

Investigation of bacterial secondary metabolite pathways

Dissertation
zur Erlangung des Grades
des Doktors der Naturwissenschaften
der Naturwissenschaftlich-Technischen Fakultät
der Universität des Saarlandes

eingereicht von

Alexander von Tesmar

2017

Tag des Kolloquiums 18.08.2017

Dekan: Prof. Guido Kickelbick

Berichterstatter: Prof. Rolf Müller

Prof. Uli Kazmaier

Vorsitz: Prof. Christian Ducho

akad. Mitarbeiter: Dr. Martin Litzenburger

Erklärung

Hiermit versichere ich an Eides Statt, dass ich die Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Der Anteil weiterer Autoren an den Kapiteln ist jeweils auf der Rückseite des entsprechenden Titelblatts angegeben.

.....

Alexander von Tesmar, Saarbrücken, 23.06.2017

Summary

This work covers studies on three bacterial secondary metabolites, which were investigated with bioinformatic, biochemical and biotechnological methods to obtain insights into the biosynthesis and their general relevance for pharmaceutical research. The biosynthesis of tomaymycin, a pyrrolo[4,2]benzodiazepine (PBD) relevant in anticancer drug development, was elucidated using *in vitro* reconstitution of the underlying enzymes. Furthermore, all biochemical steps were visualized on the intact protein using a novel mass spectrometry based method. Tilivalline, a toxin from *K. oxytoca*, is the causative agent in antibiotic associated hemorrhagic colitis (AAHC). The biosynthesis of the compound was elucidated using *in vitro* reconstitution as well as heterologous expression of the underlying gene cluster. Furthermore, inhibitors were identified that impair tilivalline production in *K. oxytoca* liquid culture and thus a road was paved for an anti-virulence strategy for AAHC. Corramycin, a peptide antibiotic from *Corallococcus coralloides*, is currently under investigation by Sanofi. This work identified the biosynthetic gene cluster and established a rigid biosynthetic model by bioinformatic analysis and labeling studies. The analysis of the corramycin gene cluster led to the identification of two independent resistance mechanisms that were further elucidated in an heterologous system and by *in vitro* reconstitution.

Zusammenfassung

Die vorliegende Arbeit enthält Studien zu drei bakteriellen Sekundärmetaboliten, die mit bioinformatischen, biochemischen und biotechnologischen Methoden untersucht wurden, um die zugrundeliegende Biosynthese und die allgemeine Relevanz für die pharmazeutische Forschung aufzuklären. Die Biosynthese von Tomaymycin, einem für die Entwicklung von Krebsmedikamenten relevantem Pyrrolo[4,2]benzodiazepin (PBD), wurde durch *in vitro* Rekonstitution der entsprechenden Enzyme aufgeklärt. Weiter wurden alle biochemischen Zwischenschritte auf dem intakten Protein durch den Einsatz einer neuen massenspektrometrischen Methode visualisiert. Tilivalline, ein Toxin aus *K. oxytoca*, ist der verantwortliche Stoff in der antibiotika assoziierten hämorrhagischen Colitis (AAHC). Die Biosynthese des Stoffes wurde durch *in vitro* Rekonstitution und dem Einsatz heterologer Expressionssysteme aufgeklärt. Weiter wurde Inhibitoren identifiziert, die die Tilivallinproduktion in Flüssigkultur von *K. oxytoca* beeinträchtigen. Dieses Ergebnis ermöglicht die Entwicklung einer anti-virulenz Strategie für die Behandlung von AAHC. Corramycin, ein Peptidantibiotika produziert von *Corallococcus coralloides*, wird aktuell von Sanofi entwickelt. Die vorliegende Arbeit identifizierte das zugehörige Gencluster und etablierte ein schlüssiges Biosynthesemodell durch bioinformatische Methoden und Labeling-Studien. Die Analyse des Corramycin-Genclusters führte zur Identifikation von zwei unabhängigen Resistenzmechanismen, die durch heterologe Systeme und *in vitro* Rekonstitution weiter aufgeklärt wurden.

Danksagung

Mein besonderer Dank gilt meinem Doktorvater Prof. Rolf Müller für die Aufnahme in die Arbeitsgruppe, die Überlassung der herausfordernden und spannenden Themen sowie die gute Zusammenarbeit. Bei Prof. Uli Kazmaier bedanke ich mich für die stets gute Begleitung und die Übernahme des Zweitgutachtens. Kirsten Harmrolfs danke ich für die vielen Ratschläge und die gute Betreuung meiner Arbeit. Stéphane Renard danke ich für sein stetes Bemühen um meine Interessen und seine ehrlichen Worte.

Ich möchte mich bei der gesamten Arbeitsgruppe für die gute Zusammenarbeit und die vielen großen und kleinen Kooperationen bedanken.

Bedanken möchte ich mich auch bei meiner Familie sowie Ralph und Martina für die fortwährende Unterstützung.

Zutiefst in Dankbarkeit verbunden bin ich den NFM und der VE - pugna continuat!

Mein größter Dank gebührt Dana.

Publications

The following results from this thesis were prepared for publication or have been submitted to scientific journals:

Alexander von Tesmar, Michael Hoffmann, Jan Pippel, Antoine Abou Fayad, Stefan Werner, Armin Bauer, Wulf Blankenfeldt and Rolf Müller: **"Total biosynthesis of tomaymycin comprehensively monitored on the nonribosomal peptide megasynthetase"** submitted to *Cell Chemistry & Biology* in May 2017

Alexander von Tesmar, Michael Hoffmann, Viktoria Schmitt, Antoine Abou Fayad, Jennifer Herrmann and Rolf Müller: **"In vitro reconstitution and heterologous expression of an enterotoxin produced by *Klebsiella oxytoca*"**, in preparation, June 2017

Contents

1	Introduction	9
1.1	Natural Products	10
1.1.1	Relevance of natural products as therapeutics	10
1.1.2	Microbes as source of natural products	12
1.1.3	Myxobacteria as potent producers of natural products	13
1.1.4	Natural products as virulence factors	16
1.1.5	Biosynthesis of natural products	18
1.2	Antiinfectives	23
1.2.1	Antibiotics	23
1.2.2	Antivirulence agents	24
1.3	General Aims and Scope of this Thesis	25
1.4	Bibliography	26
2	Total biosynthesis of tomaymycin comprehensively monitored on the nonribosomal peptide megasynthetase	37
2.1	Introduction	40
2.2	Results and Discussion	41
2.3	Significance	49
2.4	Bibliography	52
2.5	Experimental Procedures	60
2.6	Supporting Information	71
3	In vitro reconstitution and heterologous expression of an enterotoxin produced by <i>Klebsiella oxytoca</i>	79
3.1	Introduction	82
3.2	Results	83
3.3	Discussion	90
3.4	Bibliography	92
3.5	Experimental Procedures	96
3.6	Supporting Information	102
4	Biosynthesis of Corramycin	109
4.1	Introduction	112
4.2	Results	113
4.3	Discussion	123

4.4	Bibliography	125
4.5	Experimental Procedures	131
4.6	Supporting Information	135
5	Corramycin self-resistance mechanism of <i>Corallococcus coralloides</i>	147
5.1	Introduction	150
5.2	Results	150
5.3	Discussion	156
5.4	Bibliography	160
5.5	Experimental Procedures	163
5.6	Supporting Information	167
6	Discussion	171
6.1	Summary	172
6.2	Obstacles in antibiotic development	173
6.3	Microbial sources of compounds in antibiotic drug discovery	174
6.3.1	Myxobacteria as industrial source for antibiotics	174
6.3.2	Activation of silent gene clusters	175
6.3.3	Heterologous expression	177
6.4	Anti-virulence drugs	180
6.4.1	Development of anti-virulence drugs	180
6.4.2	Blocking the NRPS based biosynthesis of virulence factors	183
6.5	Conclusion	185
6.6	Bibliography	186

Chapter 1

Introduction

1.1 Natural Products

The common definition of natural products is broad and describes all products that are made by a living organism, hence includes complex and non pure substances like wood, meat or silk.¹ The scientific definition is more precise and means the compound class of organic substances produced by living organisms and was further divided into two fundamental sections by the physiologist and nobel price winner Albrecht Kossler in 1891.² In his talk "Ueber die chemische Zusammensetzung der Zelle" (the chemical composition of the cell) Kossler proposed the term primary metabolite for substances that can be found in every cell and therefore play an essential role for survival. Substances that are only present in distinct cell types were termed secondary metabolites.³

Since the 19th century our understanding about the "chemical composition of the cell" has grown dramatically, yet the 125 year old terminology of primary and secondary metabolites remains, although the border between the two sections is long known to be blurred. Secondary metabolites, meaning small bioactive molecules from natural sources, are perpetual in the focus of countless drug development efforts.

1.1.1 Relevance of natural products as therapeutics

Today it is estimated that around half of the approved drugs in use are natural products or are direct derivatives. Pure synthetical pharmaceutical molecules are scarce, for though medicinal chemistry provided a large quantity of useful compounds, the starting point for the majority of drug design efforts remained a natural product.⁴

Before the development of synthetic chemistry and the rational design of molecules, humanity had to rely solely on substances that were provided by nature. The oldest known record instructing the use of natural products as medicine was found in Mesopotamia and dates back to 2600 BC. The clay tablets written in cuneiform describe the production and application of plant oils to treat coughs, colds, inflammations or parasite infections. Many of the described plants such as cedar, cypress, poppy seeds and myrr are still in pharmaceutical use today. The famous "Eber Papyrus" from Theben in Egypt dates back to 1500 BC and lists over 700 drugs, their formulation in beer, milk, wine or honey as well as the application itself. The first pharmaceutical documents from China date back to 1100 BC, the foundation of the indian ayurvedic system took place 1000 BC. The first known rational approach to the medicinal use of natural products was conducted in 300 BC by Theophrastus, a greek philosopher and natural scientist. His "History of Plants" is a compendium of medicinal plants, their storage and cultivation techniques. 100 AD the greek physician Dioscoridis traveled alongside the roman army and bequeathed a comprehensive pharmaceutical record of the then known world. After the fall of the west roman empire the knowledge survived in monasteries north of the alpes as well as in the arabic area, where the first known privately owned drug store is documented in 800 AD.⁵ The collective ancient knowledge of medicinal plants set the stage for the scientific approach starting in the 20th century. In 1828 the german pharmacologist Johann Buchner identified salicin as the active compound of willow bark, which analgetic and antipyretic effects were described already in roman times.⁶ Many of our today's first line drugs are of natural origin, including quinine, theophylline, penicillin, morphine, paclitaxel, digoxin, vincristine, doxorubicin, cyclosporine and

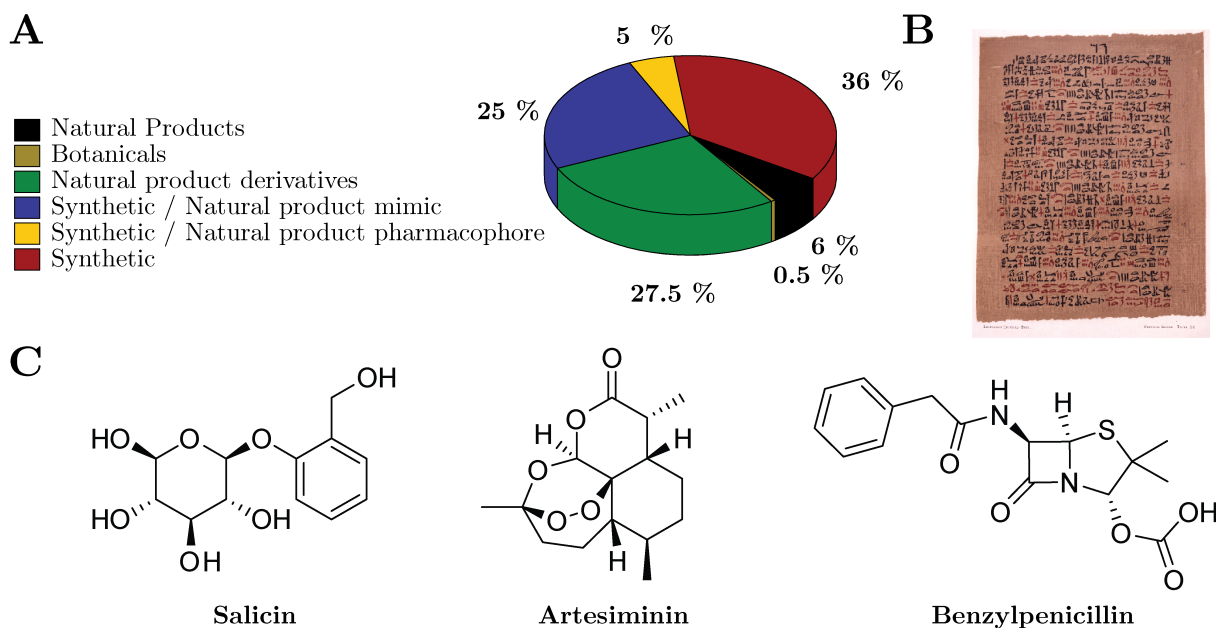


Figure 1.1: **A** Origin of approved drugs in the year 2013.⁹ **B** The "Eber Papyrus" found in Theben, Egypt dates back to 1500 BC and describes over 700 drugs and their formulation (Photography published by the U.S. National Library of Medicine, 22 March 1999) **C** Salicin, artesimin and benzylpenicillin are natural products that represent milestones for modern medicine.

vitamin A.⁷ In 2015 the nobel price for physiology was partly awarded for the discovery and isolation of the anti malaria drug artemisinin from the plant *Artemisia annua*.⁸ The recent report of Newmann and Cragg states that 35% of the approved drugs since 1981 are of pure synthetic origin. 6.5% are unaltered natural products or botanicals, while the remaining 58.5% are either natural product mimics, derivatives or use a natural product pharmacophore (Figure 1.1).⁴

Consequently about two thirds of all approved drugs in the last 35 years are in one or the other way dependent on a natural product. This dependency proved to be vital, since after the introduction of high throughput screening (HTS) and compound libraries based on combinatorial chemistry in the beginning of the 1990s the interest in natural product research dwindled. Natural products seemed to be "library unfriendly" and unsuitable for the new drug discovery pipelines. Additionally, academic and industrial interest in antiinfectiva research declined, a traditional stronghold of natural products.¹⁰ Subsequently the output of new molecules from the R&D departments of the big pharma companies dwindled and the rather fruitless approach of statistically combining of chemical reactions detached from any natural inspiration together with the reemerging need for new antiinfectiva due to the rise of resistant pathogens led to a revived interest in natural products. With the hypothesis that "all natural products have some receptor-binding activity"⁹ the extraction of an organism produces a "natural combinatorial library", which compounds are already preselected by nature.⁴ If a natural product with the desired activity is found, it is rarely applicable directly, as the low percentage of unaltered substances in use demonstrates. Most of the time several iterative cycles of chemical altering the molecule and studies on structure activity relation (SAR), pharmacokinetics, cytotoxicity are necessary to generate an approved drug. Nevertheless, natural products still build up a significant basis for a majority of drug design efforts in industry as well as academia.

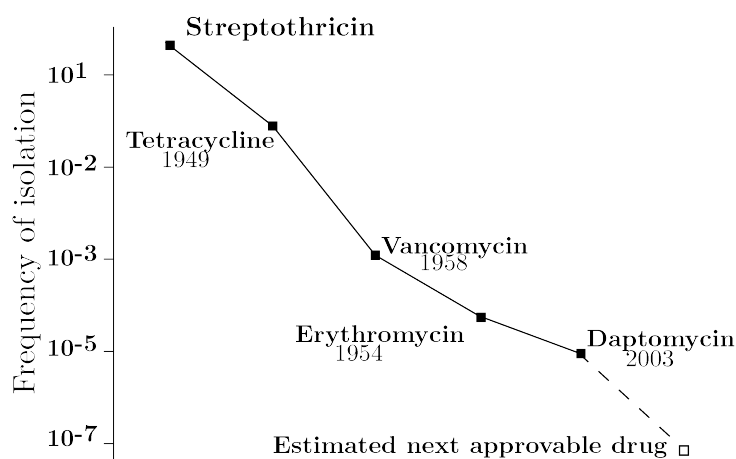


Figure 1.2: Graphical display of the frequency of isolation of approved drugs produced by *streptomyces*. The year of approval is indicated (streptothricin has none).¹³

1.1.2 Microbes as source of natural products

Bacteria and fungi represent one if not the most potent source for novel natural product scaffolds regarding chemical diversity as well as bioactivity. In 2012 it was estimated that from the 70.000 known microbial products about 50% have one or more types of bioactivity, including 16.000 antibiotics. Bacteria are responsible for 40.000 and fungi for 30.000 of this microbial natural products.¹¹ Nevertheless, only a small fraction of the existing microbial natural products is thought to be known. Watve et al. predicted that if one could sample and analyze all streptomycetes on earth one would end up with approximately 150.000 different compounds with distinct biological activity.¹² Given this numbers the potential for the successful isolation of novel and bioactive scaffolds seems endless. Nevertheless, the frequency of discovery of approved antibiotics from streptomyces declined over time. Streptothricin is found in about 20% of all streptomyces species and was therefore easy to find. Tetracycline has a frequency of 0.4%, vancomycin 0.001%, erythromycin 0.00005% and the most recent daptomycin 0.00001%. The estimation of the frequency for the next, not yet identified, approvable drug from streptomyces is 10^{-7} , equivalent to over 10 million cultures that have to be screened (see Figure 1.2).¹³ This small mathematical example highlights two of the major hurdles in identifying bioactive natural products: First, the rediscovery of already known compounds hampers the activity based isolation pipe line and has to be minimized by suitable dereplication methods, e.g. LC-MS databases. Secondly, the frequency of discovery of a compound can be smaller than its actual genetic frequency. The genome of the streptomycin producer *Streptomyces griseus* harbors 22 putative biosynthetic gene clusters, but only few compounds are actually produced in standard culture conditions.¹⁴

Since the introduction of cost efficient sequencing methods over 9000 complete sequencing projects are listed on the GOLD genome online database in the beginning of 2017. Tools like Antismash¹⁵ or Clusterfinder¹⁶ regularly identify more biosynthetic gene clusters in a given organism, bacteria or fungi, than assigned compounds from liquid culture, indicating the presence of silent gene clusters. Appropriately, different culture conditions including co-cultivation with other microbes is a standard procedure to access these hidden compounds. Furthermore, several molecular biology methods are available to directly activate a silent gene cluster, e.g. promoter

insertion, inactivation of repressor genes,¹⁷ or the heterologous expression in an easy to manipulate host strain.^{18, 19} Since the heterologous expression is only dependent on the genome data it enables a grasp on the unculturable strains, that are estimated to make up over 99% of the overall microbial population.²⁰

1.1.3 Myxobacteria as potent producers of natural products

Myxobacteria are Gram-negative δ -proteobacteria exhibiting a complex life cycle and advanced multicellular social behavior, inhabiting almost every habitat, including soils, deep-sea sediments, sweet water and hydrothermal vents. Multicellularity is thought to have occurred at several time-points during evolution and also exists in the prokaryotic domain of life. The division of labor among organized groups of cells leads to an overall higher efficiency and therefore displays an evolutionary advantage. Primitive multicellularity like formation of filaments or clusters is widespread among different unicellular fungi like *Saccharomyces cerevisiae* and different bacteria phyla like cyano- or actinobacteria. The formation of biofilms and swarms by different *bacillus* species already constitutes a more advanced form. Nevertheless, myxobacteria are able to build up patterned multicellular structures that require cell differentiation, spatial morphogenesis, division of labor as well as intercellular communication.²¹

The multicellular life cycle of myxobacteria is determined by the availability of nutrients. In a nutrient depleted environment fruiting body formation takes place. This multilayer and three dimensional biofilm with a diameter between 20-1000 μm consists of an outer layer of dormant, physically pH and temperature resistant cells that protect the inner compartment of spore cells. In nutrient rich environments swarms of myxobacteria move in a rhythmic fashion called "rippling" to reach prey organisms, which are subsequently lysed and digested. Upon complete depletion of the available resources, the life cycle returns to the formation of fruiting bodies.²⁴ The multicellular behavior of myxobacteria is non obligatory and can be lost in conditions where it is not necessary, e.g. liquid culture.²⁵

Myxobacteria lack a flagellum and employ two distinct but synergistic motility systems to move on surfaces with approximately 2 $\mu\text{m}/\text{min}$. A single cell employs A-motility (adventurous) while group of cells use S-motility (social). S-motility relies on the extension and retraction of type IV pili located at the cell poles. By adhesion to other cells or the excreted extracellular matrix of polysaccharides groups of cells are able to move forward. The mechanism of A-motility is not yet fully understood, but two models have been proposed. The "slime gun model" locate the necessary thrust in the hydration process of unilateral excreted polysaccharides at the end of the cell. Cytological observations of the protein AglZ led to the "focal adhesion complex model". It assumes an intra cellular motor complex that is attached to the membrane and the cytoskeleton. By applying tension to the cell framework from the inside a forward movement is achieved.^{24, 26}

The order myxococcales is known as a potent producer of natural products.²⁷ The observed diversity and quantity of natural products as well as the complex life cycle are reflected in the unusual large genome sizes, which average between 9-12 Mbp and thus belong to the largest known in the bacterial domain. The epothilone producing *Sorangium cellulosum* So0157-2 holds a genome of 14.7 Mbp with various duplications of regulatory genes like sigma factors and protein kinases to provide a large flexibility in environment adaption and social behavior.²⁸ The high GC genomes of myxobacteria contain on average between 8-10% of biosynthetic genes in

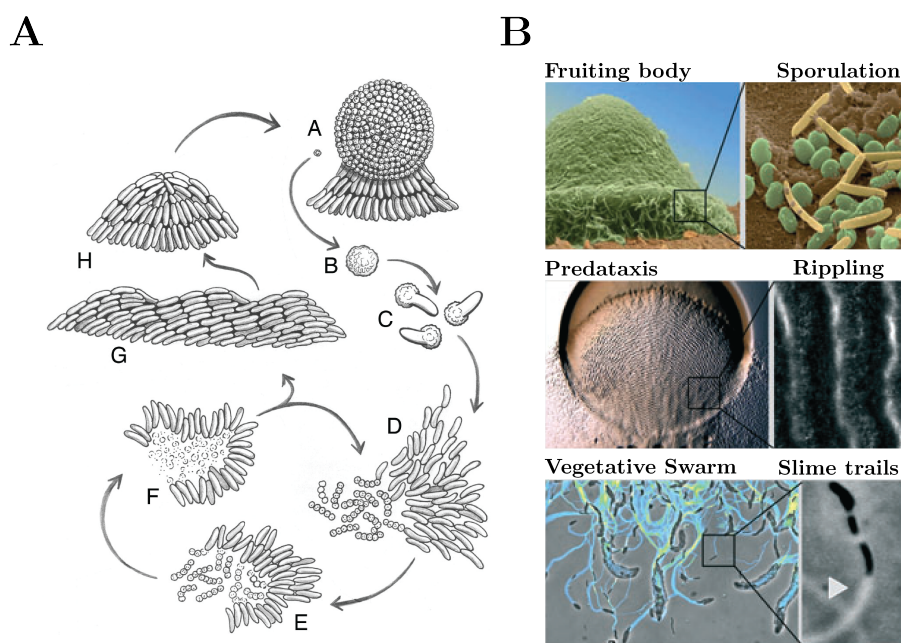
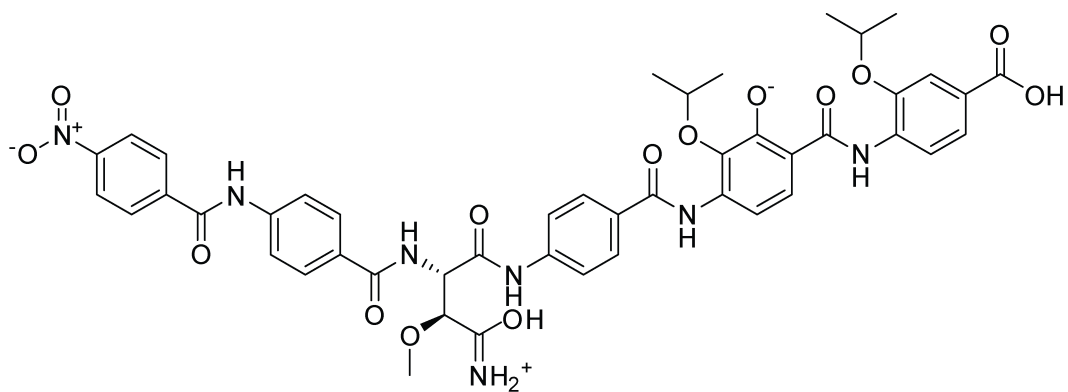


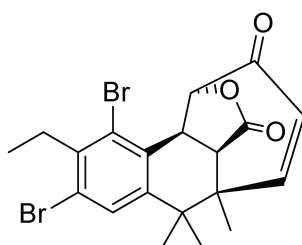
Figure 1.3: **A** Schematic depiction of the myxobacterial life cycle from B. S. Goldman showing a fruiting body (A), spherical myxospores (B), germinating spores (C), rod-shaped vegetative cells while preying (D), aggregation of vegetative cells (E), aggregated cells during prey degradation (F), accordion waves (G) and aggregation of starving cells prior to fruiting body formation (H).²² **B** Artificial colored electron microscope image of an *M. xanthus* fruiting body containing spores and light microscopy images of preying and swarming cells.²³

10-20 different clusters and therefore exhibit a greater potential for natural products than actinomycetes.^{22, 29, 30}

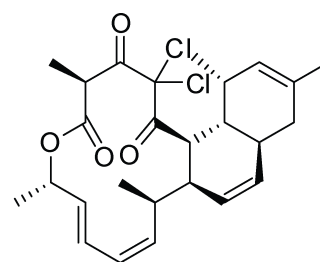
The first study of secondary metabolites from myxobacteria was published in 1947. Oxford reported the secretion of an unknown antibacterial compound from *Myxococcus virescens* into the liquid media, although he was not able to elucidate the structure due to instability.³¹ In the beginning of 2017 109 distinct core structures of myxobacterial origin with over 1000 derivatives thereof have been reported. The panel of bioactivity is large and includes antifungal, antibacterial, antiviral, immunosuppressive, antimalarial and antioxidative properties. The elucidation of the mode of action of several compound classes indicates that the cellular targets are as diverse as the bioactivity.³² Most of the myxobacterial natural products are either non ribosomal peptides, polyketides or hybrids thereof and exhibit many uncommon structural features. The linear polyketide ajudazol contains an unusual isochromanone ring and binds to complex I of the electron transfer chain.³³ Salinabromid isolated from *Enhygromyxa salina*, a weak antibacterial compound, harbors a bromonated benzene ring inside a novel tetracyclic scaffold.^{Felder.2013} The novel class of cystobactamids described in 2014 were isolated from *Cystobacter* sp.. Next to their unique NRPS-assembled para-amino-benzoic acid chain a strong antibacterial activity against gram-negative strains was observed. The minimal inhibitory concentration in the low $\mu\text{g/mL}$ range is derived from a strong inhibition of the type IIa topoisomerase. Therefore cystobactamids are a strong prospect in the search for alternatives to the quinolones, whereof multiple derivatives were rendered useless by mutations.³⁴ Another potent class of myxobacterial antibiotics are the sorangicins. The cyclic polyketide backbone contains unusual pyran moieties and a bicyclic ether group. The inhibition of eubacterial RNA polymerase gives rise



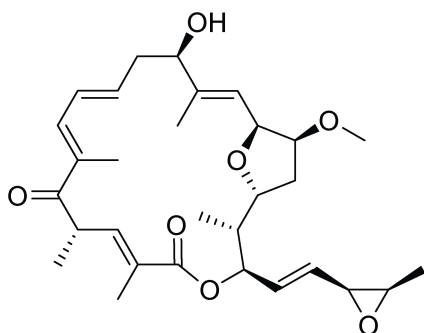
Cystobactamide 919-2



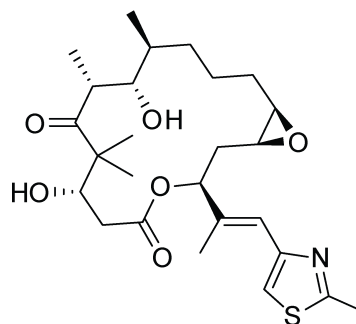
Salinabromid



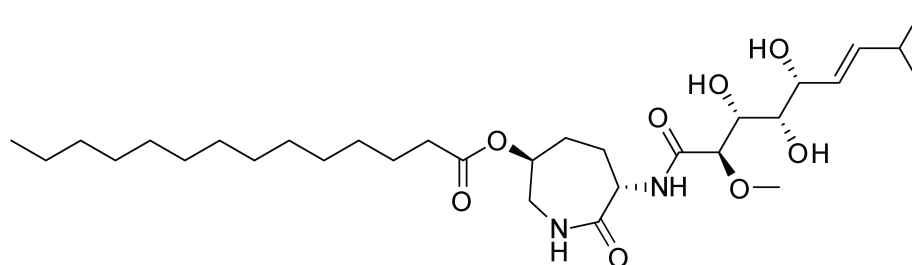
Chlorotonil A



Maltepolide A



Epothilone A



Bengamid A

Figure 1.4: Structures of the myxobacterial natural products cystobactamid 919-2, salinabromid, chlorotonil A, maltepolide A, epothilone A and bengamid A.

to the strong antibacterial activity against both, gram-negative and -positive bacteria.³⁵ The cyclic polyketide chlorotolone was extracted from a liquid culture of *Sorangium cellulosum* So ce 1525. It contains a unique gem-dichloro-1,3-dione moiety and exhibits a strong activity against the parasitic protozoa *plasmodia* that causes malaria.^{36,37}

Several myxobacterial compounds exhibit activity on eukaryotic cells. The compound family of maltepolides, as well isolated from a *Sorangium cellulosum* strain, harbors a tetrahydrofuran moiety in the macrolactone scaffold. Although exhibiting no strong antibacterial activity, a significant morphological change in eukaryotic PtK₂ cells could be observed upon incubation. The bengamids were first isolated from a marine sponge and contain a δ -hydroxylysine cycle. Inhibition of the methionine aminopeptidase is responsible for a strong anticancer activity.^{38, 39} The most prominent representative of cytotoxic and myxobacterial compounds altogether is epothilone. The PKS derived macrolide showed antifungal as well as β -tubulin binding properties.^{40, 41} Binding and stabilization of tubulin leads to cytotoxic effects on proliferating cells and is the mode of action of established drugs like taxol in anticancer therapy. Epothilone showed cytotoxic activity against several cancer cell lines.⁴² Although binding to the same binding pocket, epothilone showed to be active against taxol resistant cancer cell lines. To improve the pharmacokinetic profile as well as the water solubility, several semisynthetic derivatives of epothilone were synthesized. In 2007 the derivative ixabepilone was approved as Ixempra[®] by the Food and Drug Administration as treatment against taxol resistant breast cancer.⁴³

1.1.4 Natural products as virulence factors

Microbial natural products, exploited by humans in the quest for novel pharmaceuticals since thousands of years, are originally made by its producer to gain some kind of evolutionary advantage in the never ending competition for resources in its ecological niche. One of the richest habitats for microorganisms is the human body, where bacteria cells outnumber the human ones in the ratio of 1.3 to 1. Most of the microorganism are either commensal or mutual, which means that their presence is either neutral or even beneficial for the host. However, obligate pathogens cause diseases and produce natural products that target the host tissue. The interaction of small molecules produced by pathogens with the host is fundamental for the understanding of a disease and the subsequent development of therapeutics.^{44–46}

In the mid 16th century the idea that microbes cause diseases occurred for the first time. In the late 17th century Antony van Leewenhoek was the first person to see microbes upon inspecting his own dental plaque with a microscope. The advancement in the understanding of diseases provided by the work of scientist like Louis Pasteur or Robert Koch in the second half of the 19th century were fundamental for the modern medicine of today.⁴⁷ However, as our understanding of the genomic basis of bacteria and their social mechanisms grows, the importance of natural products with special function involved in both, healthy and dis-balanced microflora, becomes evident.⁴⁸ The Human Microbiome Project identified over 3000 biosynthetic gene clusters for small molecules in the genomes of human associated bacteria whereby only a small amount was widely distributed.⁴⁹ Some of the identified compounds are connected to various diseases (1.5).⁵⁰ *Staphylococcus aureus* is a common bacteria that lives on the skin and mucous membranes of humans. Although harmless for healthy individuals it can cause severe infections if they penetrate a broken skin barrier. *S. aureus* is known to produce the carotenoid staphyloxanthin that inter-

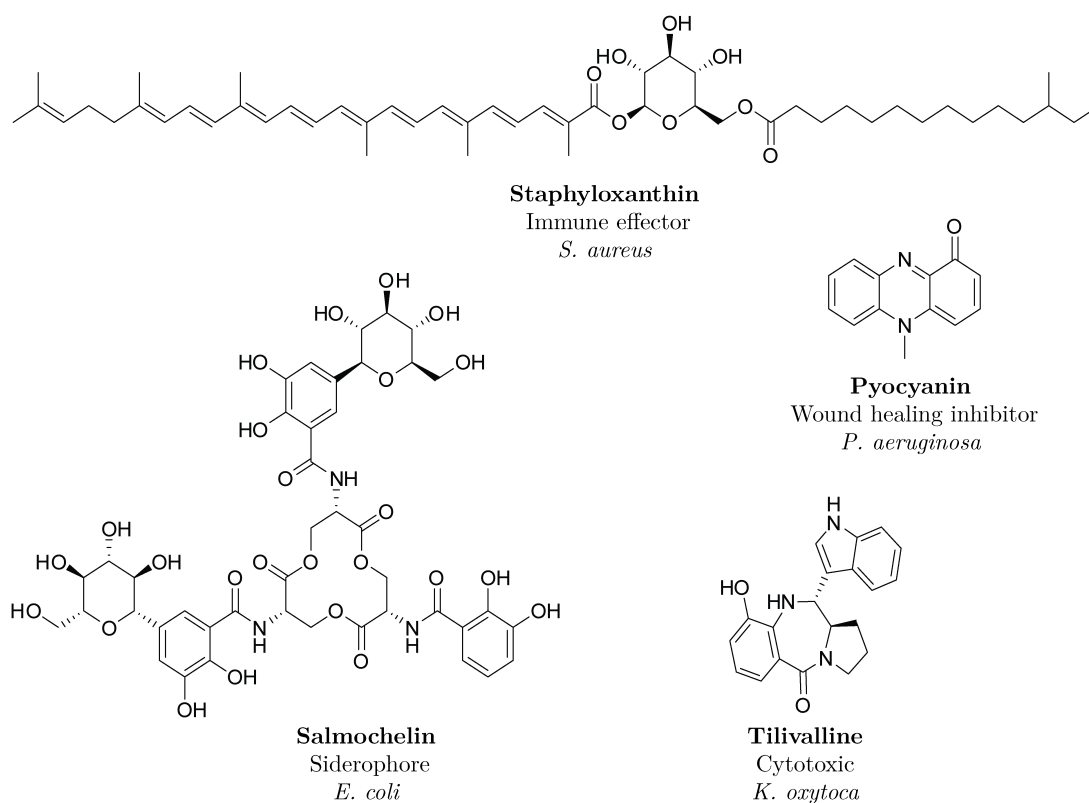


Figure 1.5: Structure of virulence factors produced by human pathogens and their role in infection.

feres with the immune response of the host and thus helps its producer to circumvent its natural defense system.⁵¹ The whopping cough causing *Bordetella pertussis* produces the disaccharide tetrapeptide tracheal cytotoxin, a integral part of the observed virulence.^{52–57} The relevant nosocomial pathogen *Pseudomonas aeruginosa* can infect various parts of the human body. Several natural products including homoserine lactones, quinolones, phenazines, rhamnolipids and siderophores have been identified. During skin infection the phenazine pyocyanine produced by *P. aeruginosa* impairs the wound healing process and therefore prolongs the infection. Several compounds produced during respiratory tract infections are known to modulate the gene expression of the host cells, produce free radicals or scavenging iron from epithelial cells.^{58, 59}

The human habitat with the densest microbial population is the gastro intestinal tract (GI tract). The healthy microflora includes about 1000 different species and can be disturbed by different factors, including stress, infection or antibiotic treatment.^{60, 61}

The iron limiting conditions in the intestinal tract induce the productions of various siderophores. The pathogens *Klebsiella pneumoniae*, *Escherichia coli* and *Salmonella typhimurium* all secrete enterobactin to scavenge iron. To circumvent the inactivation by the host secreted protein siderocalin, that irreversible binds enterobactin, pathogens glycosilate the siderophore to generate salmochelin.^{62, 63} However, this system is not only used to harm the host cells. To gain an advantage against other microbes relying on the enterobactin system, *E. coli* produces the class of microcins (RiPP) that act as antibiotic on related species. The microcins are attached to the C-terminal end of enterobactin and subsequently actively imported by other species in a "trojan horse" fashion.^{64–66}

Recent insights into the secondary metabolome of gut bacteria revealed the less obvious interaction between microbial natural products and mental health. The connection between the

GI-tract and the central nervous system (CNS) is called the gut-brain axis.⁶⁷ The amino acid metabolism of gut bacteria includes neuro transmitter like serotonin,⁶⁸ γ -aminobutyric acid⁶⁹ or β -phenylethylamin.⁷⁰ Although still a controversial field, autism spectrum disorder (ASD) related diseases have been shown to be often accompanied by chronic constipation, diarrhea or abdominal pain.^{71–74} The serum level of the microbial produced 4-ethylphenylsulfat was increased in ASD model mice. Subsequently, the ASD related behavior could be induced by injection of the compound into healthy control mice.⁷⁵

The identification of microbiome-derived natural products and the characterization of the respective biological function in healthy individuals as well as in diseases is an important step towards understanding the complex habitat in the human body and the development of respective therapeutics.

1.1.5 Biosynthesis of natural products

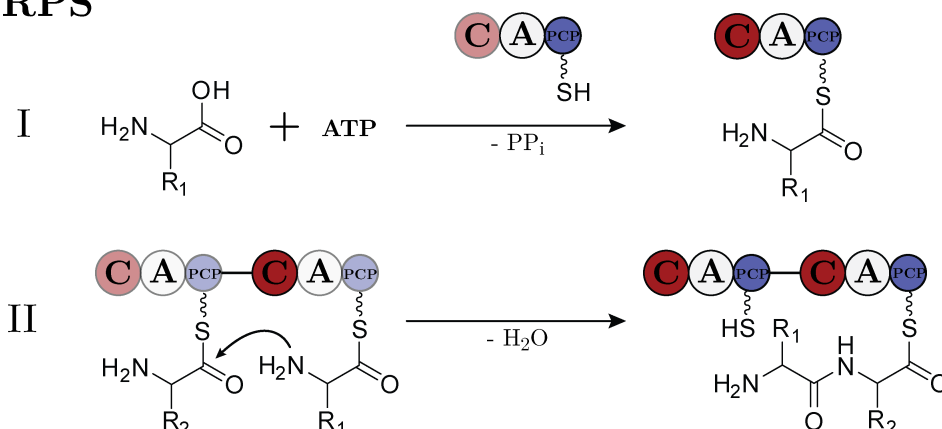
The majority of the pharmaceutical relevant bacterial produced secondary metabolites is either a nonribosomal peptide, a polyketide or a hybrid thereof. The one thing that these natural products have in common is their assembly by a large, multimodular megaenzyme from small monomers in a conveyor belt like fashion: The non-ribosomal peptide synthetase (NRPS) builds up peptides from single amino acid building blocks. The polyketide synthase is closely related to the fatty acid synthases from the primary metabolism and builds up polyketides from small acyl monomers. Commonly several tailoring enzymes are encoded in the surrounding of the NRPS or PKS genes, that introduce modifications into the product chain that are often vital for bioactivity. Various modifications are known, including glycosylation, hydroxylation, methylation or oxydative crosslinking by members of the cytochrome P450 superfamily. The tailoring enzymes can exist as independent protein or be included as a distinctive domain in the assembly line itself.⁷⁶ Outside the realm of NRPS and PKS other biosynthetical sources for natural products exist, such as the ribosome or distinct enzymes like cyclodipeptide synthases (CDPS).

NRPS

NRPS *de novo* assemble peptide chains with a length of up to 22 monomers. For this biosynthetic process neither the ribosome nor any kind of RNA is involved to direct the substrates. Instead, a multi-modular protein assembly line activates and couples single amino acid monomers to subsequently release the final peptide, whereby each module is responsible for one building block. While the ribosome is restricted to the 20 proteinogenic amino acids or rather the respective tRNAs, that are loaded by the aminoacyl-tRNA synthetases, the NRPS can activate non-proteinogenic amino acids.^{77–80} Up to date over 200 different building blocks including unusual α -amino acids but also β - and γ -derivatives have been described.⁸¹

The assembly process on an NRPS can be divided into several phases. First, the ribosomal produced and inactive apo-enzyme has to be modified with a coenzyme A (CoA) derived phosphopanthine (PPant) arm on specific serine sidechains in the peptide carrier protein domain (PCP) in each module. These modifications act as the arms of the machinery and covalently bind the growing substrate chain as thioester. The next phase includes substrate recognition, activation and loading onto the assembly line. Each NRPS module has a 50 kDa adenylation domain (A) upstream of the PCP-domain. The A-domain selects its respective substrate and

NRPS



PKS

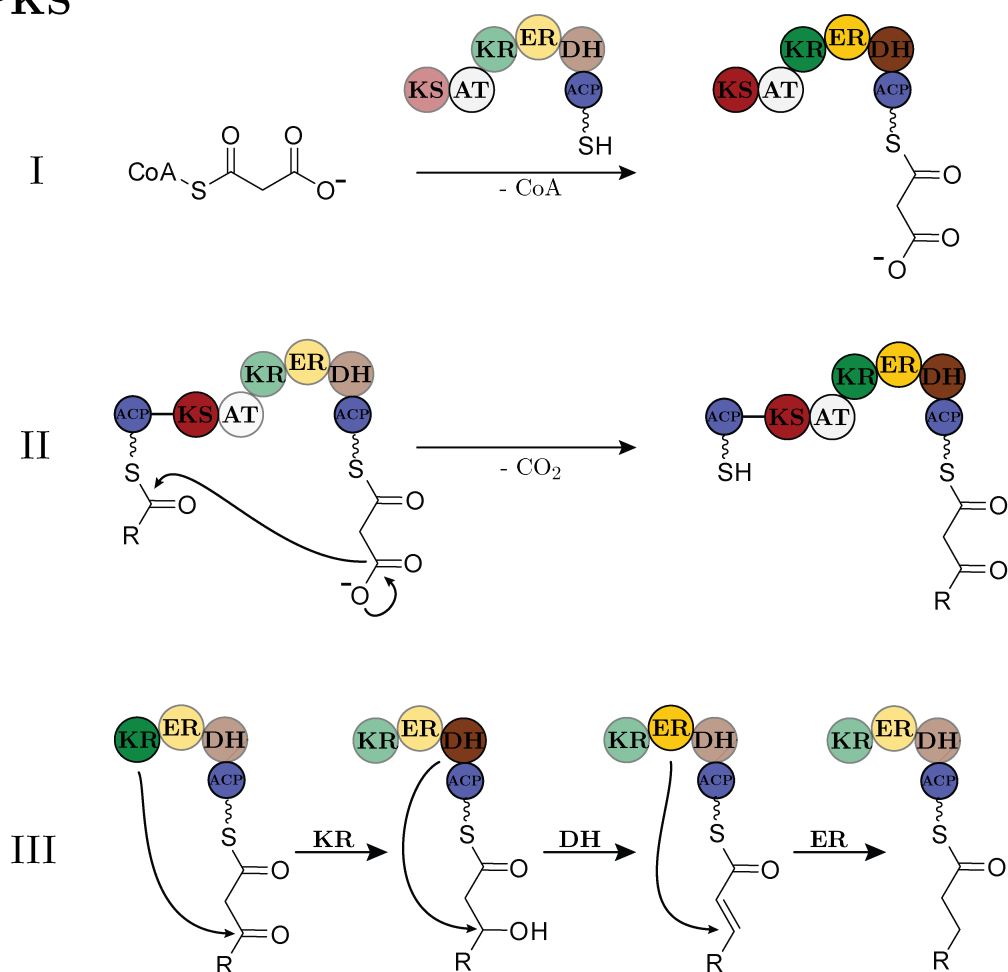


Figure 1.6: Comparison of the chain elongation methodology of NRPS and PKS assembly lines: **NRPS** The monomeric amino acid is selected and activated by the adenylation domain (A) and subsequently tethered to the peptide carrier protein (PCP) as thioester (I). The condensation domain catalyzes the nucleophilic attack of the terminal aminogroup of the downstream bound monomer and forms the peptide bond (II). **PKS** The already activated precursor is loaded on the acyl carrier protein (ACP) by the acyltransferase domain (AT) (I). The ketosynthase domain (KS) catalyzes the nucleophilic attack of the downstream monomer to generate the polyketide intermediate (II). The optional action of three domains, the ketoreductase (KR), the dehydratase (DH) and the enoylreductase (ER) lead to four possible functional groups in the final molecule (III).

generates an activated aminoacyl adenylate that is subsequently loaded onto the PPant arm of the PCP-domain. The loaded substrates are now connected by the condensation domain (C), which is present in all chain elongation modules. This 50 kDa domain catalyzes the peptide forming step between the amino group of the downstream tethered substrate with the thioester carbonyl of the upstream one. The last module harbors a domain responsible for the release of the grown substrate chain. Usually this domain is a thioesterase, a member of the serine protease superfamily, that transfers the peptide chain from the thiol group of the PPant arm to the hydroxy group of a serine sidechain. This bond is subsequently either hydrolysed by water to yield a free carboxylic acid like in the vancomycin biosynthesis⁸² or attacked by an internal nucleophile. The macrolactam ring of tyrocidine is generated by nucleophilic attack of the amino group of the first monomer phenylalanine.^{83, 84} A side chain hydroxy group as nucleophile yields a macrolacton ring, like in telomycin.⁸⁵ Other strategies including different domains are also known. A variant of a C-domain facilitates the macrolactamization in the cyclosporin⁸⁶ biosynthesis. Crocacin is as well released by a C-domain, although as free carboxylic acid.⁸⁷ Another chemical option is the release by a reductive domain. The peptide chain can be released as aldehyde, that may undergo intramolecular cyclisation to form the imine like in tomaymycin.⁸⁸ The consecutive reduction of the terminal carboxylic acid group by a reduction domain to the respective alcohol has also been described.⁸⁹ The different release domains play a vital role in the generation of chemical diversity and bioactivity e.g. by providing a rigid three dimensional structure to a otherwise flexible peptide chain through cyclisation.⁹⁰

Another important entity for modifications is the first module of the assembly line. Lipopeptides are generated by a C-domain in the first module that can use fatty acid acyl-CoAs or loaded acyl carrier proteins (ACP) from the fatty acid biosynthesis as a donor to generate an N-acylated peptide chain. The class of polymyxins⁹¹ or friulimicin^{92, 93} are both lipopeptides generated by an NRPS with a C-domain as first module.

