.* (2014a) reported that one-third of the 3,339 examined gene clusters are hybrid PKS/NRPS systems with diverse organization.

In addition to that, the hybrid PKS/NRPS pathway has given further insights to PKS/NRPS modularity and colinear assembly line, which inspired the combinatorial biosynthesis of known PKSs and NRPSs to produce “unnatural” natural products. Combinatorial biosynthesis uses genetic engineering approach to reprogram biosynthetic pathway by exchanging domains, modules, or tailoring enzymes (Beck *et al.*, 2020). For example, the exchange of an AT domain of the erythromycin PKS with an AT domain from the rapamycin PKS alters the monomer specificity from methylmalonyl-CoA to malonyl-CoA, yielding 6-desmethyl erythromycin D as the final product (Petkovic *et al.*, 2003). Meanwhile, Calcott *et al.* (2014) showed that the alteration of an A domain in one of the pyoverdine NRPS module from *Pseudomonas aeruginosa* PAO1 was not enough to modify the end product. The modification was successful only when both C and A domains of the same module were substituted by the domains from other *Pseudomonas* strains. In combination with this developing approach of combinatorial biosynthesis, the PKS, NRPS, and hybrid PKS/NRPS pathways bear immense potency for discovery of diverse novel core structures as the basis for new drug leads.

I.2.3. Ribosomally synthesized and post-translationally modified peptides (RiPPs)

In contrast to prevailing natural product classes like polyketides, non-ribosomal peptides, terpenoids, and alkaloids that were discovered in the 20th century, ribosomally synthesized and post-translationally modified peptides is a relatively new major class of natural products and have gained growing research interest in the 21st century. The members of this compound class have been discovered earlier, but the recognition that these compounds are generated via common biosynthetic pathways came later, when the genome sequencing methods have advanced. As an example, the antibiotic thiostrepton was first described in 1955 (Donovick *et al.*, 1955), but the biosynthesis pathway remained elusive until 2009, when it was revealed that the thiopeptide is synthesized ribosomally and then modified post-translationally (Liao *et al.*, 2009). In 2013, a community of scientists, whose works are related to this compound class, proposed the name ribosomally synthesized and post-translationally modified peptides (RiPPs) to classify this new compound class based on their biosynthetic commonality and set the upper size limit to 10 kDa for a peptide to be categorized as a RiPP to distinguish from post-translationally modified proteins. The biosynthetic pathways to produce RiPPs was proposed to be referred as Post-Ribosomal Peptide Synthesis (PRPS) (Arnison *et al.*, 2013).

In accordance with the proposed name, biosynthesis of RiPPs starts with the assembly of a linear precursor peptide. The precursor peptides are divided into different regions: leader region at the N-terminus, core region in the C-terminus, and follower region that is not necessarily always present. The leader and follower region act as the recognition and binding site of the post translational modification (PTM) enzymes, and the core region is modified by PTM into the mature RiPP (**Figure I.5**) (Hudson and Mitchell, 2018).

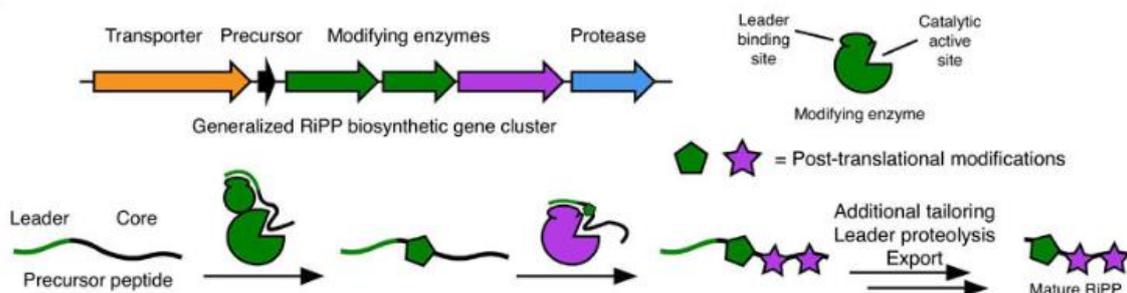


Figure I.5: Schematic representation of RiPP biosynthesis. The generalized RiPP biosynthetic gene cluster consists of genes that encode for precursor peptide, modifying enzymes, transporter, and protease. The modifying enzymes possess leader and/or follower region binding site(s) and a catalytic active site that modifies the core region of the precursor peptide. The additional tailoring enzymes complete the biosynthetic pathway and proteolysis releases the modified core region from the leader peptide, generating a mature RiPP (Hudson and Mitchell, 2018).

RiPPs form an expanding natural products class, which currently comprises 41 subclasses defined by their PTMs (Montalbán-López *et al.*, 2021). The diverse RiPP classes contribute to the structural diversity and bioactivity of the compounds. RiPPs have been reported to be potentially effective in cystic fibrosis therapy (lancovutide), treating neuropathic pain (ω -conotoxin MVIIA), and killing pathogenic bacteria (darobactin) (Imai *et al.*, 2019; Teichert and Olivera, 2010; Zeitlin *et al.*, 2004). For their antimicrobial and cytotoxic activities, RiPPs exerted various mode of actions, which inhibit major cellular processes such as DNA replication, transcription, translation, and cell envelope biogenesis (**Figure I.6**) (Cao *et al.*, 2021).

Despite their high potential, clinical implementation of RiPPs is difficult due to their poor solubility, stability, and bioavailability (Hudson and Mitchell, 2018; Rowe and Spring, 2021). Therefore, chemical synthesis was employed to modify the compound to increase thermal and proteolytic stability, solubility, production yield and purity (Rowe and Spring, 2021). In addition to that, molecular engineering of the precursor peptide encoding gene has become a prospective strategy to generate new compounds or analogues with increased medicinal value (Hudson and Mitchell, 2018). For example, the fusion of lantipeptide core regions from cryptic gene clusters to the nisin leader region produced 5 new lantibiotics that were highly active against *Micrococcus flavus* and vancomycin-resistant *Enterococcus faecalis*

(van Heel *et al.*, 2016). Moreover, generation of a RiPP library with amino acid replacements at three positions in the lasso peptide microcin J25, followed by zone inhibition assay, reported to yield analogues with improved efficacy against *Escherichia coli* and *Salmonella* strains (Pan and Link, 2011). These studies show the advantages of genome mining and heterologous expression of BGCs to get access to new natural products or the modification of known ones to improve their pharmacological features and bioactivity for human and medical usage.

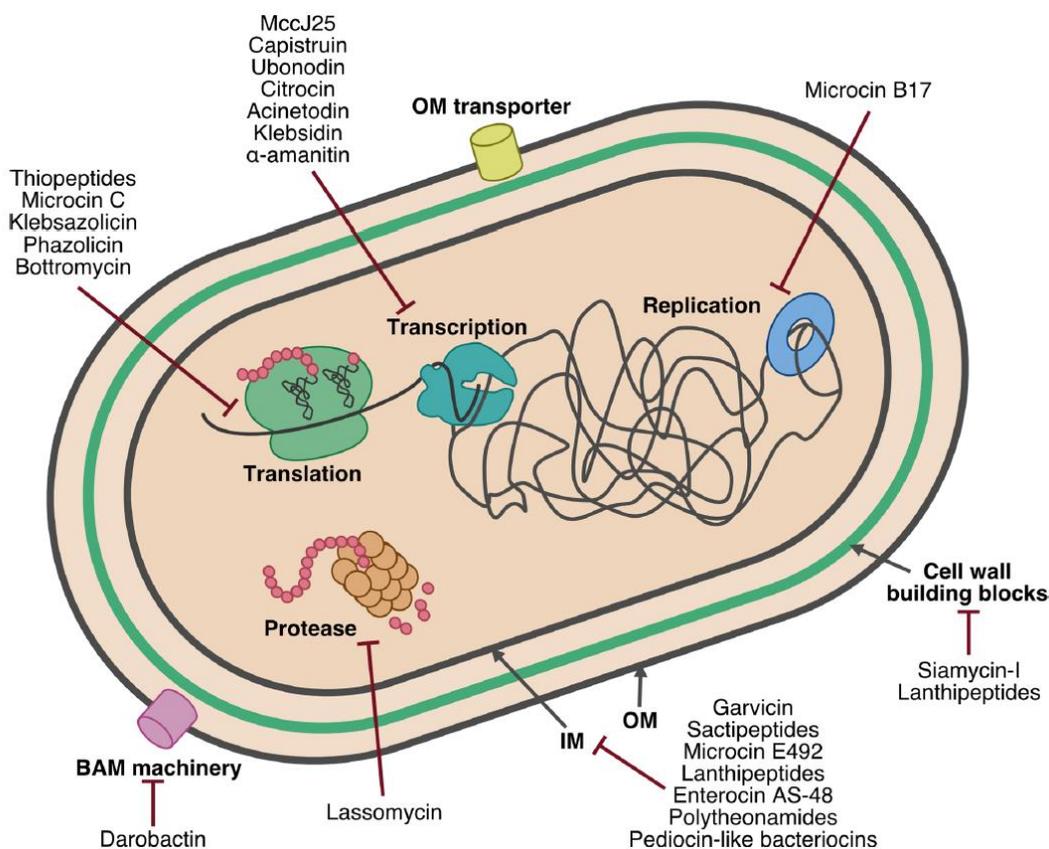


Figure I.6: Mode of actions underlying RiPPs antibiotic activities. RiPPs target various cellular processes. They inhibit the biogenesis of cell envelope, DNA replication, transcription, translation, and proteases in the targeted cell. BAM: β-barrel assembly machinery, IM: inner membrane, OM: outer membrane (Cao *et al.*, 2021).

I.3. Heterologous expression strategy for secondary metabolite production

Advances in sequencing technology have significantly reduced the whole genome sequencing price, thus more genome data are available and analysis of these revealed genes that are putatively involved in the biosynthesis of the previously discussed secondary metabolites. These biosynthetic genes are often clustered and collectively referred to as “biosynthetic gene cluster” (BGC).

The *in silico* analysis of the genome sequence of *Streptomyces coelicolor* A3(2) revealed that this model actinomycete carries cryptic gene clusters, whose products could not be identified by culturing the bacterium in a laboratory environment (Bentley *et al.*, 2002). Similarly, the cryptic ability to produce secondary metabolites was also reported from the analysis of the whole genome sequences of other bacteria, such as *Pseudomonas fluorescens*, *Streptomyces avermitillis*, *Salinispora tropica*, and *Saccharospora erythraea*. It was observed that much more BGCs are available than known compounds that had been isolated (Oliylyk *et al.*, 2007; Ōmura *et al.*, 2001; Paulsen *et al.*, 2005; Udvary *et al.*, 2007). These observations have led to the redirection of the classic antibiotic discovery platform, from the conventional Waksman’s method to a synthetic biology-based strategy, which includes genome mining, followed by genetic engineering and heterologous expression to identify novel compounds. (Lee *et al.*, 2019). The main idea of this redirected approach is to acquire compounds from cryptic BGC that was detected by genome mining (**Figure I.7**).

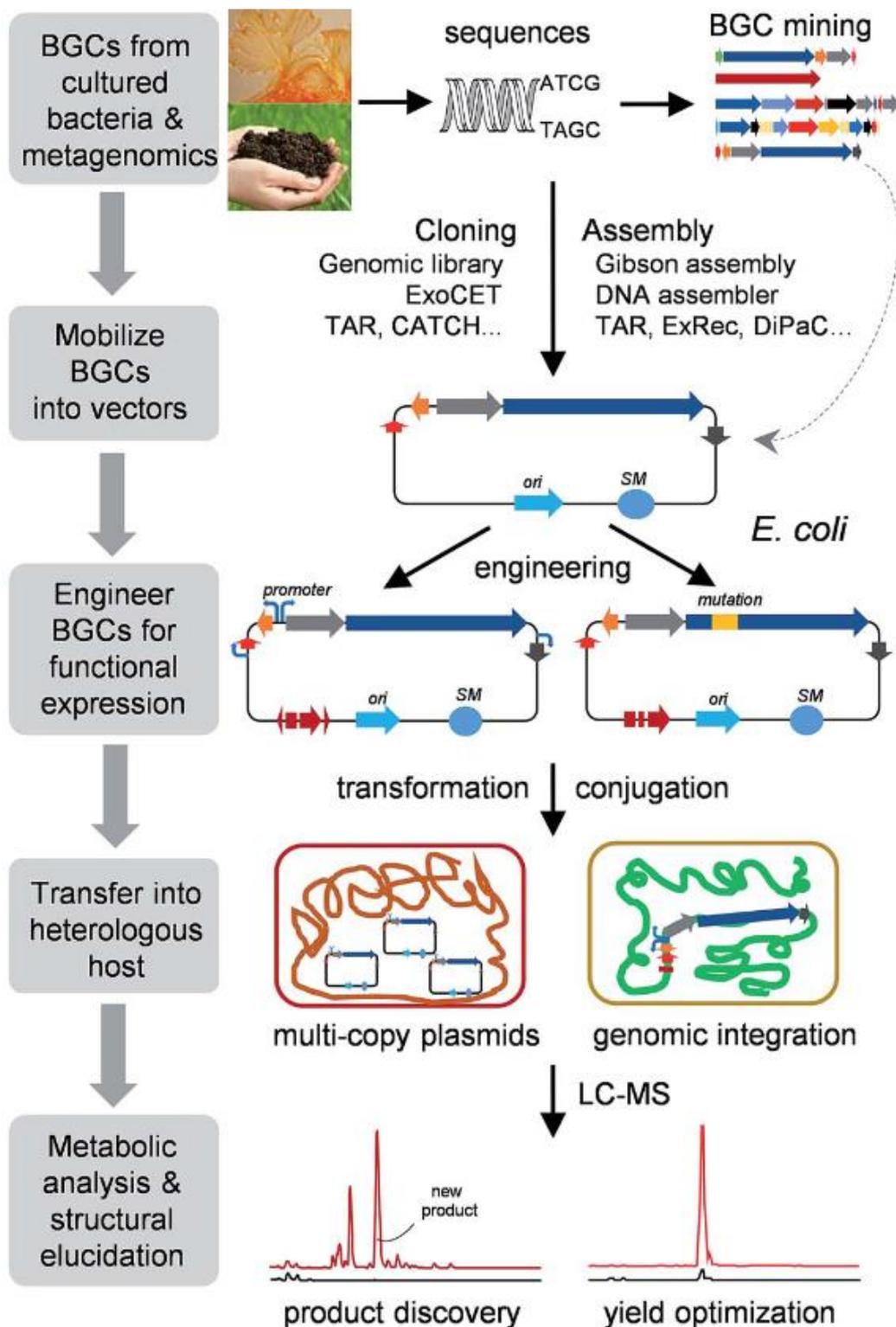


Figure I.7: General outline of heterologous expression approaches for genome-mined BGCs. BGCs that were identified by genome mining are cloned into expression vectors and engineered for functional expression. The construct is then transferred into suitable heterologous hosts, as multi-copy plasmid or integrated to the genomic DNA. The production of compounds encoded by the BGCs is confirmed by LC-MS analysis, followed by structural elucidation for novel compounds (Huo *et al.*, 2019).

I.3.1. Genome mining

Genome mining is defined as “the use of bioinformatics, molecular genetics, and natural product analytical chemistry to access the metabolic product of a gene cluster found in the genome of an organism” (Gomez-Escribano and Bibb, 2014). Some genome mining strategies have been proven to be able to link genomic data to secondary metabolites by, e.g. mining for core or specific biosynthetic genes; *in silico* prediction of compound structure and chemical synthesis; metabolomics analysis for BGC prioritization; mining based on the presence of resistance gene within the BGC; mining for RiPPs cluster; and mining based on the evolutionary relation to known natural product producer (phylogeny-based) (Figure I.8) (Scherlach and Hertweck, 2021).

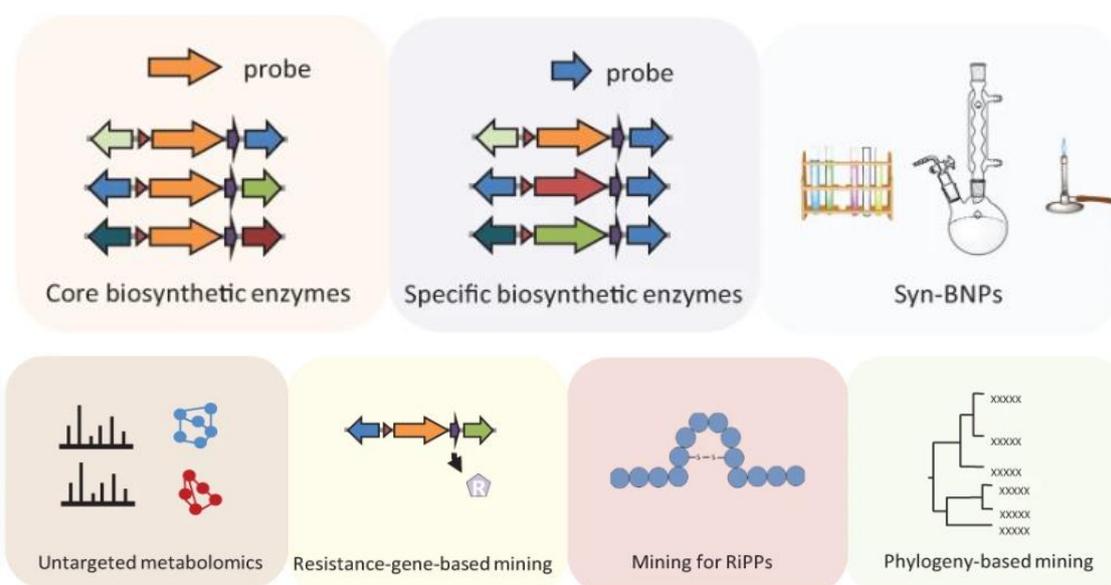


Figure I.8: Genome mining strategies for the discovery of natural products. Abbreviations: syn-BNP: synthetic-bioinformatic natural products; RiPPs: ribosomally synthesized and post translationally modified peptides (Scherlach and Hertweck, 2021).

Automated bioinformatic tools are now available for the above-mentioned strategies. Most tools identify the BGCs based on homologies to signature features of formerly elucidated BGCs by using hidden Markov models (HMMs) and/or Basic Local Alignment Search Tool (BLAST) analysis (Ren *et al.*, 2020). Genome mining tools that are commonly used for identifying PKS, NRPS, and RiPPs clusters as the

main biosynthetic pathways used for antibiotic production are: antiSMASH (Blin *et al.*, 2021), PRISM (Skinnider *et al.*, 2020), and BAGEL (van Heel *et al.*, 2013). In addition to the identification of core and additional biosynthetic genes in the predicted BGC, antiSMASH provides NRPS/PKS domain analysis, identifies similarity to BGCs with known products, predicts the NRPS/PKS monomer based on the Stachelhaus code (for NRPSs), ATsignature and minowa algorithm (for PKSs), and provides core structure prediction based on the predicted monomers (Medema *et al.*, 2011; Weber *et al.*, 2015).

I.3.2. Cloning and heterologous expression of BGC

Heterologous expression has been used to find the product of a given cryptic BGC, as well as to prove the link between a BGC and a known compound (Gomez-Escribano *et al.*, 2019; Linares-Otoya *et al.*, 2017). The clustering of the biosynthetic genes in the microbial genomes facilitates the cloning of the complete biosynthetic pathway into heterologous host for subsequent expression and identification of the products. However, the BGCs of modular systems, *i.e.* PKS and NRPS-based ones, usually have a large size with repetitive sequences and in many talented producer genera high GC-content, which complicates the cloning of the BGC. Three main types of cloning methods that have been developed for cloning of NP BGCs are (i) library-based, (ii) assembly, and (iii) direct cloning methods (Zhang *et al.*, 2019a). These will be shortly described in the next paragraphs.

The library-based methods represent a conventional approach for BGC cloning that is advantageous when the whole genome sequence is not available, as well as when processing metagenomic samples. Therefore, the (meta)genomic DNA is sheared randomly to the size of 40 kb or larger to be cloned into cosmids or fosmids, and to the size of more than 100 kb to be cloned into bacterial artificial chromosomes (BACs) or P1 artificial chromosomes (PACs) (Nah *et al.*, 2017; Zhang *et al.*, 2019a).

The assembly methods for BGC cloning involve DNA fragments with overlapping regions. *In vitro* assembly usually uses shorter DNA fragments than the *in vivo* assembly does, which are generated by polymerase chain reaction (PCR). Gibson assembly became one of the most commonly used *in vitro* assembly methods. It is

an isothermal assembly system that uses T5 exonuclease to remove nucleotides from the 5' DNA termini and to create single-strand DNA overhangs, Phusion DNA polymerase to fill the gaps in the annealed overhangs, and *Taq* DNA ligase to covalently seal the nicks (**Figure I.9**) (Gibson *et al.*, 2009).

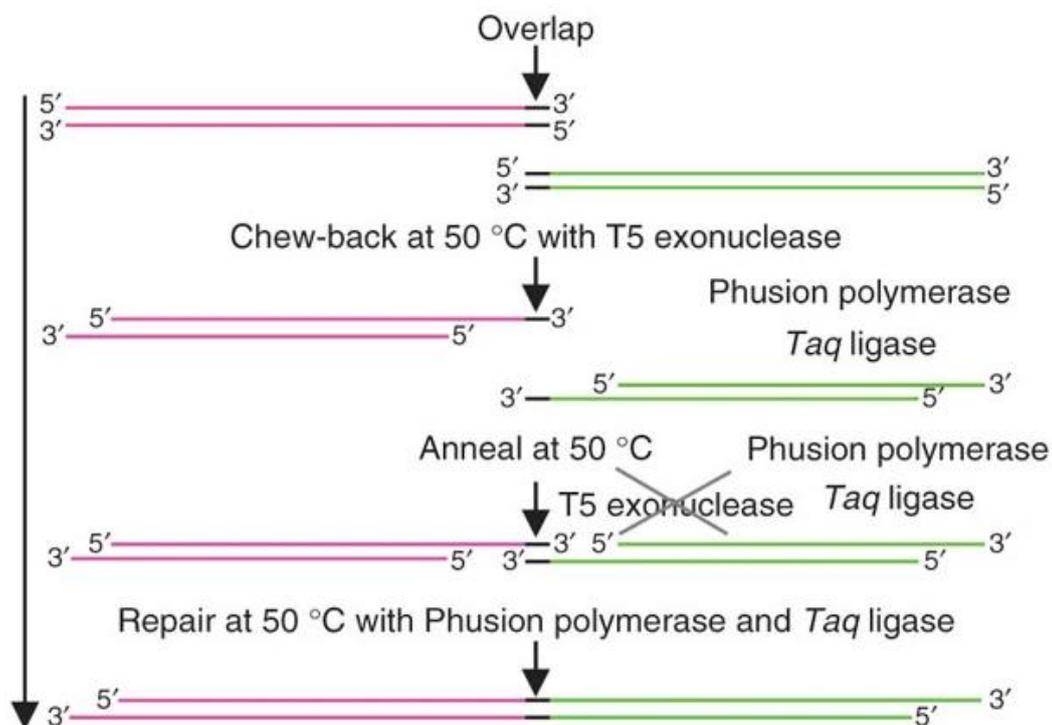


Figure I.9: Reaction scheme for one-step isothermal *in vitro* assembly. The isothermal assembly requires overlapping regions between the fragments to be assembled. The reaction occurs at 50°C, where T5 exonuclease chews back 5' ends of the double-stranded DNA, followed by the annealing of single-stranded DNA overhangs and DNA repair by Phusion polymerase and *Taq* ligase (Gibson *et al.*, 2009).

Transformation-associated recombination (TAR) cloning is an example of a direct cloning method. TAR cloning utilises the homologous recombination ability of *Saccharomyces cerevisiae* to capture a target BGC into the capture vector for later integration of the target BGC into the heterologous expression host by using the ϕ C31 integrase system (**Figure I.10**). The capture vector can be generated by the Gibson assembly method to insert 50 bp homologous regions to the capture vector (e.g. pCAP03, for more details see section III.2.5) (Zhang *et al.*, 2019b). TAR cloning can recover DNA fragments up to 250 kb and was reported to have a yield of 1–5% positive clones carrying the desired target DNA sequence when used to directly capture target DNA from mammalian a chromosome (Kouprina and

Larionov, 2006). NRPS and PKS BGCs have been successfully cloned by using the TAR cloning approach, as exemplified by the heterologous expression of the 30-kb marinopyrrole BGC, the 67-kb taromycin A BGC, and the 33-kb scleric acid BGC (Alberti *et al.*, 2018; Yamanaka *et al.*, 2014).

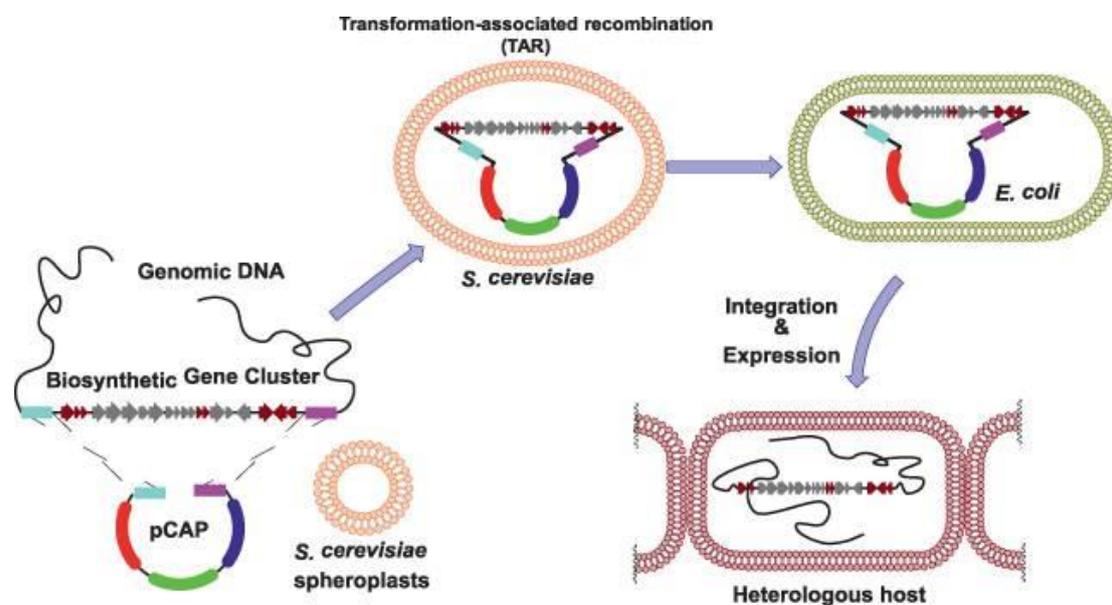


Figure I.10: Transformation-associated recombination (TAR) for direct cloning of BGCs. A capture vector with 50 bp homology arms is used to capture a target BGC from genomic DNA via homologous recombination in the yeast *S. cerevisiae*. The generated heterologous expression construct can be propagated in *E. coli* cells, which are applied for the conjugation of the construct into the heterologous host, *e.g.* *Streptomyces* strains (Zhang *et al.*, 2019b).

Prior to heterologous expression, genetic engineering can be done by deleting or activating transcriptional regulators, addition of promoters, reporter genes, or ribosomal binding sites, as well as rare codons optimization (Tan and Liu, 2017). One commonly used method for genetic engineering of target BGCs is the λ Red-mediated recombination (**Figure I.11**). This method uses the product of phage-derived genes γ , β , and *exo*, which are referred to as Gam, Beta, and Exo, respectively. Gam inhibits exonucleases from degrading the linear DNA cassette introduced into the *E. coli*, Exo acts as the exonuclease that chews back the dsDNA from the 5' ends, and Beta protects the single strand overhang and promote the recombination (Datsenko and Wanner, 2000).

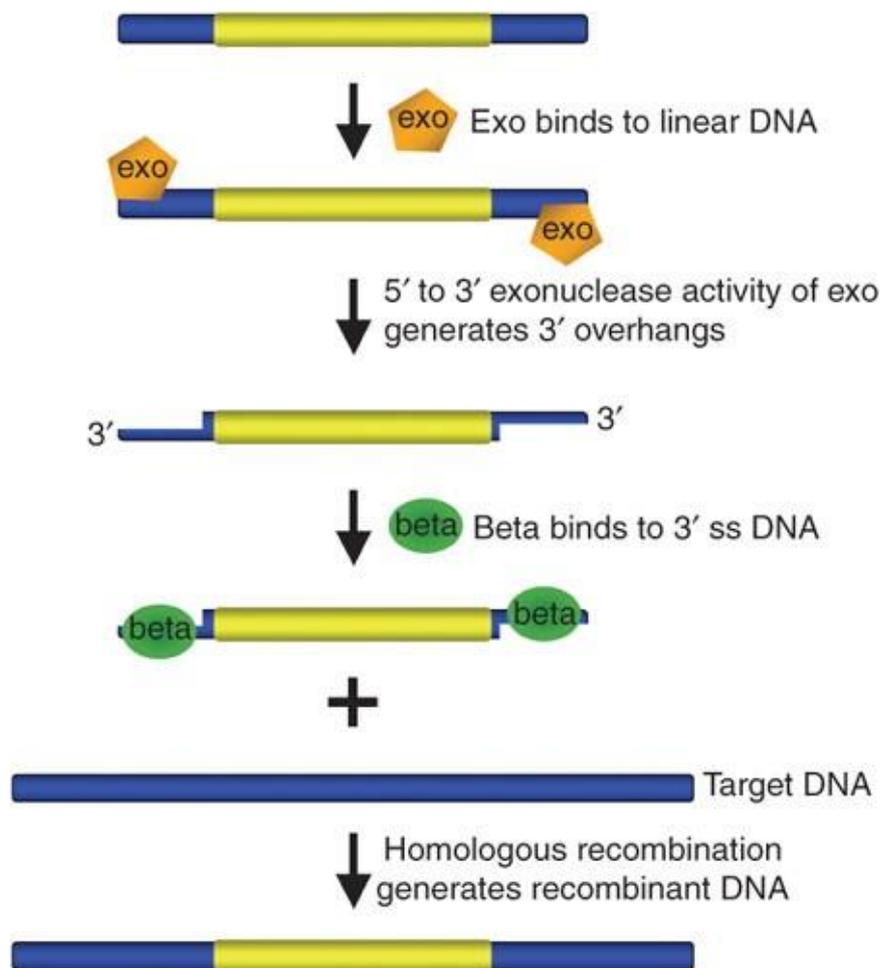


Figure I.11: λ Red-mediated recombination. The bacteriophage λ recombination system consists of three main components, *i.e.* Exo, Beta, and Gam. Gam (not shown on the figure) inhibits RecBCD nucleases from degrading the linear DNA, Exo creates 3' overhang through its exonuclease activity, and Beta binds to the 3' overhang and promotes its annealing to complementary target DNA (Sharan *et al.*, 2009).

Once the desired constructs carrying the BGC of interest are generated, the heterologous expression host(s) needs to be carefully selected. *Streptomyces* strains have been proven to be the most widely used heterologous expression host for polyketides and NRPs from actinomycetes (Nah *et al.*, 2017). *Streptomyces* strains are advantageous, due to the availability of genetic tools for convenient cloning and genetic manipulation of the target BGC and the availability of many biosynthetic precursors, since they possess an intrinsic ability to produce PKs and NRPs (Zhang *et al.*, 2019a).

II. Scope of the study

Genome mining to identify putative biosynthetic gene clusters (BGCs) that encode for the biosynthesis of metabolites and heterologous expression of these BGCs are promising approaches for the discovery of novel natural products. In addition to discovery of compounds, heterologous expression can confirm the correlation between a specific BGC and the compound of interest, and enables detailed investigation of the biosynthetic pathway. Thereby, production yield can be increased and derivatives of the natural product can be generated. In this study, it is aimed to apply heterologous expression of polyketide synthase (PKS), non-ribosomal peptide synthetase (NRPS), and ribosomally synthesized and post-translationally modified peptide (RiPP) BGCs from different sources to study biosynthetic pathways and to discover novel natural products. To reach this goal, genome mining for the *in silico* detection of putative BGSs, state-of-the-art cloning techniques to transfer the BGC to a heterologous host, and optimization of the subsequent heterologous expression itself will be applied.

In order to discover novel potentially active natural products, it is projected to explore so far under-investigated bacterial habitats in this study. Therefore, the bacterial strains and sequence information that will form the basis for the study will be obtained from two projects that were started before this work. First, the INDOBIO project in which the biodiversity of sea slugs and their associated bacteria were investigated, and second the Burying Beetle project in which members of the *Nicrophorus vespilloides* gut microbiome were isolated and tested for their potential to inhibit Gram-positive and Gram-negative test organisms. Bacterial strains that showed antibacterial activity in screening assays will be selected and their genomic DNA will be searched for the presence of NRPS and PKS BGCs. Once candidate BGCs are identified *in silico*, they will be subsequently cloned and heterologously expressed in suitable hosts to prospectively discover new bacterial secondary/specialized metabolites.

In a second part of the project, it is planned to first prove the functionality of the predicted darobactin BGC and second, to design and generate a heterologous expression platform for this compound (class). Darobactin A represents a new

antibiotic that selectively targets Gram-negative bacteria with compelling *in vivo* activities in mouse infection models. A corresponding BGC, predicted to encode for the production of this RiPP compound, was identified through analysis of whole genome DNA sequence data. Once the involvement of the BGC in darobactin A production is verified, it is envisaged to increase the production yield. This will ensure compound supply for further studies.

III. Heterologous expression of NRPS and PKS BGCs

III.1. Introduction

III.1.1. INDOBIO project

The project “Indonesian Opisthobranchs and Associated Microorganisms – From Biodiversity to Drug Lead Discovery (INDOBIO)” was a collaborative initiative between Germany and Indonesia that was funded by the Federal Ministry of Education and Research (BMBF). Dedicated funding was available for fellowships via the German Academic Exchange Service (DAAD). The main goal of the project was to explore the biodiversity of Indonesian sea slugs as the source of novel drug lead with potential bioactivity.

The sea slug biodiversity from the yearly sampling trips to North Sulawesi in 2015 to 2018 was well-documented and the published data revealed that the most commonly collected specimens belonged to the Nudibranchia taxa (Eisenbarth *et al.*, 2018; Kaligis *et al.*, 2018; Papu *et al.*, 2020; Undap *et al.*, 2019). Nudibranchia is a member of the Opisthobranch. In the “new euthyneuran tree”, this taxa was redefined as a member of the Heterobranchia under Nudipleura (Wägele *et al.*, 2014; Wägele and Willan, 2000). Morphologically, distinctive characteristics of the Nudibranchia are the loss of the protective shell and bright colourful patterns (Gosliner *et al.*, 2015). In addition to these characteristics that are not favourable toward physical defence, the nudibranchs are living in exposed habitats and are easily spotted by their predators. Therefore, they utilize chemical defence system to protect not only the adult sea slugs, but also their egg mass. Their bright colours are often associated with unpleasant taste, and they store toxic metabolites in their mantle or glands, which can also be released via mucous secretion. Their chemical armoury is derived from their food source, *de novo* synthesis, or synthesized by their symbiotic bacteria (Avila *et al.*, 2018).

The potential of the sea slugs that were discovered in North Sulawesi, Indonesia, to produce bioactive compounds had been elaborated in a review paper (Fisch *et al.*, 2017). Out of 15 families that were represented by the collected samples from Bunaken National Park (BNP), around 150 active compounds have been reported. On the other hand, only 14 species had been studied for bioactive compounds out of 76 divergent sea slug species that were collected around BNP (**Figure III.1**). This

finding highlighted sea slugs as understudied source for the discovery of bioactive compounds.



Figure III.1: *Hexabranchnus sanguineus* and its egg mass. Left: *Hexabranchnus sanguineus* specimen that was collected in September 2017 from Bangka Archipelago, North Sulawesi. The identity of this juvenile *H. sanguineus*, which mimics the coloration of *Hypselodoris maculosa*, was corroborated by molecular analysis (Papu *et al.*, 2020). **Right:** *Hexabranchnus sanguineus* egg mass that was collected in August 2015 from Bunaken National Park, North Sulawesi. Photos courtesy of the Zoologische Forschungsmuseum Alexander Koenig (ZFMK).

Most of the bioactive compounds that were reported from sea slugs were produced by PKS, NRPS, terpenoid, or hybrid of the aforementioned biosynthesis pathway (Fisch *et al.*, 2017). One example of the bioactive compounds are the trisoxazole macrolides, *i.e.* ulapualide A and B, halichondramides, and kabiramide A-C, isolated from *Hexabranchnus sanguineus*, its egg mass, and *Halichondria* sponge as its food source (Kernan *et al.*, 1988; Matsunaga *et al.*, 1989; Pawlik *et al.*, 1988; Roesener and Scheuer, 1986). These compounds were reported to have antifungal activity and cytotoxic effect. The study on biosynthesis of similar actin-binding macrolides such as misakinolides and luminaolides indicated that they are produced by yet unidentified microbes through PKS biosynthetic pathway (Ueoka *et al.*, 2015). Another compound that was reported to be produced through PKS biosynthetic pathway is bryostatin 1, isolated from *Polycera atra* MacFarland, 1905 and its egg mass. The 80 kb PKS cluster was identified from the unculturable bacterial symbiont of the bryozoan *Bugula neritina*, “*Candidatus* Endobugula sertula” (Sudek *et al.*, 2007). This compound has been extensively studied in clinical trials for cancer, Alzheimer’s disease, and HIV infection; and has been granted orphan drug status

by the FDA (Food and Drug Administration) (Raghuvanshi and Bharate, 2020). Both examples showed that the bioactive compounds were sequestered by the sea slugs from their food sources, and transferred to their egg mass for chemical protection against their predator. Moreover, both examples displayed that the bioactive compounds were produced by yet to be identified or cultivated microbes. Therefore, a genomic approach would be a suitable strategy to mine the potency of bioactive natural products from the sea slugs-associated bacteria, by identification of putative PKS and NRPS gene clusters using degenerative primers and heterologous expression of the identified putative clusters.

A study was conducted to investigate the sea slug-associated bacteria for the presence of PKS and NRPS genes, as well as for their antimicrobial activity (Böhringer *et al.*, 2017). As a result, 35 strains showed activity against Gram-positive and Gram-negative pathogens. In addition to that, 16 NRPS A-domain and 3 PKS KS-domain sequences could be detected by PCR with degenerated primers. From these detected genes, seven distinct A domains and all three different KS domains were obtained from the *Hexabranchnus sanguineus* egg mass symbionts, *Gordonia* sp. Bu15_44 and *G. terrae* Bu15_45 (**Figure III.2**). These strains also showed activity against the Gram-positive bacterium *Arthrobacter psychrolactophilus* in the antimicrobial screening. Thus, these *Gordonia* strains were presumed to hold a potential for the discovery of new bioactive compounds.



Figure III.2: Axenic culture of *Gordonia* strains isolated from *Hexabranchnus sanguineus* egg mass. Left: *Gordonia* sp. Bu15_44, right: *G. terrae* Bu15_45. Both strains were grown on ISP2 agar supplemented with 2% NaCl.

Gordona is a new and understudied genus that was first proposed in 1971 (Tsukamura, 1971), but then redefined by 16S rRNA analysis and renamed as *Gordonia* in 1997 (Stackebrandt *et al.*, 1997). The first whole-genome sequence of this genus was published in 2009 from human pathogen *G. bronchialis* 3410^T (GenBank accession: CP001802), isolated from the phlegm of a patient with pulmonary disease (Ivanova *et al.*, 2010). Sponge-associated *Gordonia* strains were reported to have antimicrobial activity against *Bacillus subtilis*, *Bacillus cereus*, *Staphylococcus aureus* MRSA, *Enterococcus faecalis*, *Escherichia coli*, *Pseudomonas aeruginosa*, *Salmonella typhi*, and *Candida albicans* (Elfalah *et al.*, 2013; Graça *et al.*, 2013). However, there are no reports regarding the compounds related to these antimicrobial activities. The only compound that was reported to have antimicrobial activity is gordonic acid, which was isolated from a co-cultivation of *Gordonia* sp. and *Streptomyces tendae* (Park *et al.*, 2017). The PCR screening for the presence of PKS and NRPS domains in the genomic DNA of *Gordonia* strains, as well as the *in silico* analysis of the whole genome sequence of *G. alkanivorans* suggested for the huge potential of natural product discovery from this genus (Gontang *et al.*, 2010; Graça *et al.*, 2013; Nahurira *et al.*, 2019).

The whole genomes of both *Gordonia* strains were sequenced in order to explore their potential from the genomic level. Indeed, *in silico* analysis of the *Gordonia* strains whole-genome sequence revealed the presence of complete PKS and NRPS clusters. Four out of five BGCs from *G. terrae* Bu15_45 could not be 100% identified to known BGC. Thus, the potential of finding novel compound from this strain was high. Here, it was projected to investigate the potential of *G. terrae* Bu15_45 further by heterologous expression attempts of the detected putative BGCs with the aim of discovering new bioactive compounds.

III.1.2. Burying beetle project

The burying beetle *Nicrophorus vespilloides* that belongs to the family Silphidae was first described by Johann Friedrich Wilhelm Herbst in 1783. This species is used as the model organism in the study of burying beetles. *N. vespilloides* was differentiated from *N. hebes* by investigations in molecular level, ecological

preference, reproductive incompatibility, and morphological appearance, and *N. vespilloides* was determined to inhabit Palearctic ecozone (Sikes *et al.*, 2016).

Burying beetle has a unique ecological role in recycling small vertebrates back into the ecosystem by feeding and breeding on their carcasses. In order to fulfil this role, it uses its antennae to sense infochemicals from the carcasses to locate the cadaver from several kilometres distance (Kalinová *et al.*, 2009; von Hoermann *et al.*, 2013). Once a cadaver with proper size and decomposition degree has been found, the burying beetles prepare the carrion as a food source for their future offspring – a behavioural act known as biparental care (Eggert and Müller, 1997).

The animal carrions are heavily colonized by diverse microorganisms that are competing for the nutrients in the carcasses. The burying beetle uses a broad spectrum of antimicrobial compounds in synergy with lysozymes to detoxify microbial toxin in the carrion as well as to digest, preserve, and protect the carrion from other beetles. The expression of antimicrobial peptides (AMPs) and lysozymes by the male and female beetles are tightly regulated, depending on the presence of the carrions and/or larvae (Jacobs *et al.*, 2016).

In addition to that, Vogel *et al.* (2017) suggested the role of the gut microbiome, which consists of Firmicutes, Proteobacteria, and ascomycetous yeast, to be working synergistically with the enzymes and compounds that are produced by the beetle to sanitize, digest, and preserve the carcasses. Transcriptomic and metabolic analysis of *Yarrowia*, the most dominant yeast among the fungal community in *N. vespilloides*, displayed the presence of digestive enzymes, such as extracellular lipases, proteases, and creatinases, as well as four low-molecular-weight compounds with bioactivity (Degenkolb *et al.*, 2011; Vogel *et al.*, 2017). Further studies are needed to explain the mechanism that is used by the beetles to preserve and utilise cadaver as their food source on the molecular level.



Figure III.3: Burying beetle *Nicrophorus vespilloides*. Photo was taken at Vogelsberg, Germany, courtesy of Dr. Philipp Heise.

Previously, a study on the gut microbiome of the burying beetle *Nicrophorus vespilloides* (**Figure III.3**) to find new bioactive compounds was conducted by Philipp Heise (Heise, 2019). The isolation project of the bacterial symbionts in the gut of the burying beetle *Nicrophorus vespilloides* resulted in 113 bacterial isolates with a broad spectrum of bioactivity against *Mycobacterium smegmatis*, *Escherichia coli*, *Pseudomonas aeruginosa*, *Staphylococcus aureus* and *Candida albicans* (Heise, 2019). Several strains with activity against Gram-negative bacteria or with strong activity against Gram-positive bacteria were further investigated. However, the activity against Gram-negative bacteria could not be retained after HPLC separation and the purification of one crude extract that had activity against *S. aureus* and *M. smegmatis* led to re-discovery of a known compound, serrawettin W2. Therefore, genome mining was pursued as the next attempt to find novel bioactive compounds. Out of three prioritized strains that were genome-sequenced, only strain 39 was identified to carry a hybrid type I PKS/NRPS.



Figure III.4: Axenic culture of strain 39 isolated from the gut of *Nicrophorus vespilloides*. The strain was grown on a lysogeny agar plate.

Strain 39 (**Figure III.4**) has 98% identity to *Arthrobacter* sp. based on its 16SrRNA, but its whole genome sequence has the highest homology to *Glutamicibacter arilaitensis* (88.47% identity with 12.8% coverage) (Heise, 2019). This result was not considered redundant because *Glutamicibacter* is a reclassification of *Arthrobacter* (Busse, 2016). Currently, only twelve species were classified under *Glutamicibacter*, which makes it difficult to confirm the taxonomic classification of strain 39 (Busse and Schumann, 2019; Das *et al.*, 2020).

Despite the difficult classification, this understudied bacterium might hold unveiled potential for the discovery of new bioactive compounds. In addition to that, the bioactivity assay of strain 39 displayed activity against *M. smegmatis* and *E. coli*. Since there was only one NRPS/PKS hybrid cluster identified from the *in silico* analysis, this putative cluster was hypothesized to be a candidate to encode for the production of a secondary metabolite responsible for the observed bioactivity. This study aims to answer the hypothesis that the NRPS/PKS hybrid cluster is corresponding to the production of the bioactive compound, by heterologous expression of the putative cluster in *Streptomyces* heterologous host(s).

III.2. Material and methods

III.2.1. *In silico* analysis of putative BGC

Genomic DNA sequence of *Gordonia terrae* Bu15_45 and a bacterial symbiont from *Nicrophorus vespilloides* (strain 39) were analyzed by antiSMASH to identify putative NRPS and PKS clusters (Blin *et al.*, 2019). The predicted clusters were manually checked for the presence of minimal modules, *i.e.* a condensation (C) domain, an adenylation (A) domain, and a peptidyl carrier protein (PCP) domain for NRPS cluster; and an acyltransferase (AT) domain, a ketosynthase (KS) domain, and an acyl carrier protein (ACP) domain for PKS cluster. Based on the presence of the putative modules, the BGCs were classified as complete and not complete. The complete BGCs similarity to known cluster were observed to find out whether the putative BGCs consist of homologous genes to known BGC. In addition to that, region-to-region similarity to MIBiG cluster was also investigated to check the sequence similarity of each region in putative BGCs to the region of BGC in MIBiG. Then, the additional biosynthetic genes as well as genes other than the core biosynthetic genes were studied to predict their function in the putative BGC. Lastly, the NRPS/PKS monomer prediction was investigated. If the integrated monomer prediction in antiSMASH did not provide specific monomer, the Stachelhaus code for the A domains of NRPS clusters were manually analysed to find the closest match (Challis *et al.*, 2000; Stachelhaus *et al.*, 1999). The final product core structure was predicted according to the predicted monomer and some additional biosynthetic genes whose function are well-studied.

III.2.2. Construction of *Gordonia* library

Construction of artificial metagenomic library from genomic DNA of *Gordonia* sp. Bu15_44 and *Gordonia terrae* Bu15_45 (**Table S2**) was done according to Gurgui and Piel (2010) with some modifications.

Genomic DNA isolation from Gordonia sp. Bu15_44 and Gordonia terrae Bu15_45.

Gordonia sp. Bu15_44 and *Gordonia terrae* Bu15_45 were inoculated into 50 mL ISP2 medium (10 g/L malt extract, 4 g/L yeast extract, 4 g/L glucose) supplemented with 2% of NaCl at 30°C with 180 rpm shaking. After three days of cultivation, the

cells were harvested by centrifugation (10,000 x *g* for 15 minutes at 4°C), washed with 10 mL of TE (25 mM Tris-HCl pH8, 25 mM EDTA) in seawater, and the cell pellet was frozen at -20°C overnight. Then, the pellet was resuspended in 10 mL lysis buffer (3 mg/mL lysozyme and 100µg/mL RNase in TE25S buffer [25 mM Tris-HCl pH8, 25 mM EDTA, 0.3 M sucrose]) and incubated at 37°C for 1 hour. Thereafter, 100 µL proteinase K (20 mg/mL) was added and mixed by inversion. Subsequently, 1 mL of 10% SDS was added, mixed carefully by inversion, and incubated at 2 hour at 55°C. Prior to addition of 2.5 mL 5M potassium acetate, mixture was placed on ice for 20 minutes. 8 mL phenol:chloroform:isoamylalcohol (25:24:1, pH8) was added, and mixed carefully by inversion for 6 minutes. The mixture was then centrifuged (5,000 x *g* for 15 minutes at 4°C) to separate the aqueous layer from the organic layer. The aqueous upper layer was carefully transferred to a new tube using cut-off tips. 8 mL of chloroform was added, mixed carefully for 6 minutes and centrifuged (5,000 x *g* for 15 minutes at 4°C). The aqueous layer was transferred into new tube using cut-off tips, 0.6 vol isopropanol was added, mixed by inversion, and incubated on ice for 3 minutes. The genomic DNA was spooled by a sealed Pasteur pipette, rinsed with 10 mL ice-cold 70% ethanol, and dissolved in 1 mL TE at 4°C overnight.

To check the genomic DNA quality, gel electrophoresis was done with 0.7% agarose at 100 V for 2 hour 20 minutes in TBE buffer (92 mM Tris, 2 mM EDTA, 88 mM boric acid). 1µL of isolated genomic DNA was mixed with 1 µL of 6x DNA loading dye (Thermo Scientific) and 4 µL of deionized water. GeneRuler DNA ladder mix (Thermo Scientific) and 1 kb DNA extension ladder (Invitrogen) were used as DNA ladder and Fosmid Control DNA from the CopyControl HTP Fosmid Library Production Kit with pCC2FOS Vector (Epicentre Biotechnologies) was used as size and concentration control. The Fosmid Control DNA has a size of 40 kb with 100ng/µL concentration.

Preparation of genomic DNA for ligation to pCC2FOS fosmid.

After confirming the quality and quantity of isolated genomic DNA, 1 µg DNA from *Gordonia* sp. Bu15_44 and 1 µg DNA of *Gordonia terrae* Bu15_45 were combined and loaded into 1% low melting point (LMP) agarose gel with fosmid control DNA loaded on each side of the gel. The electrophoresis was run for 30 minutes at 60V,

then the voltage was lowered to 40 V for 16 hours. Thereafter, each side of the gel were sliced including a small portion of the well where genomic DNA was loaded. Both slices were stained in ethidium bromide (EtBr) solution (0.5 $\mu\text{g}/\text{mL}$), and the position of the fosmid control DNA band were marked. The gel slices were reassembled back as in **Figure III.5**, and the genomic DNA was excised according to the markings on the side gel slices. This was done to prevent the genomic DNA for library construction to be exposed to EtBr or UV light.