PKS

A large part of bacterial natural products are in part or as a whole PKS produced polyketides. Although the formation of a C-C bond instead of an peptide bond is catalyzed by PKS, the underlying assembly line principle resembles the one of NRPS: One building block per module is tethered to an PPant arm bound to the acyl carrier protein domain (ACP) and the chain growth is subsequently facilitated by the nucleophilic attack of the downstream substrate. The chain release variations in PKS systems also resemble the NRPS ones. Consequently both types of assembly lines, NRPS and PKS, can form hybrids to produce peptide-polyketide natural products.

PKS assembly lines derive from the fatty acid synthases (FAS) found in the primary metabolism. Therefore malonyl-CoA is the key substrate for PKS, in contrast to the substrate diversity observed in NRPS systems. Malonyl-CoA is formed by the carboxylation of acetyl-CoA in an ATP dependent manner by the acetyl-CoA carboxylase. This activated building block is tethered to the ACP domain by an acyltransferase domain (AT). The downstream C₂ malonyl carbon acts as nucleophile to attack the upstream acyl group to form a C-C bond in a claisen condensation that is driven by decarboxylation and catalyzed by the ketosynthase domain (KS). The resulting

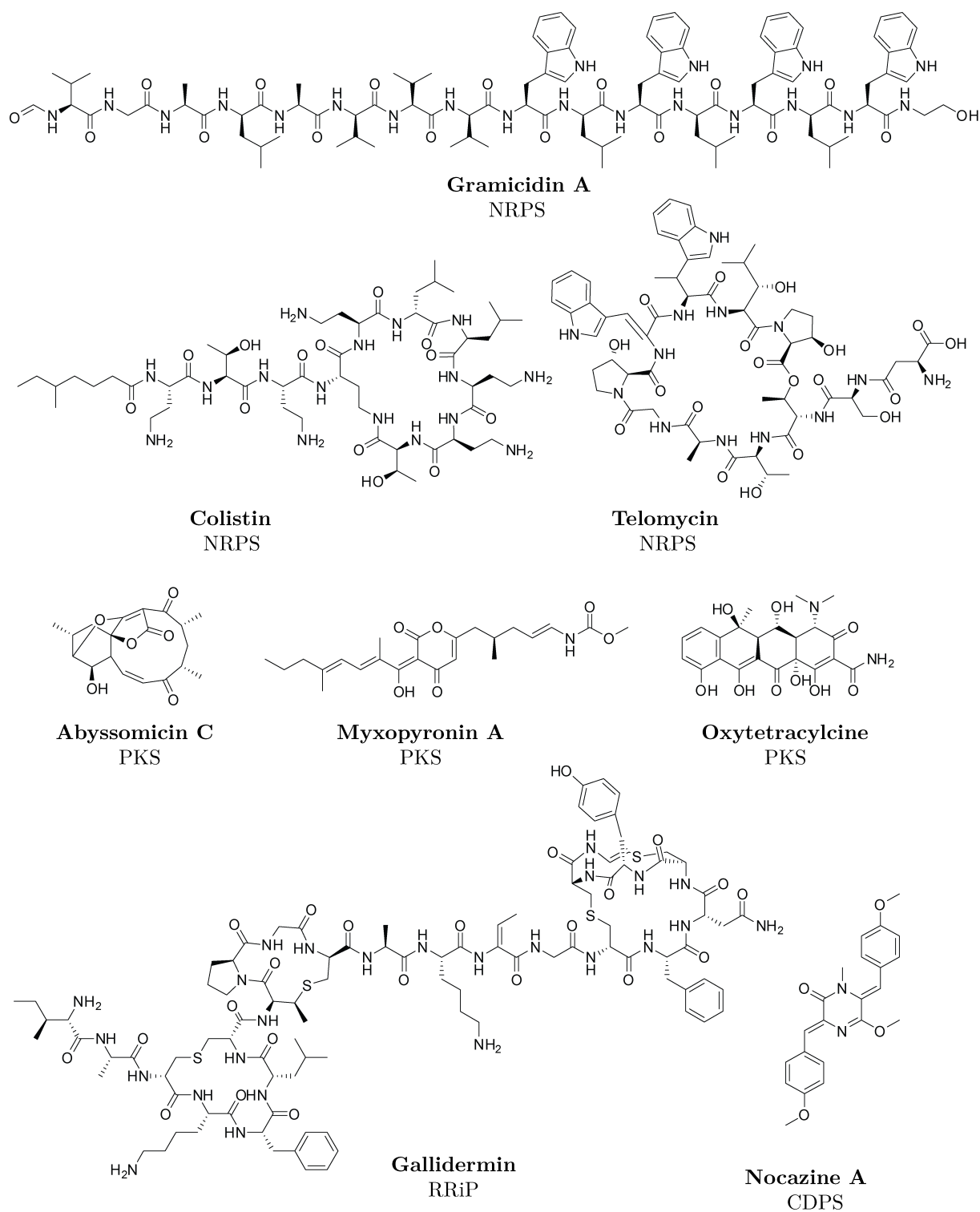


Figure 1.7: Structure of different secondary metabolites and their biosynthetic origin.

initial β -ketone gives rise to the term polyketide. In FAS three enzymes act subsequently on the β -ketone to catalyze a 4-electron reduction to the final carbon chains made from CH_2 groups. First, the keto reductase (KR) reduces the ketone NADP dependent to the β -alcohol. Water dissociation catalyzed by the dehydratase (DH) yields an α - β -enoyl that is reduced by the enoyl reductase (ER) under FADH_2 consumption to the final CH_2 - CH_2 group. In this way the surplus of energy resources in form of reduction equivalents and acetyl-CoA building blocks is stored as fatty acid chains by the cell. PKS systems use the same set of basic catalytic domains to form a specific product, whereby their proteins can be arranged in two distinct ways. Type I PKS are made of large proteins harboring one module for each elongation step similar to NRPS systems. Type II PKS resemble the iterative work flow of the FAS and each catalytic function is only present once and harbored by a single protein.^{76, 77, 94–96}

In contrast to the fatty acids produced by FAS, PKS systems generate chemical diversity by the incomplete processing of the keto-intermediates. By partially omitting parts of the three processing steps of the β -ketone four different functional groups can be introduced in the final product chain. If no processing takes place, the β -ketone remains intact, action of only the KR domain produces an alcohol. The presence of an KR and DH domain introduces a double bond. If all three domains, KR, DH and ER are present, a fatty acid like CH_2 group is incorporated. Additional to this intrinsic possibility several other mechanisms are known to introduce chemical diversity into polyketides. The large class of macrolactone polyketides (macrolides) is based on intramolecular cyclisation after chain release and includes several antibiotics like erythromycin.⁹⁷ Precursor directed derivatisation by unusual starter units⁹⁸ or the incorporation of alternative building blocks like methylmalonyl-CoA or hexyl-malonyl-CoA⁹⁹ are also known.

Other biosynthetic sources for natural products

Multi modular enzymes are not the only source for microbial secondary metabolites and several other biosynthetic pathways outside of the realm of NRPS and PKS systems are known: Terpenes or plant metabolites like flavonoides are for example produced by enzymes distinct from NRPS or PKS systems. But also the ribosome, responsible for the protein synthesis in the cell, is a source for small bioactive peptides. To overcome the restriction to the 20 proteinogenic amino acids and their ultimately DNA-templated linear assembly, intensive post-translational modification takes place. These class of natural products is termed ribosomally synthesized and post-translationally modified peptides (RiPPs). The post-translational modification allows for a similar degree of chemical diversity like NRPS derived peptides and consequently RiPPs with diverse bioactive properties are known. RiPPs are synthesized on the ribosome harboring an N-terminal leader peptide and the C-terminal core region, where the final peptide sequence is located. The leader peptide is recognized by the enzyme machinery introducing the post-translational modifications on the core peptide, before it is cleaved off to release the final product.^{100, 101} Several classes of RiPPs are known including lasso peptides,¹⁰² cyanobactins¹⁰³ and lanthipeptides.¹⁰⁴ RiPPs as well as NRPS derived peptides are synthesized in a templated fashion, since either the genetic code or the organisation of the respective enzyme determines the structure of the natural product. Furthermore non-templated biosynthetic pathways exist that also generate bioactive peptides independent from the ribosome or NRPS.¹⁰⁵ Cyclodipeptide synthases (CDPS) for example detour loaded tRNA from the primary metabolism to assemble diketopiperazines.¹⁰⁶

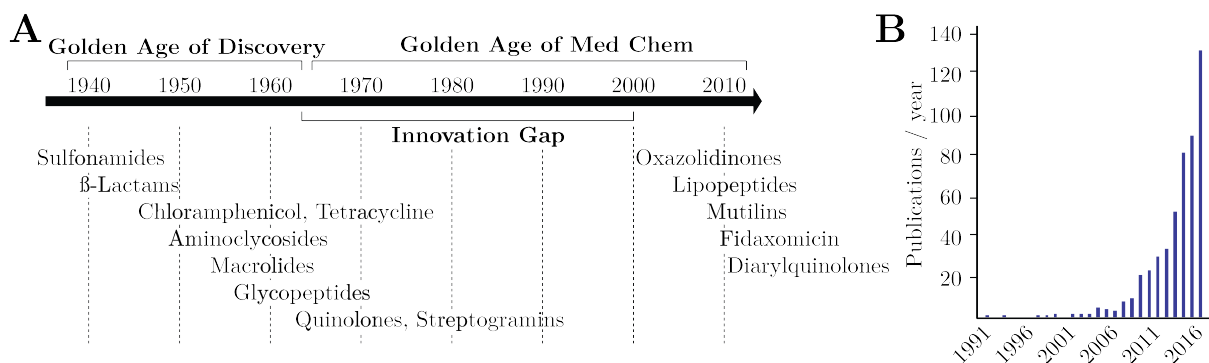


Figure 1.8: **A** Timeline of antibiotic discovery since the introduction of the sulfonamides in the 1930s. The lack of research efforts led to the innovation gap between 1960 and 2000 in which the medicinal chemistry took the lead role in the generation of novel antibiotic derivatives. The introduction of oxazolidinones in 2000 started the second period of antibiotic discovery which continues upon today. Adapted from Walsh and Wencewicz.¹⁰⁷ **B** Plot of publications per year that included the term anti-virulence, antivirulence or phatoblocker listed in pubmed visualizes the increased research in virulence factors and their inhibition. Data was generated using the webserver Medline trend by Alexandru Dan Corlan (<http://dan.corlan.net/medline-trend.html>).

1.2 Antiinfectives

1.2.1 Antibiotics

Every year 14.9 million people are killed by an infectious disease. Tuberculosis accounts for 1.5 millions alone, diarrheal diseases for 1.8 million.¹⁰⁸ The enforcing of high hygienic standards or the introduction of vaccines have lowered the chance of an infection in the first place, nevertheless acute infectious diseases have to be treated with antibiotics. Antibiotics are molecules with a low molecular weight that specific target bacterial life and either kill it (bacteriocidal) or prevent further growth (bacteriostatic). The abundant resource of antibiotic compounds in the realm of natural products shows that nature has always used this compound class. Since the work of Louis Pasteur and Robert Koch that identified the microbes as the cause for infectious diseases, humanity has searched for antibiotic substances.^{47, 109–111} In 1909 the Robert Koch student Paul Ehrlich discovered salvarsan as treatment against syphilis.¹¹² 20 years later Alexander Fleming observed the antibacterial properties of an penicillium mold on the famous agar plates and thus discovering the first antibiotic.¹¹³ Broad use of penicillin however did not take place until the 1940s. The first widespread used antibiotic class were the sulfonamides,¹¹⁴ introduced in the early 1930s and initiating the so called "golden age of antibiotic discovery" (Figure 1.8). In the upcoming 30 years several new classes of antibiotics were discovered that improved the treatment of infectious diseases significantly. Since the 1960s the development of novel antibiotics was discontinued due to the general impression that infectious diseases have been defeated once and for all. This assumption proved to be wrong, since many antibiotic classes were rendered useless by the development of resistance and the emergence of novel pathogens. The delayed resumption of serious research and development efforts on the field of antibiotics led to the introduction of linezolid class in 2000, almost 40 years after the preceding class of quinolones in 1962. The year long lack of novel antibiotic classes and the widespread misuse in human and life stock treatment has assisted the development of resistances leading to the recent concern of the world health organization that the era of antibiotics is coming to an end. Especially the

pathogens of the ESKAPE group, namely *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacter species*, have acquired resistance against several if not all reserve antibiotics available. The insight that resistance is only a question of time and all antibiotics therefore have a limited life-span induces a constant need for novel therapeutics to successfully counter infectious diseases.^{107, 115, 116}

1.2.2 Antivirulence agents

The continuing threat to public health by antibiotic resistant pathogens has led to the development of alternative strategies like nanoparticles, anti-virulence agents, immuno- and phage therapy. Anti-virulence agents have an intrinsic advantage over antibiotics. Since only the virulence is disrupted and the growth and viability of the pathogen is not affected, the selective pressure is kept low and therefore the probability of resistance development is narrowed. Furthermore, the beneficial microbiota that is severely damaged by an antibiotic therapy, is not affected and serious side effects like antibiotic-associated colitis are prevented.¹¹⁷ The good theoretical premise has led to an yearly increasing number of publications in the area of anti-virulence research from a single one in 1991 to 131 in 2016 (Figure 1.8).

Obviously the ESKAPE group of pathogens is in the focus of research efforts in anti-virulence agents. One major problem of several nosocomial pathogens is the rapid formation of biofilms that provides resistance to antibiotics even to actual sensitive strains. Several mechanisms involved in biofilm formation are known and can be targeted by small molecules. Sugar like molecules can bind to the lectins of *P. aeruginosa* and thus inhibiting the interconnection between the single cells that make up the backbone of the biofilm. The sugars were able to inhibit biofilm formation in a mouse lung, although critical side effects by binding of host lectins were observed.¹¹⁸ A different approach targets the signaling pathways that are critical for biofilm formation. The cyclic-di-GMP pathway, vital for biofilm formation in several pathogens can be inhibited with ebselen, although no *in vivo* data is available.¹¹⁹ Potent inhibitors with MIC values of 200-300 nM like 3-acyltetronic acid or benzamide-benzimidazole derivatives can break up the antibiotic resistant biofilm of *S. aureus* or *P. aeruginosa* in mice.^{120, 121} Inhibiting the biosynthesis of a toxic virulence factor has also been reported. The production of the anti-immune compound staphyloxanthin in *S. aureus* can be inhibited by phosphonoacetamide derivatives *in vivo*.¹²² Another target is the well-conserved T3SS apparatus in *P. aeruginosa* and other pathogens that is responsible for the delivery of toxins into the host cell.^{123, 124}

Summarized, anti-virulence agents have the potential to constitute an additional weapon against bacterial infections. Especially the combination with antibiotics, for example in chronic infections with extensive biofilm formation, seems promising. Nevertheless, only few trials with anti-virulence agents have been made and no approved drug is on the market. Several T3SS inhibitors have been trialled, but failed.^{125, 126} Lectin inhibitors were successful in small scale studies with cystic fibrosis patients, but no further data is available.¹²⁷ Several reasons are responsible for this scenario. First, with the word "anti-virulence" first appearing in 1991 the field of anti-virulence research is very young, especially compared to the century long antibiotic research. Furthermore, the assessment of an anti-virulence drug is more pretentious than for antibiotics. Pathogen cell numbers do not decline and secondary effects in the host like a resulting septic shock have to be considered.¹²⁸ Finally, no big industry partners are currently involved in a way that would pro-

vide a required framework for developing an anti-virulence agent to a approved drug.¹¹⁷ If these hurdles are overcome, anti-virulence agents might provide a sorely needed secondary pipeline in the ad infinitum ongoing fight against bacterial pathogens.

1.3 General Aims and Scope of this Thesis

The general aim of this thesis was the elucidation of the biosynthesis of three different bacterial natural products: tomaymycin, tilivalline and corramycin. To achieve the intended goals, methods from molecular biology, microbiology, enzymology, bioanalytics and biotechnology were applied. Within this interdisciplinary approach, heterologous expression of biosynthetic gene clusters and the *in vitro* reconstitution of proteins were the crucial methods used to address the central questions of this thesis.

Tomaymycin is a long known pyrrolo[4,2]benzodiazepine used as a scaffold in anti-cancer drug development. Although the tomaymycin gene cluster has been identified previously,¹²⁹ a rigid elucidation of the biosynthesis using *in vitro* reconstitution of participating enzymes has not been pursued. Furthermore, this study aimed to apply recent advantages in mass spectrometry of intact proteins to NRPS systems, visualizing all biochemical processes during biosynthesis on the megaenzyme itself.

Tilivalline, the causative toxin in *K. oxytoca* associated AAHC, is structurally related to tomaymycin and was also linked to its gene cluster previously.¹³⁰ Central aim of this thesis was the elucidation of the mechanism responsible for the incorporation of the indole moiety. Furthermore, since the indole substitution is contraindicative to the mode-of-action known for tomaymycin, this study also aspired to investigate the mechanism of cytotoxicity of the compound class. Insights into toxin biosynthesis and its mode-of-action are relevant for pharmaceutical research and thus this study also aimed to exploit gained knowledge for the development of novel treatment strategies, e. g. anti-virulence drugs.

Corramycin, an antibiotic produced by *Corallococcus coralloides*, was previously isolated by Sanofi. This study aimed to correlate this natural product to its biosynthetic gene cluster and furthermore to identify possible self-resistance mechanisms to aid the ongoing efforts by Sanofi to elucidate the to-date unknown mode-of-action of the corramycin antibacterial activity. Insights into the biosynthesis of natural products often establishes access to novel derivatives and thus this study also aimed to identify or generate corramycin derivatives.

1.4 Bibliography

- [1] Editorial. All natural. *Nature chemical biology*, 3(7):351, 2007.
- [2] Kara Rogers. *The components of life: From nucleic acids to carbohydrates*. Biochemistry, cells, and life. Britannica Educational Pub. in association with Rosen Educational Services, New York, NY, 1st ed. edition, 2011.
- [3] Albrecht Kossel. Ueber die chemische zusammensetzung der zelle (the chemical composition of the cell). *Archiv fuer Anatomie und Physiologie*, pages 181–186, 1891.
- [4] David J. Newman and Gordon M. Cragg. Natural products as sources of new drugs from 1981 to 2014. *Journal of natural products*, 79(3):629–661, 2016.
- [5] Gordon M. Cragg and David J. Newman. Drugs from nature: past achievements, future prospects. *Ethnomedicine and Drug Discovery*, pages 23–37, 2002.
- [6] Olivier Lafont. Du saule a l’aspirine. *Revue d’histoire de la pharmacie*, 55(354):209–216, 2007.
- [7] A. M. Clark. Natural products as a resource for new drugs. *Pharmaceutical research*, 13(8):1133–1144, 1996.
- [8] The Nobel Assembly at Karolinska Institutet. 2015 nobel prize in physiology or medicine, 2015.
- [9] Gordon M. Cragg and David J. Newman. Natural products: a continuing source of novel drug leads. *Biochimica et biophysica acta*, 1830(6):3670–3695, 2013.
- [10] Frank E. Koehn and Guy T. Carter. The evolving role of natural products in drug discovery. *Nature reviews. Drug discovery*, 4(3):206–220, 2005.
- [11] Janos Berdy. Thoughts and facts about antibiotics: Where we are now and where we are heading. *The Journal of antibiotics*, 65(8):441, 2012.
- [12] M. G. Watve, R. Tickoo, M. M. Jog, and B. D. Bhole. How many antibiotics are produced by the genus streptomyces? *Archives of microbiology*, 176(5):386–390, 2001.
- [13] Richard Baltz. Antimicrobials from actinomycetes: Back to the future. *Microbe*, 2(3):125–131, 2007.
- [14] Yasuo Ohnishi, Jun Ishikawa, Hirofumi Hara, Hirokazu Suzuki, Miwa Ikenoya, Haruo Ikeda, Atsushi Yamashita, Masahira Hattori, and Sueharu Horinouchi. Genome sequence of the streptomycin-producing microorganism streptomyces griseus ifo 13350. *Journal of bacteriology*, 190(11):4050–4060, 2008.
- [15] Tilmann Weber, Kai Blin, Srikanth Duddela, Daniel Krug, Hyun Uk Kim, Robert Brucoleri, Sang Yup Lee, Michael A. Fischbach, Rolf Muller, Wolfgang Wohlleben, Rainer Breitling, Eriko Takano, and Marnix H. Medema. antismash 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic acids research*, 43(W1):W237–43, 2015.

-
- [16] Peter Cimermancic, Marnix H. Medema, Jan Claesen, Kenji Kurita, Laura C. Wieland Brown, Konstantinos Mavrommatis, Amrita Pati, Paul A. Godfrey, Michael Koehrsen, Jon Clardy, Bruce W. Birren, Eriko Takano, Andrej Sali, Roger G. Linington, and Michael A. Fischbach. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell*, 158(2):412–421, 2014.
- [17] Gang Liu, Keith F. Chater, Govind Chandra, Guoqing Niu, and Huarong Tan. Molecular regulation of antibiotic biosynthesis in streptomyces. *Microbiology and molecular biology reviews : MMBR*, 77(1):112–143, 2013.
- [18] Yunzi Luo, Behnam Enghiad, and Huimin Zhao. New tools for reconstruction and heterologous expression of natural product biosynthetic gene clusters. *Natural product reports*, 33(2):174–182, 2016.
- [19] Axel A. Brakhage and Volker Schroeckh. Fungal secondary metabolites - strategies to activate silent gene clusters. *Fungal genetics and biology : FG & B*, 48(1):15–22, 2011.
- [20] Brajesh K. Singh and Catriona A. Macdonald. Drug discovery from uncultivable microorganisms. *Drug discovery today*, 15(17-18):792–799, 2010.
- [21] Jose Munoz-Dorado, Francisco J. Marcos-Torres, Elena Garcia-Bravo, Aurelio Moraleda-Munoz, and Juana Perez. Myxobacteria: Moving, killing, feeding, and surviving together. *Frontiers in microbiology*, 7:781, 2016.
- [22] B. S. Goldman, W. C. Nierman, D. Kaiser, S. C. Slater, A. S. Durkin, J. A. Eisen, C. M. Ronning, W. B. Barbazuk, M. Blanchard, C. Field, C. Halling, G. Hinkle, O. Iartchuk, H. S. Kim, C. Mackenzie, R. Madupu, N. Miller, A. Shvartsbeyn, S. A. Sullivan, M. Vaudin, R. Wiegand, and H. B. Kaplan. Evolution of sensory complexity recorded in a myxobacterial genome. *Proceedings of the National Academy of Sciences of the United States of America*, 103(41):15200–15205, 2006.
- [23] Yong Zhang, Adrien Ducret, Joshua Shaevitz, and Tam Mignot. From individual cell motility to collective behaviors: insights from a prokaryote, myxococcus xanthus. *FEMS microbiology reviews*, 36(1):149–164, 2012.
- [24] Emilia M. F. Mauriello, Tam Mignot, Zhaomin Yang, and David R. Zusman. Gliding motility revisited: how do the myxobacteria move without flagella? *Microbiology and molecular biology reviews : MMBR*, 74(2):229–249, 2010.
- [25] G. J. Velicer, L. Kroos, and R. E. Lenski. Loss of social behaviors by myxococcus xanthus during evolution in an unstructured habitat. *Proceedings of the National Academy of Sciences of the United States of America*, 95(21):12376–12380, 1998.
- [26] David R. Zusman, Ansley E. Scott, Zhaomin Yang, and John R. Kirby. Chemosensory pathways, motility and development in myxococcus xanthus. *Nature reviews. Microbiology*, 5(11):862–872, 2007.
- [27] H. Reichenbach. Myxobacteria, producers of novel bioactive substances. *Journal of industrial microbiology & biotechnology*, 27(3):149–156, 2001.

- [28] Kui Han, Zhi-feng Li, Ran Peng, Li-ping Zhu, Tao Zhou, Lu-guang Wang, Shu-guang Li, Xiao-bo Zhang, Wei Hu, Zhi-hong Wu, Nan Qin, and Yue-zhong Li. Extraordinary expansion of a sorangium cellulosum genome from an alkaline milieu. *Scientific reports*, 3:2101, 2013.
- [29] Haruo Ikeda, Jun Ishikawa, Akiharu Hanamoto, Mayumi Shinose, Hisashi Kikuchi, Tadayoshi Shiba, Yoshiyuki Sakaki, Masahira Hattori, and Satoshi Omura. Complete genome sequence and comparative analysis of the industrial microorganism streptomyces avermitilis. *Nature biotechnology*, 21(5):526–531, 2003.
- [30] Markiyani Oliynyk, Markiyani Samborskyi, John B. Lester, Tatiana Mironenko, Nataliya Scott, Shilo Dickens, Stephen F. Haydock, and Peter F. Leadlay. Complete genome sequence of the erythromycin-producing bacterium saccharopolyspora erythraea nrrl23338. *Nature biotechnology*, 25(4):447–453, 2007.
- [31] A. E. Oxford. Observations concerning the growth and metabolic activities of myxococci in a simple protein-free liquid medium. *Journal of bacteriology*, 53(2):129–138, 1947.
- [32] J. Herrmann, A. Abou Fayad, and R. Muller. Natural products from myxobacteria: novel metabolites and bioactivities. *Natural product reports*, 2016.
- [33] Brigitte Kunze, Rolf Jansen, Gerhard Hofle, and Hans Reichenbach. Ajudazols, new inhibitors of the mitochondrial electron transport from chondromyces crocatus. production, antimicrobial activity and mechanism of action. *The Journal of antibiotics*, 57(2):151–155, 2004.
- [34] Sascha Baumann, Jennifer Herrmann, Ritesh Raju, Heinrich Steinmetz, Kathrin I. Mohr, Stephan Huttel, Kirsten Harmrolfs, Marc Stadler, and Rolf Muller. Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *Angewandte Chemie (International ed. in English)*, 53(52):14605–14609, 2014.
- [35] Herbert Irschik, Maren Kopp, Kira J. Weissman, Kathrin Buntin, Jorn Piel, and Rolf Muller. Analysis of the sorangicin gene cluster reinforces the utility of a combined phylogenetic/retrobiosynthetic analysis for deciphering natural product assembly by trans-acting pkcs. *Chembiochem : a European journal of chemical biology*, 11(13):1840–1849, 2010.
- [36] Katrin Jungmann, Rolf Jansen, Klaus Gerth, Volker Huch, Daniel Krug, William Fenical, and Rolf Muller. Two of a kind—the biosynthetic pathways of chlorotoniol and anthracimycin. *ACS chemical biology*, 10(11):2480–2490, 2015.
- [37] Herbert Irschik, Peter Washausen, Florenz Sasse, Jorg Fohrer, Volker Huch, Rolf Muller, and Evgeny V. Prusov. Isolation, structure elucidation, and biological activity of maltopolides: remarkable macrolides from myxobacteria. *Angewandte Chemie (International ed. in English)*, 52(20):5402–5405, 2013.
- [38] Patrick McDougal and Joseph Rico Don VanDerveer. Bengamides, heterocyclic anthelmintics from a jaspidae marine sponge. *Journal of Organic Chemistry*, 51:4494–4497, 1986.

-
- [39] Silke C. Wenzel, Holger Hoffmann, Jidong Zhang, Laurent Debussche, Sabine Haag-Richter, Michael Kurz, Frederico Nardi, Peer Lukat, Irene Kochems, Heiko Tietgen, Dietmar Schummer, Jean-Paul Nicolas, Loreley Calvet, Valerie Czepczor, Patricia Vrignaud, Agnes Muhlenweg, Stefan Pelzer, Rolf Muller, and Mark Bronstrup. Production of the bengamide class of marine natural products in myxobacteria: Biosynthesis and structure-activity relationships. *Angewandte Chemie (International ed. in English)*, 54(51):15560–15564, 2015.
- [40] D. M. Bollag, P. A. McQueney, J. Zhu, O. Hensens, L. Koupal, J. Liesch, M. Goetz, E. Lazarides, and C. M. Woods. Epothilones, a new class of microtubule-stabilizing agents with a taxol-like mechanism of action. *Cancer research*, 55(11):2325–2333, 1995.
- [41] K. Gerth, N. Bedorf, G. Hofle, H. Irschik, and H. Reichenbach. Epothilons a and b: antifungal and cytotoxic compounds from *sorangium cellulosum* (myxobacteria). production, physico-chemical and biological properties. *The Journal of antibiotics*, 49(6):560–563, 1996.
- [42] P. Fumoleau, B. Coudert, N. Isambert, and E. Ferrant. Novel tubulin-targeting agents: anticancer activity and pharmacologic profile of epothilones and related analogues. *Annals of oncology : official journal of the European Society for Medical Oncology*, 18 Suppl 5:v9–15, 2007.
- [43] John T. Hunt. Discovery of ixabepilone. *Molecular cancer therapeutics*, 8(2):275–281, 2009.
- [44] Junjie Qin, Ruiqiang Li, Jeroen Raes, Manimozhiyan Arumugam, Kristoffer Solvsten Burgdorf, Chaysavanh Manichanh, Trine Nielsen, Nicolas Pons, Florence Levenez, Takuji Yamada, Daniel R. Mende, Junhua Li, Junming Xu, Shaochuan Li, Dongfang Li, Jianjun Cao, Bo Wang, Huiqing Liang, Huisong Zheng, Yinlong Xie, Julien Tap, Patricia Lepage, Marcelo Bertalan, Jean-Michel Batto, Torben Hansen, Denis Le Paslier, Allan Linneberg, H. Bjorn Nielsen, Eric Pelletier, Pierre Renault, Thomas Sicheritz-Ponten, Keith Turner, Hongmei Zhu, Chang Yu, Shengting Li, Min Jian, Yan Zhou, Yingrui Li, Xiuqing Zhang, Songgang Li, Nan Qin, Huanming Yang, Jian Wang, Soren Brunak, Joel Dore, Francisco Guarner, Karsten Kristiansen, Oluf Pedersen, Julian Parkhill, Jean Weissenbach, Peer Bork, S. Dusko Ehrlich, and Jun Wang. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285):59–65, 2010.
- [45] Ron Sender, Shai Fuchs, and Ron Milo. Revised estimates for the number of human and bacteria cells in the body. *PLoS biology*, 14(8):e1002533, 2016.
- [46] Fredrik Backhed, Josefine Roswall, Yangqing Peng, Qiang Feng, Huijue Jia, Petia Kovatcheva-Datchary, Yin Li, Yan Xia, Hailiang Xie, Huanzi Zhong, Muhammad Tanweer Khan, Jianfeng Zhang, Junhua Li, Liang Xiao, Jumana Al-Aama, Dongya Zhang, Ying Shiuan Lee, Dorota Kotowska, Camilla Colding, Valentina Tremaroli, Ye Yin, Stefan Bergman, Xun Xu, Lise Madsen, Karsten Kristiansen, Jovanna Dahlgren, and Jun Wang. Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell host & microbe*, 17(5):690–703, 2015.

- [47] Michael T. Madigan. *Brock Mikrobiologie*. Pearson Deutschland, s.l., 13., aktualisierte aufl. edition, 2013.
- [48] Anne E. Clatworthy, Emily Pierson, and Deborah T. Hung. Targeting virulence: a new paradigm for antimicrobial therapy. *Nature chemical biology*, 3(9):541–548, 2007.
- [49] Mohamed S. Donia, Peter Cimermancic, Christopher J. Schulze, Laura C. Wieland Brown, John Martin, Makedonka Mitreva, Jon Clardy, Roger G. Linington, and Michael A. Fischbach. A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell*, 158(6):1402–1414, 2014.
- [50] Neha Garg, Tal Luzzatto-Knaan, Alexey V. Melnik, Andres Mauricio Caraballo-Rodriguez, Dimitrios J. Floros, Daniel Petras, Rachel Gregor, Pieter C. Dorrestein, and Vanessa V. Phelan. Natural products as mediators of disease. *Natural product reports*, 2016.
- [51] George Y. Liu, Anthony Essex, John T. Buchanan, Vivekanand Datta, Hal M. Hoffman, John F. Bastian, Joshua Fierer, and Victor Nizet. Staphylococcus aureus golden pigment impairs neutrophil killing and promotes virulence through its antioxidant activity. *The Journal of experimental medicine*, 202(2):209–215, 2005.
- [52] W. E. Goldman, D. G. Klapper, and J. B. Baseman. Detection, isolation, and analysis of a released bordetella pertussis product toxic to cultured tracheal cells. *Infection and immunity*, 36(2):782–794, 1982.
- [53] B. T. Cookson, H. L. Cho, L. A. Herwaldt, and W. E. Goldman. Biological activities and chemical composition of purified tracheal cytotoxin of bordetella pertussis. *Infection and immunity*, 57(7):2223–2229, 1989.
- [54] B. T. Cookson, A. N. Tyler, and W. E. Goldman. Primary structure of the peptidoglycan-derived tracheal cytotoxin of bordetella pertussis. *Biochemistry*, 28(4):1744–1749, 1989.
- [55] K. E. Luker, J. L. Collier, E. W. Kolodziej, G. R. Marshall, and W. E. Goldman. Bordetella pertussis tracheal cytotoxin and other muramyl peptides: distinct structure-activity relationships for respiratory epithelial cytopathology. *Proceedings of the National Academy of Sciences of the United States of America*, 90(6):2365–2369, 1993.
- [56] K. E. Luker, A. N. Tyler, G. R. Marshall, and W. E. Goldman. Tracheal cytotoxin structural requirements for respiratory epithelial damage in pertussis. *Molecular microbiology*, 16(4):733–743, 1995.
- [57] Chung-I Chang, Yogarany Chelliah, Dominika Borek, Dominique Mengin-Lecreulx, and Johann Deisenhofer. Structure of tracheal cytotoxin in complex with a heterodimeric pattern-recognition receptor. *Science (New York, N.Y.)*, 311(5768):1761–1764, 2006.
- [58] Alexandria A. Reinhart and Amanda G. Oglesby-Sherrouse. Regulation of pseudomonas aeruginosa virulence by distinct iron sources. *Genes*, 7(12), 2016.
- [59] Yi-Chia Liu, Kok-Gan Chan, and Chien-Yi Chang. Modulation of host biology by pseudomonas aeruginosa quorum sensing signal molecules: Messengers or traitors. *Frontiers in microbiology*, 6:1226, 2015.

-
- [60] Andrew B. Shreiner, John Y. Kao, and Vincent B. Young. The gut microbiome in health and in disease. *Current opinion in gastroenterology*, 31(1):69–75, 2015.
- [61] Catherine A. Lozupone, Jesse I. Stombaugh, Jeffrey I. Gordon, Janet K. Jansson, and Rob Knight. Diversity, stability and resilience of the human gut microbiota. *Nature*, 489(7415):220–230, 2012.
- [62] Victoria I. Holden and Michael A. Bachman. Diverging roles of bacterial siderophores during infection. *Metallomics : integrated biometal science*, 7(6):986–995, 2015.
- [63] Wilma Neumann, Anmol Gulati, and Elizabeth M. Nolan. Metal homeostasis in infectious disease: recent advances in bacterial metallophores and the human metal-withholding response. *Current opinion in chemical biology*, 37:10–18, 2016.
- [64] Elizabeth M. Nolan, Michael A. Fischbach, Alexander Koglin, and Christopher T. Walsh. Biosynthetic tailoring of microcin e492m: post-translational modification affords an antibacterial siderophore-peptide conjugate. *Journal of the American Chemical Society*, 129(46):14336–14347, 2007.
- [65] Erwin Strahsburger, Marcelo Baeza, Octavio Monasterio, and Rosalba Lagos. Cooperative uptake of microcin e492 by receptors fepa, fiu, and cir and inhibition by the siderophore enterochelin and its dimeric and trimeric hydrolysis products. *Antimicrobial agents and chemotherapy*, 49(7):3083–3086, 2005.
- [66] Xavier Thomas, Delphine Destoumieux-Garzon, Jean Peduzzi, Carlos Afonso, Alain Blond, Nicolas Birlirakis, Christophe Goulard, Lionel Dubost, Robert Thai, Jean-Claude Tabet, and Sylvie Rebuffat. Siderophore peptide, a new type of post-translationally modified antibacterial peptide with potent activity. *The Journal of biological chemistry*, 279(27):28233–28242, 2004.
- [67] Stephen M. Collins, Michael Surette, and Premysl Bercik. The interplay between the intestinal microbiota and the brain. *Nature reviews. Microbiology*, 10(11):735–742, 2012.
- [68] Miles Berger, John A. Gray, and Bryan L. Roth. The expanded biology of serotonin. *Annual review of medicine*, 60:355–366, 2009.
- [69] Javier A. Bravo, Paul Forsythe, Marianne V. Chew, Emily Escaravage, Helene M. Savignac, Timothy G. Dinan, John Bienenstock, and John F. Cryan. Ingestion of lactobacillus strain regulates emotional behavior and central gaba receptor expression in a mouse via the vagus nerve. *Proceedings of the National Academy of Sciences of the United States of America*, 108(38):16050–16055, 2011.
- [70] Meredith Irsfeld, Matthew Spadafore, and Birgit M. Pruss. beta-phenylethylamine, a small molecule with a large impact. *WebmedCentral*, 4(9), 2013.
- [71] Elaine Y. Hsiao. Gastrointestinal issues in autism spectrum disorder. *Harvard review of psychiatry*, 22(2):104–111, 2014.

- [72] Daniel L. Coury, Paul Ashwood, Alessio Fasano, George Fuchs, Maureen Geraghty, Ajay Kaul, Gary Mawe, Paul Patterson, and Nancy E. Jones. Gastrointestinal conditions in children with autism spectrum disorder: developing a research agenda. *Pediatrics*, 130 Suppl 2:S160–8, 2012.
- [73] James B. Adams, Leah J. Johansen, Linda D. Powell, David Quig, and Robert A. Rubin. Gastrointestinal flora and gastrointestinal status in children with autism—comparisons to typical children and correlation with autism severity. *BMC gastroenterology*, 11:22, 2011.
- [74] Paul M. Ridker, Eleanor Danielson, Francisco A. H. Fonseca, Jacques Genest, Antonio M. Gotto, JR, John J. P. Kastelein, Wolfgang Koenig, Peter Libby, Alberto J. Lorenzatti, Jean G. MacFadyen, Borge G. Nordestgaard, James Shepherd, James T. Willerson, and Robert J. Glynn. Rosuvastatin to prevent vascular events in men and women with elevated c-reactive protein. *The New England journal of medicine*, 359(21):2195–2207, 2008.
- [75] Elaine Y. Hsiao, Sara W. McBride, Sophia Hsien, Gil Sharon, Embriette R. Hyde, Tyler McCue, Julian A. Codelli, Janet Chow, Sarah E. Reisman, Joseph F. Petrosino, Paul H. Patterson, and Sarkis K. Mazmanian. Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell*, 155(7):1451–1463, 2013.
- [76] Michael A. Fischbach and Christopher T. Walsh. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chemical reviews*, 106(8):3468–3496, 2006.
- [77] D. E. Cane, C. T. Walsh, and C. Khosla. Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science (New York, N.Y.)*, 282(5386):63–68, 1998.
- [78] D. E. Cane and C. T. Walsh. The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. *Chemistry & biology*, 6(12):R319–25, 1999.
- [79] Robert Finking and Mohamed A. Marahiel. Biosynthesis of nonribosomal peptides1. *Annual review of microbiology*, 58:453–488, 2004.
- [80] Christopher T. Walsh and Michael A. Fischbach. Natural products version 2.0: connecting genes to molecules. *Journal of the American Chemical Society*, 132(8):2469–2493, 2010.
- [81] Christopher T. Walsh, Robert V. O’Brien, and Chaitan Khosla. Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angewandte Chemie (International ed. in English)*, 52(28):7098–7124, 2013.
- [82] Brian K. Hubbard and Christopher T. Walsh. Vancomycin assembly: nature’s way. *Angewandte Chemie (International ed. in English)*, 42(7):730–765, 2003.
- [83] H. D. Mootz and M. A. Marahiel. The tyrocidine biosynthesis operon of bacillus brevis: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *Journal of bacteriology*, 179(21):6843–6850, 1997.
- [84] Florian Kopp and Mohamed A. Marahiel. Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis. *Natural product reports*, 24(4):735–749, 2007.