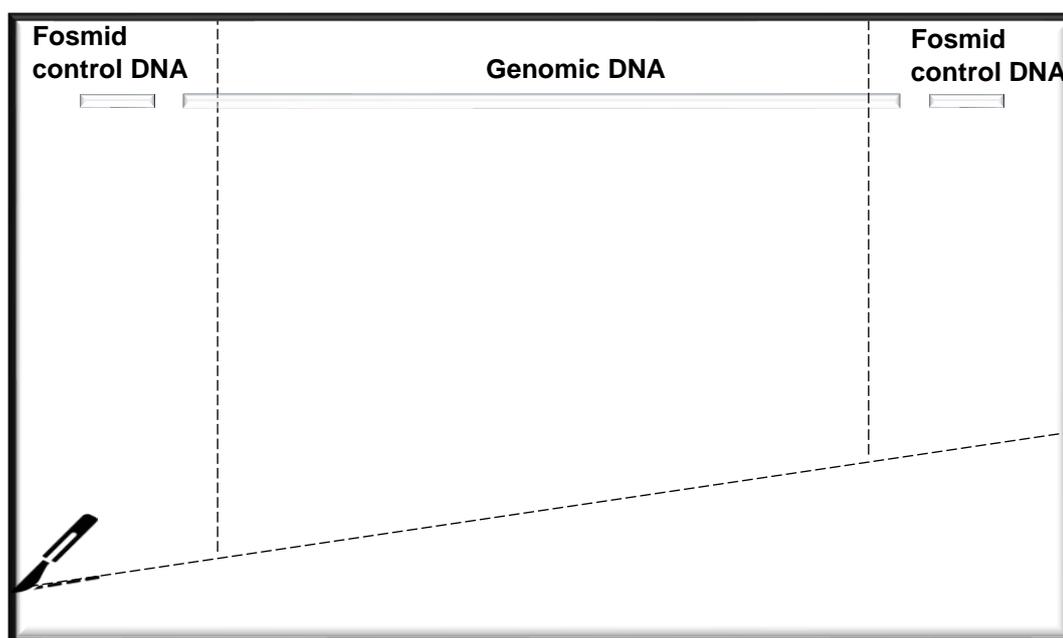


Figure III.5: Illustration of LMP agarose gel for size selection of genomic DNA. Genomic DNA was loaded into 1% LMP agarose gel. Fosmid control DNA was used as size control. After overnight run at 40 V, the LMP agarose gel was sliced according to the dashed line, and each slice from both sides of the gel were stained with EtBr.

The excised gel was placed in 15 mL tube and weighed. Then, it was washed twice with 2 volumes of 1x b-Agarase I Buffer (New England Biolabs) on ice for 30 minutes each and the LMP was melted by incubation at 65°C for 10 minutes. The tube was quickly transferred to 42°C water bath and incubated for 1 hour with 1 unit of β -Agarase I. Then, it was centrifuged (10,000 $\times g$ for 20 minutes at room temperature). 90-95% of the supernatant was transferred to a new 15 mL tube while avoiding the gelatinous pellet. The DNA was precipitated by addition of 1/10 volume of 3 M sodium acetate (pH 7), gently mixed, then addition of 2.5 volumes of 100% ethanol,

mixed by inversion. After incubation for 10-15 minutes at room temperature, the precipitated DNA was centrifuged (10,000 x *g* for 30 minutes at room temperature) to achieve DNA pellet. 95% of the supernatant was removed carefully to not disturb the DNA pellet. The pellet was then washed twice with 70% ethanol with centrifugation (10,000 x *g* for 20 minutes at 4°C). After the second wash, the supernatant was removed carefully and the pellet was air-dried to remove remaining ethanol. The DNA was then resuspended in 55 µL Tris-HCl buffer (10mM Tris-HCl, pH 8.5). To check DNA quality and concentration, analytical gel electrophoresis was done as previously described.

The size-selected DNA was blunt-ended using end-repair enzyme mix (Epicentre Biotechnologies) with the reaction described in (**Table III.1**)

Table III.1 DNA blunt-ending reaction mix

| Component | Volume | Final Concentration |
|--|--------|---------------------|
| 10x end-repair buffer | 8 µL | 1x |
| 2.5 mM dNTP mix | 8 µL | 0.25 mM |
| 10 mM ATP | 8 µL | 1 mM |
| Genomic DNA | 20 µL | 25 ng/µL |
| End-repair enzyme mix (T4 DNA polymerase and T4 polynucleotide kinase) | 4 µL | - |
| Sterile water | 32 µL | - |

The reaction was incubated for 45 minutes at room temperature, and was stopped by adding 3.33 µL of 250 mM EDTA, then incubation at 70°C for 10 minutes. Thereafter, 120 µL sterile deionized water, 20 µL of 3M sodium acetate (pH 5.2), and 140 µL isopropanol was added and mixed by inversion. After 30 minutes of precipitation at room temperature, the blunt-ended DNA was centrifuged at maximum speed for 30 minutes at 4°C. The supernatant was removed very carefully, 500 µL of ice-cold 70% ethanol was added without disrupting the pellet, and then centrifuged at maximum speed for 10 minutes at 4°C. The supernatant was removed very carefully and completely. Then, the DNA pellet was air-dried at room temperature. 20 µL of Tris-HCl buffer (pH 8.5) was added to the dried pellet,

and was incubated for 30 minutes on ice to let the DNA pellet dissolve. To check DNA quality and concentration, analytical gel electrophoresis was done as previously described.

Next, blunt-ended DNA was ligated to pCC2FOS fosmid with 1:10 molar ratio. T4 DNA ligase (New England Biolabs) was used according to the reaction described in **Table III.2**. The reaction was incubated at 16°C overnight and stopped by incubation at 65°C for 15 minutes.

Table III.2 Ligation reaction of blunt-ended DNA to pCC2FOS fosmid

| Component | Volume | Final Concentration |
|--------------------------|--------|---------------------|
| 10x ligase buffer | 2 µL | 1x |
| 0.5 µg/µL pCC2FOS fosmid | 0.6 µL | 15 ng/µL |
| Blunt-ended insert DNA | 15 µL | 7.5 ng/µL |
| T4 DNA ligase | 2 µL | 800 U |
| Sterile water | 0.4 µL | - |

Packaging and plating of fosmid clones

An overnight culture of *E. coli* EPI300-T1^R was diluted 1:10 in 50 mL lysogeny broth (LB: 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, pH 7.5) supplemented with 10 mM MgSO₄ and incubated at 37°C with 180 rpm shaking to an OD₆₀₀ of 0.8 – 1.0. The cells were kept at 4°C until needed.

MaxPlax lambda packaging extract (Epicentre Biotechnologies) was thawed on ice and 25 µL of each packaging reaction was transferred to a new 1.5 mL tube. The remaining 25 µL in the original tube was returned to -80°C. 10 µL ligation reaction was added to each 25 µL of the MaxPlax packaging extract and mixed gently by pipetting without introducing air bubbles. The mixture was spun for a few seconds to pull down the liquid from the lid and tube wall. After the packaging reaction was incubated at 30°C for 90 minutes in water bath, the remaining 25 µL MaxPlax packaging extract was taken from -80°C and added into the incubated reaction. The reaction was continued to be incubated for another 90 minutes at 30°C. Thereafter,

940 μL of phage dilution buffer (PDB: 10 mM Tris-HCl, pH 8.3, 100 mM NaCl, 10 mM MgCl_2) was added and mixed gently.

Before plating, the packaged phage particles was titrated by making serial dilutions (1:10, 1:100, and 1:1000) of the 1 mL packaged phage with PDB. 10 μL of each dilution were added separately into 100 μL of previously prepared *E. coli* EPI300-T1^R and incubated at 37°C for 20 minutes in a water bath. The infected *E. coli* cells were plated on LB plates with 12.5 $\mu\text{g}/\text{mL}$ chloramphenicol and incubated at 37°C overnight. The titer of packaged phage particles was calculated by: number of colonies x dilution factor x 100.

The 1 mL MaxPlax packaging reaction was mixed with 10 mL previously prepared *E. coli* EPI300-T1^R and incubated at 37°C for 20 minutes in a water bath. After the incubation, it was diluted into 500 mL of LB containing 5 g/L SeaPrep agarose (Lonza) and 12.5 $\mu\text{g}/\text{mL}$ chloramphenicol to achieve 30 cfus/mL. Since the library should cover 20-fold of the *Gordonia* genomic DNA size (6 Mb each), 6000 cfus were required. Therefore, 1 mL aliquots of the *E. coli* EPI300-T1^R with packaged fosmid were distributed into 217 pre-cooled 2 mL screw-cap vials. The vials were kept on ice water for 1 hour to let the soft agar solidify. Thereafter, the vials were carefully removed from ice to 37°C incubator to be incubated for 16-18 hours. The colonies should grow spatially as white dots. The colonies were counted to confirm the number of colonies per vial. Then, the vials were vortexed, and 30 μL aliquots from each vial in the same row were taken to make a row pool. 0.5 mL of 50% glycerol was added to the vials and the library was stored at -80°C.

III.2.3. Colony PCR

Colony PCR was performed for library screening and for checking Gibson assembly results. Primers for library screening were designed to recognize the beginning and the end of the putative BGC, while primers for corroboration of Gibson assembly were designed to bind in different fragments (usually forward primers bind to the vector and the reverse primer binds to the insert). In the first round of library screening, row pool was screened by PCR (**Table III.3**). When the row pool yielded the expected amplicon, the individual vials were screened in the second round of PCR. An aliquot of 10 μL from the single vial that yielded expected amplicon was

plated on LB containing 12.5 µg/mL chloramphenicol. Single colonies were screened in the third round of PCR to obtain a clone that carries target BGC. In the case of target BGC larger than 40 kb, the complete BGC could not be carried by a single clone. Therefore, the screening primers were designed to recognize the BGC in two or more fragments (**Table S1**). The PCR was performed in a Biometra TRIO thermocycler (Analytik Jena AG) using the following program: 95°C for 10 minutes; 34 cycles of 95°C for 45 seconds, 50-60°C for 45 seconds (applied annealing temperature was depending on the primer sequence), 72°C for 1 minute/kb (extension time was varied depending on the length of the fragment to be amplified), followed by a final extension step at 72°C for 5 minutes. The PCR result was visualized in gel electrophoresis with 1% agarose gel in TAE buffer.

Table III.3 PCR reaction mix

| Component | Final concentration |
|-----------------------------|---------------------|
| Sterile Water | ad to 20 µL |
| Green GoTaq® Flexi Buffer | 1x |
| MgCl ₂ | 1.25 mM |
| DMSO | 5% |
| dNTPs | 50 µM each |
| Forward primer | 500 nM |
| Reverse primer | 500 nM |
| GoTaq® Flexi DNA polymerase | 1 U |
| DNA template | from colony |

III.2.4. Fosmid/plasmid isolation

Positive clones from the library screening were inoculated in 2 mL LB supplemented with 12.5 µg/mL chloramphenicol and 4 µL of CopyControl™ Fosmid Autoinduction Solution and incubated at 37°C overnight. The fosmid carrying target BGC was isolated by alkaline lysis method.

Alkaline lysis plasmid isolation method was used to isolate fosmid and pCAP03 vector. 2 mL of an overnight culture was centrifuged at 10,000 rpm for 5 minutes. The supernatant was removed, and the cell pellet was resuspended in 100 µL of

buffer P1 (50 mM Tris-HCl pH 8, 10 mM EDTA). Subsequently, buffer P2 (200 mM NaOH, 1% SDS) was added and mixed gently by inverting the tube five to six times. Ice-cold buffer P3 (3M potassium acetate pH 5.5) should be added in less than 5 minutes after the addition of buffer P2. Then, the tube was mixed by inversion and stored on ice for 5 minutes before centrifugation at 13,000 rpm for 5 minutes. The supernatant was transferred into a new tube, and 450 μ L of phenol:chloroform:isoamylalcohol (24:25:1) was added and mixed by vortexing briefly. Thereafter, it was centrifuged at 10,000 rpm for 5 minutes. The aqueous upper phase was transferred into a new tube, 900 μ L of ice-cold isopropanol was added, and the tube was kept on ice for 2 minutes before a centrifugation step at 10,000 rpm for 30 minutes. The supernatant was discarded, and 300 μ L ice-cold 70% ethanol was added without disturbing the pellet. The tube was centrifuged for 30 seconds at 10,000 rpm, and the ethanol was removed carefully. The pellet was air-dried at room temperature for 5 minutes, then dissolved in 50 μ L TE buffer (10 mM Tris-HCl pH 8).

Isolated fosmid was sent for sequencing using sequencing primer pCC2FOS-For and pCC2FOS-Rev to confirm the borders of the DNA insert.

III.2.5. TAR cloning

TAR cloning was performed according to Zhang *et al.*, (2019b) with some modifications. As the capture vector, pCAP03 (**Figure III.6A**) was digested with *Xho*I and *Nde*I to create ready-to-use pCAP03 (RTU-pCAP03). Then, a synthetic double strand DNA that contains *Pme*I recognition site flanked by 50 bp homologous regions to the target BGC was cloned into the linearized vector using Gibson Assembly method (Gibson *et al.*, 2009) (**Figure III.6B**). After the isothermal assembly, the reaction was dialyzed with a 0.025 μ m nitrocellulose membrane (Merck™ MF-Millipore™), and subsequently transferred to *E. coli* TOP10 cells as a plasmid maintenance host by electroporation using Micropulser Electroporator (Bio-Rad) in a 0.2 cm electroporation cuvette at a voltage of 2.5 kV.

The capturing of the target BGC into the capture vector was done in *S. cerevisiae* VL6-48N strain. To do this, a single colony of the yeast strain was inoculated in YPD medium (Yeast extract peptone dextrose medium: yeast extract 10 g/L, peptone 20

g/L, and 2% glucose added after sterilization) supplemented with adenine (100 ng/mL, 74 mM HCl). After incubation in shaker incubator at 30°C, 0.5 mL of the overnight culture was inoculated in 50 mL YPD medium supplemented with adenine and incubated at 30°C with 200 rpm shaking. When an OD₆₀₀ of 0.7 – 1.0 was reached, the cultivation flask was put on ice for 10 minutes. Then, the cells were washed with 50 mL ice-cold sterile water (centrifugation for 3 minutes at 1800 x g, 4°C) and then resuspended in 50 mL 1 M ice-cold sorbitol. The cell suspension was kept at 4°C overnight to equilibrate osmotic pressure.

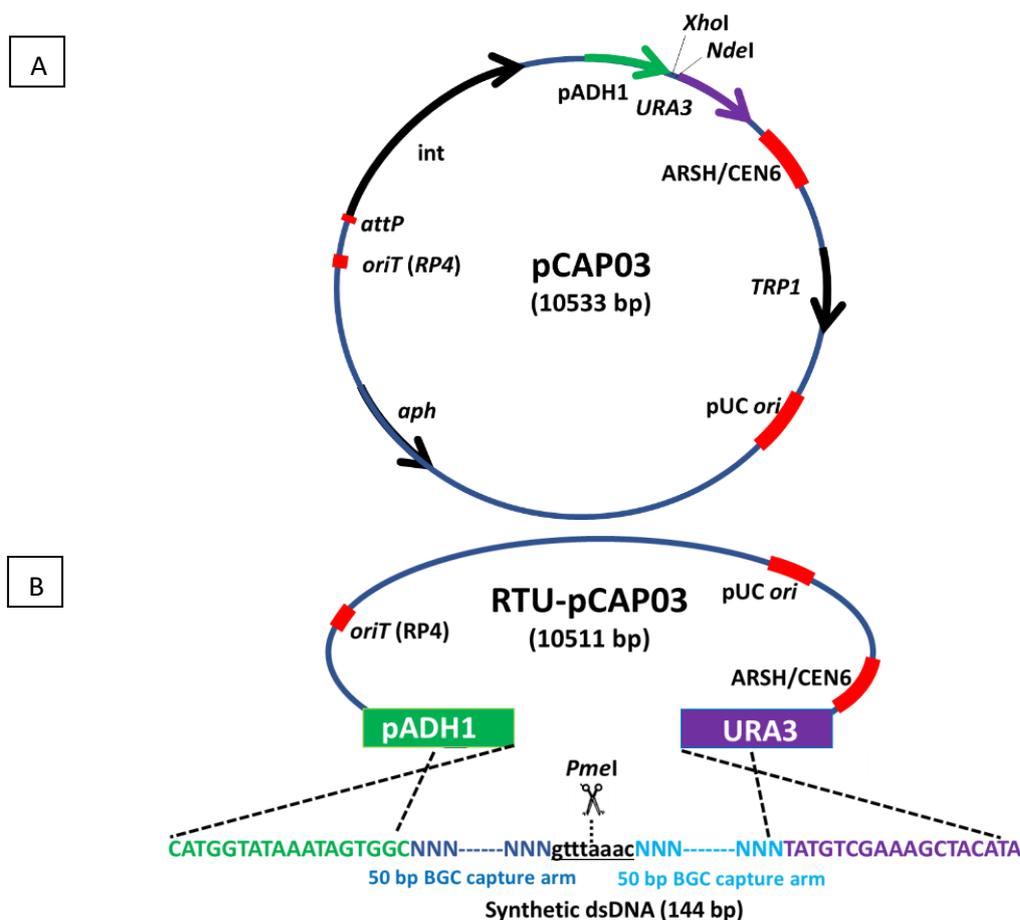


Figure III.6: Vector map of pCAP03 (A) and TAR capture vector (B). **A:** pCAP03 contains promoter of *Schizosaccharomyces pombe ADH1* gene (pADH1) upstream of *URA3* gene, which encodes orotidine 5'-phosphate decarboxylase as a selectable marker. This vector also contains *ARSH/CEN6*, a low-copy number yeast origin of replication; *pUC ori*, a high-copy number *E. coli* origin of replication; *oriT*, an origin of transfer of RP4 plasmid. In addition to those, *TRP1* is available as yeast selection marker on tryptophan-deficient medium, *aph* is present as a kanamycin resistant gene, and *attP* site and integrase encoding gene (*int*) are available for ϕ C31 integrase system. **B:** TAR capture vector was generated by inserting a 144 bp synthetic DNA capture arm into RTU-pCAP03 using Gibson Assembly method.

On the next day, the tube containing the yeast cells was inverted several times and spun for 3 minutes at 1800 x *g*, 4°C. The supernatant was removed completely, and the cells were resuspended in 20 mL of SPE (2 mM HEPES buffer pH 7.5, 10 mM EDTA pH 8.0, 182 g/L sorbitol) by vortexing. 40 µL of 2-mercaptoethanol was added and mixed before the addition of 80 µL zymolyase-20T (10 mg/mL zymolyase-20T, 25% (w/v) glycerol, 50 mM Tris-HCl pH 7.5). The mixture was incubated for 40 minutes in 30°C incubator with 20 rpm shaking.

OD₆₀₀ of the cell suspension was measured to determine the spherolasting level of the yeast cells. 800 µL of 1 M sorbitol and 200 µL of zymolase treated cells were added into a 1 cm-wide cuvette for OD₆₀₀ measurement. Then, 200 µL of 10% SDS was added into the cuvette, and the OD₆₀₀ was remeasured. If the difference between the first and second OD₆₀₀ measurement has reached five- to twenty-fold difference, the spherolasts had been formed and were ready to use. Otherwise, the incubation is extended for another 10 minutes and the OD₆₀₀ measurement was done again to check the spherolasting level. The spheroplasts were then prepared to be transformed by *Pmel*-linearized capture vector and the plasmid that carries target BGC according to the protocol by Zhang *et al.* (2019b). The transformed yeast cells were screened by PCR to select the colony that carries pCAP03-BGC.

The plasmid was isolated from the positive colony by alkaline lysis method according to Zhang *et al.* (2019b) with the following modifications. After the addition of potassium acetate and a centrifugation step, the supernatant was removed into a new tube. 5 µL RNase A (10 mg/mL) was added into the supernatant and incubated for 15 minutes at room temperature. Subsequently, 20 µL proteinase K (20 mg/mL) was added and incubated for 30 minutes at 55°C. The DNA extraction was then proceeded with phenol-chloroform protein removal, and DNA washing steps by isopropanol and 70% ethanol. The isolated plasmid was then dissolved in 50 µL of 10 mM Tris-HCl pH 8 for storage. The construction of the plasmid by yeast homologous recombination was confirmed by checking the restriction pattern of the plasmid. Prior to electroporation into *E. coli* TOP10 as maintenance host, the DNA was dialyzed with a 0.025 µm nitrocellulose membrane (Merck™ MF-Millipore™) for 30 minutes.

III.2.6. Gibson assembly

5x Gibson assembly reaction buffer was made by mixing 3 mL of 1M Tris-HCl pH 7.5, 200 μ L of 1 M $MgCl_2$, 60 μ L of 100 mM dGTP, 60 μ L of 100 mM dATP, 60 μ L of 100 mM dTTP, 60 μ L of 100 mM dCTP, 300 μ L of 1M DTT, 1.5 g PEG-8000, 300 μ L of 100 mM NAD, in a final volume of 6 mL. The Gibson assembly master mix was made by mixing 320 μ L of Gibson assembly reaction buffer, 0.64 μ L of 10 U/ μ L T5 exonuclease, 20 μ L of 2 U/ μ L Phusion DNA polymerase, 160 μ L deionized water, in a final volume of 1.2 mL. The Gibson assembly master mix was aliquoted into 15 μ L each, and kept in -20°C until needed.

The amplification of the fragments for Gibson assembly were performed by using Q5 polymerase (New England Biolabs) according to manufacturer's recommendation, with annealing temperature that was calculated by using New England Biolabs (NEB) T_m calculator. The concentration of the DNA fragments that were going to be assembled was quantified by measuring the absorbance at 260 nm using a microvolume spectrophotometer (Eppendorf BioSpectrometer®). The DNA fragments were added into the 15 μ L Gibson assembly master mix in equimolar amounts to reach the final reaction volume of 20 μ L. Then, the reaction was incubated at 50°C for two hours, followed by a dialysis step with a 0.025 μ m nitrocellulose membrane (Merck™ MF-Millipore™) for 20 minutes and transformation of electrocompetent cells.

To create pGM1202-C1, three fragments were assembled: (i) *orf118* that was amplified by using the primer pair pGM1202-ermE*-F/118-119-R, (ii) *nbtT* that was amplified using the primer pair 118-119-F/119-pGM-R, (iii) pGM1202 that was digested with *HindIII* and *BamHI*. To generate pGM1202-C1, four fragments were assembled: (i) *orf118* that was amplified by using the primer pair pGM1202-ermE*-118/118-119-R, (ii) *nbtT* that was amplified using the primer pair 118-119-F/119-123-R, (iii) *nbtH* that was amplified using the primer pair 119-123-F/123-pGM-R, (iv) pGM1202 that was digested with *HindIII* and *BamHI*.

III.2.7. λ -Red Recombination for promoter insertion

Addition of *ermE** or *tcp830* promoter in front of the cloned BGC was performed by λ -Red Recombination according to the protocol by Gust *et al.* (2006) with some modifications. Instead of using pIJ790, pKD46 was used as the λ -Red recombination plasmid. Therefore, carbenicillin (50 μ g/mL) was used as a selective marker for pKD46. After transformation of the *E. coli* BW25113/pKD46 by pCAP03-BGC, the positive colonies were checked by PCR and by growing them in selective LB agar containing 50 μ g/mL carbenicillin and 50 μ g/mL kanamycin.

10 mL LB liquid containing above mentioned selective antibiotics was inoculated with 5% of *E. coli* BW25113/pKD46/pCAP03-BGC overnight culture and incubated at 30°C for 1 hour. Then, 100 μ L of 1 M L-arabinose was added to induce the expression of the λ -Red genes and the incubation was continued until OD₆₀₀ of ~0.5. The cells were prepared to be electrocompetent by washing with ice-cold 10% glycerol, and transformed with linear dsDNA that contains the promoter with apramycin or streptomycin resistance cassette as selective marker, FRT sites, and 40 bp homologous region on each ends. Transformants were grown on LB containing kanamycin and apramycin/streptomycin (50 μ g/mL) at 37°C to promote the loss of pKD46. The integration of promoter cassette upstream of the captured BGC in pCAP03 was confirmed by PCR and restriction pattern analysis.

III.2.8. Triparental conjugation

Triparental conjugation to transfer pCAP03-BGC into *S. coelicolor* M1146 and *S. lividans* TK24 was performed according to Zhang *et al.* (2019b). To confirm the integration of pCAP03-BGC into the genomic DNA of expression host, a few colonies were grown in 3 mL ISP2 at 30°C overnight for genomic DNA isolation and PCR was performed to check the presence of target BGC. The colony with positive result was grown on mannitol soy (MS) agar plate without any antibiotics for spore suspension preparation. After three days of incubation, the spores were collected from the plate by addition of 2 mL sterile 20% glycerol and collection of the spore suspension using sterile cotton. The cotton was put into a syringe, and the spore suspension was pushed out of the cotton into cryogenic tubes.

III.2.9. *Gordonia* cultivation for nocobactin NA dereplication

G. terrae Bu15_45 was cultured in 100 mL minimal medium (MM) according to Hoshino *et al.*, (2011) (MM: 1.5% Na₂HPO₄·2H₂O, 0.3% KH₂PO₄, 0.5% Na₃C₆H₅O₇·2H₂O, 0.02% NH₄Cl, 0.1% MgSO₄·7H₂O, and 0.05% NaCl) at 37°C. After three days of cultivation, the culture was harvested and freeze-dried. Into the dried culture, 20 mL ethyl acetate was added, followed by a sonication step for 10 minutes. The mixture was centrifuged for 10 minutes to separate the ethyl acetate extract and the ethyl acetate-insoluble part. The ethyl acetate extract was collected in a new round-bottom flask, and evaporated to dryness by using rotary evaporator. The crude extract were collected by 1 mL methanol from the flask and sent to be analysed by LCMS as the ethyl acetate crude extract. To prepare the methanol crude extract, 20 mL methanol was added to the ethyl acetate-insoluble part and sonicated for 10 minutes. The methanol extract was subsequently concentrated into 1 mL by rotary evaporator, and sent to be analysed by LCMS.

III.2.10. Heterologous expression and compound detection

Streptomyces expression host was cultivated in 10 mL TSB (17 g/L peptone from casein, 3 g/L peptone from soymeal, 2.5 g/L D-glucose, 5 g/L NaCl, 2.5 g/L K₂HPO₄, pH 7.3) at 30°C for three days as a pre-culture for heterologous expression of the cloned BGC. 1% of the pre-culture was used to inoculate TSB and ISP2. Expression construct that has tcp830 promoter was induced with 2 µg/mL anhydrotetracycline after certain incubation period. Expression cultures were incubated for 7-10 days at 30°C before being harvested.

The heterologous expression cultures were centrifuged to separate the medium and cell pellet, and both parts were freeze-dried. Subsequently, ethyl acetate was added into the dried medium, and methanol was added to the dried pellet. The ethyl acetate and methanol crude extracts were subjected to mass spectrometer for detection of expressed compound.

Mass spectra were recorded on a micrOTOF-Q mass spectrometer (Bruker) with an ESI-source coupled with a HPLC Dionex Ultimate 3000 (Thermo Scientific) using an EC10/2 Nucleoshell C18 2.7 µm column (Macherey-Nagel). The column temperature was 25°C. MS data were acquired over a range from 100 to 1000 *m/z*

in positive mode. Auto MS/MS fragmentation was achieved with rising collision energy (35–50 kV over a gradient from 500 to 2000 m/z) with a frequency of 4 Hz for all of the ions over a threshold of 100. The injection volume was 2 μL with a concentration of 1 mg/mL.

III.3. Results

III.3.1. INDOBIO Project

Previously, it was reported that sea slug associated bacteria could produce bioactive secondary metabolite. Two of them, *Gordonia terrae* Bu15_45 and *Gordonia* sp. Bu15_44 that were isolated from *Hexabranchnus sanguineus* egg mass showed moderate activity against *Arthrobacter psychrolactophilus* (Böhringer *et al.*, 2017). Here, the genomic approach was taken to elucidate the antimicrobial potential of these strains. The PCR screening for A-domains (NRPS) and KS-domains (PKS) with degenerated primers could discover seven A-domain sequences and three KS domains from the *Gordonia* strains (Böhringer *et al.*, 2017). Thus, the genomic DNA of these two *Gordonia* strains were subjected to whole-genome sequencing and analyzed for the presence of putative NRPS or PKS cluster by antiSMASH.

III.3.1.1 *In silico* analysis of putative BGC

In silico analysis of *G. terrae* Bu15_45 genome sequence data using the tool antiSMASH resulted in the detection of 15 biosynthetic gene clusters (BGCs). Ten of the identified BGCs only consisted of a few domains and could not be regarded as a complete BGC. The remaining five BGCs possessed all necessary putative domains needed to constitute a complete BGC (**Table III.4**). These clusters were screened from the artificial metagenomic library to be subsequently heterologously expressed in the host strains belonging to the genus *Streptomyces*.

Five putative BGCs that were detected by antiSMASH were analyzed closer and will be discussed in the following section. If the A domain prediction by the integrated NRPSPredictor2 could not specify the monomer, the Stachelhaus code was analyzed manually to find the closest match (Challis *et al.*, 2000; Stachelhaus *et al.*, 1999). Similar MIBiG clusters were also analyzed to be compared to the detected cluster. NCBI@blastx was also run on some additional biosynthetic genes and on genes that seem to be not related to compound biosynthesis to confirm their function.

Table III.4. Putative biosynthetic BGCs detected from the genomic DNA of *Gordonia terrae* Bu15_45 by *in silico* analysis using antiSMASH

| Cluster numbering | Type | Size (kb) | Region-to-region similarity to MIBiG cluster |
|-------------------|-----------------|-----------|--|
| 52 | NRPS/PKS hybrid | 26 | Nocobactin NA (70%) |
| 23 | NRPS | 41 | Massetolide A (71%) |
| 37 | NRPS | 39 | Brabantamide A (60%) |
| 20 | NRPS | 17 | Pyreudiones (66%) |
| 9 | NRPS | 129 | 9A: Icosalides (67%) 9B: Sevadacin (72%) |

III.3.1.1.1 Cluster 52: Nocobactin NA

Cluster 52 is an NRPS/PKS hybrid cluster with the size of 25,316 bp. Although the sequence similarity is only 70% to the nocobactin NA BGC, all genes of the nocobactin NA BGC have their homolog in cluster 52 based on the significant blast hits comparison. Therefore, cluster 52 was identified as a nocobactin NA BGC. Hoshino *et al.* (2011) identified the nocobactin NA BGC in *Nocardia farcinica* to be separated in two genetic loci that are 180 kb away from each other. Interestingly, in *G. terrae* Bu15_45 the BGC was identified as one cluster (**Figure III.7**). In addition to the biosynthetic genes that were previously described, a gene for a periplasmic binding protein of an ABC transporter (*orf118*) and a gene encoding for a protein of unknown function (*orf127*) were also identified in cluster 52 (**Table S3**). A BLAST search using the translated amino acid sequence of *orf127* as query only gave hits for hypothetical proteins from *Gordonia* spp.

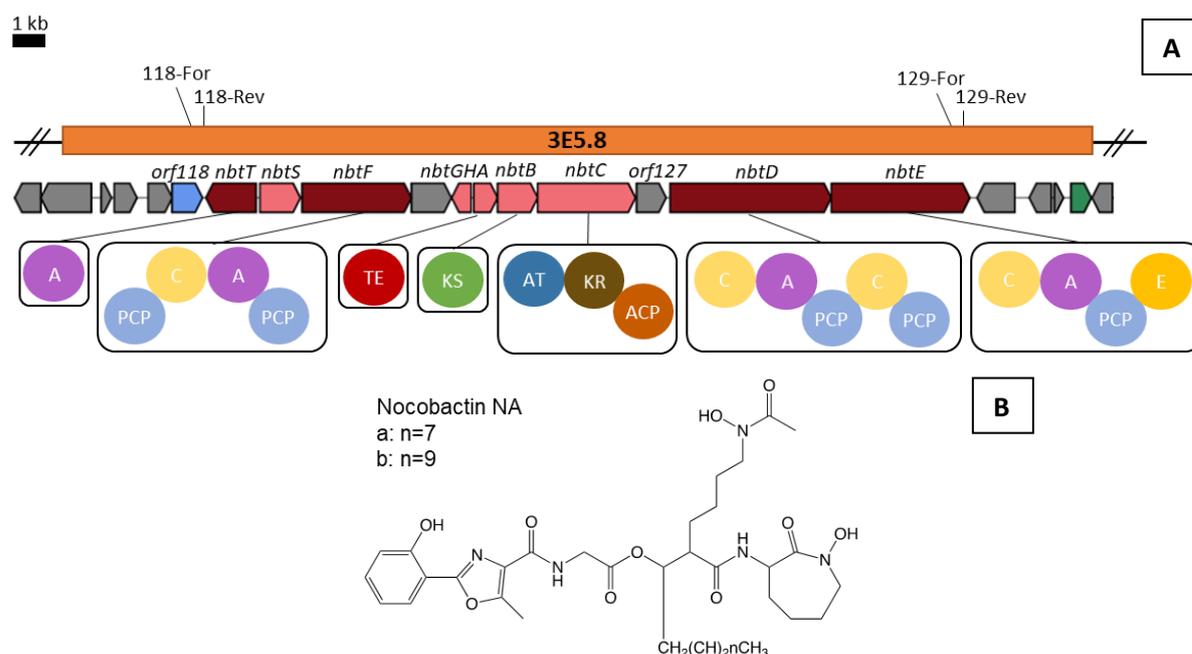


Figure III.7: Cluster 52. A: Cluster 52 consists of 12 genes, where 10 of them were described as nocobactin NA BGC, one transporter gene (*orf118*) and one gene for protein with unknown function (*orf127*). NbtTFDE are NRPS modules and NbtABC are PKS modules. *nbtS* encodes for salicylate synthase, *nbtG* for lysine monooxygenase, and *nbtH* for lysine acetyltransferase. Cluster 52 was isolated from the *Gordonia* library by PCR screening using two primer pairs, 118-For/118-Rev and 129-For/129-Rev. The orange bar marked the DNA insert that is carried in the fosmid of colony 3E5.8. **B:** Nocobactin NA is the predicted product of cluster 52.

III.3.1.1.2 Cluster 23

Cluster 23 is quite conserved in *Gordonia* strains. It has 90% similarity to an orphan BGC from *G. polyisoprenivorans* VH2, 88% similarity to a BGC from *G. insulae* MMS17-SY073, and 88% similarity to a BGC from *Gordonia* sp. Bu15_44, a strain that was also isolated from *Hexabanchus sanguineus* egg mass. antiSMASH analysis of both clusters from *G. terrae* Bu15_45 and *Gordonia* sp. Bu15_44 are identical. They are NRPS clusters, with seven A domains and seven genes that encode for putative tailoring enzymes (**Figure III.8, Table III.5**).

Although cluster 23 has 71% sequence similarity to the massetolide A BGC, the A domain specificities are different. The A domain in *orf23.4* was predicted by antiSMASH to have substrate specificity to 2,3-dihydroxybenzoic acid (80% Stachelhaus code match) and the A domain in *orf23.8* was predicted to activate glycine (100% Stachelhaus code match). *orf23.7* has two A domains, which were predicted to activate β -alanine and threonine, with 90% Stachelhaus code match

for both A domains. Meanwhile, *orf23.6* has three A domains. The first one shows a 90% Stachelhaus code match to ornithine, while the second and third domain have low similarities to known Stachelhaus codes that were available in the database. The second A domain was predicted to have specificity for hydrophobic-aromatic amino acid, e.g. phenylalanine, tryptophan, phenylglycine, tyrosine, β -hydroxytyrosine, while the third A domain was predicted to have specificity for serine/threonine/glutamine/glutamate.

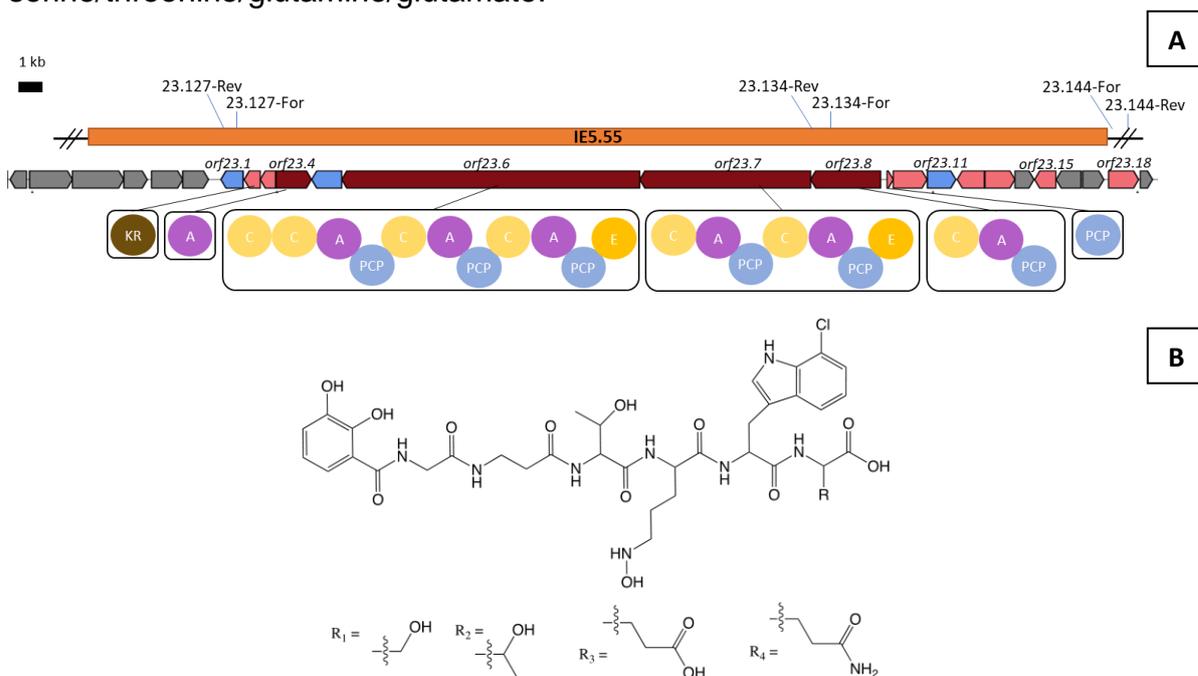


Figure III.8: Cluster 23. A: Cluster 23 consists of 18 open reading frames. Three of them (*orf23.6*, *orf23.7* and *orf23.8*) encode for NRPS modules, while *orf23.4* and *orf23.9* are standalone A and PCP domains, respectively. Three transporter-related genes were predicted in this cluster (in blue), and seven additional biosynthetic genes that modify the monomers were annotated (in pink). Cluster 23 was isolated from the *Gordonia* library by PCR screening using three primer pairs, 23.127-For/23.127-Rev, 23.134-For/23.134-Rev, and 23.144-For/23.144-Rev. The orange bar marked the DNA insert that is carried in the fosmid of colony IE5.55. *orf23.18* that was supposed to be included in cluster 23, was not included in the fosmid IE5.55 and would be cloned separately. **B:** Core structure prediction of the compound that is produced by cluster 23. Due to the unspecificity of the third A domain in 23.6, R was used to mark the side chain of the last amino acid.

In addition to the NRPS modules, additional genes that modify the NRP building blocks are also present in the BGC, which are also different to the tailoring enzymes found in massetolide A BGC. *orf23.2* was predicted as an inactive PKS-KR domain by antiSMASH. BLASTp of the translated gene gave hits for SDR family oxidoreductase (98.02%) and 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase

(69.31%) from *Gordonia* strains. Since the A domain in *orf23.4* has specificity to 2,3-dihydroxybenzoate, the latter blastp prediction fits better to cluster 23 assembly line. *orf23.12*, *orf 23.3* and *orf 23.2* were proposed to modify chorismate to 2,3-dihydroxybenzoate. *orf23.10* was annotated as tryptophan halogenase. Therefore, it is more likely that the second A domain in *orf23.6* has the specificity towards tryptophan.

Interestingly, these clusters do not have a TE domain that is required for the release of the biosynthetic product. Instead of a TE domain, an additional C domain is available at the C-terminus of the NRPS domains. Although unusual, a C domain can replace the TE domain function by catalyzing an intramolecular cyclization reaction, as proposed in the biosynthesis of rapamycin (Schwecke *et al.*, 1995), FK506 (Motamedi and Shafiee, 1998), mirubactin (Giessen *et al.*, 2012), and cyclosporine (Weber *et al.*, 1994).

Table III.5 Predicted function of *orfs* in cluster 23.

| Gene | Size (bp) | Protein function prediction |
|-----------------|-----------|--|
| <i>orf23.1</i> | 996 | Periplasmic binding protein of ABC transporter |
| <i>orf23.2</i> | 720 | Short-chain dehydrogenase/reductase (SDR) |
| <i>orf23.3</i> | 684 | Isochorismatase |
| <i>orf23.4</i> | 1578 | NRPS A domain for 2,3-dihydroxybenzoic acid |
| <i>orf23.5</i> | 1344 | Transmembrane secretion effector (MFS transporter) |
| <i>orf23.6</i> | 13311 | NRPS domains (monomer prediction: ornithine-tryptophan-serine/threonine/glutamate/glutamine) |
| <i>orf23.7</i> | 7629 | NRPS domains (monomer prediction: beta-alanine-threonine) |
| <i>orf23.8</i> | 3129 | NRPS domains (monomer prediction: glycine) |
| <i>orf23.9</i> | 282 | NRPS PCP domain |
| <i>orf23.10</i> | 1512 | Tryptophan halogenase |
| <i>orf23.11</i> | 1272 | Sodium/hydrogen exchanger family |
| <i>orf23.12</i> | 1080 | Isochorismate synthase |
| <i>orf23.13</i> | 1335 | Lysine/ornithine N-monooxygenase |
| <i>orf23.14</i> | 855 | Siderophore-interacting FAD-binding domain |
| <i>orf23.15</i> | 951 | Oxidoreductase |
| <i>orf23.16</i> | 1113 | Nucleoside hydrolase |
| <i>orf23.17</i> | 945 | Mechanosensitive ion channel |
| <i>orf23.18</i> | 1347 | Acetylornithine deacetylase |
| <i>orf23.19</i> | 549 | Phosphatidylethanolamine-binding protein |

III.3.1.1.3 Cluster 37

When compared to MIBiG cluster by region to region analysis, cluster 37 has 60% similarity score to brabantamide A, an antimicrobial compound that is biosynthesized by an NRPS and required a rhamnose BGC (*rml* gene cluster) that is located on different loci (Schmidt *et al.*, 2014). Cluster 37 consists of NRPS domains and saccharide tailoring enzymes (**Figure III.9**). The NRPS gene encodes for three modules, with substrate specificity to threonine (80%), leucine (60%), and tyrosine (60%) based on Stachelhaus code prediction. Upstream of the NRPS module, a saccharide biosynthetic cluster was annotated to contain two glycosyltransferase. Thus, the final product of cluster 37 was predicted to contain two glycosyl groups. Downstream of the NRPS module, an oxidoreductase tailoring enzyme was also expected to be involved in the biosynthesis of cluster 37 product (**Table III.6**). Cluster 37 was initially predicted to start from *orf37.0* to *orf37.22*. Thereafter, the following genes are predicted to be not related to compound biosynthesis. *orf37.34* and the transporter gene behind it was expected to be part of another BGC.

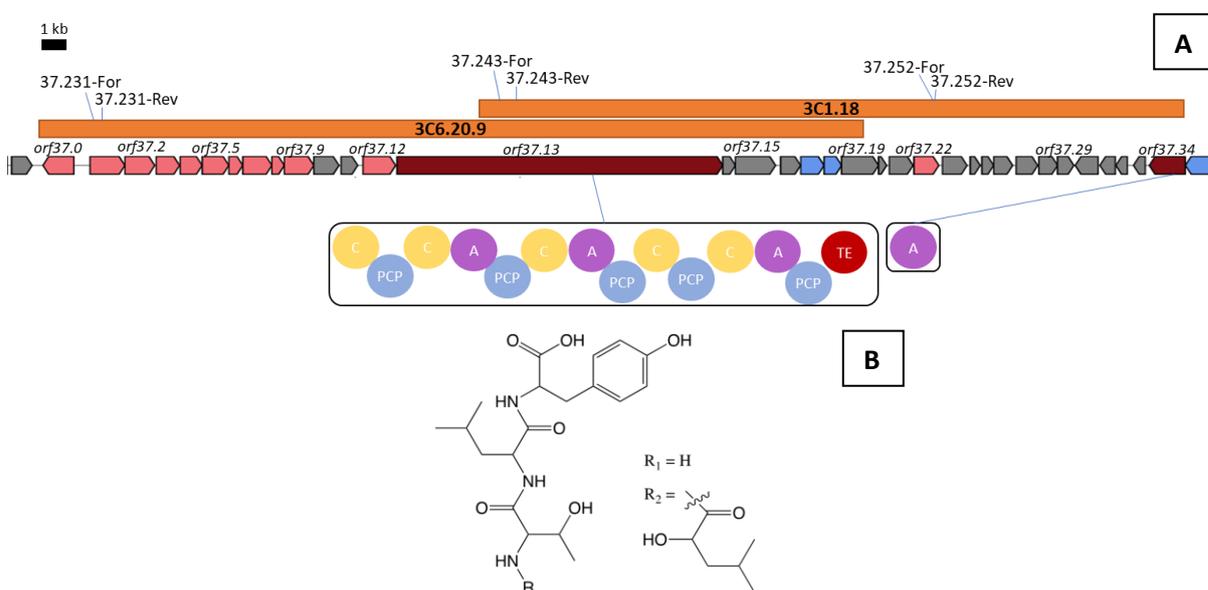


Figure III.9: Cluster 37. A: The putative BGC 37 consists of a saccharide BGC (13 orfs) and an NRPS cluster (10 orfs). The NRPS cluster has five condensation domains, three adenylation domains, five PCP domains, and one thioesterase domain. The 39 kb cluster was covered by two fosmids, 3C6.20.9 and 3C1.18. Three primer pairs were used to screen cluster 37 from the *Gordonia* library, i.e. 37.231-For/37.231-Rev, 37.243-For/37.243-Rev and 37.252-For/37.252-Rev. **B:** Core structure prediction of the compound that is produced by cluster 37 based on the specificity of the A domains. The incorporation of α -hydroxyisocaproic acid is elusive due to the distance of the A domain in *orf37.34* that is far away from the core NRPS cluster.

Table III.6 Predicted function of orfs in cluster 37.

| Gene | Size (bp) | Protein function prediction (%similarity)^a |
|-----------------|------------------|---|
| <i>orf37.0</i> | 1353 | O-antigen ligase family protein (97.55%) |
| <i>orf37.1</i> | 1521 | exopolysaccharide biosynthesis polyprenyl glycosylphosphotransferase (99.36%) |
| <i>orf37.2</i> | 1302 | nucleotide sugar dehydrogenase (99.54%) |
| <i>orf37.3</i> | 1023 | GDP-mannose 4,6-dehydratase (100%) |
| <i>orf37.4</i> | 960 | NAD-dependent epimerase/dehydratase family protein (99.37%) |
| <i>orf37.5</i> | 1167 | Glycosyltransferase (98.45%) |
| <i>orf37.6</i> | 612 | Acetyltransferase (98.92%) |
| <i>orf37.7</i> | 1275 | Glycosyltransferase (97.17%) |
| <i>orf37.8</i> | 546 | putative colanic acid biosynthesis acetyltransferase (98.34%) |
| <i>orf37.9</i> | 1290 | lipopolysaccharide biosynthesis protein (98.37%) |
| <i>orf37.10</i> | 1107 | polysaccharide pyruvyl transferase family protein (98.37%) |
| <i>orf37.11</i> | 729 | SGNH/GDSL hydrolase family protein (99.59%) |
| <i>orf37.12</i> | 1431 | polysaccharide biosynthesis tyrosine autokinase (98.11%) |
| <i>orf37.13</i> | 14217 | NRPS domains (monomer prediction: threonine-leucine-tyrosine) |
| <i>orf37.14</i> | 555 | Low molecular weight phosphotyrosine protein phosphatase |
| <i>orf37.15</i> | 1752 | Protein of unknown function |
| <i>orf37.16</i> | 876 | Protein of unknown function |
| <i>orf37.17</i> | 1020 | ABC transporter, permease protein |
| <i>orf37.18</i> | 768 | ABC transporter ATP-binding protein |
| <i>orf37.19</i> | 1617 | Fumarate reductase flavoprotein |
| <i>orf37.20</i> | 360 | 4Fe-4S binding domain |
| <i>orf37.21</i> | 1038 | NMT1/THI5 like |
| <i>orf37.22</i> | 1092 | oxidoreductase |
| <i>orf37.23</i> | 1077 | Hypothetical protein |
| <i>orf37.24</i> | 444 | Hypothetical protein |
| <i>orf37.25</i> | 516 | Hypothetical protein |
| <i>orf37.26</i> | 828 | NAD synthase |
| <i>orf37.27</i> | 945 | Manganese-containing catalase |
| <i>orf37.28</i> | 822 | Virulence factor BrkB |
| <i>orf37.29</i> | 720 | CDP-alcohol phosphatidyltransferase |
| <i>orf37.30</i> | 1002 | Putative zinc-binding metallo-peptidase |
| <i>orf37.31</i> | 678 | Unknown protein |
| <i>orf37.32</i> | 501 | Hemerythrin HHE cation binding domain |
| <i>orf37.33</i> | 516 | Unknown protein |
| <i>orf37.34</i> | 1578 | NRPS A domain for α -hydroxy-isocaproic acid |
| <i>orf37.35</i> | 1356 | Major facilitator superfamily transporter |

^a) % similarity was identified by NCBI@blastx

III.3.1.1.4 Cluster 20

Cluster 20 is a bimodular NRPS cluster with four additional putative biosynthetic genes (**Table III.7, Figure III.10A**). The NRPS part of the cluster has 66% sequence similarity to pyreudiones BGC, a monomodular NRPS without additional tailoring enzymes. The single A domain in the pyreudiones BGC has only 40% and 50% Stachelhaus code match to the first and second A domain in cluster 20, respectively. On the other hand, the additional biosynthetic enzyme that is encoded in *orf20.2*, phosphoenolpyruvate phosphomutase/carboxyphosphonopyruvate phosphomutase, is a unique tailoring enzyme that was found in the BGC of phosphinothricin tripeptide (PTT) (Blodgett *et al.*, 2005). The presence of this tailoring enzyme indicates that the product of cluster 20 might also contain a phosphonate group. Instead of the addition of two alanine residues as in PTT biosynthesis, the A domains in cluster 20 was predicted to have specificity towards valine with 60% Stachelhaus code match. To be noted is that the two A domains in cluster 20 have 90% Stachelhaus code match to each other, suggesting that they activate similar monomers. Therefore, the product of cluster 20 is roughly predicted in **Figure III.10B**.