-
- [85] Chengzhang Fu, Lena Keller, Armin Bauer, Mark Bronstrup, Alexandre Froidbise, Peter Hammann, Jennifer Herrmann, Guillaume Mondesert, Michael Kurz, Matthias Schiell, Dietmar Schummer, Luigi Toti, Joachim Wink, and Rolf Muller. Biosynthetic studies of telomycin reveal new lipopeptides with enhanced activity. *Journal of the American Chemical Society*, 137(24):7692–7705, 2015.
- [86] A. Lawen and R. Zocher. Cyclosporin synthetase. the most complex peptide synthesizing multienzyme polypeptide so far described. *The Journal of biological chemistry*, 265(19):11355–11360, 1990.
- [87] Stefan Muller, Shwan Rachid, Thomas Hoffmann, Frank Surup, Carsten Volz, Nestor Zuburranyi, and Rolf Muller. Biosynthesis of crocacin involves an unusual hydrolytic release domain showing similarity to condensation domains. *Chemistry & biology*, 21(7):855–865, 2014.
- [88] Wei Li, ShenChieh Chou, Ankush Khullar, and Barbara Gerratana. Cloning and characterization of the biosynthetic gene cluster for tomaymycin, an sjg-136 monomeric analog. *Applied and environmental microbiology*, 75(9):2958–2963, 2009.
- [89] Arush Chhabra, Asfarul S. Haque, Ravi Kant Pal, Aneesh Goyal, Rajkishore Rai, Seema Joshi, Santosh Panjikar, Santosh Pasha, Rajan Sankaranarayanan, and Rajesh S. Gokhale. Nonprocessive 2 + 2e- off-loading reductase domains from mycobacterial nonribosomal peptide synthetases. *Proceedings of the National Academy of Sciences of the United States of America*, 109(15):5681–5686, 2012.
- [90] Liangcheng Du and Lili Lou. Pks and nrps release mechanisms. *Natural product reports*, 27(2):255–278, 2010.
- [91] Fatoumata Tambadou, Thibault Caradec, Anne-Laure Gagez, Antoine Bonnet, Valerie Sopena, Nicolas Bridiau, Valerie Thiery, Sandrine Didelot, Cyrille Barthelemy, and Romain Chevrot. Characterization of the colistin (polymyxin e1 and e2) biosynthetic gene cluster. *Archives of microbiology*, 197(4):521–532, 2015.
- [92] C. Muller, S. Nolden, P. Gebhardt, E. Heinzelmann, C. Lange, O. Puk, K. Welzel, W. Wohlleben, and D. Schwartz. Sequencing and analysis of the biosynthetic gene cluster of the lipopeptide antibiotic friulimicin in *actinoplanes friuliensis*. *Antimicrobial agents and chemotherapy*, 51(3):1028–1037, 2007.
- [93] Subrata Paul, Hiroaki Ishida, Leonard T. Nguyen, Zhihong Liu, and Hans J. Vogel. Structural and dynamic characterization of a freestanding acyl carrier protein involved in the biosynthesis of cyclic lipopeptide antibiotics. *Protein science : a publication of the Protein Society*, 2017.
- [94] C. Khosla, R. S. Gokhale, J. R. Jacobsen, and D. E. Cane. Tolerance and specificity of polyketide synthases. *Annual review of biochemistry*, 68:219–253, 1999.
- [95] J. Staunton and K. J. Weissman. Polyketide biosynthesis: a millennium review. *Natural product reports*, 18(4):380–416, 2001.

- [96] Elizabeth S. Sattely, Michael A. Fischbach, and Christopher T. Walsh. Total biosynthesis: in vitro reconstitution of polyketide and nonribosomal peptide pathways. *Natural product reports*, 25(4):757–793, 2008.
- [97] B. J. Rawlings. Type i polyketide biosynthesis in bacteria (part a—erythromycin biosynthesis). *Natural product reports*, 18(2):190–227, 2001.
- [98] Lauren Ray and Bradley S. Moore. Recent advances in the biosynthesis of unusual polyketide synthase substrates. *Natural product reports*, 33(2):150–161, 2016.
- [99] Nick Quade, Liujie Huo, Shwan Rachid, Dirk W. Heinz, and Rolf Muller. Unusual carbon fixation gives rise to diverse polyketide extender units. *Nature chemical biology*, 8(1):117–124, 2011.
- [100] Manuel A. Ortega and van der Donk, Wilfred A. New insights into the biosynthetic logic of ribosomally synthesized and post-translationally modified peptide natural products. *Cell chemical biology*, 23(1):31–44, 2016.
- [101] John A. McIntosh, Mohamed S. Donia, and Eric W. Schmidt. Ribosomal peptide natural products: bridging the ribosomal and nonribosomal worlds. *Natural product reports*, 26(4):537–559, 2009.
- [102] Julian D. Hegemann, Marcel Zimmermann, Xiulan Xie, and Mohamed A. Marahiel. Lasso peptides: an intriguing class of bacterial natural products. *Accounts of chemical research*, 48(7):1909–1919, 2015.
- [103] Clarissa Melo Czekster, Ying Ge, and James H. Naismith. Mechanisms of cyanobactin biosynthesis. *Current opinion in chemical biology*, 35:80–88, 2016.
- [104] Lindsay M. Repka, Jonathan R. Chekan, Satish K. Nair, and van der Donk, Wilfred A. Mechanistic understanding of lanthipeptide biosynthetic enzymes. *Chemical reviews*, 2017.
- [105] Tobias W. Giessen and Mohamed A. Marahiel. Ribosome-independent biosynthesis of biologically active peptides: Application of synthetic biology to generate structural diversity. *FEBS letters*, 586(15):2065–2075, 2012.
- [106] Pascal Belin, Mireille Moutiez, Sylvie Lautru, Jerome Seguin, Jean-Luc Pernodet, and Muriel Gondry. The nonribosomal synthesis of diketopiperazines in trna-dependent cyclodipeptide synthase pathways. *Natural product reports*, 29(9):961–979, 2012.
- [107] Christopher Walsh and Timothy Wencewicz. *Antibiotics: Challenges, mechanisms, opportunities*. American Society for Microbiology, 2016.
- [108] David M. Morens, Gregory K. Folkers, and Anthony S. Fauci. The challenge of emerging and re-emerging infectious diseases. *Nature*, 430(6996):242–249, 2004.
- [109] Louis Pasteur, J. Joubert, and Ch. Chamberland. *La théorie des germes et ses applications a la médecine et a la chirurgie: Lecture faite a l’Académie de médecine*. G. Masson, Paris, 1878.

-
- [110] Robert Koch. *Zur Aetiologie des Milzbrandes*. Gerschel, Berlin, 1881.
- [111] Robert Koch. *Zur Untersuchung von pathogenen Organismen*. [Norddeutschen Buchdruckerei und Verlagsanstalt], [Berlin], 1881.
- [112] Paul Ehrlich and S. Hata. *Salvarsan, Arsenobenzol, Hiperideal*. Leipzig, 2. auflage edition, 1907.
- [113] Alexander Fleming. On the antibacterial action of cultures of a penicillium: With special reference to their use in isolation of b. influenzæ. *Britis Journal of Experimental Pathology*, 10(3):266–236, 1929.
- [114] Gerhard Domagk. Ein beitrag zur chemotherapie der bakteriellen infektionen. *Angewandte Chemie*, 46:657–667, 1935.
- [115] Michael A. Fischbach and Christopher T. Walsh. Antibiotics for emerging pathogens. *Science (New York, N.Y.)*, 325(5944):1089–1093, 2009.
- [116] Christopher T. Walsh and Timothy A. Wencewicz. Prospects for new antibiotics: a molecule-centered perspective. *The Journal of antibiotics*, 67(1):7–22, 2014.
- [117] Damien Maura, Alicia E. Ballok, and Laurence G. Rahme. Considerations and caveats in anti-virulence drug development. *Current opinion in microbiology*, 33:41–46, 2016.
- [118] Iwona Bucior, Jason Abbott, Yuanlin Song, Michael A. Matthay, and Joanne N. Engel. Sugar administration is an effective adjunctive therapy in the treatment of pseudomonas aeruginosa pneumonia. *American journal of physiology. Lung cellular and molecular physiology*, 305(5):L352–63, 2013.
- [119] Ori J. Lieberman, Mona W. Orr, Yan Wang, and Vincent T. Lee. High-throughput screening using the differential radial capillary action of ligand assay identifies ebselen as an inhibitor of diguanylate cyclases. *ACS chemical biology*, 9(1):183–192, 2014.
- [120] Ewan J. Murray, Rebecca C. Crowley, Alex Truman, Simon R. Clarke, James A. Cottam, Gopal P. Jadhav, Victoria R. Steele, Paul O’Shea, Catharina Lindholm, Alan Cockayne, Siri Ram Chhabra, Weng C. Chan, and Paul Williams. Targeting staphylococcus aureus quorum sensing with nonpeptidic small molecule inhibitors. *Journal of medicinal chemistry*, 57(6):2813–2819, 2014.
- [121] Melissa Starkey, Francois Lepine, Damien Maura, Arunava Bandyopadhyaya, Biljana Lesic, Jianxin He, Tomoe Kitao, Valeria Righi, Sylvain Milot, Aria Tzika, and Laurence Rahme. Identification of anti-virulence compounds that disrupt quorum-sensing regulated acute and persistent pathogenicity. *PLoS pathogens*, 10(8):e1004321, 2014.
- [122] Yongcheng Song, Chia-I Liu, Fu-Yang Lin, Joo Hwan No, Mary Hensler, Yi-Liang Liu, Wen-Yih Jeng, Jennifer Low, George Y. Liu, Victor Nizet, Andrew H-J Wang, and Eric Oldfield. Inhibition of staphyloxanthin virulence factor biosynthesis in staphylococcus aureus: in vitro, in vivo, and crystallographic results. *Journal of medicinal chemistry*, 52(13):3869–3880, 2009.

- [123] Paul Warrener, Reena Varkey, Jessica C. Bonnell, Antonio DiGiandomenico, Maria Camara, Kimberly Cook, Li Peng, Jingying Zha, Partha Chowdury, Bret Sellman, and C. Kendall Stover. A novel anti-pcrv antibody providing enhanced protection against pseudomonas aeruginosa in multiple animal infection models. *Antimicrobial agents and chemotherapy*, 58(8):4384–4391, 2014.
- [124] Lun K. Tsou, Paul D. Dossa, and Howard C. Hang. Small molecules aimed at type iii secretion systems to inhibit bacterial virulence. *MedChemComm*, 4(1):68–79, 2013.
- [125] Inc KaloBios Pharmaceuticals. Kalobios reports top-line data for phase 2 study of kb001-a to treat pseudomonas aeruginosa lung infections in cystic fibrosis patients, 2016.
- [126] Bruno Francois, Charles-Edouard Luyt, Anthony Dugard, Michel Wolff, Jean-Luc Diehl, Samir Jaber, Jean-Marie Forel, Denis Garot, Eric Kipnis, Alexandre Mebazaa, Benoit Misset, Antoine Andreumont, Marie-Cecile Ploy, Alan Jacobs, Geoffrey Yarranton, Tillman Pearce, Jean-Yves Fagon, and Jean Chastre. Safety and pharmacokinetics of an anti-pcrv pegylated monoclonal antibody fragment in mechanically ventilated patients colonized with pseudomonas aeruginosa: a randomized, double-blind, placebo-controlled trial. *Critical care medicine*, 40(8):2320–2326, 2012.
- [127] Hans-Peter Hauber, Maria Schulz, Almuth Pforte, Dietrich Mack, Peter Zabel, and Udo Schumacher. Inhalation with fucose and galactose for treatment of pseudomonas aeruginosa in cystic fibrosis patients. *International journal of medical sciences*, 5(6):371–376, 2008.
- [128] Won-Young Kim and Sang-Bum Hong. Sepsis and acute respiratory distress syndrome: Recent update. *Tuberculosis and respiratory diseases*, 79(2):53–57, 2016.
- [129] Wei Li, ShenChieh Chou, Ankush Khullar, and Barbara Gerratana. Cloning and characterization of the biosynthetic gene cluster for tomaymycin, an sjg-136 monomeric analog. *Applied and environmental microbiology*, 75(9):2958–2963, 2009.
- [130] Georg Schneditz, Jana Rentner, Sandro Roier, Jakob Pletz, Kathrin A. T. Herzog, Roland Bückner, Hanno Troeger, Stefan Schild, Hansjörg Weber, Rolf Breinbauer, Gregor Gorkiewicz, Christoph Högenauer, and Ellen L. Zechner. Enterotoxigenicity of a nonribosomal peptide causes antibiotic-associated colitis. *Proceedings of the National Academy of Sciences of the United States of America*, 111(36):13181–13186, 2014.

Chapter 2

Total biosynthesis of tomaymycin comprehensively monitored on the nonribosomal peptide megasyntetase

ALEXANDER VON TESMAR*, MICHAEL HOFFMANN*, JAN PIPPEL, ANTOINE ABOU
FAYAD, STEFAN WERNER, ARMIN BAUER, WULF BLANKENFELDT, ROLF MÜLLER[†]

*contributed equally

[†]To whom correspondence should be addressed.

Author's efforts

The author significantly contributed to the conception of the study, designed and performed experiments, evaluated and interpreted resulting data . The author performed cloning, expression and purification of all proteins used for the in vitro reconstitution. Furthermore, the author designed and performed all in vitro assays for reconstitution of tomaymycin biosynthesis and determination of underlying substrate specificity. The author contributed equally to conceiving and writing of the manuscript.

Contributions by others

Michael Hoffmann significantly contributed to the conception of the study, designed and performed experiments, evaluated and interpreted resulting data. He developed the LC-MS method used to analyze intact proteins and performed all in vitro assays regarding intact protein measurements. All LC-MS measurements described as well as the evaluation and interpretation of the respective data were conducted by him. Furthermore he contributed equally to conceiving and writing of the manuscript.

Jan Pippel cloned, purified and crystalized TomG, solved the structure by X-Ray analysis, interpreted the results and contributed to the writing and editing of the manuscript. All described synthesis including the respective purification were conducted by Antoine Abou-Fayad. Stefan Werner provided the initial cloning for the NRPS constructs. The anthranilic acid derivatives used were provided by Armin Bauer. The project was supervised by Wulf Blankenfeldt and Rolf Müller, who also contributed to conceiving and proofreading of the manuscript.

Abstract

In vitro reconstitution and subsequent biochemical analysis of natural product biosynthetic pathways remains a challenging endeavor, especially if megaenzymes of the polyketide synthase (PKS) and/or nonribosomal peptide synthetase (NRPS) type are involved. In theory, all biosynthetic steps may be deciphered using mass spectrometry based analyses of both the carrier protein-coupled intermediates and the free intermediates. We here report the total biosynthesis of the pyrrolo[4,2]benzodiazepine scaffold tomaymycin using an *in vitro* reconstituted NRPS system. Direct LC-MS analyses of proteoforms were employed to directly decipher every step of the biosynthesis on its respective megasynthetase with up to 170 kDa in size. To the best of our knowledge, this is the first report of a comprehensive analysis of virtually all chemical steps involved in the biosynthesis of a nonribosomally synthesized natural product catalyzed by an NRPS. The study includes experiments to determine substrate specificities of the corresponding A-domains in competition assays by analyzing the adenylation step as well as the transfer to the respective carrier protein domain. Additionally, condensation reactions were monitored and the formation of the final product was analyzed.

2.1 Introduction

Natural products account for a large part of the known therapeutics and thus remain to be a major topic of drug discovery.¹ Understanding the underlying biosynthetic pathways for the production of these compounds in Nature sheds light on their assembly, regulation and mode of action and facilitates all further efforts towards their use as chemical tools or even in drug development. Hence, the advancement, expansion and application of methods to analyze biosynthetic pathways is still a crucial process ideally involving *in vitro* reconstitution and (bio)chemical analysis of complex multistep reaction sequences.

Tomaymycin, a pyrrolo[4,2]benzodiazepine (PBD), is an antitumor antibiotic produced among others by *Streptomyces achromogenes*.² PBD core structures are found in many natural products and occur in a variety of substitution patterns.³ Covalent bonding to the minor groove of the DNA double helix in a sequence specific manner is enabled by the PBD imine. This gives rise to the observed *in vitro* inhibition of the phage T7 RNA polymerase⁴ and antitumor activity.^{4, 5} The underlying structure activity relationship is well studied and has led to an array of synthetic PBDs in pursuit of a potent antitumor drug.^{6–8} Previous biosynthesis studies on tomaymycin, sibiromycin and anthramycin^{9–11} revealed common genes encoding a bi-modular nonribosomal peptide synthetase (NRPS) as the basis of the biosynthetic pipeline for PBD production.

NRPSs are multimodular megasynthetases that connect single building blocks in a protein-templated fashion to generate diverse groups of natural products. Independent of the ribosome, NRPSs are not limited to proteinogenic amino acids and thus unusual modifications are often incorporated, leading to highly diverse structural features. The minimal module for chain elongation consists of three parts: the adenylation domain (A), responsible for activating the substrate; the thiolation domain (T), onto which the amino acid is covalently tethered to the terminal thiol of a 4-phosphopantethein prosthetic group (PPant); and the condensation domain (C) that catalyzes formation of the peptide bond between two adjacent T-bound moieties. Additional domains can be present to facilitate a variety of chemical variations to the growing peptide chain, such as methylation, hydroxylation, or epimerization.¹² The majority of nonribosomal peptides is released by the action of either a thioesterase (T) or a reductase domain (Re).^{13, 14}

Although a large number of NRPS systems have been identified based on sequencing results, only a small portion is connected to their encoded small molecule products. Detailed studies were described for a total of twelve NRPS systems that have been reconstituted *in vitro*,¹⁵ including aureusimine,^{16, 17} PF1022,^{18, 19} pacidamycin,²⁰ antimycin,²¹ beauvericin²² and the siderophores enterobactin,²³ yersinibactin,²⁴ pyochelin,²⁵ vibriobactin,²⁶ myxochelin^{27, 28} and pseudomonine.¹⁵ Recently, the activity of the 652 kDa valinomycin NRPS could be fully restored in *E. coli*, proving the capability of the host to express very large megasynthetases in a functional form.²⁹ Although previously studied NRPS systems such as the one producing aureusimine AusA^{16, 17} share some features of domain architecture with tomaymycin, no *in vitro* reconstitution of a full PBD biosynthetic pathway has been reported so far.

Detailed study of NRPS as well as PKS (polyketide synthase) *in vitro* systems on the molecular level became mainly accessible in the last two decades by crucial developments in instrumental analytics.³⁰ The high sensitivity detection of small molecules (< 2000 Da) by LC-MS measurements is the basis of most published NRPS *in vitro* studies. The detection of complete intact

protein complexes has been described up to a mass of 18 MDa³¹ under certain conditions and in general with this approach it is feasible to identify the molecular weight or post-translational modifications of proteins > 100 kDa.^{32, 33} The interests in NRPS/PKS proteomics are centered on proteoform analysis and thus especially on the PPant moiety, as it covalently binds the growing substrate chain during the assembly of natural products.

We here provide a methodology to directly visualize each intermediate step on a NRPS assembly line. All proteoforms (up to 170 kDa in size) constituting the cycle of tomaymycin biosynthesis that includes substrate loading, dipeptide formation and product release, have been detected in a direct intact protein LC-MS experiment, thereby deciphering all chemical steps involved. In conjunction with conventional LC-MS analysis as well as biochemical characterization of the substrate specificity, we display a comprehensive monitoring of a full NRPS system. Additionally, the methyltransferase TomG was characterized biochemically and structurally, allowing insights in the timing of the tailoring reactions involved in tomaymycin biosynthesis.

2.2 Results and Discussion

In vitro Reconstitution of TomA and TomB

The two adjacent NRPS-encoding genes tomA and tomB were previously proposed to facilitate the biosynthesis of the tomaymycin PBD core structure.¹¹ However, until now the biochemical proof for the proposed biochemistry of PBD formation has not been provided.

To examine the biosynthetic machinery in detail, both genes constituting the NRPS, *tomA* and *tomB*, as well as the isolated TomA A-domain were heterologously expressed in *E. coli* BL21 (DE3) together with a vector harboring the promiscuous PPant-transferase MtaA.³⁴ An N-terminal, HRV3C cleavable, 6xHis-MBP tag was used as it showed maximized yield and solubility. All proteins were purified to homogeneity by a three-step protocol including nickel-affinity chromatography, a reverse nickel-affinity chromatography step after HRV3C digestion to remove the tags and a polishing gel filtration step.

Initially, the adenylation activity of the TomA A-domain was examined *in vitro* in a hydroxamate methylthioguanosine (MesG) based assay, revealing a relaxed substrate specificity (Figure 2.1). Hydroxylamine acts as surrogate acceptor molecule and mimics the missing PPant-arm, releasing the tightly bound acyl-adenylate from the A-domain by hydroxamate formation. Adenylation activity is finally quantified using the pyrophosphatase-purine nucleoside phosphorylase coupling system.³⁵ Interestingly, the strongest adenylation activity was observed for 3-trifluoromethyl and 3-bromo anthranilic acid, both non-natural derivatives. Within the tomaymycin related substrates, the 5-hydroxy derivative shows faster adenylation rate followed by 4-hydroxy, 5-methoxy and 4,5-dihydroxy anthranilic acid. The adenylation of fully tailored 4-hydroxy-5-methoxy anthranilic acid by the TomA A-domain is very slow, making the role as a natural substrate implausible (Figure S 2.9). Due to the promiscuous substrate specificity, further conclusions for the tomaymycin biosynthetic pathway by the examination of the excised A-domain alone are not meaningful.

Thus, the purified proteins TomA and TomB were used to reconstitute the NRPS activity *in vitro*. Incubation of TomA and TomB with the corresponding anthranilic acid derivative, 4-methylene-proline, ATP and NADPH, followed by LC-MS/MS analysis verified the presence of

the corresponding carbinolamine (Figure S 2.8). The lack of either cofactor, ATP or NADPH, abolished product formation. TomB is able to use NADPH as well as NADH for the reductive release in equivalent efficiency. The NRPS-released aldehyde immediately reacts to a circularized imine that could be detected in trace amounts. The equilibrium of the imine and its corresponding carbinolamine form in aqueous solution showed to be strongly in favor of the latter. In addition to the observed tomaymycin derivatives, two by-products appeared and were identified as anthranilic acid proline dimers. Isolated incubation showed that each module is solely responsible for the formation of one dimer, probably by transesterification of the loaded cognate substrate to the carboxylic acid group of the remaining free substrate (Figure S 2.10). The *in vitro* reconstituted NRPS TomA/TomB readily assembled the corresponding carbinolamine for five out of fourteen tested anthranilic acid derivatives. Several analyzed derivatives of anthranilic acid, e.g. carrying substituents with sterically demanding groups like -CF₃, iodine, bromine, chlorine and various hydroxyl and methoxy substitution patterns showed no incorporation (Figure 2.1). Incorporation of 4,5-dihydroxy anthranilic acid, previously hypothesized as a potential intermediate by Li et al., could not be detected (Li et al., 2009a). Further, incorpo-

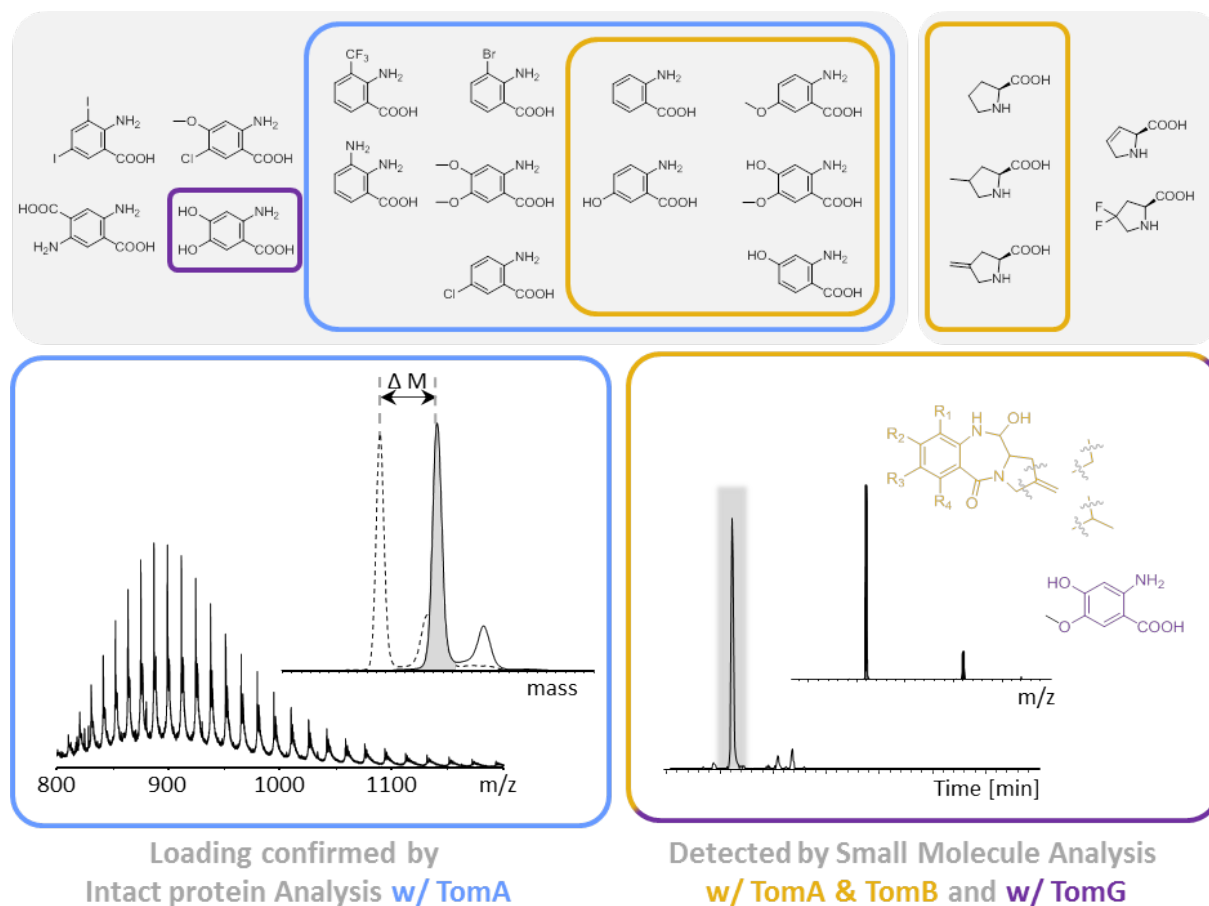


Figure 2.1: Chemical structures of the fourteen tested anthranilic acid as well as the five tested proline derivatives including a schematic overview of the applied methodology. Derivatives, which could serve as potential substrates based on the results of the respective methods are highlighted with colored frames. Based on *in vitro* loading assays with TomA and subsequent LC-MS measurements on intact protein level ten derivatives would be conceivable (blue frames). Further reduction to five could be observed by performing *in vitro* reconstitution assays with respective LC-MS measurements on small molecule level (yellow frames). *In vitro* assays with TomG with respective small molecule analytics showed that only one derivative is processed (purple frames).

ration of three proline derivatives to the corresponding carbinolamine could be observed in trace amounts (Figure 2.1) whereas 3-difluoro and 2,3-dehydro proline was not accepted by TomB. In summary, eight different tomaymycin derivatives were generated *in vitro*. The tomaymycin NRPS is thus capable to generate PBDs of large chemical diversity, making it a useful tool for synthetic biology approaches. The adenylation domain, although responsible for the loading pro-

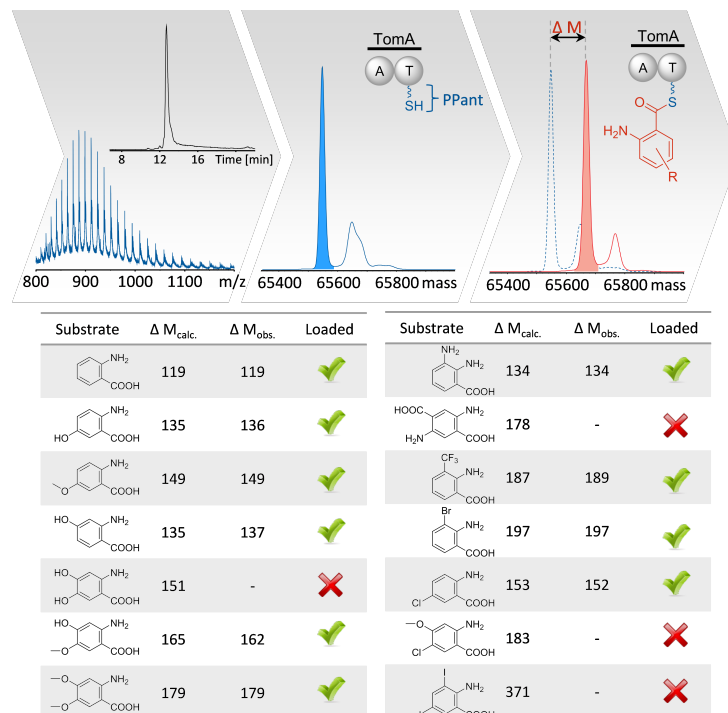


Figure 2.2: Scheme showing the workflow of TomA loading assays including the results of loading attempts with different, available anthranilic acid derivatives: The top panel shows exemplary LC-MS results obtained for loading assays of TomA. Charge state distributions were typically found to be in the range of 800 to 1200 m/z and deconvoluted mass spectra allowed to obtain the respective protein mass typically within < 10 ppm. Mass shifts shown in the lower panel are given as delta masses with respect to holo TomA.

cess itself, is not the only instance in selecting the building block. Especially the condensation domain is known to influence the final product, e.g. by combined epimerization-condensation reactions,^{36, 37} formation of a -lactam rings³⁸ or the recruitment of in trans acting tailoring enzymes.³⁹ Additionally, it is long known that the C-domain possesses substrate specificity and thus can act as additional gatekeeper beside the A-domain.^{40, 41} In the tomaymycin NRPS the broad substrate tolerance of the TomA adenylation domain seems to be constrained by the condensation domain. Similar effects on the adenylation specificity by a condensation domain have been shown in a recent study of the microcystin NRPS.⁴²

Intact Protein MS of TomA and TomB

To address the intrinsically limited significance of experiments based on isolated domains, intact protein MS analysis of TomA and TomB was conducted to examine the biosynthetic machinery as a whole. Evidence based on MS data has become an important tool to elucidate biochemical processes that are tethered to NRPS or PKS systems, but such experiments were mainly performed by utilizing proteolysis prior to measurements rather than analyzing the intact protein directly.^{43–49} In many of these studies, the PPant-ejection assay⁵⁰ was used. Its applicability

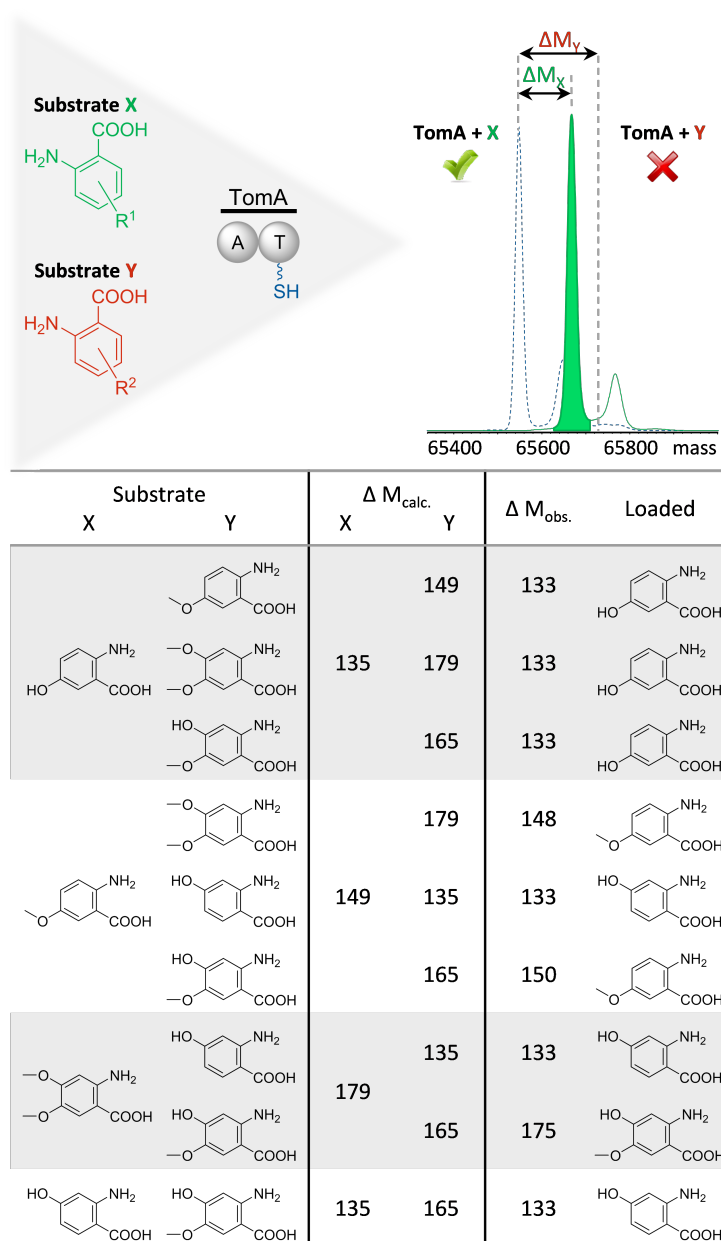


Figure 2.3: Approaches towards the analysis of substrate specificity of TomA by competitive loading assays with pairs of anthranilic acid derivatives in equimolar ratio. Detected delta masses are related to holo TomA and showed which substrate is preferred.

for large multiple charged proteins is challenging due to a limited total energy deposition by the standard collision-induced dissociation process (CID), resulting in low fragmentation efficiency.^{51–53} Nevertheless, this approach led to the successful identification of several intermediates on the fungal iterative polyketide synthases responsible for the biosynthesis of aflatoxin B1.⁵⁴ However, the method requires an additional laborious sample preparation step. Up to now, only few studies have reported intact protein measurements on megasynthases for proteins with a limited size (< 125 kDa) in combination with direct infusion (DI) experiments such as in the recent determination of the kinetic profile of the gramicidin S synthetase A by quantifying the substrate tethered proteoform⁵⁵ or the visualization of inhibitor binding.⁵⁶ Generally MS-based analysis of substrate loading and processing of intact NRPS-proteins faces two major difficulties: the limited solubility of large proteins in salt-free solutions, which are mandatory for

direct infusion experiments, as well as the analytical difficulties connected to the occurring high charge states (up to $z=200$ for $m=170$ kDa). In our study, an LC separation using a monolithic PS-DVB column was instrumental to overcome the solubility issues in DI-MS experiments by providing desalting, separation and focusing prior to data acquisition using a time of flight mass spectrometer (ToF-MS) and similar to bottom-up proteomics, the addition of DMSO to the eluents (1-2 % (v/v)) resulted in an increase of spectrum quality.⁵⁷ This observation is based on narrowing down the wide charge state distribution which leads to an increase of the remaining ion species intensity by approximately factor 3 (Figure S 2.7). In summary, the described methodology provides significant advantages in its applicability for larger sample numbers because it circumvents most laborious sample preparation techniques. Although it leads to limited spectra accumulation in comparison to DI experiments, the chromatographic focusing plus the addition of DMSO increases the spectra quality and as such compensates sufficiently the drawbacks. Based on a substantial number of in-house experiments we conclude that the developed workflow allows the detection of mass shifts of > 100 Da for proteins of 170 kDa with a typical mass accuracy of < 10 ppm. Deconvoluted mass spectra resulting from this method confirmed the presence of the holo-form of both, TomA and TomB, as indicated by an observed mass shift of 340 Da relative to the primary sequence-derived protein mass (Figure 2.4).

In vitro assay conditions were developed to perform a fast and reliable LC-MS based substrate specificity testing for TomA. Fourteen different anthranilic acid derivatives were tested and the resulting deconvoluted protein spectra analyzed for the envisaged mass shifts (Figure 2.2). In total, 10 out of 14 tested anthranilic acid derivatives were loaded onto the carrier domain. Notably, 4,5-dihydroxy anthranilic acid is not processed by TomA. In contrast, only five derivatives were completely processed in the above described *in vitro* reconstitution assays towards the respective PBDs. For example, halogenated anthranilic acids were activated and loaded onto the carrier protein but were not further processed (Figure 2.1 and 2.2). This again indicates that downstream domains restrict the tolerance of tomaymycin biosynthesis towards carrier protein loaded anthranilic acid derivatives. To examine the relative substrate preference between two given derivatives, competition assays were set up using equimolar amounts of two substrates with TomA simultaneously. 5-hydroxy anthranilic acid was preferred by TomA over all other derivatives, followed by 4-hydroxy- and 4-hydroxy-5-methoxy anthranilic acid (Figure 2.3). Remarkably, no mixed loading could be observed in any experiment, emphasizing a clear preference in all tested scenarios.

In addition to the comprehensive analysis of TomA by the described LC-MS method it was our main goal to show the feasibility of this method for larger proteins and to monitor every single step of the biosynthesis by means of MS detection. Indeed, we were able to apply this approach also for TomB, despite the large size of approximately 170 kDa, and could detect the calculated mass shift for the holo-form as well as the subsequent methylene proline loading (+110 Da) (Figure 2.4). To achieve a comprehensive monitoring of the assembly of tomaymycin, an *in trans* assay was developed. TomA and TomB were incubated with anthranilic acid, methylene proline, and ATP, but without NADPH. Direct LC-MS analysis of the *in vitro* system lead to the identification of a signal corresponding to the mass of the predicted anthranilic acid-proline dipeptidic intermediate tethered to TomB (Figure 2.4). Upon addition of NADPH all signals corresponding to loaded forms of TomB disappeared with only the holo-form remaining. Small

molecule analysis by LCMS after protein removal proved the presence of trace amounts of the respective tomaymycin derivative and thus the completion of the biosynthetic pathway (Figure 2.4). However, significant amounts of hetero-dipeptides of the two substrates were detected and were shown to be formed by TomA or TomB alone with slightly different retention times. The rapidly activated substrate is probably transferred to the free amino group of the unactivated species, in case the correct protein partner is missing (Figure 2.10). It is not surprising that no full conversion from the proline-loaded TomB to the dipeptide-loaded TomB species could be detected as substrate activation and loading might be faster than the *in trans* condensation plus product release reactions (the latter may also be limited due to suboptimal protein folding). Preferred formation of the dipeptide species in the *in vitro* assay may be due to insufficient formation of a TomAB complex, which is likely required for efficient substrate channeling. Such a complex is expected to be formed in the native producer. The reason for the low efficiency of product formation currently remains unclear but obviously, the *in vitro* assays do not perfectly mimic the *in vivo* situation in the producing actinomycete species. Naturally, the resulting low amount of the species of interest complicates the detection with high mass accuracy.

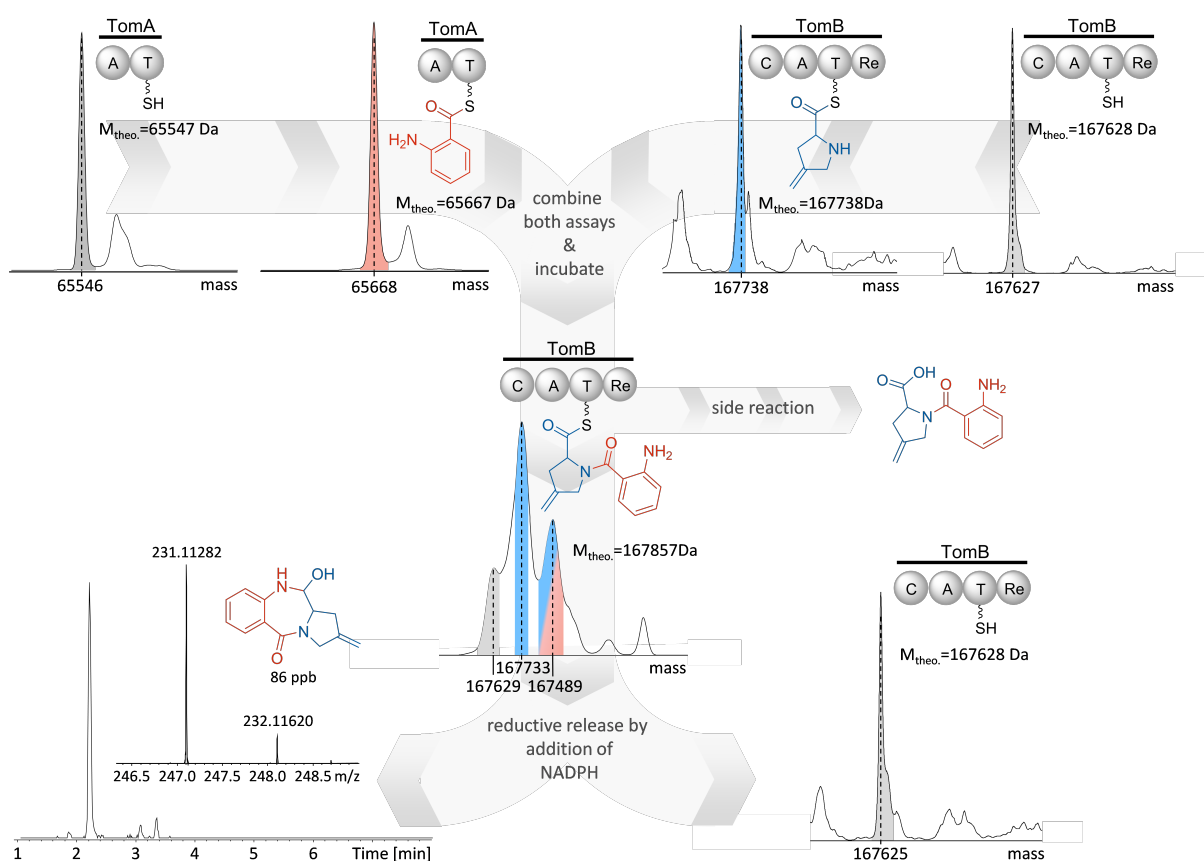


Figure 2.4: Composed steps of the tomaymycin *in vitro* reconstitution including the respective deconvoluted mass spectra for each protein species to confirm the substrate binding and thus monitoring the complete biosynthesis. The observed mass deviation in case of the dipeptide-loaded TomB is attributable to overlapping MS signals of the occurring species due to incomplete conversion. Hence, the limits of maximum entropy deconvolution algorithm is reached leading to inaccurate mass detection.

***In vitro* Reconstitution and Structural Analysis of TomG**

The *in vitro* reconstitution of the tomaymycin NRPS revealed a high substrate tolerance for the anthranilic acid activating module TomA. The question which anthranilic acid derivative is activated and processed by the NRPS to produce tomaymycin could therefore not be answered unambiguously by examining the isolated megasynthetase. We therefore additionally analyzed the influence of the O-methyltransferase (OMT) TomG that has previously been speculated by Li et al. to be responsible for the methylation of the 5-hydroxy position of the anthranilic acid derivative.

For this purpose, two TomG constructs were generated. For biochemical analysis, *tomG* was cloned into pET28b, expressed in *E. coli* BL21 (DE3) and the respective 6xHis-tagged protein was purified to homogeneity in a two-step protocol comprising nickel-affinity and size exclusion chromatography. Additionally, for crystallization experiments, TomG was analogously expressed as an N-terminal fusion protein with a 6xHis-T7-lysozyme-tag and purified to homogeneity by a three-step protocol comprising two nickel-affinity chromatography steps and a polishing gel filtration step.

The crystal structure of TomG was determined by molecular replacement and revealed the Rossman-like fold of a homodimeric SAM-dependent class I OMT (Figure 2.5). Common features of these OMTs are a conserved SAM binding site and a methyl group transfer from SAM to its substrate via a Sn2 reaction. This mechanism involves an Mg²⁺ ion which positions the respective target group by coordinating the hydroxyl group(s) of the substrate.⁵⁸⁻⁶¹ In both chains of the TomG crystal structure, additional electron density corresponding to SAM could be identified whereas no Mg²⁺ ion was detected in the active site.

In LC-MS analysis SAM-dependent methylation of 4,5-dihydroxy anthranilic acid at the 5-hydroxy position by TomG could be verified *in vitro* by and comparison with authentic standards. In contrast, no activity on 5-hydroxy anthranilic acid could be observed (Figure 2.2), likely because the missing second hydroxy group in ortho-position is essential for the reaction as also described for OMTs in the catechol metabolism.^{62, 63} 4-hydroxylation of the tomaymycin anthranilic acid moiety is most likely accomplished by TomE and TomF as already supposed by Li et al. and the high homology of these proteins to the 4-nitrophenol oxidation system from *Rhodococcus* sp.⁶⁴ supports this hypothesis. However, despite considerable efforts, both proteins could not be reconstituted *in vitro*.

Given the clear substrate preference of TomA for 5-hydroxy anthranilic acid and the fact that TomG activity requires two adjacent hydroxyl groups at the anthranilic acid moiety, two main scenarios for the timing of the tailoring reactions are conceivable. Firstly, hydroxylation and methylation could occur on-line, while 5-anthranilic acid is tethered covalently to the megasynthetase, as it is reported for the anthranilate precursor in sibiromycin biosynthesis.⁶⁵ Secondly, the tailoring reactions could take place after product release with the PBD as substrate. Possible on-line methylation and subsequent processing of 5-methoxy anthranilic acid without prior hydroxylation of the 4-position can be excluded as incubation with TomA, TomB, and TomG alone did not yield the respective 5-methoxy-PBD.