Table III.7 Predicted function of orfs in cluster 20.

| Gene | Size (bp) | Protein function prediction |
|----------------|-----------|--|
| <i>orf20.1</i> | 1038 | iron compound ABC transporter, periplasmic |
| <i>orf20.2</i> | 759 | Carboxyvinyl-carboxyphosphonate phosphorylmutase/Phosphoenolpyruvate phosphomutase |
| <i>orf20.3</i> | 492 | Carbonic anhydrase |
| <i>orf20.4</i> | 936 | methionyl-tRNA formyltransferase |
| <i>orf20.5</i> | 8616 | NRPS domains (monomer prediction: valine-valine) |
| <i>orf20.6</i> | 1824 | ABC transporter related protein, ATP-binding domain |
| <i>orf20.7</i> | 1881 | ABC transporter related protein, ATP-binding domain |
| <i>orf20.8</i> | 1371 | Lysine/ornithine N-monooxygenase |

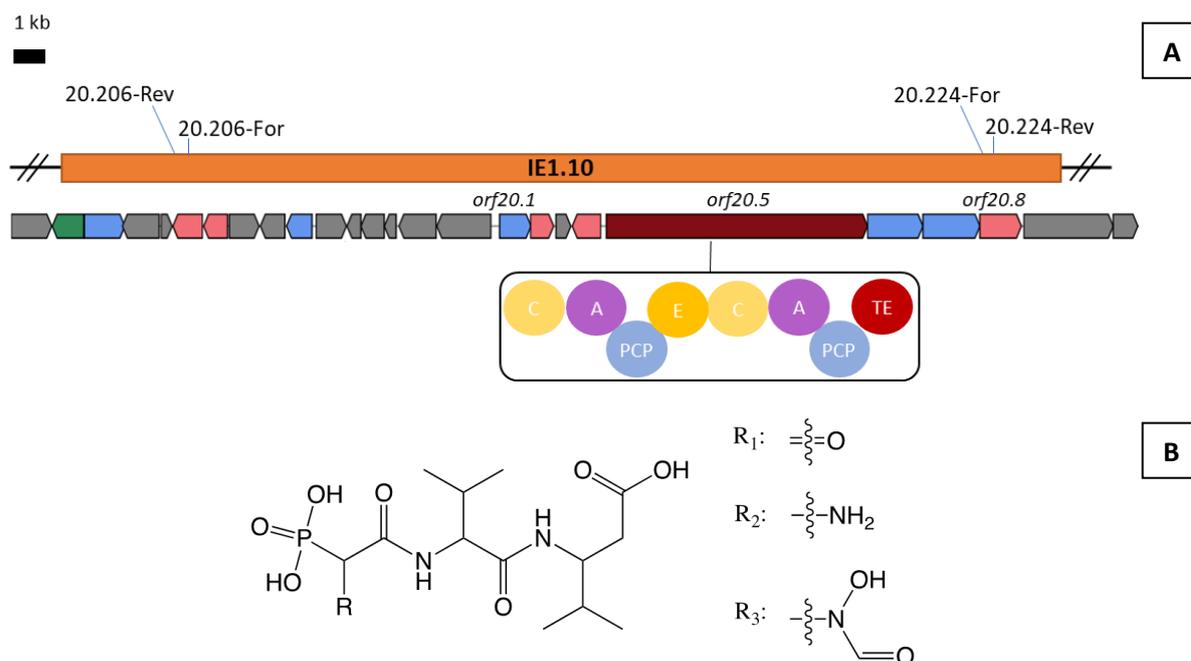


Figure III.10: Cluster 20. Cluster 20 is a 17.2 kb NRPS cluster that consists of eight *orfs*. Three *orfs* encode for transport related protein, four *orfs* for tailoring enzymes, and one *orf* for two NRPS modules. 11.8 kb upstream of *orf20.1*, two additional biosynthetic genes were annotated as methyltransferase and enoyl-CoA hydratase genes. Cluster 20 is carried by colony IE1.10, which was screened from the *Gordonia* library using the primer pair 20.206-For/20.206-Rev and 20.224-For/20.224-Rev. **B**: Core structure prediction of cluster 20 product.

III.3.1.1.5 Cluster 9

Cluster 9 is annotated by antiSMASH as a huge NRPS cluster with the size of 169,165 kb. However, on a closer look, it can be predicted that the annotated cluster 9 actually consists of three NRPS clusters, each with their own TE domain (**Figure III.11**).

The first cluster, cluster 9A, is the smallest cluster among them with the size of 22,155 bp (**Table III.8**). This cluster consists of an NRPS cluster with four modules. The first A domain has 100% Stachelhaus code match to serine. The second, third, and fourth A domains has 70% similarity to each other, with Stachelhaus code match to leucine/tyrosine (70%/80%, 60%/80%, 70%/60%, respectively). The involvement of *orf9A.1* to *orf9A.7* in the biosynthesis of cluster 9A product are inconclusive, due to the predicted protein function and the distant of *orf9A.2* to the main NRPS cluster. Cluster 9A has 67% similarity when compared to icosalide A and icosalide B BGC (Jenner *et al.*, 2019). The BGC of icosalide A also consists of

four A domains, but with five C domains. The fourth C domain in icosalide A BGC that is responsible for the addition of a lipid chain is missing from cluster 9A.

The second cluster, 9B, is a 37,670 bp gene cluster that encodes for nine NRPS modules (**Table III.8**). The amino acid sequence of the substrate binding pocket of the A domains in this cluster have 70-80% match to predicted Stachelhaus code, except for the first A domain that has 60% match to threonine, and the sixth and eighth A domain that has 100% match to threonine and serine, respectively. Although cluster 9B was analysed by antiSMASH to have 72% region to region similarity to sevadicin, the number of modules on both BGC differs a lot. Sevadicin BGC consists of only three modules, with A domain specificity to phenylalanine, alanine, and tyrosine (Garcia-Gonzalez *et al.*, 2014). Similar to sevadicin BGC, cluster 9B does not have tailoring enzymes and the modules consists of the minimal NRPS domains, C-A-PCP, with two epimerization domains.

The last cluster, 9C, is a 63,650 bp gene cluster that encodes for seventeen NRPS modules (**Table III.8**). In comparison to that, the largest NRPS cluster that was reported from the Atlas database is plu2670 from *Photorhabdus luminescens*, which comprises of 15 modules (Wang *et al.*, 2014). The A domain specificity for this cluster was also predicted by antiSMASH and by manual comparison of the Stachelhaus code. If the match is above 60%, it will be annotated in **Table III.8** according to the antiSMASH prediction. If not, two highest match will be mentioned, as in the case for the second, eighth, twelfth, fourteenth, and fifteenth A domain. Since the gene direction was going from *orf9C.5* to *orf9C.1*, the incorporation of the monomers is starting with the A domains in *orf9C.5*, and continued by the A domains in *orf9C.3*, which give the predicted core scaffold of the product to be the peptide: serine-threonine-ornithine-serine/glutamate-serine-phenylalanine-serine-serine-serine-phenylalanine-leucine-serine-ornithine-serine-serine/glutamate-leucine-serine/glutamate-valine/N-(1,1-dimethyl-1-allyl) tryptophan.

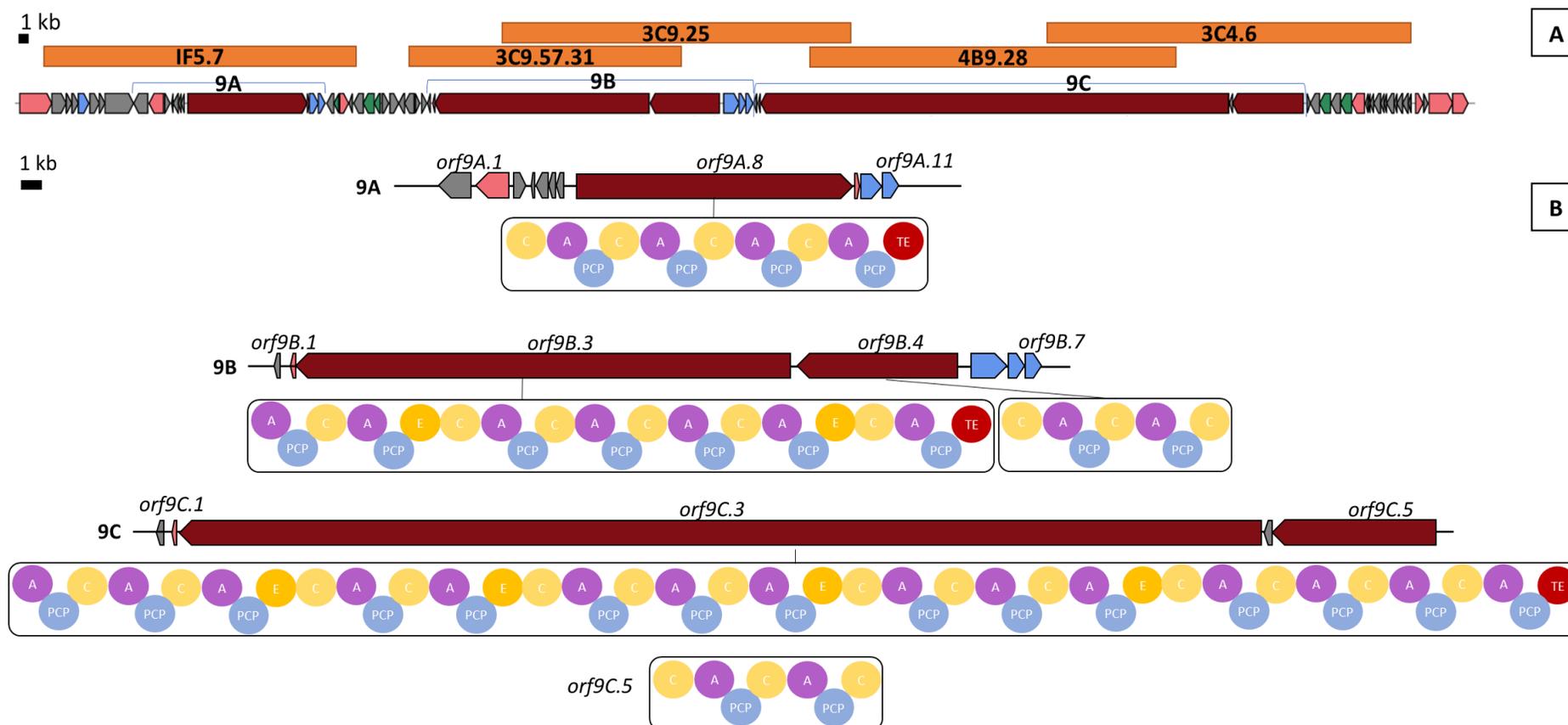


Figure III.11: Cluster 9. A: Cluster 9 consists of three NRPS clusters, named cluster 9A, 9B and 9C. These clusters are covered by the inserts of five fosmids from the *Gordonia* library (in orange). **B:** Cluster 9A, 9B and 9C are considered as separate BGC due to their possession of a thioesterase domain at the end of each NRPS assembly line. In dark red are the core NRPS genes, in pink are additional biosynthetic genes, in blue are transporter genes, and in grey are genes that might not be involved in the biosynthesis of the product, or genes with unknown function. Cluster 9A only has one core NRPS gene that encode for four modules, while cluster 9B and 9C have two core NRPS genes each, that encode for nine and seventeen modules each. All three clusters have *orfs* encoding for an mbtH-like protein (*i.e.* *orf9A.9*, *orf9B.2*, and *orf9C.2*).

Table III.8 Predicted function of orfs in cluster 9.

| Gene | Size (bp) | Protein function prediction |
|-----------------|-----------|--|
| <i>orf9A.1</i> | 1626 | Domain of unknown function |
| <i>orf9A.2</i> | 1677 | Pyruvate oxidase/decarboxylase |
| <i>orf9A.3</i> | 603 | Domain of unknown function |
| <i>orf9A.4</i> | 186 | Ribosomal protein S14p/S29e |
| <i>orf9A.5</i> | 612 | Ribosomal protein L5 |
| <i>orf9A.6</i> | 318 | Ribosomal protein 50S L24 |
| <i>orf9A.7</i> | 369 | Ribosomal protein L14p/L23e |
| <i>orf9A.8</i> | 13821 | NRPS domains (monomer prediction: serine-leucine/tyrosine- leucine/tyrosine-leucine/tyrosine) |
| <i>orf9A.9</i> | 240 | MbtH-like protein |
| <i>orf9A.10</i> | 1038 | ABC transporter, ATP-binding protein |
| <i>orf9A.11</i> | 780 | ABC-2 type transporter |
| <i>orf9B.1</i> | 309 | Hypothetical protein |
| <i>orf9B.2</i> | 258 | MbtH-like protein |
| <i>orf9B.3</i> | 24777 | NRPS domains (threonine-phenylalanine-serine-serine-phenylalanine-threonine-leucine) |
| <i>orf9B.4</i> | 8022 | NRPS domains (serine-threonine) |
| <i>orf9B.5</i> | 1776 | ABC transporter, ATP-binding protein |
| <i>orf9B.6</i> | 795 | ABC-2 type transporter |
| <i>orf9B.7</i> | 828 | ABC-2 type transporter |
| <i>orf9C.1</i> | 357 | Hypothetical protein |
| <i>orf9C.2</i> | 234 | MbtH-like protein |
| <i>orf9C.3</i> | 54159 | NRPS domains (ornithine-serine/glutamate-serine-phenylalanine-serine-serine-serine-phenylalanine-leucine-serine-ornithine-serine- serine/glutamate-leucine- serine/glutamate-valine/N-(1,1-dimethyl-1-allyl)Tryptophan |
| <i>orf9C.4</i> | 390 | Hypothetical protein |
| <i>orf9C.5</i> | 8181 | NRPS domains (serine-threonine) |

The *in silico* analysis of the *Gordonia* sp. Bu15_44 also yielded three putative natural product BGCs, which are part of another study and will not be discussed here. The putative BGCs in the *Gordonia* strains potentially encode for new natural products. To tap this potential, an artificial 'metagenomic' library from the genomic DNA of *G. terrae* Bu15_45 and *Gordonia* sp. Bu15_44 was prepared. This library would serve as the DNA source for TAR cloning to capture the BGC into the integrative vector for heterologous expression in *Streptomyces* host strains.

III.3.1.2 *Gordonia* library construction and screening

The construction of an artificial ‘metagenomic’ library of *Gordonia* was preceded by isolation of genomic DNA from the two species, *Gordonia* sp. Bu15_44 and *Gordonia terrae* Bu15_45. The isolated DNA had to be randomly sheared to the size of around 40 kb in a high concentration. To estimate the size and concentration of the isolated DNA, it was run in an agarose gel electrophoresis and the size and concentration was compared to a fosmid control DNA that has a known size of 40 kb and known concentration of 100 ng/ μ L (**Figure III.12**). The isolated DNA from *Gordonia* sp. Bu15_44 was estimated to be one-tenth of the fosmid control DNA, which equalled to 10 ng/ μ L, while the isolated DNA from *Gordonia* sp. Bu15_45 was estimated to be equal to the fosmid control DNA, which was 100 ng/ μ L.

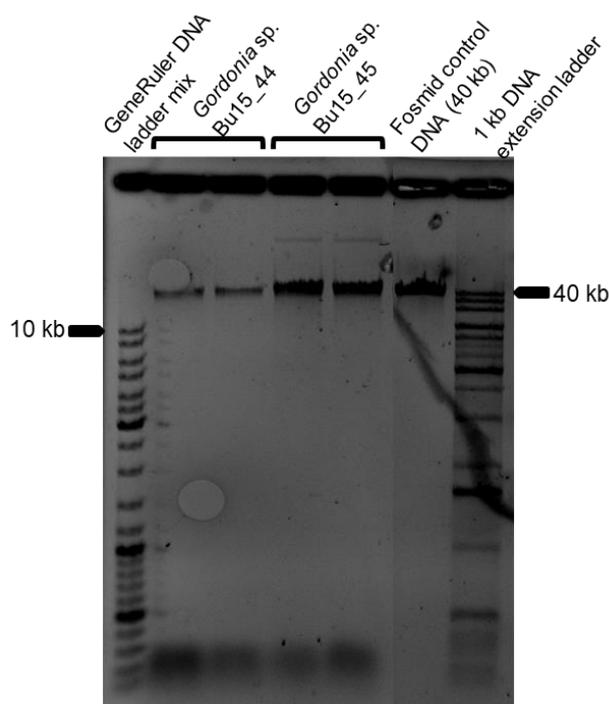


Figure III.12: Genomic DNA from *Gordonia* sp. Bu15_44 and *Gordonia terrae* Bu15_45. Isolated genomic DNA was sheared to an expected size of 40 kb. 100 ng/ μ L fosmid control DNA (40 kb) was used to approximate isolated DNA size and concentration.

The isolated DNA was then prepared to be ligated to the pCC2FOS fosmid. Firstly, isolated genomic DNA was run on an low-melting point (LMP) agarose gel to separate the high molecular weight DNA and the sheared DNA. Then, the genomic DNA with the desired size of 40 kb was purified from the gel slice by digesting the gel with β -Agarase I and precipitation of DNA by sodium acetate and ethanol.

Subsequently, it was blunt-ended by T4 DNA polymerase and T4 polynucleotide kinase. **Figure III.13** shows the analytical gel after blunt-ending of the DNA. It shows that the blunt ended DNA had a good quality and necessary length of 40 kb, with a concentration of approximately around 10 ng/ μ L.

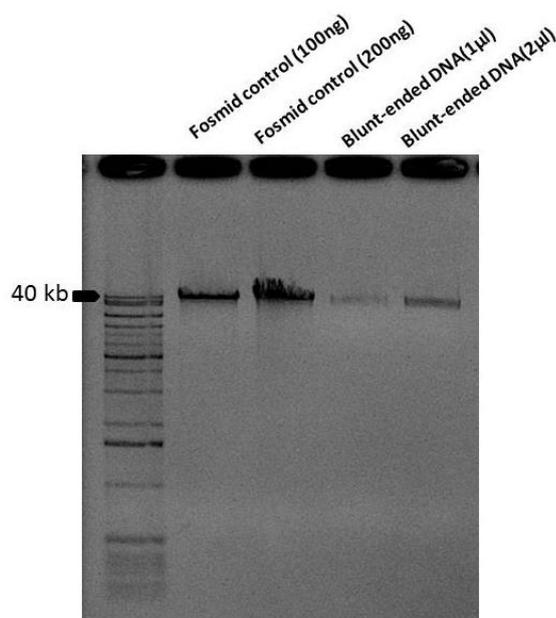


Figure III.13: Analytical gel of blunt-ended DNA. A mixture of genomic DNA from *Gordonia* sp. Bu15_44 and *Gordonia terrae* Bu15_45 was blunt-ended and run on electrophoresis gel. 1 μ L and 2 μ L fosmid DNA control (100 ng/ μ L) was used to approximate the DNA size and concentration. The band intensity of the blunt-ended DNA was estimated to be one-tenth concentration of the fosmid DNA.

The blunt-ended DNA was ligated to pCC2FOS fosmid and introduced into *E. coli* EPI300-T1^R by transduction using lambda phage. Since each of the *Gordonia* genome has a size of 6 Mbp, and 20-fold coverage is needed to ensure that the complete sequence is covered in the artificial metagenomic library, the library should contain DNA inserts with a total size of 240 Mbp. Because each fosmid clone carries 40 kb of DNA insert, a minimum of 6000 colonies had to be contained in the library (**Figure III.14**).

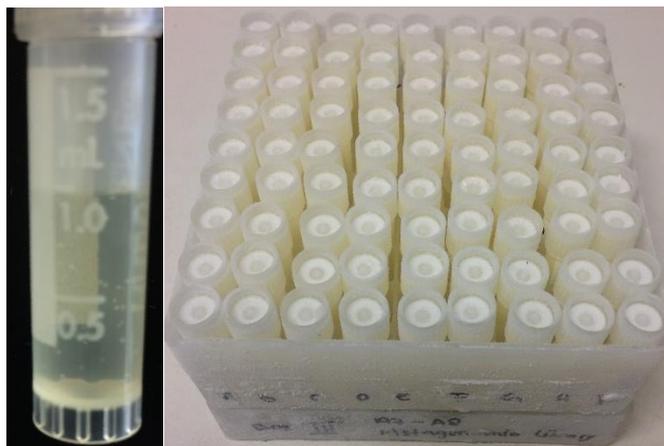


Figure III.14: Artificial metagenomic *Gordonia* library. Left: Single vial from *Gordonia* library containing *E. coli* EPI300-T1^R colonies with packaged fosmid that were growing spatially. Right: Artificial metagenomic *Gordonia* library consists of 217 vials with 27-28 colonies/vial and 23 vials of row pools.

The artificial metagenomic library was then ready to be screened for any targeted BGC (**Table III.4**). Here, the PCR screening approach for cluster 52 is elaborated as the representation of the screening approach for other BGCs from the artificial metagenomic library. First of all, the 23 row pools were screened using two primer pairs that target each end of the BGC (**Figure III.7**), *i.e.* 118-For/118-Rev and 129-For/129-Rev, that give a 270 bp and a 558 bp amplified product, respectively (**Figure III.15**). Row pool 1C, 3D, 3E, 4A, 4B, and 4DE showed positive PCR result for both targets. The screening was continued by checking each vial in row 1C. Vial 1C8 that gave the expected amplicons was plated on LB-chloramphenicol plates, so that single colonies could be picked and screened for the presence of the target BGC.

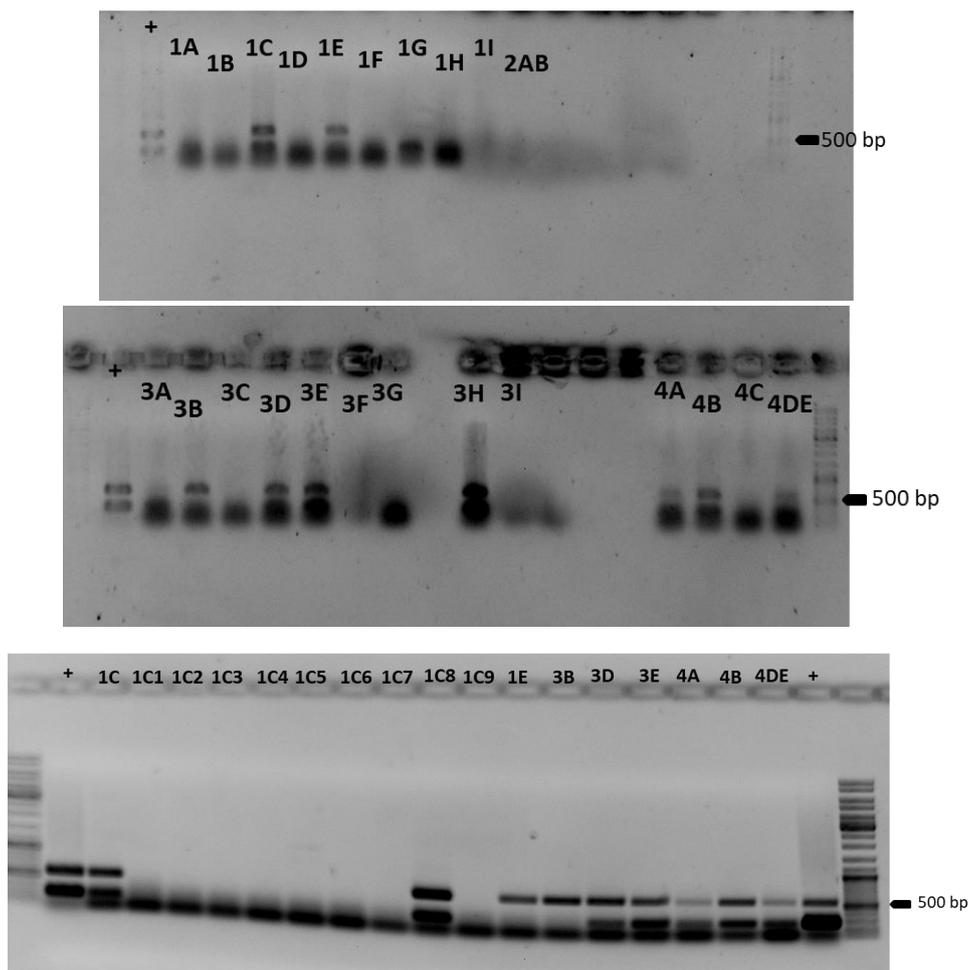


Figure III.15: Row pool and vial screening for cluster 52. *Upper pictures* show the screening result of 23 row pools of the *Gordonia* library. Row pool 1C and 4B have both target bands with the size of 270 bp and 558 bp. *Lower picture* shows the screening result of the row 1C, which contains 9 vials, and the repetition of some row pools that did not show clear result in the previous screening round. Vial 1C8 has both target bands. *Gordonia terrae* Bu15_45 genomic DNA was used as positive control and GeneRuler DNA ladder Mix was used as DNA size marker. Gel electrophoresis was run in 1% agarose gel at 90 V for 40 minutes.

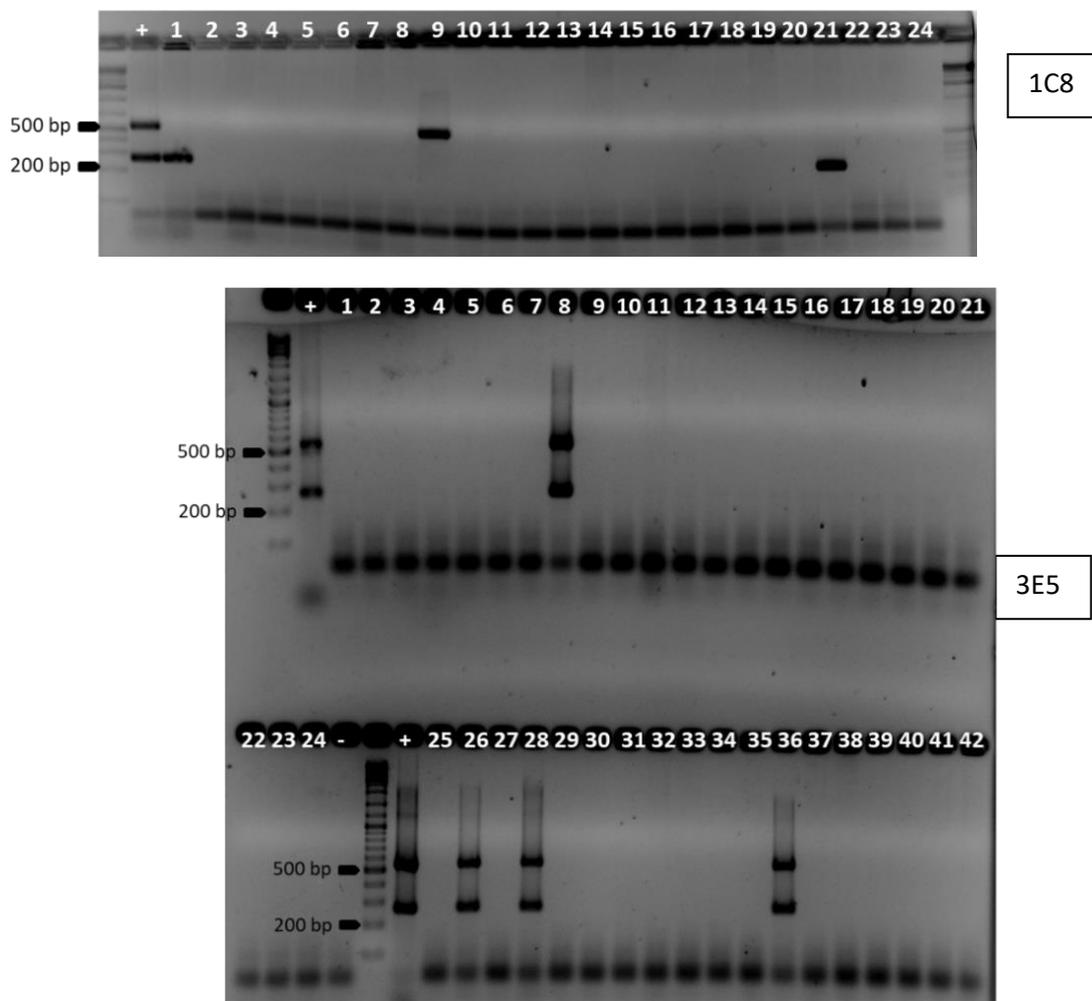


Figure III.16: PCR result for single colonies screening. *Upper picture* shows colony screening result from vial 1C8. *Lower picture* shows colony screening result from vial 3E5. Out of 48 colonies that were picked, four colonies had both target amplicons with the size of 270 and 558 bp. *Gordonia terrae* Bu15_45 genomic DNA was used as positive control. Gel electrophoresis was run in 3% agarose gel at 100 V for 50 minutes.

Colonies from vial 1C8 were checked by PCR to find a single colony that gives amplicons for 118-For/118-Rev and 129-For/129-Rev, which means that the colony carries the whole sequence of cluster 52. **Figure III.16** shows the screening result of 24 out of the 87 checked colonies from vial 1C8. No single colony gave amplicons for both primer pairs. The colonies only gave positive band for one primer pair, e.g. colony 1 and 21 only gave amplicon for 118-For/118-Rev, while colony 9 gave amplicon for 129-For/129-Rev only. This result shows that the checked colonies carried only a part of cluster 52, not the full length. Therefore, the vials in another row that had shown both amplicons in the row pool screening, i.e. row 3E, was

checked. Vial 3E5 showed both target amplicons and was plated on LB-chloramphenicol plate for single colony screening. Out of 48 colonies, target amplicons of both primer pairs (118-For/118-Rev and 129-For/129-Rev) could be observed from 4 colonies (**Figure III.16**).

Table III.9 Primer pairs for BGCs screening from *Gordonia* artificial metagenomic library and colonies that carry the BGCs.

| Target Cluster | Primer Name | Colony |
|-------------------|---------------------------|-----------|
| 52 | 118-For and 118-Rev | 3E5.8 |
| | 129-For and 129-Rev | |
| 23 | 23.127-For and 23.127-Rev | 1E5.55 |
| | 23.134-For and 23.134-Rev | |
| | 23.134-For and 23.134-Rev | 1E6.12 |
| | 23.144-For and 23.144-Rev | |
| 37 | 37.231-For and 37.231-Rev | 3C6.20.9 |
| | 37.243-For and 37.243-Rev | |
| | 37.243-For and 37.243-Rev | 3C1.18 |
| | 37.252-For and 37.252-Rev | |
| 20 | 20.206-For and 20.206-Rev | 1E1.19 |
| | 20.224-For and 20.224-Rev | |
| 9 | 9A215F and 9A215R | 1F5.7 |
| | 9A227F and 9A227R | |
| | 9B239F and 9B239R | 3C9.57.31 |
| | 9B242F and 9B242R | |
| | 9C242F and 9C242R | 3C9.25 |
| | 9CD247F2 and 9CD247R2 | |
| | 9CD247F2 and 9CD247R2 | 4B9.28 |
| | 9CD247F3 and 9CD247R3 | |
| | 9CD247F3 and 9CD247R3 | 3C4.6 |
| 9D255F and 9D255R | | |

Similar to the screening process of cluster 52, the artificial metagenomic *Gordonia* library was also screened by PCR for four other putative BGCs that were identified by AntiSMASH, and the colonies that carry the BGCs are listed in **Table III.9**. Clusters that are larger than 30 kb could not be found to be located on one fosmid only. Cluster 23 (40,971 bp) was found in one fosmid except for the last gene, *orf23.18*. Cluster 37 (49,656 bp) is covered by the inserts of two fosmids, while cluster 9 (135,426 bp) is covered by the inserts of 5 fosmids.

III.3.1.3 TAR cloning and triparental conjugation

The fosmid from colony 3E5.8 was isolated, yielding a concentration of 1196.6 µg/mL, and sent for sequencing. The sequencing result revealed a 33,679 bp insert, spanning from five genes upstream of *nbtT* to four genes downstream of *nbtE* (**Figure III.7**). TAR cloning was performed to capture cluster 52 (from the transporter gene upstream of *nbtT* to *nbtE*). The correct pCAP03-cl.52 was confirmed by PCR and restriction pattern (**Figure III.17**), and was maintained in *E. coli* TOP10. Then, pCAP03-cl.52 was introduced to *S. coelicolor* M1146 and *S. lividans* TK24 by triparental conjugation, and the integration of cluster 52 into genomic DNA of the expression host was confirmed by doing genomic DNA isolation and PCR check (**Figure III.17**). Thereafter, the positive colony was grown on MS-agar medium, and stored as a spore suspension (*S. coelicolor*-cl.52).

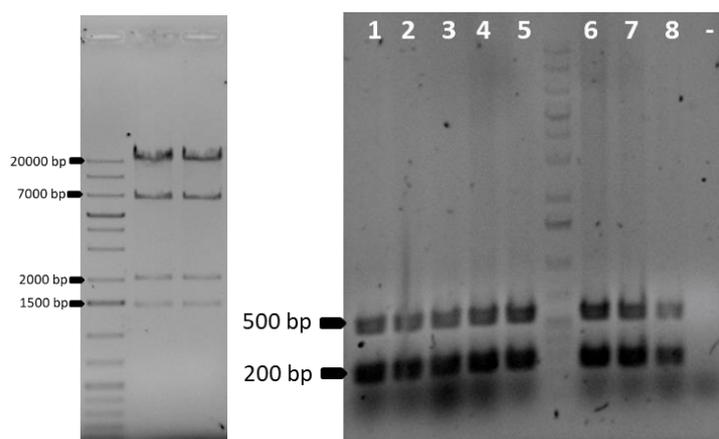


Figure III.17: TAR cloning of pCAP03-cl.52 and its integration to *Streptomyces* expression host. **Left picture** shows the restriction pattern of pCAP03-cl.52 to check the success of TAR cloning. Two plasmids that gave positive PCR result after TAR cloning were isolated and digested by *EcoRV*. Both plasmids showed the correct restriction pattern, with fragment sizes of 27085 bp, 6691 bp, 2020 bp, and 1443 bp. **Right picture** shows PCR test result with the primer pair 118-For/118-Rev (270 bp) and 129-For/129-Rev (558 bp) to check the successful integration of cluster 52 into the *Streptomyces* expression host. GeneRuler 1 kb plus DNA ladder was used as DNA size marker, and genomic DNA of wild type *S. coelicolor* M1146 was used as the negative control for the test PCR.

TAR cloning has also been performed for cluster 23 and cluster 20. Similar to the work for cluster 52, the capture of the clusters from pCC2FOS fosmid into the

pCAP03 was done in *S. cerevisiae* VL6-48N. The yeast cells that were carrying pCAP03 with DNA insert would be able to grow in selective medium with 5-Fluoroorotic acid (5-FOA), while the ones without DNA insert or that carried re-ligating vector would not grow in the presence of 5-FOA. The DNA insert was corroborated by PCR and restriction pattern. The correct heterologous expression vectors were isolated from the yeast cells, and introduced to heterologous expression hosts (*S. coelicolor* M1146 and *S. lividans* TK24) by triparental conjugation. The integration of the clusters were checked by PCR.

III.3.1.4 Introduction of promoters in heterologous expression hosts

To overexpress the BGC, constitutive promoter (P_{ermE} and P_{ermE^*}) and synthetic tetracycline-inducible promoter (P_{tcp830}) were introduced in front of the BGCs by λ -Red recombination. For cluster 52, three constructs were made by integrating different promoters at different locations. Construct 52.Rec3 has P_{tcp830} in front of *orf118* and construct 52.Rec4 has the same promoter in front of *nbtS*. Meanwhile, construct 52.Rec6 has P_{ermE} upstream of *nbtS* (**Figure III.18**). In construct 52.Rec4 and 52.Rec6, *orf118* and *nbtT* were deleted and later cloned into pGM1202. After the addition of the promoter in *E. coli* BW25113, the constructs were integrated into the *Streptomyces* heterologous host genome and corroborated by PCR. The primer pair tcp830-unitest/118-Rev was used to check 52.Rec3, tcp830-unitest/52.120-R for 52.Rec4, and ermE-unitest/52.120-R for 52.Rec6, giving expected amplicons with the size of 1030 bp, 742 bp, and 797 bp, respectively (**Figure III.19**).

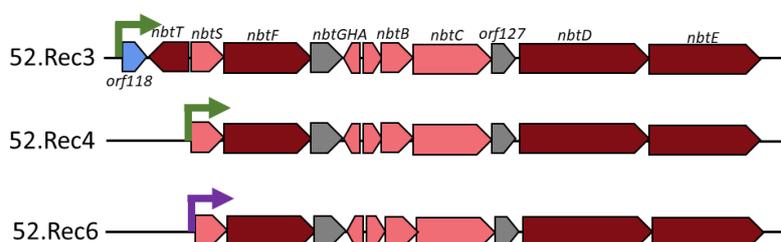


Figure III.18: Construct for overexpression of cluster 52. P_{tcp830} (green arrow) was inserted upstream *orf118* and *nbtS* in 52.Rec3 and 52.Rec4, respectively. P_{ermE} (purple arrow) was inserted upstream *nbtS* in 52.Rec6.

Two additional constructs were generated to carry the genes that were deleted in 52.Rec4 and 52.Rec6, as well as *nbtH* which has different orientation than the other genes in cluster 52. These two constructs were cloned into pGM1202 under a constitutive promoter P_{ermE^*} by isothermal assembly. The first construct contained *orf118* and *nbtT* (pGM1202-C1), and the second construct carried *orf118*, *nbtT*, and *nbtH* (pGM1202-C2). PCR was performed using the primer pair pGM-seq/pGM-Rseq to check that the constructs were successfully transformed into the *S. lividans* TK24 heterologous host. The expected amplicons from pGM1202-C1 was 3162 bp, and from pGM1202-C2 was 3820 bp (**Figure III.19**).

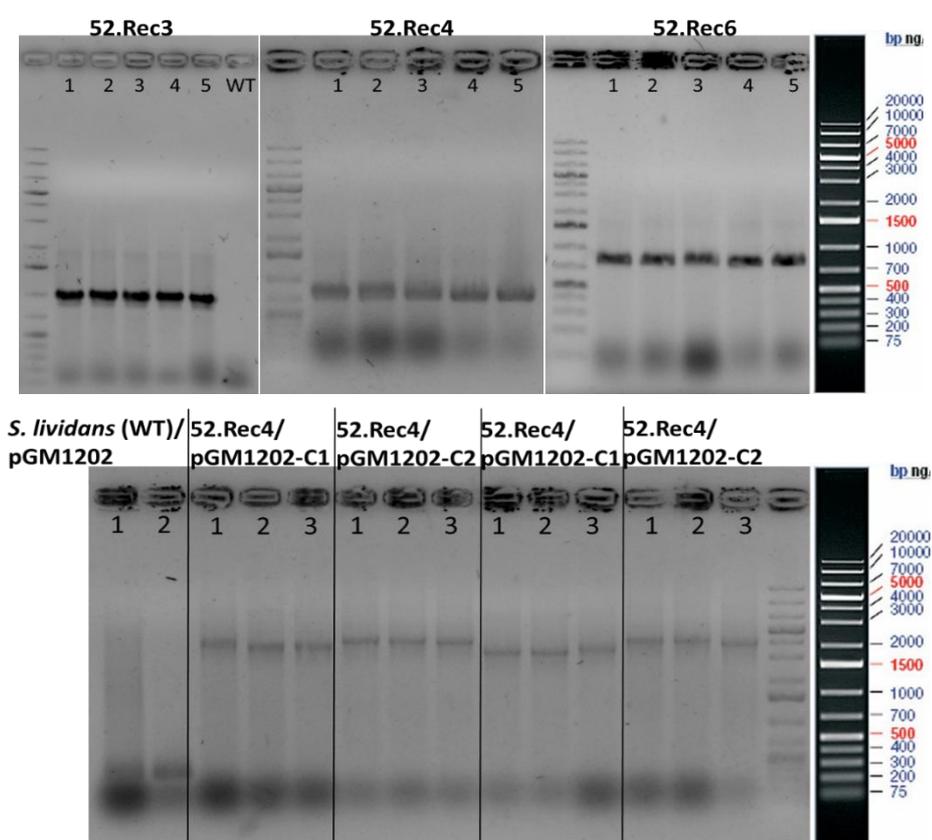


Figure III.19: PCR results to check the integration of overexpression construct and introduction of additional plasmid into *Streptomyces* heterologous host. *Upper picture* shows PCR result from isolated DNA of *S. coelicolor* M1146 and *S. lividans* TK24. 52.Rec3 was integrated into *S. coelicolor* M1146, while 52.Rec4 and 52.Rec6 was integrated into *S. lividans* TK24. Five colonies were checked for each construct, and all gave the expected amplicons, i.e. 1030 bp for 52.Rec3, 742 bp for 52.Rec4, and 797 bp for 52.Rec6. *Lower picture* showed the colony PCR result to check the transformation of 52.Rec4 and 52.Rec6 by pGM1202-C1 and pGM1202-C2. Three colonies were checked for each transformation, and all gave expected amplicons, i.e. 3162 bp for pGM1202-C1 and 3820 bp for pGM1202-C2. As negative control, wild type *S. lividans* TK24 was transformed with empty pGM1202, and gave amplicons of 365 bp in the colony PCR check. Gene Ruler 1 kb plus DNA ladder was used as DNA size marker.

III.3.1.5 Heterologous expression result

In a next step, it was planned to analyse the effect of the integration of the respective BGC into the heterologous host(s). Therefore, the transgenic *Streptomyces* expression host was cultivated in different media and conditions. Thereafter, the organic extract was evaluated by LC-MS analysis. First of all, cluster 52 that has 100% similarity to a reported nocobactin NA BGC was studied as proof of principle of the whole cloning and expression method.

As a positive control for cluster 52 heterologous expression, nocobactin NA production by *Gordonia terrae* Bu15_45 in minimal medium was investigated. From LC-MS analysis of the ethyl acetate and methanol crude extract, nocobactin NA-a ($C_{38}H_{57}N_5O_{10}$) and nocobactin NA-b ($C_{40}H_{61}N_5O_{10}$) were detected as their ferric complexes, with m/z 797.3293 $[M+H]^+$ and m/z 825.3606 $[M+H]^+$, respectively (**Figure III.20**). The peak for nocobactin NA-a appeared at 42.9 minutes, while the peak for nocobactin NA-b appeared at 44.6 minutes. Beside their proton adduct, the sodium adduct for both molecules could also be detected at the same retention time.

Likewise, *S. coelicolor*-cl.52 was also cultivated in minimal medium, extracted by ethyl acetate and methanol, and analyzed by LCMS. However, no mass for nocobactin NA could be observed. Heterologous expression using other media, *i.e.* ISP2 and TSB, with different incubation period (3 and 10 days) also did not result in production of nocobactin NA.

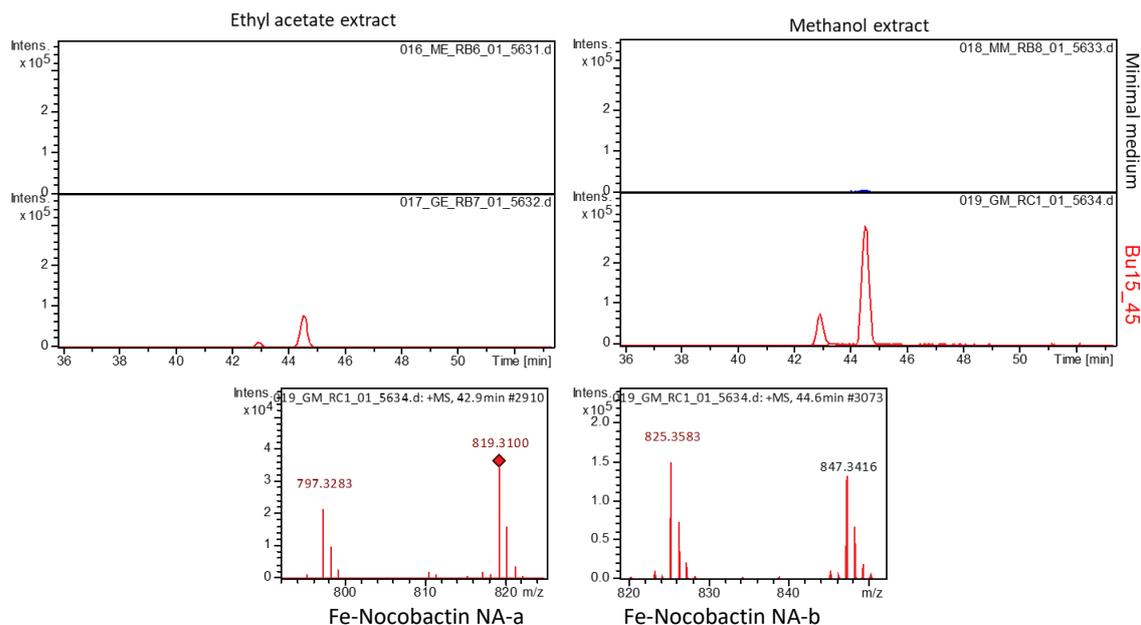


Figure III.20: LCMS analysis of nocobactin NA production by *G. terrae* Bu15_45. Ferri-nocobactin NA-a (m/z 797.3283 $[M+H]^+$, 819.3100 $[M+Na]^+$; at 42.9 minutes) and ferri-nocobactin NA-b (m/z 825.3583 $[M+H]^+$, 847.3416 $[M+Na]^+$; at 44.6 minutes) were detected in both ethyl acetate and methanol extract of *G. terrae* Bu15_45. These masses could not be found in the medium control.

Therefore, in a next step the synthetic promoter *tcp830* was inserted into the expression constructs to overexpress the BGC. After induction with anhydrotetracycline, at different growth periods, still no nocobactin mass could be detected by LCMS analysis of the ethyl acetate and methanol crude extracts. Next, cluster 52 was introduced to another expression host, *i.e.* *S. lividans* TK24, and another promoter, P_{ermE} , was inserted in front of *nbtS*. During the insertion of the P_{ermE} promoter, the two genes in front of *nbtS* were deleted, and therefore pGM1202-C1 and pGM1202-C2 were introduced to *S. lividans* TK24 with integrated cluster 52.Rec4 and 52.Rec6. Despite all attempts with different constructs, conditions, medium, and heterologous host (**Table III.10, Figure S1-S3**), the heterologous expression of cluster 52 to produce nocobactin NA was still unsuccessful.

Table III.10: Different attempts of heterologous expression of cluster 52 to produce nocobactin NA

| Heterologous host | Construct | Pre-induction incubation | Total cultivation period | Medium |
|----------------------------|---------------------|--------------------------|--------------------------|----------------|
| <i>S. coelicolor</i> M1146 | Cluster 52 | - | 3 days | Minimal medium |
| <i>S. coelicolor</i> M1146 | Cluster 52 | - | 7 days | Minimal medium |
| <i>S. coelicolor</i> M1146 | Cluster 52 | - | 3 days | TSB |
| <i>S. coelicolor</i> M1146 | Cluster 52 | - | 3 days | ISP2 |
| <i>S. coelicolor</i> M1146 | Cluster 52 | - | 10 days | ISP2 |
| <i>S. coelicolor</i> M1146 | 52.Rec3 | 30 hours | 4.5 days | ISP2 |
| <i>S. coelicolor</i> M1146 | 52.Rec3 | 3 days | 4.5 days | ISP2 |
| <i>S. coelicolor</i> M1146 | 52.Rec3 | 7 days | 8 days | ISP2 |
| <i>S. coelicolor</i> M1146 | 52.Rec3 | 3 days | 14 days | ISP2 |
| <i>S. lividans</i> TK24 | 52.Rec4 | 5 days | 8 days | ISP2 |
| <i>S. lividans</i> TK24 | 52.Rec6 | - | 8 days | ISP2 |
| <i>S. lividans</i> TK24 | 52.Rec4/ pGM1202-C1 | 1 day | 8 days | ISP2 |
| <i>S. lividans</i> TK24 | 52.Rec4/ pGM1202-C2 | 1 day | 8 days | ISP2 |
| <i>S. lividans</i> TK24 | 52.Rec6/ pGM1202-C1 | - | 8 days | ISP2 |
| <i>S. lividans</i> TK24 | 52.Rec6/ pGM1202-C2 | - | 8 days | ISP2 |
| <i>S. lividans</i> TK24 | 52.Rec4/ pGM1202-C1 | 4 days | 7 days | ISP2 |
| <i>S. lividans</i> TK24 | 52.Rec4/ pGM1202-C2 | 4 days | 7 days | ISP2 |
| <i>S. lividans</i> TK24 | 52.Rec6/ pGM1202-C1 | - | 7 days | ISP2 |
| <i>S. lividans</i> TK24 | 52.Rec6/ pGM1202-C2 | - | 7 days | ISP2 |
| <i>S. coelicolor</i> M1146 | 52.Rec4/ pGM1202-C1 | 4 days | 7 days | ISP2 |
| <i>S. coelicolor</i> M1146 | 52.Rec6/ pGM1202-C1 | - | 7 days | ISP2 |
| <i>S. coelicolor</i> M1146 | 52.Rec4/ pGM1202-C2 | 4 days | 7 days | ISP2 |
| <i>S. coelicolor</i> M1146 | 52.Rec6/ pGM1202-C2 | - | 7 days | ISP2 |
| <i>S. lividans</i> TK24 | Cluster 52 | - | 7 days | Minimal medium |

III.3.2. Burying beetle project

The study of the gut microbiome of the burying beetle *Nicrophorus vespilloides* revealed bacterial symbionts that showed antimicrobial activity against bacterial pathogens (Heise *et al.*, 2019). 16S rDNA analysis of strain 39 that was previously reported to possess activity against *E. coli* and *M. smegmatis*, showed 98% identity to *Arthrobacter* sp. The genomic DNA of this strain was sequenced, and the whole genome DNA sequence has the highest homology to *Glutamicibacter arilaitensis* (88.47% identity with 12.8% coverage) (Heise, 2019).

III.3.2.1 *In silico* analysis of BGC in strain 39

The whole genome sequence (3.96 Mb) of strain 39 was analyzed by antiSMASH, and only one NRPS/PKS cluster could be detected, which was named cluster 39. Region-to-region similarity to MIBiG cluster showed 61% similarity to myxochromide S1 BGC from *Stigmatella aurantiaca*. The myxochromide S1 BGC consists of one iterative PKS module and six NRPS modules, where the fourth module was predicted to be not involved in the biosynthesis of myxochromide S1. In comparison to that, cluster 39 has one PKS module and three NRPS modules, with additional biosynthetic genes (**Figure III.21**, **Table III.11**).

The closest Stachelhaus code match for the A domain in Orf39.11 is for phenylalanine; however, only by 50%. Next, the AT domain in Orf39.12 was predicted to specifically incorporate malonyl-CoA based on its AT signature (87.5%) and minowa score (96.5). Meanwhile, the A domain in Orf39.13 has highest Stachelhaus code match to histidine (70%), and the one in Orf39.18 was predicted to activate N-(1,1-dimethyl-1-allyl) tryptophan (70%). The core structure for cluster 39 product was predicted only according to the monomer predictions for the three A domains and one AT domain.

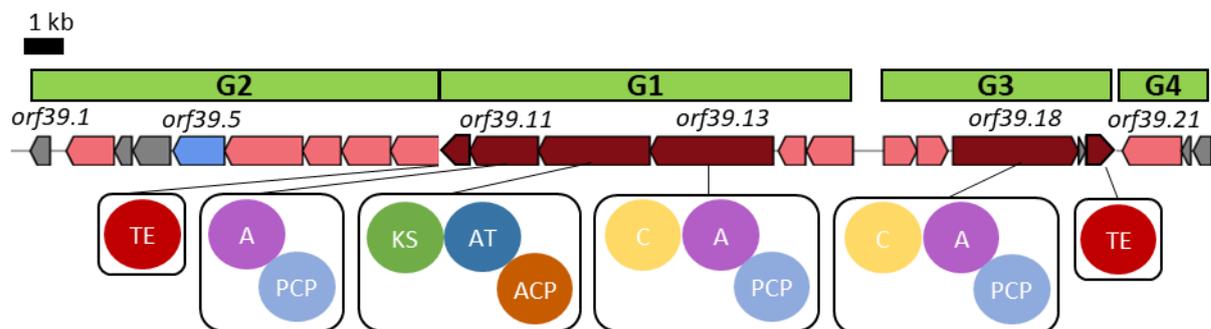


Figure III.21: Cluster 39. Cluster 39 is a 29,937 bp NRPS/PKS cluster. This cluster contains 23 open reading frames, three of which are NRPS modules, one PKS module, and two thioesterase domains (in dark red). Ten genes were predicted as additional biosynthetic genes, which could modify the building block for PKS and NRPS modules (in pink). In addition to these, one transporter-related gene could be identified (in blue), and six other genes were predicted to be not involved in the compound biosynthesis (in grey). G1 to G4 in green blocks indicate the amplified fragments used for Gibson assembly of the cluster into the heterologous expression vector pCAP03.