To evaluate the option of a possible post-NRPS methylation of the PBD core structure, the corresponding tomaymycin analogs of 5-hydroxy and 4,5-dihydroxy anthranilic acid were chemically synthesized. Retrosynthetic analysis revealed substituted 2-amino benzoic acids and L-proline

as suitable main building blocks. The first step of the synthesis involved a standard BOC protection of the amino group of the benzoic acid followed by a peptide coupling between the latter compound and L-proline methyl ester. The aldehyde was obtained by reducing the methyl ester functional group with diisobutylaluminum hydride (DIBAL-H). Upon de-protection of the BOC group of the dipeptide moiety, an intramolecular reductive amination took place forming the 1,4-diazepin-2-one-ring which is the core of tomaymycin. These synthetic tomaymycin derivatives were used as substrate for TomG, but no methylation could be observed. These results hint towards an on-line scenario, in which the covalently tethered 5-hydroxy anthranilic acid is first processed by TomE and TomF to provide the required 4-hydroxylation before methylation proceeds by TomG.

In this light, the availability of the crystal structure of TomG allowed us to examine the active site for possible steric factors that could be accountable for the observed substrate specificity. TomG shares highest structural similarity with an OMT from *Bacillus cereus* (BcOMT2,⁶⁶ pdb code: 3duw, Z-score = 34.2, sequence identity = 48 %, Figure S 2.12). BcOMT2 shows promiscuous substrate tolerance for flavonoids,^{66, 67} which have a size similar to PBDs. Accordingly, both enzymes exhibit an solvent-exposed substrate binding pocket with the electronegative bottom being formed by residues Lys210, Gly211 and Tyr212 (TomG numbering) (Figure 2.5). However, in TomG, the walls of this pocket are composed of less flexible, nonpolar residues (backbone atoms of Asp38 and side chain atoms of Pro37, Leu39, Leu169, and Gly170) (Figure 5B). Thus, in TomG the hydrophobic substrate binding pocket is less narrow and more exposed to the solvent compared to BcOMT2 and might therefore not be able to completely host flavonoid sized molecules. Because no prominent steric factors were identified that could discriminate between 4,5-dihydroxy anthranilic acid and the corresponding 4,5-dihydroxy-PBD, additional *in-silico* modeling on the basis of the binding mode of 3,5-dinitrocatechol (DNC) in human catechol-OMT (COMT) (pdb code: 3bwm)⁶⁸ was performed. TomG and human COMT share a highly conserved metal binding site and both enzymes methylate substrate(s) with similar scaffolds by the same mechanism.^{11, 63, 69} According to these calculations, the substrate binding site of TomG can accommodate both 4,5-dihydroxy anthranilic acid and the corresponding 4,5-dihydroxy-PBD without requiring distortion of the ligands.⁶⁶ The benzyl ring of both compounds resides in a narrow cleft, where it forms hydrophobic contacts with the protein and is oriented by polar chelating interactions between the two hydroxyl groups and the Mg²⁺ cation. The amino group of 4,5-dihydroxy anthranilic acid is surrounded by residues with electronegative potential. However, these interactions seem rather weak such that this basic and the negatively charged carboxylate group are largely solvent-exposed. Together, this indicates that the active site of TomG is perfectly suited to interact with 4,5-dihydroxy anthranilic acid. Simultaneously, given the solvent-exposed binding mode, productive binding of 4,5-dihydroxy anthranilic acid covalently linked to the PPant moiety of TomA in the proposed on-line scenario is also feasible. The predicted binding mode of the 4,5-dihydroxy-PBD, on the other hand, seems less plausible, since large parts of its more hydrophobic pyrrolo moiety would be solvent-exposed (Figure 2.5), which may explain why TomG does not methylate PBDs *in vitro*.

Based on our results presented above, a biosynthetic scenario for the anthranilate moiety with on-line hydroxylation and subsequent methylation appears the most plausible one. The *in vitro*

reconstitution of the NRPS TomA and TomB and the intact protein MS experiments excluded 4,5-dihydroxy anthranilic acid as NRPS substrate. Furthermore, methylation of a 5-hydroxy anthranilic acid moiety without prior hydroxylation during the assembly or after product release could be dismissed by a combined assay of TomA, TomB and TomG and the synthesis of the respective tomaymycin derivatives for the utilization in the TomG *in vitro* assay, respectively.⁶⁶ However, since TomA also accepts 4-hydroxy-5-methoxy anthranilic acid as substrate, tailoring prior to NRPS assembling cannot be completely ruled out. Thus, although the on-line scenario is in best accordance with the experimental data, a mixture of both scenarios might be observed *in vivo*.

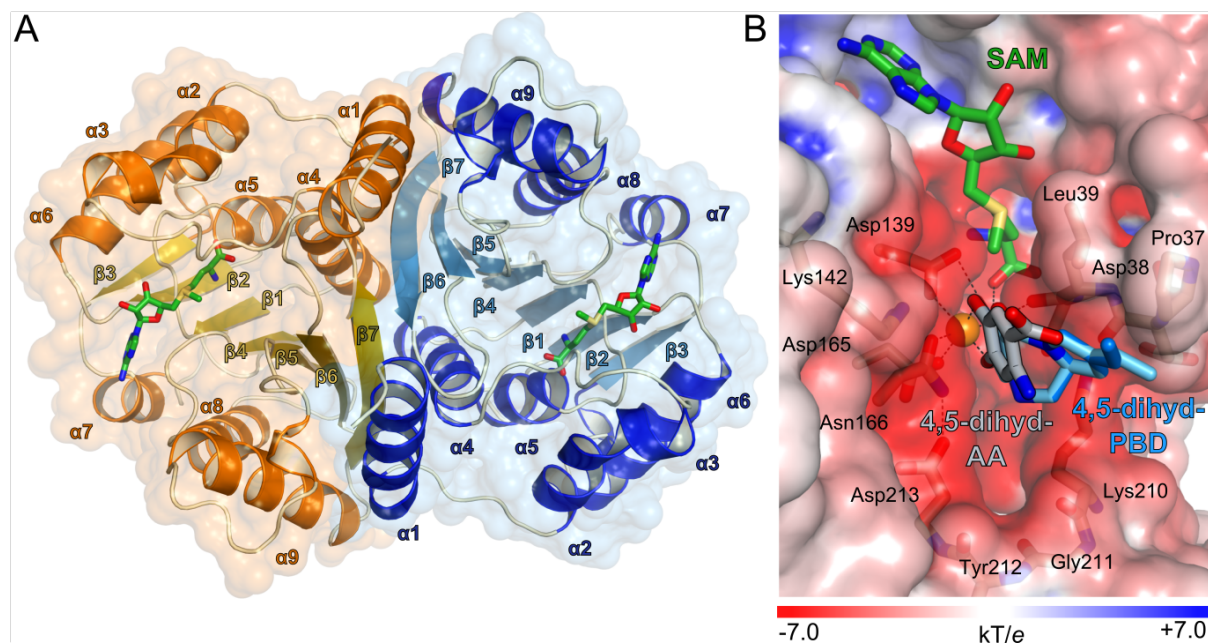


Figure 2.5: **A** Overview of the dimeric crystal structure of TomG (monomer A: yellow and orange; monomer B: blue) with the bound cofactor S-adenosyl methionine (SAM) (stick presentation in green); **B** Proposed binding modes of 4,5-dihydroxy anthranilic acid (4,5-dihyd-AA; in light grey) and 4,5-dihydroxy-PBD (4,5-dihyd-PBD; two binding modes in stick presentation in light and dark blue) to the TomG monomer (Mg²⁺ ion shown as orange sphere; molecular surface colored by electrostatic potential); Figures were prepared using PYMOL (The PyMOL Molecular Graphics System, Version 1.8.2.3 Schrödinger, LLC). Secondary structure elements were assigned with the aid of the STRIDE webserver.⁷⁰ Electrostatic potential was calculated using the APBS 2.1 plugin⁷¹ under PYMOL and the PDB2PQR webserver (http://nbc-222.ucsd.edu/pdb2pqr_2.0.0/)⁷²

2.3 Significance

We were able to establish a generally applicable intact protein LC-MS approach, allowing the direct detection of NRPS proteins with masses up to 170 kDa and high mass accuracy. Besides the validation of the protein sizes and the presence of the PPant moiety, comprehensive substrate loading experiments including competitive loading assays were performed. The results showed that TomA is able to load ten out of fourteen anthranilic acid derivatives and prefers 5-hydroxy anthranilic acid. In addition, TomB loading with methylene proline and subsequent transfer of activated anthranilic acid from TomA to form the dipeptidic intermediate could be unambiguously detected. We were thereby able to verify virtually all chemical steps involved in the biosynthesis of the nonribosomally synthesized natural product tomaymycin by LC-MS

measurements. In vitro reconstitution and the X-ray structure of the OMT TomG combined with the extensive analysis of the NRPS led to insights towards the timing of tailoring reactions in tomaymycin biosynthesis. Previous reports postulated two possible methylation time points, either before or after NRPS templated PBD assembly.¹¹ Completion of the tailoring reaction prior NRPS processing could not be completely ruled out but seems unlikely due to poor activity on 4-hydroxy-5-methoxy anthranilic acid by TomA. Our results strongly favor the following reaction sequence: The starting species 5-hydroxy anthranilic acid is loaded by TomA and firstly hydroxylated by TomE and TomF to be subsequently methylated by TomG, before downstream processing by TomB and maturing of tomaymycin takes place.

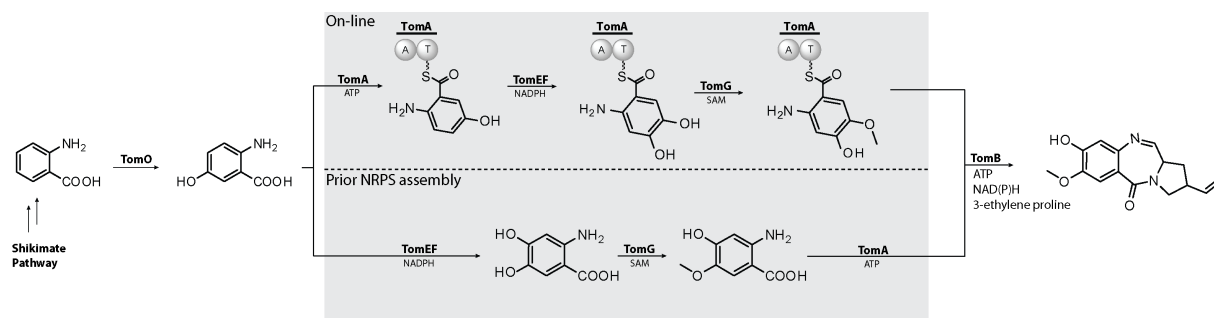


Figure 2.6: Tomaymycin biosynthesis comprising hydroxylation of 5-OH anthranilic acid by TomE and TomF, methylation by TomG and subsequent PBD assembly by TomAB. According to our data tailoring of anthranilic acid likely takes place either on-line while tethered to TomA or prior NRPS assembly.

Acknowledgement

This work was generously supported by a grant from the German Federal Ministry of Education and Research to Saarland University and Sanofi Aventis (FKZ: 0315198). We thank Kirsten Harmrolfs and Thomas Hoffmann for discussion and critical reading of the manuscript. We thank Silke Herok-Schoepper for technical assistance. We also kindly thank Christian Feiler for providing the expression vector for TomG for crystallization experiments. Additionally, parts of this research were carried out at the PETRA III at DESY, a member of the Helmholtz Association (HGF). We would thus also like to thank the beamline staff for assistance in using beamline P11.

2.4 Bibliography

- [1] David J. Newman and Gordon M. Cragg. Natural products as sources of new drugs from 1981 to 2014. *Journal of natural products*, 79(3):629–661, 2016.
- [2] K. Arima, M. Kosaka, G. Tamura, H. Imanaka, and H. Sakai. Studies on tomaymycin, a new antibiotic. i. isolation and properties of tomaymycin. *The Journal of antibiotics*, 25(8):437–444, 1972.
- [3] Barbara Gerratana. Biosynthesis, synthesis, and biological activities of pyrrolbenzodiazepines. *Medicinal research reviews*, 32(2):254–293, 2012.
- [4] M. S. Puvvada, S. A. Forrow, J. A. Hartley, P. Stephenson, I. Gibson, T. C. Jenkins, and D. E. Thurston. Inhibition of bacteriophage t7 rna polymerase in vitro transcription by dna-binding pyrrolo2,1-c1,4benzodiazepines. *Biochemistry*, 36(9):2478–2484, 1997.
- [5] L. H. Hurley, T. Reck, D. E. Thurston, D. R. Langley, K. G. Holden, R. P. Hertzberg, J. R. Hoover, G. Gallagher, JR, L. F. Faucette, and S. M. Mong. Pyrrolo1,4benzodiazepine antitumor antibiotics: relationship of dna alkylation and sequence specificity to the biological activity of natural and synthetic compounds. *Chemical research in toxicology*, 1(5):258–268, 1988.
- [6] D. E. Thurston, D. S. Bose, P. W. Howard, T. C. Jenkins, A. Leoni, P. G. Baraldi, A. Guiotto, B. Cacciari, L. R. Kelland, M. P. Foloppe, and S. Rault. Effect of a-ring modifications on the dna-binding behavior and cytotoxicity of pyrrolo2,1-c1,4benzodiazepines. *Journal of medicinal chemistry*, 42(11):1951–1964, 1999.
- [7] S. J. Gregson, P. W. Howard, K. E. Corcoran, S. Barcella, M. M. Yasin, A. A. Hurst, T. C. Jenkins, L. R. Kelland, and D. E. Thurston. Effect of c2-exo unsaturation on the cytotoxicity and dna-binding reactivity of pyrrolo2,1-c1,4benzodiazepines. *Bioorganic & medicinal chemistry letters*, 10(16):1845–1847, 2000.
- [8] Rohtash Kumar and J. William Lown. Recent developments in novel pyrrolo2,1-c1,4benzodiazepine conjugates: synthesis and biological evaluation. *Mini reviews in medicinal chemistry*, 3(4):323–339, 2003.
- [9] Yunfeng Hu, Vanessa Phelan, Ioanna Ntai, Chris M. Farnet, Emmanuel Zazopoulos, and Brian O. Bachmann. Benzodiazepine biosynthesis in streptomyces refuineus. *Chemistry & biology*, 14(6):691–701, 2007.
- [10] Wei Li, Ankush Khullar, ShenChieh Chou, Ashley Sacramo, and Barbara Gerratana. Biosynthesis of sibiromycin, a potent antitumor antibiotic. *Applied and environmental microbiology*, 75(9):2869–2878, 2009.
- [11] Wei Li, ShenChieh Chou, Ankush Khullar, and Barbara Gerratana. Cloning and characterization of the biosynthetic gene cluster for tomaymycin, an sjg-136 monomeric analog. *Applied and environmental microbiology*, 75(9):2958–2963, 2009.

-
- [12] Gene H. Hur, Christopher R. Vickery, and Michael D. Burkart. Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Natural product reports*, 29(10):1074–1098, 2012.
- [13] Robert Finking and Mohamed A. Marahiel. Biosynthesis of nonribosomal peptides1. *Annual review of microbiology*, 58:453–488, 2004.
- [14] Michael A. Fischbach and Christopher T. Walsh. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chemical reviews*, 106(8):3468–3496, 2006.
- [15] Elizabeth S. Sattely, Michael A. Fischbach, and Christopher T. Walsh. Total biosynthesis: in vitro reconstitution of polyketide and nonribosomal peptide pathways. *Natural product reports*, 25(4):757–793, 2008.
- [16] M. A. Wyatt, M. C. Mok, M. Junop, and N. A. Magarvey. Heterologous expression and structural characterisation of a pyrazinone natural product assembly line. *Chembiochem : a European journal of chemical biology*, 13(16):2408–2415, 2012.
- [17] D. J. Wilson, C. Shi, Am Teitelbaum, Am Gulick, and C. C. Aldrich. Characterization of ausa: a dimodular nonribosomal peptide synthetase responsible for the production of aureusimine pyrazinones. *Biochemistry*, 52(5):926–937, 2013.
- [18] Sven C. Feifel, Timo Schmiederer, Till Hornbogen, Holger Berg, Roderich D. Sussmuth, and Rainer Zocher. In vitro synthesis of new enniatins: probing the alpha-d-hydroxy carboxylic acid binding pocket of the multienzyme enniatin synthetase. *Chembiochem : a European journal of chemical biology*, 8(15):1767–1770, 2007.
- [19] Jane Muller, Sven C. Feifel, Timo Schmiederer, Rainer Zocher, and Roderich D. Sussmuth. In vitro synthesis of new cyclodepsipeptides of the pf1022-type: probing the alpha-d-hydroxy acid tolerance of pf1022 synthetase. *Chembiochem : a European journal of chemical biology*, 10(2):323–328, 2009.
- [20] Wenjun Zhang, Ioanna Ntai, Megan L. Bolla, Steven J. Malcolmson, Daniel Kahne, Neil L. Kelleher, and Christopher T. Walsh. Nine enzymes are required for assembly of the pacidamycin group of peptidyl nucleoside antibiotics. *Journal of the American Chemical Society*, 133(14):5240–5243, 2011.
- [21] Moriah Sandy, Zhe Rui, Joe Gallagher, and Wenjun Zhang. Enzymatic synthesis of dilactone scaffold of antimycins. *ACS chemical biology*, 7(12):1956–1961, 2012.
- [22] Diana Matthes, Lennart Richter, Jane Muller, Alexander Denisiuk, Sven C. Feifel, Yuquan Xu, Patricia Espinosa-Artiles, Roderich D. Sussmuth, and Istvan Molnar. In vitro chemoenzymatic and in vivo biocatalytic syntheses of new beauvericin analogues. *Chemical communications (Cambridge, England)*, 48(45):5674–5676, 2012.
- [23] A. M. Gehring, I. Mori, and C. T. Walsh. Reconstitution and characterization of the escherichia coli enterobactin synthetase from entb, ente, and entf. *Biochemistry*, 37(8):2648–2659, 1998.

- [24] Deborah Ann Miller, Lusong Luo, Nathan Hillson, Thomas A. Keating, and Christopher T. Walsh. Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of yersinia pestis. *Chemistry & biology*, 9(3):333–344, 2002.
- [25] H. M. Patel and C. T. Walsh. In vitro reconstitution of the pseudomonas aeruginosa non-ribosomal peptide synthesis of pyochelin: characterization of backbone tailoring thiazoline reductase and n-methyltransferase activities. *Biochemistry*, 40(30):9023–9031, 2001.
- [26] T. A. Keating, C. G. Marshall, and C. T. Walsh. Reconstitution and characterization of the vibrio cholerae vibriobactin synthetase from vibb, vibe, vibf, and vibh. *Biochemistry*, 39(50):15522–15530, 2000.
- [27] N. Gaitatzis, B. Kunze, and R. Muller. In vitro reconstitution of the myxochelin biosynthetic machinery of stigmatella aurantiaca sg a15: Biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases. *Proceedings of the National Academy of Sciences of the United States of America*, 98(20):11136–11141, 2001.
- [28] Yanyan Li, Kira J. Weissman, and Rolf Müller. Myxochelin biosynthesis: direct evidence for two- and four-electron reduction of a carrier protein-bound thioester. *Journal of the American Chemical Society*, 130(24):7554–7555, 2008.
- [29] Jennifer Jaitzig, Jian Li, Roderich D. Sussmuth, and Peter Neubauer. Reconstituted biosynthesis of the nonribosomal macrolactone antibiotic valinomycin in escherichia coli. *ACS synthetic biology*, 3(7):432–438, 2014.
- [30] Fred W. McLafferty. A century of progress in molecular mass spectrometry. *Annual review of analytical chemistry (Palo Alto, Calif.)*, 4:1–22, 2011.
- [31] Joost Snijder, Rebecca J. Rose, David Veessler, John E. Johnson, and Albert J. R. Heck. Studying 18 mda virus assemblies with native mass spectrometry. *Angewandte Chemie (International ed. in English)*, 52(14):4020–4023, 2013.
- [32] Natalie J. Thompson, Sara Rosati, Rebecca J. Rose, and Albert J. R. Heck. The impact of mass spectrometry on the study of intact antibodies: from post-translational modifications to structural analysis. *Chemical communications (Cambridge, England)*, 49(6):538–548, 2013.
- [33] Timothy K. Toby, Luca Fornelli, and Neil L. Kelleher. Progress in top-down proteomics and the analysis of proteoforms. *Annual review of analytical chemistry (Palo Alto, Calif.)*, 9(1):499–519, 2016.
- [34] N. Gaitatzis, A. Hans, R. Müller, and S. Beyer. The mtaa gene of the myxothiazol biosynthetic gene cluster from stigmatella aurantiaca dw4/3-1 encodes a phosphopantetheinyl transferase that activates polyketide synthases and polypeptide synthetases. *Journal of biochemistry*, 129(1):119–124, 2001.
- [35] Daniel Wilson and Courtney Aldrich. A continuous kinetic assay for adenylation enzyme activity and inhibition. *Analytical biochemistry*, 404(1):56–63, 2010.

-
- [36] Carl J. Balibar, Frederic H. Vaillancourt, and Christopher T. Walsh. Generation of d amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. *Chemistry & biology*, 12(11):1189–1200, 2005.
- [37] Christian Rausch, Ilka Hoof, Tilmann Weber, Wolfgang Wohlleben, and Daniel H. Huson. Phylogenetic analysis of condensation domains in nrps sheds light on their functional evolution. *BMC evolutionary biology*, 7:78, 2007.
- [38] Nicole M. Gaudelli, Darcie H. Long, and Craig A. Townsend. beta-lactam formation by a non-ribosomal peptide synthetase during antibiotic biosynthesis. *Nature*, 520(7547):383–387, 2015.
- [39] Kristina Haslinger, Madeleine Peschke, Clara Brieke, Egle Maximowitsch, and Max J. Cryle. X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis. *Nature*, 521(7550):105–109, 2015.
- [40] T. Stachelhaus, H. D. Mootz, V. Bergendahl, and M. A. Marahiel. Peptide bond formation in nonribosomal peptide biosynthesis. catalytic role of the condensation domain. *The Journal of biological chemistry*, 273(35):22773–22781, 1998.
- [41] P. J. Belshaw, C. T. Walsh, and T. Stachelhaus. Aminoacyl-coas as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science (New York, N.Y.)*, 284(5413):486–489, 1999.
- [42] Sabine Meyer, Jan-Christoph Kehr, Andi Mainz, Daniel Dehm, Daniel Petras, Roderich D. Sussmuth, and Elke Dittmann. Biochemical dissection of the natural diversification of microcystin provides lessons for synthetic biology of nrps. *Cell chemical biology*, 23(4):462–471, 2016.
- [43] Matthew T. Mazur, Christopher T. Walsh, and Neil L. Kelleher. Site-specific observation of acyl intermediate processing in thiotemplate biosynthesis by fourier transform mass spectrometry: the polyketide module of yersiniabactin synthetase. *Biochemistry*, 42(46):13393–13400, 2003.
- [44] Leslie M. Hicks, Sarah E. O’Connor, Matthew T. Mazur, Christopher T. Walsh, and Neil L. Kelleher. Mass spectrometric interrogation of thioester-bound intermediates in the initial stages of epothilone biosynthesis. *Chemistry & biology*, 11(3):327–335, 2004.
- [45] Shaun M. McLoughlin and Neil L. Kelleher. Kinetic and regiospecific interrogation of covalent intermediates in the nonribosomal peptide synthesis of yersiniabactin. *Journal of the American Chemical Society*, 126(41):13265–13275, 2004.
- [46] Sylvie Garneau, Pieter C. Dorrestein, Neil L. Kelleher, and Christopher T. Walsh. Characterization of the formation of the pyrrole moiety during clorobiocin and coumermycin a1 biosynthesis. *Biochemistry*, 44(8):2770–2780, 2005.
- [47] Leah M. Miller, Matthew T. Mazur, Shaun M. McLoughlin, and Neil L. Kelleher. Parallel interrogation of covalent intermediates in the biosynthesis of gramicidin s using high-resolution mass spectrometry. *Protein science : a publication of the Protein Society*, 14(10):2702–2712, 2005.

- [48] Leslie M. Hicks, Matthew T. Mazur, Leah M. Miller, Pieter C. Dorrestein, Nathan A. Schnarr, Chaitan Khosla, and Neil L. Kelleher. Investigating nonribosomal peptide and polyketide biosynthesis by direct detection of intermediates on 70 kda polypeptides by using fourier-transform mass spectrometry. *Chembiochem : a European journal of chemical biology*, 7(6):904–907, 2006.
- [49] Jordan L. Meier, Anand D. Patel, Sherry Niessen, Michael Meehan, Roland Kersten, Jane Y. Yang, Michael Rothmann, Benjamin F. Cravatt, Pieter C. Dorrestein, Michael D. Burkart, and Vineet Bafna. Practical 4'-phosphopantetheine active site discovery from proteomic samples. *Journal of proteome research*, 10(1):320–329, 2011.
- [50] Dario Meluzzi, Wei Hao Zheng, Mary Hensler, Victor Nizet, and Pieter C. Dorrestein. Top-down mass spectrometry on low-resolution instruments: Characterization of phosphopantetheinylated carrier domains in polyketide and non-ribosomal biosynthetic pathways. *Bioorganic and Medicinal Chemistry Letters*, 18(10):3107–3111, 2008.
- [51] J. Mitchell Wells and Scott A. McLuckey. Collision-induced dissociation (cid) of peptides and proteins. In *Biological Mass Spectrometry*, volume 402 of *Methods in Enzymology*, pages 148–185. Elsevier, 2005.
- [52] Xuemei Han, Yueju Wang, Aaron Aslanian, Marshall Bern, Mathieu Lavalley-Adam, and John R. 3rd Yates. Sheathless capillary electrophoresis-tandem mass spectrometry for top-down characterization of pyrococcus furiosus proteins on a proteome scale. *Analytical chemistry*, 86(22):11006–11012, 2014.
- [53] Jennifer S. Brodbelt. Ion activation methods for peptides and proteins. *Analytical chemistry*, 88(1):30–51, 2016.
- [54] Jason M. Crawford, Paul M. Thomas, Jonathan R. Scheerer, Anna L. Vagstad, Neil L. Kelleher, and Craig A. Townsend. Deconstruction of iterative multidomain polyketide synthase function. *Science (New York, N.Y.)*, 320(5873):243–246, 2008.
- [55] Xun Sun, Hao Li, Jonas Alfermann, Henning D. Mootz, and Haw Yang. Kinetics profiling of gramicidin s synthetase a, a member of nonribosomal peptide synthetases. *Biochemistry*, 53(50):7983–7989, 2014.
- [56] Michael J. Tarry, Asfarul S. Haque, Khanh Huy Bui, and T. Martin Schmeing. X-ray crystallography and electron microscopy of cross- and multi-module nonribosomal peptide synthetase proteins reveal a flexible architecture. *Structure (London, England : 1993)*, 25(5):783–793.e4, 2017.
- [57] Hannes Hahne, Fiona Pachl, Benjamin Ruprecht, Stefan K. Maier, Susan Klaeger, Dominic Helm, Guillaume Medard, Matthias Wilm, Simone Lemeer, and Bernhard Kuster. Dmso enhances electrospray response, boosting sensitivity of proteomic experiments. *Nature methods*, 10(10):989–991, 2013.
- [58] R. W. Woodard, M. D. Tsai, H. G. Floss, P. A. Crooks, and J. K. Coward. Stereochemical course of the transmethylation catalyzed by catechol o-methyltransferase. *The Journal of biological chemistry*, 255(19):9124–9127, 1980.

-
- [59] G. Schluckebier, M. O’Gara, W. Saenger, and X. Cheng. Universal catalytic domain structure of adomet-dependent methyltransferases. *Journal of molecular biology*, 247(1):16–20, 1995.
- [60] Jennifer L. Martin and Fiona M. McMillan. Sam (dependent) i am: the s-adenosylmethionine-dependent methyltransferase fold. *Current opinion in structural biology*, 12(6):783–793, 2002.
- [61] David K. Liscombe, Gordon V. Louie, and Joseph P. Noel. Architectures, mechanisms and molecular evolution of natural product methyltransferases. *Natural product reports*, 29(10):1238–1250, 2012.
- [62] J. Axelrod and R. Tomchick. Enzymatic o-methylation of epinephrine and other catechols. *The Journal of biological chemistry*, 233(3):702–705, 1958.
- [63] P. T. Mannisto and S. Kaakkola. Catechol-o-methyltransferase (comt): biochemistry, molecular biology, pharmacology, and clinical efficacy of the new selective comt inhibitors. *Pharmacological reviews*, 51(4):593–628, 1999.
- [64] Masahiro Takeo, Masumi Murakami, Sanae Niihara, Kenta Yamamoto, Munehiro Nishimura, Dai-ichiro Kato, and Seiji Negoro. Mechanism of 4-nitrophenol oxidation in rhodococcus sp. strain pn1: characterization of the two-component 4-nitrophenol hydroxylase and regulation of its expression. *Journal of bacteriology*, 190(22):7367–7374, 2008.
- [65] Tobias W. Giessen, Femke I. Kraas, and Mohamed A. Marahiel. A four-enzyme pathway for 3,5-dihydroxy-4-methylanthranilic acid formation and incorporation into the antitumor antibiotic sibiromycin. *Biochemistry*, 50(25):5680–5692, 2011.
- [66] Jang-Hee Cho, Younghee Park, Joong-Hoon Ahn, Yoongho Lim, and Sangkee Rhee. Structural and functional insights into o-methyltransferase from bacillus cereus. *Journal of molecular biology*, 382(4):987–997, 2008.
- [67] Hyo Jung Lee, Bong Gyu Kim, and Joong Hoon Ahn. Molecular cloning and characterization of bacillus cereus o-methyltransferase. *Journal of Microbiology and Biotechnology*, 16(4):619–622, 2006.
- [68] K. Rutherford, I. Le Trong, R. E. Stenkamp, and W. W. Parson. Crystal structures of human 108v and 108m catechol o-methyltransferase. *Journal of molecular biology*, 380(1):120–130, 2008.
- [69] P. Lautala, I. Ulmanen, and J. Taskinen. Molecular mechanisms controlling the rate and specificity of catechol o-methylation by human soluble catechol o-methyltransferase. *Molecular pharmacology*, 59(2):393–402, 2001.
- [70] D. Frishman and P. Argos. Knowledge-based protein secondary structure assignment. *Proteins*, 23(4):566–579, 1995.
- [71] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*, 98(18):10037–10041, 2001.

- [72] Todd J. Dolinsky, Paul Czodrowski, Hui Li, Jens E. Nielsen, Jan H. Jensen, Gerhard Klebe, and Nathan A. Baker. Pdb2pqr: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic acids research*, 35(Web Server issue):W522–5, 2007.
- [73] Elisabeth Gasteiger, Christine Hoogland, Alexandre Gattiker, S'everine Duvaud, Marc R. Wilkins, Ron D. Appel, and Amos Bairoch. Protein identification and analysis tools on the expasy server. In John M. Walker, editor, *The proteomics protocols handbook*, pages 571–607. Humana Press, Totowa, N.J., 2005.
- [74] Anja Burkhardt, Tim Pakendorf, Bernd Reime, Jan Meyer, Pontus Fischer, Nicolas Stübe, Saravanan Panneerselvam, Olga Lorbeer, Karolina Stachnik, Martin Warmer, Philip Rödiger, Dennis Göries, and Alke Meents. Status of the crystallography beamlines at petra iii. *The European Physical Journal Plus*, 131(3), 2016.
- [75] Wolfgang Kabsch. Xds. *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 2):125–132, 2010.
- [76] Philip R. Evans and Garib N. Murshudov. How good are my data and what is the resolution? *Acta crystallographica. Section D, Biological crystallography*, 69(Pt 7):1204–1214, 2013.
- [77] Martyn D. Winn, Charles C. Ballard, Kevin D. Cowtan, Eleanor J. Dodson, Paul Emsley, Phil R. Evans, Ronan M. Keegan, Eugene B. Krissinel, Andrew G. W. Leslie, Airlie McCoy, Stuart J. McNicholas, Garib N. Murshudov, Navraj S. Pannu, Elizabeth A. Potterton, Harold R. Powell, Randy J. Read, Alexei Vagin, and Keith S. Wilson. Overview of the ccp4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography*, 67(Pt 4):235–242, 2011.
- [78] Airlie J. McCoy, Ralf W. Grosse-Kunstleve, Paul D. Adams, Martyn D. Winn, Laurent C. Storoni, and Randy J. Read. Phaser crystallographic software. *Journal of applied crystallography*, 40(Pt 4):658–674, 2007.
- [79] Jakub Grzegorz Kopycki, Milton T. Stubbs, Wolfgang Brandt, Martin Hagemann, Andrea Porzel, Jurgen Schmidt, Willibald Schliemann, Meinhart H. Zenk, and Thomas Vogt. Functional and structural characterization of a cation-dependent o-methyltransferase from the cyanobacterium *synechocystis* sp. strain pcc 6803. *The Journal of biological chemistry*, 283(30):20888–20896, 2008.
- [80] Jay Painter and Ethan A. Merritt. Optimal description of a protein structure in terms of multiple groups undergoing tls motion. *Acta crystallographica. Section D, Biological crystallography*, 62(Pt 4):439–450, 2006.
- [81] Pavel V. Afonine, Ralf W. Grosse-Kunstleve, Nathaniel Echols, Jeffrey J. Headd, Nigel W. Moriarty, Marat Mustyakimov, Thomas C. Terwilliger, Alexandre Urzhumtsev, Peter H. Zwart, and Paul D. Adams. Towards automated crystallographic structure refinement with phenix.refine. *Acta crystallographica. Section D, Biological crystallography*, 68(Pt 4):352–367, 2012.

- [82] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan. Features and development of coot. *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 4):486–501, 2010.
- [83] Vincent B. Chen, W. Bryan 3rd Arendall, Jeffrey J. Headd, Daniel A. Keedy, Robert M. Immormino, Gary J. Kapral, Laura W. Murray, Jane S. Richardson, and David C. Richardson. Molprobity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 1):12–21, 2010.
- [84] Paul D. Adams, Pavel V. Afonine, Gabor Bunkoczi, Vincent B. Chen, Ian W. Davis, Nathaniel Echols, Jeffrey J. Headd, Li-Wei Hung, Gary J. Kapral, Ralf W. Grosse-Kunstleve, Airlie J. McCoy, Nigel W. Moriarty, Robert Oeffner, Randy J. Read, David C. Richardson, Jane S. Richardson, Thomas C. Terwilliger, and Peter H. Zwart. Phenix: a comprehensive python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 2):213–221, 2010.
- [85] Liisa Holm and Paivi Rosenstrom. Dali server: conservation mapping in 3d. *Nucleic acids research*, 38(Web Server issue):W545–9, 2010.
- [86] Heidi L. Schubert, Robert M. Blumenthal, and Xiaodong Cheng. Many paths to methyl-transfer: a chronicle of convergence. *Trends in biochemical sciences*, 28(6):329–335, 2003.
- [87] Piotr Z. Kozbial and Arcady R. Mushegian. Natural history of s-adenosylmethionine-binding proteins. *BMC structural biology*, 5:19, 2005.
- [88] Xavier Robert and Patrice Gouet. Deciphering key features in protein structures with the new endsript server. *Nucleic acids research*, 42(Web Server issue):W320–4, 2014.

2.5 Experimental Procedures

Protein Preparation Procedure

Expression and Purification of TomA, TomB and TomA Adenylation domain

The respective gene was inserted into a pETM44 expression vector with an N-terminal 6xHis-MBP tag using according primers (see Key Source Table). The construct was coexpressed with MtaA in *E. coli* BL21 (DE3) cells (500 mL LB, 0.1 mM IPTG, 16 °C, 16h). The cell pellet was resuspended in lysis buffer (150 mM NaCl, 25 mM Tris, 40 mM Imidazole, pH 7.5), sonicated and centrifuged. The supernatant was loaded onto a gravity flow column containing NiNTA loaded sepharose, washed with lysis buffer and subsequently eluted in one step (250 mM Imidazole). The tag was cleaved off using HRV3C protease during overnight dialysis against SEC buffer (150 mM NaCl, 25 mM Tris, pH 7.5) at 4 °C, and removed by a second Ni-NTA chromatography step. After passing through a Superdex 200 16/60 pg column (GE Healthcare Life Sciences) the protein was concentrated using a 30 kDa cutoff filter and stored at -80 °C in 10 % glycerol. Protein purity was determined by SDS-PAGE. Protein concentration was determined spectrophotometrically upon determining the respective extinction coefficient from the amino acid sequence using the PROTPARAM webserver (<http://web.expasy.org/protparam/>).⁷³

Expression and Purification of TomG for biochemical experiments

The respective gene was inserted into pET28b with an N-terminal 6xHis tag using according primers (Table S 2.2). The construct was expressed in *E. coli* BL21 (DE3) (500 mL LB, 0.1 mM IPTG, 16 °C, 16h). The cell pellet was resuspended in lysis buffer, sonicated and centrifuged. The supernatant was loaded onto a gravity flow column containing Ni-NTA loaded sepharose, washed with lysis buffer and subsequently eluted in one step (250 mM Imidazole). 6xHis-fusion protein was passed through a Superdex 200 16/60 pg column in SEC buffer (150 mM NaCl, 25 mM Tris, pH 7.5), concentrated using a 10 kDa cutoff filter and stored at -80 °C in 10 % glycerol. Protein purity was determined by SDS-PAGE. Protein concentration was determined spectrophotometrically as described above.

Expression and Purification of recombinant TomG for crystallization experiments

The respective gene was inserted into a modified pET-19b plasmid (Feiler & Blankenfeldt, unpublished results) using according primers (table S 2.2), yielding a fusion protein with an N-terminal 6xHis-T7-lysozyme-tag and a HRV3C protease cleavage site. The fusion protein was expressed in *E. coli* BL21 (DE3) (2SOC-medium, 0.1 mM IPTG, 16 °C, 16 h). The cell pellet was resuspended in lysis buffer, homogenized and centrifuged. The supernatant was applied onto a 5 mL HisTrap HP chelating column (GE Healthcare Life Sciences) loaded with nickel sulfate on an KTA protein purification system (GE Healthcare Life Sciences). Protein was eluted using 500 mM imidazole and analogously to conditions described above, the N-terminal tag of TomG was cleaved off using HRV3C protease. To remove the tag, a second nickel affinity chromatography step was performed and the flow-through containing TomG was concentrated using a Vivaspin 6 10 kDa cutoff concentrator (GE Healthcare Life Sciences). As a polishing step, the protein was passed through a Superdex 75 16/60 pg column (GE Healthcare Life Sciences)

and fractions containing TomG were concentrated and dialyzed against 200 mM NaCl, 1 mM CaCl₂, 10 % glycerol, 50 mM bicine, pH 8.0. Protein purity was assessed by SDS-PAGE and protein concentration was determined spectrophotometrically as described above.

Activity Assays

Hydroxamate-MesG Assay

Adenylation activity of TomA A-domain was tested using the hydroxamate-MesG assay as described previously.³⁵ Assays in 100 μ L reaction buffer (75 mM Tris HCl, 150 mM NaCl, 10 mM MgCl₂, 0.5 mM TCEP, pH 7.5) containing TomA A domain (350 nM), 150 mM hydroxylamine, 0.1 U nucleoside phosphorylase, 0.04 pyrophosphatase, 0.25 mM MesG and 0.08 mM of the respective substrate were incubated at 24 °C in a 96-well plate (Greiner). Cleavage of MesG was monitored by measuring the absorption at 360 nm over 40 min.

In vitro Reconstitution Assay of TomA and TomB

TomA and TomB (0.5 μ M each) were incubated with the respective cofactors (0.5 mM ATP, 0.25 mM NADPH) and substrates (0.25 mM 4-methylene proline, anthranilic acid derivate) in a total of 100 μ L reaction buffer (150 mM NaCl, 50 mM Tris, 5 mM MgCl₂, pH 7.5) at 30 °C for 12h. The mixture was dried in a vacuum condenser and subsequently resuspended in 20 μ L water:acetonitrile (1:1), centrifuged and submitted to LC-MS analysis.

In vitro Reconstitution Assay of TomG

10 μ M TomG was incubated with 1 mM SAM, 5 mM MgCl₂ and 1 mM substrate in a total of 30 μ L reaction buffer for 2h at 37 °C. The mixture was dried in a vacuum condenser and subsequently resuspended in 20 μ L water:acetonitrile (1:1), centrifuged and submitted to LC-MS analysis.

MS-based Analysis

LC-MS of tomaymycin analogues

The measurements to detect all types of tomaymycin derivatives were performed on a Dionex Ultimate 3000 RSLC system using a BEH C18, 50 x 2.1 mm, 1.7 μ m dp column (Waters, Germany). Separation of 5 μ L sample was achieved by a linear gradient from (A) H₂O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 600 μ L/min and 45 °C. The gradient was initiated by a 0.5 min isocratic step at 5 % B, followed by an increase to 95 % B in 9 min to end up with a 2 min step at 95 % B before reequilibration under initial conditions. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 75 μ L/min before entering the solarix XR (7T) FT-ICR mass spectrometer (Bruker Daltonics, Germany) using the Apollo II ESI source. In the source region, the temperature was set to 200 °C, the capillary voltage was 4500 V, the dry-gas flow was 4.0 L/min and the nebulizer was set to 1.1 bar. After the generated ions passed the quadrupole with a low cutoff at 150 m/z they were trapped in the collision cell for 500 ms, if applicable precursors fragmented with 15 V, and finally transferred within 0.9 ms through the hexapole into the ICR cell. Captured ions were excited by applying

a frequency sweep from 100 to 1000 m/z and detected in broadband mode by acquiring a 122 ms transient. Transient length was intentionally low to ensure a sufficient number of points per peak.

LC-MS of Anthranilic Acid Derivatives

Analysis to detect anthranilic acid derivatives were performed on a Dionex Ultimate 3000 RSLC system using a Phenomenex Luna PFP, 100 x 2.0 mm, 3 μ m dp column. Separation of 1 μ L sample was achieved by a linear gradient with (A) H₂O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 600 μ L/min and 45 °C. The gradient was initiated by a 1 min isocratic step at 2 % B, followed by an increase to 95 % B in 9 min to end up with a 1.5 min step at 95 % B before reequilibration under the initial conditions. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 75 μ L/min before entering the solarix XR (7T) FT-ICR mass spectrometer (Bruker Daltonics, Germany) using the Apollo ESI source. In the source region, the temperature was set to 200 °C, the capillary voltage was 4500 V, the dry-gas flow was 4.0 L/min and the nebulizer was set to 1.1 bar. After the generated ions passed the quadrupole with a low cutoff at 150 m/z they were trapped in the collision cell for 100 ms and finally transferred within 1.0 ms through the hexapole into the ICR cell. Captured ions were excited by applying a frequency sweep from 100 to 1600 m/z and detected in broadband mode by acquiring a 489 ms transient.

LC-MS measurements on intact protein level

All ESI-MS-measurements of intact protein samples were performed on a Dionex Ultimate 3000 RSLC system using a ProSwift RP-4H (monolithic PS-DVB), 250 x 1 mm column (Thermo, USA). Separation of 1 μ L sample was achieved by a multistep gradient from (A) H₂O + 0.1 % FA + 1 % DMSO to (B) ACN + 0.1 % FA + 1 % DMSO at a flow rate of 200 μ L/min and 45 °C. The gradient was initiated by a 0.5 min isocratic step at 5 % B, followed by an increase to 65 % B in 18 min, followed by an increase to 98 % in 0.5 min to end up with a 3 min step at 98 % B before reequilibration with initial conditions. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 75 μ L/min before entering the maXis 4G hr-ToF mass spectrometer (Bruker Daltonics, Bremen, Germany) using the standard Bruker ESI source. In the source region, the temperature was set to 180 °C, the capillary voltage was 4000 V, the dry-gas flow was 6.0 l/min and the nebulizer was set to 1.1 bar. Mass spectra were acquired in positive ionization mode ranging from 600–1800 m/z at 2.5 Hz scan rate. Protein masses were deconvoluted by using the Maximum Entropy algorithm (Copyright 1991-2004 Spectrum Square Associates, Inc.).

Loading Assays for LC-MS with TomA, TomB, TomA and TomB

Substrate loading assays for TomA and B were performed by incubating mixtures of protein and substrate including the respective cofactors for 60 min at 30 °C. The 14 μ L assay mixtures contained 120 pmol of TomA or B, 500 pmol substrates, 500 pmol ATP and 200 nmol MgCl₂ in reaction buffer. After incubation, mixtures were directly submitted for intact protein analyses. To determine the intermediate on TomB during the assembly, TomA and TomB assay

mixtures were combined, also incubated for 60 min at 30 °C and submitted for LC-MS measurements. Reductive product release was achieved by adding 10 nmol NADPH to the leftovers from the combined protein assays and repeated incubation for 30 min. The resulted mixtures were measured on small molecule as well as intact protein level.