The three open reading frames *orf39.2*, *orf39.8*, and *orf39.9* were annotated to encode for argininosuccinate lyase. The argininosuccinate lyase from *orf39.8* has 30% identity with 73.4% coverage to DabC. Together with DabA (cysteine synthase) and DabB (argininosuccinate lyase), they are responsible for the synthesis of 2,3-diaminobutyric acid, one of the building blocks for friulimicin biosynthesis (Müller *et al.*, 2007).

orf39.3 encodes for a glyoxalase-like domain, which is not involved in compound biosynthesis, but is part of a glyoxalase system for detoxification of methylglyoxal (He *et al.*, 2020). Proline dehydrogenase (Orf39.4) is an oxidoreductase that is involved in arginine and proline biosynthesis, and is never reported to be involved in non-ribosomal peptide nor polyketide biosynthesis. Mandelate racemase (Orf39.7) catalyzes the first step of benzoate biosynthesis to convert (R)-mandelate to (S)-mandelate, while 4'-phosphopantetheinyl transferase (Orf39.14) is an enzyme that activates apo-ACP or apo-PCP to their holo-form by transferring 4'-phosphopantetheine group from coenzyme A.

Peptidase S9 is a family of serine peptidase with diverse activity that is included in clan SC peptidase (Page and Di Cera, 2008). The peptidase S9 in Orf39.6 and in Orf39.21 has 45% and 34% identity to BelC, respectively. BelC was assigned to be

involved in belactosin biosynthesis, although its functionality has not been experimentally confirmed (Wolf *et al.*, 2017).

Meanwhile, the other tailoring enzymes like cytochrome P450 (Orf39.15), acyl-CoA dehydrogenase (Orf39.16 and Orf39.17), and N-acetyltransferase (Orf39.23) are commonly found to be involved in the biosynthesis of natural products.

Table III.11: Predicted function of orfs in cluster 39.

| Gene | Size (bp) | Protein function prediction |
|-----------------|-----------|--|
| <i>orf39.1</i> | 522 | Hypothetical protein |
| <i>orf39.2</i> | 1251 | Argininosuccinate lyase/adenylosuccinate lyase |
| <i>orf39.3</i> | 453 | Glyoxalase-like domain |
| <i>orf39.4</i> | 975 | Proline dehydrogenase |
| <i>orf39.5</i> | 1344 | Major facilitator transporter |
| <i>orf39.6</i> | 2040 | Peptidase S9 family |
| <i>orf39.7</i> | 1002 | Mandelate racemase |
| <i>orf39.8</i> | 1278 | Argininosuccinate lyase/adenylosuccinate lyase |
| <i>orf39.9</i> | 1284 | Argininosuccinate lyase/adenylosuccinate lyase |
| <i>orf39.10</i> | 741 | Thioesterase domain |
| <i>orf39.11</i> | 1791 | NRPS domains (monomer prediction: phenylalanine) |
| <i>orf39.12</i> | 2907 | PKS domains (monomer prediction : malonyl-CoA) |
| <i>orf39.13</i> | 3219 | NRPS domains (monomer prediction: histidine) |
| <i>orf39.14</i> | 711 | 4'-phosphopantetheinyl transferase superfamily |
| <i>orf39.15</i> | 1215 | Cytochrome P450 |
| <i>orf39.16</i> | 873 | Acyl-CoA dehydrogenase |
| <i>orf39.17</i> | 807 | Acyl-CoA dehydrogenase |
| <i>orf39.18</i> | 3282 | NRPS domains (monomer prediction: N-(1,1-dimethyl-1-allyl) tryptophan) |
| <i>orf39.19</i> | 189 | Hypothetical protein |
| <i>orf39.20</i> | 744 | Thioesterase domain |
| <i>orf39.21</i> | 1545 | Peptidase S9 family |
| <i>orf39.22</i> | 252 | Hypothetical protein |
| <i>orf39.23</i> | 474 | N-acetyltransferase |

III.3.2.2 Generation of heterologous expression system

Cluster 39 was cloned into the shuttle vector pCAP03 by using the Gibson assembly method. The cluster was amplified in four fragments with 25 bp overlapping regions on each ends and was rearranged so that all genes are facing the same direction

(**Figure III.21**). Fragment G1 was amplified by the primer pair: pCAP03-3603-F/3598-3597-R, G2: 3598-3597-F/3589-3604-R, G3: 3589-3604-F/3608-3611-R, G4: 3608-3611-F/3609-pCAP03-R, and the pCAP03 vector was digested with *XhoI* and *NdeI* (**Figure III.22**). The fragments G1, G2, and G3 needed to be amplified at 70°C annealing temperature to prevent unspecific amplification, while fragment G4 was amplified at 60°C annealing temperature. The Gibson assembly result was expected to be the vector pCAP03-Cluster39 that has the size of 40,346 bp.

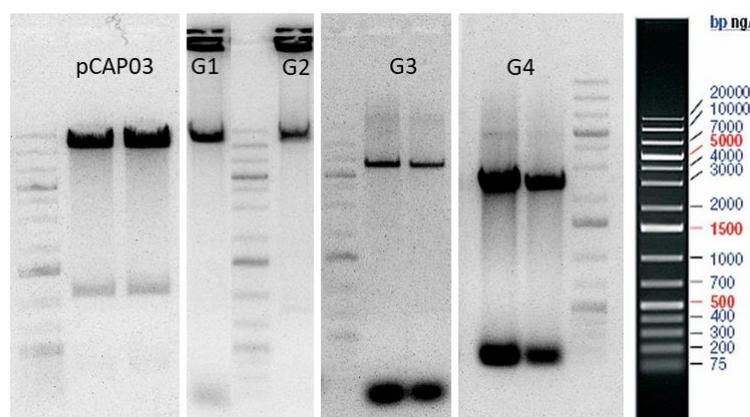


Figure III.22: Gibson assembly fragments for inserting cluster 39 into pCAP03 vector. pCAP03 was purified from maintenance host, *E. coli* TOP10 and digested with *XhoI* and *NdeI*. The expected size is 10,511 bp. Fragment G1 has the expected size of 10,813 bp, G2: 10,751 bp, G3: 6,069 bp, and G4: 2,372 bp. GeneRuler DNA ladder 1 kb plus was used to estimate the size of each fragment.

After the isothermal assembly, multiplex colony PCR was performed to check which *E. coli* TOP10 cells were carrying the construct pCAP03-Cluster39. The primers were designed to be on different fragments, so that the correct amplicons could confirm that the assembly was successful. Three bands were expected with the size of 810 bp, 1067 bp, and 706 bp. The colony with three correct bands were cultured, and the plasmid was isolated to check the restriction pattern (**Figure III.23**). Colony 18 was confirmed to carry the correct assembled plasmid pCAP03-Cluster39.

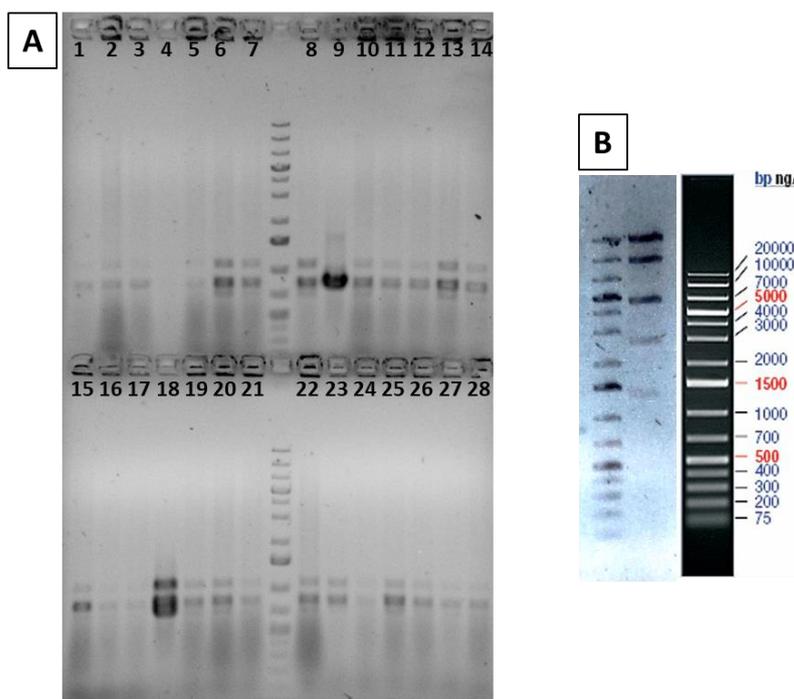


Figure III.23: Corroboration of pCAP03-cluster39 isothermal assembly result. A: Multiplex colony PCR of the correct assembled plasmid was expected to have three bands with the size of 810 bp, 1067 bp, and 706 bp. Colony 6, 13, and 18 showed three positive bands. **B:** The plasmid in colony 18 was isolated and digested with *Scal*, and gave the correct restriction pattern, with the fragments size of: 17300 bp, 9534 bp, 4813 bp, 4716 bp, 2613 bp, and 1370 bp. GeneRuler DNA ladder 1 kb plus was used as DNA size marker.

Next, the constitutive promoter (P_{ermE^*}) and synthetic tetracycline-inducible promoter (P_{tcp830}) were introduced into pCAP03-cluster39, upstream of *orf39.15* by λ -Red recombination, creating pCAP39-ea and pCAP39-ta. As a selective marker, the apramycin resistance cassette was also introduced together with the promoters. The correct recombination result was confirmed by colony PCR and restriction pattern. The PCR check was performed by using *tcp830-unitest/3603-R* for checking the insertion of P_{tcp830} and *ermE-unitest/3603* for checking the insertion of P_{ermE^*} . The restriction pattern was performed by digesting the plasmid with *XbaI*, which should give DNA fragments with the size of 15253 bp (for pCAP39-ea) / 15185 bp (for pCAP39-ta), 12030 bp, 9294 bp, and 4725 bp (**Figure III.24**). The restriction pattern of pCAP39-ea showed one additional band that is higher than the 20 kb DNA ladder, which was interpreted as the undigested plasmid.

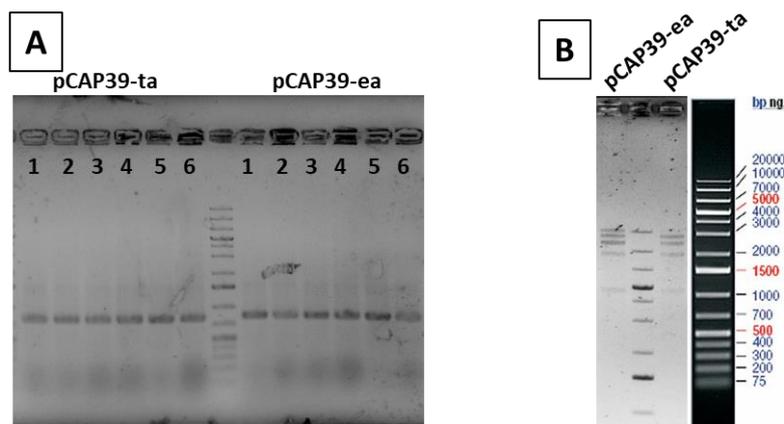


Figure III.24: λ -Red recombination for the insertion of P_{ermE^*} and P_{tcp830} into pCAP03-Cluster 39, creating pCAP39-ea and pCAP39-ta. A: Colony PCR for checking the insertion of the promoters. All colonies showed amplicons with the expected size of 837 bp. **B:** Digestion of pCAP39-ea and pCAP39-ta with *Xba*I showed the correct restriction pattern. GeneRuler DNA ladder 1 kb plus was used as DNA size marker.

pCAP03-cluster39, pCAP39-ea, and pCAP39-ta were introduced into *S. coelicolor* M1146 as the expression host by triparental conjugation. The integration of the plasmid into the *S. coelicolor* genome was confirmed by PCR from isolated genomic DNA.

III.3.2.3 Heterologous expression result

Three heterologous expression system that were generated (*S. coelicolor*/pCAP03-cluster 39, *S. coelicolor*/pCAP39-ea, *S. coelicolor*/pCAP39-ta) were inoculated into TSB medium. It was clearly visible that the control, *S. coelicolor*/pCAP03, was growing faster than the three heterologous expression hosts, which carry an integrated BGC in their genome. After four days of growth, the control and *S. coelicolor*/pCAP39-ta were induced with anhydrotetracycline. Then, the incubation was continued for another four days before lyophilization of the whole culture and extraction with ethyl acetate and methanol.

The base peak chromatogram (BPC) of the ethyl acetate and methanol crude extract of *S. coelicolor*/ pCAP03-cluster 39, *S. coelicolor*/pCAP39-ea, and *S. coelicolor*/pCAP39-ta were compared to that of the empty vector control, *i.e.* *S.*

coelicolor/pCAP03. The peaks that did not appear in the BPC of the control were investigated closer, by checking the spectra of each peak to find different masses.

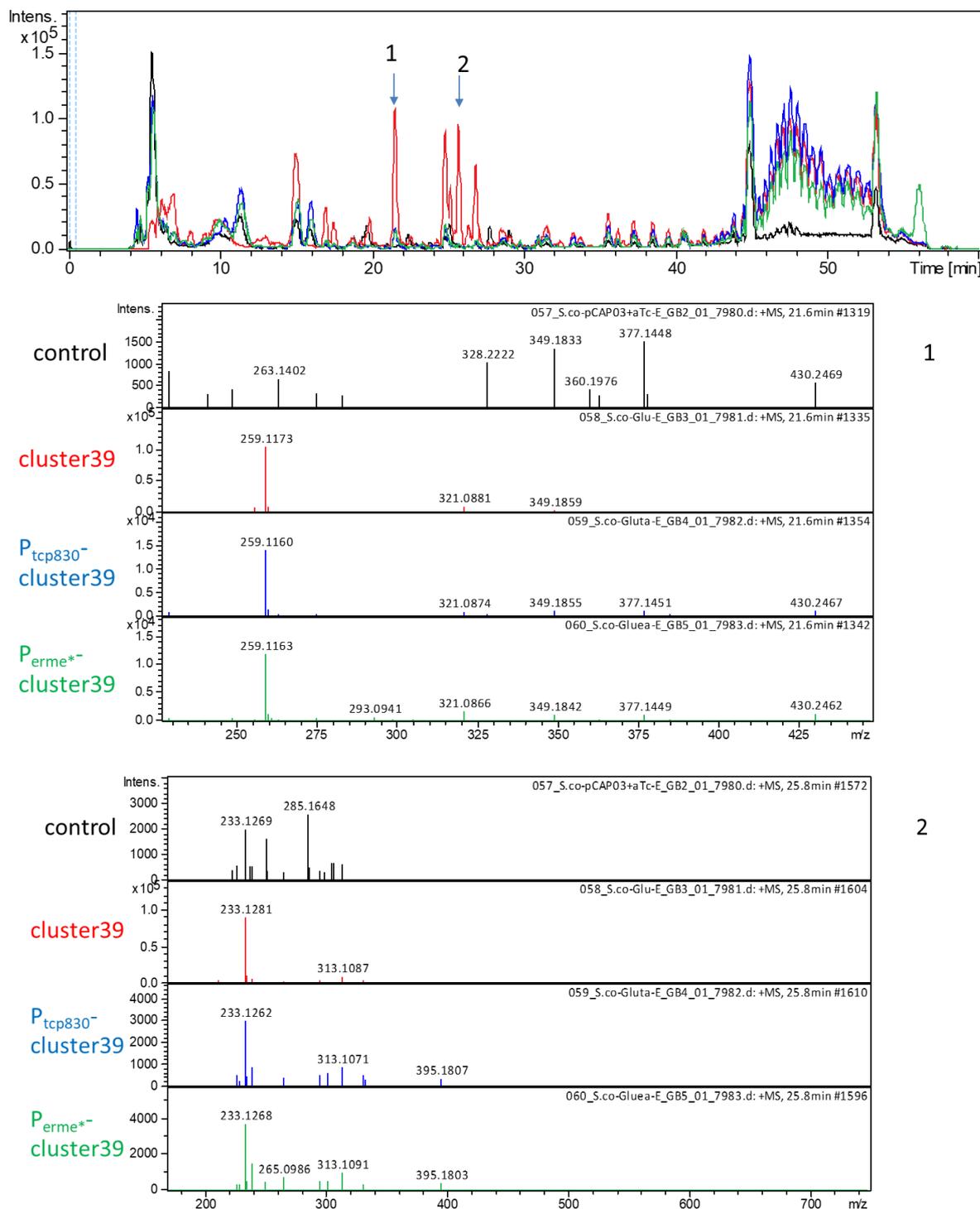


Figure III.25: LCMS data for ethylacetate extract of the heterologous expression of cluster 39 in *S. coelicolor*. Topmost figure showed the base peak chromatogram of control empty pCAP03 vector, cluster 39, P_{tcp830^-} -cluster39 and P_{erne^*} -cluster39. 1 and 2 are the spectra for the peaks at 21.6 minutes and 25.8 minutes, respectively. For peak 1, the ion with m/z 259.11 was only detected in extracts of heterologous expressions and not in the control. The dominant ion of peak 2, m/z 233.12 could be detected in all samples.

Figure III.25 showed the spectra of two peaks. These two peaks were chosen as the example of how the BPC was analyzed. All other peaks that appeared from the heterologous expression sample and not from the control were also investigated in the same manner. The first example is the peak from 21.6 minutes. At this retention time, the BPC for the control was flat and the BPC of the heterologous expression samples had a peak. At a closer look at the mass spectra in this peak, the main mass is 259.11 Da. However, when this mass was checked in another heterologous expression sample, *e.g.* the heterologous expression of cluster 52 and 23, the same mass appeared at the same retention time. Therefore, this was not regarded as the product of cluster 39. The second example is the peak from 25.8 minutes. The main mass of this peak is 233.12 Da, but it also appeared in much lower amount in the control. Therefore, this mass was also not considered as the product of cluster 39. Some other masses could be detected only in the heterologous expression samples and not in the control, *e.g.* m/z 321.08 at 21.7 min, 327.04 at 17.6 min, and 384.11 at 17.5 min. However, these masses could be dereplicated in the other heterologous expression samples from other clusters, *i.e.* cluster 52 and cluster 23. Therefore, these masses could not be assigned as the product of cluster 39.

III.4. Discussion

III.4.1. INDOBIO Project

As an attempt to find novel bioactive compounds to overcome the antibiotic resistance issue, understudied organisms were investigated. The conventional method of microorganism cultivation/co-cultivation and activity-guided fractionation often leads to the re-discovery of known compounds (Riyanti *et al.*, 2020). Meanwhile, the study of genetic modification tools and heterologous expression platform advanced rapidly in the past two decades (Zhang *et al.*, 2019a). The previous study of *Hexabranchnus sanguineus* egg mass revealed potentially active *Gordonia* strains (Böhringer *et al.*, 2017). The whole genome sequence of these strains revealed several putative NRPS and hybrid NRPS-PKS clusters when analysed by AntiSMASH. Other than cluster 52 from *G. terrae* Bu15_45, all putative clusters showed low similarity to known cluster, which harbour the potential of new compounds discovery.

AntiSMASH analysis of cluster 52 showed that it comprised of all genes that are necessary for the biosynthesis of nocobactin NA with 70% sequence similarity to the known BGC from *Nocardia farcinica*. Therefore, *G. terrae* Bu15_45 was cultivated to confirm the production of nocobactin NA. The LCMS analysis of ethyl acetate and methanol extract showed that nocobactin NA could be detected in its iron complexes (ferri-nocobactin NA-a and ferri-nocobactin NA-b). Since nocobactin NA could be dereplicated, cluster 52 was considered to be functional and used as the proof of principle for the methods that were applied for the heterologous expression approach.

Other than cluster 52, all other putative clusters that were regarded as complete BGCs are classified as NRPS with the smallest size of 17 kb to the largest size of 129 kb. Although these clusters have 60% to 72% region-to-region similarity to known cluster in MIBiG database, they do not have complete homologous genes to the known BGC and the same A domain specificities. These putative clusters were then further analysed to assign the borders of the clusters for cloning approaches and to predict the core structure for estimation of target mass to be detected in LCMS analysis.

The first factor to be considered for heterologous expression of these putative clusters was the host. Since *Gordonia* belongs to the phylum actinobacteria, *Streptomyces* was chosen as the heterologous expression host. *Streptomyces* is commonly used as heterologous expression host for actinomycete-derived natural products due to its versatility to produce polyketides and non-ribosomal peptides. In addition to that, molecular tools for genetic modification in *Streptomyces* strains as well as engineered strains for the purpose of heterologous expression are extensively studied. In this study, two strains were used as the expression host, *S. coelicolor* M1146 and *S. lividans* TK24. *S. coelicolor* M1146 is derived from *S. coelicolor* A3(2) with the deletion of four endogenous secondary metabolite gene clusters, *i.e.* actinorhodin, prodiginine, CPK (uncharacterized type I PKS) and CDA (calcium-dependent antibiotic) (Gomez-Escribano and Bibb, 2011). In contrast to *S. coelicolor* A3(2) whose complete genome sequence was available since 2002 (Bentley *et al.*, 2002), whole genome sequence for *S. lividans* TK24 was only available in 2015 (Rückert *et al.*, 2015). Only recently, the genetic engineering of the TK24 strain was proven to increase heterologous production of natural products by knocking out endogenous gene clusters and introduction of additional ϕ C31 integration sites (Ahmed *et al.*, 2020; Novakova *et al.*, 2018).

Various methods have been developed for the cloning of natural product BGCs into heterologous expression hosts, *e.g.* library-based methods (bacterial artificial chromosome, cosmid, fosmid, P1-derived artificial chromosome), Gibson assembly, site-specific recombination-based tandem assembly (SSRTA), transformation-associated recombination (TAR) cloning, linear-linear homologous recombination (LLHR) (Huo *et al.*, 2019; Zhang *et al.*, 2019a). In this work, TAR cloning was chosen as the cloning approach of the putative BGCs into the *Streptomyces* expression hosts due to its robustness to directly capture large BGC, up to 250 kb. However, it has the downside of low efficiency, where only 1%-5% of the yeast colonies carried the correct target gene (Kouprina and Larionov, 2008). In order to bypass this problem, an artificial metagenomic library from the genomic DNA of *Gordonia* sp. Bu15_44 and *G. terrae* Bu15_45 was generated.

The artificial metagenomic *Gordonia* library was a fosmid library that carries \pm 40 kb DNA inserts. In order to construct a good library, the quality of the genomic DNA was the most crucial aspect. Therefore, the DNA was handled carefully with wide-

bore tips and several analytical gels were run to check the DNA quality. The first one was the gel electrophoresis analysis of the isolated DNA from both *Gordonia* strains, where the DNA size was confirmed to be sheared to the size of around 40 kb with adequate concentration for further steps of library construction. After checking with analytical gel, another gel electrophoresis was performed on an LMP agarose gel to actually cut the gel to select DNA with the target size. This gel was not exposed to ethidium bromide and UV light to prevent DNA damage. A second analytical gel was performed after blunt-ending of the DNA, prior to ligation to pCC2FOS fosmid, which showed that the DNA size was maintained at 40 kb and no sheared DNA was observed.

The artificial metagenomic *Gordonia* library that has 20-fold coverage of the *Gordonia* sp. Bu15_44 and *G. terrae* Bu15_45 genomic DNA was screened for target BGCs in order to increase the TAR cloning effectiveness. Although the DNA insert in the fosmid has the size of 31-46 kb, it was difficult to find the whole BGC in one fosmid clone if the target BGC was more than 30 kb. Only cluster 52 (25,316 bp) and cluster 20 (17,198 bp) could be completely found as a whole BGC in one fosmid clone. Cluster 23 (40,971 bp) could be found in one fosmid clone without the last gene *orf23.18*. Cluster 37 (49,656 bp) could be found in one fosmid clone without the last four genes, *orf37.19–orf37.22*. Coincidentally, a second fosmid that has 16,419 bp overlap with the first fosmid and carried the four missing genes could be easily obtained from the PCR screening. Lastly, cluster 9 that was predicted to be a huge NRPS cluster, actually contain three independent NRPS clusters could be covered by the inserts of five fosmid clones. The cluster 9A (22,155 bp) was covered in one fosmid clone, while cluster 9B and 9C were covered in four fosmid clones that were overlapping with each other.

The fosmid that carries the target BGC was used as the DNA source for TAR cloning instead of directly using the genomic DNA from *G. terrae* Bu14_45. This method increased the TAR cloning efficiency significantly. 13 out of 27 (48%) yeast colonies gave positive bands from PCR corroboration of the presence of cluster 52, while 9 out of 14 colonies (64.2%) and 13 out of 14 colonies (92.8%) were confirmed as positive TAR clones for the cloning of cluster 23 and cluster 20, respectively. In comparison to the library-mediated TAR cloning that was used in this study, a CRISPR/Cas9-mediated TAR cloning was reported to increase the TAR cloning

efficiency up to 32% (Lee *et al.*, 2015). Therefore, the library-mediated TAR cloning was a better choice of method, considering that more than five target BGCs were intended to be cloned.

After TAR cloning, cluster 52 was integrated into *S. coelicolor* M1146 and *S. lividans* TK24 genomic DNA, mediated by Φ C31 integrase system. However, cultivation of both strains in different media did not yield expected masses for ferri-nocobactin NA-a nor ferri-nocobactin NA-b from LCMS analysis. Therefore, several constructs were made to introduce constitutive promoter *ermE* (Bibb *et al.*, 1985) and inducible promoter *tcp830* (Rodríguez-García *et al.*, 2005) at two different positions. The first position was in front of *orf118* which was considered as the start of the BGC (52.Rec3) and the second position was in front of *nbtS*, based on the operonic structure of cluster 52 (52.Rec4 and 52.Rec6). These constructs were integrated to the genome of the *Streptomyces* heterologous hosts. Additionally, two constructs were generated to carry two or three genes that are not in the same direction as the promoter (pGM1202-C1 and pGM1202-C2). The genes in these constructs were cloned in the same orientation under the constitutive promoter *ermE*^{*} and introduced as a plasmid to the respective *Streptomyces* hosts with integrated 52.Rec4 and 52.Rec6 (Table 7). The cultivation of these strains using different cultivation periods also did not result in production of the target compound.

Since the expression of cluster 52 failed to produce the expected product, the heterologous expression of the other BGCs was de-prioritized. The prospective work that could be done to reach the goal of NP production through heterologous expression of genome-mined BGC will be discussed in section **III.4.3**.

III.4.2. Burying beetle Project

An understudied bacterium was isolated from the gut of *Nicrophorus vespilloides*, namely strain 39, which showed activity against Gram-positive and Gram-negative bacteria. The classification of this strain remained to be hypothetical due to the low number of published genome sequence from the newly reclassified genus, *Glutamicibacter* (Busse and Schumann, 2019; Das *et al.*, 2020). The *in silico* analysis of its genomic DNA could only identify one NRPS/PKS hybrid cluster, which was hypothesized to produce the secondary metabolite responsible for the observed bioactivity.

In order to prove the hypothesis, the NRPS/PKS cluster was meant to be heterologously expressed in *Streptomyces* because *Glutamicibacter* and *Arthrobacter* belong to the same order of Actinomycetales. Instead of using direct cloning method, DNA assembly method was performed to clone the putative BGC into the integrative vector pCAP03. Additionally, a constitutive promoter (*tcp830*) was intended to be introduced upstream of the first gene of the construct, *orf39.15*. In total, six fragments were prepared for the isothermal assembly which includes one promoter cassette with streptomycin resistance gene, four fragments of the putative BGC that were amplified based on the gene directions, and one fragment of the vector. However, after several attempts, no correct assembly could be achieved. The efficiency of Gibson assembly was known to be maintained up to 75% for four fragments (Storch *et al.*, 2015), the first step assembly of *Mycoplasma genitalium* genome was also performed with maximum five fragments, each with the size of 5 to 7 kb (Gibson *et al.*, 2008). Therefore, the promoter cassette was removed from the Gibson assembly design, leaving five fragments to be assembled. By reducing the number of fragments to be assembled, the correct construct could be obtained (pCAP03-cluster39), although in a very low efficiency (3.57%).

After the assembly of the putative BGC to have all genes facing to the same direction in the pCAP03 vector, the next step would be the introduction of the promoter cassette. Initially, the promoter insertion was planned to be done by Gibson assembly, since the promoter fragment was already available from the previous attempt. In addition, the putative BGC could now be amplified from pCAP03-cluster39 in three fragments, since they were already arranged in the correct direction. However, the restriction digestion of the assembled constructs

always showed the wrong pattern. The obstacles that could prevent the generation of the right construct could be the low efficiency of Gibson assembly for more than four fragments, the repeated sequence that was available in the promoter cassette (FRT sites) which could form secondary structures, as well as the presence of *oriT* in the antibiotic resistance cassette, which is also present in the pCAP03 vector (Hillson, 2011; Storch *et al.*, 2015; Yamanaka *et al.*, 2014).

To overcome the low efficiency problem of Gibson assembly, λ -Red recombination was chosen as the method for the promoter insertion. This method allows the insertion of a linear DNA fragment into a circular plasmid based on recombination of 60 bp homology on each ends, mediated by bacteriophage-derived λ -Red proteins (Gust *et al.*, 2004; Mosberg *et al.*, 2010). The promoter cassette that was going to be inserted was also slightly modified by removing one FRT site and the *oriT* sequence. This method proved to be efficient, whereby all colonies that were picked carried the correct construct. By using this method, the constitutive promoter (*ermE**) and the inducible promoter (*tcp830*) were introduced upstream of *orf39.15*, creating pCAP39-ea and pCAP39-ta.

After the integration of the three constructs (pCAP03-cluster39, pCAP39-ea and pCAP39-ta) into *S. coelicolor* M1146 as the expression host, the transgenic strains and a control (empty pCAP03 integrated into *S. coelicolor* M1146) were cultivated to investigate the product of the cloned BGC. Thus, LCMS chromatograms were analysed thoroughly by manual comparison of the base peak chromatogram of the control and the transgenic strains. Although some masses could be identified as a candidate to be the product of cluster 39, they were later found in other heterologous expression samples, *i.e.* cluster 52 and cluster 23, and had to be removed as the candidate mass. In the end, this method could not identify the product of cluster 39 yet. The prospective work that is proposed to be done to achieve the heterologous production as well as for improvement of compound detection will be discussed in the following section.

III.4.3. Prospective work to improve heterologous expression of NP BGC

Since cluster 52 is a siderophore BGC, it was expected that this cluster is regulated by DtxR (diphtheria toxin regulator) family, which is conserved in Gram-positive bacteria with high-GC-content DNA (Waldron and Robinson, 2009). Characterization of IdeR, a member of DtxR family, from *Streptomyces avermitilis* showed that in the presence of iron, *ideR* expression was activated and the IdeR-Fe²⁺ complex repressed the transcription of target genes by binding specifically to the 'iron box' sequence within their promoter region. This way, IdeR controls iron homeostasis in *Streptomyces avermitilis* by regulating the expression of siderophore BGCs, iron transporter genes, iron storage genes, and iron utilization genes (Cheng *et al.*, 2018).

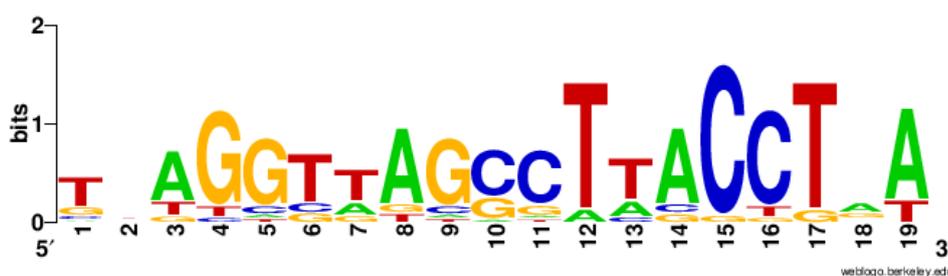


Figure III.26: Position weight matrix of twenty iron box sequences from *S. avermitilis*. Letter size is equivalent to the weight of the nucleotide base at a certain position (Crooks *et al.*, 2004).

Twenty iron box sequences from *S. avermitilis* that regulate iron metabolism, transcriptional regulation and secondary metabolism (Cheng *et al.*, 2018) were merged to create a consensus sequence (**Figure III.26**). The degenerated sequence (5'- TNAGGTWAGSCTWACCTRA-3') was used as the query to scan cluster 52 for IdeR binding site using Virtual Footprint (Münch *et al.*, 2005). As a result, five putative binding sites were identified, *i.e.* upstream of *orf117*, *nbtT*, *nbtA*, *orf127* and *nbtD* (**Figure III.27**). Since IdeR binds to both DNA strands of the iron box to form a double dimer complex (Ghosh *et al.*, 2015; Wisedchaisri *et al.*, 2004), two IdeR binding sites repressed nocobactin biosynthesis genes in cluster 52 bi-directionally (**Table S4**). In total, the location of the five putative binding sites

suggests that IdeR would repress the transcription of all genes in cluster 52. Therefore, these iron boxes should be removed in order to activate the expression of the genes in cluster 52.

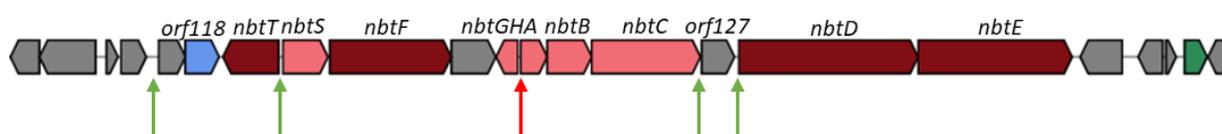


Figure III.27: Prediction of IdeR binding site in cluster 52. Green arrow indicates iron box sequence in DNA minus strand and red arrow indicates iron box sequence in DNA plus strand.

Previously, Luo *et al.* (2013) and Shao *et al.* (2013) showed that knocking out the BGC regulatory gene was not enough to increase the expression of the biosynthesis genes, while Bauman *et al.* (2019) showed that streptophenazines were produced in trace amounts when the regulatory gene was removed. Even when the *ermE** promoter was inserted upstream of the whole PTM BGC or upstream of the predicted streptophenazine BGC regulator, the biosynthetic genes were still expressed in a very low level. The heterologous expression of PTM and spectinabilin was successfully achieved when well-characterized promoters were introduced in front of each gene in the BGC. However, it seems that it is unnecessary to introduce a promoter in front of each gene in the BGC, because the heterologous expression of streptophenazine showed that the *sp44* promoter could promote expression of the 25 kb BGC and the heterologous expression of novobiocin showed that the use of a single *tcp830* promoter was sufficient to promote the transcription of the respective 18 kb BGC (Bauman *et al.*, 2019; Dangel *et al.*, 2010). Therefore, one pathway refactoring strategy that could be applied to cluster 52 in order to heterologously produce nocobactin is by studying the expression of the biosynthetic genes in cluster 52 to find which genes are not transcribed in the heterologous environment and replacing the iron box with promoters with different strength to fine-tune the gene expression. The promoters could be selected from housekeeping gene promoters or from a synthetic promoter library of which the ones should be selected that are known to function in the culturing conditions. (Shao *et al.*, 2013; Siegl *et al.*, 2013).

Since the product of cluster 39 from *Glutamicibacter arilaitensis* and the other BGCs from *G. terrae* Bu15_45 are not known, it is difficult to predict the regulatory pathway that is involved in the BGC. Moreover, no regulatory gene was predicted to exist in any of the putative BGC. The most robust way to improve the heterologous expression of these clusters is to introduce promoters based on the operonic structure of the genes and remove all intergenic region where the regulatory protein usually binds. In addition to that, if production does not occur, a transcriptomic study could be performed to find the genes that are not transcribed in the heterologous expression environment and put an additional promoter upstream of the gene.

Once the biosynthetic genes are confirmed to be transcribed, another aspect that needs to be improved is the LCMS data analysis. One tool that could be used for the analysis of the LCMS data is GNPS (Global Natural Products Social Molecular Networking). This online tool can be used to group the masses based on the fragmentation pattern so that the data set from LCMS measurements can be compared to each other as well as to spectral library for compound dereplication (Wang *et al.*, 2016). However, the limitation for the usage of this tool was the quality of the MS/MS data, where the LCMS data that was previously measured did not have enough fragments for sufficient molecular networking.

Additionally, the use of heterologous expression hosts with cleaner metabolic background and higher production of secondary metabolites would ease the product detection. For example, *S. coelicolor* M1154 could be a better choice than *S. coelicolor* M1146, since *S. coelicolor* M1154 proved to produce more heterologous metabolites due to the mutation in the *rpoB* and *rpsL* which leads to the increase in gene expression (Gomez-Escribano and Bibb, 2011; Jones *et al.*, 2013; Ochi and Hosaka, 2013). Recently, an improved *S. lividans* strain was also published, where 10 of the endogenous secondary metabolite BGCs were removed and three additional integration sites for ϕ C31 system were introduced. This engineered strain could produce deoxycinnamycin 3.5- to 4.5-fold more than the parental *S. lividans* TK24 and 7- to 10-fold more than *S. coelicolor* M1154 (Ahmed *et al.*, 2020). Since the engineered *S. lividans* Δ YA10 could produce higher level of amino acid-derived compounds, this strain would be suitable to be the heterologous expression host of most of the putative BGCs that were described in this study.

III.5. Summary

In silico analysis of the genomic DNA sequence data of *Gordonia terrae* Bu15_45, obtained from the INDOBIO project, and of strain 39, obtained from the burying beetle project, revealed five and one putative BGCs, respectively. One BGC from *G. terrae* Bu15_45 (cluster 52) harbors genes with high identity to the nocobactin NA BGC. Meanwhile, the other five identified BGCs show low similarities to known BGCs. Since *G. terrae* Bu15_45 and strain 39 were previously proven to possess antibacterial activities, the identified BGCs of the secondary/specialized metabolism hold immense potential for the discovery of new bioactive compounds.

The cloning of target BGCs proved to be challenging due to the large size and repetitive sequences of the NRPS- and PKS-modular systems. Different techniques were applied for cloning of the respective target BGC into the expression vector. In the INDOBIO project, the cloning of target BGCs was achieved by generation of an artificial metagenomic library to increase the cloning efficiency, followed by TAR cloning. In the burying beetle project, the target BGC was successfully cloned into the expression vector using the Gibson assembly method.

After the successful cloning of the respective target BGCs, as well as the introduction and integration of the BGCs into the genome of the expression hosts *Streptomyces coelicolor* M1146 and *Streptomyces lividans* TK24, these heterologous hosts were fermented and subsequently analysed with HR-LCMS to detect and identify compounds produced by the system. Cluster 52 that was linked by *in silico* analysis to the known metabolite nocobactin NA was planned to be used as proof of principle. However, the expected product could not be detected by HR-LCMS analysis. Similarly, the analysis of fermentations of the heterologous expression strains of cluster 23 from *G. terrae* Bu15_45 (**Figure S3**) and cluster 39 from strain 39 (*Glutamicibacter arilaitensis*) did not result in the detection of compounds that could be linked to the respective BGC.

Even though the cloning and heterologous expression of the BGCs did not result in the biosynthesis of natural products in the system tested, the constructs might form the basis for future studies. Biosynthetic pathway refactoring was discussed in section III.4.3 to be a prospective approach to improve heterologous expression of natural product BGCs. Moreover, recent advancements in the engineering of

Streptomyces heterologous expression strains provided new strains such as *S. lividans* Δ YA10 and *S. coelicolor* M1154 that could be used as more suitable hosts.

IV. Heterologous expression of RiPP BGC

IV.1. Introduction

Antibiotic resistance has become one of the major threats on global health, as infectious diseases are becoming more and more difficult to be treated by the market available antibiotics. The world health organization (WHO) puts critical priority to the research and development of new antibiotics against the multi-drug-resistant (MDR) Gram-negative pathogens, e.g. *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, *Escherichia coli*, *Klebsiella pneumoniae* (WHO News Release, 2017). The outer membrane (OM) of Gram-negative bacteria plays a major role in their resistance against certain drugs, due to the fact that the OM functions as a permeability barrier, which restricts drug penetration into the cell (Silhavy *et al.*, 2010). Hence, there is an obvious need to fill the antibiotic development pipeline with candidate molecules that can be evolved into medicinal drugs for the treatment of infectious diseases caused by Gram-negative bacteria.

Darobactin A (DAR) is a novel antibiotic that selectively kills Gram-negative pathogens, e.g. *Acinetobacter baumannii* (MIC, 8 µg/ml), *Pseudomonas aeruginosa* PAO1 (MIC, 2 µg/ml), *Escherichia coli* (MIC, 4 µg/ml), and *Klebsiella pneumoniae* (MIC, 2 µg/ml). The inhibitory activity of DAR against *E. coli* was initially observed from the 15x concentrated extract of *Phototrhobdus kharii* HGB1456. DAR was also observed to be produced by other *Phototrhobdus* strains, but a long fermentation period (10-14 days) was required to obtain DAR with a maximum yield obtained from *P. kharii* DSM3369, which was only 3 mg/L (Imai *et al.*, 2019).

DAR, which has a molecular formula of $C_{47}H_{55}O_{12}N_{11}$ and molecular mass of 966.41047, is a ribosomally synthesized peptide that consists of seven amino acids $W^1-N^2-W^3-S^4-K^5-S^6-F^7$, with post-translational modifications that create an ether bond between W^1 and W^3 , and a C-C bond between W^3 and K^5 (**Figure IV.1**). The tryptophan-to-tryptophan ether bond is a unique link that has not been described prior to the discovery of DAR, while the tryptophan-to-lysine bond was first observed in streptide, a compound produced by *Streptococcus thermophilus* (Schramma *et al.*, 2015).

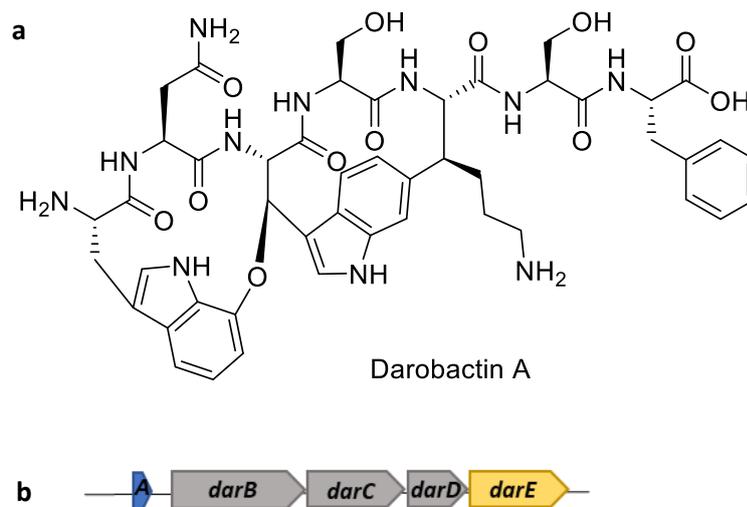


Figure IV.1: Darobactin A (DAR) is a ribosomally synthesized and post-translationally modified peptide (RiPP), encoded by the *dar* operon. **a**, DAR is a modified heptapeptide consisting of the seven amino acids W^1 - N^2 - W^3 - S^4 - K^5 - S^6 - F^7 , with an ether bond between W^1 and W^3 , and a C-C bond between W^3 and K^5 . **b**, The *dar* BGC (in total 6.2 kb in length) consists of *darA* that encodes the precursor peptide; *darBCD* that encode subunits of an ABC transporter; and *darE* that encodes a radical S-adenosylmethionine enzyme (RaS).

A co-crystallization model was created and further experimental proof was provided that DAR binds to BamA, which is the central component of the OM β -barrel assembly machinery (Kaur *et al.*, 2021). It helps to fold and insert β -barrel proteins such as porins into the OM. If this chaperone-like function is impaired, it will result in the disruption of OM formation. In addition to its good *in vitro* activity, DAR showed promising efficacy in mouse septicemia and thigh infection models without cytotoxic effects (Imai *et al.*, 2019). Therefore, DAR has emerged as a promising drug lead.

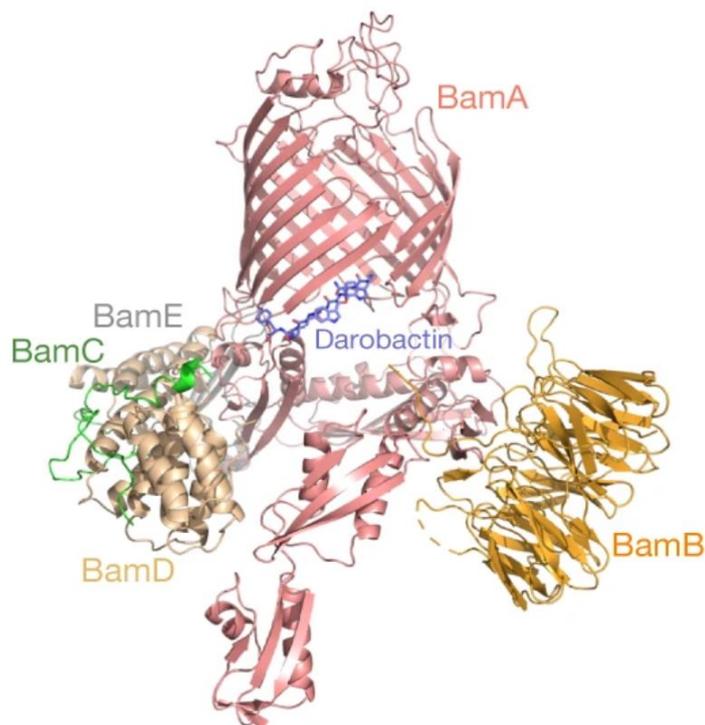


Figure IV.2: Darobactin A (DAR) mode of action. DAR binds to the β 1 strand of the BamA β -barrel lateral gate and blocks the β -signal binding site. Thus, the nascent OMP chain interaction to the BamA lateral gate is compensated and the OM formation is impaired (Kaur *et al.*, 2021).

Heterologous expression is often used to express a silent gene cluster (Hegemann *et al.*, 2013; Yamanaka *et al.*, 2014), express BGCs from unculturable bacteria (Crüsemann *et al.*, 2018; Ongley *et al.*, 2013), increase yield (Flinspach *et al.*, 2010; Sucipto *et al.*, 2017), as well as to proof the functionality of a putative BGC (Linares-Otoya *et al.*, 2017). In addition to that, the genetic tools for modification of the BGC in a heterologous expression system are more easily available to study the biosynthetic pathway (Bouhired *et al.*, 2014; Crüsemann *et al.*, 2018) and to produce new derivatives of the natural product (Paulus *et al.*, 2018).

We identified a 6.2 kb biosynthetic gene cluster (BGC) that is responsible for DAR production (**Figure IV.1b**). The BGC consists of five genes, *darABCDE*. *darA* encodes for a precursor peptide consisting 58 amino acids, *darB* encodes for ABC transporter permease protein, *darC* encodes for periplasmic adaptor protein of the tripartite efflux pump, *darD* encodes for the ATP binding domain of the inner membrane ABC transporter, and *darE* encodes for radical SAM enzyme which

belongs to the SPASM subfamily, SPASM subfamily founding members are the enzymes that are involved in the maturation of Subtilosin, Pyrroloquinoline quinolone (PQQ), Anaerobic sulfatases, and Mycofactocin. Knocking out the BGC abolished DAR production, but reinstatement of the BGC into *P. khanii* DSM3369 could not restore DAR production, indicating that the regulatory framework in *P. khanii* DSM3369 does not allow interference.

The present study aims to further investigate the identified BGC by heterologous expression approach. As the first step, it is projected to confirm that the identified BGC is a complete cluster and can be used to produce DAR heterologously. Secondly, it is envisaged that the major limitation for further investigation of DAR, *i.e.* the compound supply due to the long fermentation period and low DAR production yield by native producer, to be overcome. Therefore, the goal was improvement of DAR production by optimization of the biosynthesis using a heterologous production system. We choose several different heterologous hosts, cloned the respective DAR BGC from different species, including upstream regions of the BGC, and lastly created a DAR-resistant heterologous host to boost DAR production. Furthermore, *in silico* predicted DAR BGCs were experimentally confirmed and the minimal BGC necessary for heterologous production was identified.

IV.2. Material and methods

IV.2.1. Plasmid and strain construction

Several constructs were generated during this project for heterologous expression of DAR (**Figure IV.3**). The vector used for the expression constructs was in all cases pRSFDuet™-1 (Merck KGaA). Chromosomal DNA used as template for amplification of the DAR BGC was isolated using the innuPREP Bacteria DNA Kit (Analytik Jena AG). In general, fragments were amplified using Q5 DNA polymerase (New England Biolabs) and purified from the agarose gel using Large Fragment DNA Recovery Kit (Zymo Research). The polymerase chain reaction (PCR) was performed in a Biometra TRIO thermocycler (Analytik Jena AG) using the following program: 95°C for 2 minutes; 34 cycles of 95°C for 45 seconds, 60-70°C for 45 seconds (applied annealing temperature was depending on the primer sequence), 72°C for 30 seconds/kb (extension time was varied depending on the length of the fragment to be amplified), followed by a final extension step at 72°C for 5 minutes. All primer sequences can be found in **Table S1**.

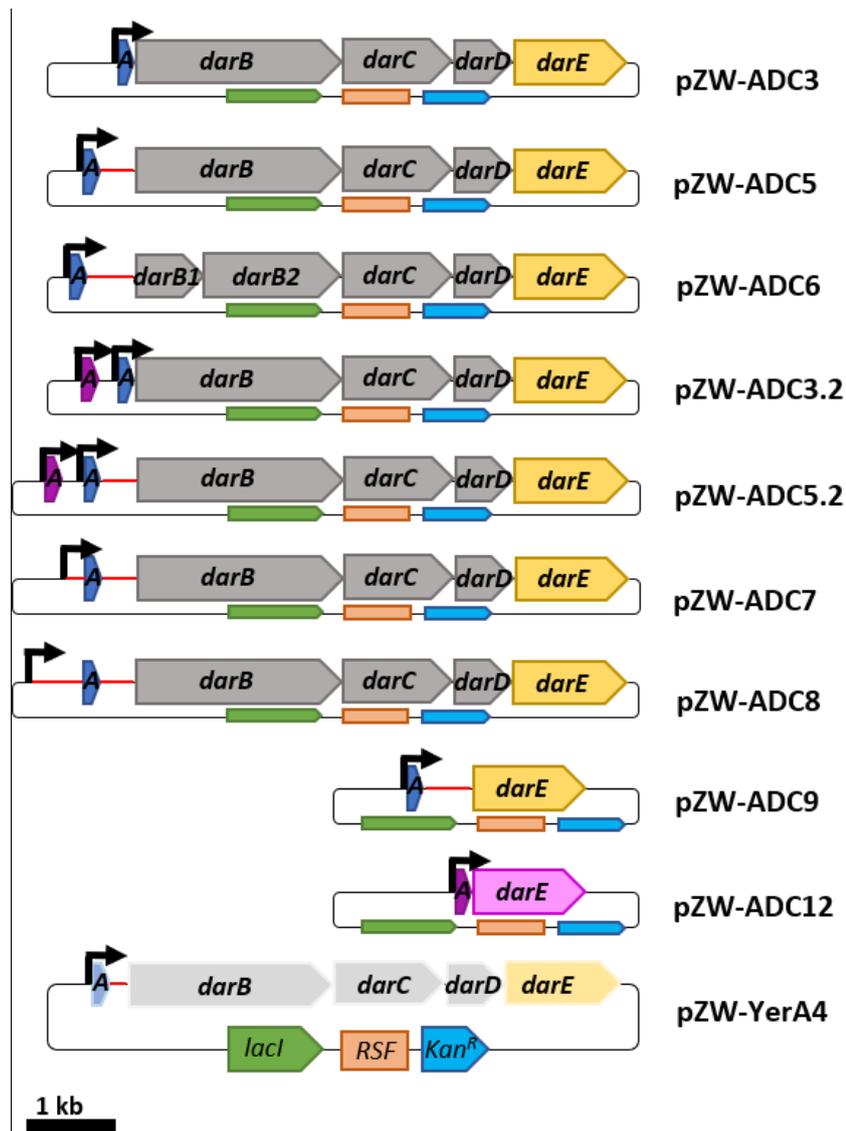


Figure IV.3 : Expression constructs created in this study. The DAR BGC was cloned into the pRSFDuet™-1 vector under control of the T7-*lac* promoter. pZW-ADC3 has a ‘streamlined’ DAR BGC from *P. khanii* HGB1456, where all intergenic regions were removed, while pZW-ADC5 and pZW-ADC6 carry the native cluster from *P. khanii* HGB1456 and *P. khanii* DSM3369, respectively. Addition of a codon optimized *darA* from *P. namnaonensis* (in purple) to pZW-ADC3 and pZW-ADC5 created pZW-ADC3.2 and pZW-ADC5.2. Plasmid pZW-ADC7 carries the *P. khanii* HGB1456-derived DAR BGC with an additional 200 bp upstream region of *darA* and pZW-ADC8 harbors a 605 bp upstream region. Plasmid pZW-ADC9 harbors the *P. khanii* DSM3369-derived DAR BGC, omitting *darBCD*, while pZW-ADC12 carries codon optimized *darA* and *darE* (in pink) from *P. namnaonensis* and pZW-YerA4 carries the DAR BGC from *Yersinia frederiksenii* ATCC 33641. The black arrow indicates the T7-*lac* promoter and the color code for *lacl*, RSF and the kanamycin resistance cassette (Kan^R) is kept constant. A red line indicates sequence derived from the BGC, while a black line indicates the vector backbone. The *lacl* gene encodes for the *lac* operon repressor and RSF is an origin of replication that was derived from RSF1030, which allows the plasmid to be maintained at a high copy number in the cell.

Plasmid pZW-ADC3 carries the genes of the DAR BGC without the intergenic region between *darA* and *darB*. Therefore, *darA* was amplified from *Photorhabdus khanii* HGB1456 (**Table S2**, GenBank accession number WHZZ00000000) using the primer pair ADC-F7 and ADC-R1; and *darB* to *darE* was amplified using the primers ADC-F2 and ADC-R7. pRSFDuet™-1 was linearized using *NdeI* and *AvrII* restriction enzymes (New England Biolabs) to insert the two purified fragments into the second multiple cloning site of the vector under the control of the T7-*lac* promoter. To do this, the one-step isothermal DNA assembly protocol described by Gibson was followed (Gibson *et al.*, 2009), with the minor modification that 1.2 µl of 10 U/µL T5 exonuclease was added instead of 0.64 µL. Therefore, the final concentration of the Gibson reaction mix was the following: 100 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 0.2 mM each dNTP, 10 mM DTT, 5% PEG-8000, 1 mM NAD, 7.5 U/mL T5 exonuclease, 25 U/mL Phusion polymerase, 4 U/µL Taq DNA ligase, 0.02 – 0.5 pmol DNA fragments. This reaction mix was then incubated at 50°C for one hour. After the isothermal assembly, the reaction was dialyzed using a 0.025 µm nitrocellulose membrane (Merck™ MF-Millipore™), and subsequently transferred to *E. coli* TOP10 cells as a plasmid maintenance host by electroporation using Micropulser Electroporator (Bio-Rad) in a 0.2 cm electroporation cuvette at a voltage of 2.5 kV.

The second and third construct, carries the native DAR BGC from *Photorhabdus khanii* HGB1456 (pZW-ADC5) and *Photorhabdus khanii* DSM3369 (GenBank accession number AYSJ00000000) (pZW-ADC6), respectively. For both constructs, the respective BGC was amplified using the primer pair ADC-F7 and ADC-R7; therefore, the respective bacterial genomic DNA was used as template. Gibson Assembly was performed as described above.

Plasmids pZW-ADC3.2 and pZW-ADC5.2 were created by restriction-digest of pZW-ADC3 and pZW-ADC5. Each plasmid was restricted using *NcoI* and *NotI* (New England Biolabs). Then, an additional codon optimized version of *Photorhabdus namnaonensis darA* (**Sequence S1**) was inserted into the first multiple cloning site of the pRSFDuet™-1 based vectors by Gibson Assembly.

The plasmids pZW-ADC7 and pZW-ADC8 carry the native DAR BGC from *P. khanii* HGB1456 with an additional 200 bp and 605 bp upstream region of *darA*,

respectively. The BGC were amplified from the bacterial genomic DNA using the primer pair ADC-F9 and ADC-R7 for the pZW-ADC7 insert, and the primer pair ADC-F10 and ADC-R7 for the pZW-ADC8 insert. The respective inserts were assembled to the *NdeI-AvrII* linearized pRSFDuet™-1 using Gibson Assembly.

The plasmid pZW-ADC9 carries the DAR BGC from *P. khanii* DSM3369 without the transporter genes, *i.e.* only *darA* and *darE*. *darA* was amplified from the bacterial genomic DNA using the primer pair ADC-F7 and ADC9-R and *darE* was amplified using the primer pair ADC9-F and ADC-R7. Both fragments were assembled to the *NdeI-AvrII* linearized pRSFDuet™-1 using Gibson Assembly.

Synthetic codon-optimized DNA of *darA* and *darE* from *P. namnaonensis* (**Sequence S1** and **S2**) were amplified using the primer pairs pro.CO.F – pro.CO.R and RS.CO.F and RS.CO.R, respectively. Both amplicons were assembled to the *NdeI-AvrII* linearized pRSFDuet™-1 vector by Gibson Assembly, creating pZW-ADC12.

pZW-YerA4 carries the DAR BGC derived from *Yersinia frederiksenii* ATCC 33641 (GenBank accession number NZ_JPPS000000000) that was amplified using the primer pair YerA-F4 and YerA-R4. The amplicon was also assembled to the *NdeI-AvrII* linearized pRSFDuet™-1 by using Gibson assembly.

Following assembly and propagation in *E. coli* TOP10 cells, all constructs were checked by test PCR and by their restriction pattern.

IV.2.2. Heterologous expression for functional proof of DAR BGC

The expression construct pZW-ADC3 was transferred from the maintenance host, *E. coli* TOP10, into the expression host *E. coli* BL21(DE3) by electroporation. The transformation of *E. coli* expression hosts was performed as described above for *E. coli* TOP10. After the electroporation, the transformed cells were incubated for 1 hour in LB medium at 37°C. Then, they were plated on LB plates containing 50 mg/L kanamycin. Single colonies were picked from the selective plates and the presence of pZW-ADC3 was confirmed by colony PCR. Colony with correct insert was then inoculated in 3 mL Terrific Broth (24 g/L yeast extract, 20 g/L tryptone, 4 mL/L glycerine, 0.017 M KH₂PO₄, 0.072 M K₂HPO₄) containing 50 mg/L kanamycin and

incubated at 37°C overnight. 300 µL of this pre-culture were used to inoculate 30 mL of fresh Terrific Broth containing 50 mg/L kanamycin, incubated at 37°C until an OD₆₀₀ of ~0.5 was reached, and then induced with IPTG (final concentration of 0.5 mM). After IPTG induction, the cultures were incubated at 30°C with 120 rpm shaking.

After three days of fermentation, 1 mL aliquot was taken from the expression culture, 1 mL methanol was added, and the mixture was vortexed for 5-10 seconds. Then, it was centrifuged at top speed for 5 minutes, and 200 µL of the supernatant was sent to be analysed on a micrOTOF-Q mass spectrometer (Bruker) as described in section III.2.10, with slight alteration that 5 µL of the samples was injected, and the sample concentration was not adjusted to 1 mg/mL. As a control for DAR production, *P. khanii* HGB1456 was grown in LB medium for 10 days at 30°C with 120 rpm shaking, and LC-MS sample was prepared in the same manner as the heterologous expression culture.

IV.2.3. Qualitative analysis of different promoters

The qualitative analysis of different promoters was done by visualization using green fluorescence protein (GFP) as reporter gene, and by testing the inhibition of *E. coli* MG1655 BamA6 (Ruiz *et al.*, 2006). To create the construct pZW-ADC5-GFP, *gfpmut3.1* was amplified from pFU95 (GenBank accession number: JF796092.1) using the primer pair GFP-F and GFP-R2 and pZW-ADC5 was amplified using the primer pair ADC5-GFP-F2 and ADC5-GFP-R. Then, these two fragments were assembled by Gibson Assembly. Thereafter, plasmid pZW-ADC5-GFP was amplified using the primer pair T5-F and T5-R. Afterwards, one-fragment-Gibson-Assembly was performed to replace the T7-*lac* with the T5 promoter (Shibui *et al.*, 1988).

The J23101 promoter (Davis *et al.*, 2011) was amplified using overlap extension PCR. The first primer pair, *i.e.* J23101-F and J23101-R, in which the primers complement each other, has a 20 bp 5' overhang region on each primer. The first PCR was performed to fill the overhang using the following conditions: 95°C for 2 minutes; 15 cycles of 95°C for 45 seconds, 60°C for 45 seconds, 72°C for 15 seconds. Subsequently, the resulting double stranded DNA template was amplified

using the second primer pair J23101-F2 and J23101-R2. Hence, this primer pair was added and the following PCR program was applied: 95°C for 2 minutes; 30 cycles of 95°C for 45 seconds, 55°C for 45 seconds, 72°C for 15 seconds, followed by a final extension step at 72°C for 5 minutes. The therefrom-resulting 125 bp fragment was assembled with pZW-ADC5-GFP (**Figure S4**) that was amplified using the primer pair ADC-F7 and J23101-ADC5-GFP-R.

The *prpB* promoter and its regulatory gene *prpR* were amplified from *E. coli* MG1655 genomic DNA (GenBank accession number NC_000913.3) with the primer pair pPro-F and pPro-R, and the primer pair pPro-A5-GFP-F and pProA5-GFP-R was used to amplify pZW-ADC5-GFP (**Figure S4**). The two fragments were assembled using Gibson Assembly.

Overnight cultures of *E. coli* Rosetta™(DE3) that carry the respective constructs with different promoters were grown in LB with kanamycin and chloramphenicol at 37°C. 5 µL of the culture was spotted on LB agar with necessary antibiotics and promoter inducer. The constructs with T5 and J23101 promoter did not need any inducer, T7 promoter was induced with 1 mM IPTG, and *prpB* promoter was induced with 20 mM sodium propionate. The agar plates were incubated at 30°C for three days, and GFP-depending fluorescence was observed by a blue-light transilluminator at 470 nm. Subsequently, the plates were overlaid with LB-agar containing 1% overnight culture of *E. coli* MG1655 BamA6, kanamycin, and necessary inducer. The plates were incubated at 30°C, and the inhibition zones were documented on the next day.

IV.2.4. Generation of *E. coli* mutant strains as heterologous hosts

Two *E. coli* mutant strains were created in this study: (i) *E. coli* Rosetta™(DE3)(Dar^R) and (ii) *E. coli* BAP1 $\Delta toIC::aac(3)-IV$ (Apr^R); by means of lambda Red recombination. The DAR resistant strain was generated by introduction of three point mutation into the *bamA* gene, which are 1300A>G, 1334A>C, and 2113G>A. That these three point mutations result in a DAR resistant phenotype was confirmed by a previous study, whereby DAR resistance increased to 128 µg/mL (Imai *et al.*, 2019). A linear PCR product of the *bamA* gene with the three point mutation was amplified from this *E. coli* DAR-resistant mutant (strain3) (Imai *et al.*,

2019) using the *bamA-recF* and *bamA-recR* primer pair. On the other hand, an apramycin resistance cassette (*aac(3)-IV*) flanked with FRT regions was amplified using the TolCKO-F and TolCKO-R primer pair to be introduced to *E. coli* BAP1 to create the *tolC* knockout mutant.

The λ -Red recombination to create the mutant strain was done according to Datsenko and Wanner (Datsenko and Wanner, 2000), with some modifications. The target *E. coli* strains were transformed with pKD46, a heat-sensitive plasmid that carries λ red genes under the control of the *araBAD* promoter. They were grown in lysogeny broth (LB) with carbenicillin (50 μ g/mL) at 30°C for one hour, supplemented with 20 mM L-arabinose, then further incubated to reach the OD₆₀₀ of \approx 0.6. Thereafter, the cells were made electrocompetent by three times washing with ice-cold 10% glycerol. The linear PCR amplicon (\sim 100 ng) was added into 50 μ L of electrocompetent *E. coli* cell solution and electroporation was performed as previously described in section IV.2. After 1 hour incubation for cell recovery, the *E. coli* Rosetta™(DE3)(Dar^R) strain was plated on LB with DAR (32 μ g/mL) and the *E. coli* BAP1 Δ *tolC*::*aac(3)-IV* (Apr^R) was plated on LB with apramycin (50 μ g/mL) as selection markers, and incubated at 37°C to promote the loss of pKD46. The DAR resistant strain was confirmed by sequencing of the *bamA* gene and the *tolC* knockout strain was confirmed by PCR.

IV.2.5. Heterologous host strains, media and maintenance

All bacterial strains used in this study were maintained in glycerol stocks (25% (v/v)) at -80°C. *E. coli* and *P. khanii* strains were grown in LB medium at 37°C and 30°C, respectively, unless stated otherwise. The LB medium was prepared with 10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl, and supplemented with 50 mg/L kanamycin and 25 mg/L chloramphenicol when required for the purpose of selection. Kanamycin supplementation was done for *E. coli* strains carrying the expression constructs and chloramphenicol was supplemented for *E. coli* Rosetta™(DE3) cultivation.

To generate a growth curve, 50 mL of LB medium were supplemented with required antibiotics, inoculated with 500 μ L of an overnight culture, and cultivated at 30°C

with 180 rpm shaking. Samples for OD₆₀₀ measurement were taken at 0, 2, 4, 6, 9, 12, 16, 25, 54, 78, and 102 hours after inoculation.

Vibrio natriegens Vmax™ was grown in LB medium with artificial seawater (LB-ASW) at 30°C. The artificial seawater was prepared with 0.1 g/L KBr, 23.48 g/L NaCl, 10.61 g/L MgCl₂·6H₂O, 1.47 g/L CaCl₂·2H₂O, 0.66 g/L KCl, 0.04 g/L SrCl₂·6H₂O, 3.92 g/L Na₂SO₄, 0.19 g/L NaHCO₃, 0.03 g/L H₃BO₃. 200 mg/L kanamycin was supplemented to the LB-ASW for the growth of *V. natriegens* Vmax™ that carries the expression constructs.

To make electrocompetent cells for electroporation, *E. coli* strains and *V. natriegens* Vmax™ were grown in LB and LB-ASW, respectively, to OD₆₀₀ between 0.4 – 0.7. Then, the cells were washed two times with electroporation buffer (10% glycerol for *E. coli* strains, and 680 mM sucrose, 7 mM K₂HPO₄, pH 7 for *V. natriegens* Vmax™) (Weinstock *et al.*, 2016). Subsequently, the cells were resuspended in 1:100 volume of electroporation buffer compared to the initial culture volume, and aliquoted into chilled tubes. The electrocompetent cells were always made fresh prior to electroporation.

IV.2.6. Heterologous expression for optimization of DAR production

The expression constructs were transferred from the maintenance host, *E. coli* TOP10, into different expression hosts by electroporation. The transformation of *E. coli* expression hosts was performed as described above for *E. coli* TOP10, while the transformation of *V. natriegens* Vmax™ was done in a 0.1 cm electroporation cuvette at a voltage of 900 V. After the electroporation, the strains were incubated for 1 hour in their respective growth medium and temperature. Then, they were plated on LB(-ASW) plates containing kanamycin as the selective agent with the aforementioned concentration, with the addition of chloramphenicol when *E. coli* Rosetta™(DE3) was used as the host. Single colonies were picked from the selective plates and the presence of the respective expression plasmid was confirmed by PCR. Colonies with correct constructs were then inoculated in 3 mL LB(-ASW) containing required antibiotics and incubated at 37°C or 30°C overnight (see point 2.2). 500 µL of this pre-culture were used to inoculate 50 mL of fresh LB(-

ASW) medium containing kanamycin, incubated at 37°C or 30°C until an OD₆₀₀ of 0.4 – 0.6 was reached, and then induced with IPTG (final concentration of 0.5 mM). After IPTG induction, the cultures were incubated at 30°C with 180 rpm shaking.

IV.2.7. UPLC-HRMS analysis

DAR production was analyzed by UPLC-HRMS. From the expression culture, a 1 mL aliquot was taken and centrifuged to separate the medium and the bacterial cell. The medium was lyophilized, 1 mL methanol was added, the mixture was sonicated in a Bandelin Sonorex RK255 ultrasonic bath for 30 minutes, and centrifuged at 10,000 x *g* for 5 minutes. The methanol was removed, and the pellet was resuspended in 1 mL deionized water. After a final centrifugation at 10,000 x *g* for 5 minutes, the sample was ready to be injected to the UPLC-HRMS system. To prepare the sample from the cell pellet, 500 µL of methanol was added prior to sonication for 30 minutes. Then, 500 µL of deionized water was added, and the sonication was continued for another 15 minutes. Thereafter, the solution was centrifuged to pellet the insoluble part and the supernatant was injected to the UPLC-HRMS system.

The UPLC-HRMS system was an Agilent Infinity 1290 UPLC system equipped with an Acquity UPLC BEH C18 1.7 µm (2.1x100 mm) column (Waters) and an Acquity UPLC BEH C18 1.7 µm VanGuard Pre-Column (2.1x5 mm; Waters) setup coupled to a DAD detector and a micrOTOFQ II mass spectrometer (Bruker.). The LC part was run using a gradient (A: H₂O, 0.1% FA; B: MeCN, 0.1% FA; Flow: 600 µL/min): 0 min: 95%A; 0.80 min: 95%A; 18.70 min: 4.75%A; 18.80 min: 0%A; 23.00 min: 0%A; 23.10 min: 95%A; 25.00 min: 95%A and the column oven temperature was set to 45°C. MS parameters were as follows: nebulizer gas 1.6 bar; gas temperature, 200°C; gas flow, 8 L/min; capillary voltage, 4500 V; endplate offset, 500 V; measurement was done in positive ion mode.

A DAR standard curve was generated by plotting the peak area of DAR from the extracted ion chromatogram (EIC) (for the *m/z* of 483.7089 and 475.1956 ± 0.01) to a series of DAR concentrations (2, 3, 4, 5, 10, 15, 20, 30, 40 mg/L). The DAR concentration from a heterologous expression culture was quantified by calculating the peak area and interpolating it to the DAR standard curve. The linear range for

this quantification method was 3 µg/mL to 30 µg/mL. Therefore, the peak area below the border was not converted to concentration. The standard curve was measured with all batches that were analyzed by UPLC-HRMS to exclude technical differences between measurements.

IV.2.8. Determination of *E. coli* survival rate in the presence of DAR

Overnight cultures of *E. coli* strains were diluted 1:100 with LB-chloramphenicol and cultivated at 30°C with 180 rpm shaking. After reaching an optical density at 600 nm (OD₆₀₀) of 0.6 – 0.7, these cultures were further diluted to OD₆₀₀ of 0.001, 0.1, and 0.5. The diluted cultures were pipetted into eight wells in a 96-well plate, 100 µL final volume in each well. Different concentrations of antibiotics were adjusted. The first three wells: 6 µg/mL DAR, the second three wells: 30 µg/mL DAR, the seventh well: 128 µg/mL gentamycin as negative control, and the eighth well was not treated with an antibiotic as a positive control of the bacterial growth. The plate was incubated overnight at 30°C with 180 rpm shaking, and OD₆₀₀ was measured using a Tecan infinite®200 plate reader (Tecan Group Ltd., Männedorf, Switzerland). The survival rate was calculated by the following equation:

$$\frac{OD_{600} \text{ of sample} - OD_{600} \text{ of negative control}}{OD_{600} \text{ of positive control} - OD_{600} \text{ of negative control}} \times 100\%$$

IV.2.9. Expression of DarA

Our first approach to heterologously express *darA* as a His-tagged version was unsuccessful. Therefore, a Maltose Binding Protein (MBP)-encoding sequence was fused to the upstream region. First, pET24c was digested with *NdeI* (New England Biolabs) overnight and dephosphorylated (Thermo Scientific) for 1 hour. Then, the solution was directly purified using the Large fragment Zymo-research KIT yielding pET24c-*NdeI* as the 1st DNA fragment. The MBP sequence was amplified by PCR using pMAL-c5x (New England Biolabs) as template and primers Daro.MBP.Xa.F and Daro.MBP.Xa.R. Subsequently, it was purified by agarose gel purification, yielding an 1181 bp fragment. The The gene *darA* was amplified by PCR using a synthetic codon-optimized DNA version (**Sequence S1**) with primers Daro.Xa.pro.F

and Daro.Xa.pro.R, yielding a 217 bp DNA fragment. Then, these three fragments were ligated via Gibson Assembly (GA), yielding pET24c.MBP.Xa.darA. Thereafter, the GA reaction was transformed to *E.coli* top10 cells via electroporation and colonies were selected on LB agar plates containing 50 µg/mL kanamycin. Clones were picked and tested by PCR using primers Daro.MBP.Xa.F and Daro.Xa.pro.R, yielding a 1378 bp fragment consisting of *MBP* and *darA*. Verified plasmid DNA was isolated from an overnight *E.coli* top10/ pET24c.MBP.Xa.darA in LB + Kan culture and subsequently transferred to *E.coli* BL21 as expression host.

The expression of the MBP-DarA fusion protein was performed following the manufacturer's instruction (NEB protein expression manual) with the following modifications. After induction, the culture was incubated at 18°C at 180 rpm overnight. Purification was done using amylose resin (New England Biolabs). Then, the protein was concentrated using Amicon ultra-15 columns (Merck).

IV.2.10. Trypsin digestion

Trypsin digestion of MBP-DarA was performed following manufacturer's instruction (Promega). Protein solution (50 µL) was mixed with the same volume of 100 mM ammonium bicarbonate. Cysteine reduction was done by addition of 1 µL 500 mM DTT and incubation at 56°C for 30 min. Thereafter, the sample was cooled down to room temperature and 2 µL of 750 mM iodoacetamide were added. Incubation was done in the dark at room temperature for 30 min. 5 µL of trypsin (1 µg/µL) were added, and the sample was incubate at 37°C overnight. Thereafter, 0.5% (vol/vol) TFA was added to the solution and 1 µL of the sample was applied onto pH paper to verify a pH <2. The sample was then centrifuged in a tabletop centrifuge at max speed for 2 min. Clear supernatant was transferred to a new 1.5 mL Eppendorf tube. Further, C18 material was added to a 200 µL pipette tip (stage tip) and the column was equilibrated with 20 µL MeOH (centrifugation at 1000 x *g*, 2 min). Then, 20 µL solution B (0.5% formic acid in 80/20% MeCN/water) was added to the stage tip and centrifugation was done at 1000 x *g*, 2 min. 20 µL solution A (0.5% formic acid in 5/95% MeCN/water) were added to the stage tip and centrifugation was done at 1000 x *g*, 2 min. The last step was repeated). Next, the trypsin-digested sample was added to the stage tip and centrifuged at 1000 x *g* for 4 min. The stage tip was then

washed with 20 μ L solution A and centrifuged again at 1000 x *g*, 2 min. The stage tip was then moved to a new 1.5 mL Eppendorf tube and 100 μ L of solution B were added, centrifuged at 1000 x *g*, 2 min. the resulting elution fraction was sent for LCMS analysis.

IV.2.11. Quantification of *darA* and *darE* expression level

Overnight cultures of *E. coli* Rosetta™(DE3) strains carrying pZW-ADC3, pZW-ADC5, pZW-ADC7, and pZW-ADC8 were diluted 100-fold with fresh LB supplemented with kanamycin and chloramphenicol, and cultivated at 30°C at 180 rpm shaking to reach an OD₆₀₀ of 0.4±0.05. Then, IPTG was added (final concentration 0.5 mM) and the cultures were incubated at 30°C at 180 rpm shaking overnight.

7.5 mL samples were taken from the induced overnight culture for RNA isolation using the Quick-RNA Fungal/Bacterial Miniprep Kit (Zymo Research) according to the manufacturer's protocol with modifications as following. The 7.5 mL samples were distributed to 5 tubes, each containing 1.5 mL culture for better cell lysis. The samples were combined again during the Zymo-Spin™IICRCColumn centrifugation step. As the final step, RNA was eluted in 50 μ L of nuclease-free water.

25 μ L of isolated RNA was then treated with 5 units of DNase I (Zymo Research) in 50 μ L reaction volume (40 mM Tris-HCl, pH 8.0, 10 mM NaCl, 6 mM MgCl₂, and 10 mM CaCl₂) for 45 min at room temperature, and the reaction was stopped by incubation at 65°C for 10 min. Thereafter, 50 μ L of 5 M ammonium acetate was added and mixed well before 300 μ L of 99% ethanol were added. The mixture was incubated overnight at -20°C for RNA precipitation. Then, it was centrifuged at 12000 rpm, 15 min, 4°C and the supernatant was removed. The pellet was washed with 99% ethanol, 70% ethanol and subsequently air-dried. The RNA pellet was dissolved in 40 μ L of nuclease-free water, then the concentration and purity was measured using a microvolume spectrophotometer (Eppendorf BioSpectrometer®). cDNA was synthesized using LunaScript® RT SuperMix Kit (New England Biolabs) from 1 μ g RNA for each sample according to manufacturer's protocol. The synthesis reaction was incubated as follows: 25°C for 2 min, 55°C for 15 min, 95°C for 1 min. cDNA was directly used in the qPCR reaction and the same procedure was also done for the no-reverse transcriptase (no-RT) control.

For qPCR, the primer pair A-F3 and A-R3 was designed to target *darA* and the primer pair E-F3 and E-R3 was designed for *darE*. Primer pairs amplified 108 bp and 212 bp fragments, respectively. *rrsA* was used as reference gene, amplified by the primer pair *rrsA*-F3 and *rrsA*-R3 to generate 105 bp amplicons (Zhou *et al.*, 2011). Each sample was analyzed for *darA*, *darE*, and *rrsA* in 5-fold dilution series in triplicates. The qPCR reaction (10 μ L volume) contained 1x Luna Universal qPCR Master Mix (New England Biolabs), 0.25 μ M forward primer, 0.25 μ M reverse primer, and 1 μ L of cDNA. No-RT control was also done in 5-fold dilution series, and no template control (NTC) was analyzed in triplicates (qPCR plate scheme design is shown in **Table S5**). The qPCR was performed in StepOnePlus™ instrument (Applied Biosystems) with SYBR®Green detection, using the following thermocycler run: 10 min at 95°C, followed by 40 cycles of 15 s at 95°C and 1 min at 60°C. At the end of the run, a melt curve was measured by the following run: 15 s at 95°C, 1 min at 60°C, then increasing the temperature from 60°C to 95°C with a thermal ramp rate of 0.3°C/s.

StepOne™ Software v2.3 was used to analyze the qPCR data and to determine quantification cycle (C_q). C_q was then plotted against the cDNA dilution series to create a standard curve for relative quantification of each gene. The qPCR efficiency (E) was calculated by the following equation:

$$E = 10^{\left(-\frac{1}{\text{slope}}\right)}$$

Relative amount of each gene was then calculated by the following equation:

$$\text{Relative amount} = E^{-C_q}$$

The *darA:darE* transcript ratio was calculated from the relative amount of each gene.

IV.3. Results

The starting point of this project was the identification of 6.2 kb BGC from the genomic DNA of *P. khanii* DSM3369, which can be linked to DAR biosynthesis by knockout experiments (Imai *et al.*, 2019). However, to proof that the gene cluster is complete and functional to produce DAR, a heterologous expression approach was necessary. Thereafter, some aspects that affected heterologous expression was investigated to increase DAR production yield, such as the selection of heterologous expression hosts, generation of expression host with increased DAR resistance level, comparison of DAR BGCs from different sources, effect of the transporter genes, involvement of dedicated peptidase, and integration of additional *darA* copy and intergenic regions.

For the optimization of heterologous expression of DAR, HR-LCMS analysis was used to quantify the DAR yield. The HR-LCMS chromatogram of DAR showed two main peaks: (i) m/z 483.7089 (calcd.) that corresponds to $[M+2H]^{2+}$ and (ii) m/z 475.1956 (calcd.) that corresponds to $[M+H-NH_2]^{2+}$ (**Figure IV.4**). Therefore, an EIC was generated from the combination of these two masses, and the peak area was integrated. The DAR concentration was calculated by interpolating the peak area on the DAR standard curve.

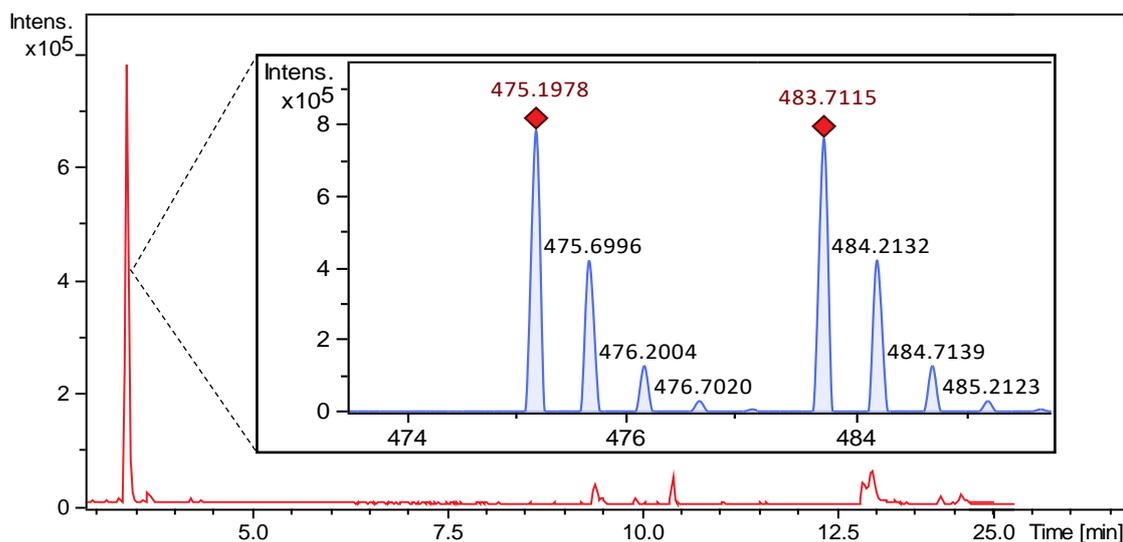


Figure IV.4: DAR mass spectra. DAR peak in HR-LCMS ionized into two masses: m/z 483.7115 (theoretical value: m/z 483.7089, difference of 0.0026) that corresponds to $[M+2H]^{2+}$ and m/z 475.1978 (theoretical value: m/z 475.1956, difference 0.0022) that corresponds to $[M+H-NH_2]^{2+}$. For quantification, the EIC of these two masses was integrated.

IV.3.1. Functional prove of the DAR BGC in *E. coli*

As a first step, the expression plasmid pZW-ADC3 was generated. The genes of the BGC were transferred into a 'streamlined' BGC, meaning all intergenic regions were removed, all genes have the same orientation and are under the control of the same promoter, *i.e.* T7-*lac*. This construct was transferred into *E. coli* BL21(DE3) as the heterologous host, and then cultivated for three days with the presence of IPTG as the inducer of T7-*lac* promoter. Cultivation of this heterologous expression system could produce DAR and confirmed that the BGC is complete and functional (Figure IV.5).

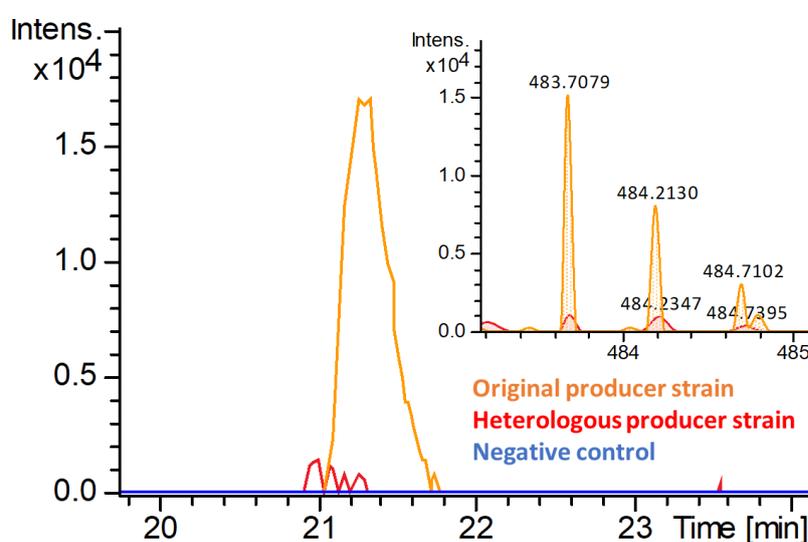


Figure IV.5: Extracted Ion Chromatograms (EICs) of natural and heterologous DAR expression. DAR BGC is functional to produce DAR in the heterologous expression system. Orange shows DAR peak from the original producer strain, *P. khanii* HGB1456 after 10 days of cultivation. Red shows DAR peak from the heterologous producer strain *E. coli* BL21(DE3)/pZW-ADC3 that was induced with 0.5 mM IPTG and cultivated for three days. Blue shows the heterologous expression control without the addition of IPTG in the fermentation.

However, the resulting DAR yield with this system was much lower than that of the native producer strains *P. khanii* HGB1456 and *P. khanii* DSM3369. Therefore, we tried to compare with other promoters, including inducible and constitutive ones. The *gfp* reporter gene was cloned downstream of the DAR BGC to enable a fast and easy readout. It was observed that green fluorescence correlated with the inhibition zone against a DAR-sensitive test strain (Figure S4). Out of four promoters, *i.e.* T5, J23101, T7-*lac* and *prpB* (the latter propionate-inducible), the

T7-*lac* promoter proved to produce the highest DAR levels (**Figure S4**). The T7 RNA polymerase system under control of the *lac* operator generally has a higher level of protein expression compared to the other systems, due to the selective target of the T7 RNA polymerase for the T7 promoter sequence (Studier and Moffatt, 1986). Therefore, T7-*lac* promoter will be used in the following experiments.

IV.3.2. Selection of the heterologous producer strain

As a next step to improve the DAR production yield, the construct pZW-ADC3 was also transferred into other common *E. coli* strains, *i.e.* *E. coli* BAP1 and *E. coli* Rosetta™(DE3), as well as into *Vibrio natriegens* Vmax™Express as expression host. *E. coli* BL21(DE3) is the most widely used strain for protein expression that has a prophage carrying a chromosomally encoded T7 RNA polymerase under the control of *lacUV5* (Studier and Moffatt, 1986). *E. coli* BAP1 is a derivative of *E. coli* BL21(DE3) that has been genetically modified to produce complex polyketides and non-ribosomal peptides (Pfeifer *et al.*, 2001a). Even though DAR is not synthesized by such a system, this BL21(DE3) derivative was tested as well. *E. coli* Rosetta™(DE3) is a further derivative of *E. coli* BL21(DE3), which carries a plasmid that encodes tRNA genes for rare codons. *Vibrio natriegens* Vmax™Express is an optimized *Vibrio natriegens* strain for protein expression of genes regulated by the T7 promoter. It should be mentioned that *in silico* analysis revealed the presence of the DAR BGC in several *Vibrio* strains (Imai *et al.*, 2019).

Fermentation vessels of all heterologous hosts were sampled from 1-8 days. The strain *E. coli* BL21(DE3) produced the least amount of DAR during the complete fermentation time, on average 5 mg/L were detected from the cell pellet and 4.3 mg/L from the medium (**Figure IV.6**). On the other hand, DAR could barely be detected from the cell pellet of *V. natriegens* and could only be observed from the medium. The DAR yield from *V. natriegens* showed a clear increase from day 1 to day 4, with no further increase thereafter. Compared to the other two *E. coli* strains, the DAR yield from *V. natriegens* was the least, with a maximum yield of 11.6 mg/L that was reached on the fourth day of fermentation. *E. coli* BAP1 and *E. coli* Rosetta™(DE3) produced the similar amount of DAR on the first day, but the yield increment of *E. coli* Rosetta™(DE3) was higher. When the yield of DAR in the medium and the cell pellet was combined, we achieved the highest production by

E. coli Rosetta™(DE3) (26.1 mg/L) on the seventh day of fermentation. Therefore, *E. coli* Rosetta™(DE3) was selected as the heterologous expression host for the DAR BGC in the following experiments.

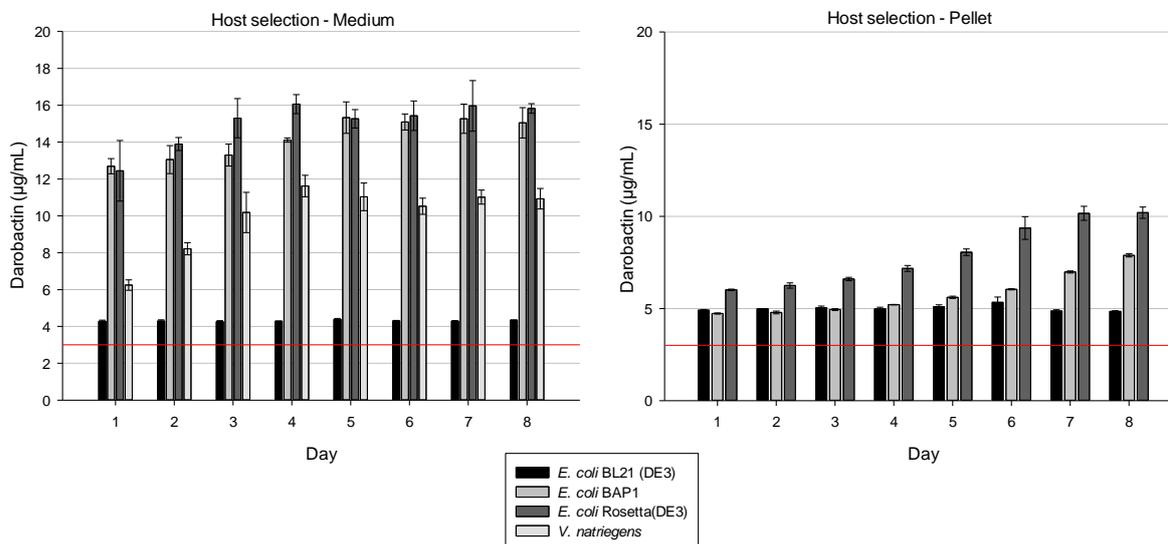


Figure IV.6: DAR expression level from different heterologous hosts. Expression plasmid pZW-ADC3 was introduced to four different heterologous host, *i.e.* *E. coli* BL21(DE3), *E. coli* BAP1, *E. coli* Rosetta™(DE3), and *Vibrio natriegens* Vmax™ Express. *E. coli* Rosetta™(DE3) was producing the highest yield in the medium and in the cell pellet. The red line marks the border of the linear range of the calibration curve. Therefore, values below the border are not shown. Data were collected from three biological replicates at day 1 to 8; error bars show standard deviation.

It should be mentioned that in later experiments it was investigated if the difference in production yield between *E. coli* BL21(DE3) and Rosetta™(DE3) might be due to the rare codons present in the DAR BGC. Therefore, the construct pZW-ADC12 that carries codon-optimized versions of *darA* and *darE* (**Sequence S1** and **S2**) was generated. This plasmid was introduced into the heterologous hosts *E. coli* BL21(DE3) and *E. coli* Rosetta™(DE3), and DAR production was measured. Using this construct, a comparable amount of DAR could be detected in both strains, in the medium as well as in the cell pellet (**Figure S5**).

IV.3.3. Generation of an *E. coli* strain with increased DAR resistance level

Considering that *E. coli* cells are sensitive against the antibiotic to be biosynthesized, a producer strain with an increased resistance level towards DAR was generated. As observed by passaging experiments, DAR resistant strains could be generated. These strains accumulated mutations, *i.e.* more than one, in the *bamA* gene. A high-level resistant strain possessed three point mutations in *bamA*. By inserting these three point mutations into a clean background *E. coli* strain, it was proven that these create a high-level resistant phenotype (Imai *et al.*, 2019). Therefore, the mutations were introduced to *E. coli* Rosetta™(DE3) to create *E. coli* Rosetta™(DE3)(Dar^R). The DAR minimum inhibitory concentration (MIC) was determined for both strains to validate the results. Indeed, growth of the original *E. coli* Rosetta™(DE3) strain was inhibited at 2 µg/mL and the *E. coli* Rosetta™(DE3)(Dar^R) strain was able to grow at DAR concentrations >64 µg/mL. Here, it has to be noted that the MIC determination was done for *E. coli* strains under standard conditions, using an OD₆₀₀ of 0.001, while the heterologous expression of DAR was induced when the OD₆₀₀ has reached 0.4 – 0.6. Therefore, the parent strain without an increased DAR resistance level only had a survival rate of 18% (no growth was visible by eye), when 6 µg/mL DAR were introduced at an OD₆₀₀ of 0.001. However, the strain could maintain 57% growth when the same amount of DAR was added to a culture with an OD₆₀₀ of 0.5 (**Figure S6**).

Plasmid pZW-ADC3 was then introduced to *E. coli* Rosetta™(DE3)(Dar^R), and the DAR production was compared to *E. coli* Rosetta™(DE3). However, the original *E. coli* Rosetta™(DE3) strain could produce almost two-fold DAR compared to the strain that has increased resistance to DAR (**Figure IV.7**).

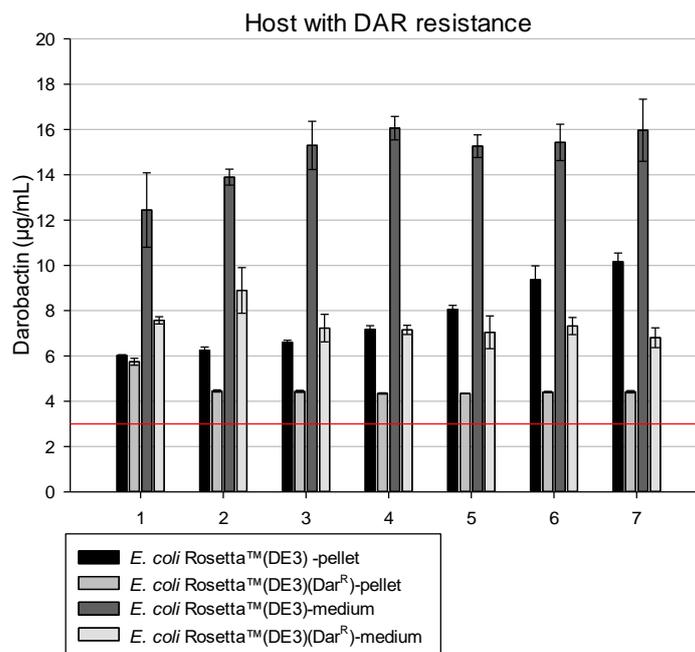


Figure IV.7: DAR production in *E. coli* Rosetta™(DE3) with DAR resistance. Heterologous expression of the streamlined DAR BGC in *E. coli* Rosetta™(DE3)/ pZW-ADC3 and *E. coli* Rosetta™(DE3)(Dar^R)/pZW-ADC3. The original *E. coli* Rosetta™(DE3) strain produced about two-fold more DAR as *E. coli* Rosetta™(DE3)(Dar^R). Data were collected from three biological replicates at day 1 to 7; error bars show standard deviation. The red line is the linearity limit of the DAR standard curve.

We observed that during the cultivation period the strain *E. coli* Rosetta™(DE3)(Dar^R) was not growing as well as *E. coli* Rosetta™(DE3). To test if the difference in DAR production is due to impaired growth, growth curves of these strains were generated. This revealed that the *E. coli* strain with increased resistance to DAR had an impaired growth compared to its parent strain (**Figure IV.8**). The *E. coli* Rosetta™(DE3) strain had a significantly longer log phase (up to 54 hours after inoculation), while the DAR resistant strain has already started its stationary phase at 25 hours after inoculation. This means that the parent strain reached a cell density twice as high as the DAR resistant strain. In addition to that, the *E. coli* strains that carry the empty pRSFDuet™-1 vector control were cultivated, and the growth curve showed a similar trend, i.e. growth of *E. coli* Rosetta™(DE3)(Dar^R) was more hindered compared to *E. coli* Rosetta™(DE3) (**Figure S7**).

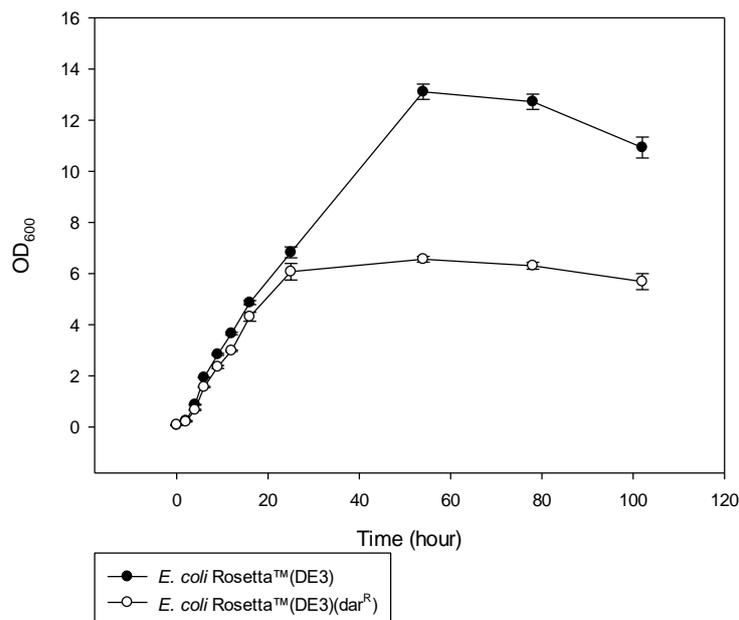


Figure IV.8: Growth curve of *E. coli* Rosetta™(DE3) and *E. coli* Rosetta™(DE3)(Dar^R). Optical density (OD) was measured at 600 nm to compare the growth of *E. coli* Rosetta™(DE3) and *E. coli* Rosetta™(DE3)(Dar^R). The *E. coli* strain with increased resistance to DAR has a shorter log phase compared to its parent strain and does not reach the same OD. Data were collected from three biological replicates at the time points given; error bars show the standard deviation.

IV.3.4. Comparison of different DAR BGCs

It was observed before that DAR production yields of *Photorhabdus khanii* DSM3369 were two-fold higher compared to *P. khanii* HGB1456. Furthermore, *in silico* analysis revealed the presence of DAR BGCs in *Yersinia* strains (Imai *et al.*, 2019). Therefore, we cloned the native DAR BGC from *P. khanii* HGB1456, *P. khanii* DSM3369 and *Yersinia frederiksenii* ATCC33641, yielding pZW-ADC5, pZW-ADC6 and pZW-YerA4 respectively. All constructs were transferred into the selected heterologous host, *E. coli* Rosetta™(DE3).

In the heterologous expression system, the production yield difference between the DAR BGCs from both *Photorhabdus* strains were not so prominent. Both strains produced over 20 mg/L already after one-day cultivation and the strain carrying pZW-ADC5 reached the highest yield on the fourth cultivation day, with a total of 27.7 mg/L. Meanwhile, DAR production in the strain carrying pZW-ADC6 was in the same range and remained steady on average of 24.9 mg/L throughout the third to seventh cultivation day (Figure IV.9). On the other hand, *E. coli* Rosetta™(DE3)

carrying pZW-YerA4 produced only a total of 10.3 mg/L DAR on the second day of cultivation, which was 2.7-fold less, compared to the maximum yield of pZW-ADC5.

A comparison of the DAR BGCs from both *Photorhabdus* strains on DNA sequence level revealed 99% identity between both BGCs. The most evident difference between the two BGCs is an additional 136 bp intergenic region between the *darA* and *darB* genes in *P. khanii* DSM3369 (**Figure S8**). To test if this intergenic region will have an impact on DAR production levels, we compared the native DAR BGC from *P. khanii* HGB1456 (pZW-ADC5) to the streamlined BGC, in which the intergenic region was removed (pZW-ADC3). In general, DAR yield is higher in medium than in pellet (**Figure IV.9**). In the cell pellet, DAR production from the host with pZW-ADC3 was accumulating throughout the cultivation period with the most significant difference to pZW-ADC5 occurred on the seventh day. In medium, pZW-ADC5 produced better than pZW-ADC3 and achieved its highest production peak on the fourth day. Overall, pZW-ADC5 could produce more DAR in a shorter period compared to pZW-ADC3 and was used for further experiments.

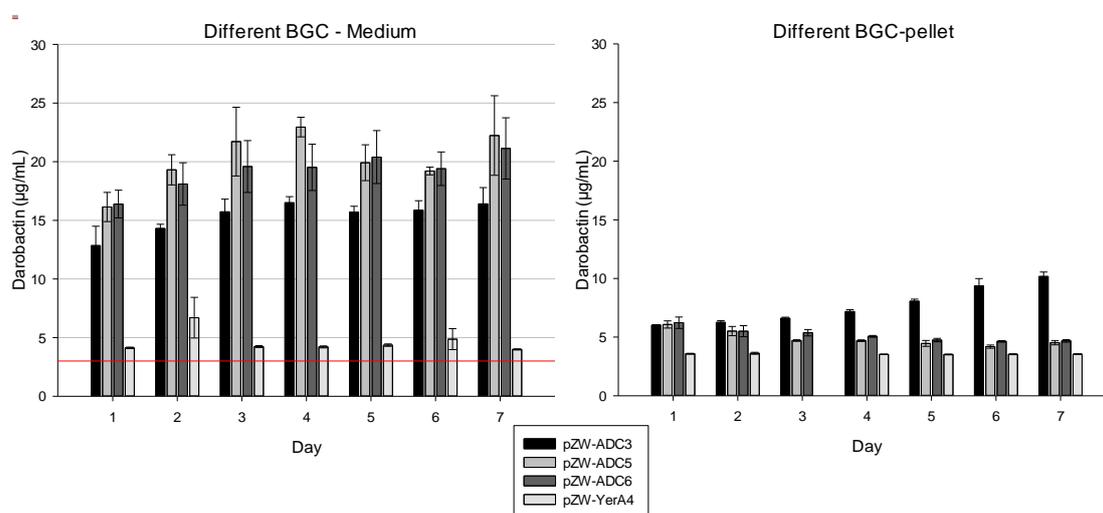


Figure IV.9: Bar diagram showing DAR production of different BGCs in *E. coli* Rosetta™(DE3). Heterologous expression of the streamlined DAR BGC from *P. khanii* HGB1456 (pZW-ADC3), the native BGC from *P. khanii* HGB1456 (pZW-ADC5), the native BGC from *P. khanii* DSM3369 (pZW-ADC6), and the native BGC from *Y. frederiksenii* ATCC33641 (pZW-YerA4) in *E. coli* Rosetta™(DE3). Data were collected from three biological replicates at day 1 to 7; error bars show standard deviation. The red line is the linearity limit of the DAR standard curve. Values below this line are not given. Therefore, the data point of the third day of pZW-YerA4 was not converted to concentration. Expression of the native BGCs from *P. khanii* HGB1456 and of *P. khanii* DSM3369 were in the same range. Heterologous expression of the *Yersinia*-derived BGC resulted also in DAR production, although in much lower amount.

IV.3.5. Influence of the transporter-encoding genes *darBCD*

The three genes *darB*, *darC* and *darD* are coding for subunits of an ABC transporter. To answer the question whether these transporter genes play an additional role to the biosynthesis of DAR and to define the minimum DAR BGC, these genes were removed from the expression construct. Therefore, pZW-ADC9, a construct that carries only *darA* and *darE* from *P. khanii* DSM3369, was created. This experiment showed that without *darBCD*, DAR could still be produced. DAR accumulation in the medium of pZW-ADC6 was higher ($\pm 50\%$) than in pZW-ADC9. However, even though pZW-ADC6 possessed the transporter genes, the accumulation of DAR in the pellet was slightly higher than in the strain carrying pZW-ADC9 (**Figure IV.10**).