Protein Structure Determination and Modeling

Crystallization, X-ray data collection and structure determination of TomG

Vapor diffusion trials were set up in a 96-well sitting drop format at 19 °C using a dispensing robot (Zinsser Analytics). 0.1 μ L of the protein solution with a concentration of 5 mg/mL TomG containing 0.6 mM of S-adenosyl methionine (SAM) were mixed with an equal amount of reservoir solution to give concentrations listed in supplementary Table S1. The drops were equilibrated against 70 μ L of reservoir solution and crystals appeared after 5-8 days. Before flash cooling the crystals in liquid nitrogen, 10 % (v/v) of (2R,3R)-(-)-2,3-butanediol was added as cryoprotectant. X-ray data collection was carried out at 100 K on beamline P11 of the PETRA III Synchrotron (DESY, Hamburg, Germany).⁷⁴ The diffraction data were indexed, integrated and scaled with XDS⁷⁵ and AIMLESS⁷⁶ of the CCP4 suite.⁷⁷ The structure of TomG was solved by molecular replacement with PHASER⁷⁸ by using the coordinates of residues 44-220 of a methyltransferase from Cyanobacterium *Synechocystis* (pdb entry: 3CBG)⁷⁹ as search model (32 % sequence identity). The model was further improved by iterative cycles of translation liberation screw-motion (TLS)⁸⁰ refinement using PHENIX.REFINE⁸¹ and manual rebuilding with COOT.⁸² Final structure building and interpretation was aided by MOLPROBITY⁸³ as well as by feature enhanced maps calculated with PHENIX.^{81, 84} Relevant crystallographic parameters of the structures are listed in Table S1. The DALI webserver was used to analyze structural similarities and calculate Z-scores as well as sequence similarities to other proteins.⁸⁵ Figures were prepared using PyMOL (The PyMOL Molecular Graphics System, Version 1.8.2.3 Schrodinger, LLC).

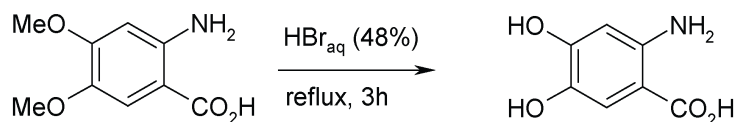
In-silico Modelling

The software MOE (Molecular Operating Environment (MOE, 2013.08; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2017) was used for *in-silico* modeling. Chain A of the TomG structure was superimposed on a model of human catechol-OMT (HsCOMT;⁶⁸ pdb code: 3bwm) using only the metal-coordinating residues with the program SUPERPOSE under CCP4.⁷⁷ Subsequently, coordinates of the ligand 3,5-dinitrocatechol (DNC), the Mg²⁺ ion as well as a coordinating water molecule were extracted and added to the TomG model. To ensure reasonable distances for the metal-coordinating atoms towards Mg²⁺, the coordinates of the active site residue Asp143 in TomG were further adjusted. Using DNC as scaffold, the ligands 4,5-dihydroxy anthranilic acid and unmethylated tomaymycin were built in low energy conformations that did not exhibit strong clashes with the protein. For unmethylated tomaymycin, low energy conformations were determined using a conformational search approach. For this purpose, ligand coordinates were extracted and the 4,5-dihydroxyphenyl moiety was kept fixed during calculations. Charges were assigned using the MMFF94x force field. Protons were added via the Protonate 3D option at pH 7.5, 310 K and a salt concentration of

0.15 M. A stochastic conformational search of the attached substituents was performed using default parameters together with a RMS gradient of 0.0001, a rejection limit of 1000, a RMSD limit of 0.25, and an energy window of 15. Planar systems were treated as rigid bodies. For the following energy minimization processes, the structures were prepared as follows: except for the coordinating water, all water molecules were excluded from the model and protons and charges were added as described above. The ligand and protein residues within an 8 Å sphere around the ligand were subjected to energy minimization in the MMFF94x force field. The atoms in the 8 Å sphere were restrained to their starting positions of the crystal structure. During minimization, the coordinates of the metal-coordinating atoms, Mg²⁺ and the hydroxyl oxygens of the ligands were kept fixed to ensure a productive binding mode. Supplementary information about the TomG crystal structure TomG was obtained as a homodimer which crystallizes at a concentration of 5 mg/mL with 0.6 mM of the cofactor S-adenosyl methionine (SAM) and 2.1 M NH₄(SO₄)₂ as precipitant in space group P2₁2₁2₁ and two molecules in the asymmetric unit. The crystal structure was determined by molecular replacement using the structure of a cation-dependent O-methyltransferase (OMT) from *Cyanobacterium synechocystis* exhibiting a sequence similarity of 33 % (SynOMT;⁷⁹ pdb code: 3cbg). As described in the main section, common features of SAM-dependent class I OMTs is a conserved SAM binding site and a S_N2-based transfer of the methyl group of SAM to the substrate in the presence of an Mg²⁺ ion.^{58–60, 86} Consequently, the active site architecture is conserved among OMTs that utilize this mechanism, and in TomG, the highly conserved residue Lys142 (Figure S 2.12) would be positioned to deprotonate the hydroxyl group of potential substrates. Furthermore, the conserved amino acids Asp139, Asp165, Asn166 together with a water molecule in the immediate vicinity are expected to octahedrally coordinate an Mg²⁺ ion (Figure 2.6), although no divalent ion was observed in the electron density for TomG. In proximity to the proposed metal binding site, additional electron density corresponding to SAM was identified in both chains of the structure. The cofactor is bound by a set of H-bonds to the side chain or backbone atoms of residues Val41, Gly65, Ser71, Glu89, Asp139, and water molecules. Furthermore, the adenine moiety forms hydrophobic interactions with residues Trp90, Ala118, and Ala140. Among these residues, Gly65, Ser71, Glu89, Asp139, Ala118 and Ala140 are highly conserved between related OMTs⁸⁷ (Figure S 2.12). Via these interactions, SAM is positioned such that its methyl group points towards the proposed active site of TomG.

Synthesis

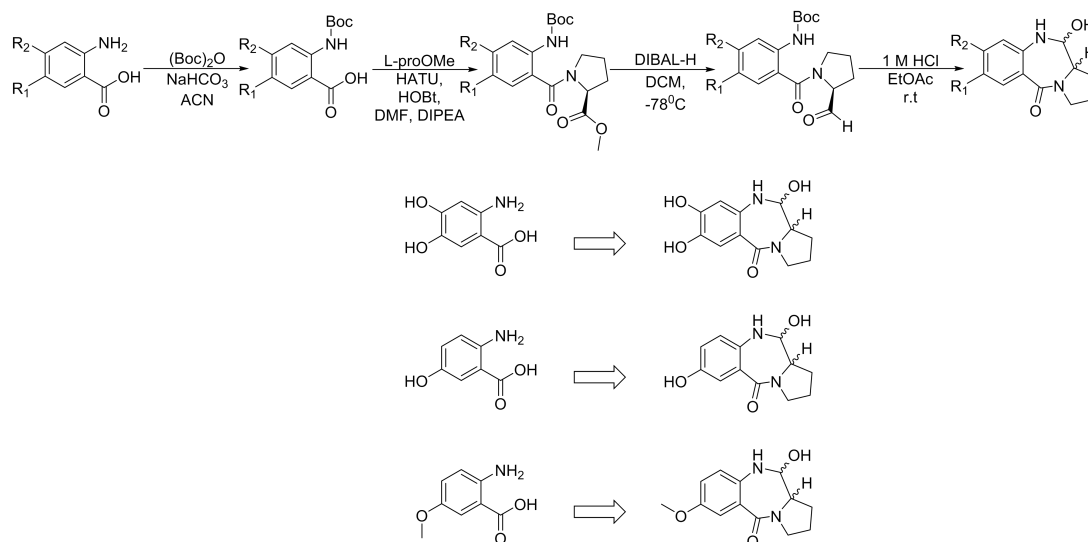
Synthesis of 2-Amino-4,5-dihydroxy-benzoic acid



2-Amino-4,5-dihydroxy-benzoic acid was prepared by ether cleavage from commercially available 2-amino-4,5-dimethoxy-benzoic acid under acidic conditions. A suspension of 2-amino-4,5-dimethoxy-benzoic acid (5.0 g, 24.85 mmol) in aqueous hydro-bromic acid (48 %, 127mL, 1.12 mol) was heated under reflux for 3h after which the mixture was concentrated under reduced pressure. After being cooled to room temperature the residue was triturated with diethyl ether.

Ice-cold water was added and the phases were separated. The aqueous phase was adjusted to pH5 by addition of saturated aqueous NaHCO₃ solution. The precipitate was filtered off, washed with water and dried in vacuo to yield 2.38 g (57 %) of 2-amino-4,5-dihydroxy-benzoic acid as a dark brown solid. ¹H NMR (600MHz, DMSO-d₆) 9.39 (s, 1H), 8.20 (s, 1H), 7.96 (br s, 3H), 7.06 (s, 1H), 6.12 (s, 1H); ¹³C NMR (150MHz, DMSO-d₆): 169.08, 152.05, 146.87, 135.48, 116.28, 102.09, 100.72 ; MS (ESI-) C₇H₆NO₄ [M-H]⁻ calculated for 168.0, found 168.0.

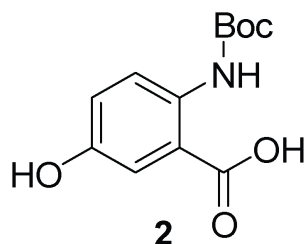
Synthesis of the tomaymycin derivatives



2-N-boc-amino-5-methoxybenzoic acid (1):

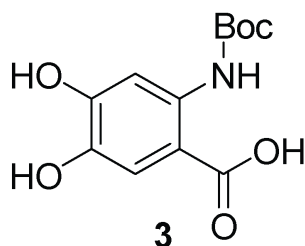
To a stirred solution of 2-amino-5-methoxybenzoic acid (50 mg, 0.3 mmol) in acetonitrile (20 ml) at room temperature, a mixed solution of di-tert-butyl dicarbonate (98 mg, 0.45 mmol), saturated solution of sodium bicarbonate (NaHCO₃) (19.6 ml) and acetonitrile (10 ml) was added. Following addition, the mixture left stirring for 15 h at room temperature. When no trace of starting material remained (TLC: silica on alumina, UV, ninhydrin, hexane : ethyl acetate :: 1:1 R_f 0.27), solvent was removed under reduced pressure resulting in white sticky foam that was partitioned between ethyl acetate (40 ml) and 0.1M HCl (30 ml). The organic layer was washed with 0.1M HCl (30 ml) five times. The organic extract was then dried over sodium sulfate, filtered off, and concentrated under reduced pressure to yield pale yellow oil (1). The compound was used without any further purification. HRMS (ESI, +ve) C₁₃H₁₇NO₅ [M+H]⁺ calculated for 268.1179, found 268.1168.

2-N-boc-amino-5-hydroxybenzoic acid (2):



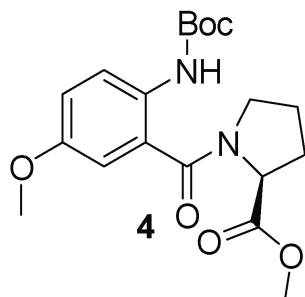
2-N-boc-amino-5-hydroxy benzoic acid (2) was prepared following the same protocol as for compound (1) and was used without any further purification. HRMS (ESI, +ve) C₁₂H₁₅NO₅ [M+H]⁺ calculated for 254.1023, found 254.1027.

2-N-boc-amino-4,5-dihydroxybenzoic acid (3):



2-N-boc-amino-4,5-dihydroxybenzoic acid (3) was prepared following the same protocol as for compound (1) and was used without any further purification. HRMS (ESI, +ve) C₁₂H₁₅NO₆ [M+H]⁺ calculated for 270.0972, found 270.0967.

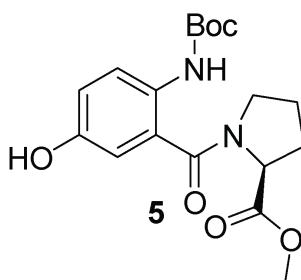
2-N-boc-amino-5-methoxybenzoyl- L-proline methyl ester (4):



2-N-boc-amino-5-methoxy benzoic acid (1) (50 mg, 0.18 mmol), HATU (68.4 mg, 0.18 mmol) and HOBt (5 mg, 0.03 mmol) was dissolved in anhydrous DMF (1 ml) under nitrogen atmosphere. DIPEA (61 μ L, 0.36 mmol) was added to the mixture and the solution was left stirring at room temperature for 5 minutes under nitrogen atmosphere. A mixture of L-proline methyl ester hydrochloride (30 mg, 0.18 mmol) and DIPEA (61 μ L, 0.36 mmol) in anhydrous DMF (1

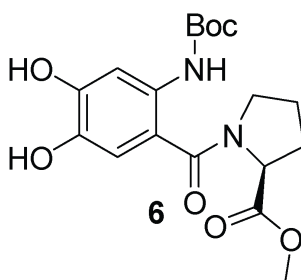
ml) was added to the latter solution and the reaction was left stirring for 16 hours under nitrogen atmosphere at room temperature. The reaction mixture was then diluted in ethyl acetate (80 ml) and 0.1M HCl (30 ml). The organic layer was washed with 0.1M HCl (30 ml) five times. The organic extract was then dried over sodium sulfate, filtered off, concentrated under reduced pressure, and purified using flash chromatography (silica, ninhydrin/UV, hexane: ethyl acetate:: 4:1; 2:1) to afford 2-N-boc-amino-5-methoxy benzoyl- L-proline methyl ester (4) as a light yellow powder. ¹H NMR (500 MHz, CDCl₃) 8.51 (br, 1H), 7.77 (d, J = 7.9 Hz, 1H), 7.19 (d, J = 2.6 Hz, 1H), 7.08 (dd, J = 7.8, 2.5 Hz, 1H), 4.67 (dd, J = 8.5, 5.3 Hz, 1H), 3.79 (s, 3H), 3.72 (s, 3H), 3.61-3.49 (m, 2H), 2.37-2.28(m, 1H), 2.07-1.97 (m, 3H), 1.92-1.88(m, 1H), 1.52(s, 9H); ¹³C NMR (126 MHz, CDCl₃) 172.4, 168.4, 155.5, 153.1, 135.5, 125.1, 122.9, 119.4, 113.2, 79.8, 68.3, 58.1, 55.1, 52.1, 31.7, 28.5, 24.2 ; HRMS (ESI, +ve) C₁₉H₂₆N₂O₆ [M+H]⁺ calculated for 379.1864, found 379.1860.

2-N-boc-amino-5-hydroxybenzoyl- L-proline methyl ester (5):



2-N-boc-amino-5-hydroxybenzoyl- L-proline methyl ester (5) was prepared following the same protocol as for compound (4). ¹H NMR (500 MHz, CDCl₃) 9.20 (br, 1H), 8.48 (br, 1H), 7.81 (d, J = 8.3 Hz, 1H), 7.02 (d, J = 2.9 Hz, 1H), 6.82 (dd, J = 8.2, 2.7 Hz, 1H), 4.66 (dd, J = 8.3, 5.2 Hz, 1H), 3.73 (s, 3H), 3.60-3.47 (m, 2H), 2.37-2.28(m, 1H), 2.09-1.94 (m, 3H), 1.92-1.88(m, 1H), 1.52(s, 9H); ¹³C NMR (126 MHz, CDCl₃) 172.8, 168.7, 153.4, 151.9, 133.1, 125.1, 123.2, 122.4, 118.2, 79.2, 68.6, 57.8, 52.4, 31.5, 28.2, 24.5 ; HRMS (ESI, +ve) C₁₈H₂₄N₂O₆ [M+H]⁺ calculated for 365.1707, found 365.1699.

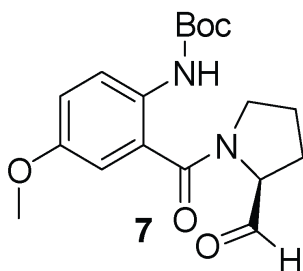
2-N-boc-amino-4,5-dihydroxybenzoyl- L-proline methyl ester (6):



2-N-boc-amino-4,5-dihydroxybenzoyl- L-proline methyl ester (6) was prepared following the

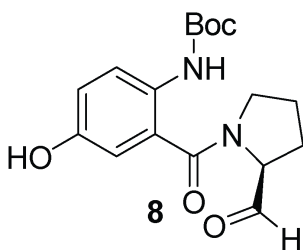
same protocol as for compound (4). ^1H NMR (500 MHz, CDCl_3) 9.18 (br, 1H), 8.58 (br, 1H), 7.81 (s, 1H), 7.02 (s, 1H), 4.67 (dd, $J = 8.6, 5.3$ Hz, 1H), 3.73 (s, 3H), 3.61-3.49 (m, 2H), 2.37-2.28(m, 1H), 2.07-1.97 (m, 3H), 1.92-1.88(m, 1H), 1.52(s, 9H); ^{13}C NMR (126 MHz, CDCl_3) 172.8, 168.7, 155.4, 151.9, 141.3, 133.1, 125.1, 122.4, 118.2, 79.2, 68.6, 57.8, 52.4, 31.5, 28.2, 24.5 ; HRMS (ESI, +ve) $\text{C}_{18}\text{H}_{24}\text{N}_2\text{O}_7$ $[\text{M}+\text{H}]^+$ calculated for 381.1656, found 381.1652.

S-1- (2-N-boc-amino-5-methoxybenzoyl)pyrrolidine-2-carbaldehyde (7):



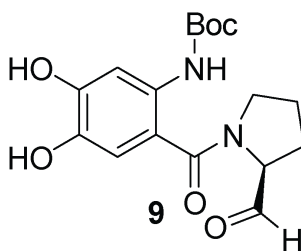
To a stirred solution of 2-N-boc-amino-5-methoxybenzoyl- L-proline methyl ester (4) (30 mg, 0.08 mmol) in anhydrous ether (10 ml) at -78°C , a solution of 1 M DIBAL hydride in hexane (70 μl , 0.08 mmol) was added over a period of 1 h under an atmosphere of nitrogen. The mixture was left stirring at -78°C for 5 h under an atmosphere of nitrogen. When no trace of starting material remained (TLC: silica, ninhydrin/UV, hexane : ethyl acetate:: 3:1, R_f 0.45), the reaction was quenched with cold methanol (1 ml). A white precipitate was formed. Reaction mixture was diluted with ethyl acetate (50 ml) and washed with saturated solution of sodium bicarbonate (20 ml) three times. The organic layer was then dried over sodium sulfate, filtered off, concentrated under reduced pressure and purified using flash chromatography (silica, ninhydrin/UV, hexane: ethyl acetate:: 5:1; 4:1) to afford S-1- (2-N-boc-amino-5-methoxybenzoyl)pyrrolidine-2-carbaldehyde (7). ^1H NMR (500 MHz, CDCl_3) 9.62 (s, 1H), 8.50 (br, 1H), 7.78 (d, $J = 8.1$ Hz, 1H), 7.17 (d, $J = 2.8$ Hz, 1H), 7.10 (dd, $J = 8.0, 2.7$ Hz, 1H), 4.28 (m, 1H), 3.79 (s, 3H), 3.71-3.50 (m, 2H), 2.45-2.30(m, 1H), 2.25-2.1 (m, 3H), 2.04-1.95(m, 1H), 1.52(s, 9H),; ^{13}C NMR (126 MHz, CDCl_3) 198.2, 168.6, 155.6, 153.4, 135.2, 124.9, 123.7, 119.1, 113.5, 80.1, 66.3, 55.6, 51.3, 31.9, 28.3, 24.2; HRMS (ESI, +ve) calculated for $\text{C}_{18}\text{H}_{24}\text{N}_2\text{O}_5$ m/z $[\text{M}+\text{H}]^+$: 349.1758; found: 349.1746.

S-1- (2-N-boc-amino-5-hydroxybenzoyl)pyrrolidine-2-carbaldehyde (8):



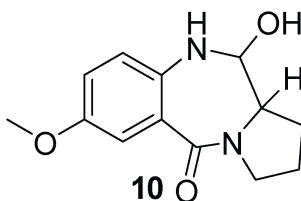
S-1- (2-N-boc-amino-5-hydroxybenzoyl)pyrrolidine-2-carbaldehyde (8) was prepared following the same protocol as for compound (7). ¹H NMR (500 MHz, CDCl₃) 9.64 (s, 1H), 9.22 (br, 1H), 8.51 (br, 1H), 7.79 (d, J = 8.0 Hz, 1H), 7.12 (d, J = 3.0 Hz, 1H), 6.85 (dd, J = 8.1, 3 Hz, 1H), 4.31 (m, 1H), 3.68-3.47 (m, 2H), 2.47-2.35(m, 1H), 2.29-2.11 (m, 3H), 2.01-1.91(m, 1H), 1.54(s, 9H); ¹³C NMR (126 MHz, CDCl₃) 198.7, 168.1, 154.1, 152.3, 132.5, 125.1, 123.4, 122.2, 118.2, 80.0, 66.5, 52.4, 31.3, 28.1, 24.5 ; HRMS (ESI, +ve) C₁₇H₂₂N₂O₅ [M+H]⁺ calculated for 335.1605, found 335.1609.

S-1- (2-N-boc-amino-4,5-dihydroxybenzoyl)pyrrolidine-2-carbaldehyde (9):



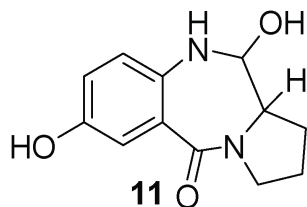
S-1- (2-N-boc-amino-4,5-dihydroxybenzoyl)pyrrolidine-2-carbaldehyde (9) was prepared following the same protocol as for compound (7). ¹H NMR (500 MHz, CDCl₃) 9.63 (s, 1H), 9.2 (br, 1H), 8.58 (br, 1H), 7.84 (s, 1H), 7.1 (s, 1H), 4.3 (m, 1H), 3.65-3.49 (m, 2H), 2.51-2.31(m, 1H), 2.28-1.88 (m, 4H), 1.52(s, 9H); ¹³C NMR (126 MHz, CDCl₃) 198.6, 168.3, 155.7, 152.1, 141.3, 136.1, 119.7, 115.4, 110.2, 80.1, 68.1, 52.4, 31.6, 26.2, 24.5 ; HRMS (ESI, +ve) C₁₇H₂₂N₂O₆ [M+H]⁺ calculated for 351.1551, found 351.1546.

Carbinolamine from 5-methoxy anthranilic acid (10):



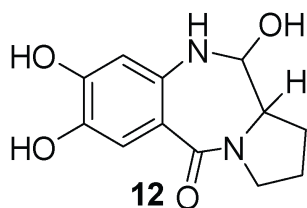
S-1- (2-N-boc-amino-5-methoxybenzoyl)pyrrolidine-2-carbaldehyde (7) (10 mg, 0.03 mmol) was dissolved in ethyl acetate (2 ml) and 1 M HCl (2 ml) was added while stirring vigorously. The mixture was left stirring at high speed for 16 hours at room temperature. The reaction mixture was then concentrated under reduced pressure and lyophilized to yield a yellow powder (10). ¹H NMR (500 MHz, CD₃OD) 7.33-7.21 (m, 1H), 6.93-6.91 (m, 2H), 4.95 (dd, J = 6.3, 2 Hz, 1H), 4.52-4.44 (m, 2H), 3.82 (s, 3H), 2.30-2.24(m, 1H), 2.15-2.06 (m, 1H), 1.95-1.88(m, 2H); ¹³C NMR (126 MHz, CD₃OD) 165.6, 154.3, 132.2, 124.1, 122.6, 119.1, 112.8, 84.1, 60.8, 53.6, 50.3, 33.9, 22.3; HRMS (ESI, +ve) calculated for C₁₃H₁₆N₂O₃ m/z [M+H]⁺: 249.1234; found: 249.1239.

Carbinolamine from 5-hydroxy anthranilic acid (11):



Carbinolamin X (11) was prepared following the same protocol as for compound (10). ¹H NMR (500 MHz, CD₃OD) 7.42-7.38 (m, 1H), 7.13-6.94 (m, 2H), 4.93 (dd, J = 6.7, 2.3 Hz, 1H), 4.54-4.45 (m, 2H), 2.32-2.26(m, 1H), 2.19-2.08 (m, 1H), 1.95-1.88(m, 2H); ¹³C NMR (126 MHz, CD₃OD) 165.3, 152.3, 132.1, 122.9, 122.2, 118.7, 113.9, 84.9, 60.8, 50.3, 33.7, 22.2; HRMS (ESI, +ve) calculated for C₁₂H₁₄N₂O₂ m/z [M+H]⁺: 235.1077; found: 235.1080.

Carbinolamine from 4,5-dihydroxy anthranilic acid (12):



Carbinolamin X (12) was prepared following the same protocol as for compound (10). ¹H NMR (500 MHz, CD₃OD) 7.2 (s, 1H), 6.93 (s, 1H), 4.96 (dd, J = 6.7, 2.3 Hz, 1H), 4.56-4.49 (m, 2H), 2.33-2.24 (m, 1H), 2.19-2.08 (m, 1H), 1.95-1.86 (m, 2H); ¹³C NMR (126 MHz, CD₃OD) 165.6, 157.3, 140.5, 139.5, 118.5, 116.7, 103.3, 84.6, 61, 50.4, 33.9, 22.1; HRMS (ESI, +ve) calculated for C₁₂H₁₄N₂O₄ m/z [M+H]⁺: 251.1026; found: 251.1018.

2.6 Supporting Information

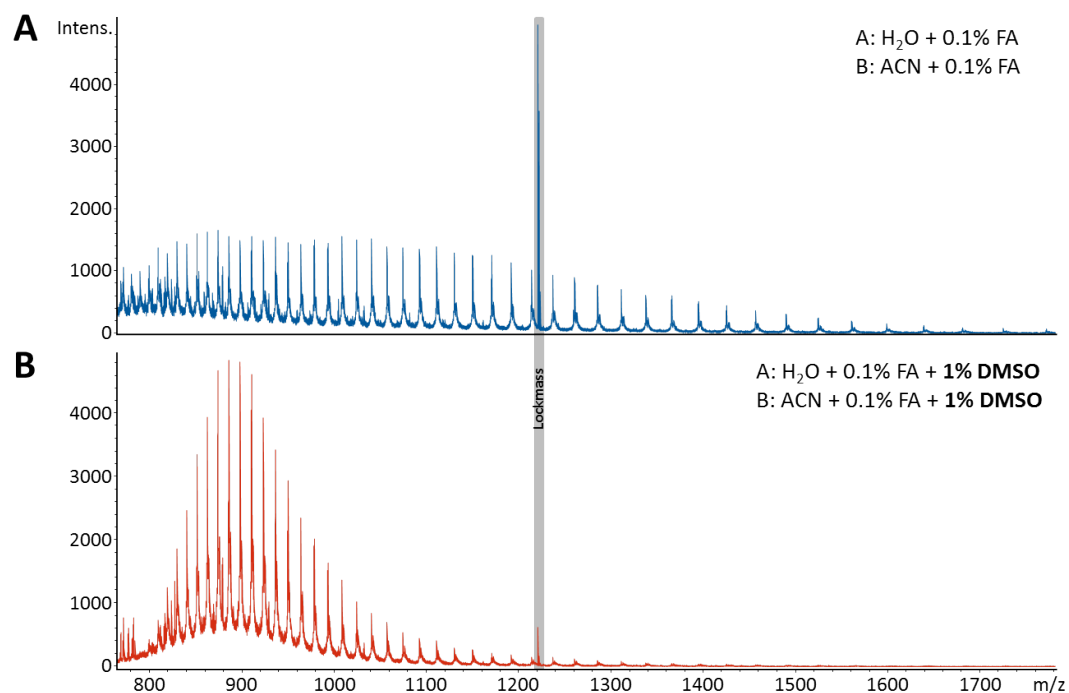


Figure 2.7: Comparative mass spectra of TomA show the influence on spectra quality owing to the addition of DMSO (**B**) to the standard LC eluents (**A**). Reduction of charge state signals leads to an intensity increase for the remaining signals by a factor of approx. 3.

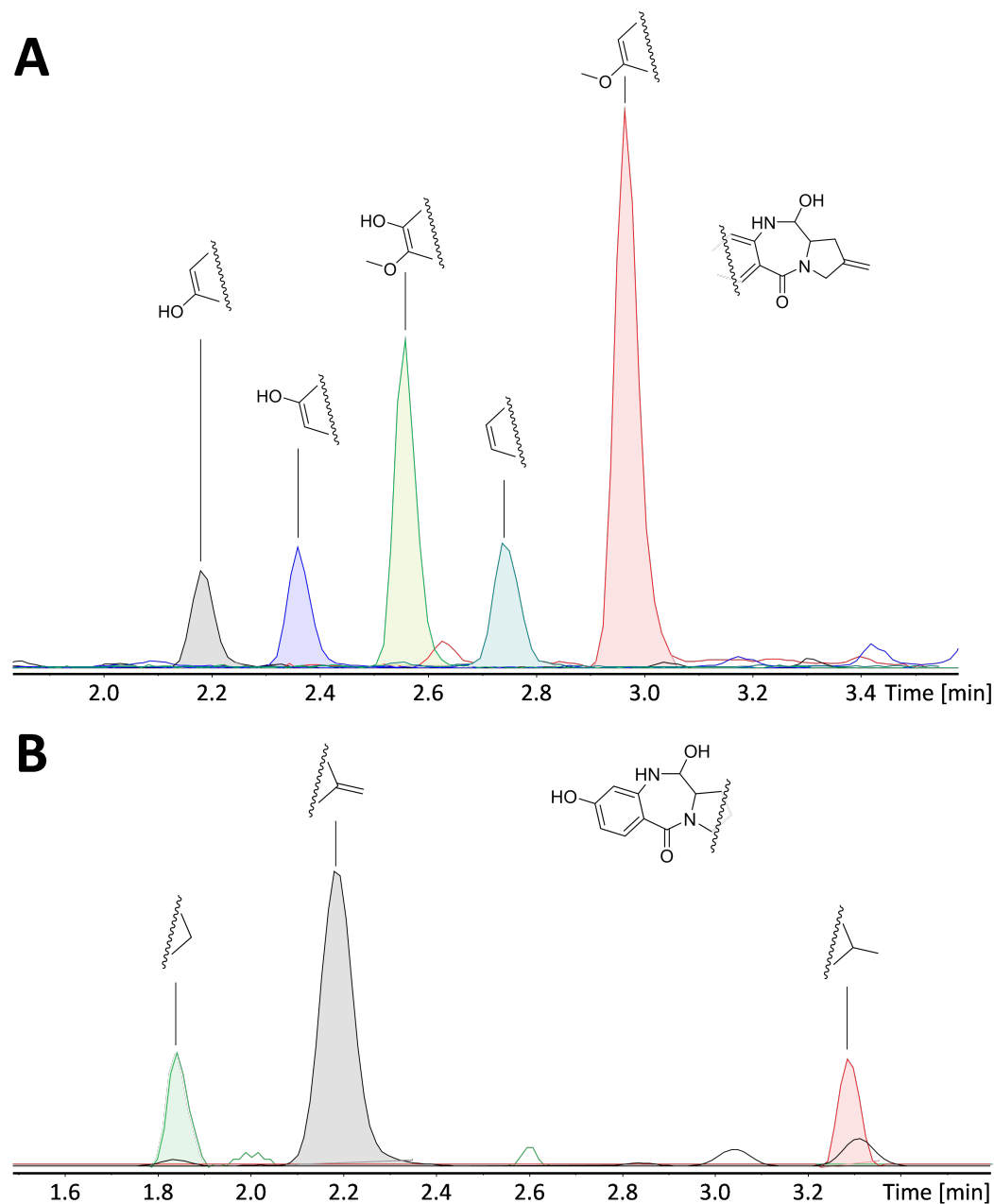


Figure 2.8: **A** LC-MS data of tomaymycin derivatives derived from in vitro reconstitution of TomA and TomB with different anthranilic acid derivatives combined with methylene proline. Overlaid EIC traces of the respective derivatives with a m/z width of 0.002. **B** LC-MS data of tomaymycin derivatives derived from in vitro reconstitution of TomA and TomB with different proline derivatives combined with 5-hydroxy anthranilic acid. Overlaid EIC traces of the respective derivatives with a m/z width of 0.002.

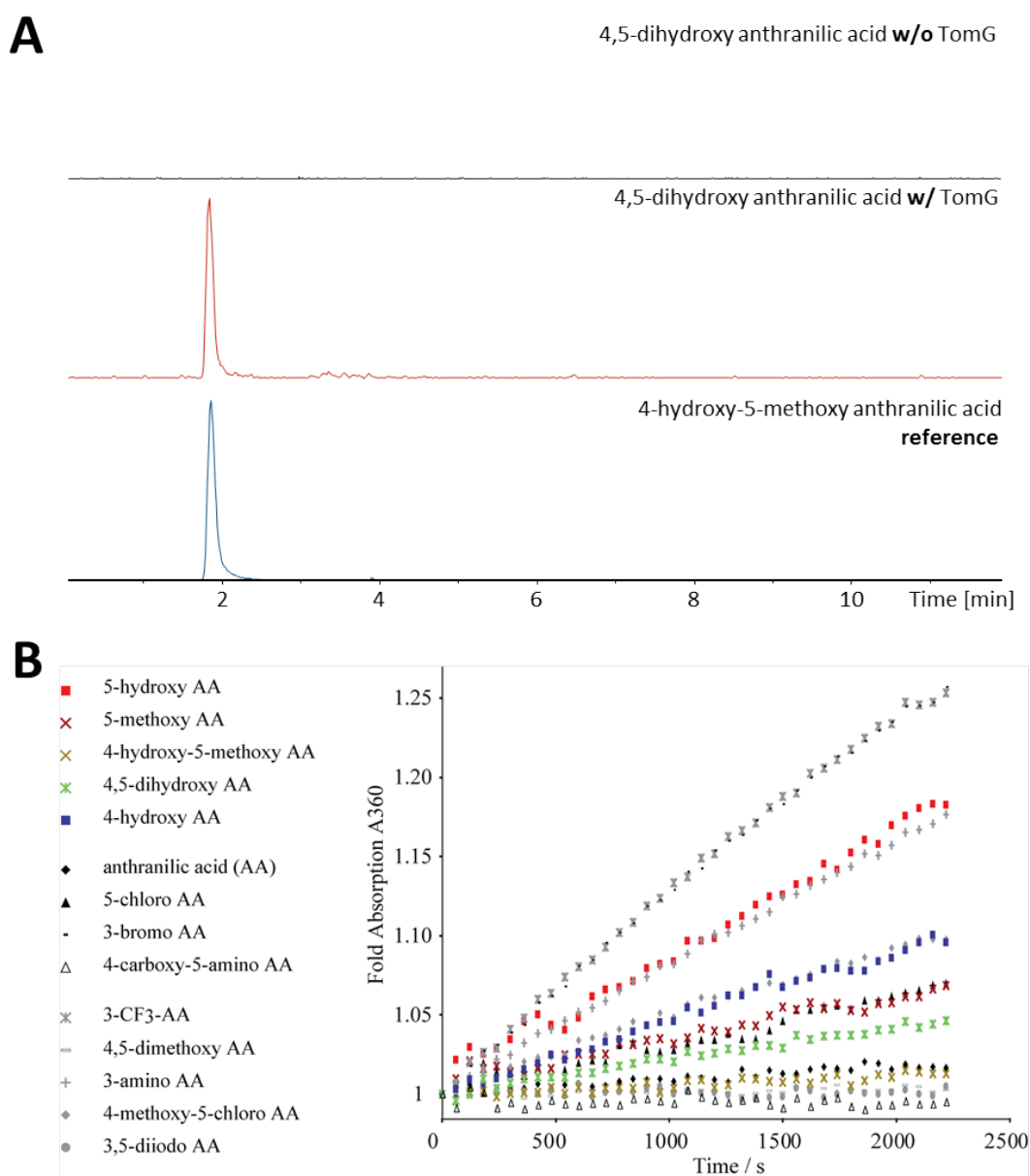


Figure 2.9: **A** EIC traces of 4-hydroxy-5-methoxy anthranilic acid (m/z 184.06043 0.002) to prove the functionality of TomG on 4,5-dihydroxy anthranilic acid. Substrate was incubated without and with TomG. The observed signal was validated by measuring a reference of 4-hydroxy-5-methoxy anthranilic acid. **B** Hydroxamate-MESG assay of TomA A-domain reveals relaxed substrate specificity while 5-hydroxy anthranilic acid has the fastest adenylation rate within the possible natural substrates. Phosphate production was quantified by measuring A360. Fold increase of absorption relative to control reaction omitting TomA A-domain is displayed.

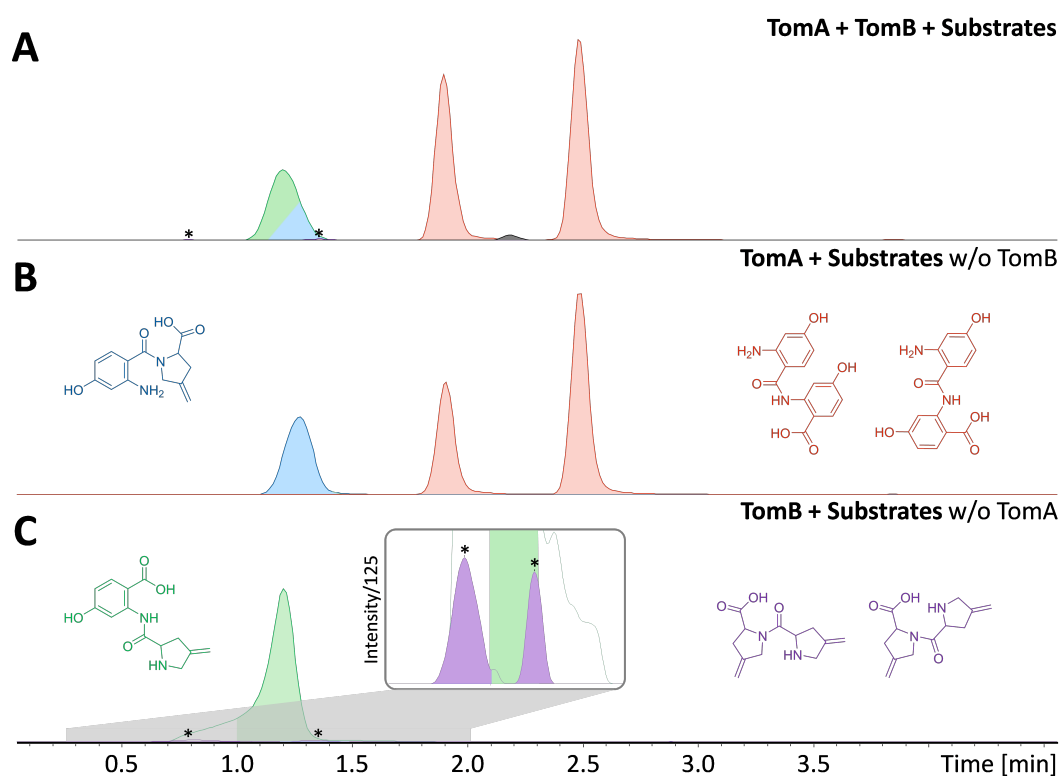


Figure 2.10: Overview of detected dimeric side products in the in vitro reconstitution assays. Overlaid EIC traces of the hetero-dimeric peptides with proposed different linkage (m/z 263.10263 0.002), proposed homo-dimeric side product conformers of two anthranilic acids (m/z 289.08190 0.002) and two prolines (m/z 237.12337 0.002). A Assay with TomA and B including 5-hydroxyanthranilic acid, methylene proline and the respective co-factors. B Same mixture but lacking TomB. C Same mixture as in A but lacking TomA.

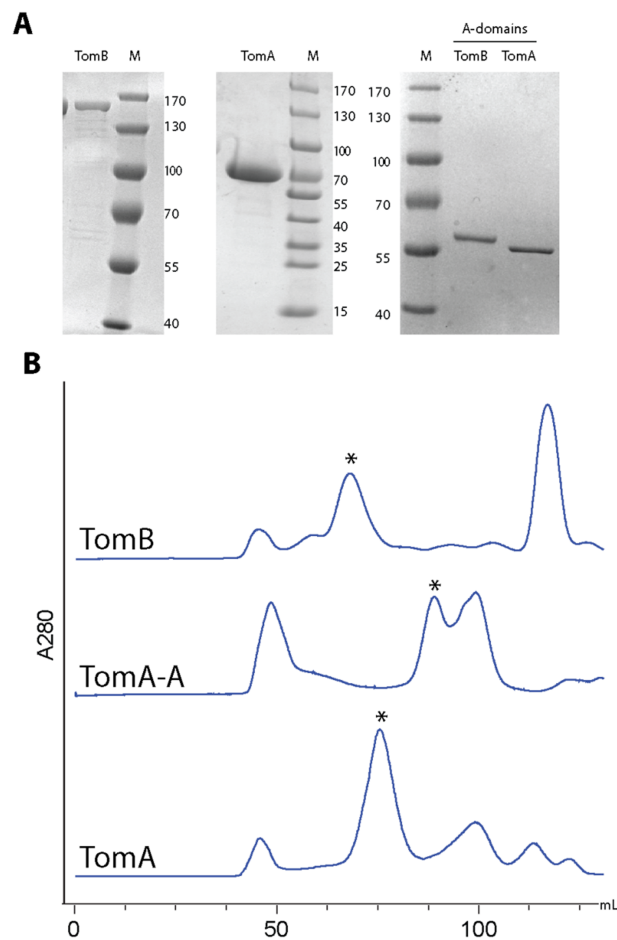


Figure 2.11: **A** SDS-PAGE analysis of SEC purified proteins TomA (65 kDa), TomB (167 kDa) and the TomA A-domain (55 kDa). **B** SEC traces of TomA, TomA-A (A-domain) and TomB.

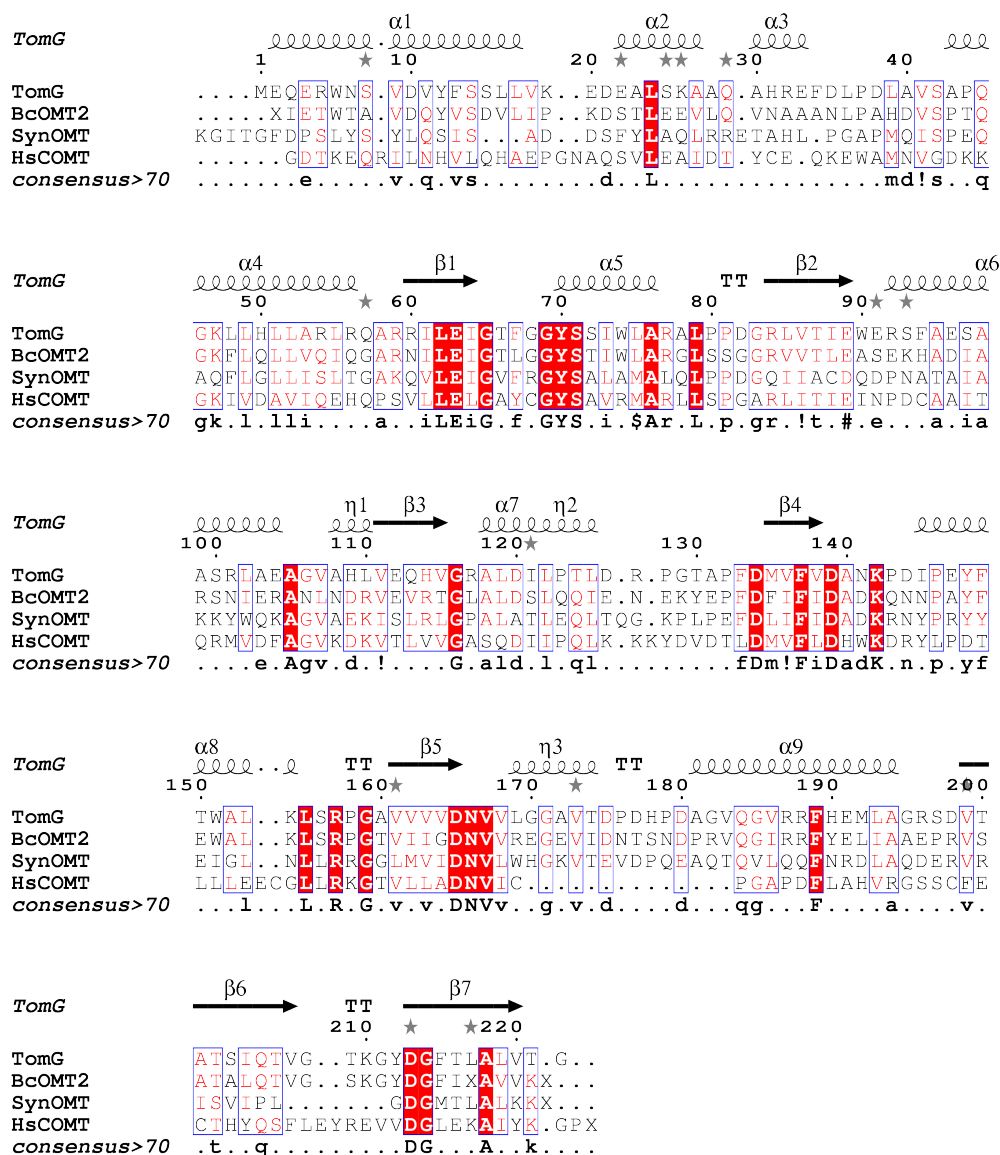


Figure 2.12: Multiple structural alignment and consensus sequence of diverse OMTs. Secondary structure assignment and sequence numbering corresponds to the structure of TomG. Sequences were aligned using the DALI webserver.⁸⁵ The alignment was manually edited and visualized using the ESPript 3.0 webserver.⁸⁸ Abbreviations used: BcOMT2 - O-methyltransferase from *Bacillus cereus*; SynOMT - O-methyltransferase from *Cyanobacterium synechocystis*; HsCOMT - catechol-O-methyltransferase from *Homo sapiens*.