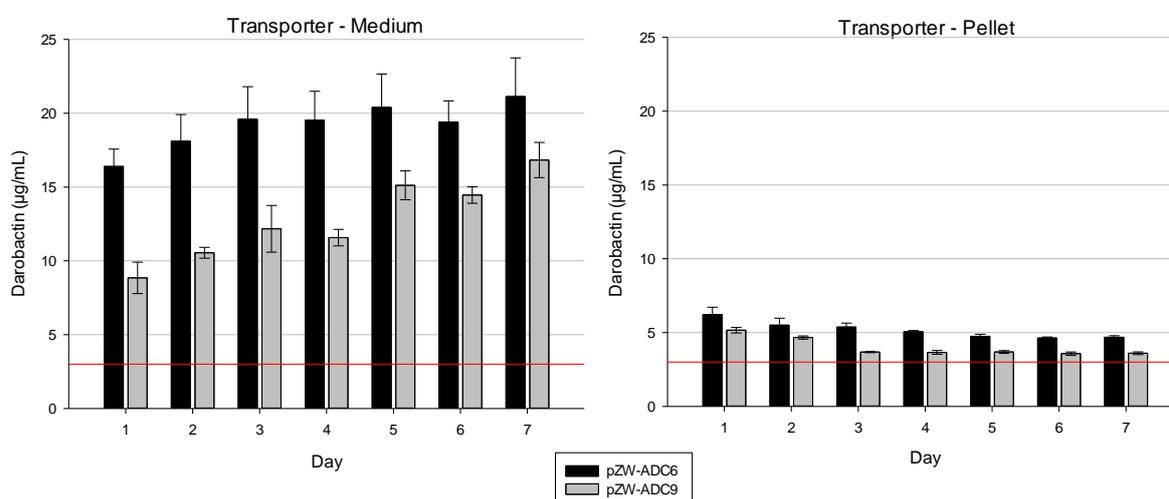


Figure IV.10: Comparison of DAR production from constructs that carry transporter genes and without transporter genes. pZW-ADC6 is a construct with the native DAR BGC from *P. khanii* DSM3369, and pZW-ADC9 is constructed from pZW-ADC6 omitting *darBCD*. Red line is the linearity limit of DAR standard curve. Data were collected from three biological replicates at day 1 to 7; error bars show standard deviation. DAR was detectable outside of the cells without *darBCD*.

The genes *darBCD* encode for proteins that compose a tripartite efflux pump (TEP), which is suspected to work together with the outer membrane efflux protein TolC. To investigate if TolC is involved in the export of DAR into the medium, a *tolC* knockout strain (*E. coli* $\Delta tolC$) was created and the plasmid pZW-ADC5 was transferred into this host. However, there was no significant difference detectable (i) in DAR production and (ii) in the ratio between the intra- and extracellular concentration. (**Figure S9**).

IV.3.6. A dedicated peptidase is not part of the BGC

To obtain DAR from the precursor DarA, it has to be modified and cleaved twice, upstream and downstream of the core heptapeptide. DarE introduces the modification, *i.e.* the formation of the bicyclic ring structure, and proteolytic processing is usually achieved by one or more peptidase(s). However, we did not detect any dedicated peptidase within the DAR BGC. Therefore, it was in question whether the peptidic cleavage is also catalyzed by DarE, or if a peptidase from the host strain can catalyze this proteolysis. To get insights into this, DarA alone was heterologously expressed and purified in *E. coli*. An His-tagged version of DarA could not be obtained. This is commonly observed in heterologous expression of RiPP precursor proteins, most probably due to proteolytic instability. Fusion of a MBP (Maltose Binding Protein) to the N-terminal side improved stability and enabled purification using an amylose resin (**Figure S10**). Then, the purified MBP-DarA was digested with trypsin and analyzed by LCMS (**Figure IV.11**).

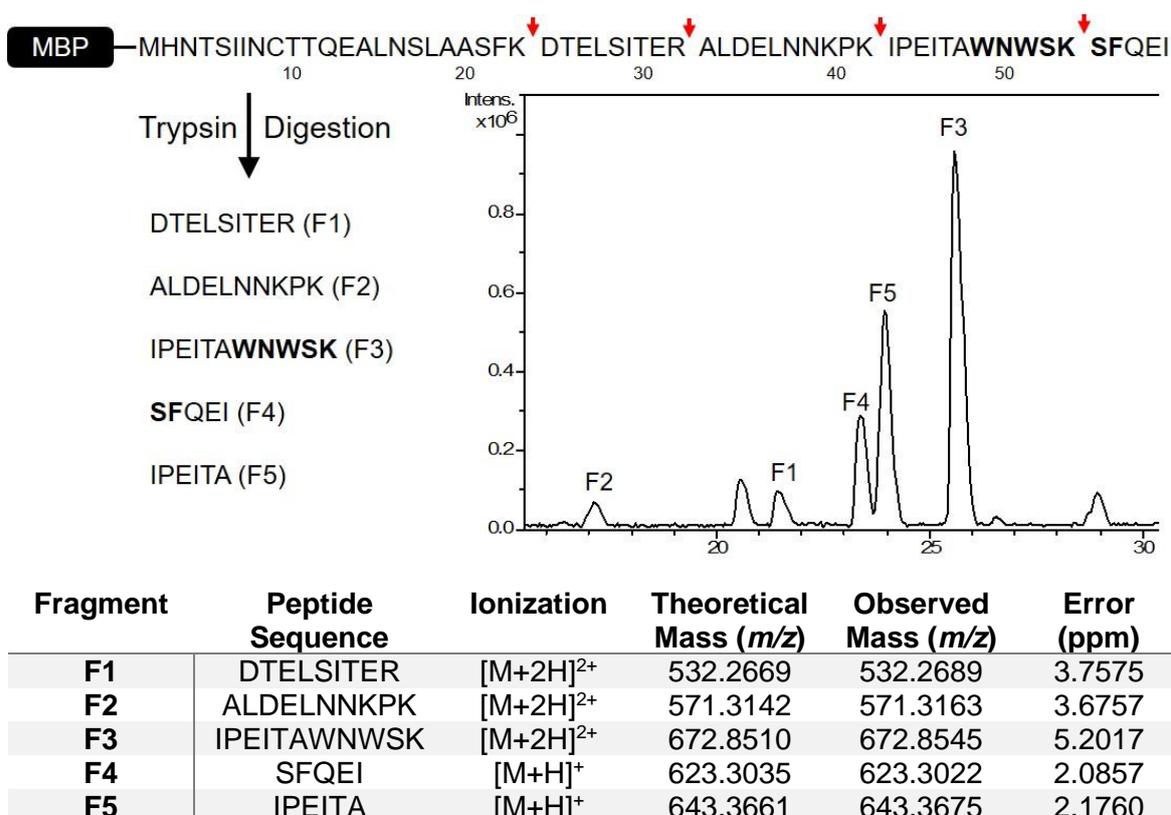


Figure IV.11: Scheme and LCMS chromatogram of trypsin-digested MBP-DarA. Red arrows represent trypsin-cutting sites; the core peptide of DarA is in bold. The table shows the fragments including calculated and observed m/z ratio.

In addition to the predicted trypsin-digested fragments (F1, F2, F3 and F4), the LCMS analysis showed another abundant fragment (F5). This is in accordance with a part of F3 that was cleaved exactly at the position where the core peptide should be released from the leader peptide. Since only DarA was introduced into the *E. coli* host strain, it can be concluded that this proteolytic activity relies on a peptidase from the host and not by any of DarBCDE. The cleavage of the C-terminal amino acids might be also catalyzed by an *E. coli* peptidase. However, these end-standing amino acids were not observed. The peptidase(s) processing DarA remain elusive and by this experiment, the option that the peptide is catalyzing a self-cleavage cannot be ruled out. Furthermore, to test if the last three C-terminal amino acids are essential for heterologous expression it was envisaged if these can be omitted. Therefore, we analyzed the expression constructs carrying natural DarA, or a variant in which the last three amino acids were removed. There were no significant changes in production observed (**Figure S11**).

IV.3.7. Integration of an additional *darA* copy

Since *darA* encodes for the precursor peptide that will be modified to DAR, it was investigated whether having two copies of *darA* will improve the DAR yield. A second copy of the *darA* gene was introduced to pZW-ADC3 and pZW-ADC5, creating pZW-ADC3.2 and pZW-ADC5.2. If the intergenic region between *darA* and *darB* was not present in the expression construct, indeed more DAR was produced when a second *darA* copy was introduced (pZWADC3.2). The increase of DAR yield due to the addition of the second *darA* copy was evident in the cell pellet as well as in the medium. On the third day of cultivation, twice the amount of DAR could be detected in the medium of *E. coli* Rosetta™(DE3) pZW-ADC3.2 in comparison to pZW-ADC3. On the other hand, if the native BGC as it is present in *Photorhabdus* was used and an additional *darA* gene was added to this construct (pZW-ADC5.2), the maximum DAR yield that could be obtained was 17.5 mg/L. This is about two third of the maximum yield of *E. coli* Rosetta™(DE3) pZW-ADC5 that reached 27.7 mg/L (**Figure IV.12**).

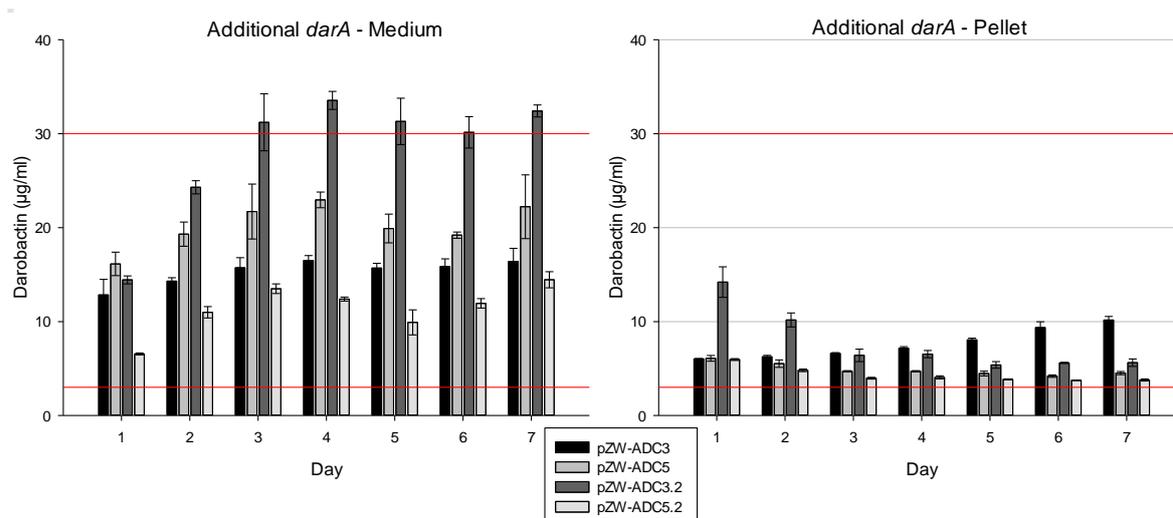


Figure IV.12: DAR production in the medium and cell pellet for strains carrying an additional *darA* copy. Bar diagram showing the DAR production from *E. coli* Rosetta™(DE3) that carries streamlined DAR BGC with a single (pZW-ADC3) and double copy of *darA* (pZW-ADC3.2), as well as native DAR BGC from *P. khanii* HGB1456 with a single (pZW-ADC5) and double copy of *darA* (pZW-ADC5.2). Red line is the linearity limit of DAR standard curve. Data were collected from three biological replicates at day 1 to 7; error bars show standard deviation. Double copy of *darA* boosted DAR production in the streamlined DAR BGC, but halved the DAR production from the DAR BGC with intergenic region between *darA* and *darB*.

IV.3.8. Integration of native *darA* upstream regions

It was observed that in the streamlined construct an additional *darA* copy could double the yield; while in the construct based on the native BGC, this effect was not so prominent. To get an additional data point, in a next step the upstream region *darA* was integrated into the expression construct. To study the effect of putative regulatory elements present in this region concerning heterologous DAR production, two additional constructs were created. The *in silico* analysis of the region up- and downstream *darA* with bacterial promoter prediction tools, e.g. BPROM (Victor Solovyev and Salamov, 2011) and CNNprom (Umarov and Solovyev, 2017a) revealed several putative transcriptional factor binding sites (TFBS) (**Figure S12**) and housekeeping sigma factor binding sites (RpoD16 and RpoD17, Gruber and Gross, 2003). Furthermore, the presence of LexA and FNR (transcriptional repressor at non-stress and aerobic condition, respectively; Butala *et al.*, 2009 and Unden and Schirawski, 1997) binding sites between *darA* and *darB* were detected. Beside LexA, a H-NS (global transcriptional silencer of genes with

high AT content; Dorman, 2004) binding site was predicted 393 bp upstream of *darA* (Figure S12).

Hence, 200 bp and 605 bp located upstream of *darA* were added to pZW-ADC5 yielding pZW-ADC7 and pZW-ADC8, respectively. Interestingly, the DAR production was doubled in pZW-ADC7 in comparison to pZW-ADC5. In contrast, including a 605 bp long region upstream of *darA* in the expression construct (pZW-ADC8) resulted in a DAR production yield half as much as pZW-ADC5 (Figure IV.13).

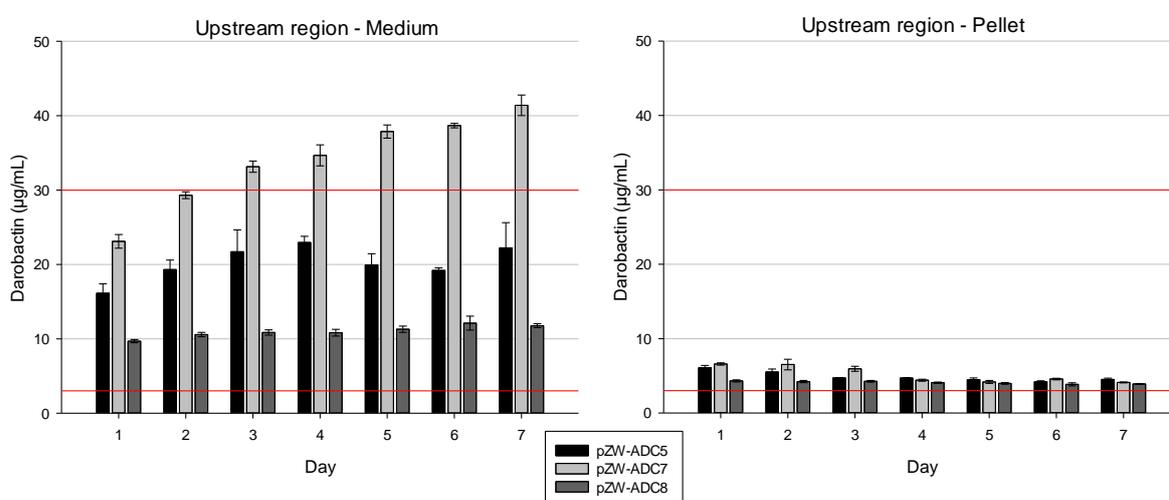


Figure IV.13: Comparison of constructs carrying different length of *darA* upstream regions of *Photorhabdus khanii* HGB1456. Bar diagram depicted DAR production from *E. coli* Rosetta™(DE3) that carries DAR BGC without upstream region (pZW-ADC5), with 200 bp upstream region (pZW-ADC7), and with 605 bp upstream region (pZW-ADC8). Data were collected from three biological replicates at day 1 to 7; error bars show standard deviation. The red line is the linearity limit of the DAR standard curve. It can be seen that the 200 bp upstream region improves heterologous expression.

IV.3.9. Quantification of *darA* and *darE* transcription

From these experiments, including upstream regions or additional gene copies, it became clear that the *darA* transcript level should have an impact on DAR production. Hence, quantification of *darA* and *darE* transcript level was aimed by qPCR experiments. This revealed that pZW-ADC3, which has no intergenic region, had a *darA:darE* transcription ratio of 6.89. Meanwhile, pZW-ADC5 that has an intergenic region between *darA* and *darB* showed a ratio of 23.56 and produced higher DAR levels than pZW-ADC3.

Next, all constructs that have the intergenic region between *darA* and *darB*, but carry different lengths of the *darA* upstream region (pZW-ADC5, pZW-ADC7, and pZW-ADC8) were compared. Also here, total DAR production indicated a relation to the *darA:darE* transcript ratio. Plasmid pZW-ADC7 that produced the highest amount of DAR (29.7 mg/L) had the highest *darA:darE* transcript ratio of 37.42. On the other hand, pZW-ADC8, which resulted in the lowest DAR production among the three constructs, had the lowest *darA:darE* transcript ratio of 17.05 (Figure IV.14, Table S6). The overall transcript level of *darA* had no influence of the observed DAR yield in these experiments.

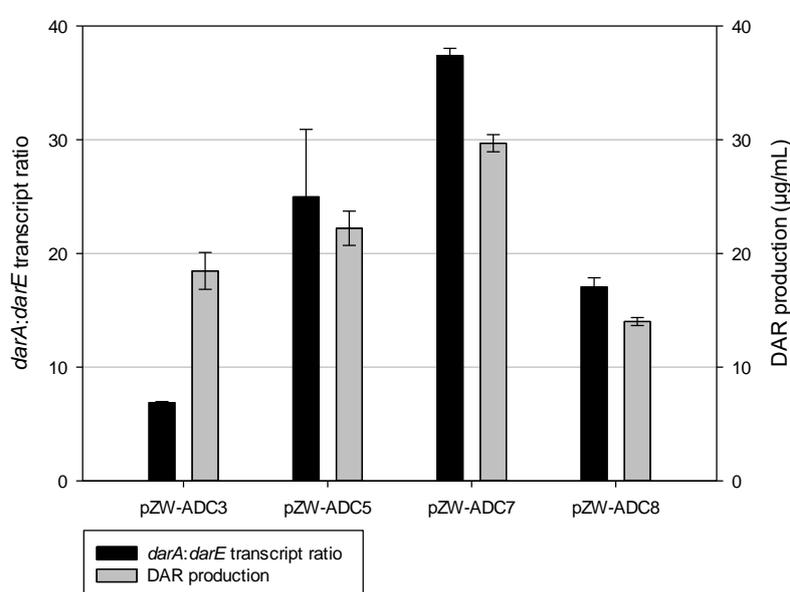


Figure IV.14: Ratio of *darA:darE* transcripts from expression constructs with different non-coding regions in relation to their DAR production. Construct pZW-ADC7, which gave the highest DAR yield, showed the highest *darA:darE* ratio. Transcripts of *darA* and *darE* were quantified by qPCR in triplicates and DAR production was measured from three biological replicates; error bars show standard deviation.

IV.4. Discussion

DAR was recently discovered as a natural product most promising for further development as antibiotic lead structure (Imai *et al.*, 2019). A RiPP BGC from *P. khanii* DSM3369 could be linked to DAR, since knocking out the BGC abolished DAR production. Here, the final confirmation of the BGC was achieved by heterologous expression in *E. coli* BL21(DE3), which resulted in production of DAR. This result proves that the BGC that contains five genes, *darABCDE*, are sufficient and functional to produce DAR. Moreover, even though the production yield was low (<1 mg/L), it clearly indicated the general suitability of *E. coli* as expression host for DAR.

In the previous experiments for the discovery of DAR, it was hard to spot the initial activity of the compound in classic overlay assays, since it seemed that the product of interest was only poorly expressed. For further investigation of this compound, we aimed to increase the production yield to enable biotechnological studies, e.g. derivatization of the compound, in the future. Although the heterologous expression in *E. coli* BL21(DE3) could shorten the fermentation period to three days, the yield was still lower than the one of the native producer strains. Therefore, optimization of DAR heterologous expression system were envisaged.

First of all, an experiment was designed to discover a more suitable expression host. The originally identified producer strains, *i.e.* *Photorhabdus* species, are Proteobacteria of the order Enterobacterales. Therewith, *Vibrio natriegens* and especially *E. coli* strains still emerged as the first option, since these strains are Enterobacterales as well and are extensively studied and established molecular tools exist. *E. coli* as a promising host for heterologous expression of *Photorhabdus*-derived natural products was exemplified before, e.g. by the expression of luminmycin (Bian *et al.*, 2012) and anthraquinone (Zhou *et al.*, 2019) from *P. luminescens*, as well as phototemtide A from *P. temperata* (Zhao *et al.*, 2020). Among the host strains tested, the result showed that *Vibrio natriegens* was outperformed by *E. coli* strains, especially Rosetta™(DE3) which produced DAR the best (**Figure IV.6**). An explanation why *E. coli* Rosetta™(DE3) produced higher amounts of DAR in comparison to other related *E. coli* strains might be the presence of the pRARE plasmid, which carries the genes for rare tRNA. The *darE* gene sequence from *P. khanii* HGB1456 has 33 codons that are rarely used in *E. coli*.

The result that utilizing codon-optimized *darA* and *darE* resulted in similar production yields of *E. coli* BL21(DE3) and Rosetta™(DE3) supports this hypothesis (**Figure S5**). *E. coli* as an established workhorse of modern microbiology and biotechnology represents an excellent choice for future optimization, since reproducibility and process speed can be improved, downstream processing is facilitated by a reduced chemical background and simultaneously regulatory concerns are avoided if the project advances to GMP studies.

***bamA* mutation that confers resistance to DAR impaired the growth of heterologous host**

A point that has to be considered to select a heterologous host is that the product DAR is efficiently killing *E. coli* cells. Therefore, in preceding studies, DAR-resistant strains were generated and the resistance was traced back to mutations in the *bamA* gene. Hence, the reported three point mutations that confer high-level DAR resistance were introduced into the *bamA* gene of *E. coli* Rosetta™(DE3) to enable higher production titers without killing the producer strain in batch fermentation. BamA is an essential component of the β -barrel assembly machinery (BAM) complex that catalyze the assembly of outer membrane protein (OMP) in Gram-negative bacteria. Modifications in BamA affects the viability of the cell (Konovalova *et al.*, 2017; Wu *et al.*, 2005). The three point mutations in the *bamA* gene were also affecting the *E. coli* virulence, thereby indicating the reduced fitness of the mutant strain (Imai *et al.*, 2019). In the present study, it was observed that the generated mutant strain showed impaired growth compared to the strain with wild type *bamA*. In consequence, the lower viability of the DAR-resistant strain and the lower OD reached during fermentation, resulted in lower DAR production. This result indicates that having a well-functioning BamA for viability is more beneficial in DAR production than having a mutated BamA through DAR resistance. Therefore, we selected *E. coli* Rosetta™(DE3) with native *bamA* over *E. coli* Rosetta™(DE3) with mutated *bamA*.

Minimum DAR BGC consists of only *darA* and *darE*

In a next step, it was evaluated if the transporter-encoding genes *darBCD* are essential for the heterologous DAR expression or if *darA*, encoding the precursor peptide and *darE*, encoding the radical SAM modification enzyme (Imai *et al.*, 2019), are sufficient. By deletion of *darBCD*, DAR production was not abolished. However, the yield was lower (1.5-fold) than the one reached with the construct including transporter genes (pZW-ADC6). Most interestingly, DAR was also present in the medium, even without the transporter-encoding genes. On the one hand, this clearly defines the minimum DAR BGC, which consists of only *darA* and *darE*. On the other hand, it became clear that in *E. coli*, DAR is present outside the cell even without the specific heterologous transporter genes *darBCD*.

In general, in addition to *darBCD*, which encode for an ABC transporter permease protein, a periplasmic adaptor protein, and an ATP binding domain of the inner membrane ABC transporter, respectively, an outer membrane efflux protein, *e.g.* TolC, would be needed to transport a molecule to the outside. This is for example the case when molecules toxic to the cell are transported out as a form of antibiotic resistant mechanism in *E. coli* (Greene *et al.*, 2018). However, a Δ *tolC* *E. coli* mutant strain did not exhibit any difference in production level to the parent *E. coli* strain with a functional *tolC* gene, and maintained the ratio between intra- and extracellular DAR. This indicates that DAR is not actively transported to the outside of the cell by a DarBCD-TolC complex in the heterologous expression system. It can be assumed that in a natural system the transporters are needed to guarantee a regulated efflux. However, for the heterologous system, the presence of the transporters is not essential, but the transport system utilized is not known. Beside TolC, three other outer membrane efflux protein (OEP) were identified in *E. coli*, *i.e.* YjcP, YohG, and YlcB (Sulavik *et al.*, 2001). In addition to that, *E. coli* harbors TolC-independent efflux systems that belong to the major facilitator superfamily (MFS), small multidrug resistance family (SMR), multidrug and toxin extrusion (MATE) family, cation diffusion facilitator family (CDF), and proteobacterial antimicrobial compound efflux family (PACE) (Slipki *et al.*, 2018). Due to the size of DAR, it can be assumed that passive diffusion via porins does not take place.

The role of unspecific peptidase in the maturation of DarA to DAR

To release DAR from modified DarA by DarE, the modified core peptide in DarA needs to be cleaved. However, we did not find any dedicated peptidase from DAR BGC. Furthermore, we could show that DarBCD were not involved in the DAR biosynthesis. It can be hypothesized that either DarE might also catalyze a peptidolytic cleavage, or an unspecific peptidase from the host is involved. By expression of DarA fused to MBP without DarE and purification through an amylose column, we could observe that a fraction of DarA was already cleaved exactly before the core peptide (**Figure IV.11**). This result clearly favors the second explanation that the cleavage of DarA is not catalyzed by DarE. Most likely, a peptidase from the host can catalyze the reaction and DarA can be cleaved at this position even though no modification by DarE took place. This cleavage seems to be specific, since neither ions corresponding to peptides with the sequence IPEIT, nor to IPEITAW were detected. However, the SF (Ser-Phe) fragment, which should be liberated by the cleavage behind the core peptide sequence, was not detected. Either this cleavage is not catalyzed by a peptidase in *E. coli* that can be acquired, or a peptidase cuts first upstream of the core peptide. Then, this fragment will be lost during purification of MBP-fused peptides. Anyway, the fact that the end-standing amino acids can be omitted in expression constructs will facilitate the generation of derivatives, since a cleavage site must not be considered.

Native DAR BGC from different bacterial strains and the effect of non-coding region in DAR BGC to DAR production

Besides the BGC from *P. khanii* HGB1456 and *P. khani* DSM3369, a *Yersinia*-derived DAR BGC was chosen as an alternative source to be expressed in *E. coli* Rosetta™ (DE3), since the GC content is similar to *Photorhabdus* species. The DAR production result showed that DAR BGC from *Photorhabdus* strains resulted in a more than 3-fold higher DAR production compared to DAR BGC from *Y. frederiksenii* ATCC33641 (**Figure IV.9**). The DAR BGC of the two *Photorhabdus* strains is highly similar; it revealed 99.49% similarity on nucleotide level with 99% query coverage. Despite this high similarity, the intergenic region between *darA* and *darB* of the two *Photorhabdus* strains differs in length. *P. khanii* DSM3369 carries a 136 bp insertion that is lacking in *P. khanii* HGB1456 (**Figure S8**). This difference

seems to have a positive effect in heterologous production of DAR from *P. khanii* HGB1456 BGC. Meanwhile, the DAR BGC from *Y. frederiksenii* ATCC33641 showed only 63.64% identity with 31% query covered on nucleotide level compared to the DAR BGC from *P. khanii* HGB1456. Hence, the difference in DAR production from these different sources could be attributed to the difference in DAR BGC sequence.

Anyway, for the regulation of a BGC that consists mainly of a precursor and an enzyme necessary for the modification of the first a fine-tuned mechanism to adjust the ration between these two can be expected. In the bottromycin BGC for example, the precursor peptide is enhanced specifically by a regulatory protein that does not affect other genes in the BGC (Vior *et al.*, 2020). In the DAR BGCs, there is an intergenic non-coding region between *darA* and *darB*. It can be speculated that this region carries recognition sequences, which could play a role in regulating the expression level of *darA* and the genes located downstream of it, thus affecting DAR production. The importance of the transcript ratio of *darA* and *darE* was shown by comparison of pZW-ADC3 in which the intergenic region was removed and pZW-ADC5 as a control. Plasmid pZW-ADC5, which encodes production of higher DAR levels than pZW-ADC3, has a higher *darA:darE* ratio than pZW-ADC3 (**Figure IV.14**). This indicates that the intergenic region plays a role in adjusting this transcript ratio and the higher the ratio, the better the DAR production is. A further experiment to increase the *darA:darE* ratio, in which an additional *darA* copy was introduced to pZW-ADC3 (pZW-ADC3.2), increased DAR production as well (**Figure IV.12**). However, introducing an additional *darA* copy in pZW-ADC5 (pZW-ADC5.2 that carries the intergenic region) decreased DAR production (**Figure IV.12**). It seemed like the presence of the intergenic region regulates *darA:darE* transcription level and the system was put out of action by the artificial introduction of an additional *darA* copy.

Interestingly, the 200 bp region upstream of *darA* was quite conserved in the *Phototrhabdus* genus (94% identity and 70%-94% query coverage) and was therefore integrated into the expression construct, creating pZW-ADC7. In addition, the 605 bp upstream region of *darA* was also integrated in a construct, despite the fact that this region was less conserved throughout *Phototrhabdus* strains carrying the BGC (94% identity and 24%-64% query coverage). Integration of the 200 bp

upstream region significantly increased DAR production, while extending the integrated upstream region up to 605 bp diminished it (**Figure IV.13**). Consequently, the effect of the region upstream of *darA* towards the *darA:darE* transcript level was also investigated. Indeed, pZW-ADC7 showed the highest *darA:darE* ratio observed in this study. Hence, it should be concluded that the non-coding region upstream of *darA* has a positive regulatory effect on transcription level and DAR production. Thereby, supporting the hypothesis that a higher *darA:darE* transcript level goes in line with higher DAR production. Within the 200 bp upstream region of *darA*, three open reading frames (ORFs) with the size of 52 aa, 44 aa and 28 aa, respectively, were detected. The first two ORF were located in the opposite strand, while the last ORF was located on the same strand as the DAR BGC. The first two ORFs were only annotated as hypothetical proteins, while the 28 aa ORF had 92.9% identity to the partial sequence of *relE* (135 aa).

Several putative TFBS, housekeeping sigma factor binding sites (RpoD16 and RpoD17) were predicted up- and downstream *darA*. Regulation of a RiPP BGC by an RNA polymerase sigma factor (RpoD) was before reported for microcin “MccB17” (Duquesne *et al.*, 2007). Furthermore, the presence of LexA and FNR binding sites between *darA* and the following cluster parts enabled transcription repression of downstream genes; thereby, increasing the *darA:darE* transcript ratio. In addition, the presence of an H-NS binding site 393 bp upstream of *darA* could indicate a downregulation of *darABCDE* transcription, as it was shown for the construct carrying this region. Future detailed analysis of DAR BGC regulation will support the further fine-tuning of the optimal transcript ratio for heterologous expression.

IV.5. Summary

The heterologous expression of the DAR BGC provided the final confirmation that the BGC is complete and functional. Moreover, the heterologous expression system developed in this study forms the basis for future bioengineering approaches, which enable further exploitation of this compound class as a promising drug lead structure for antibiotic development. Using this platform, generation of novel derivatives, e.g. with improved pharmaceutical properties and altered activity profile becomes feasible. Heterologous expression of DAR was optimized to reach a higher yield in a shorter period compared to the original producer strains. The optimization strategies included the use of different expression hosts, comparison of different DAR BGCs originating from different strains, as well as modifications of the BGC. The minimal BGC, consisting solely of *darA* and *darE* was determined and it was shown that to release DAR from the leader peptide, no designated DAR peptidase has to be co-expressed. A high *darA:darE* transcript ratio is necessary to obtain a high DAR yield. Currently, a heterologous production system using *E. coli* Rosetta™(DE3) is available that enables production of >30 mg/L DAR within 2 days.

V. Conclusion

The genome-mining approach is an advancing approach for novel antibiotic discovery. Using this approach, the buried potential of bioactive natural products as the base for drug development can be unveiled. This study exemplified the use of genome-mining to detect cryptic BGCs from understudied bacteria that showed antibacterial activity.

The genetic tools for cloning putative BGCs into heterologous expression systems are widely available and are proven to be effective in this study. This genetic platform allows a plug-and-play strategy to refactor the biosynthetic pathway of interest. However, the application of molecular techniques should be coupled with a justified experimental design to reach the goal of compound production. As experienced in this study, compound production by heterologous expression of BGCs from understudied bacteria was especially more challenging due to the lack of information about the bacteria and the gene regulations might be more complex. Moreover, a suitable heterologous host for understudied bacteria is still missing. Although the heterologous expression of large multimodular BGCs, like many NRPS and PKS systems, had been previously exemplified, this is a growing field in which advancements are always available and more insights are provided to improve the heterologous expression system as discussed in section III.4.3 (Prospective work to improve heterologous expression of NP BGC).

On the other hand, the *E. coli* heterologous expression system, which had been comprehensively studied, was proven to be effective in confirmation of the DAR BGC functionality. Moreover, a functional heterologous expression platform for DAR was generated in this study. This platform can be utilized in the future to (i) produce DAR for further studies, (ii) produce DAR derivatives with improved activity and/or pharmacokinetics, and (iii) further investigate the biosynthesis of DAR and related molecules.

VI. References

Abraham, E.P., Chain, E., 1940. An Enzyme from Bacteria able to Destroy Penicillin. *Nature* 146, 837–837. <https://doi.org/10.1038/146837a0>

Ahmed, Y., Rebets, Y., Estévez, M.R., Zapp, J., Myronovskyi, M., Luzhetskyy, A., 2020. Engineering of *Streptomyces lividans* for heterologous expression of secondary metabolite gene clusters. *Microbial Cell Factories* 19, 5. <https://doi.org/10.1186/s12934-020-1277-8>

Alberti, F., Leng, D.J., Wilkening, I., Song, L., Tosin, M., Corre, C., 2018. Triggering the expression of a silent gene cluster from genetically intractable bacteria results in scleric acid discovery †Electronic supplementary information (ESI) available: Supplementary methods and results; Tables S1–S6; Fig. S1–S22. See DOI: 10.1039/c8sc03814g. *Chem Sci* 10, 453–463. <https://doi.org/10.1039/c8sc03814g>

Arnison, P.G., Bibb, M.J., Bierbaum, G., Bowers, A.A., Bugni, T.S., Bulaj, G., Camarero, J.A., Campopiano, D.J., Challis, G.L., Clardy, J., Cotter, P.D., Craik, D.J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P.C., Entian, K.-D., Fischbach, M.A., Garavelli, J.S., Göransson, U., Gruber, C.W., Haft, D.H., Hemscheidt, T.K., Hertweck, C., Hill, C., Horswill, A.R., Jaspars, M., Kelly, W.L., Klinman, J.P., Kuipers, O.P., Link, A.J., Liu, W., Marahiel, M.A., Mitchell, D.A., Moll, G.N., Moore, B.S., Müller, R., Nair, S.K., Nes, I.F., Norris, G.E., Olivera, B.M., Onaka, H., Patchett, M.L., Piel, J., Reaney, M.J.T., Rebuffat, S., Ross, R.P., Sahl, H.-G., Schmidt, E.W., Selsted, M.E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Süßmuth, R.D., Tagg, J.R., Tang, G.-L., Truman, A.W., Vederas, J.C., Walsh, C.T., Walton, J.D., Wenzel, S.C., Willey, J.M., van der Donk, W.A., 2013. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* 30, 108–160. <https://doi.org/10.1039/c2np20085f>

Avila, C., Núñez-Pons, L., Moles, J., 2018. From the Tropics to the Poles: Chemical Defense Strategies in Sea Slugs (Mollusca: Heterobranchia). pp. 71–163. <https://doi.org/10.1201/9780429453465-3>

Barlow, M., Hall, B.G., 2002. Phylogenetic analysis shows that the OXA beta-lactamase genes have been on plasmids for millions of years. *J Mol Evol* 55, 314–321. <https://doi.org/10.1007/s00239-002-2328-y>

Bauman, K.D., Li, J., Murata, K., Mantovani, S.M., Dahesh, S., Nizet, V., Luhavaya, H., Moore, B.S., 2019. Refactoring the Cryptic Streptophenazine Biosynthetic Gene Cluster Unites Phenazine, Polyketide, and Nonribosomal Peptide Biochemistry. *Cell Chem Biol* 26, 724-736.e7. <https://doi.org/10.1016/j.chembiol.2019.02.004>

Beck, C., Garzón, J.F.G., Weber, T., 2020. Recent Advances in Re-engineering Modular PKS and NRPS Assembly Lines. *Biotechnol Bioproc E* 25, 886–894. <https://doi.org/10.1007/s12257-020-0265-5>

Bentley, S.D., Chater, K.F., Cerdeño-Tárraga, A.-M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown,

References

- 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., Rabbinowitsch, E., Rajandream, M.-A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J., Hopwood, D.A., 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417, 141. <https://doi.org/10.1038/417141a>
- Bérdy, J., 2005. Bioactive Microbial Metabolites. *J Antibiot* 58, 1–26. <https://doi.org/10.1038/ja.2005.1>
- Bian, X., Plaza, A., Zhang, Y., Müller, R., 2012. Luminmycins A–C, Cryptic Natural Products from *Phototrhabdus luminescens* Identified by Heterologous Expression in *Escherichia coli*. *J. Nat. Prod.* 75, 1652–1655. <https://doi.org/10.1021/np300444e>
- Bibb, M.J., Janssen, G.R., Ward, J.M., 1985. Cloning and analysis of the promoter region of the erythromycin resistance gene (*ermE*) of *Streptomyces erythraeus*. *Gene* 38, 215–226. [https://doi.org/10.1016/0378-1119\(85\)90220-3](https://doi.org/10.1016/0378-1119(85)90220-3)
- Bisang, C., Long, P.F., Cortés, J., Westcott, J., Crosby, J., Matharu, A.L., Cox, R.J., Simpson, T.J., Staunton, J., Leadlay, P.F., 1999. A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature* 401, 502–505. <https://doi.org/10.1038/46829>
- Blin, K., Shaw, S., Kloosterman, A.M., Charlop-Powers, Z., van Wezel, G.P., Medema, M.H., Weber, T., 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkab335>
- Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H., Weber, T., 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* 47, W81–W87. <https://doi.org/10.1093/nar/gkz310>
- Blodgett, J.A.V., Zhang, J.K., Metcalf, W.W., 2005. Molecular Cloning, Sequence Analysis, and Heterologous Expression of the Phosphinothricin Tripeptide Biosynthetic Gene Cluster from *Streptomyces viridochromogenes* DSM 40736. *Antimicrob Agents Chemother* 49, 230–240. <https://doi.org/10.1128/AAC.49.1.230-240.2005>
- Böhringer, N., Fisch, K.M., Schillo, D., Bara, R., Hertzler, C., Grein, F., Eisenbarth, J.-H., Kaligis, F., Schneider, T., Wägele, H., König, G.M., Schäberle, T.F., 2017. Antimicrobial Potential of Bacteria Associated with Marine Sea Slugs from North Sulawesi, Indonesia. *Front Microbiol* 8. <https://doi.org/10.3389/fmicb.2017.01092>
- Bouhired, S.M., Crüsemann, M., Almeida, C., Weber, T., Piel, J., Schäberle, T.F., König, G.M., 2014. Biosynthesis of Phenylannolone A, a Multidrug Resistance Reversal Agent from the Halotolerant Myxobacterium *Nannocystis pusilla* B150. *ChemBioChem* 15, 757–765. <https://doi.org/10.1002/cbic.201300676>

- Busse, H.-J., 2016. Review of the taxonomy of the genus *Arthrobacter*, emendation of the genus *Arthrobacter sensu lato*, proposal to reclassify selected species of the genus *Arthrobacter* in the novel genera *Glutamicibacter* gen. nov., *Paeniglutamicibacter* gen. nov., *Pseudoglutamicibacter* gen. nov., *Paenarthrobacter* gen. nov. and *Pseudarthrobacter* gen. nov., and emended description of *Arthrobacter roseus*. *Int J Syst Evol Microbiol* 66, 9–37. <https://doi.org/10.1099/ijsem.0.000702>
- Busse, H.-J., Schumann, P., 2019. Reclassification of *Arthrobacter endophyticus* (Wang *et al.* 2015) as *Glutamicibacter endophyticus* comb. nov. *Int J Syst Evol Microbiol* 69, 1057–1059. <https://doi.org/10.1099/ijsem.0.003268>
- Butala, M., Zgur-Bertok, D., Busby, S.J.W., 2009. The bacterial LexA transcriptional repressor. *Cell Mol Life Sci* 66, 82–93. <https://doi.org/10.1007/s00018-008-8378-6>
- Calcott, M.J., Owen, J.G., Lamont, I.L., Ackerley, D.F., 2014. Biosynthesis of Novel Pyoverdines by Domain Substitution in a Nonribosomal Peptide Synthetase of *Pseudomonas aeruginosa*. *Applied and Environmental Microbiology* 80, 5723–5731. <https://doi.org/10.1128/AEM.01453-14>
- Cao, L., Do, T., Link, A.J., 2021. Mechanisms of action of ribosomally synthesized and posttranslationally modified peptides (RiPPs). *Journal of Industrial Microbiology and Biotechnology* 48. <https://doi.org/10.1093/jimb/kuab005>
- Challis, G.L., Ravel, J., Townsend, C.A., 2000. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chemistry & Biology* 7, 211–224. [https://doi.org/10.1016/S1074-5521\(00\)00091-0](https://doi.org/10.1016/S1074-5521(00)00091-0)
- Cheng, Y., Yang, R., Lyu, M., Wang, S., Liu, X., Wen, Y., Song, Y., Li, J., Chen, Z., 2018. IdeR, a DtxR Family Iron Response Regulator, Controls Iron Homeostasis, Morphological Differentiation, Secondary Metabolism, and the Oxidative Stress Response in *Streptomyces avermitilis*. *Appl Environ Microbiol* 84. <https://doi.org/10.1128/AEM.01503-18>
- Crooks, G.E., Hon, G., Chandonia, J.-M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. *Genome Res* 14, 1188–1190. <https://doi.org/10.1101/gr.849004>
- Crüsemann, M., Reher, R., Schamari, I., Brachmann, A.O., Ohbayashi, T., Kuschak, M., Malfacini, D., Seidinger, A., Pinto-Carbó, M., Richarz, R., Reuter, T., Kehraus, S., Hallab, A., Attwood, M., Schiöth, H.B., Mergaert, P., Kikuchi, Y., Schäberle, T.F., Kostenis, E., Wenzel, D., Müller, C.E., Piel, J., Carlier, A., Eberl, L., König, G.M., 2018. Heterologous Expression, Biosynthetic Studies, and Ecological Function of the Selective Gq-Signaling Inhibitor FR900359. *Angewandte Chemie International Edition* 57, 836–840. <https://doi.org/10.1002/anie.201707996>
- Cushnie, T.P.T., Cushnie, B., Lamb, A.J., 2014. Alkaloids: An overview of their antibacterial, antibiotic-enhancing and antivirulence activities. *International Journal of Antimicrobial Agents* 44, 377–386. <https://doi.org/10.1016/j.ijantimicag.2014.06.001>

References

- Dangel, V., Westrich, L., Smith, M.C.M., Heide, L., Gust, B., 2010. Use of an inducible promoter for antibiotic production in a heterologous host. *Appl Microbiol Biotechnol* 87, 261–269. <https://doi.org/10.1007/s00253-009-2435-4>
- Das, L., Deb, S., Das, S.K., 2020. *Glutamicibacter mishrai* sp. nov., isolated from the coral *Favia veroni* from Andaman Sea. *Arch Microbiol* 202, 733–745. <https://doi.org/10.1007/s00203-019-01783-0>
- Datsenko, K.A., Wanner, B.L., 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *PNAS* 97, 6640–6645.
- Davis, J.H., Rubin, A.J., Sauer, R.T., 2011. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res* 39, 1131–1141. <https://doi.org/10.1093/nar/gkq810>
- de Crécy-Lagard, V., Saurin, W., Thibaut, D., Gil, P., Naudin, L., Crouzet, J., Blanc, V., 1997. Streptogramin B biosynthesis in *Streptomyces pristinaespiralis* and *Streptomyces virginiae*: molecular characterization of the last structural peptide synthetase gene. *Antimicrob Agents Chemother* 41, 1904–1909.
- Degenkolb, T., Düring, R.-A., Vilcinskas, A., 2011. Secondary Metabolites Released by The Burying Beetle *Nicrophorus vespilloides*: Chemical Analyses and Possible Ecological Functions. *J Chem Ecol* 37, 724–735. <https://doi.org/10.1007/s10886-011-9978-4>
- Demain, A.L., Fang, A., 2000. The natural functions of secondary metabolites. *Adv Biochem Eng Biotechnol* 69, 1–39. https://doi.org/10.1007/3-540-44964-7_1
- Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J., Katz, L., 1991. Modular organization of genes required for complex polyketide biosynthesis. *Science* 252, 675–679. <https://doi.org/10.1126/science.2024119>
- Donovick, R., Pagano, J.F., Stout, H.A., Weinstein, M.J., 1955. Thiostrepton, a new antibiotic. I. In vitro studies. *Antibiot Annu* 3, 554–559.
- Dorman, C.J., 2004. H-NS: a universal regulator for a dynamic genome. *Nat Rev Microbiol* 2, 391–400. <https://doi.org/10.1038/nrmicro883>
- Duquesne, S., Destoumieux-Garzón, D., Peduzzi, J., Rebuffat, S., 2007. Microcins, gene-encoded antibacterial peptides from enterobacteria. *Nat. Prod. Rep.* 24, 708–734. <https://doi.org/10.1039/B516237H>
- Eggert, A.-K., Müller, J.K., 1997. Biparental care and social evolution in burying beetles: lessons from the larder, in: Crespi, B.J., Choe, J.C. (Eds.), *The Evolution of Social Behaviour in Insects and Arachnids*. Cambridge University Press, Cambridge, pp. 216–236. <https://doi.org/10.1017/CBO9780511721953.011>
- Ehrlich, J., Bartz, Q.R., Smith, R.M., Joslyn, D.A., Burkholder, P.R., 1947. Chloromycetin, a New Antibiotic From a Soil Actinomycete. *Science* 106, 417. <https://doi.org/10.1126/science.106.2757.417>

- Eisenbarth, J.-H., Undap, N., Papu, A., Schillo, D., Dialao, J., Reumschüssel, S., Kaligis, F., Bara, R., Schäberle, T.F., König, G.M., Yonow, N., Wägele, H., 2018. Marine Heterobranchia (Gastropoda, Mollusca) in Bunaken National Park, North Sulawesi, Indonesia—A Follow-Up Diversity Study. *Diversity* 10, 127. <https://doi.org/10.3390/d10040127>
- Elfalah, H., Usup, G., Ahmad, A., 2013. Anti-Microbial Properties of Secondary Metabolites of Marine *Gordonia tearrae* Extract. *Journal of Agricultural Science* 5, p94. <https://doi.org/10.5539/jas.v5n6p94>
- Felnagle, E.A., Jackson, E.E., Chan, Y.A., Podevels, A.M., Berti, A.D., McMahon, M.D., Thomas, M.G., 2008. Nonribosomal Peptide Synthetases Involved in the Production of Medically Relevant Natural Products. *Mol Pharm* 5, 191–211. <https://doi.org/10.1021/mp700137g>
- Fisch, K.M., Hertzner, C., Böhringer, N., Wuisan, Z.G., Schillo, D., Bara, R., Kaligis, F., Wägele, H., König, G.M., Schäberle, T.F., 2017. The Potential of Indonesian Heterobranchs Found around Bunaken Island for the Production of Bioactive Compounds. *Mar Drugs* 15. <https://doi.org/10.3390/md15120384>
- 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.
- Flinspach, K., Westrich, L., Kaysser, L., Siebenberg, S., Gomez-Escribano, J.P., Bibb, M., Gust, B., Heide, L., 2010. Heterologous expression of the biosynthetic gene clusters of coumermycin A(1), clorobiocin and caprazamycins in genetically modified *Streptomyces coelicolor* strains. *Biopolymers* 93, 823–832. <https://doi.org/10.1002/bip.21493>
- Funa, N., Ohnishi, Y., Fujii, I., Shibuya, M., Ebizuka, Y., Horinouchi, S., 1999. A new pathway for polyketide synthesis in microorganisms. *Nature* 400, 897–899. <https://doi.org/10.1038/23748>
- Garcia-Gonzalez, E., Müller, S., Ensle, P., Süssmuth, R.D., Genersch, E., 2014. Elucidation of sevadicin, a novel non-ribosomal peptide secondary metabolite produced by the honey bee pathogenic bacterium *Paenibacillus larvae*. *Environmental Microbiology* 16, 1297–1309. <https://doi.org/10.1111/1462-2920.12417>
- Ghosh, S., Chandra, N., Vishveshwara, S., 2015. Mechanism of Iron-Dependent Repressor (IdeR) Activation and DNA Binding: A Molecular Dynamics and Protein Structure Network Study. *PLOS Computational Biology* 11, e1004500. <https://doi.org/10.1371/journal.pcbi.1004500>
- Gibson, D.G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E.A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M.A., Merryman, C., Young, L., Noskov, V.N., Glass, J.I., Venter, J.C., Hutchison, C.A., Smith, H.O., 2008. Complete Chemical Synthesis, Assembly, and Cloning of a *Mycoplasma genitalium* Genome. *Science* 319, 1215–1220. <https://doi.org/10.1126/science.1151721>

- Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6, 343–345. <https://doi.org/10.1038/nmeth.1318>
- Giessen, T.W., Franke, K.B., Knappe, T.A., Kraas, F.I., Bosello, M., Xie, X., Linne, U., Marahiel, M.A., 2012. Isolation, Structure Elucidation, and Biosynthesis of an Unusual Hydroxamic Acid Ester-Containing Siderophore from *Actinosynnema mirum*. *J. Nat. Prod.* 75, 905–914. <https://doi.org/10.1021/np300046k>
- Gomez-Escribano, J.P., Bibb, M.J., 2014. Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: from genome mining to manipulation of biosynthetic pathways. *Journal of Industrial Microbiology and Biotechnology* 41, 425–431. <https://doi.org/10.1007/s10295-013-1348-5>
- Gomez-Escribano, J.P., Bibb, M.J., 2011. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microbial Biotechnology* 4, 207–215. <https://doi.org/10.1111/j.1751-7915.2010.00219.x>
- Gomez-Escribano, J.P., Castro, J.F., Razmilic, V., Jarmusch, S.A., Saalbach, G., Ebel, R., Jaspars, M., Andrews, B., Asenjo, J.A., Bibb, M.J., 2019. Heterologous Expression of a Cryptic Gene Cluster from *Streptomyces leeuwenhoekii* C34T Yields a Novel Lasso Peptide, Leepeptin. *Applied and Environmental Microbiology* 85, e01752-19. <https://doi.org/10.1128/AEM.01752-19>
- Gontang, E.A., Gaudêncio, S.P., Fenical, W., Jensen, P.R., 2010. Sequence-Based Analysis of Secondary-Metabolite Biosynthesis in Marine Actinobacteria. *Appl. Environ. Microbiol.* 76, 2487–2499. <https://doi.org/10.1128/AEM.02852-09>
- Gosliner, T., Valdés, Á., Behrens, D.W., 2015. *Nudibranch & Sea Slug Identification: Indo-Pacific*. New World Publications, Inc., Jacksonville, Florida USA.
- Graça, A.P., Bondoso, J., Gaspar, H., Xavier, J.R., Monteiro, M.C., Cruz, M. de la, Oves-Costales, D., Vicente, F., Lage, O.M., 2013. Antimicrobial Activity of Heterotrophic Bacterial Communities from the Marine Sponge *Erylus discophorus* (Astrophorida, Geodiidae). *PLOS ONE* 8, e78992. <https://doi.org/10.1371/journal.pone.0078992>
- Greene, N.P., Kaplan, E., Crow, A., Koronakis, V., 2018. Antibiotic Resistance Mediated by the MacB ABC Transporter Family: A Structural and Functional Perspective. *Front. Microbiol.* 0. <https://doi.org/10.3389/fmicb.2018.00950>
- Griffith, R.S., 1981. Introduction to vancomycin. *Rev Infect Dis* 3 suppl, S200-204.
- Gruber, T.M., Gross, C.A., 2003. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. *Annu. Rev. Microbiol.* 57, 441–466. <https://doi.org/10.1146/annurev.micro.57.030502.090913>
- Gurgui, C., Piel, J., 2010. Metagenomic Approaches to Identify and Isolate Bioactive Natural Products from Microbiota of Marine Sponges, in: Streit, W.R., Daniel, R. (Eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 247–264. https://doi.org/10.1007/978-1-60761-823-2_17