Supplementary information about the TomG crystal structure

TomG was obtained as a homodimer which crystallizes at a concentration of 5 mg/mL with 0.6 mM of the cofactor S-adenosyl methionine (SAM) and 2.1 M $\text{NH}_4(\text{SO}_4)_2$ as precipitant in space group P212121 and two molecules in the asymmetric unit. The crystal structure was determined by molecular replacement using the structure of a cation-dependent O-methyltransferase (OMT) from *Cyanobacterium synecocystis* exhibiting a sequence similarity of 33 % (SynOMT, pdb code: 3cbg?).⁷⁹ As described in the main section, common features of SAM-dependent class I OMTs is a conserved SAM binding site and a Sn^{2+} -based transfer of the methyl group of SAM to the substrate in the presence of an Mg^{2+} ion.^{58–60, 86} Consequently, the active site architecture is conserved among OMTs that utilize this mechanism, and in TomG, the highly conserved residue Lys142 (Figure S 2.12) would be positioned to deprotonate the hydroxyl group of potential substrates. Furthermore, the conserved amino acids Asp139, Asp165, Asn166 together with a water molecule in the immediate vicinity are expected to octahedrally coordinate an Mg^{2+} ion (Figure 2.6), although no divalent ion was observed in the electron density for TomG. In proximity to the proposed metal binding site, additional electron density corresponding to SAM was identified in both chains of the structure. The cofactor is bound by a set of H-bonds to the side chain or backbone atoms of residues Val41, Gly65, Ser71, Glu89, Asp139, and water molecules. Furthermore, the adenine moiety forms hydrophobic interactions with residues Trp90, Ala118, and Ala140. Among these residues, Gly65, Ser71, Glu89, Asp139, Ala118 and Ala140 are highly conserved between related OMTs (Figure S 2.12).⁸⁷ Via these interactions, SAM is positioned such that its methyl group points towards the proposed active site of TomG.

Table 2.1: Details of the crystal structure analysis, ¹ values in parenthesis correspond to highest resolution shell

Crystallization and data Collection	TomG
Wavelength (Å)	1.033
Resolution range (Å) ¹	50.47 1.55 (1.58 1.55)
Space group	P212121
Unit cell parameters a, b, c (Å), α , β , γ (°)	67.597, 75.877, 86.784, 90.00 90.00 90.00
Total reflections ¹	859146 (40703)
Unique reflections ¹	65091 (3142)
Multiplicity ¹	13.2 (13.0)
Completeness (%)	1 99.6 (98.7)
Mean I/sigma(I) ¹	22.9 (1.7)
R-meas ¹	0.079 (1.691)
R-pim ¹	0.021 (0.464)
CC1/2 ¹	1.000 (0.783)
CC* ¹	1.000 (0.937)
Wilson B-factor (Å ²)	11.53
Monomers/asymmetric unit	2
Refinement	
Resolution range (Å) ¹	50.47 1.55 (1.60 1.55)
R-work ¹	0.1482 (0.2791)
R-free ¹	0.1661 (0.2889)
Protein residues	447
Number of non-hydrogen atoms	
Total	4198
Protein	3577
Ligands	127
Water	494
Rmsd	
Bonds (Å)	0.004
Angles (°)	0.715
Ramachandran statistics	
Favored (%)	97.07
Allowed (%)	2.93
Outliers (%)	0.00
B-factor (Å ²)	
Average	22.77
Protein	20.96
Ligands	33.39
Water	33.15

Table 2.2: Oligonucleotides used in this study

Primer	Sequence
TomA-FP	AAAAAAGCTTCTAGTGACCTTTCTTCGGATCGCG
TomA-RP	AAAAAAGCTTCTACTGGGTCTGTGCGGCG
TomB-FP	AAAAACATGTTAATGAACTCCCCCTCCGAAC
TomB-RP	AAAAAAGCTTTTCAGTCGTCGGCCAGTTCTCC
TomG-FP	AAAAAACATATGATGGAACAAGAGCGATGGAA
TomG-RP	AAAAAAGGATCCCTAGCCGGTGACCAG
TomGcrystal-FP	CGCCATATGATGGAACAAGAGCG
TomGcrystal-RP	ATTCTCGAGCTAGCCGGTGACC

Chapter 3

In vitro reconstitution and heterologous expression of an enterotoxin produced by *Klebsiella* *oxytoca*

ALEXANDER VON TESMAR, MICHAEL HOFFMANN, VIKTORIA SCHMITT, ANTOINE
ABOU-FAYAD, JENNIFER HERRMANN, ROLF MÜLLER*

*To whom correspondence should be addressed.

Author's efforts

The author was responsible for the conception of the study, designed and performed experiments, evaluated and interpreted resulting data. The author contributed the identification and isolation of tilivalline. The author designed and generated all presented constructs for heterologous production of tilivalline and was responsible for the evaluation of the respective analytical data. The author cloned, expressed and purified all proteins for *in vitro* reconstitution. The author furthermore designed and performed all feeding and inhibitor studies. The author was responsible for conceiving and writing of the manuscript.

Contributions by others

Michael Hoffmann contributed partly to the conception of the study, performed experiments, evaluated and interpreted resulting data. He isolated tilivalline from the clinical isolate of *Klebsiella oxytoca* by preparative LC-MS. The majority of all LC-MS measurements were performed, evaluated and interpreted by him. Michael Hoffmann conducted also the experiments on the analytical scale to prove the spontaneous indole incorporation. Furthermore he contributed to the writing and editing of the manuscript. All bioactivity assays and cultivations regarding the clinical isolate or other pathogens (biosafety level 2) were conducted by Viktoria Schmitt, who was supervised by Jennifer Herrmann. All described synthesis including the respective purification were conducted by Antoine Abou-Fayad. The project was supervised by Rolf Müller, who also contributed to conceiving and proofreading of the manuscript.

Abstract

Tilvalline is a pyrrolo[4,2]benzodiazepine derivative produced by the pathobiont *Klebsiella oxytoca* and the causative toxin in antibiotic associated hemorrhagic colitis (AAHC). Heterologous expression of the tilivalline biosynthetic gene cluster along with *in vitro* reconstitution of the respective NRPS (NpsA, ThdA, NpsB) was employed to reveal a non-enzymatic indole incorporation via a spontaneous Friedel-Crafts-like alkylation reaction. Cytotoxicity screening of tilivalline and its carbinolamine precursor, both derived from total synthesis, revealed tilivalline to be the more active species and thus the actual toxin. Furthermore, the heterologous system was used to generate novel tilivalline derivatives by supplementation of respective anthranilate and indole precursors. Finally, it could be shown that salicylic and acetylsalicylic acid inhibit the biosynthesis of tilivalline in *K. oxytoca* liquid culture, presumably by blocking the peptide carrier protein ThdA, pointing towards a potential application as anti-virulence drug for AAHC.

3.1 Introduction

Tilivalline, a pyrrolo[4,2]benzodiazepine (PBD) derivative, is produced by *Klebsiella oxytoca*¹ and has been identified as the causative virulence factor in antibiotic associated hemorrhagic colitis (AAHC).² PBDs are a common class of natural products produced by *Streptomyces* species and occur with various substitution patterns. Compounds like anthramycin, sibiromycin or tomaymycin can bind covalently to the minor groove of double stranded DNA and hence exhibit a potent antineoplastic activity. In the past a considerable effort to produce novel synthetic PBD derivatives has been expended, resulting in PBD conjugates like DSB-120 and SJG-136, dimeric compounds with strong antitumor activity.³ Tilivalline likewise shows activity against mouse leukemia L1210 cells⁴ and has been identified to be causative for the cytotoxic effects of *K. oxytoca* extracts which have been initially described in 1989.^{5, 6} Recently, tilivalline has been reported to be fundamental for the course of AAHC caused by *K. oxytoca*.^{2, 7} AAHC is triggered by antibiotic treatment and a resulting enterobacterial *K. oxytoca* overgrowth in the colon that is inherently resistant to amino and carboxy penicillins. Such a dysbiotic population is dominated by the pathobiont *K. oxytoca* and causes bloody diarrhea and abdominal cramps.^{8–10} The biosynthetic gene cluster of tilivalline reported by Schneditz et al. contains a bi-modular non ribosomal peptide synthetase (NRPS) that is thought to synthesize the PBD core from an anthranilate and a pyrrolo precursor. This assumption is based on the finding that the tilivalline cluster has the same module organization as the sibiromycin and tomaymycin gene clusters.^{11, 12} NRPSs are multimodular megasynthetases that produce a diverse group of natural products by catalyzing peptide bonds between monomeric building blocks. Owing to the independence from the ribosome, NRPS building blocks are not limited to the proteinogenic amino acids. Hence, unusual blocks with diverse modifications are often incorporated. The minimal module for the chain elongation process consists of three domains: The adenylation domain (A) activates the substrate that subsequently is covalently tethered to the terminal thiol 4phosphopanthien prosthetic group (PPant) on the thiolation domain (T). The condensation domain (C) catalyzes the peptide bond between two adjacent T domain tethered substrates. The last module usually contains a thioesterase (TE) or reductase (R) domain that releases the substrate chain from the enzyme.^{13, 14} In addition, optional domains included in distinct modules can facilitate a variety of modifications such as epimerization, hydroxylation or methylation and thus increase the diversity of structural elements in natural products.¹⁵ Here we present the elucidation of the tilivalline biosynthesis using heterologous expression of the gene cluster and the *in vitro* reconstitution of the underlying NRPS NpsA, ThdA, and NpsB. Our data show a non-enzymatic incorporation of the tilivalline indole moiety by a Friedel-Crafts-like alkylation. Furthermore, the heterologous system was used to generate novel tilivalline derivatives in a convenient 96-well plate based synthetic biology platform. In addition, we identified competitive inhibitors to block the tilivalline NRPS in the natural producer *K. oxytoca* and thus prevented toxin production in liquid culture. This finding is especially intriguing as tilivalline biosynthesis inhibition could serve as a valuable pathoblocker strategy in colitis treatment.

3.2 Results

Clinical isolates of *K. oxytoca* produces tilivalline

K. oxytoca isolate #6 was isolated from stool sample at the University Hospital Homburg Saar, Germany. The presence of the tilivalline biosynthetic gene cluster was confirmed by colony PCR using primers for NpsA. An initial small-scale culture of the respective isolate in liquid medium was extracted with ethyl acetate and fractionated into 96 well plates while a split flow was used to acquire LC-MS data. Eventually, the fractions were screened towards cytotoxicity against KB-3.1 human cervix carcinoma cells. The combined screening effort revealed one active fraction containing primarily a compound with 334.1550 m/z. Comparison with an authentic standard obtained by total synthesis revealed the detected compound as tilivalline, which was finally confirmed by purification using preparative HPLC and subsequent NMR analysis (Figure 3.1). To gain further insights into the genome of *K. oxytoca* isolate #6 the Illumina sequencing technology was used for sequencing and the gene data were analyzed towards similarity to the already reported tilivalline biosynthetic gene cluster. The tilivalline biosynthetic gene cluster was identified using Antismash 2.0¹⁶ and revealed an identical locus organization as the one reported by Schneditz et al.. The DNA sequence identity of the two clusters spanning 16.264 bp is striking 99.4 %. The lowest observed identity on protein level is 96.7 % for UvrX (Table S 3.3). The three genes *npsA*, *thdA* and *npsB* encode a two-modular NRPS including a reductive release domain. They share identical domain organization with the already described PBD producing synthetase of tomaymycin from *Streptomyces*.¹² Similar to the tomaymycin cluster, copies of shikimate pathway genes and other housekeeping enzymes are present to provide the anthranilic acid precursor for the biosynthesis of the core structure. The homologue of the 4-nitrophenol-2-monooxygenase NphA1, HmoX, is putatively responsible for the hydroxylation at position 3 of the anthranilic acid. Other NphA1 homologues can be found in the biosynthetic gene clusters of tomaymycin or sibiromycin, which both harbor a hydroxylated anthranilic acid moiety.¹⁷ The comparison with the tomaymycin biosynthesis further indicated a release of the formed dipeptide as an aldehyde followed by an immediate circularization to an imine which reacts spontaneously to the respective carbinolamine by hydration. The nature of the indole incorporation remained elusive as no candidate genes for the indole incorporation could be identified in the tilivalline biosynthetic gene cluster (Figure 3.1).

Heterologous expression of the tilivalline biosynthetic gene cluster reveals non-enzymatic indole incorporation

To generate a heterologous expression construct a TAR-based assembly of PCR amplified DNA fragments was employed (see methods section for details). The two operons containing the NRPS, the tailoring, and the precursor biosynthetic genes were PCR amplified from the *K. oxytoca* isolate #6 genomic DNA. The promoter cassette sequences for Ptet and PBAD as well as the vector backbone were PCR amplified while suitable homologous regions at the 5' end of the primers were introduced.¹⁸ The resulting five fragments were assembled by TAR in *Saccharomyces cerevisiae* to the plasmid pCly-Til. Plasmids exhibiting the correct restriction pattern were verified by employing Illumina sequencing technology (Figure 3.2). Upon induction of *E. coli* BL21 (DE3) pCLY-Til-24 in M9 minimal medium containing tetracycline and L-arabinose, tilivalline was

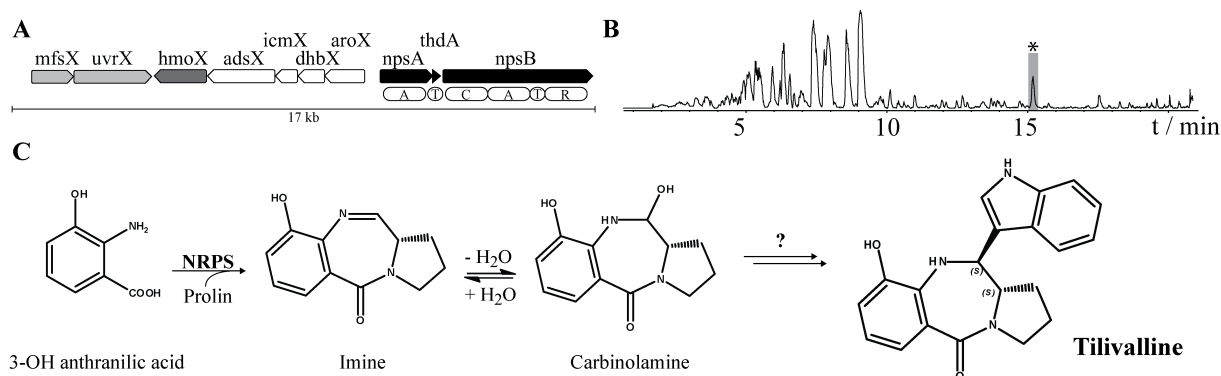


Figure 3.1: **A** Biosynthetic gene cluster of tilivalline from *K. oxytoca* isolate #6 spans 17 kb. It contains ten open reading frames constituting a two modular NRPS (black), the putative 3-hydroxylase (dark grey), genes for providing anthranilic acid (white) and regulators (light grey). **B** UV (200-600 nm) trace of the ethyl acetate extract of an *K. oxytoca* isolate #6 fermentation. The asterisk indicates the cytotoxic fraction harboring tilivalline. **C** Proposed biosynthesis of tilivalline: The NRPS (NpsA,ThdA,NpsB) generates the tilivalline backbone from 3-hydroxy anthranilic acid and proline. The biochemistry of the subsequent indole incorporation was elusive.

not detectable in the supernatant. However, trace amounts of the carbinolamine derived from 3-hydroxy anthranilic acid and L-proline were identified by LC-MS. The heterologous system apparently lacked the required indole source and/or enzyme for tilivalline production. Since no indole formation related genes are present in the tilivalline gene cluster and the *E. coli* tryptophan degradation pathway is repressed in minimal media,¹⁹ indole was supplemented to the medium by feeding 5 mM indole. Repeatedly performed LC-MS analysis of the corresponding cultures then indeed confirmed the presence of tilivalline indicating a non-enzymatic mechanism for tilivalline formation from the imine. To facilitate indole generation in vivo the gene of the wild type tryptophanase TnaA of *K. oxytoca* was cloned tag-free into the pET-28b expression vector. Upon induction of an *E. coli* transformant harboring both constructs, pCLY-Til-24 and pET-28b-tnaA, tilivalline production was observed without any addition of indole (Figure 3.2). To circumvent the poor growth and low yields in *E. coli* transformants harboring two plasmids with a combined size of over 20 kb and two resistance markers, a second approach was established. The operon containing the core NRPS NpsA, ThdA, and NpsB was cloned entirely and tag-free into pET-28b, which yielded a 10-fold increased tilivalline production in M9 minimal media supplemented with 3-hydroxy anthranilic acid, indole, and L-proline. To prove the non-enzymatic nature of the indole incorporation, all enzymes in the supernatant of an *E. coli* pET-28b-Til-NRPS transformant grown without supplementation of indole were inactivated by three different methods. The supernatant containing the carbinolamine was either heat inactivated at 100 °C for 5 minutes or supplemented with SDS or methanol to final concentrations of 1 % and 20 %, respectively. Subsequently performed incubation with indole confirmed the tilivalline formation in all three scenarios and thus manifested the probability of a spontaneous reaction of the NRPS produced carbinolamine with free indole (Figure 3.3).

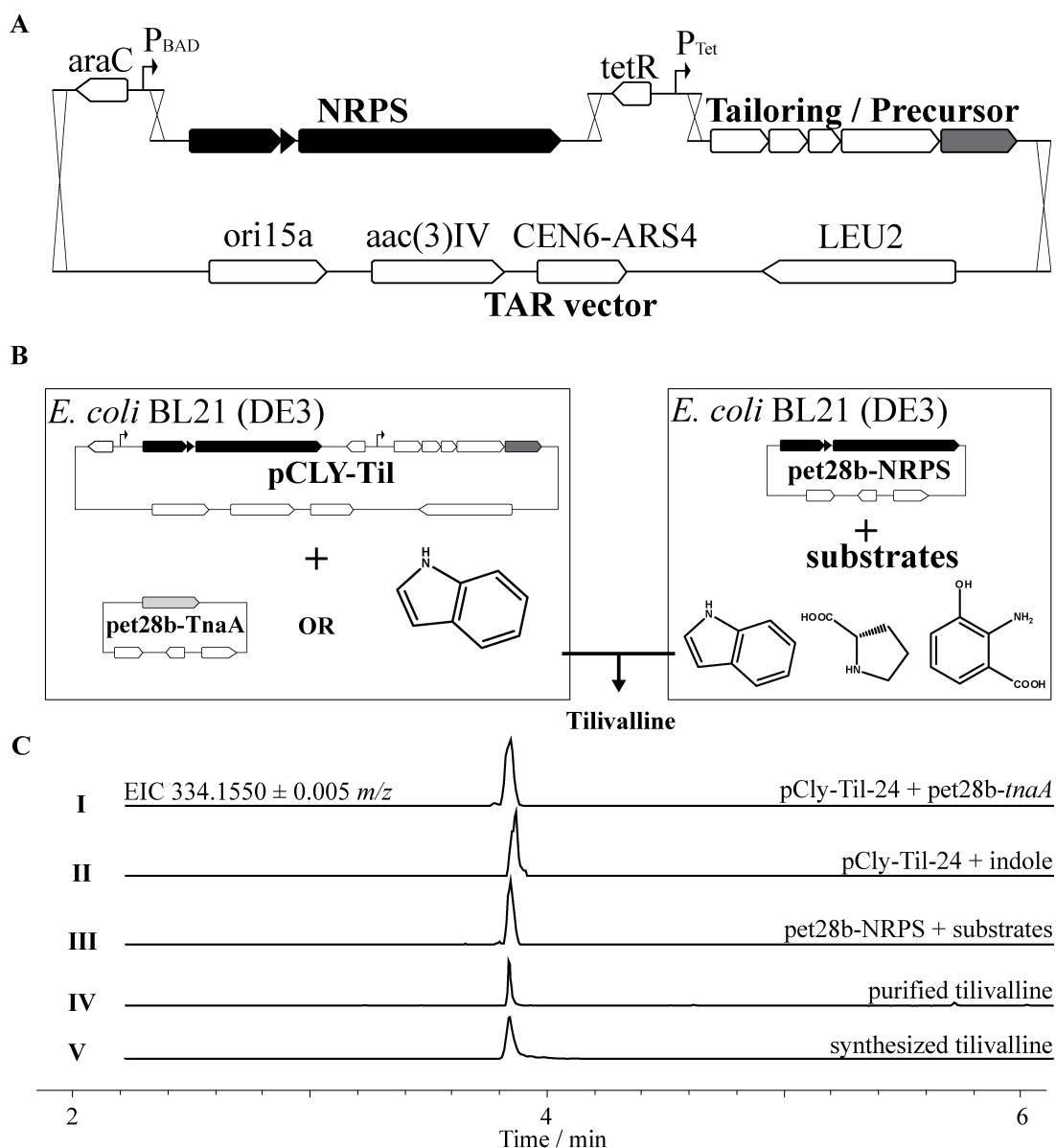


Figure 3.2: **A** Schematic display of the TAR assembly of the five PCR generated DNA fragments to yield the heterologous expression construct pCLY-Til. **B** Two different heterologous expression systems for tilivalline in *E. coli* BL21 (DE3) have been generated. Simultaneous expression of pCLY-Til and pET-28b-tnaA or indole supplementation led to successful restoration of tilivalline production. Expression of the construct pET-28b-NRPS while supplementation of all three precursors indole, 3-hydroxy anthranilic acid, and L-proline restored tilivalline production as well. **C** LC-MS analysis of the different heterologous systems shows the successful production of tilivalline in M9 minimal media. *E. coli* BL21 (DE3) pCly-Til-24 produces tilivalline when indole is either produced in the cell by the tryptophanase construct pET-28b-tnaA (I) or supplemented in the culture media (II). *E. coli* BL21 (DE3) pET-28b-NRPS produces tilivalline when the substrates are supplemented (III). Synthesized (V) and purified tilivalline from *K. oxytoca* isolate #6 (IV) were used as standard.

In vitro reconstitution of the tilivalline NRPS

The three genes constituting the tilivalline NRPS, *npsA*, *thdA*, and *npsB* and the *K. oxytoca* tryptophanase *tnaA* were heterologously expressed in *E. coli* BL21 (DE3). For the NRPS enzymes a N-terminal, HRV3C cleavable, 6xHis-MBP tag was used as it showed to enhance the yields and solubility. The obtained NRPS proteins were purified to homogeneity by a three-step protocol using nickel-affinity chromatography, a reverse nickel-affinity chromatography step upon

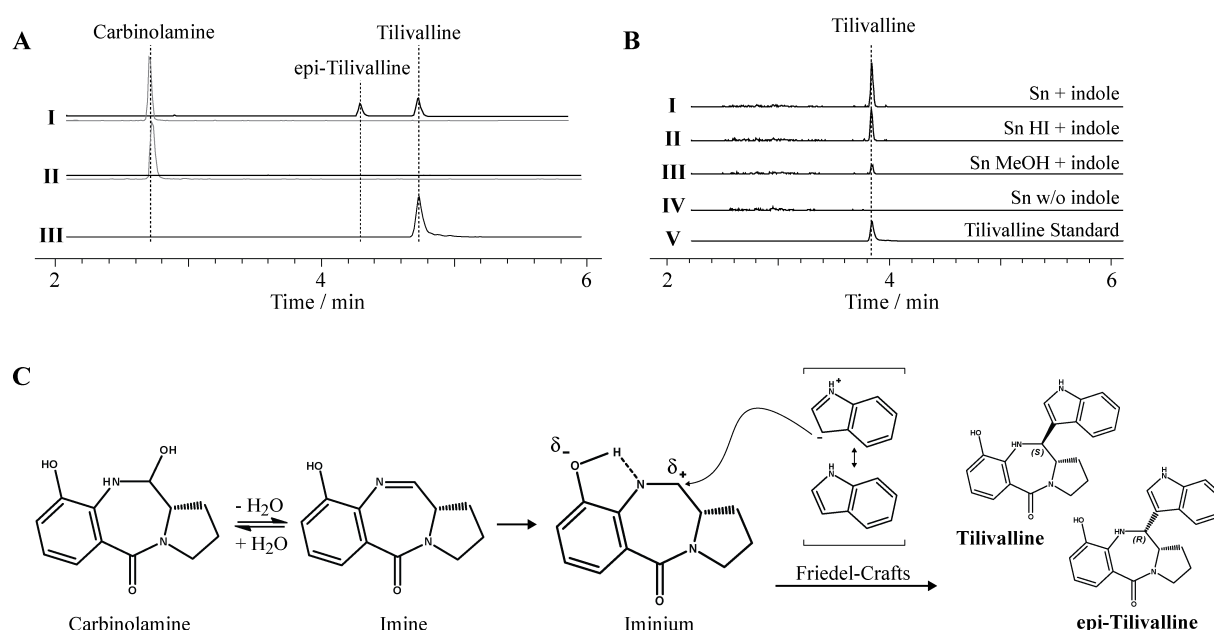


Figure 3.3: **A** Formation of tilivalline and epi-tilivalline can be detected upon incubation of indole with the purified PBD precursor in the carbinolamine form (I). PBD-precursor was purified from the *in vitro* reconstitution of NpsA, ThdA and NpsB without indole supplementation (II). Synthesized tilivalline was used as a standard (III). EIC traces for tilivalline (black, 334.1500 0.002 m/z) and carbinolamine (grey, 235.10772 0.002 m/z) are shown. **B** LC-MS derived EICs prove formation of tilivalline in the supernatant of induced *E. coli* BL21 (DE3) pET-28b-NRPS cultures in M9 minimal media (IV) upon addition of indole (I). Inactivation of enzymes by heat (HI) (II) or methanol (III) did not impair product formation. Synthesized tilivalline was used as standard (V) **C** Theoretical reaction mechanism for indole incorporation include the iminium transition state stabilized by the acidic phenolic hydrogen in ortho-position. Tilivalline is formed after Friedel-Crafts-like alkylation of nucleophilic indole at the carbeniumion takes place. Non-enzymatic incorporation leads to loss of stereochemical control and both epimers, tilivalline and epi-tilivalline are formed, while the latter proved instable and degrades over time.

HRV-3C digestion to remove the tags, and a polishing gel filtration step (Figure S 3.9). TnaA was enriched by affinity chromatography using an N-terminal 6xHis-tag and purified to homogeneity by a final gel filtration step without removing the tag (Figure S 3.10). The purified proteins NpsA, ThdA, and NpsB were used to reconstitute the biosynthesis of tilivalline *in vitro*. Upon incubation of the NRPS proteins with the substrates 3-hydroxy anthranilic acid, L-proline, and indole together with the cofactors ATP and NADPH, tilivalline formation could be observed by LC-MS. The addition of TnaA, tryptophan and the cofactor PLP instead of indole also restored production. The activity of TnaA was previously reconstituted *in vitro* (Figure S 3.10).²⁰ Trace amounts of the carbinolamine could be detected in all assays but it was further accumulated once the indole source was omitted. To prove the non-enzymatic nature of the indole incorporation, the carbinolamine was purified from the *in vitro* reconstitution assay in an analytical scale. *In vitro* reaction with indole yielded tilivalline and a second, isobaric compound with a slightly different retention time (Figure 3.3). This additional signal is most likely owing to the formation epi tilivalline, the diastereomer of tilivalline, during indole addition, since no enzyme controls the stereochemistry. In contrast to tilivalline, the signal of epi tilivalline decreased in aqueous solution over time, indicating degradation. Instability of epi tilivalline also explains its absence in *K. oxytoca* extracts and in the heterologous system.

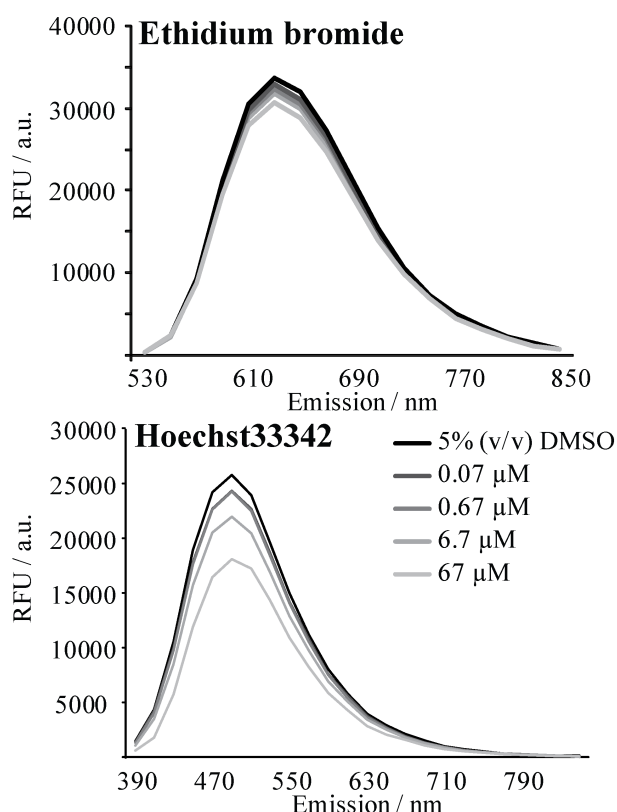
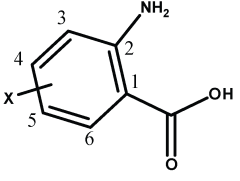


Figure 3.4: Tilivalline does not displace ethidium bromide but displaces Hoechst33342 at concentrations higher than $6.7 \mu\text{M}$. This indicates a weak binding of tilivalline to the minor groove of the DNA but excludes an intercalation.

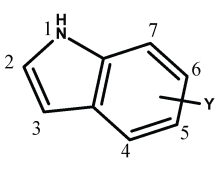
Tilivalline and not the carbinolamine is the active species

Schneditz et al. reported an IC_{50} of $10 \mu\text{M}$ for tilivalline against Hep2 cells. This weak cytotoxicity and the non-enzymatic indole incorporation raised the question if tilivalline or rather the carbinolamine or even the imine precursor is the actual virulence factor. The carbinolamine precursor of tilivalline was synthesized (see synthetic procedures for details) and an IC_{50} value for KB-3.1 human cervix carcinoma cells of $23.65 \mu\text{M}$ was determined for the synthetic material. Thus, it is 20-fold less active compared to tilivalline with a determined IC_{50} value of $1.63 \mu\text{M}$ in our system. Consequently, the indole moiety seems to be essential for the virulence of *K. oxytoca*. The imine that must be formed first could not be detected in any assay and could neither be synthesized, presumably because of the high reactivity. However, the carbinolamine and tilivalline are magnitudes less active than the imine tomaymycin that exhibits an IC_{50} value of $0.041 \mu\text{M}$. We speculate that the substitution pattern found in tomaymycin allows for stability of the imine which is not the case in the hypothetical tilivalline imine precursor. In addition to the cytotoxicity assays, we investigated potential tilivalline-DNA interaction which is the known mode of action of many PBDs, for instance anthramycin. Accordingly performed DNA displacement assays with tilivalline and ethidium bromide showed no activity while when using Hoechst33342 a slight displacement at high concentrations of $> 6.7 \mu\text{M}$ was observed. This suggested that no DNA intercalation takes place and only a low affinity to the minor groove of the DNA exists (Figure 3.4). Consequently, the mode of action of tilivalline remains elusive.



anthranilic acid

Derivative	Production
AA	+++
4-OH	+++
3-Me	++
3-OMe	+++
5-OMe	++
3-F	++
5-F	+++
3-Cl	++
4-Cl	+++
5-Cl	+++
5-Br	++
3-CF3	tr.



indole

Derivative	Production
4-OH	+++
5-OH	+++
6-OH	+++
4-NO2	+
5-NO2	++
6-NO2	++
7-NO2	+
8-F	+++
9-F	+++
10-F	+++
5-Br	+
5-I	tr.

Figure 3.5: Mutasynthesis of tilivalline derivatives by supplementing *E. coli* BL21 (DE3) pET-28b-NRPS fermentations with anthranilic acid or indole derivatives. Incorporation efficiency was estimated by comparing LC-MS derived EIC signal intensities compared to tilivalline production: similar production (+++), one magnitude less (++), little (+) and trace amount (tr).

Tilivalline NRPS as heterologous synthetic biology platform

A heterologous expression system was utilized to set up a 96-well plate based platform for a precursor directed mutasynthesis approach (Figure 3.5). Twelve different halogenated, hydroxylated, or nitrated indole derivatives were supplemented with 3-hydroxy anthranilic acid and L-proline and supernatants were subsequently analyzed by LC-MS (Figure S 3.7). Fluoro- and hydroxyl substituted indoles were incorporated with comparable efficiencies to the native, unsubstituted indole. Although nitro- and bromo derivatives were also accepted, the signal intensities were one magnitude lower. The incorporation of 5-iodo indole was extremely inefficient and only trace amounts of the respective tilivalline derivative could be detected. Furthermore, the tilivalline NRPS accepts twelve different anthranilic acid derivatives including halogen-, hydroxy-, methoxy-, and trifluoromethyl substituted ones (Figure S 3.8). The halogenated as well as the hydroxylated and methoxylated anthranilic acid derivatives were readily accepted with product signal intensities comparable to the natural tilivalline. The substitution of the 3-hydroxy group with the sterically demanding trifluoromethyl led to a nearly complete abolishment of production and only trace amounts of the respective derivative could be detected. All anthranilic acid derivatives were supplemented together with indole and L-proline. For all tilivalline derivatives, the emergence of a second signal with an approximately 100-fold lower intensity compared to the main product signal was observed. This indicates the formation of both epimers during indole addition and the subsequent degradation of the epi-tilivalline derivative, as already observed for tilivalline.

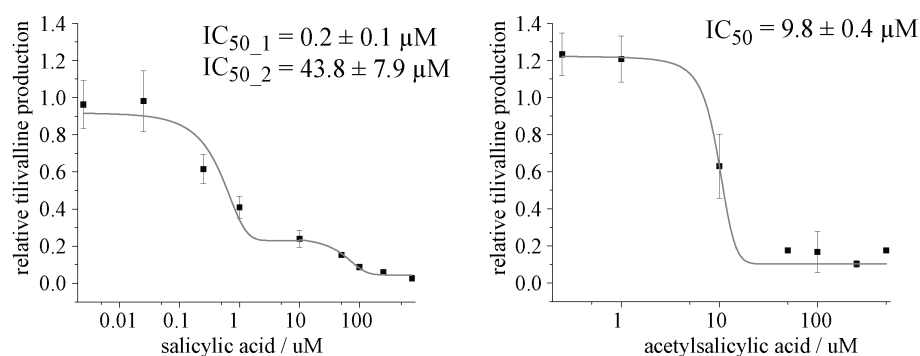


Figure 3.6: Salicylic acid and acetylsalicylic acid inhibit tilivalline biosynthesis in *K. oxytoca* isolate #6 liquid culture. Experiments were carried out in triplicates and mean values and standard deviations of tilivalline EIC signal areas are given. A two-step and one-step dose-response model was used to fit salicylic acid and acetylsalicylic acid data, respectively.

Salicylic acid based compounds inhibit tilivalline production in *K. oxytoca*

Tilivalline production in *K. oxytoca* isolate #6 was evaluated for different fermentation conditions, whereby incubation for 16 h at 37 °C turned out to be most reliable. Tilivalline production in presence of different potential inhibitors were subsequently quantified relative to each other by LC-MS. As could be predicted rationally from the biosynthetic route, benzoic acid derivatives lacking the amino group might be loaded to the NRPS and act as inhibitors of the overall reaction. Indeed, salicylic acid inhibited production completely in two steps with IC₅₀ values of 0.3 µM and 43 µM. Acetylsalicylic acid showed an IC₅₀ value of 10 µM for the reduction of tilivalline production to a final level of 15 % (Figure 3.6).

3.3 Discussion

Antibiotic resistant pathogens are an increasing threat to public health. Consequently, the development of alternative antibacterial strategies has become vital. Virulence of pathogens is often dependent on distinct natural products that act as toxins or siderophores.²¹⁻²⁴ Hence, the underlying biosynthetic machinery has emerged as a promising antibacterial target.^{25, 26} Substances that are capable to block secondary metabolite pathways have less impact on the overall fitness of the pathogen compared to antibiotics, which usually target vital cell functions. This is thought to reduce the probability of resistance development and the problematic side effects associated with an antibiotic regime.²⁷ *K. oxytoca* is a human intestinal pathobiont prevalent in 2-10 % of healthy individuals. The toxin producing *K. oxytoca* has been identified to cause AAHC by overgrowing the colon as a consequence to penicillin treatment.^{8, 9, 28} A causative link between the NRPS-produced PBD tilivalline and AAHC was recently established, making the underlying biosynthesis a viable target for drug design efforts.² Here, we present the elucidation of tilivalline biosynthesis and the setup of a synthetic biology platform to produce a variety of derivatives which eventually opens up the possibility to inhibit the toxin biosynthesis in *K. oxytoca* liquid culture using competitive inhibitors. The structural backbone of tilivalline is produced by a two-modular NRPS, similar to the ones found in PBD producing actinomycetes. Interestingly, tilivalline harbors an indole moiety that is incorporated by a non-enzymatic process, as we show in this paper. The reaction occurs spontaneously when incubating the respective PBD with indole in aqueous solution. It can be assumed that the positive resonance effect of the ortho-hydroxyl group increases the basicity of the imine. A five-membered ring transition state including the phenolic hydrogen and the imine nitrogen presumably promotes iminium formation. The resulting mesomeric carbeniumion undergoes a Friedel-Crafts like alkylation with the electron rich indole-carbon (Figure 3.3). Since *K. oxytoca* is indole-positive the required free indole is provided by degrading tryptophan.²⁹ Other PBDs like tomaymycin, sibiromycin, and anthramycin show promising cytotoxic activities and thus has been further developed towards antitumor agents.^{30, 31} The underlying mode-of-action is the covalent bond formation between the imine carbon of the PBD with an amino group of a guanidine in the minor groove of the DNA.³² Apparently, an analogous mode-of-action for tilivalline is not realistic since the imine-carbon of the PBD is already occupied by the indole, what raised the question whether tilivalline or rather the PBD precursor is the actual virulence factor. The synthesized PBD carbinolamine precursor of tilivalline surprisingly exhibited significantly lower cytotoxic activity compared to tilivalline. Furthermore, isolation of the imine form of the PBD precursor during total synthesis proved impossible due to instant and quantitative conversion to the carbinolamine form in aqueous solution, probably a result of the different substitution pattern on the aromatic moiety compared to the PBD compounds from actinomycetes. Existence of the imine form *in vivo* is therefore improbable and although bacterial secretion systems that inject toxins directly into the host cell exist,^{33, 34} the high instability of the imine makes a participation in *K. oxytoca* virulence unlikely. Our findings provide evidence that tilivalline indeed is the bioactive species and the underlying mode-of-action differs from the one described for other PBDs such as tomaymycin. Furthermore, tilivalline is a prominent biomarker and thus valuable for the development for a rapid screening method for the detection of AAHC associated *K. oxytoca* infections. Although the molecular details of tilivalline induced apoptosis remain elusive, the respective NRPS con-

stitutes a promising target for anti-virulence strategies. NRPS assembly lines use adenylation domains to activate the amino acid substrates as acyl-AMP that remains tightly bound to the active site for downstream processing. Therefore, non-hydrolyzable acyl-AMP mimics like the acyl-sulfamoyl-adenosine (acyl-AMS) scaffold have been commonly used for selective inhibition of adenylation enzymes.³⁵ Despite the established application in in vitro experiments, the application in liquid culture is hampered due to poor cell-permeability.³⁶ Nevertheless, inhibition of the biosynthesis of salicyl-capped non-ribosomal-peptide-polyketide siderophores in *M. tuberculosis* and *Y. pestis* using salicyl AMS has been reported where successful inhibition impaired bacterial growth under iron-limiting conditions.³⁷ Along those lines we here present an alternative inhibition approach for the NRPS produced toxin tilivalline. The relaxed substrate specificity of NpsA, the first A-domain of the tilivalline assembly line, became obvious in the presented synthetic biology platform and inspired us to test salicylic acid as competitive inhibitor. Activation and transfer of salicylic acid, that misses the amino group compared to the native substrate 3-hydroxy anthranilic acid, blocks the PCP-domain to be no longer available for tilivalline biosynthesis. Our work provides first evidence that salicylic and acetylsalicylic acid inhibits tilivalline biosynthesis in *K. oxytoca* liquid cultures. Acetylsalicylic acid has been shown to hydrolyze to salicylic acid in aqueous solution and thus probably acts as a prodrug.³⁸ The low cytotoxicity values of tilivalline and the high production titer in liquid culture indicate a high occurrence of the toxin in the colon as well. Consequently, reduction of this pathogenicity factor is a promising approach to cure AAHC. The observed inhibition of its biosynthesis by salicylic acid, a well-established drug of such long-standing history, poses a remarkable opportunity for future treatments.

3.4 Bibliography

- [1] Nikolaus Mohr and Herbert Budzikiewicz. Tilivalline, a new pyrrolo[2, 1-c][1,4] benzodiazepine metabolite from klebsiella. *Tetrahedron*, 38(1):147–152, 1982.
- [2] Georg Schneditz, Jana Rentner, Sandro Roier, Jakob Pletz, Kathrin A. T. Herzog, Roland Bückner, Hanno Troeger, Stefan Schild, Hansjörg Weber, Rolf Breinbauer, Gregor Gorkiewicz, Christoph Högenauer, and Ellen L. Zechner. Enterotoxicity of a nonribosomal peptide causes antibiotic-associated colitis. *Proceedings of the National Academy of Sciences of the United States of America*, 111(36):13181–13186, 2014.
- [3] Barbara Gerratana. Biosynthesis, synthesis, and biological activities of pyrrolobenzodiazepines. *Medicinal research reviews*, 32(2):254–293, 2012.
- [4] T. Shioiri, T. Aoyama, N. Yamagami, N. Shimizu, N. Mori, and K. Kohda. Structure-cytotoxicity relationship of tilivalline derivatives. *Anti-cancer drug design*, 10(2):167–176, 1995.
- [5] J. Minami, A. Okabe, J. Shiode, and H. Hayashi. Production of a unique cytotoxin by klebsiella oxytoca. *Microbial pathogenesis*, 7(3):203–211, 1989.
- [6] M. Higaki, T. Chida, H. Takano, and R. Nakaya. Cytotoxic component(s) of klebsiella oxytoca on hep-2 cells. *Microbiology and immunology*, 34(2):147–151, 1990.
- [7] Alison Darby, Kvin Lertpiriyapong, Ujjal Sarkar, Uthpala Seneviratne, Danny S. Park, Eric R. Gamazon, Chara Batchelder, Cheryl Cheung, Ellen M. Buckley, Nancy S. Taylor, Zeli Shen, Steven R. Tannenbaum, John S. Wishnok, and James G. Fox. Cytotoxic and pathogenic properties of klebsiella oxytoca isolated from laboratory animals. *PloS one*, 9(7):e100542, 2014.
- [8] Laurent Beaugerie, Michaela Metz, Frédéric Barbut, Guy Bellaiche, Yoram Bouhnik, Laurent Raskine, Jean-Claude Nicolas, François-Patrick Chatelet, Norbert Lehn, and Jean-Claude Petit. Klebsiella oxytoca as an agent of antibiotic-associated hemorrhagic colitis. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association*, 1(5):370–376, 2003.
- [9] Christoph Högenauer, Cord Langner, Eckhard Beubler, Irmgard T. Lippe, Rudolf Schicho, Gregor Gorkiewicz, Robert Krause, Nikolas Gerstgrasser, Guenter J. Krejs, and Thomas A. Hinterleitner. Klebsiella oxytoca as a causative organism of antibiotic-associated hemorrhagic colitis. *The New England journal of medicine*, 355(23):2418–2426, 2006.
- [10] Janet Chow, Haiqing Tang, and Sarkis K. Mazmanian. Pathobionts of the gastrointestinal microbiota and inflammatory disease. *Current opinion in immunology*, 23(4):473–480, 2011.
- [11] Wei Li, Ankush Khullar, ShenChieh Chou, Ashley Sacramo, and Barbara Gerratana. Biosynthesis of sibiromycin, a potent antitumor antibiotic. *Applied and environmental microbiology*, 75(9):2869–2878, 2009.