- Gust, B., Chandra, G., Jakimowicz, D., Yuqing, T., Bruton, C.J., Chater, K.F., 2004. Lambda red-mediated genetic manipulation of antibiotic-producing *Streptomyces*. *Adv. Appl. Microbiol.* 54, 107–128. [https://doi.org/10.1016/S0065-2164\(04\)54004-2](https://doi.org/10.1016/S0065-2164(04)54004-2)
- Gust, B., O'Rourke, S., Bird, N., Kieser, T., Chater, K., 2006. Recombineering in *Streptomyces coelicolor*. *FEMS Online Protocols* 22. <http://streptomyces.org.uk/redirect/RecombineeringFEMSMP-2006-5.pdf>
- He, Y., Zhou, C., Huang, M., Tang, C., Liu, X., Yue, Y., Diao, Q., Zheng, Z., Liu, D., 2020. Glyoxalase system: A systematic review of its biological activity, related-diseases, screening methods and small molecule regulators. *Biomedicine & Pharmacotherapy* 131, 110663. <https://doi.org/10.1016/j.biopha.2020.110663>
- Hegemann, J.D., Zimmermann, M., Zhu, S., Klug, D., Marahiel, M.A., 2013. Lasso peptides from proteobacteria: Genome mining employing heterologous expression and mass spectrometry. *Biopolymers* 100, 527–542. <https://doi.org/10.1002/bip.22326>
- Heise, P., 2019. The microbiome of the burying beetle *Nicrophorus vespilloides* as an untapped source for the screening of bioactive small molecules (Dissertation). Justus-Liebig-Universität Gießen, Gießen.
- Heise, P., Liu, Y., Degenkolb, T., Vogel, H., Schäberle, T.F., Vilcinskas, A., 2019. Antibiotic-Producing Beneficial Bacteria in the Gut of the Burying Beetle *Nicrophorus vespilloides*. *Front Microbiol* 10. <https://doi.org/10.3389/fmicb.2019.01178>
- Hertweck, C., Luzhetskyy, A., Rebets, Y., Bechthold, A., 2007. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat. Prod. Rep.* 24, 162–190. <https://doi.org/10.1039/B507395M>
- Hillson, N.J., 2011. DNA Assembly Method Standardization for Synthetic Biomolecular Circuits and Systems, in: Koepl, H., Setti, G., di Bernardo, M., Densmore, D. (Eds.), *Design and Analysis of Biomolecular Circuits: Engineering Approaches to Systems and Synthetic Biology*. Springer, New York, NY, pp. 295–314. https://doi.org/10.1007/978-1-4419-6766-4_14
- Hochlowski, J.E., Swanson, S.J., Ranfranz, L.M., Whittern, D.N., Buko, A.M., McAlpine, J.B., 1987. Tiacumicins, a novel complex of 18-membered macrolides. II. Isolation and structure determination. *J Antibiot (Tokyo)* 40, 575–588. <https://doi.org/10.7164/antibiotics.40.575>
- Hoshino, Y., Chiba, K., Ishino, K., Fukai, T., Igarashi, Y., Yazawa, K., Mikami, Y., Ishikawa, J., 2011. Identification of Nocobactin NA Biosynthetic Gene Clusters in *Nocardia farcinica*. *J Bacteriol* 193, 441–448. <https://doi.org/10.1128/JB.00897-10>
- Hudson, G.A., Mitchell, D.A., 2018. RiPP antibiotics: biosynthesis and engineering potential. *Curr Opin Microbiol* 45, 61–69. <https://doi.org/10.1016/j.mib.2018.02.010>
- Huo, L., Hug, J.J., Fu, C., Bian, X., Zhang, Y., Müller, R., 2019. Heterologous expression of bacterial natural product biosynthetic pathways. *Nat. Prod. Rep.* 36, 1412–1436. <https://doi.org/10.1039/C8NP00091C>

- Imai, Y., Meyer, K.J., Iinishi, A., Favre-Godal, Q., Green, R., Manuse, S., Caboni, M., Mori, M., Niles, S., Ghiglieri, M., Honrao, C., Ma, X., Guo, J.J., Makriyannis, A., Linares-Otoya, L., Böhringer, N., Wuisan, Z.G., Kaur, H., Wu, R., Mateus, A., Typas, A., Savitski, M.M., Espinoza, J.L., O'Rourke, A., Nelson, K.E., Hiller, S., Noinaj, N., Schäberle, T.F., D'Onofrio, A., Lewis, K., 2019. A new antibiotic selectively kills Gram-negative pathogens. *Nature* 576, 459–464. <https://doi.org/10.1038/s41586-019-1791-1>
- Ivanova, N., Sikorski, J., Jando, M., Lapidus, A., Nolan, M., Lucas, S., Del Rio, T.G., Tice, H., Copeland, A., Cheng, J.-F., Chen, F., Bruce, D., Goodwin, L., Pitluck, S., Mavromatis, K., Ovchinnikova, G., Pati, A., Chen, A., Palaniappan, K., Land, M., Hauser, L., Chang, Y.-J., Jeffries, C.D., Chain, P., Saunders, E., Han, C., Detter, J.C., Brettin, T., Rohde, M., Göker, M., Bristow, J., Eisen, J.A., Markowitz, V., Hugenholtz, P., Klenk, H.-P., Kyrpides, N.C., 2010. Complete genome sequence of *Gordonia bronchialis* type strain (3410T). *Stand Genomic Sci* 2, 19–28. <https://doi.org/10.4056/sigs.611106>
- Jacobs, C.G.C., Steiger, S., Heckel, D.G., Wielsch, N., Vilcinskis, A., Vogel, H., 2016. Sex, offspring and carcass determine antimicrobial peptide expression in the burying beetle. *Sci Rep* 6. <https://doi.org/10.1038/srep25409>
- Jenner, M., Jian, X., Dashti, Y., Masschelein, J., Hobson, C., Roberts, D.M., Jones, C., Harris, S., Parkhill, J., Raja, H.A., Oberlies, N.H., Pearce, C.J., Mahenthalingam, E., Challis, G.L., 2019. An unusual *Burkholderia gladioli* double chain-initiating nonribosomal peptide synthetase assembles 'fungal' icosalide antibiotics. *Chem. Sci.* 10, 5489–5494. <https://doi.org/10.1039/C8SC04897E>
- Jones, A.C., Gust, B., Kulik, A., Heide, L., Buttner, M.J., Bibb, M.J., 2013. Phage P1-Derived Artificial Chromosomes Facilitate Heterologous Expression of the FK506 Gene Cluster. *PLoS One* 8. <https://doi.org/10.1371/journal.pone.0069319>
- Jukes, T.H., 1985. Some Historical Notes on Chlorotetracycline. *Reviews of Infectious Diseases* 7, 702–707. <https://doi.org/10.1093/clinids/7.5.702>
- Kaligis, F., Eisenbarth, J.-H., Schillo, D., Dialao, J., Schäberle, T.F., Böhringer, N., Bara, R., Reumschüssel, S., König, G.M., Wägele, H., 2018. Second survey of heterobranch sea slugs (Mollusca, Gastropoda, Heterobranchia) from Bunaken National Park, North Sulawesi, Indonesia - how much do we know after 12 years? *Marine Biodiversity Records* 11, 2. <https://doi.org/10.1186/s41200-018-0136-3>
- Kalinová, B., Podskalská, H., Růžička, J., Hoskovec, M., 2009. Irresistible bouquet of death—how are burying beetles (Coleoptera: Silphidae: *Nicrophorus*) attracted by carcasses. *Naturwissenschaften* 96, 889–899. <https://doi.org/10.1007/s00114-009-0545-6>
- Katsuyama, Y., Ohnishi, Y., 2012. Type III polyketide synthases in microorganisms. *Methods Enzymol* 515, 359–377. <https://doi.org/10.1016/B978-0-12-394290-6.00017-3>
- Kaur, H., Jakob, R.P., Marzinek, J.K., Green, R., Imai, Y., Bolla, J.R., Agustoni, E., Robinson, C.V., Bond, P.J., Lewis, K., Maier, T., Hiller, S., 2021. The antibiotic

darobactin mimics a β -strand to inhibit outer membrane insertase. *Nature* 593, 125–129. <https://doi.org/10.1038/s41586-021-03455-w>

Kautsar, S.A., Blin, K., Shaw, S., Navarro-Muñoz, J.C., Terlouw, B.R., van der Hooft, J.J.J., van Santen, J.A., Tracanna, V., Suarez Duran, H.G., Pascal Andreu, V., Selem-Mojica, N., Alanjary, M., Robinson, S.L., Lund, G., Epstein, S.C., Sisto, A.C., Charkoudian, L.K., Collemare, J., Linington, R.G., Weber, T., Medema, M.H., 2020. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Research* 48, D454–D458. <https://doi.org/10.1093/nar/gkz882>

Keatinge-Clay, A.T., 2017. Polyketide Synthase Modules Redefined. *Angew Chem Int Ed Engl* 56, 4658–4660. <https://doi.org/10.1002/anie.201701281>

Keatinge-Clay, A.T., 2012. The structures of type I polyketide synthases. *Nat Prod Rep* 29, 1050–1073. <https://doi.org/10.1039/c2np20019h>

Kernan, M.R., Molinski, T.F., Faulkner, D.J., 1988. Macrocyclic antifungal metabolites from the Spanish dancer nudibranch *Hexabranhus sanguineus* and sponges of the genus *Halichondria*. *J. Org. Chem.* 53, 5014–5020. <https://doi.org/10.1021/jo00256a021>

Konovalova, A., Kahne, D.E., Silhavy, T.J., 2017. Outer Membrane Biogenesis. *Annu Rev Microbiol* 71, 539–556. <https://doi.org/10.1146/annurev-micro-090816-093754>

Kouprina, N., Larionov, V., 2008. Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*. *Nat. Protocols* 3, 371–377. <https://doi.org/10.1038/nprot.2008.5>

Kouprina, N., Larionov, V., 2006. TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. *Nat Rev Genet* 7, 805–812. <https://doi.org/10.1038/nrg1943>

Lee, N., Hwang, S., Lee, Y., Cho, S., Palsson, B., Cho, B.-K., 2019. Synthetic Biology Tools for Novel Secondary Metabolite Discovery in *Streptomyces*. *J Microbiol Biotechnol* 29, 667–686. <https://doi.org/10.4014/jmb.1904.04015>

Lee, N.C.O., Larionov, V., Kouprina, N., 2015. Highly efficient CRISPR/Cas9-mediated TAR cloning of genes and chromosomal loci from complex genomes in yeast. *Nucleic Acids Research* 43, e55–e55. <https://doi.org/10.1093/nar/gkv112>

Lewis, K., 2013. Platforms for antibiotic discovery. *Nat Rev Drug Discov* 12, 371–387. <https://doi.org/10.1038/nrd3975>

Liao, R., Duan, L., Lei, C., Pan, H., Ding, Y., Zhang, Q., Chen, D., Shen, B., Yu, Y., Liu, W., 2009. Thiopeptide Biosynthesis Featuring Ribosomally Synthesized Precursor Peptides and Conserved Posttranslational Modifications. *Chemistry & Biology* 16, 141–147. <https://doi.org/10.1016/j.chembiol.2009.01.007>

- Lim, Y.P., Go, M.K., Yew, W.S., 2016. Exploiting the Biosynthetic Potential of Type III Polyketide Synthases. *Molecules* 21. <https://doi.org/10.3390/molecules21060806>
- Linares-Otoya, L., Linares-Otoya, V., Armas-Mantilla, L., Blanco-Olano, C., Crüsemann, M., Ganoza-Yupanqui, M.L., Campos-Florian, J., König, G.M., Schäberle, T.F.Y. 2017, 2017. Identification and heterologous expression of the kocurin biosynthetic gene cluster. *Microbiology* 163, 1409–1414. <https://doi.org/10.1099/mic.0.000538>
- Ling, L.L., Schneider, T., Peoples, A.J., Spoering, A.L., Engels, I., Conlon, B.P., Mueller, A., Schäberle, T.F., Hughes, D.E., Epstein, S., Jones, M., Lazarides, L., Steadman, V.A., Cohen, D.R., Felix, C.R., Fetterman, K.A., Millett, W.P., Nitti, A.G., Zullo, A.M., Chen, C., Lewis, K., 2015. A new antibiotic kills pathogens without detectable resistance. *Nature* 517, 455–459. <https://doi.org/10.1038/nature14098>
- Liras, P., Demain, A.L., 2009. Chapter 16. Enzymology of beta-lactam compounds with cephem structure produced by actinomycete. *Methods Enzymol* 458, 401–429. [https://doi.org/10.1016/S0076-6879\(09\)04816-2](https://doi.org/10.1016/S0076-6879(09)04816-2)
- Luo, Y., Huang, H., Liang, J., Wang, M., Lu, L., Shao, Z., Cobb, R.E., Zhao, H., 2013. Activation and characterization of a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster. *Nature Communications* 4, 2894. <https://doi.org/10.1038/ncomms3894>
- Magarvey, N.A., Haltli, B., He, M., Greenstein, M., Hucul, J.A., 2006. Biosynthetic pathway for mannopeptimycins, lipoglycopeptide antibiotics active against drug-resistant gram-positive pathogens. *Antimicrob Agents Chemother* 50, 2167–2177. <https://doi.org/10.1128/AAC.01545-05>
- Mahizan, N.A., Yang, S.-K., Moo, C.-L., Song, A.A.-L., Chong, C.-M., Chong, C.-W., Abushelaibi, A., Lim, S.-H.E., Lai, K.-S., 2019. Terpene Derivatives as a Potential Agent against Antimicrobial Resistance (AMR) Pathogens. *Molecules* 24, 2631. <https://doi.org/10.3390/molecules24142631>
- Malpartida, F., Hopwood, D.A., 1984. Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotic and its expression in a heterologous host. *Nature* 309, 462–464. <https://doi.org/10.1038/309462a0>
- Margalith, P., Beretta, G., 1960. Rifomycin. XI. taxonomic study on *Streptomyces mediterranei* nov. sp. *Mycopathologia et Mycologia Applicata* 13, 321–330. <https://doi.org/10.1007/BF02089930>
- Matsunaga, S., Fusetani, N., Hashimoto, K., Koseki, K., Noma, M., Noguchi, H., Sankawa, U., 1989. Bioactive marine metabolites. 25. Further kabiramides and halichondramides, cytotoxic macrolides embracing trisoxazole, from the *Hexabranhus* egg masses. *J. Org. Chem.* 54, 1360–1363. <https://doi.org/10.1021/jo00267a024>
- McGUIRE, J.M., Bunch, R.L., Anderson, R.C., Boaz, H.E., Flynn, E.H., Powell, H.M., Smith, J.W., 1952. Ilotycin, a new antibiotic. *Antibiot Chemother (Northfield)* 2, 281–283.

Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M.A., Weber, T., Takano, E., Breitling, R., 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research* 39, W339–W346. <https://doi.org/10.1093/nar/gkr466>

Medema, M.H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J.B., Blin, K., de Bruijn, I., Chooi, Y.H., Claesen, J., Coates, R.C., Cruz-Morales, P., Duddela, S., Düsterhus, S., Edwards, D.J., Fewer, D.P., Garg, N., Geiger, C., Gomez-Escribano, J.P., Greule, A., Hadjithomas, M., 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., Kötter, P., Krug, D., Masschelein, J., Melnik, A.V., Mantovani, S.M., Monroe, E.A., Moore, M., Moss, N., Nützmann, 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-Gómez, 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., Höfte, 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., Luzhetskyy, A., Mahmud, T., Mast, Y., Méndez, C., Metsä-Ketelä, M., Micklefield, J., Mitchell, D.A., Moore, B.S., Moreira, L.M., Müller, R., Neilan, B.A., Nett, M., Nielsen, J., O’Gara, F., Oikawa, H., Osbourn, A., Osburne, M.S., Ostash, B., Payne, S.M., Pernodet, J.-L., Petricek, M., Piel, J., Ploux, O., Raaijmakers, J.M., Salas, J.A., Schmitt, E.K., Scott, B., Seipke, R.F., Shen, B., Sherman, D.H., Sivonen, K., Smanski, M.J., Sosio, M., Stegmann, E., Süssmuth, R.D., Tahlan, K., Thomas, C.M., Tang, Y., Truman, A.W., Viaud, M., Walton, J.D., Walsh, C.T., Weber, T., van Wezel, G.P., Wilkinson, B., Willey, J.M., Wohlleben, W., Wright, G.D., Ziemert, N., Zhang, C., Zotchev, S.B., Breitling, R., Takano, E., Glöckner, F.O., 2015. Minimum Information about a Biosynthetic Gene cluster. *Nat Chem Biol* 11, 625–631. <https://doi.org/10.1038/nchembio.1890>

Miller, B.R., Gulick, A.M., 2016. Structural Biology of Non-Ribosomal Peptide Synthetases. *Methods Mol Biol* 1401, 3–29. https://doi.org/10.1007/978-1-4939-3375-4_1

Montalbán-López, M., Scott, T.A., Ramesh, S., Rahman, I.R., van Heel, A.J., Viel, J.H., Bandarian, V., Dittmann, E., Genilloud, O., Goto, Y., Grande Burgos, M.J., Hill, C., Kim, S., Koehnke, J., Latham, J.A., Link, A.J., Martínez, B., Nair, S.K., Nicolet, Y., Rebuffat, S., Sahl, H.-G., Sareen, D., Schmidt, E.W., Schmitt, L., Severinov, K., Süssmuth, R.D., Truman, A.W., Wang, H., Weng, J.-K., van Wezel, G.P., Zhang, Q., Zhong, J., Piel, J., Mitchell, D.A., Kuipers, O.P., van der Donk, W.A., 2021. New developments in RiPP discovery, enzymology and engineering. *Nat Prod Rep* 38, 130–239. <https://doi.org/10.1039/d0np00027b>

- Mortality and Cause of Death, 1900 v. 2010 [WWW Document], 2014. . Carolina Demography. URL <https://www.ncdemography.org/2014/06/16/mortality-and-cause-of-death-1900-v-2010/> (accessed 6.11.21).
- Mosberg, J.A., Lajoie, M.J., Church, G.M., 2010. Lambda Red Recombineering in *Escherichia coli* Occurs Through a Fully Single-Stranded Intermediate. *Genetics* 186, 791–799. <https://doi.org/10.1534/genetics.110.120782>
- Motamedi, H., Hutchinson, C.R., 1987. Cloning and heterologous expression of a gene cluster for the biosynthesis of tetracenomycin C, the anthracycline antitumor antibiotic of *Streptomyces glaucescens*. *Proc Natl Acad Sci U S A* 84, 4445–4449.
- Motamedi, H., Shafiee, A., 1998. The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506. *European Journal of Biochemistry* 256, 528–534. <https://doi.org/10.1046/j.1432-1327.1998.2560528.x>
- Müller, C., Nolden, S., Gebhardt, P., Heinzemann, E., Lange, C., Puk, O., Welzel, K., Wohlleben, W., Schwartz, D., 2007. Sequencing and Analysis of the Biosynthetic Gene Cluster of the Lipopeptide Antibiotic Friulimicin in *Actinoplanes friuliensis*. *Antimicrob Agents Chemother* 51, 1028–1037. <https://doi.org/10.1128/AAC.00942-06>
- Münch, R., Hiller, K., Grote, A., Scheer, M., Klein, J., Schobert, M., Jahn, D., 2005. Virtual Footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes. *Bioinformatics* 21, 4187–4189. <https://doi.org/10.1093/bioinformatics/bti635>
- Nah, H.-J., Pyeon, H.-R., Kang, S.-H., Choi, S.-S., Kim, E.-S., 2017. Cloning and Heterologous Expression of a Large-sized Natural Product Biosynthetic Gene Cluster in *Streptomyces* Species. *Front. Microbiol.* 8. <https://doi.org/10.3389/fmicb.2017.00394>
- Nahurira, R., Wang, J., Yan, Y., Jia, Y., Fan, S., Khokhar, I., Eltoukhy, A., 2019. In silico genome analysis reveals the metabolic versatility and biotechnology potential of a halotolerant phthalic acid esters degrading *Gordonia alkanivorans* strain YC-RL2. *AMB Expr* 9, 21. <https://doi.org/10.1186/s13568-019-0733-5>
- Novakova, R., Núñez, L.E., Homerova, D., Knirschova, R., Feckova, L., Rezuchova, B., Sevcikova, B., Menéndez, N., Morís, F., Cortés, J., Kormanec, J., 2018. Increased heterologous production of the antitumoral polyketide mithramycin A by engineered *Streptomyces lividans* TK24 strains. *Appl Microbiol Biotechnol* 102, 857–869. <https://doi.org/10.1007/s00253-017-8642-5>
- NVSS - Mortality - Historical Data [WWW Document], 2019. URL https://www.cdc.gov/nchs/nvss/mortality_historical_data.htm (accessed 6.11.21).
- Ochi, K., Hosaka, T., 2013. New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters. *Appl Microbiol Biotechnol* 97, 87–98. <https://doi.org/10.1007/s00253-012-4551-9>
- Oliynyk, M., Samborsky, M., Lester, J.B., Mironenko, T., Scott, N., Dickens, S., Haydock, S.F., Leadlay, P.F., 2007. Complete genome sequence of the

- erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. *Nat Biotechnol* 25, 447–453. <https://doi.org/10.1038/nbt1297>
- Ōmura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., 2001. Genome sequence of an industrial microorganism *Streptomyces avermitilis*: Deducing the ability of producing secondary metabolites. *PNAS* 98, 12215–12220.
- Ongley, S.E., Bian, X., Zhang, Y., Chau, R., Gerwick, W.H., Müller, R., Neilan, B.A., 2013. High-titer heterologous production in *E. coli* of lyngbyatoxin, a protein kinase C activator from an uncultured marine cyanobacterium. *ACS Chem Biol* 8, 1888–1893. <https://doi.org/10.1021/cb400189j>
- Page, M.J., Di Cera, E., 2008. Serine peptidases: Classification, structure and function. *Cell. Mol. Life Sci.* 65, 1220–1236. <https://doi.org/10.1007/s00018-008-7565-9>
- Pan, S.J., Link, A.J., 2011. Sequence Diversity in the Lasso Peptide Framework: Discovery of Functional Microcin J25 Variants with Multiple Amino Acid Substitutions. *J. Am. Chem. Soc.* 133, 5016–5023. <https://doi.org/10.1021/ja1109634>
- Papu, A., Undap, N., Martinez, N.A., Segre, M.R., Datang, I.G., Kuada, R.R., Perin, M., Yonow, N., Wägele, H., 2020. First Study on Marine Heterobranchia (Gastropoda, Mollusca) in Bangka Archipelago, North Sulawesi, Indonesia. *Diversity* 12, 52. <https://doi.org/10.3390/d12020052>
- Park, H.B., Park, J.-S., Lee, S.I., Shin, B., Oh, D.-C., Kwon, H.C., 2017. Gordonic Acid, a Polyketide Glycoside Derived from Bacterial Coculture of *Streptomyces* and *Gordonia* Species. *J. Nat. Prod.* 80, 2542–2546. <https://doi.org/10.1021/acs.jnatprod.7b00293>
- Paulsen, I.T., Press, C.M., Ravel, J., Kobayashi, D.Y., Myers, G.S.A., Mavrodi, D.V., DeBoy, R.T., Seshadri, R., Ren, Q., Madupu, R., Dodson, R.J., Durkin, A.S., Brinkac, L.M., Daugherty, S.C., Sullivan, S.A., Rosovitz, M.J., Gwinn, M.L., Zhou, L., Schneider, D.J., Cartinhour, S.W., Nelson, W.C., Weidman, J., Watkins, K., Tran, K., Khouri, H., Pierson, E.A., Pierson, L.S., Thomashow, L.S., Loper, J.E., 2005. Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat Biotechnol* 23, 873–878. <https://doi.org/10.1038/nbt1110>
- Paulus, C., Rebets, Y., Zapp, J., Rückert, C., Kalinowski, J., Luzhetskyy, A., 2018. New Alpiniamides From *Streptomyces* sp. IB2014/011-12 Assembled by an Unusual Hybrid Non-ribosomal Peptide Synthetase Trans-AT Polyketide Synthase Enzyme. *Front Microbiol* 9, 1959. <https://doi.org/10.3389/fmicb.2018.01959>
- Pawlik, J.R., Kernan, M.R., Molinski, T.F., Harper, M.K., Faulkner, D.J., 1988. Defensive chemicals of the Spanish dancer nudibranch *Hexabranchus sanguineus* and its egg ribbons: macrolides derived from a sponge diet. *Journal of Experimental Marine Biology and Ecology* 119, 99–109. [https://doi.org/10.1016/0022-0981\(88\)90225-0](https://doi.org/10.1016/0022-0981(88)90225-0)

References

- Petkovic, H., Lill, R.E., Sheridan, R.M., Wilkinson, B., McCORMICK, E.L., McARTHUR, H. a. I., Staunton, J., Leadlay, P.F., Kendrew, S.G., 2003. A Novel Erythromycin, 6-Desmethyl Erythromycin D, Made by Substituting an Acyltransferase Domain of the Erythromycin Polyketide Synthase. *J. Antibiot.* 56, 543–551. <https://doi.org/10.7164/antibiotics.56.543>
- Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E., Khosla, C., 2001a. Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of *E. coli*. *Science* 291, 1790–1792. <https://doi.org/10.1126/science.1058092>
- Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E., Khosla, C., 2001b. Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of *E. coli*. *Science* 291, 1790–1792. <https://doi.org/10.1126/science.1058092>
- Raghuvanshi, R., Bharate, S.B., 2020. Preclinical and Clinical Studies on Bryostatins, A Class of Marine-Derived Protein Kinase C Modulators: A Mini-Review. *Curr Top Med Chem* 20, 1124–1135. <https://doi.org/10.2174/1568026620666200325110444>
- Raja, A., LaBonte, J., Lebbos, J., Kirkpatrick, P., 2003. Daptomycin. *Nature Reviews Drug Discovery* 2, 943–944. <https://doi.org/10.1038/nrd1258>
- Rammelkamp, C.H., Maxon, T., 1942. Resistance of *Staphylococcus aureus* to the Action of Penicillin. *Proceedings of the Society for Experimental Biology and Medicine* 51, 386–389. <https://doi.org/10.3181/00379727-51-13986>
- Rath, C.M., Scaglione, J.B., Kittendorf, J.D., Sherman, D.H., 2010. 1.11 - NRPS/PKS Hybrid Enzymes and Their Natural Products, in: Liu, H.-W. (Ben), Mander, L. (Eds.), *Comprehensive Natural Products II*. Elsevier, Oxford, pp. 453–492. <https://doi.org/10.1016/B978-008045382-8.00725-5>
- Ren, H., Shi, C., Zhao, H., 2020. Computational Tools for Discovering and Engineering Natural Product Biosynthetic Pathways. *iScience* 23. <https://doi.org/10.1016/j.isci.2019.100795>
- Resistance Map: Antibiotic Resistance [WWW Document], 2021. . The Center for Disease, Dynamics Economics & Policy. URL <https://resistancemap.cddep.org/> (accessed 6.15.21).
- Review on Antimicrobial Resistance: Tackling drug-resistant infections globally (Report commissioned by the UK Prime Minister), 2014. , *Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations*.
- Reygaert, W.C., 2018. An overview of the antimicrobial resistance mechanisms of bacteria. *AIMS Microbiol* 4, 482–501. <https://doi.org/10.3934/microbiol.2018.3.482>
- Riyanti, Balansa, W., Liu, Y., Sharma, A., Mihajlovic, S., Hartwig, C., Leis, B., Rieuwpassa, F.J., Ijong, F.G., Wägele, H., König, G.M., Schäberle, T.F., 2020. Selection of sponge-associated bacteria with high potential for the production of antibacterial compounds. *Scientific Reports* 10, 19614. <https://doi.org/10.1038/s41598-020-76256-2>

- Rodríguez-García, A., Combes, P., Pérez-Redondo, R., Smith, M.C.A., Smith, M.C.M., 2005. Natural and synthetic tetracycline-inducible promoters for use in the antibiotic-producing bacteria *Streptomyces*. *Nucleic Acids Res* 33, e87. <https://doi.org/10.1093/nar/gni086>
- Roesener, J.A., Scheuer, P.J., 1986. Ulapualide A and B, extraordinary antitumor macrolides from nudibranch eggmasses. *J. Am. Chem. Soc.* 108, 846–847. <https://doi.org/10.1021/ja00264a052>
- Rowe, S.M., Spring, D.R., 2021. The role of chemical synthesis in developing RiPP antibiotics. *Chem. Soc. Rev.* 50, 4245–4258. <https://doi.org/10.1039/D0CS01386B>
- Rückert, C., Albersmeier, A., Busche, T., Jaenicke, S., Winkler, A., Friðjónsson, Ó.H., Hreggviðsson, G.Ó., Lambert, C., Badcock, D., Bernaerts, K., Anne, J., Economou, A., Kalinowski, J., 2015. Complete genome sequence of *Streptomyces lividans* TK24. *Journal of Biotechnology* 199, 21–22. <https://doi.org/10.1016/j.jbiotec.2015.02.004>
- Ruiz, N., Wu, T., Kahne, D., Silhavy, T.J., 2006. Probing the Barrier Function of the Outer Membrane with Chemical Conditionality. *ACS Chem. Biol.* 1, 385–395. <https://doi.org/10.1021/cb600128v>
- Schatz, A., Bugle, E., Waksman, S.A., 1944. Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria.*†. *Proceedings of the Society for Experimental Biology and Medicine* 55, 66–69. <https://doi.org/10.3181/00379727-55-14461>
- Scherlach, K., Hertweck, C., 2021. Mining and unearthing hidden biosynthetic potential. *Nat Commun* 12, 3864. <https://doi.org/10.1038/s41467-021-24133-5>
- Schmidt, Y., Voort, M. van der, Crüsemann, M., Piel, J., Josten, M., Sahl, H.-G., Miess, H., Raaijmakers, J.M., Gross, H., 2014. Biosynthetic Origin of the Antibiotic Cyclocarbamate Brabantamide A (SB-253514) in Plant-Associated *Pseudomonas*. *ChemBioChem* 15, 259–266. <https://doi.org/10.1002/cbic.201300527>
- Schramma, K.R., Bushin, L.B., Seyedsayamdost, M.R., 2015. Structure and biosynthesis of a macrocyclic peptide containing an unprecedented lysine-to-tryptophan crosslink. *Nat Chem* 7, 431–437. <https://doi.org/10.1038/nchem.2237>
- Schwecke, T., Aparicio, J.F., Molnár, I., König, A., Khaw, L.E., Haydock, S.F., Oliynyk, M., Caffrey, P., Cortés, J., Lester, J.B., 1995. The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *PNAS* 92, 7839–7843. <https://doi.org/10.1073/pnas.92.17.7839>
- Shao, Z., Rao, G., Li, C., Abil, Z., Luo, Y., Zhao, H., 2013. Refactoring the Silent Spectinabilin Gene Cluster Using a Plug-and-Play Scaffold. *ACS Synth Biol* 2, 662–669. <https://doi.org/10.1021/sb400058n>
- Sharan, S.K., Thomason, L.C., Kuznetsov, S.G., Court, D.L., 2009. Recombineering: a homologous recombination-based method of genetic engineering. *Nat Protoc* 4, 206–223. <https://doi.org/10.1038/nprot.2008.227>

- Shen, B., 2003. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Current Opinion in Chemical Biology* 7, 285–295. [https://doi.org/10.1016/S1367-5931\(03\)00020-6](https://doi.org/10.1016/S1367-5931(03)00020-6)
- Shibui, T., Uchida, M., Teranishi, Y., 1988. A New Hybrid Promoter and Its Expression Vector in *Escherichia coli*. *Agricultural and Biological Chemistry* 52, 983–988. <https://doi.org/10.1080/00021369.1988.10868770>
- Siegl, T., Tokovenko, B., Myronovskiy, M., Luzhetskyy, A., 2013. Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes. *Metabolic Engineering* 19, 98–106. <https://doi.org/10.1016/j.ymben.2013.07.006>
- Sikes, D.S., Trumbo, S.T., Peck, S., 2016. Cryptic diversity in the New World burying beetle fauna: *Nicrophorus hebes* Kirby; new status as a resurrected name (Coleoptera: Silphidae: Nicrophorinae). *Arthropod Systematics and Phylogeny* 74, 299–309.
- Silhavy, T.J., Kahne, D., Walker, S., 2010. The bacterial cell envelope. *Cold Spring Harb Perspect Biol* 2, a000414. <https://doi.org/10.1101/cshperspect.a000414>
- Skinninger, M.A., Johnston, C.W., Gunabalasingam, M., Merwin, N.J., Kieliszek, A.M., MacLellan, R.J., Li, H., Ranieri, M.R.M., Webster, A.L.H., Cao, M.P.T., Pfeifle, A., Spencer, N., To, Q.H., Wallace, D.P., Dejong, C.A., Magarvey, N.A., 2020. Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. *Nat Commun* 11, 6058. <https://doi.org/10.1038/s41467-020-19986-1>
- Slipski, C.J., Zhanel, G.G., Bay, D.C., 2018. Biocide Selective TolC-Independent Efflux Pumps in Enterobacteriaceae. *J Membr Biol* 251, 15–33. <https://doi.org/10.1007/s00232-017-9992-8>
- Smith, S., Tsai, S.-C., 2007. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat Prod Rep* 24, 1041–1072. <https://doi.org/10.1039/b603600g>
- Solovyev, Victor, Salamov, A., 2011. Automatic Annotation of Microbial Genomes and Metagenomic Sequences, in: Li, R.W. (Ed.), *Metagenomics and Its Applications in Agriculture, Biomedicine and Environmental Studies*. Nova Science Publishers, Hauppauge, NY, pp. 61–78.
- Solovyev, V, Salamov, A., 2011. Automatic Annotation of Microbial Genomes and Metagenomic Sequences, in: *Metagenomics and Its Applications in Agriculture, Biomedicine and Environmental Studies* (Ed. R.W. Li). Nova Science Publishers, Hauppauge, NY, pp. 61–78.
- Stachelhaus, T., Mootz, H.D., Marahiel, M.A., 1999. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* 6, 493–505. [https://doi.org/10.1016/S1074-5521\(99\)80082-9](https://doi.org/10.1016/S1074-5521(99)80082-9)
- Stackebrandt, E., Rainey, F.A., Ward-Rainey, N.L., 1997. Proposal for a New Hierarchic Classification System, *Actinobacteria classis nov.* *International Journal*

References

of Systematic and Evolutionary Microbiology 47, 479–491. <https://doi.org/10.1099/00207713-47-2-479>

Staunton, J., Weissman, K.J., 2001. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* 18, 380–416. <https://doi.org/10.1039/A909079G>

Staunton, J., Wilkinson, B., 1997. Biosynthesis of Erythromycin and Rapamycin. *Chem. Rev.* 97, 2611–2630. <https://doi.org/10.1021/cr9600316>

Storch, M., Casini, A., Mackrow, B., Fleming, T., Trewhitt, H., Ellis, T., Baldwin, G.S., 2015. BASIC: A New Biopart Assembly Standard for Idempotent Cloning Provides Accurate, Single-Tier DNA Assembly for Synthetic Biology. *ACS Synth. Biol.* 4, 781–787. <https://doi.org/10.1021/sb500356d>

Studier, F.W., Moffatt, B.A., 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology* 189, 113–130. [https://doi.org/10.1016/0022-2836\(86\)90385-2](https://doi.org/10.1016/0022-2836(86)90385-2)

Sucipto, H., Pogorevc, D., Luxenburger, E., Wenzel, S.C., Müller, R., 2017. Heterologous production of myxobacterial α -pyrone antibiotics in *Myxococcus xanthus*. *Metab Eng* 44, 160–170. <https://doi.org/10.1016/j.ymben.2017.10.004>

Sudek, S., Lopanik, N.B., Waggoner, L.E., Hildebrand, M., Anderson, C., Liu, H., Patel, A., Sherman, D.H., Haygood, M.G., 2007. Identification of the putative bryostatin polyketide synthase gene cluster from “*Candidatus endobugula sertula*”, the uncultivated microbial symbiont of the marine bryozoan *Bugula neritina*. *J. Nat. Prod.* 70, 67–74. <https://doi.org/10.1021/np060361d>

Sulavik, M.C., Houseweart, C., Cramer, C., Jiwani, N., Murgolo, N., Greene, J., DiDomenico, B., Shaw, K.J., Miller, G.H., Hare, R., Shimer, G., 2001. Antibiotic susceptibility profiles of *Escherichia coli* strains lacking multidrug efflux pump genes. *Antimicrob Agents Chemother* 45, 1126–1136. <https://doi.org/10.1128/AAC.45.4.1126-1136.2001>

Süssmuth, R.D., Mainz, A., 2017. Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew Chem Int Ed Engl* 56, 3770–3821. <https://doi.org/10.1002/anie.201609079>

Tan, G.-Y., Liu, T., 2017. Rational synthetic pathway refactoring of natural products biosynthesis in actinobacteria. *Metabolic Engineering* 39, 228–236. <https://doi.org/10.1016/j.ymben.2016.12.006>

Teichert, R.W., Olivera, B.M., 2010. Natural Products and Ion Channel Pharmacology. *Future Med Chem* 2, 731–744. <https://doi.org/10.4155/fmc.10.31>

Tsukamura, M., 1971. Proposal of a new genus, *Gordona*, for slightly acid-fast organisms occurring in sputa of patients with pulmonary disease and in soil. *J Gen Microbiol* 68, 15–26. <https://doi.org/10.1099/00221287-68-1-15>

Udwary, D.W., Zeigler, L., Asolkar, R.N., Singan, V., Lapidus, A., Fenical, W., Jensen, P.R., Moore, B.S., 2007. Genome sequencing reveals complex secondary

metabolome in the marine actinomycete *Salinispora tropica*. PNAS 104, 10376–10381.

Ueoka, R., Uria, A.R., Reiter, S., Mori, T., Karbaum, P., Peters, E.E., Helfrich, E.J.N., Morinaka, B.I., Gugger, M., Takeyama, H., Matsunaga, S., Piel, J., 2015. Metabolic and evolutionary origin of actin-binding polyketides from diverse organisms. Nat Chem Biol 11, 705–712. <https://doi.org/10.1038/nchembio.1870>

Umarov, R.K., Solovyev, V.V., 2017a. Recognition of prokaryotic and eukaryotic promoters using convolutional deep learning neural networks. PLOS ONE 12, e0171410. <https://doi.org/10.1371/journal.pone.0171410>

Umarov, R.K., Solovyev, V.V., 2017b. Recognition of prokaryotic and eukaryotic promoters using convolutional deep learning neural networks. PLOS ONE 12, e0171410. <https://doi.org/10.1371/journal.pone.0171410>

Undap, N., Papu, A., Schillo, D., Ijong, F.G., Kaligis, F., Lepar, M., Hertzner, C., Böhringer, N., König, G.M., Schäberle, T.F., Wägele, H., 2019. First Survey of Heterobranch Sea Slugs (Mollusca, Gastropoda) from the Island Sangihe, North Sulawesi, Indonesia. Diversity 11, 170. <https://doi.org/10.3390/d11090170>

Uden, G., Schirawski, J., 1997. The oxygen-responsive transcriptional regulator FNR of *Escherichia coli*: the search for signals and reactions. Mol Microbiol 25, 205–210. <https://doi.org/10.1046/j.1365-2958.1997.4731841.x>

van Heel, A.J., de Jong, A., Montalbán-López, M., Kok, J., Kuipers, O.P., 2013. BAGEL3: automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. Nucleic Acids Research 41, W448–W453. <https://doi.org/10.1093/nar/gkt391>

van Heel, A.J., Kloosterman, T.G., Montalban-Lopez, M., Deng, J., Plat, A., Baudu, B., Hendriks, D., Moll, G.N., Kuipers, O.P., 2016. Discovery, Production and Modification of Five Novel Lantibiotics Using the Promiscuous Nisin Modification Machinery. ACS Synth. Biol. 5, 1146–1154. <https://doi.org/10.1021/acssynbio.6b00033>

Ventola, C.L., 2015. The Antibiotic Resistance Crisis. P T 40, 277–283.

Vior, N.M., Cea-Torrescassana, E., Eyles, T.H., Chandra, G., Truman, A.W., 2020. Regulation of Botromycin Biosynthesis Involves an Internal Transcriptional Start Site and a Cluster-Situated Modulator. Front Microbiol 11, 495. <https://doi.org/10.3389/fmicb.2020.00495>

Vogel, H., Shukla, S.P., Engl, T., Weiss, B., Fischer, R., Steiger, S., Heckel, D.G., Kaltenpoth, M., Vilcinskas, A., 2017. The digestive and defensive basis of carcass utilization by the burying beetle and its microbiota. Nat Commun 8. <https://doi.org/10.1038/ncomms15186>

von Hoermann, C., Steiger, S., Müller, J.K., Ayasse, M., 2013. Too Fresh Is Unattractive! The Attraction of Newly Emerged *Nicrophorus vespilloides* Females to Odour Bouquets of Large Cadavers at Various Stages of Decomposition. PLoS One 8. <https://doi.org/10.1371/journal.pone.0058524>

Wägele, H., Klusmann-Kolb, A., Verbeek, E., Schrödl, M., 2014. Flashback and foreshadowing—a review of the taxon Opisthobranchia. *Org Divers Evol* 14, 133–149. <https://doi.org/10.1007/s13127-013-0151-5>

Wägele, H., Willan, R.C., 2000. Phylogeny of the Nudibranchia. *Zoological Journal of the Linnean Society* 130, 83–181. <https://doi.org/10.1006/zjls.1999.0214>

Waldron, K.J., Robinson, N.J., 2009. How do bacterial cells ensure that metalloproteins get the correct metal? *Nature Reviews Microbiology* 7, 25–35. <https://doi.org/10.1038/nrmicro2057>

Wang, H., Fewer, D.P., Holm, L., Rouhiainen, L., Sivonen, K., 2014. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *PNAS* 111, 9259–9264. <https://doi.org/10.1073/pnas.1401734111>

Wang, J., Zhang, R., Chen, X., Sun, X., Yan, Y., Shen, X., Yuan, Q., 2020. Biosynthesis of aromatic polyketides in microorganisms using type II polyketide synthases. *Microbial Cell Factories* 19, 110. <https://doi.org/10.1186/s12934-020-01367-4>

Wang, M., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y., Nguyen, D.D., Watrous, J., Kapon, C.A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A.V., Meehan, M.J., Liu, W.-T., Crüsemann, M., Boudreau, P.D., Esquenazi, E., Sandoval-Calderón, M., Kersten, R.D., Pace, L.A., Quinn, R.A., Duncan, K.R., Hsu, C.-C., Floros, D.J., Gavilan, R.G., Kleigrew, K., Northen, T., Dutton, R.J., Parrot, D., Carlson, E.E., Aigle, B., Michelsen, C.F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., McLean, J., Piel, J., Murphy, B.T., Gerwick, L., Liaw, C.-C., Yang, Y.-L., Humpf, H.-U., Maansson, M., Keyzers, R.A., Sims, A.C., Johnson, A.R., Sidebottom, A.M., Sedio, B.E., Klitgaard, A., Larson, C.B., Boya P, C.A., Torres-Mendoza, D., Gonzalez, D.J., Silva, D.B., Marques, L.M., Demarque, D.P., Pociute, E., O'Neill, E.C., Briand, E., Helfrich, E.J.N., Granatosky, E.A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J.J., Zeng, Y., Vorholt, J.A., Kurita, K.L., Charusanti, P., McPhail, K.L., Nielsen, K.F., Vuong, L., Elfeki, M., Traxler, M.F., Engene, N., Koyama, N., Vining, O.B., Baric, R., Silva, R.R., Mascuch, S.J., Tomasi, S., Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V., Williams, P.G., Dai, J., Neupane, R., Gurr, J., Rodríguez, A.M.C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B.M., Almaliti, J., Allard, P.-M., Phapale, P., Nothias, L.-F., Alexandrov, T., Litaudon, M., Wolfender, J.-L., Kyle, J.E., Metz, T.O., Peryea, T., Nguyen, D.-T., VanLeer, D., Shinn, P., Jadhav, A., Müller, R., Waters, K.M., Shi, W., Liu, X., Zhang, L., Knight, R., Jensen, P.R., Palsson, B.Ø., Pogliano, K., Lington, R.G., Gutiérrez, M., Lopes, N.P., Gerwick, W.H., Moore, B.S., Dorrestein, P.C., Bandeira, N., 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotech* 34, 828–837. <https://doi.org/10.1038/nbt.3597>

Weber, G., Schörgendorfer, K., Schneider-Scherzer, E., Leitner, E., 1994. The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is

encoded by a giant 45.8-kilobase open reading frame. *Curr Genet* 26, 120–125. <https://doi.org/10.1007/BF00313798>

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Bruccoleri, R., Lee, S.Y., Fischbach, M.A., Müller, R., Wohlleben, W., Breitling, R., Takano, E., Medema, M.H., 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res* 43, W237–W243. <https://doi.org/10.1093/nar/gkv437>

Weinstock, M.T., Heseck, E.D., Wilson, C.M., Gibson, D.G., 2016. *Vibrio natriegens* as a fast-growing host for molecular biology. *Nat Methods* 13, 849–851. <https://doi.org/10.1038/nmeth.3970>

Weissman, K.J., 2009. Chapter 1 Introduction to Polyketide Biosynthesis, in: *Methods in Enzymology, Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates*. Academic Press, pp. 3–16. [https://doi.org/10.1016/S0076-6879\(09\)04601-1](https://doi.org/10.1016/S0076-6879(09)04601-1)

WHO News Release, 2017. WHO publishes list of bacteria for which new antibiotics are urgently needed [WWW Document]. URL <https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed> (accessed 6.13.21).

WHO News Release, 2016. At UN, global leaders commit to act on antimicrobial resistance [WWW Document]. URL <https://www.who.int/news/item/21-09-2016-at-un-global-leaders-commit-to-act-on-antimicrobial-resistance> (accessed 6.12.21).

Wisedchaisri, G., Holmes, R.K., Hol, W.G.J., 2004. Crystal structure of an IdeR-DNA complex reveals a conformational change in activated IdeR for base-specific interactions. *J Mol Biol* 342, 1155–1169. <https://doi.org/10.1016/j.jmb.2004.07.083>

Wolf, F., Bauer, J.S., Bendel, T.M., Kulik, A., Kalinowski, J., Gross, H., Kaysser, L., 2017. Biosynthesis of the β -Lactone Proteasome Inhibitors Belactosin and Cystargolide. *Angewandte Chemie International Edition* 56, 6665–6668. <https://doi.org/10.1002/anie.201612076>

Wu, T., Malinverni, J., Ruiz, N., Kim, S., Silhavy, T.J., Kahne, D., 2005. Identification of a multicomponent complex required for outer membrane biogenesis in *Escherichia coli*. *Cell* 121, 235–245. <https://doi.org/10.1016/j.cell.2005.02.015>

Yamanaka, K., Reynolds, K.A., Kersten, R.D., Ryan, K.S., Gonzalez, D.J., Nizet, V., Dorrestein, P.C., Moore, B.S., 2014. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc Natl Acad Sci U S A* 111, 1957–1962. <https://doi.org/10.1073/pnas.1319584111>

Zeitlin, P.L., Boyle, M.P., Guggino, W.B., Molina, L., 2004. A phase I trial of intranasal Moli1901 for cystic fibrosis. *Chest* 125, 143–149. <https://doi.org/10.1378/chest.125.1.143>

Zhang, H., Wang, Y., Wu, J., Skalina, K., Pfeifer, B.A., 2010. Complete Biosynthesis of Erythromycin A and Designed Analogs Using *E. coli* as a Heterologous Host.

Chemistry & Biology 17, 1232–1240.
<https://doi.org/10.1016/j.chembiol.2010.09.013>

Zhang, J.J., Tang, X., Moore, B.S., 2019a. Genetic platforms for heterologous expression of microbial natural products. *Nat. Prod. Rep.* 36, 1313–1332. <https://doi.org/10.1039/C9NP00025A>

Zhang, J.J., Yamanaka, K., Tang, X., Moore, B.S., 2019b. Direct cloning and heterologous expression of natural product biosynthetic gene clusters by transformation-associated recombination. *Methods Enzymol* 621, 87–110. <https://doi.org/10.1016/bs.mie.2019.02.026>

Zhang, L., Hashimoto, T., Qin, B., Hashimoto, J., Kozone, I., Kawahara, T., Okada, M., Awakawa, T., Ito, T., Asakawa, Y., Ueki, M., Takahashi, S., Osada, H., Wakimoto, T., Ikeda, H., Shin-ya, K., Abe, I., 2017. Characterization of Giant Modular PKSs Provides Insight into Genetic Mechanism for Structural Diversification of Aminopolyol Polyketides. *Angewandte Chemie International Edition* 56, 1740–1745. <https://doi.org/10.1002/anie.201611371>

Zhang, Z., Schwartz, S., Wagner, L., Miller, W., 2000. A Greedy Algorithm for Aligning DNA Sequences. *Journal of Computational Biology* 7, 203–214. <https://doi.org/10.1089/10665270050081478>

Zhao, L., Vo, T.D., Kaiser, M., Bode, H.B., 2020. Phototemtide A, a Cyclic Lipopeptide Heterologously Expressed from *Photorhabdus temperata* Meg1, Shows Selective Antiprotozoal Activity. *ChemBioChem* 21, 1288–1292. <https://doi.org/10.1002/cbic.201900665>

Zhou, K., Zhou, L., Lim, Q., 'En, Zou, R., Stephanopoulos, G., Too, H.-P., 2011. Novel reference genes for quantifying transcriptional responses of *Escherichia coli* to protein overexpression by quantitative PCR. *BMC Mol Biol* 12, 18. <https://doi.org/10.1186/1471-2199-12-18>

Zhou, Q., Bräuer, A., Adihou, H., Schmalhofer, M., Saura, P., Grammbitter, G.L.C., Kaila, V.R.I., Groll, M., Bode, H.B., 2019. Molecular mechanism of polyketide shortening in anthraquinone biosynthesis of *Photorhabdus luminescens*. *Chem. Sci.* 10, 6341–6349. <https://doi.org/10.1039/C9SC00749K>

VII. Appendix

Table S1. Primers used in this study

| Name | Sequence (5' → 3') |
|---------------|---|
| 118-119-F | CCGCTGTCTCCCATTCTGAGGAGGAATGGCCGATGTGTTTGA AT |
| 118-119-R | ATTCAAACACATCGGCCATTCTCCTCAGGAATGGGAGACAGC GG |
| 118-For | CGGCAAAGCTGTTCACTTCG |
| 118-Rev | AGTTGGGTGACGGCTCATTC |
| 119-123-F | TGCGGGAGTCGTTGGGCTGAGGAGGCGTCCGGAATGGATTCC GCC |
| 119-123-R | GGCGGAATCCATTCCGGACGCCTCCTCAGCCCAACGACTCCC GCA |
| 119-pGM-R | GCTCGGTACCCGGGGATCTATCAGCCCAACGACTCCCGCA |
| 123-pGM-R | GCTCGGTACCCGGGGATCTATCAGAGTCCCTGCGGGACGT |
| 129-For | TGGCACAGCATGACTATTGG |
| 129-Rev | CGAACCGACCGAGATAGTTG |
| 20.206-For | ATGCGGTGCTGCATCACATC |
| 20.206-Rev | AGCCGTTGAACGCGAACATC |
| 20.224-For | GCTGGCACACGGGAATGTTG |
| 20.224-Rev | CCAGCGTGACCTCGAAATGG |
| 23.127-For | CACTCGACATCTCCGAACTC |
| 23.127-Rev | GTCAGGTAGACGACCGTATC |
| 23.134-For | ATCGGCGAGCTCTACCTCAG |
| 23.134-Rev | GTGGTCATCCGGTGCCTTAC |
| 23.144-For | ACTGCTCGTGCATTCTCATC |
| 23.144-Rev | GATGACGTTGGCCTTGTATC |
| 3589-3604-F | CGGGCAGATCCAACAATTGAGGAGGAGGCCGTAATGGTACTCC AA |
| 3589-3604-R | TTGGAGTACCATTACGGCCTCCTCCTCAATTGTTGGATCTGCCC G |
| 3589-F | TGAACCGCTCCTTGATAACC |
| 3597-R | CTGATCTGCGCAATTAAGGC |
| 3598-3597-F | CCTAATATCTTGGTTCACTCATCTTTTGAAGGGAAACGTTCCACC |
| 3598-3597-R | GGTGAACGTTTCCCTTCAAAGATGAGTGAACCAAGATATTAGG |
| 3598-F | ACATGGGCGAGTTTATCTCC |
| 3603-R | AGGCCGTCACAGTTTAACAG |
| 3604-R | TGAAGCAGACGAGGATCTAC |
| 3608-3611-F | TCAGAGGCGATTTATGGTGAGGAGGAGATCAGAGTGTGCACTA AC |
| 3608-3611-R | GTTAGTGCACACTCTGATCTCCTCCTCACCATAAATCGCCTCTG A |
| 3609-pCAP03-R | TATGTAGCTTTTCGACATATCAAGAAATATCACCCAATA |
| 37.231-For | TCAGGGCGTTCCACACTTTG |
| 37.231-Rev | AGTGACCGCGACTGTTGTAG |
| 37.243-For | CAGGCTGATCAATGGCTATG |
| 37.243-Rev | TCGGCGAATCGAAGATATCC |
| 37.252-For | TCACTGGTCGCTGAATGAAC |
| 37.252-Rev | CTGGAAGTGCAGCATAAACG |
| 52.120-R | CGACACGTCCACGTGCATAG |

| | |
|-------------------------------|---|
| 9A215F | CCGTGTACCCGTTGTTGTTG |
| 9A215R | CAGGGCACGTCCTTCTATGG |
| 9A227F | CCAAGCTGGGCCGTAACAAG |
| 9A227R | CCGTACCCGAGGTCATGTAG |
| 9B239F | TGCGTGA CTCTGCTGTTAGAGG |
| 9B239R | ACGGTTCTCTGCTGGTGTTTC |
| 9B242F | GTGACAACGCAATCCGTCTC |
| 9B242R | CATCACGACCGCGAACAAAC |
| 9C242F | TGGCCCGAGAGGAGTAGTTG |
| 9C242R | CGGTCCGGCAGAGGAATTGTG |
| 9CD247F2 | AGCATCACCTCGAAACTAGC |
| 9CD247F3 | GTTTCGTGGCGTTTCGTTCCAG |
| 9CD247R2 | TGGCAGGTCGTCGATGATAC |
| 9CD247R3 | CCGTGGTGTCCGAGGATTAC |
| 9D255F | CTCGCACGGCATCGTAGAAC |
| 9D255R | GTCGGCCATACGGTAGACAG |
| ADC5-GFP-F2 | TGGATGAACTATACAAATAACCACCGCTGAGCAATAACTA |
| ADC5-GFP-R | CGCATGCTTAATTTCTCCTCTTACGCCGCGATGGTTTTGTT |
| ADC9-F | GTAAGGAACGGTAGCCATAAATGGACACAATAATCCCCAT |
| ADC9-R | ATGGGGATTATTGTGTCCATTTATGGCTACCGTTCCTTAC |
| ADC-F10 | GTATAAGAAGGAGATATACAAAACCTACTACATTTTAGCG |
| ADC-F2 | AAAGCTTCCAGGAAATTTAAAAGGAGAGCCATAAATGAATGCTA TT |
| ADC-F7 | GTATAAGAAGGAGATATACAATGCATAATACCTTAAATGA |
| ADC-F9 | GTATAAGAAGGAGATATACACCGATGATATACTTTTATTA |
| ADC-R1 | AATAGCATTCAATTTATGGCTCTCCTTTTAAATTTCTGGAAGCTT T |
| ADC-R7 | TGCTCAGCGGTGGCAGCAGCTTACGCCGCGATGGTTTTGTT |
| A-F3 | TCAAGAAGCACTCAATTCTC |
| A-R3 | GTGATCTCAGGGATCTTAGG |
| bamA-recF | ACTATCTGGATCGCGGTTATGC |
| bamA-recR | TTCACAGCAGTCTGGATACGAG |
| Daro.MBP.Xa.F | TTTAAGAAGGAGATATACATATGAAAATAGAAGAAGGTAACTG GTAATCTGG |
| Daro.MBP.Xa.R | CCTTCCCTCGATCCCGAGGTTGTTGTTATTGTTATTGT |
| Daro.Xa.pro.F | ACCTCGGGATCGAGGGAAGGATGCACAACACCTCTATCATCAA C |
| Daro.Xa.pro.R | TCCACCAGTCATGCTAGCCATTAGATTTCTGGAAGATTTAGA CCAGT |
| E-F3 | GGCCAACATCCCATAAAGTC |
| E-R3 | ACTTCCTCCAGGATCATCAC |
| ermE-unitest | ATCTTGACGGCTGGCGAGAG |
| GFP-F | AACAAACCATCGCGGCGTAAGAGGAGAAATTAAGCATGCG |
| GFP-R2 | TAGTTATTGCTCAGCGGTGGTTATTTGTATAGTTCATCCATGCC |
| J23101-ADC5- GFP-R | ACTCATTAGGCACCGGGATCTCGACCGATGCCCTTGAGAG |
| J23101-F | TTTACAGCTAGCTCAGTCCTAGGTATTATGCTAGCCCCATCTTA GTATATTAGTT |
| J23101-F2 | GATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTT G |
| J23101-R | GCTAGCATAATACCTAGGACTGAGCTAGCTGTAAACAACGCAAT TAATGTAAGTT |
| J23101-R2 | CATTGTATATCTCCTTCTTATACTTAACTAATACTAAGATGGG |
| pCAP03-3603-F | CATGGTATAAATAGTGCGGAGGACGAACTATGAACTCCACA |

Appendix

| | |
|------------------------|--|
| pCAP03-seq | GCTTTGCCGATGTTACTTGG |
| pCC2FOS-For | GTACAACGACACCTAGAC |
| pCC2FOS-Rev | CAGGAAACAGCCTAGGAA |
| pGM1202-ermE*-F | GAAATCTCCCGCCCTGGCCGGCTAGCCGCGGTTCGATCTTG |
| pGM-ermE*-118 | GAAATCTCCCGCCCTGGCCGGCTAGCCGCGGTTCGATCTTGAC GGCTGGCGAGAGGTGCGGGGAGGATCTGACCGACGCGGTCC ACACGTGGCACC GCGATGCTGTTGTGGGCACAATCGTGCCGG TTGGTAGGATCCAGCGGGTAGGAGGTTGGTCCCGTGCTCGGT GTG |
| pGM-Rseq | TCACTCCGCTGAAACTGTTG |
| pGM-seq | CGAGCGTTCTGAACAAATCC |
| pPro-F | AACTTACATTAATTGCGTTGCCGGATAAAGCGTTCGCG |
| pPro-R | AACTAATACTAAGATGGGAGCCCATCCTTTGTTATCAA |
| pro.CO.F | GTATAAGAAGGAGATATACAATGCACAACACCTCTATCAT |
| pro.CO.R | ATCGGGATGATGGTGTCCATTTAGATTTCTGGAAAGATTTAGA C |
| rrsA-F3 | CTCTTGCCATCGGATGTGCCCA |
| rrsA-R3 | CCAGTGTGGCTGGTCATCCTCTCA |
| RS.CO.F | GTCTAAATCTTTCCAGGAAATCTAAATGGACACCATCATCCCGA T |
| RS.CO.R | TGCTCAGCGGTGGCAGCAGCAGCAGCGATGGTCTGTTTGA |
| T5-F | TTGCTTTTCAGGAAAATTTTTCTGTATAATAGATTCCCCTAG TATATTAGTT |
| T5-R | GAATCTATTATACAGAAAATTTTCTGAAAGCAACAACGCAATT AATGTAAGTT |
| tcp830-unitest | CAGCTGTTGGCTACTCTATC |
| ToICKO-F | ATGAAGAAATTGCTCCCCATTCTTATCGGCCTGAGCCTTTATTC CGGGGATCCGTCGACC |
| ToICKO-R | TCAGTTACGGAAAGGGTTATGACCGTTACTGGTGGTAGTGTGT AGGCTGGAGCTGCTTC |
| YerA-F4 | GTATAAGAAGGAGATATACAAGGAGGTTATTTAAATGGAGAATT AT |
| YerA-R4 | TGCTCAGCGGTGGCAGCAGCTCAGTAGATACTGGCGATAT |

Table S2. Bacterial strains used in this study

| Strain | Source/Reference |
|---|--|
| <i>E. coli</i> BAP1 | (Pfeifer <i>et al.</i> , 2001b) |
| <i>E. coli</i> BAP1 Δ tolC:: aac(3)-IV (Apr ^R) | This study |
| <i>E. coli</i> BL21(DE3) | Merck KGaA (Darmstadt, Germany) |
| <i>E. coli</i> BW25113 | (Gust <i>et al.</i> , 2006) |
| <i>E. coli</i> EPI300-T1 ^R | Epicentre Biotechnologies (Madison, USA) |
| <i>E. coli</i> ET12567 | (Zhang <i>et al.</i> , 2019b) |
| <i>E. coli</i> MG1655 BamA6 | (Ruiz <i>et al.</i> , 2006) |
| <i>E. coli</i> Rosetta TM (DE3) | Merck KGaA (Darmstadt, Germany) |
| <i>E. coli</i> Rosetta TM (DE3)(Dar ^R) | This study |
| <i>E. coli</i> TOP10 | Invitrogen (California, USA) |
| <i>Gordonia</i> sp. Bu15_44 | (Böhringer <i>et al.</i> , 2017) |
| <i>Gordonia terrae</i> Bu15_45 | (Böhringer <i>et al.</i> , 2017) |
| <i>P. khanii</i> DSM3369 | DSMZ (Braunschweig, Germany) |
| <i>P. khanii</i> HGB1456 | (Imai <i>et al.</i> , 2019) |
| <i>S. cerevisiae</i> VL6-48N | (Zhang <i>et al.</i> , 2019b) |
| <i>S. coelicolor</i> M1146 | (Gomez-Escribano and Bibb, 2011) |
| <i>S. lividans</i> TK24 | (Rückert <i>et al.</i> , 2015) |
| Strain 39 (putatively <i>Glutamicibacter arilaitensis</i>) | (Heise, 2019) |
| <i>V. natriegens</i> Vmax TM | Synthetic Genomics (California, USA) |
| <i>Y. frederiksenii</i> ATCC 33641 | LGC Standards GmbH (Wesel, Germany) |

Table S3. Predicted function of *orfs* in cluster 52.

| Gene | Size (bp) | Protein function prediction |
|---------------|-----------|--|
| <i>orf118</i> | 1017 | Iron compound ABC transporter, periplasmic binding protein |
| <i>nbtT</i> | 1668 | NRPS A domain for 2,3-dihydroxybenzoic acid |
| <i>nbtS</i> | 1341 | Salicylate synthase |
| <i>nbtF</i> | 3630 | NRPS domains (monomer prediction: threonine/cysteine) |
| <i>nbtG</i> | 1326 | L-lysine 6-monooxygenase |
| <i>nbtH</i> | 645 | Lysine acetyltransferase |
| <i>nbtA</i> | 795 | Thioesterase |
| <i>nbtB</i> | 1332 | PKS ketosynthase domain |
| <i>nbtC</i> | 3246 | PKS domains for condensation of fatty acyl chain |
| <i>orf127</i> | 981 | Protein of unknown function |
| <i>nbtD</i> | 5349 | NRPS domains (monomer prediction: lysine) |
| <i>nbtE</i> | 4593 | NRPS domains (monomer prediction: lysine) |

Table S4. IdeR Iron box from cluster 52

| Target gene | Distance from start codon | Sequence |
|--------------------|----------------------------------|---------------------|
| <i>orf117</i> | -63 | GAAGGTCAGGCTAACCTGG |
| <i>nbtT</i> | -72 | TTAGGCCAGGCTAATCTAC |
| <i>nbtSFG</i> | -73 | TTAGGCCAGGCTAATCTAC |
| <i>nbtH</i> | -52 | TCACGCCAGGCTAACCCAT |
| <i>nbtABC</i> | -53 | TCACGCCAGGCTAACCCAT |
| <i>orf127</i> | -19 | TTAGCGTGACCTTACTTGA |
| <i>nbtDE</i> | -33 | AGAGGTAACCCTAACCAAC |

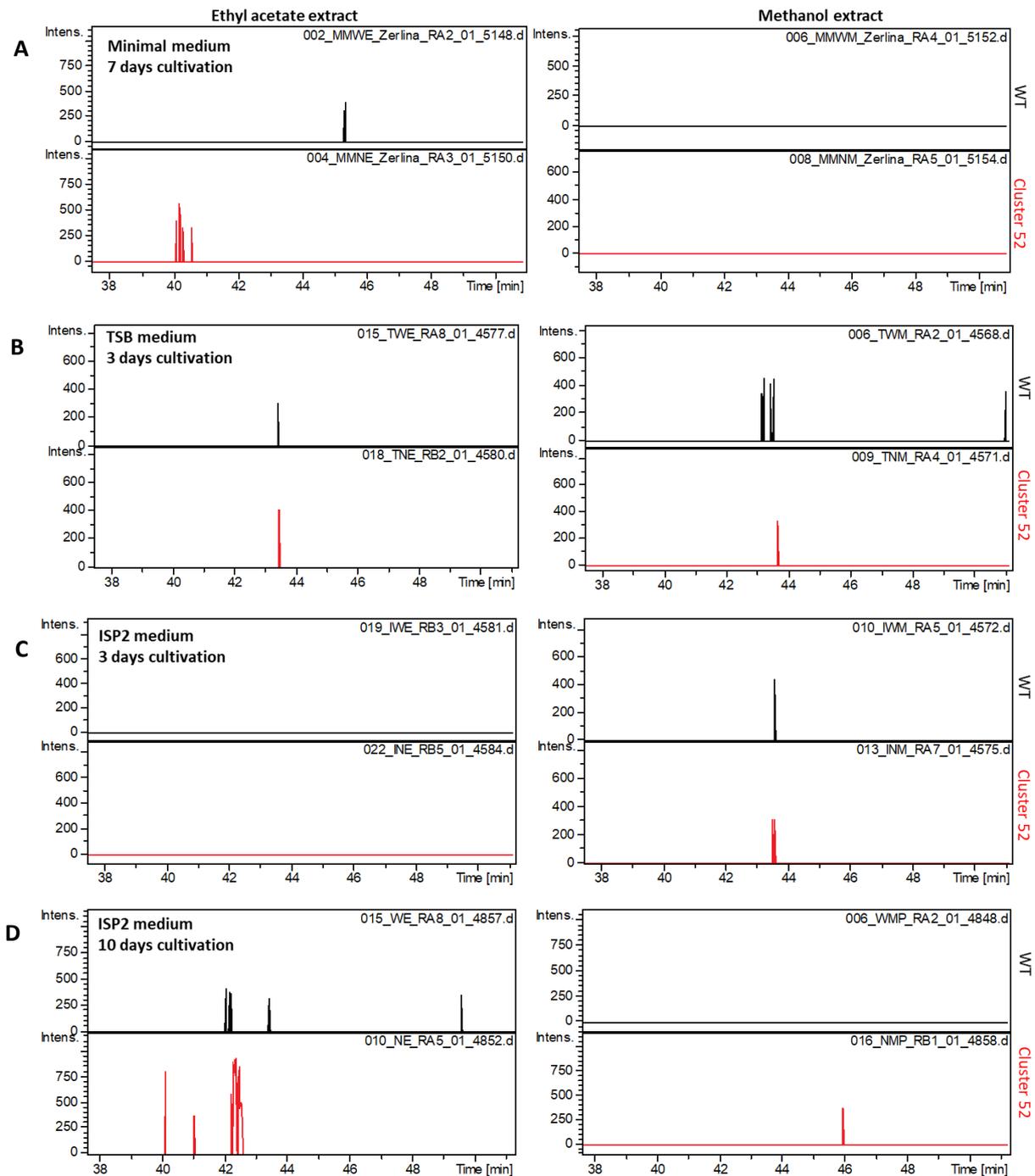


Figure S1: Extracted ion chromatogram of nocobactin NA by *S. coelicolor* M1146 with cluster 52 integrated into its genome. Wild-type *S. coelicolor* M1146 (black) and transgenic strain carrying cluster 52 (red) were cultivated in different conditions, *i.e.* in minimal medium for seven days (A), in TSB for three days (B), and in ISP2 for three (C) and ten days (D). Ferri-nocobactin NA-a (m/z 797.3283 [M+H]⁺, 819.3100 [M+Na]⁺) was expected to appear at 42.9 minutes and ferri-nocobactin NA-b (m/z 825.3583 [M+H]⁺, 847.3416 [M+Na]⁺) at 44.6 minutes (as shown in **Figure III.20**), but could not be detected in ethyl acetate and methanol extracts of the cultures. Several other cultivation conditions showed similar results and are not shown here.

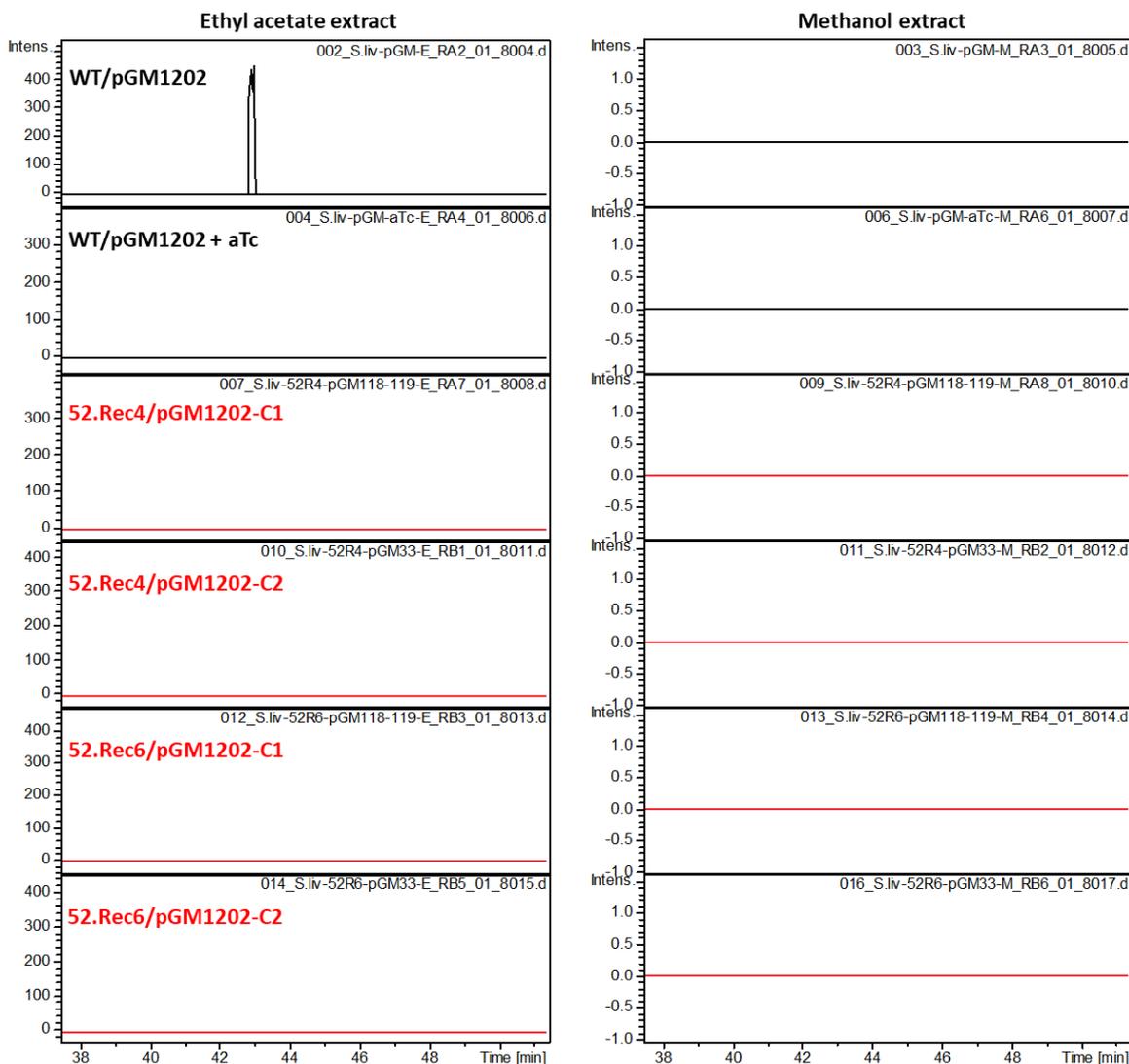
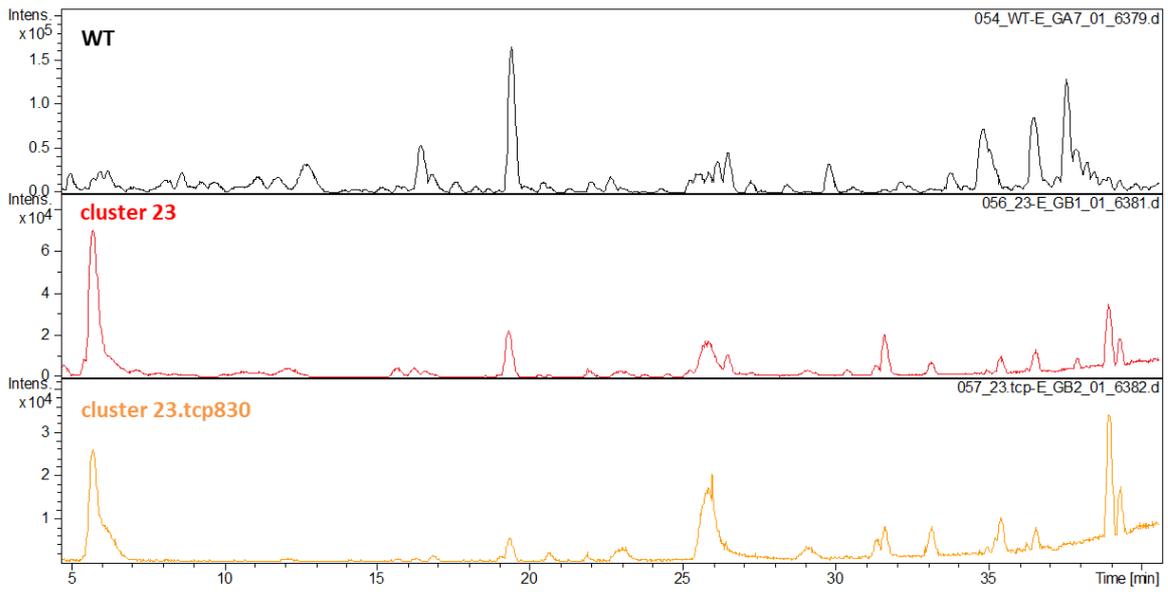


Figure S2: Extracted ion chromatogram of nocobactin NA by *S. lividans* TK24 with cluster 52 with P_{tcp830} (52.Rec4) and P_{ermE} (52.Rec6) integrated into its genome. Wild-type *S. coelicolor* M1146 (black) and transgenic strains carrying cluster 52 (red) were cultivated in ISP2 for one day, before the addition of anhydrotetracycline to induce the P_{tcp830} . The cultures were continued to be cultivated for another six days before being harvested. Ferri-nocobactin NA-a (m/z 797.3283 [M+H]⁺, 819.3100 [M+Na]⁺) was expected to appear at 42.9 minutes and ferri-nocobactin NA-b (m/z 825.3583 [M+H]⁺, 847.3416 [M+Na]⁺) at 44.6 minutes, but could not be detected in ethyl acetate and methanol extracts of the cultures. pGM1202-C1 carried *orf118* and *nbtT*, while pGM1202-C2 carried *orf11*, *nbtT* and *nbtH*. Similar experiments were also performed with *S. coelicolor* M1146 as the expression host, yielding similar results that nocobactin could not be detected from LC-MS analysis.

Ethyl acetate extract



Methanol extract

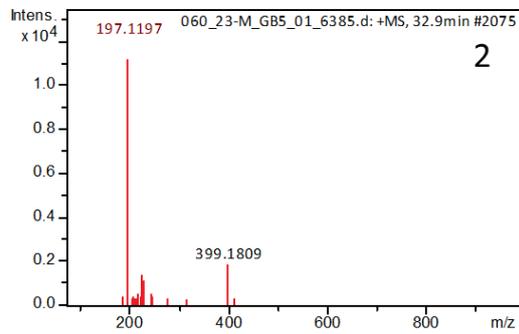
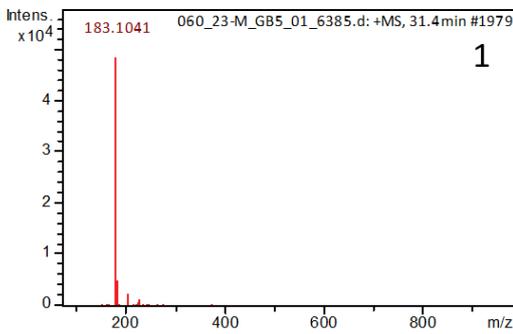
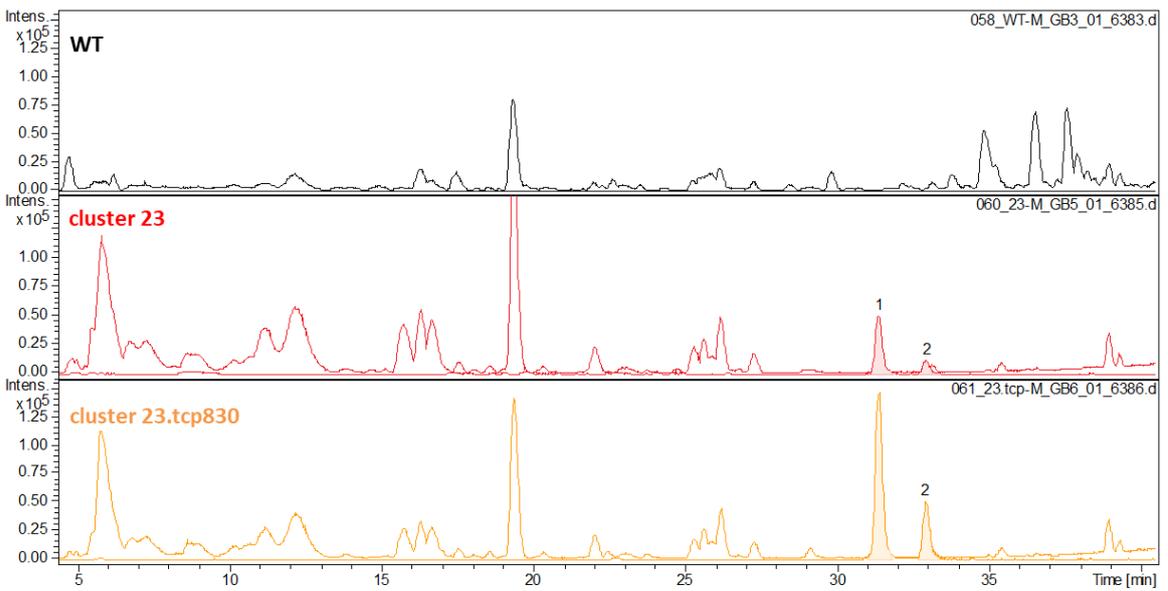


Figure S3: Heterologous expression of cluster 23 integrated into *S. coelicolor* M1146 genome. Wild-type *S. coelicolor* M1146 (black), as well as transgenic strains carrying cluster 23 (red) and cluster 23 with P_{tcp830} (orange) were cultivated in ISP2 for three days, before the addition of anhydrotetracycline to induce the P_{tcp830} . The cultures were continued to be cultivated for another three days before being harvested. The base peak chromatogram of the ethyl acetate extracts of the transgenic strains did not show additional peaks compared to the wild type control. Meanwhile, the base peak chromatogram of the methanol extracts of the transgenic strains showed two additional peaks compared to the wild type control. However, these two peaks were identified as Germicidin A/B (peak 1) and Germicidin C (peak 2), whose BGC is present in *S. coelicolor* M1146 genome.

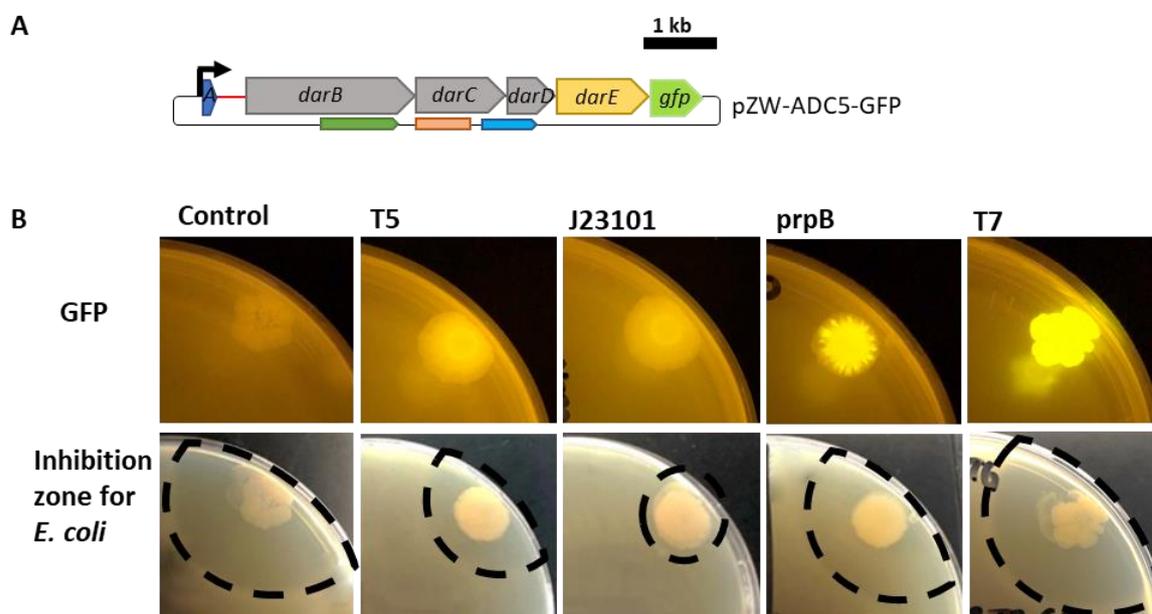


Figure S4: Qualitative analysis of different promoters used. **A**, schematic vector map; the DNA sequence encoding for GFP was inserted downstream of the DAR BGC to serve as reporter. Black arrow shows the position where the respective constitutive promoter (T5 and J23101) and inducible promoter (T7 and prpB) was integrated. **B**, Photographs of agar plates with the respective clone. GFP was observed using a blue-light transilluminator at 470 nm to compare the strength of each promoter. T7 promoter showed the brightest fluorescence, followed by prpB promoter. The inhibition zones (highlighted by the dashed lines) against *E. coli* MG1655 BamA6 also confirmed that T7 was the strongest promoter, followed by prpB, T5, and J23101. As a control without *gfp*, pZW-ADC5 was used.

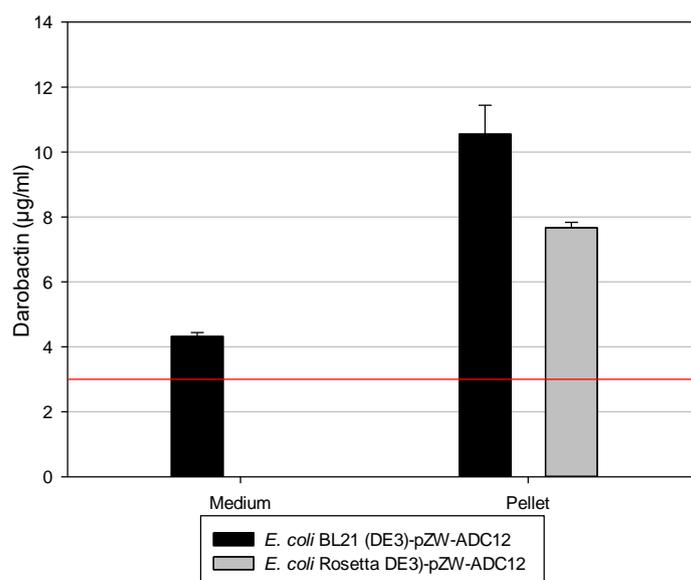


Figure S5: DAR production by codon-optimized minimum DAR BGC in *E. coli* BL21(DE3) and *E. coli* Rosetta™(DE3). *E. coli* BL21(DE3)-pZW-ADC12 produced more DAR than *E. coli* Rosetta™(DE3)-pZW-ADC12. The red line marks the lower linearity border of the calibration curve. Therefore, values below the border are not shown. Data were collected from three biological replicates from one-day culture; error bars show standard deviation.

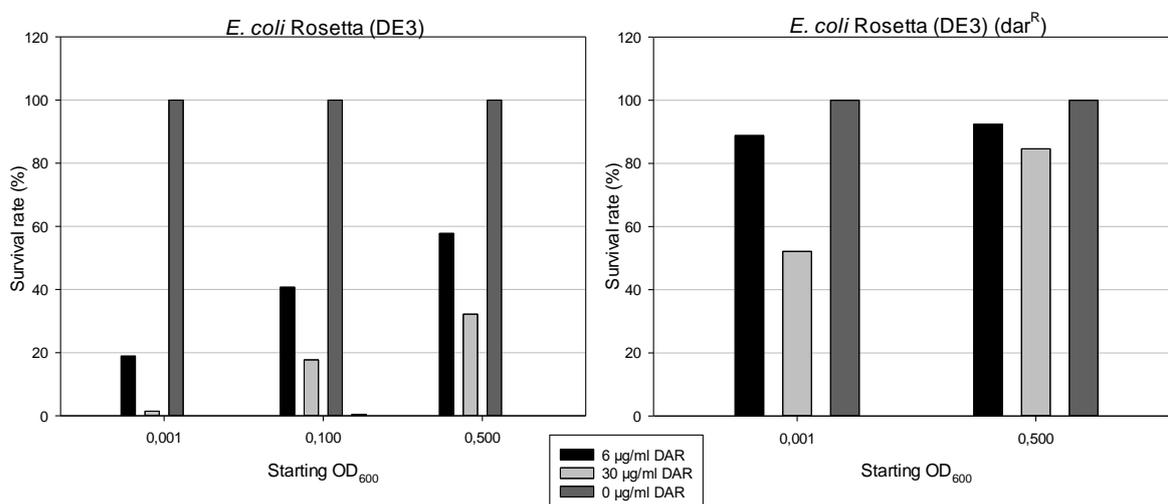


Figure S6: *E. coli* Rosetta™(DE3) and *E. coli* Rosetta™(DE3) (Dar^R) survival rate in the presence of different concentration of DAR. The growth of *E. coli* strains without DAR treatment was considered as 100% survival rate. Survival rate of both strains increased with the increase of the starting OD₆₀₀. *E. coli* Rosetta™(DE3) (Dar^R) had higher survival rate compared to its parent strain.

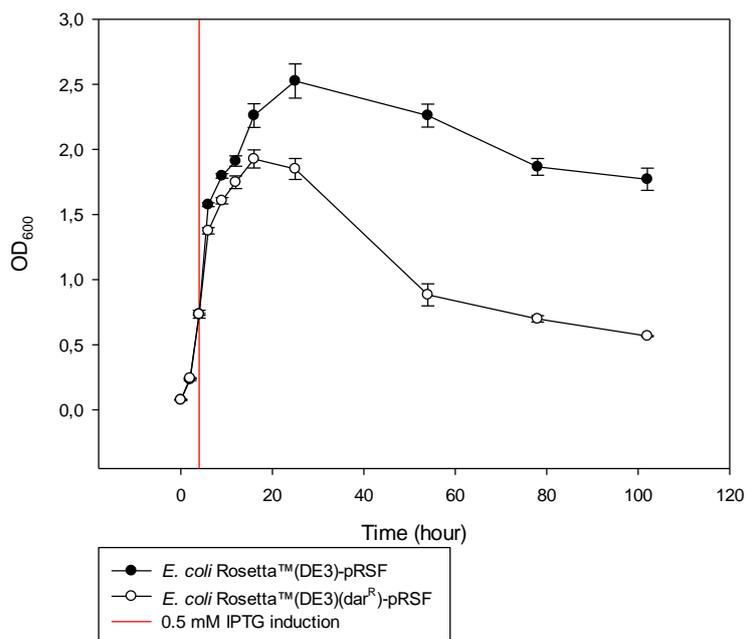


Figure S7: Growth curve of *E. coli* Rosetta™(DE3) and *E. coli* Rosetta™(DE3)(Dar^R) that carries empty expression vector, pRSFDuet™-1. Red line refers to the time point when the cultures were induced with 0.5 mM IPTG induction. Data were collected from three biological replicates; error bars show standard deviation.

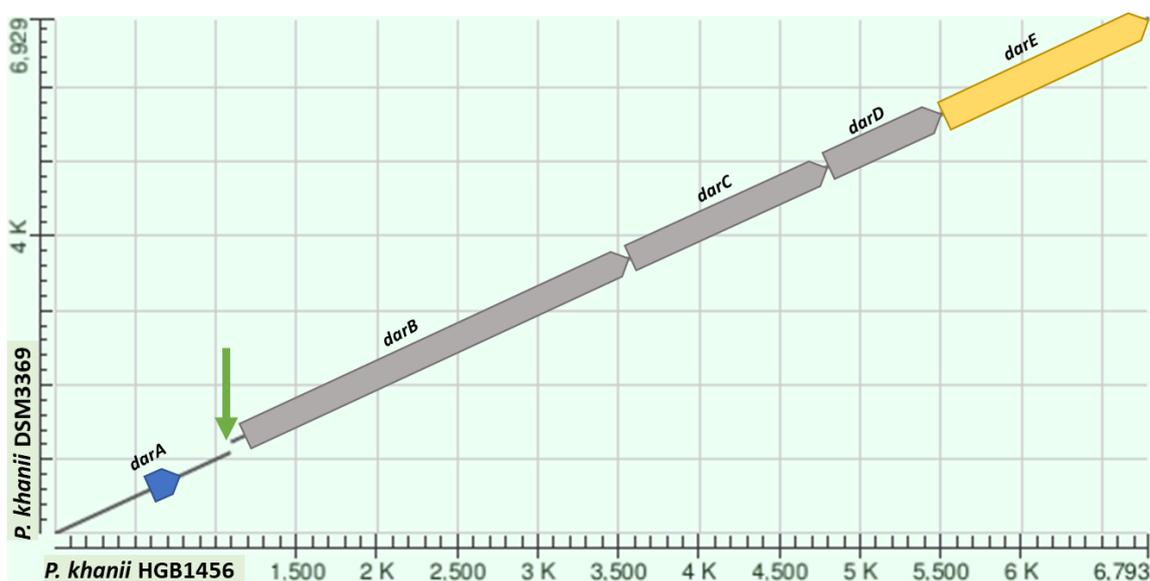


Figure S8: Dot plot depicting alignment of DAR BGC from *P. khanii* HGB1456 and *P. khanii* DSM3369. DAR BGC including 605 bp upstream region and all intergenic regions from *P. khanii* HGB1456 (6793 bp) was aligned to *P. khanii* DSM3369 (6929 bp) using blastn suite-2 sequences program (National Library of Medicine, Bethesda, MD, USA; <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Zhang *et al.*, 2000). On DNA level, they have 99% identity. DAR BGC from *P. khanii* DSM3369 has a 136 bp longer intergenic region between *darA* and *darB*, marked by green arrow.

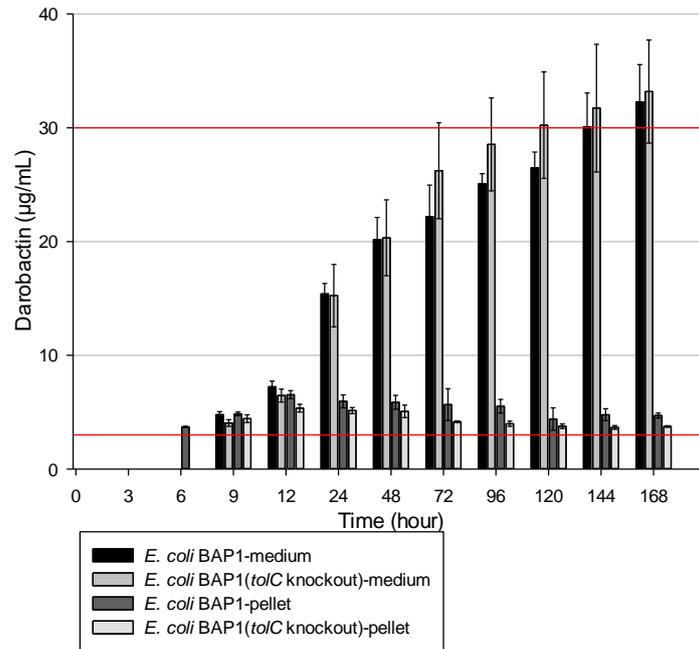


Figure S9: Comparison of DAR production in *E. coli* strains with or without TolC. *E. coli* BAP1 and *E. coli* BAP1 Δ *tolC* were transformed by pZW-ADC5, and DAR production was compared. DAR is produced in the same range by both hosts, with the same intra-/extracellular ratio. The red line indicates the linearity limit of the DAR standard curve. Data were collected from three biological replicates at mentioned time point; error bars show standard deviation.

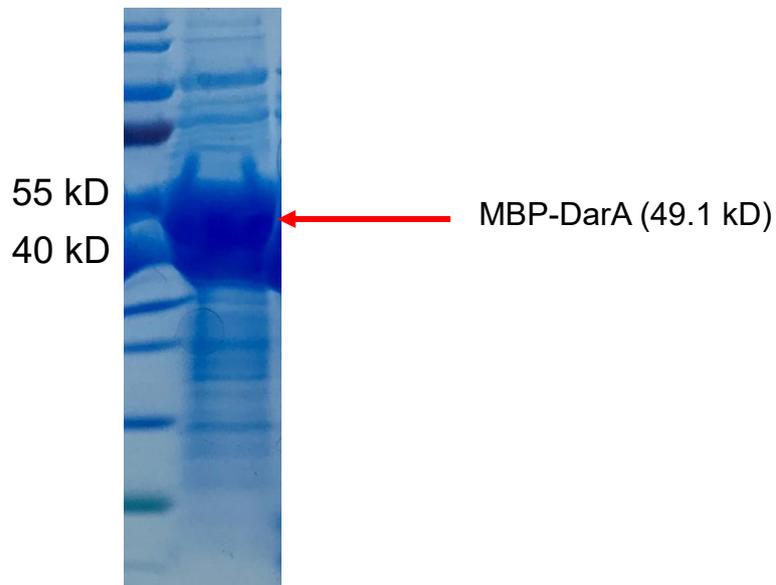


Figure S10: SDS-page analysis of MBP-DarA. The red arrow shows the target protein (MBP-DarA), which fits to the calculated molecular weight of 49.1 kDa.

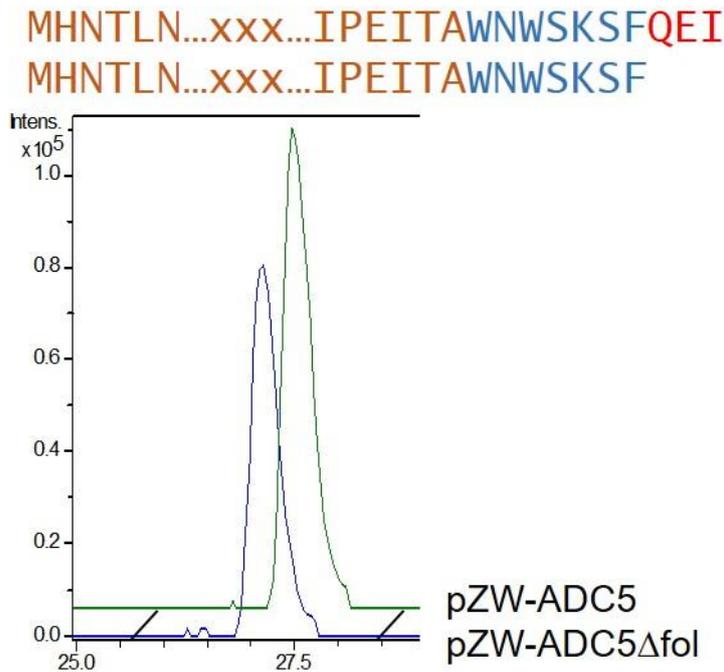


Figure S11: DAR production from DarA precursor with or without follower peptide. In the shown amino acid sequence the core peptide is indicated in blue and the follower (if present) in red. The chromatogram shows similar production for both expression constructs.

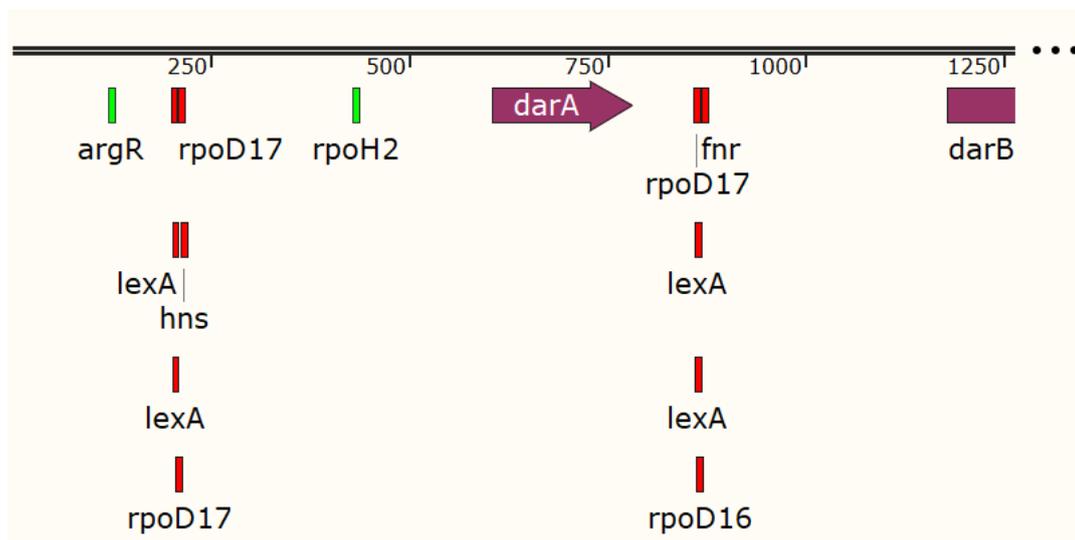


Figure S12: Transcription factor binding site prediction. *In silico* analysis of non-coding region upstream and downstream of *darA* using BPRom (V Solovyev and Salamov, 2011) and CNNprom (Umarov and Solovyev, 2017b). Red indicates transcription factor binding site by both tools, and green indicates prediction by BPRom only.

Sequence S1. Codon optimized DNA sequence of *darA* from *Phototrhobdus namnaonensis*

ATGCACAACACCTCTATCATCAACTGCACCACCCAGGAAGCTCTGAACTCTCTGGCTGCTTCT
TTCAAAGACACCCGAACTGTCTATCACCGAACGTGCTCTGGACGAACTGAACAACAAACCGAAA
ATCCCGGAAATCACCGCTTGGAAGTGGTCTAAATCTTTCCAGGAAATCTAA

Sequence S2. Codon optimized DNA sequence of *darE* from *Phototrhobdus namnaonensis*

ATGGACACCATCATCCCGATCAAATACCTGAACGCTGACGAATCTTCTATCCTGAAAAATCTC
CGAAAATCAACTACCGTCAGCTGGCTTGCCGTATCATCGGTGAAATCCCGGCTGAAAAATCC
TGGACGACGACGAACTGGCTCTGTACAACGAAGAAATCGGTATCCACTTCTCTCCGGAAATCA
TCAACGCTAACAAACTGGTTGTTGTTGTTAAAGCTACCCGTCTGTGCAACCTGCGTTGCACCT
ACTGCCACTCTTGGGCTGAAGGTAAGGTAACACCCTGACCTTCTTCAACCTGATGCGTTCTA
TCCACCGTTTCCTGTCTATCCCGAACATCAAACGTTTCGAATTCGTTTGGCACGGTGGTGAAG
TTACCCTGCTGTCTGTTAACTACTTCAAAAACTGATCTGGCTGCAGGAACAGTTCAAAAAACC
GGACCAGGTTATCACCAACTCTGTTCCAGACCAACGCTGTTAACATCCCGGAAGACTGGCTGG
TTTTCTGAAAGGTATCGGTATGGGTGTTGGTATCTCTGTTGACGGTATCCCGGAAATCCACG
ACTCTCGTCGTCTGGACTACCGTGGTCCGACCTCTCACAAAGTTGCTGCTGGTATGAAAA
AACTGCGTTCTTACGGTATCCCGTACGGTCTCTGGTTGTTGTTGACCGTGACGTTTACGAAT
CTAACATCGAAAAATGCTGTCTTACTTCTACGAAATCGGTCTGACCGACATCGAATTCCTGAA
CATCGTTCCGGACAACCGTTGCCAGCCGGGTGACGACCCGGGTGGTTCTTACATCACCTACC
ACAACATCAACTTCCTGTCTAACGTTTTCCGTGTTTGGTGGAAACGACTACCAGGACAAAAAT
CAACATCCGTCTGTTCCACGGTTTCATCGACTCTATCAAATCTTCTCAGAAAAAATCTCTGAC
TGCTACTGGGCTGGTAACTGCTCTCAGGAAATCATCACCTGGAACCGAACGGTACCGTTTCT
GCTTGCGACAAATACGTTGGTGTGCTGAAGGTAACAACCTACGGTTCTATCATCGACAACGACCTG
GGTACCTGCTGATCAAATCTAACACCAACAAAAACCACCTGAAAGAAGAAATCGAATCTTAC
GAAAAAATGCACCAAGTGCAAATGGTTCCACCTGTGCAACGGTGGTTGCCCGCACGACCGTGT
TACCAACCGTAAACACAACCCGAACTACAACGACTTCTGCTGCGGTACCGGTGGTCTGCTGG
AAATCATCAAACAGACCATCGCTGCTTAA