-
- [12] Wei Li, ShenChieh Chou, Ankush Khullar, and Barbara Gerratana. Cloning and characterization of the biosynthetic gene cluster for tomaymycin, an sjg-136 monomeric analog. *Applied and environmental microbiology*, 75(9):2958–2963, 2009.
- [13] Robert Finking and Mohamed A. Marahiel. Biosynthesis of nonribosomal peptides1. *Annual review of microbiology*, 58:453–488, 2004.
- [14] Michael A. Fischbach and Christopher T. Walsh. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chemical reviews*, 106(8):3468–3496, 2006.
- [15] Gene H. Hur, Christopher R. Vickery, and Michael D. Burkart. Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Natural product reports*, 29(10):1074–1098, 2012.
- [16] Tilmann Weber, Kai Blin, Srikanth Duddela, Daniel Krug, Hyun Uk Kim, Robert Brucconeri, Sang Yup Lee, Michael A. Fischbach, Rolf Muller, Wolfgang Wohlleben, Rainer Breitling, Eriko Takano, and Marnix H. Medema. antismash 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic acids research*, 43(W1):W237–43, 2015.
- [17] Masahiro Takeo, Masumi Murakami, Sanae Niihara, Kenta Yamamoto, Munehiro Nishimura, Dai-ichiro Kato, and Seiji Negoro. Mechanism of 4-nitrophenol oxidation in rhodococcus sp. strain pn1: characterization of the two-component 4-nitrophenol hydroxylase and regulation of its expression. *Journal of bacteriology*, 190(22):7367–7374, 2008.
- [18] Oksana Bilyk, Olga N. Sekurova, Sergey B. Zotchev, and Andriy Luzhetskyy. Cloning and heterologous expression of the greccocycline biosynthetic gene cluster. *PloS one*, 11(7):e0158682, 2016.
- [19] Thi Hiep Han, Jin-Hyung Lee, Moo Hwan Cho, Thomas K. Wood, and Jintae Lee. Environmental factors affecting indole production in escherichia coli. *Research in microbiology*, 162(2):108–116, 2011.
- [20] Shao Yang Ku, Patrick Yip, and P. Lynne Howell. Structure of escherichia coli tryptophanase. *Acta crystallographica. Section D, Biological crystallography*, 62(Pt 7):814–823, 2006.
- [21] George Y. Liu, Anthony Essex, John T. Buchanan, Vivekanand Datta, Hal M. Hoffman, John F. Bastian, Joshua Fierer, and Victor Nizet. Staphylococcus aureus golden pigment impairs neutrophil killing and promotes virulence through its antioxidant activity. *The Journal of experimental medicine*, 202(2):209–215, 2005.
- [22] Chung-I Chang, Yogarany Chelliah, Dominika Borek, Dominique Mengin-Lecreulx, and Johann Deisenhofer. Structure of tracheal cytotoxin in complex with a heterodimeric pattern-recognition receptor. *Science (New York, N.Y.)*, 311(5768):1761–1764, 2006.
- [23] Victoria I. Holden and Michael A. Bachman. Diverging roles of bacterial siderophores during infection. *Metallomics : integrated biometal science*, 7(6):986–995, 2015.

- [24] Wilma Neumann, Anmol Gulati, and Elizabeth M. Nolan. Metal homeostasis in infectious disease: recent advances in bacterial metallophores and the human metal-withholding response. *Current opinion in chemical biology*, 37:10–18, 2016.
- [25] Yongcheng Song, Chia-I Liu, Fu-Yang Lin, Joo Hwan No, Mary Hensler, Yi-Liang Liu, Wen-Yih Jeng, Jennifer Low, George Y. Liu, Victor Nizet, Andrew H-J Wang, and Eric Oldfield. Inhibition of staphyloxanthin virulence factor biosynthesis in staphylococcus aureus: in vitro, in vivo, and crystallographic results. *Journal of medicinal chemistry*, 52(13):3869–3880, 2009.
- [26] Audrey L. Lamb. Breaking a pathogen’s iron will: Inhibiting siderophore production as an antimicrobial strategy. *Biochimica et biophysica acta*, 1854(8):1054–1070, 2015.
- [27] Damien Maura, Alicia E. Ballok, and Laurence G. Rahme. Considerations and caveats in anti-virulence drug development. *Current opinion in microbiology*, 33:41–46, 2016.
- [28] Ines Zollner-Schwetz, Kathrin A. T. Herzog, Gebhard Feierl, Eva Leitner, Georg Schneditz, Hanna Sprenger, Jürgen Prattes, Wolfgang Petritsch, Heimo Wenzl, Patrizia Kump, Gregor Gorkiewicz, Ellen Zechner, and Christoph Högenauer. The toxin-producing pathobiont *klebsiella oxytoca* is not associated with flares of inflammatory bowel diseases. *Digestive diseases and sciences*, 60(11):3393–3398, 2015.
- [29] R. Podschun and U. Ullmann. *Klebsiella* spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. *Clinical microbiology reviews*, 11(4):589–603, 1998.
- [30] Rohtash Kumar and J. William Lown. Recent developments in novel pyrrolo2,1-c1,4benzodiazepine conjugates: synthesis and biological evaluation. *Mini reviews in medicinal chemistry*, 3(4):323–339, 2003.
- [31] Julia Manta, Paul J. M. Jackson, Khondaker M. Rahman, and David E. Thurston. From anthramycin to pyrrolobenzodiazepine (pbd)-containing antibody-drug conjugates (adcs). *Angewandte Chemie (International ed. in English)*, 56(2):462–488, 2017.
- [32] M. S. Puvvada, S. A. Forrow, J. A. Hartley, P. Stephenson, I. Gibson, T. C. Jenkins, and D. E. Thurston. Inhibition of bacteriophage t7 rna polymerase in vitro transcription by dna-binding pyrrolo2,1-c1,4benzodiazepines. *Biochemistry*, 36(9):2478–2484, 1997.
- [33] Meztili O. Gaytan, Veronica I. Martinez-Santos, Eduardo Soto, and Bertha Gonzalez-Pedrajo. Type three secretion system in attaching and effacing pathogens. *Frontiers in cellular and infection microbiology*, 6:129, 2016.
- [34] Erin R. Green and Joan Mecsas. Bacterial secretion systems: An overview. *Microbiology spectrum*, 4(1), 2016.
- [35] Robert Finking, Andrea Neumüller, Jens Solsbacher, Dirk Konz, Gerhard Kretzschmar, Markus Schweitzer, Thomas Krumm, and Mohamed A. Marahiel. Aminoacyl adenylate substrate analogues for the inhibition of adenylation domains of nonribosomal peptide synthetases. *Chembiochem : a European journal of chemical biology*, 4(9):903–906, 2003.

- [36] Tony D. Davis, Poornima Mohandas, Maria I. Chiriac, Glennon V. Bythrow, Luis E. N. Quadri, and Derek S. Tan. Design, synthesis, and biological evaluation of alpha-hydroxyacyl-ams inhibitors of amino acid adenylation enzymes. *Bioorganic & medicinal chemistry letters*, 26(21):5340–5345, 2016.
- [37] Julian A. Ferreras, Jae-Sang Ryu, Federico Di Lello, Derek S. Tan, and Luis E. N. Quadri. Small-molecule inhibition of siderophore biosynthesis in mycobacterium tuberculosis and yersinia pestis. *Nature chemical biology*, 1(1):29–32, 2005.
- [38] L. J. Edwards. The hydrolysis of aspirin. a determination of the thermodynamic dissociation constant and a study of the reaction kinetics by ultra-violet spectrophotometry. *Transaction of the Faraday Society*, 1950(46):723–736, 1950.
- [39] Jared T. Simpson, Kim Wong, Shaun D. Jackman, Jacqueline E. Schein, Steven J. M. Jones, and Inanc Birol. Abyss: a parallel assembler for short read sequence data. *Genome research*, 19(6):1117–1123, 2009.
- [40] Matthew Kearse, Richard Moir, Amy Wilson, Steven Stones-Havas, Matthew Cheung, Shane Sturrock, Simon Buxton, Alex Cooper, Sidney Markowitz, Chris Duran, Tobias Thierer, Bruce Ashton, Peter Meintjes, and Alexei Drummond. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)*, 28(12):1647–1649, 2012.
- [41] Amit Kunwar, Emmanuel Simon, Umang Singh, Rajnikant K. Chittela, Deepak Sharma, Santosh K. Sandur, and Indira K. Priyadarsini. Interaction of a curcumin analogue dimethoxycurcumin with dna. *Chemical biology & drug design*, 77(4):281–287, 2011.

3.5 Experimental Procedures

Sequencing of strain *K. oxytoca* isolate #6

The strain *K. oxytoca* isolate #6 was obtained from Dr. Alexander Halfmann, Saarland University Hospital, Homburg, Germany. Prior sequencing, presence of the tilivalline gene cluster was confirmed by successfully amplifying *npsA* using the primer pair NpsA-FP/NpsA-RP designed for protein purification. Draft genome sequence of *Klebsiella oxytoca* isolate #6 was obtained using Illumina sequencing technology in cooperation with the Genome Analytics group at the Helmholtz Centre for Infection Research (Braunschweig, Germany). Raw sequencing data obtained from the MiSeq platform comprised 2,849,072 paired-end reads with the length of 250 bp each. This raw data was assembled into contigs with Abyss-pe assembler software,³⁹ version 1.9.0. The resulting 192 contiguous sequences (contigs) were mapped to the reference sequence of a close relative genome of *Klebsiella oxytoca* MGH 28 (GenBank accession code: KI535631) in Geneious software,⁴⁰ version 9.0.5. The resulting draft genome circular scaffold of 5,720,468 bp was used for the downstream analysis.

Isolation of tilivalline

Preparative HPLC was performed with a Waters Autopurifier System (APS) equipped with a Waters XBridge C18 150 x 4.6 mm, 5 μ m dp analytical column for method development at 1.0 mL/min flow rate. Preparative separations were performed with a XBridge C18 150 x 19 mm, 5 μ m dp column at 25 mL/min flow rate. Both used (A) water + 0.1 % FA and (B) acetonitrile + 0.1 % FA as solvent system. Elution of an 800 μ L sample injection started with a 3 min isocratic step at 5 % B, a linear increase to 40 % B within 25 min, a steep increase to 95 % B in 2 min, followed by a 3 min plateau at 95 % prior to reequilibration to initial conditions. The overall run time was 34 min while 40 fractions were collected in a period from 19 to 22 min. Fractions were combined based on MS data and yielded tilivalline in high purity without any further purification efforts. For NMR data see table 3.4.

TAR assembly of the tilivalline biosynthetic gene cluster

PCR fragments of the vector backbone, promoter cassettes, tilivalline NRPS and tailoring operon were amplified using primers with suitable 40bp 5'overhangs (Table S 3.2). TAR assembly in *S. cerevisiae* was conducted as described previously.¹⁸ Colony PCR targeting fragment borders was used to preselect yeast colonies (see Table S 3.2). Positive clones were grown in selective leucine deficient media overnight and subsequently the plasmid DNA was isolated. Plasmid isolation by alkaline lysis from *E. coli* Gbred transformants and subsequent restriction digest was used to identify correct constructs. Isolation attempts with *E. coli* Dh10b or HS916 cells yielded low DNA concentration. Plasmids were subsequently verified by Illumina sequencing.

Heterologous expression of the tilivalline biosynthetic gene cluster

E. coli BL21 (DE3) pCly-Til-24 cells were grown overnight in LB media at 30 °C. Expression in 10 mL M9 minimal media with 2% glycerol as sole carbon source at 30 °C for 16 h was induced by addition of 2% (w/v) L-arabinose and 5 ng/mL tetracycline after cell density reached

OD₆₀₀=0.6. Subsequently the filtrated supernatant was forwarded to LCMS analysis. TnaA was coexpressed in pET-28b using the same protocol but with additional induction with 0.1 mM IPTG. Supplementation with indole was carried out simultaneously with induction with a final concentration of 1mM.

Heterologous expression of the tilivalline NRPS

The operon harboring the tilivalline NRPS was cloned into pET-28b using suitable primers allowing for tag-free expression. *E. coli* BL21 (DE3) pET-28b-NRPS cells were grown overnight in LB media at 30 °C. Expression in M9 minimal media with 2% glucose as sole carbon source at 30 °C was induced by addition of 0.1 mM IPTG, 1mM indole, 1mM 3-hydroxy anthranilic acid and 1mM L-proline after cell density reached OD₆₀₀=0.6. After incubation for 16h the filtrated supernatant was forwarded to LCMS analysis. Synthetic biology studies were carried out by substituting indole or 3-hydroxy anthranilic acid with the respective derivative. Initial fermentation was carried out in 10 mL volume. Feeding studies were carried out in 150 μ L volume using a 96-well flat bottom plate.

Protein Purification of NpsA, ThdA, and NpsB

The respective gene was inserted into a petM44 expression vector with an N-terminal 6xHis-MBP tag using according primers (Table S 3.1). The respective construct was expressed in *E. coli* BL21 (DE3) cells (500 mL LB, 0.1 mM IPTG, 16 °C, 16h). The cell pellet was resuspended in lysis buffer (150 mM NaCl, 25 mM Tris, 40 mM Imidazole, 1 mM TCEP, pH 7.5), sonicated and centrifuged. The supernatant was loaded onto a gravity flow column containing Ni-NTA loaded sepharose, washed with lysis buffer, and subsequently eluted in one step (250 mM Imidazole). The tag was cleaved using HRV3C protease during overnight dialysis against SEC buffer (150 mM NaCl, 25 mM Tris, pH 7.5) at 4 °C, and removed by a second Ni-NTA chromatography step. After passing through a Superdex 200 16/60 pg column (GE Healthcare Life Sciences) the protein was concentrated using a 30 kDa (NpsA and NpsB) or 10 kDa (ThdA) cutoff filter and stored at -80 °C in 10% glycerol. Protein purity was determined by SDS-PAGE. Protein concentration was determined spectrophotometrically upon determining the respective extinction coefficient from the amino acid sequence using the PROTPARAM webserver (<http://web.expasy.org/protparam/>).

Purification of TnaA

The respective gene was inserted into pET-28b with an N-terminal 6xHis tag for purification and without tag for heterologous expression using according primers (Table S 3.1). The construct was expressed in *E. coli* BL21 (DE3) (500 mL LB, 0.1 mM IPTG, 16 °C, 16h). The cell pellet was resuspended in lysis buffer, sonicated and centrifuged. The supernatant was loaded onto a gravity flow column containing Ni-NTA loaded sepharose, washed with lysis buffer and subsequently eluted in one step (250 mM Imidazole). 6xHis-fusion protein was passed through a Superdex 200 16/60 pg column in SEC buffer (150 mM NaCl, 25 mM Tris, pH 7.5), concentrated using a 10 kDa cutoff filter and stored at -80 °C in 10% glycerol. Protein purity was determined by SDS-PAGE. Protein concentration was determined spectrophotometrically as described above.

***In Vitro* Reconstitution of tilivalline NRPS**

NpsA, ThdA and NpsB (0.4 μ M each) were incubated with the respective cofactors (1 mM ATP, 0.5 mM NADPH) and substrates (0.5 mM L-Proline, 3-hydroxy anthranilic acid, indole) in reaction buffer (150 mM NaCl, 10 mM MgCl₂, 50 mM Tris-HCl, pH 7.5) in a total volume of 20 μ L at room temperature for 4h. The mixture was dried in a vacuum condenser and subsequently resuspended in 20 μ L water:acetonitrile (1:1), centrifuged and submitted to LC-MS analysis.

***In Vitro* Reconstitution of TnaA**

4 μ M TnaA was incubated with 0.5 mM tryptophan and 0.1 mM (or 1.0 mM) PLP in reaction buffer (150 mM NaCl, 50 mM Tris-HCl, pH 7.5) in a total volume of 20 μ L for 2h at 37 °C. The mixture was dried in a vacuum condenser and subsequently resuspended in 20 μ L water:acetonitrile (1:1), centrifuged and submitted to LC analysis. Equimolar standards of indole and tryptophan were used to identify the UV-peaks.

***In Vitro* Reconstitution of tilivalline NRPS and TnaA**

2 μ M TnaA and 0.4 μ M each of NpsA, ThdA and NpsB were incubated with the respective cofactors (0.1 mM PLP, 1 mM ATP, 0.5 mM NADPH) and substrates (0.5 mM Prolin, 3-OH-anthranilic acid and tryptophan) in reaction buffer (150 mM NaCl, 10 mM MgCl₂, 50 mM Tris-HCl, pH 7.5) in a total volume of 20 μ L at 37 °C for 2.5h. The mixture was dried in a vacuum condenser and subsequently resuspended in 20 μ L water:acetonitrile (1:1), centrifuged and submitted to LC-HRMS analysis.

DNA displacement assays

DNA displacement assays using either ethidium bromide or Hoechst33342 was conducted as described previously.⁴¹

Inhibition of tilivalline production in *K. oxytoca* isolate #6

Cryo-cultures of *K. oxytoca* isolate #6 were streaked out on TSA-S plates and incubated at 37 °C overnight and subsequently stored at 4 °C until use. Inhibition of tilivalline production was screened in 8 mL TSB media inoculated from single colonies in flasks without baffles. The inhibitor was added to its final concentration using DMSO stock solutions. Cultures were incubated for 16 h at 37 °C and 100 rpm. 10 mL ethyl acetate was added and shaking was continued for additional 30 minutes. 1 mL of the organic phase was dried and resuspended in 40 μ L acetonitrile:water (1:1), centrifuged and submitted to LCMS analysis. Experiments were carried out in triplicates.

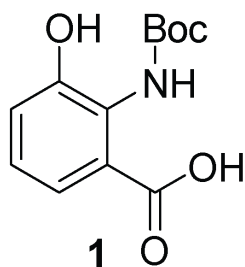
Analytical methods

All measurements regarding the tilivalline inhibition test were performed on a Dionex Ultimate 3000 RSLC system using a Waters BEH C18, 50 x 2.1 mm, 1.7 μ m dp column. Separation of 4 μ L sample was achieved by a linear gradient with (A) H₂O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 600 μ L/min and 45 °C. The gradient was initiated by a 1 min isocratic

step at 5 % B, followed by an increase to 95 % B in 6 min to end up with a 1.5 min step at 95 % B before reequilibration under the initial conditions. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 75 μ L/min before entering the solariX XR (7T) FT-ICR mass spectrometer (Bruker Daltonics, Germany) using the Apollo ESI source. In the source region, the temperature was set to 200 °C, the capillary voltage was 4500 V, the dry-gas flow was 4.0 L/min and the nebulizer was set to 1.1 bar. After the generated ions passed the quadrupole with a low cutoff at 150 m/z they were trapped in the collision cell for 100 ms and finally transferred within 1.0 ms through the hexapole into the ICR cell. Captured ions were excited by applying a frequency sweep from 100 to 1600 m/z and detected in broadband mode by acquiring a 489 ms transient. All other measurements were performed on the same setup by applying one of the following gradients but with the same MS method. The gradient was either initiated by a 1 min isocratic step at 2 or 5 % B, followed by an increase to 95 % B in 9 min to end up with a 1.5 min step at 95 % B before reequilibration under the initial conditions.

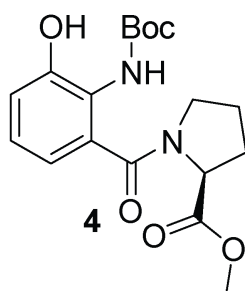
Synthesis

2-N-boc-amino-3-hydroxybenzoic acid (1):



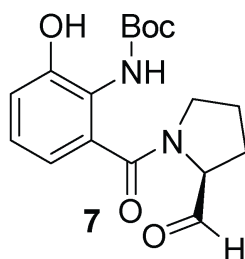
To a stirred solution of 2-amino-3-hydroxybenzoic acid (50 mg, 0.3 mmol) in acetonitrile (20 ml) at room temperature, a mixed solution of di-tert-butyl dicarbonate (98 mg, 0.45 mmol), saturated solution of sodium bicarbonate (NaHCO_3) (19.6 ml) and acetonitrile (10 ml) was added. Following addition, the mixture left stirring for 15 h at room temperature. When no trace of starting material remained (TLC: silica on alumina, UV, ninhydrin, hexane : ethyl acetate : formic acid :: 1:1:0.1 Rf 0.17), solvent was removed under reduced pressure resulting in white sticky foam that was partitioned between ethyl acetate (40 ml) and 0.1M HCl (30 ml). The organic layer was washed with 0.1M HCl (30 ml) five times. The organic extract was then dried over sodium sulfate, filtered off, and concentrated under reduced pressure to yield pale yellow oil (1). The compound was used without any further purification. HRMS (ESI, +ve) $\text{C}_{12}\text{H}_{15}\text{NO}_5$ $[\text{M}+\text{H}]^+$ calculated for 254.1023, found 254.1027.

2-N-boc-amino-3-hydroxybenzoyl- L-proline methyl ester (4):



2-N-boc-amino-3-hydroxybenzoic acid (1) (50 mg, 0.18 mmol), HATU (68.4 mg, 0.18 mmol) and HOBT (5 mg, 0.03 mmol) was dissolved in anhydrous DMF (1 ml) under nitrogen atmosphere. DIPEA (61 μ L, 0.36 mmol) was added to the mixture and the solution was left stirring at room temperature for 5 minutes under nitrogen atmosphere. A mixture of L-proline methyl ester hydrochloride (30 mg, 0.18 mmol) and DIPEA (61 μ L, 0.36 mmol) in anhydrous DMF (1 ml) was added to the latter solution and the reaction was left stirring for 16 hours under nitrogen atmosphere at room temperature. The reaction mixture was then diluted in ethyl acetate (80 ml) and 0.1M HCl (30 ml). The organic layer was washed with 0.1M HCl (30 ml) five times. The organic extract was then dried over sodium sulfate, filtered off, concentrated under reduced pressure, and purified using flash chromatography (silica, ninhydrin/UV, hexane: ethyl acetate:: 4:1; 2:1) to afford 2-N-boc-amino-3-hydroxy benzoyl- L-proline methyl ester (2) as a light yellow powder. ^1H NMR (500 MHz, CDCl_3) 9.14 (br, 1H), 8.58 (br, 1H), 7.09 (d, $J = 8.3$ Hz, 1H), 7.08 (m, 1H), 6.95 (m, 1H), 4.56 (dd, $J = 8.5, 5.3$ Hz, 1H), 3.73 (s, 3H), 3.57-3.48 (m, 2H), 2.37-2.28(m, 1H), 2.09-1.94 (m, 3H), 1.92-1.88(m, 1H), 1.52(s, 9H); ^{13}C NMR (126 MHz, CDCl_3) 172.4, 169.0, 153.2, 149.9, 124.4, 122.2, 119.4, 122.4, 118.2, 82.5, 68.6, 57.8, 52.4, 31.5, 28.2, 24.5; HRMS (ESI, +ve) $\text{C}_{18}\text{H}_{24}\text{N}_2\text{O}_6$ $[\text{M}+\text{H}]^+$ calculated for 365.1707, found 365.1699.

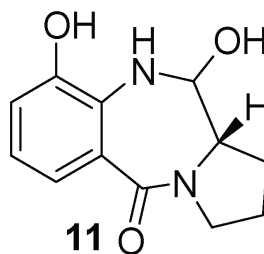
S-1- (2-N-boc-amino-3-hydroxybenzoyl)pyrrolidine-2-carbaldehyde (7):



To a stirred solution of 2-N-boc-amino-3-hydroxybenzoyl- L-proline methyl ester (4) (30 mg, 0.08 mmol) in anhydrous ether (10 ml) at -78°C , a solution of 1 M DIBAL hydride in hexane (70 μ L, 0.08 mmol) was added over a period of 1 h under an atmosphere of nitrogen. The mixture was left stirring at -78°C for 5 h under an atmosphere of nitrogen. When no trace of starting material remained (TLC: silica, ninhydrin/UV, hexane : ethyl acetate:: 3:1, R_f 0.45), the reaction was quenched with cold methanol (1 ml). A white precipitate was formed. Reaction mixture was

diluted with ethyl acetate (50 ml) and washed with saturated solution of sodium bicarbonate (20 ml) three times. The organic layer was then dried over sodium sulfate, filtered off, concentrated under reduced pressure and purified using flash chromatography (silica, ninhydrin/UV, hexane: ethyl acetate:: 5:1; 4:1) to afford S-1- (2-N-boc-amino-3-hydroxybenzoyl)pyrrolidine-2-carbaldehyde (7). ^1H NMR (500 MHz, CDCl_3) 9.68 (s, 1H), 9.22 (br, 1H), 8.50 (br, 1H), 7.78 (m, 1H), 7.17 (m, 1H), 7.10 (m, 1H), 4.28 (dd, $J = 8.5, 5.3$ Hz, 1H), (m, 1H), 3.71-3.50 (m, 2H), 2.45-2.30(m, 1H), 2.25-2.1 (m, 3H), 2.04-1.95(m, 1H), 1.52(s, 9H); ^{13}C NMR (126 MHz, CDCl_3) 198.2, 168.6, 155.6, 153.4, 135.2, 124.9, 123.7, 119.1, 113.5, 80.1, 66.3, 55.6, 51.3, 31.9, 28.3, 24.2; HRMS (ESI, +ve) calculated for $\text{C}_{17}\text{H}_{22}\text{N}_2\text{O}_5$ $[\text{M}+\text{H}]^+$ calculated for 335.1605, found 335.1609.

Carbinolamine X (11):



S-1- (2-N-boc-amino-3-hydroxybenzoyl)pyrrolidine-2-carbaldehyde (10 mg, 0.03 mmol) was dissolved in ethyl acetate (2 ml) and 1 M HCl (2 ml) was added while stirring vigorously. The mixture was left stirring at high speed for 16 hours at room temperature. The reaction mixture was then concentrated under reduced pressure and lyophilized to yield a yellow powder (11). Compound was used without any further purification. HRMS (ESI, +ve) calculated for $\text{C}_{17}\text{H}_{22}\text{N}_2\text{O}_5$ $[\text{M}+\text{H}]^+$ calculated for 235.10775, found 235.10774.

Tillivaline:

A mixture of carbinolamine X and indole were dissolved in anhydrous THF and TMS-Cl was added in the presence of 1% formic acid. Reaction mixture was refluxed for under nitrogen for 16 hours. Then compound was purified by preparative HPLC (same method as for initial purification). HRMS (ESI, +ve) calculated for $\text{C}_{20}\text{H}_{20}\text{N}_3\text{O}_2$ $[\text{M}+\text{H}]^+$ calculated for 334.1550, found 334.15493. For NMR data see Table S 3.4.

3.6 Supporting Information

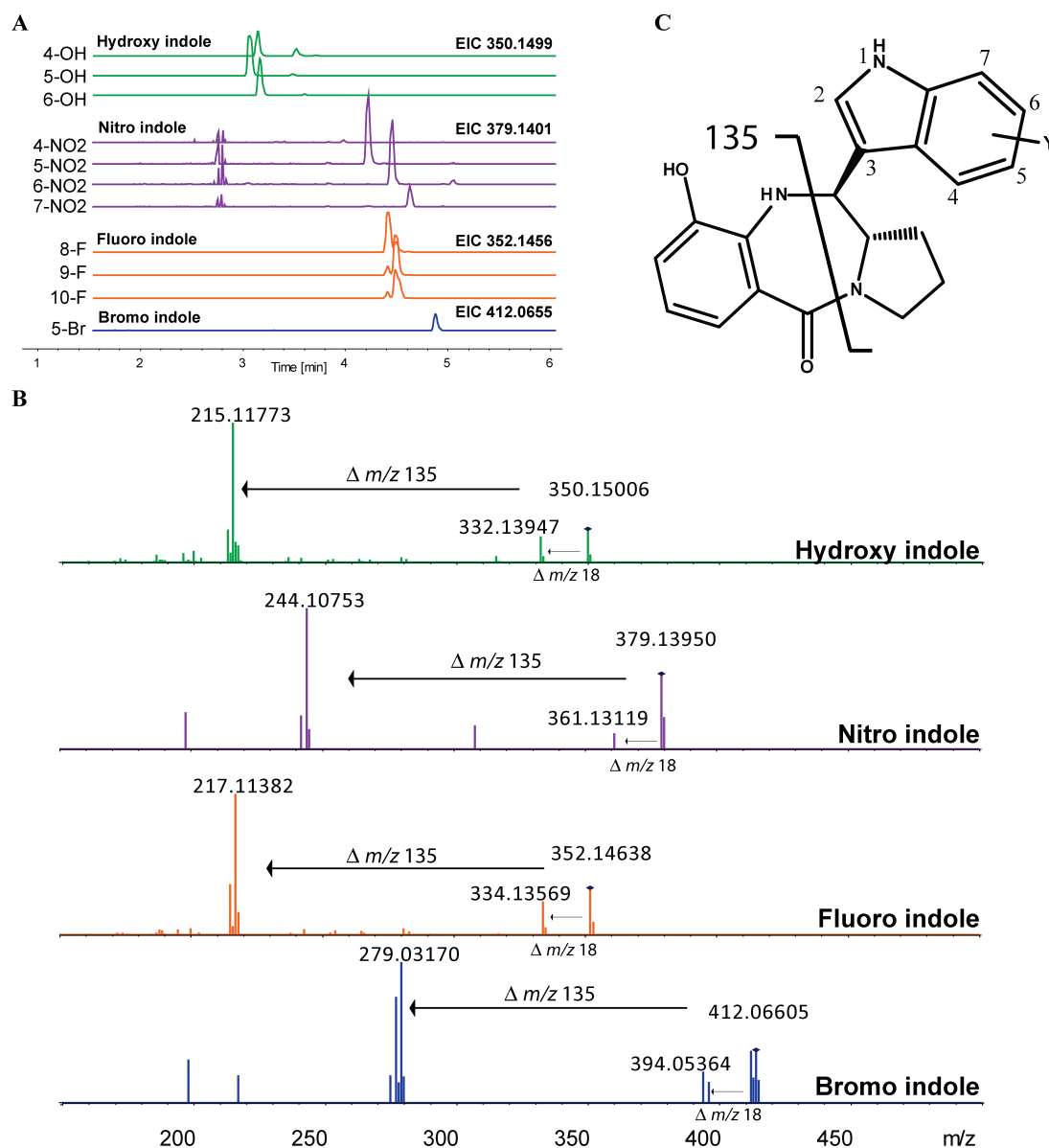


Figure 3.7: LC-MS/MS data of tilivalline derivatives derived by supplementing indole derivatives to the heterologous expression system. A Listed EICs of the expected $[M+H]^+$ -ions with a m/z width of 0.005 to prove the incorporation of the respective indole derivatives. B Exemplary fragmentation spectra of tilivalline derivatives containing a hydroxy-, nitro-, fluoro-, and bromo-substituent on the indole respectively. All spectra show two characteristic fragments caused by neutral losses of 18 and 135 Da. C Chemical structure of tilivalline derivatives with the proposed main fragmentation pathway.

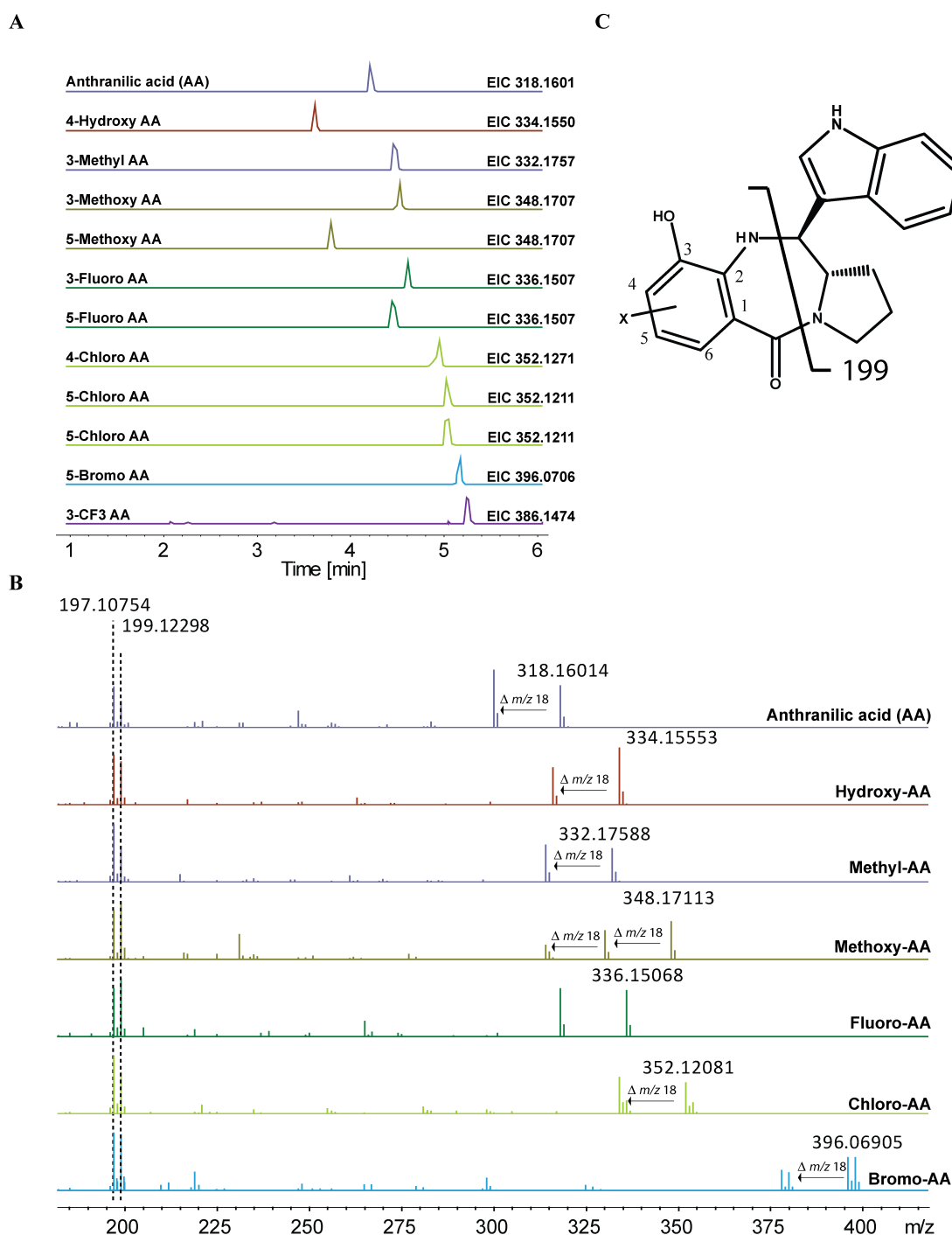


Figure 3.8: LC-MS/MS data of tilivalline derivatives derived by supplementing anthranilic acid derivatives to the heterologous expression system. A Listed EICs of the expected $[M+H]^+$ -ions with a m/z width of 0.005 to prove the incorporation of the respective anthranilic acid derivatives. B Exemplary fragmentation spectra of tilivalline derivatives which incorporated anthranilic acids without substituent, or with a hydroxy-, methyl-, methoxy-, fluoro-, chloro-, and bromo-substituent respectively. All spectra show two characteristic fragments at 197 and 199 m/z and at least one fragment caused by a neutral loss of 18 Da. C Chemical structure of tilivalline derivatives with the proposed main fragmentation pathway.

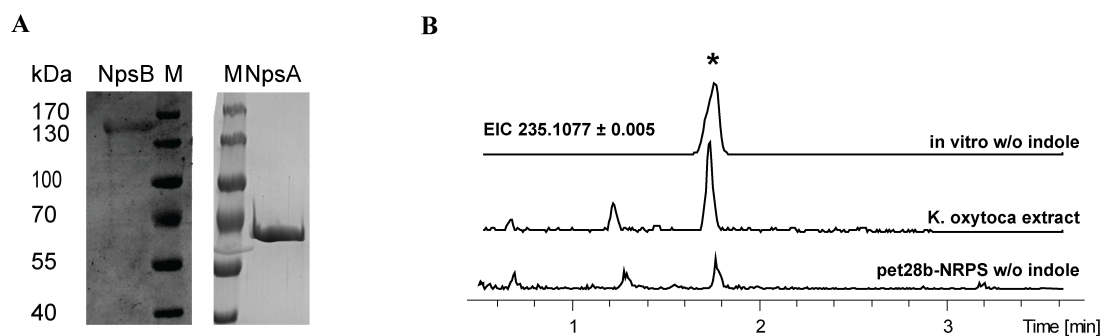


Figure 3.9: A SDS-PAGE analysis of NpsA (163 kDa) and NpsB (56 kDa). B PBD-precursor formation can be detected in the in vitro reconstitution of NpsA, ThdA and NpsB when indole is omitted. Low amounts can also be detected in the *K. oxytoca* isolate #6 extract and the *E. coli* BL21 (DE3) pET-28b-NRPS fermentation in M9 minimal media without indole.

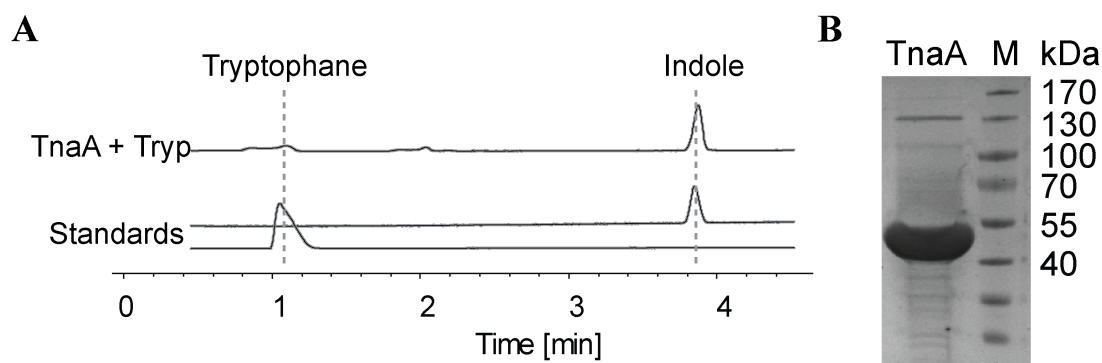


Figure 3.10: A In vitro conversion of L-tryptophane to indole by the *K. oxytoca* isolate #6 tryptophanase TnaA is indicated by LC analysis (200-600nm). L-Tryptophane and indole were used as a standard.

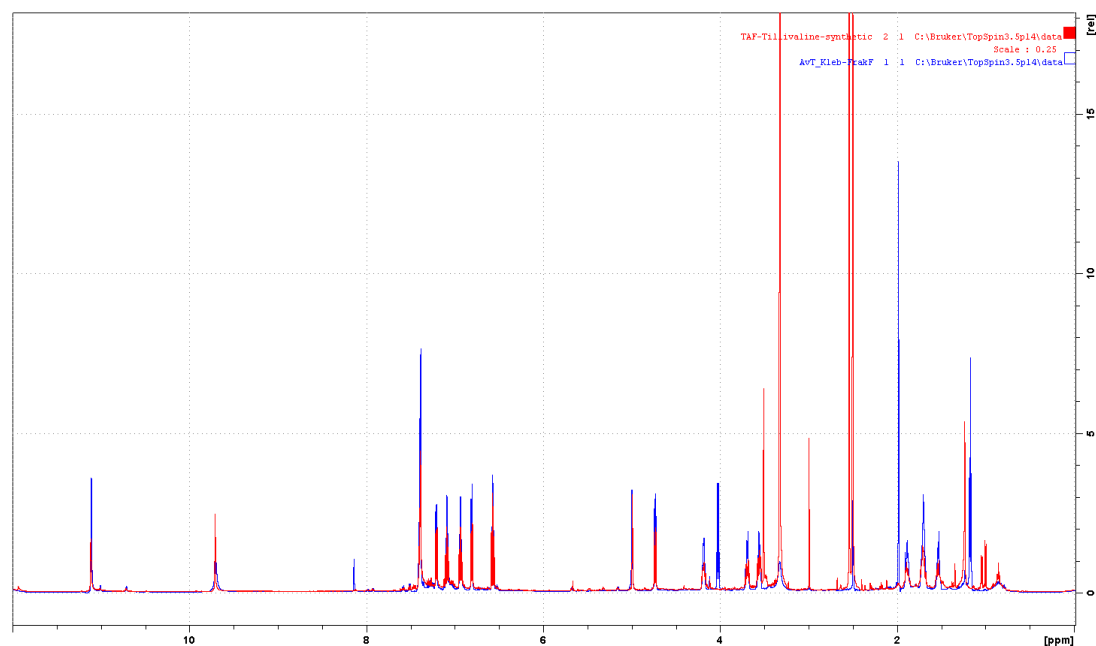


Figure 3.11: Comparison of ^1H NMR (500 MHz, DMSO) shifts of synthetic and isolated tilivalline to confirm the stereochemistry.

Table 3.1: Oligonucleotides used in this study

Primer	Sequence
NpsA-FP	AAAAACCATGGCAACGCATTTCAGCATA
NpsA -RP	AAAAAAAGCTTTTACACCTGCTCCAGTAAAGAATTT
ThdA-FP	AAAAACCATGGCAGACAACGTTGAGCAA
ThdA -RP	AAAAAAAAGCTTTTAGAGCTTTTGCCGTTGCC
NpsB-FP	AAAAACCATGGCACCATCATTCAATAGCG
NpsB-RP	AAAAAGGATCCCTAGACAATTTCTTTCTGCCGCT
TnaA-noHis-FP	AAAAAACCATGGCACGTATCCCTGAGCCGT
TnaA-RP	AAAAAAGGATCCTTATTTTCACCGGTTTTCAGTCTG
TnaA-FP	AAAAAAGCTAGCATGAAACGTATCCCTGAGC

Table 3.2: Primer sequence for TAR-assembly of the tilivalline biosynthetic gene cluster from *K. oxytoca* isolate #6 and colony PCR.

	Sequence
TAR Cloning	
TILTAR-capture-Para-RP	AAAATCATACCTGACCTCCATAGCAGAAAGTCAAAAGCCTTCGCCCTTTGACGTTGGAGT
TILTAR-NRPS -RP	CCAGCTTATCCGATGACTTC
TILTAR-NRPS-FP	ATGACGCATTTCAGCATATGT
TILTAR-Para-FP	AGGCTTTTGACTTTCTGCTA
TILTAR-Para-OL-NRPS-RP	GGACAGCCTTTAACTGATAGACATATGCTGAATGCGTCATGGGTATATCTCCTTCTTAAA
TILTAR-Para-RP	GGGTATATCTCCTTCTTAAA
TILTAR-Ptet-FP	TACGAACGGTATTAAGACCC
TILTAR-Ptet-OL-NRPS-FP	TATATTTCCGGCAGTGATGGAGAAGTCATCGGATAAGCTGGTACGAACGGTATTAAGACCC
TILTAR-Ptet-OL-Tail-RP	GCTGTTGCCAGTTGGACTTATTGCCAGGAAATGATGTCATTAGTGCCTCTTCTCTATCA
TILTAR-Ptet-RP	TTAGTGCCCTCTTCTCTATCA
TILTAR-Tail-noP-FP	ATGACATCATTTCTGGCAA
TILTAR-Tail-RP	TTATTCTGTCTAAAGGCCAGT
Colony PCR	
Check-KProm-ara-NRPS-FP	TTGCATCAGACATTGCCGTCAC
Check-KProm-ara-NRPS-FP	AATTCGCCGTTCTTTGTGCGT
Check-KProm-Ende-FP	TTCGCATGCCAAAGATTTCGCT
Check-KProm-Ende-RP	TTTCGCCACCACTGATTGAGC
Check-KProm-Start-FP	ACGATACCTGAGTATTTCCACAGT
Check-KProm-Start-RP	GCTGTGGCGATAAATCTGCTT
Check-KProm-Tet-Tail-FP	AGCACATCTAAACTTTAGCGT
Check-KProm-Tet-Tail-RP	CTTCGGGCCGGTTGACTGAATC

Table 3.3: Sequence homologies of tilivalline biosynthetic gene cluster found in *K. oxytoca* isolate #6 compared to the Schneditz et al. cluster and conferred from BLAST.

Gene	Identity* / %	Putative Function	Size / bp	Homolog	Identity / %	BLAST	
						E-value	Score
mfsX	100.0	Uncharacterized MFS-type transporter	410	ydgK	34	1,00E-69	590
uvrX	96.7	UvrABC system protein A	740	uvrA	39	2,00E-137	1117
hmoX	99.8	4-nitrophenol-monooxygenase, oxy. component	505	NphA1	62	0	654
adsX	98.8	Anthranilate synthase, phenazine specific	649	phzE	42	1,00E-149	1169
icmX	100.0	2,3 dihydro-2,3 dihydroxybenzoate synthase	210	phzD	50	1,00E-63	515
dhbX	100.0	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase	261	dhbA	46	6,00E-68	533
aroX	99.5	phospho-2-dehydro-3-deoxyheptonate aldolase	389	phzC	42	5,00E-90	725
npsA	98.8	NRPS [A]	506				
thdA	97.4	NRPS [T]	76				
npsB	99.5	NRPS [C-A-T-R]	1456				

* Pairwise identity with Schneditz et al. Cluster

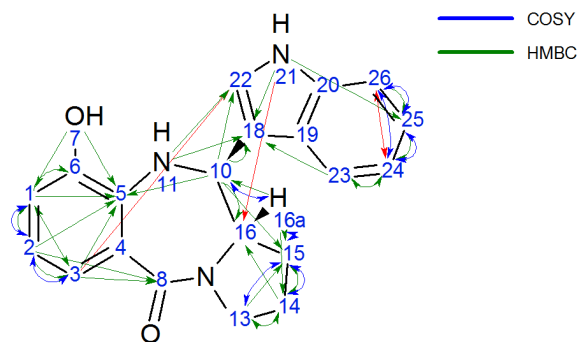


Figure 3.12: Tilivalline structure for NMR reference.

Table 3.4: NMR data for tilivalline. Atom numbers are defined in figure 3.12.

#	Atom	C			H			2D		
		Label	Shift	XHn	Label	Shift	H Multiplicity	COSY	H HMBC	C HMBC
1	14	C 12	21647	CH2	H 13	1887	m	15, 14, 15	15, 14, 15, 13, 16a	15, 13, 16
2	15, 14	C 12	21647	CH2	H 14	1700	m	14	15, 13, 16a	15
3	15	C 13	29842	CH2	H 15	1702	br s	14, 13, 16a	14, 13, 16a, 10	14
4		C 13	29842	CH2	H 16	1528	m	15, 14, 16a	15, 14, 13, 16a, 10	14, 13, 16
5	13	C 11	47109	CH2	H 12	3556	m	13	14	14, 15
6	13	C 11	47109	CH2	H 11	3693	m	15, 13	14	
7	16a	C 10	58950	CH	H 10	4187	m	15, 10	14, 10, 21	14, 15, 10
8	10	C 9	60394	CH	H 9	4734	d (9.17)	16a	16a	15, 16, 2, 18, 22, 5
9	23	C 1	111370	CH	H 3	7391	br s	24	2, 18, 24	
10	1	C 7	114560	CH	H 7	6809	dd (7.61, 1.47)	2	2, 3, 7	3, 5, 6
11	2, 18	C 8	115962	C, CH	H 8	6571	m	1, 3	10, 23, 21, 11	1, 3, 5, 8
12	26	C 6	118555	CH	H 2	7403	m	24, 25	25	
13	24	C 2	118555	CH	H 6	6938	m	25, 26	25, 23	23
14		C 20	119758	C	2					
15	25	C 5	121001	CH	H 5	7092	m	24, 26	26, 21	24
16	3	C 4	121244	CH	H 4	7207	m	2	2, 1	1, 22, 5, 8
17	22	C 3	123298	CH	H 1	7391	m	10, 3, 11		
18		C 16	124827	C	24, 23, 11					
19	5	C 17	134301	C	10, 2, 1, 3, 7					
20		C 15	136099	C	25, 23, 11					
21		C 19	144734	C	2, 21, 7					
22	6	C 18	145334	C	1					
23	8	C 14	165960	C	2, 3					
24	7			OH	H 20	9710	s	1, 5		
25	H 18	3694	m	14	14, 15					
26	21			NH	H 19	4991	s	16, 2, 18, 25		
27	11			NH	H 17	11123	br s			2, 18, 22

Chapter 4

Biosynthesis of Corramycin

ALEXANDER VON TESMAR, MICHAEL HOFFMANN, RAM PRASAD AWAL, NESTOR ZABURANNYI, KIRSTEN HARMROLFS, JOACHIM WINK, STEPHANE RENARD, ROLF MÜLLER*

*To whom correspondence should be addressed.

Author's efforts

The author was responsible for the conception of the study, designed and performed experiments, evaluated and interpreted resulting data. The author performed the presented analysis of the corramycin biosynthetic gene cluster. Furthermore, the author designed and performed all experiments to elucidate the biosynthesis, the putative predrug mechanism and the gyrase inhibition assay. The author identified the described novel corramycin derivatives. The author conceived and wrote the manuscript.

Contributions by others

Nestor Zaburannyi refined the raw sequence data. Ram Awal isolated the strain *Corallococcus coralloides* MCy10984. Michael Hoffmann purified the corramycin derivative 1348. Structural elucidation of corramycin derivative 1348 was performed by Kirsten Harmrolfs. Stephan Renard (Sanofi) provided corramycin and background information. Joachim Wink provided the producer strain *Corallococcus coralloides* ST201330. The project was supervised by Rolf Müller.

Abstract

Corramycin is a peptide antibiotic produced by *Corallococcus coralloides* and initially isolated by Sanofi in an bioactivity-guided approach. Genome sequencing of two producer strains, *C. coralloides* ST201330 and MCy10984, enabled the identification of the corresponding biosynthetic gene cluster. Corramycin is produced by a 12-modular NRPS-PKS hybrid assembly line and feeding of isotope labeled precursors was used to establish a putative biosynthesis model. Interestingly, the assembly line harbors a fatty-acid activating FAAL domain in the starter module. *In vitro* reconstitution of fatty acid activation by FAAL indicates the presence of a pre-drug mechanism, involving acylated corramycin. Furthermore, several derivatives of corramycin including hydroxylated and glycosylated ones were identified.

4.1 Introduction

Bacterial natural products remain a main resource in the drug discovery process.¹ Since cost-efficient sequencing methods made thousands of bacterial genomes available it became evident that the genetic potential for novel natural product classes is far from being depleted.² Regularly, the analysis of bacterial genomic data with tools like Antismash³ or Clusterfinder⁴ identify more biosynthetic gene clusters than assigned natural products. The hunt for novel scaffolds, hampered by the need for elaborate de-replication methods and the apparent inactivity of present gene clusters, has turned to lesser studied phyla of bacteria from exotic habitats.

Myxobacteria are gram-negative δ -proteobacteria exhibiting a complex life-cycle and advanced multicellular social behavior, inhabiting almost every habitat, including soil, deep-sea sediments, sweet water and hydrothermal vents.⁵ The order myxococcales produces a large quantity of natural products with remarkable chemical diversity.^{6, 7} The unusual large genomes, which are between 9-12 Mbp in size and thus belong to the largest known in the bacterial domain, contain an average of 8-10% of biosynthetic genes in 10-20 different clusters and therefore exhibit a greater potential for natural products than actinomycetes.⁸⁻¹⁰

The rise of bacterial resistance against the established classes of antibiotics make the discovery of novel scaffolds and the subsequent development into marketable antibiotics inevitable. Especially the pathogens of the ESKAPE group, namely *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumonia*, *Actinobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacter species*, have acquired resistance against several if not all last resort antibiotics available. The insight that resistance is only a question of time and all antibiotics therefore have a limited life-span induces a constant need for novel therapeutics to successfully counter infectious diseases.¹¹⁻¹³

Corramycin was initially isolated by Sanofi from *Coralloccoccus coralloides* ST201330 in an activity guided approach after the extracts showed antibacterial activity against *E. coli*. Structural elucidation by NMR revealed a peptidic compound with 1184.56 *m/z*. It contains eight α -amino acids including the up to date unknown δ -*N*-methyl- β -hydroxy histidine. The linear structure also harbors an unusual "aminosugar" moiety and the N-terminus is capped by dihydroxy butyric acid. Corramycin exhibits bioactivity against several Gram-negative pathogens including the multidrugresistant *E. coli* ATC35218, *E. cloacae*, *K. pneumoniae*, *A. baumannii* and *S. typhimurium* with MIC values ranging from 2-64 $\mu\text{g/mL}$. Sanofi generates corramycin either by optimized fermentation conditions at 9.5 mg/L or total synthesis.¹⁴

Here, we report the identification of the corramycin biosynthetic gene cluster spanning 55 kb and harboring a 12-modular NPRS/PKS assembly line.

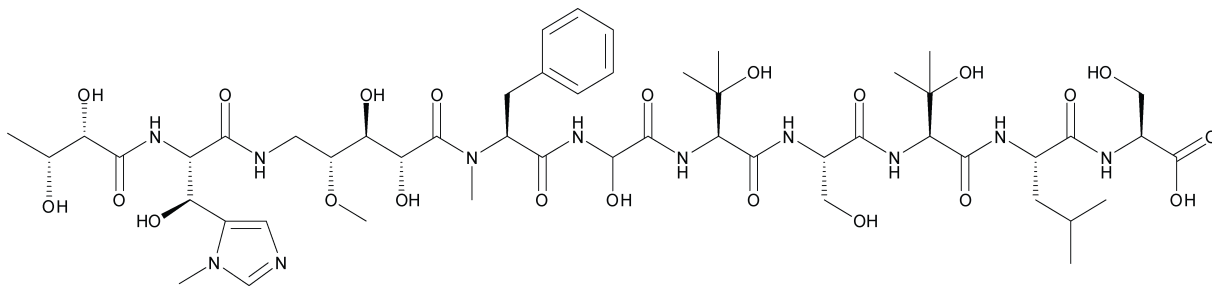


Figure 4.1: Structure of corramycin.

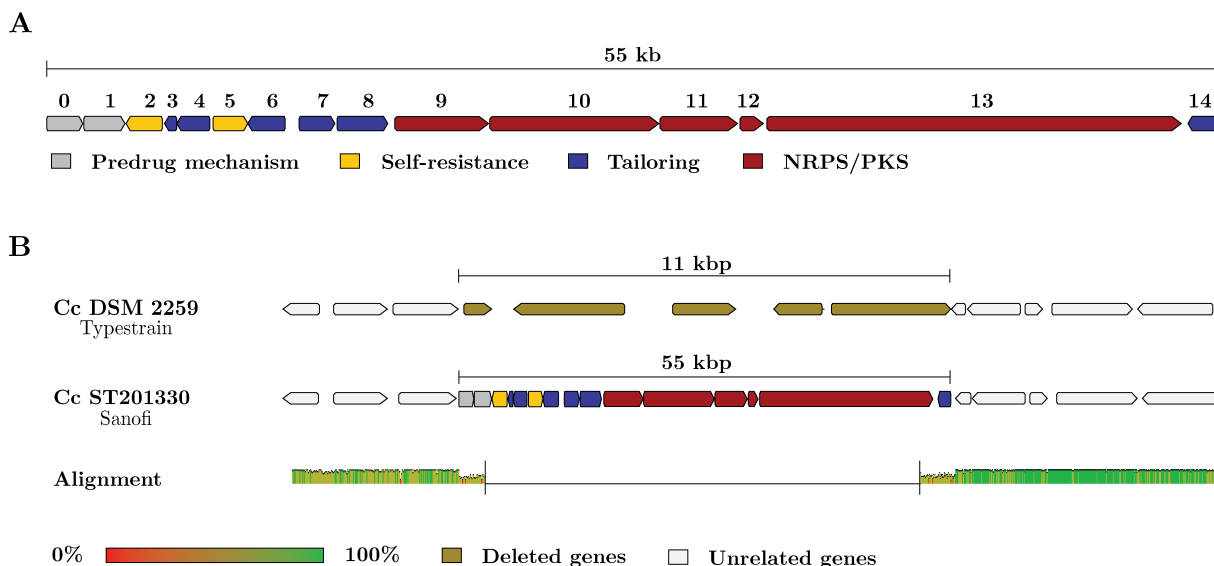


Figure 4.2: **A** The biosynthetic gene cluster of corramycin stretches over 55 kb and contains 14 genes responsible for NRPS/PKS, tailoring, self-resistance and the predrug mechanism. **B** Sequence alignment with the type strain DSM2259 reveals the cluster borders. Pairwise identity is given.

4.2 Results

Genome sequencing of *Corallococcus coralloides* ST201330 reveals corramycin biosynthetic gene cluster

Illumina sequencing of the genomic DNA of *Corallococcus coralloides* ST201330 produced 26 scaffolds with a total length of 4 772 480 bp. Antismash analysis³ of the genomic data revealed a locus harboring three NRPS modules including an *N*-methyltransferase domain with the predicted specificity of Phe-nMT-Gly-Val, representing a signature motif of corramycin. Upon closer analysis of the downstream region, additional four modules matching the corramycin structure could be identified. Since scaffold 6 ended after the proposed module seven, other scaffolds were analyzed to identify the missing part of the assembly line. A potential termination module with the predicted specificity for serine was found on the edge of scaffold 3. PCR based re-sequencing of the gap and subsequent manual assembly was employed to gather the complete sequence information of the corramycin cluster.

Five genes build up the corramycin assembly line containing a total of 12 modules. To determine the cluster border the genetic surrounding was aligned to the *Corallococcus coralloides* type strain DSM2259. A well defined insertion of a 55 kb stretch in *Corallococcus coralloides* ST201330 harboring a total of 15 ORFs represents the corramycin biosynthetic gene cluster (Figure 4.2). The first module of the assembly line is a fatty-acyl AMP ligase (FAAL), a member of the adenylate forming enzymes. FAALs are often associated with NRPS/PKS pathways as a starter unit catalyzing the acylation of the first building block.¹⁵ Sequence alignments using Clustal Ω ¹⁶ and BLAST analysis¹⁷ revealed a substrate specificity for long-chain fatty acids, ranging from 5-12 C-atoms in length. This finding is inconsistent with the final structure of corramycin that omits any fatty acid moiety. Nevertheless, all active sites of the corramycin FAAL seem to be present (see figure 4.10). Consequently, the fatty acid is thought to be incorporated in corramycin and removed in later steps of the biosynthesis.

The second module harbors a Fkbh-like domain, which has been shown to activate glycerate for incorporation in bryostatin¹⁸ and vioprolid. It belongs to the HAD-superfamily and is also found in other biosynthetic pathways, for example pellasoren and tautomycin.^{19, 20} BLAST analysis and sequence alignment with all characterized members of this family (Figure S 4.11) proved the presence of all significant motifs: The first motif harbors the nucleophilic aspartate, that facilitates the dephosphorylation of the D-1,3-bisphospho-glycerate substrate, after it was bound to the cysteine in motif two as thioester. The resulting glycerate building block is subsequently transferred to the downstream ACP. The domain structure present in the first two modules of the assembly line suggests incorporation of a fatty acid linked glycerate extender unit, resulting in a C-3 extension of the corramycin backbone. The C-4 butyric acid moiety found in the final molecule is most likely generated by an additional tailoring step of a C-methylation.

The prediction of the NRPS A-domain substrate specificities with the tool NRPSpredictor2²¹ and the Stachelhaus-code²² are in accordance with the structure (Table 4.1). Module 3 shows low identity of 60 %, caused by the activation of the unprecedented β -hydroxy histidine. β -alanine, putatively produced by the aspartate decarboxylase ORF-3, is incorporated by module 4. Module 8 and 10 incorporate γ -hydroxy valine, a common moiety in myxobacterial NRPs like myxoprimcomide.²³ Module 5 harbors a type-I PKS module. Sequence analysis with the BLAST algorithm and a conserved domain search predicted a KS-AT-KR-ACP architecture, which is consistent with the corramycin structure. Further sequence analysis of the AT domain²⁴ predicted the usual malonyl unit as substrate, resulting in a C-2 extension of the chain. Construction of a phylogenetic tree on the NaPDos online server²⁵ classified the KS domain of module 5 as hybrid type. KS domains of this type form a monophyletic group and are always situated downstream of an NRPS PCP-domain.²⁵ Consecutive action of the upstream β -alanine incorporating module 4 and the PKS of module 5 produce the observed C-5 body, termed the "sugar-moiety".

The corramycin assembly line harbors a total of 10 NRPS C-domains including one epimerisation domain (E) in module 3 that were phylogenetically analyzed on the NaPDos server. The first C-domain in module 1 groups together with other starter modules that incorporate acyl-groups. The epimerization domain of module 3 is located upstream of the D-amino acid linking

Table 4.1: Stachelhaus code extracted from the A-domains of the corramycin assembly line from *Corallococcus coralloides* ST201330. All amino acids are L-configured.

Module	Prediction			Structure
	Substrate	%	Stachelhaus	
3	Phe	60	D A G V V V A V Y K	β -OH-N-Me-His
4	β -Ala	60	G D Q I I S L C D K	β -Ala
6	Phe	60	D A Y C M V A I A K	Phe
7	Gly	90	D I L Q L G V I W K	Gly
8	Val	90	D A L W L G G T F K	γ -OH-Val
9	Ser	100	D V W H V S L V D K	Ser
10	Val	90	D A L W L G G T F K	γ -OH-Val
11	Val/Leu	80	D A Y L I G A V I K	Ile
12	Ser	100	D V W H V S L V D K	Ser

C-domain of module 4 (DCL), indicating the usual activation of an L-amino acid and the subsequent epimerization on-line. The KS domain is characterized as hybrid type, as anticipated by the localization downstream of an NRPS module (Figure S 4.8)

The assembly line includes a total of three methyltransferase (MT) domains (Figure S 4.9). Module 6 harbors an N-MT domain downstream of its A-domain that was readily annotated by standard NRPS-prediction tools and was one of the key features that led to the identification of the assembly line. A stretch of 400 amino acids in module 3 was annotated as an unknown domain, but BLAST analysis revealed a second N-MT domain architecture. The third identified MT-domain is located downstream of module 4 and was not annotated automatically by standard NRPS prediction tools. Closer examination by manual BLAST of the respective module revealed a stretch of approximately 300 amino acids as a FkbM-like methyltransferase, a family whose three analyzed members are known to be stand-alone O-MTs.²⁶ Sequence alignment and phylogenetic analysis of 54 MTs including the corramycin assembly line bound ones was employed. The MT of module 6 clusters together with NRPS associated N-MTs that have been shown to catalyze methylation of the amide nitrogen,^{27, 28} what is in accordance with the corramycin structure. The designated N-MT from module 3 does not cluster with any of the known MTs usually found within natural product gene clusters and is putatively involved in the *N*-methylation of histidine. Indeed no NRPS associated in-cis N-MT acting on an aromatic substrate is known so far. Sequence alignment of module 4 MT with the three published members of the FkbM family proves the found sequence to be complete.^{26, 29, 30} Therefore involvement in the methoxylation of the β -alanine moiety is likely.

The biosynthetic gene cluster of corramycin contains ten additional ORFs, nine upstream and one downstream of the core NRPS/PKS genes (Table 4.2). ORF-0, encoding for a hydrolase and ORF-1, a homologue of glycosyl-transferases, are probably not involved in the biosynthesis, but rather in the export and maturation process. Both harbor an N-terminal peptide signal for periplasmatic localization. The membrane bound ORF-1 might be involved in the export of the fatty acid linked acyl-corramycin to the periplasma, where hydrolysis by ORF-0 facilitates maturation to the final product (Figure 4.2). ORF-2 as well as ORF-5 could be linked to the corramycin self-resistance mechanism (see chapter 5). ORF-3 encodes an aspartate decarboxylase that provides the β -alanine for module 4. ORF-4, ORF-6 and ORF-7 encode for various hydroxylases, putatively responsible for the β -hydroxylation of histidine and the two hydroxylations of β -alanine and the PKS incorporated acetate moiety. γ -hydroxy valine is probably present in sufficient quantities in the organism making a cluster borne enzyme superfluous. ORF-8 as well as ORF-14 encode for methyltransferases. ORF-14 is homologous to usual SAM-dependent O-methyltransferases and may thus be involved in the generation of corramycin derivative 1198 (Figure 4.5). ORF-8 is part of the radical-SAM superfamily and putatively involved in the elongation of the C-3 glycerate starter unit to the observed C-4 butyric-acid by *C*-methylation.

The putative scenario for the biosynthesis of corramycin is as follows: Module 1 activates a fatty acid that is part of a pre-drug mechanism and ultimately removed. Module 2 incorporates glycerate, the actual first moiety, that is C-methylated by ORF-8. Module 3 incorporates either L-histidine or β -hydroxy histidine that is subsequently epimerized and *N*-methylated. Module 4 activates β -alanine, that is either hydroxylated before or after incorporation. The FkbM-MT

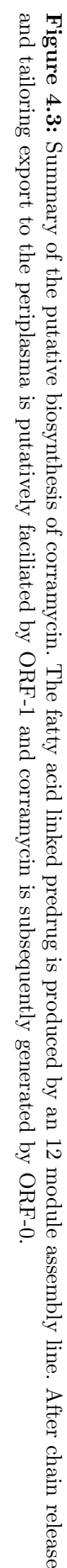


Table 4.2: Summary of the in silico analysis of all ORFs found in the corramycin biosynthetic gene cluster of *Corallococcus coralloides* ST201330

Gene	size / aa	Putative Function	Homolog	Identity	E-value
ORF-0	378	o-Phthalyl amidase / Lipase	P0C2Y0	25	$5.4 \cdot 10^{-2}$
ORF-1	424	Glycosyl-Transferase (Membrane-Protein)	arnF	31	$8.2 \cdot 10^{-1}$
ORF-2	383	DNA Replication and Repair	RecF	38	$1.0 \cdot 10^{-3}$
ORF-3	130	Aspartat-Decarboxylase	panD	78	$2.0 \cdot 10^{-68}$
ORF-4	342	Luciferase-like Monooxygenase Component	BtrO	27	$2.0 \cdot 10^{-32}$
ORF-5	364	Kinase	murA	26	$3.5 \cdot 10^{-1}$
ORF-6	388	JmjC-domain containing protein	KDM8	32	$7.0 \cdot 10^{-16}$
ORF-7	372	Limonene 1,2-monooxygenase	limB	24	$2.0 \cdot 10^{-6}$
ORF-8	524	Radical-SAM enzyme - B12 dependent	bchE	26	$8.0 \cdot 10^{-21}$
ORF-9	5379	M-1 M-2			
ORF-10	9690	M-3 M-4			
ORF-11	4554	M-5			
ORF-12	1419	M-6 C-Domain			
ORF-13	23328	M-6 M-7 M-8 M-9 M-10 M-11 M-12			
ORF-14	322	SAM-dependent O-Methyltransferase	eryG	40	$5.0 \cdot 10^{36}$

of module 4 methylates the hydroxy group. Module 5 incorporated acetate completes the heavily tailored N-terminal part of corramycin. Afterwards standard NRPS modules incorporate L-phenylalanine, glycine, γ -hydroxy L-valine, serine, γ -hydroxy L-valine, L-leucine, and L-serine before the product chain is released by the TE domain of module 12. C-methylation of the glycinate unit either takes place on-line or after product release. The fatty acid chain is putatively cleaved off by ORF-0 after export to the periplasma by ORF-1 (Figure 4.3).

Corallococcus coralloides MCy10984 produces corramycin

Standard screening methods at HIPS detected the presence of corramycin in the extract of strain *Corallococcus coralloides* MCy10984 that was isolated in Nizza, France. Illumina sequencing of the genomic DNA yielded 59 scaffolds with a total length of 1 023 083 bp. Antismash analysis revealed the presence of the corramycin gene cluster. Alignment with the genomic data of *Corallococcus coralloides* ST201330 confirmed an identical cluster organization over the 55 kbp with an overall pairwise identity of 97.7 %.

Corallococcus coralloides MCy10984 showed a more stable production than *Corallococcus coralloides* ST201330. Production was optimized in E-media using shaking flasks. Production was quantified using a calibration curve of authentic corramycin provided by Sanofi. All fermentations were supplemented with vitamin B12 that proved to be critical for corramycin production. Fermentation was carried out with and without addition of 2 % XAD. Peak production of 5 mg/L corramycin was observed after 70 h in 50 mL media. The time-point of maximum production correlates with the observed growth curve that reaches the plateau after 50 h (Figure 4.4). XAD binding capability of corramycin proved to be inefficient, with approximately 50 % of the compound remaining in the filtered supernatant. Nevertheless, overall production was increased by addition of XAD. Several other means of optimization did not improve production: The use of baffled flasks or spirals had no effect or even diminished corramycin output. Increase of the fermentation volume to 500 mL halved the production. Sanofi reported yields of up to 9.5 mg/L using complex media recipes that unfortunately hamper LC-MS analysis. E- or Amb-media with

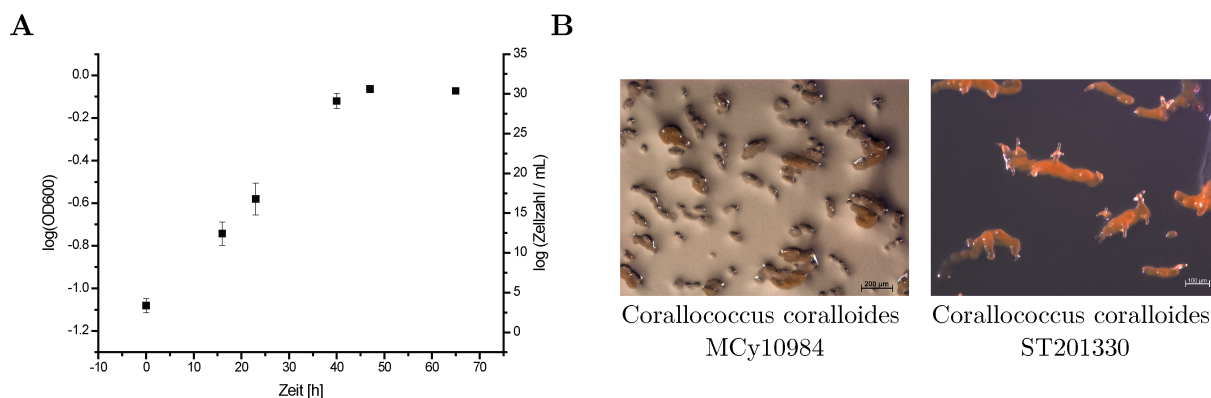


Figure 4.4: **A** Growth curve of *Coralloccoccus coralloides* MCy10984 **B** Microscopic image of *Coralloccoccus coralloides* MCy10984 and ST201330 fruiting body formation on agar plates

obtained yields of 5 mg/mL was thus used for all further experiments.

Supplementation of isotope labeled precursors confirm predicted substrates

To validate the putative biosynthetic pathway for corramycin, isotope labeled precursors were supplemented during fermentation of *Coralloccoccus coralloides* MCy10984 and analyzed using LC-MS. A total of four subsequent mass shifts by 3 Da occurred upon supplementation of d₃-methionine indicating the presence of four SAM dependent methylations in the compound. This result supports the scenario of a C-methylation at the glycerate starter unit, since only three O- and N-methylations are present elsewhere: histidine N-methylation, methylation of the β-alanine hydroxy group and N-methylation at the phenylalanine amide bond. Consequently, the methyltransferase ORF-14 is not involved in corramycin biosynthesis. Supplementation of ¹³C₄-¹⁵N-aspartate induced a mass shift of 4 Da, in accordance with the decarboxylation to β-alanine catalyzed by ORF-3 prior to assembly. Acetate incorporation by the PKS of module 5 was endorsed by the observed mass shifts upon supplementation of several labeled acetates, namely 1-¹³C-, 2-¹³C- and ¹³C₂-acetate. Incorporation of each one d₅-phenylalanine and d₂-glycine was observed. Supplementation of d₈-valine resulted in two mass shifts of 7 Da, since one deuterium is lost upon hydroxylation. No mass shift occurred when d₆-hydroxy valine or ¹³C₄-¹⁵N threonine was supplemented. Consequently, valine is the activated substrate that is hydroxylated either on-line or after product release. Two d₃-serines as well as one d₃-leucine were incorporated in accordance with the putative biosynthetic model (Table 4.3 and Figure S 4.12).

Coralloccoccus coralloides MCy10984 produces corramycin derivatives

The datasets acquired during supplementation of isotope labeled precursors together with high resolution MSⁿ analysis enabled the detection of several derivatives of corramycin (Table 4.3). Hydroxy corramycin is a non-active side product already described by Sanofi that is hydroxylated at the CH₂ group of glycine. The corresponding *m/z* signal was detected and showed the expected fragmentation and isotope mass shift patterns. Respective signals for demethylated, methylated and double hydroxylated corramycin were also identified. Interestingly, a species with 1348 *m/z* showed identical mass shifts upon supplementation of isotope labeled precursors and was present in higher intensities than native corramycin. Compound 1348 was purified from the supernatant

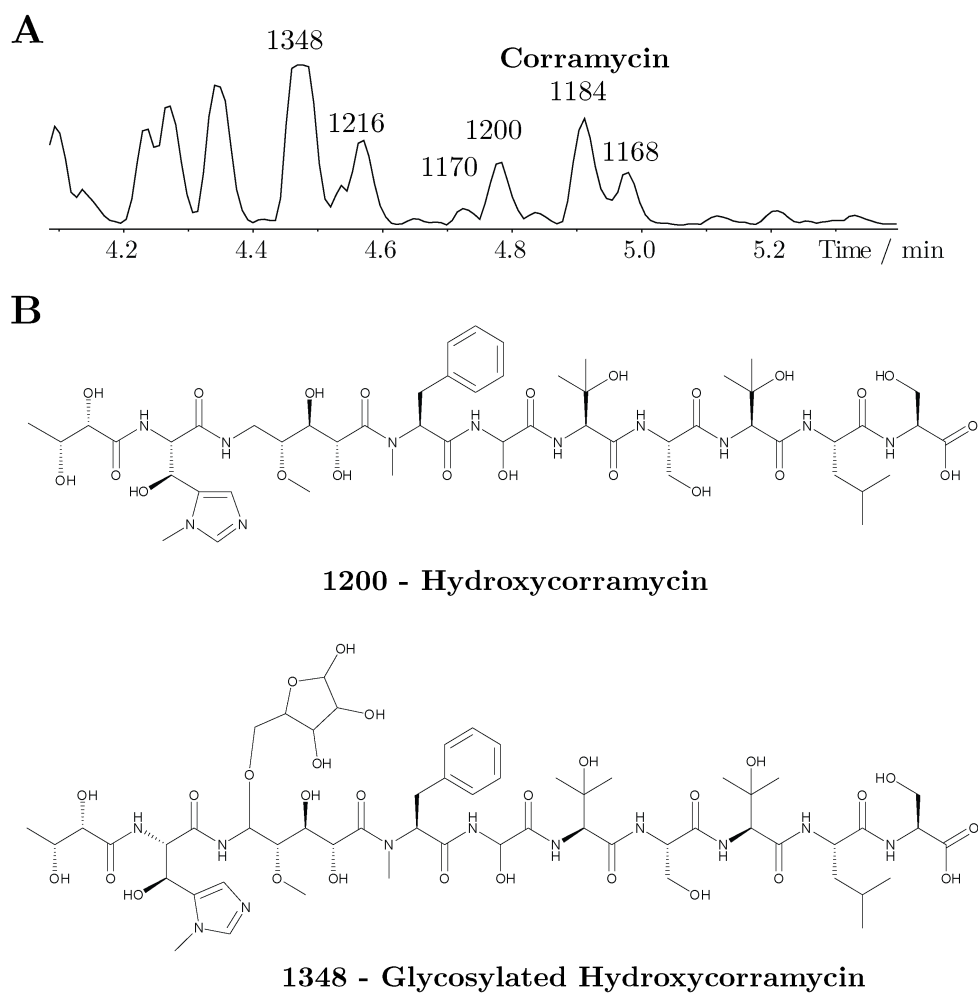


Figure 4.5: **A** Total ion count chromatogram of the *Corrallococcus coralloides* MCy10984 extract with indicated corramycin derivatives. **B** The structure of hydroxycorramycin (1200) was previously described by Sanofi and confirmed at HIPS by NMR analysis (see method section). Glycosylated corramycin (1348) was identified in MCy10984 extract, purified and the analyzed by NMR (see method section).

Table 4.3: Results of the supplementation experiments and the analysis of corramycin derivatives. For supplementation with isotope labeled precursors the used isotope, the observed mass shift $\Delta m/z$, the assigned building block and the putative module of the corramycin assembly line is given. Assigned corramycin derivatives are described by the observed m/z value, the mass shift $\Delta m/z$ to the native corramycin, the proposed modification and the relative intensity in the *Coralloccoccus coralloides* MCy10984 extract.

Supplementation				Corramycin derivatives			
Isotope	$\Delta m/z$	Moiety	Module	m/z	$\Delta m/z$	Modification	Int.
d ₃ -Met	12	4 x Me	-	1184*	-	-	++
¹³ C ₄ - ¹⁵ N-Asp	3	β -Ala	4	1170	- 14	- Me	+
1- ¹³ C-Ac	1			1198	+ 14	+ Me	tr.
2- ¹³ C-Ac	1	Ac	5	1200*	+ 16	+ OH	++
¹³ C ₂ -Ac	2			1216	+ 32	+ 2 OH	+
d ₅ -Phe	5	Phe	6			+ 2 OH	
d ₂ -Gly	2	Gly	7	1348*	+ 164	+ pentose	+++
d ₈ -Val	14	2 x OH-Val	8,10				
d ₃ -Ser	6	2 x Ser	9,12				
d ₃ -Leu	3	Leu	11				
d ₆ -OH-Val	-						
¹³ C ₄ - ¹⁵ N-Thr	-						
¹³ C ₃ -Glycerol	-						

* Structure determined by NMR: 1184 and 1348 at HIPS, 1184 and 1200 at Sanofi

due to poor XAD binding capabilities and identified as a double hydroxylated and additional glycosylated corramycin derivative by NMR analysis (Figure 4.5). One hydroxy group is located at the glycine analogous to hydroxy corramycin. The second hydroxy group is glycosylated with a pentose sugar and attached at the first CH₂ group of the β -alanine moiety. Reconciliation with the bioactivity screening of *Coralloccoccus coralloides* MCy10984 showed that 1348 eluted in an inactive fraction (4.5).

The first module of the corramycin assembly line activates fatty acids *in vitro*

The FAAL domain of corramycin was heterologously expressed in *E. coli* BL21 (DE3). An N-terminal, HRV3C cleavable, 6xHis-MBP tag was used as an efficient means to optimize yield and solubility. FAAL was purified to homogeneity by a three-step protocol using nickel-affinity chromatography, a reverse nickel-affinity chromatography step after HRV-3C digestion to remove the tag and a polishing gel filtration step. To examine the substrate specificity a malachite green based assay was employed.³¹ Quantification of the change in absorption at 620 nm indicated length dependent activation of linear fatty-acids with a maximum activity on decanoic acid (Figure 4.6). The spectrum of activated fatty acid ceased at tridecylcic acid and hexanoic acid as no ATP consumption could be detected for both.

Corramycin is a weak gyrase inhibitor

Corramycin is a compound with potent antibacterial properties, nevertheless the underlying mode-of-action remains elusive. Bacterial DNA gyrase, responsible of introducing negative supercoiling into the DNA doublestrand, is a well known target of established antibiotics like

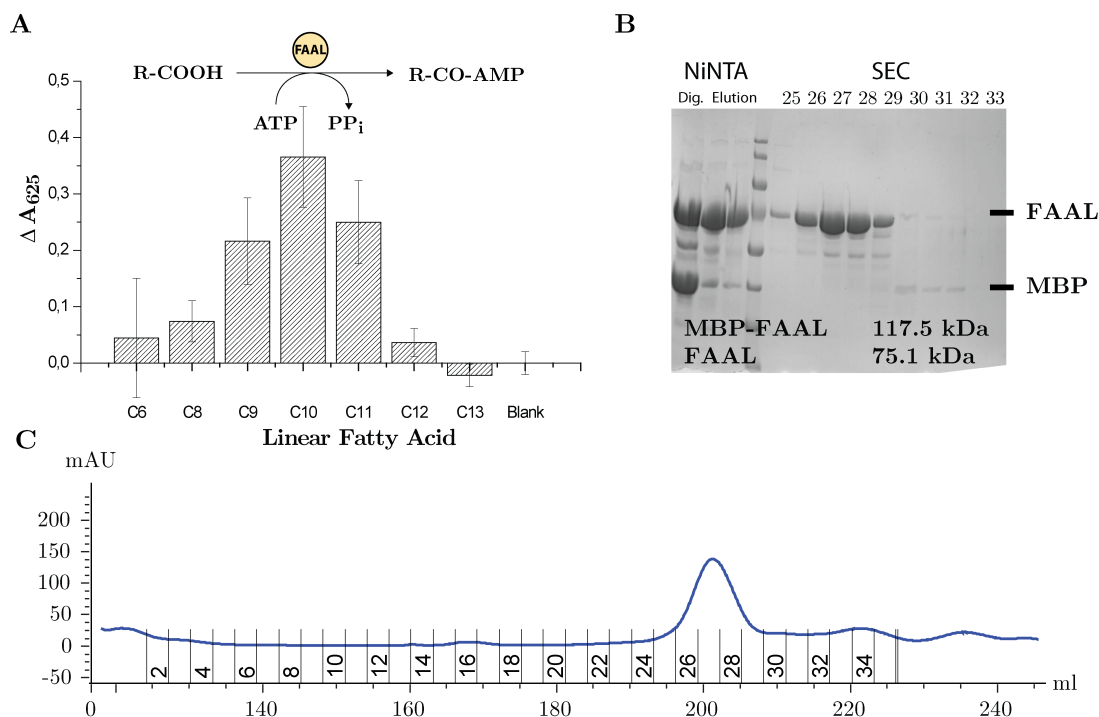


Figure 4.6: **A** Adenylation of linear fatty acid by FAAL *in vitro* produces PP_i that is detected in an malachite green based assay. Substrate specificity of FAAL is determined by monitoring the shift in absorption at 625 nm. **B** SDS-PAGE analysis of FAAL purification process comprising Ni-NTA, MBP-cleavage, second Ni-NTA and a final size exclusion step. **C** UV-trace of the size exclusion chromatography of FAAL.

chinolons. Furthermore, the recently described cystobactamid, a linear peptid-like compound from myxobacteria, as well inhibits gyrase activity.³² To verify the effect of corramycin on the *E. coli* gyrase activity a standard *in vitro* assay was employed.³³ Indeed, corramycin showed weak inhibition of the gyrase activity with an MIC of 27 μ M. Titration of ATP with constant inhibitory corramycin concentration did not restore activity, suggesting that the ATP-binding pocket is not affected. Furthermore, the effect on the activity of *E. coli* topoisomerase I was examined. Topoisomerase I catalyzes the opposing reaction of the gyrase by introducing positive supercoils into the DNA doublestrands. Nevertheless, no effect of corramycin on topoisomerase I activity could be observed (Figure 4.7).

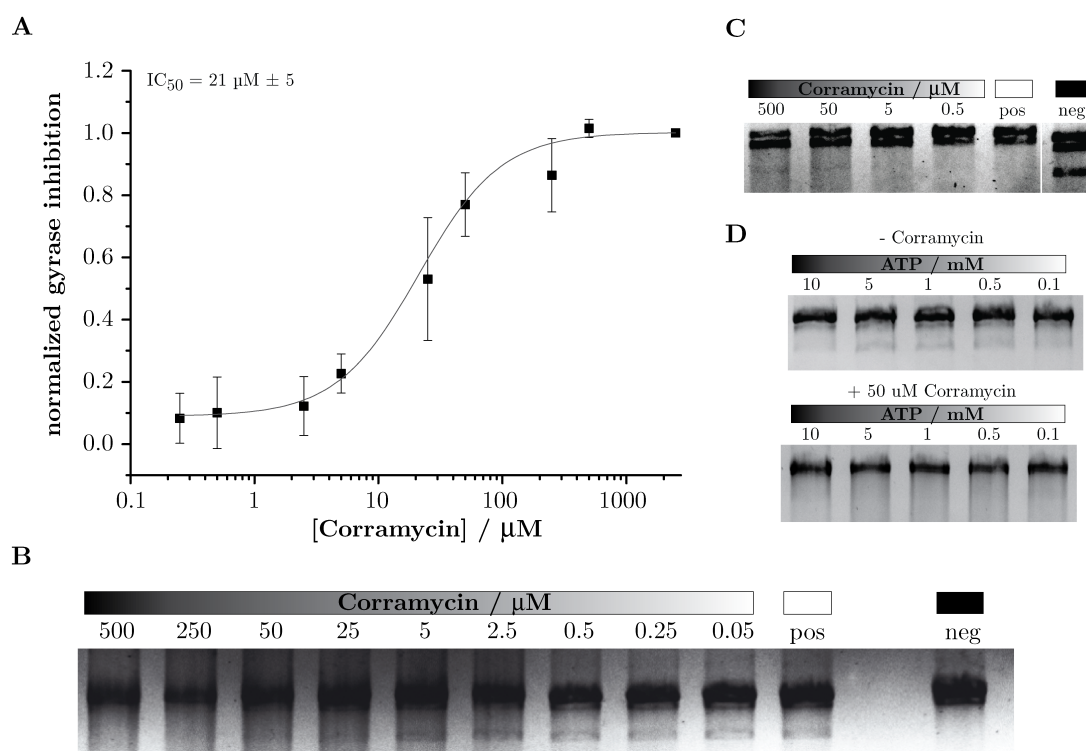


Figure 4.7: **A** Determination of the IC₅₀ of the corramycin gyrase inhibition. Relative gel intensities were measured, normalized and plotted against corramycin concentration. Mean values with bars indicating standard deviation are displayed. IC₅₀ value was determined using the Hill1 fit function. **B** Gel image of one gyrase inhibition experiment. Production of supercoiled DNA by *E. coli* gyrase inhibited by corramycin. **C** Corramycin has no effect on topoisomerase I activity. **D** Variable concentration of ATP has no effect on corramycin gyrase inhibition. Competitive binding to the ATP binding site is therefore unlikely.

4.3 Discussion

Corramycin is a peptide antibiotic with bioactivity against Gram-negative bacteria. A variety of mechanisms for the antibiotic activity of peptide based antibiotics are known. The bactericidal effect of antimicrobial peptides (AMP), a class of linear peptides between 12 and 50 amino acids in size found in all three kingdoms of life, was initially thought to be caused exclusively by membrane-lytic properties.^{34, 35} Nevertheless, research in the last ten years identified various targets inside the cell like the ribosome, cell division, DNA, RNA or cell-wall biosynthesis, and thus revealing alternative, more specific mode-of-actions for AMPs.³⁶ Microcins are a large class of AMPs produced by several enterobacteria. Microcin B17 was found to inhibit the type II DNA topoisomerase, commonly called gyrase.^{37–39} Another member of the microcins, microcin 25, is a specific inhibitor of the RNA polymerase.⁴⁰ Various mode-of-actions have also been shown for non-ribosomal peptides from myxobacteria, which are generally smaller than ribosomal AMPs. Myxovalgine, produced by *Myxococcus fulvus*, consists of 14 amino acids including several α - β -dehydro variants, and exhibits antibacterial activity against several Gram-positive strains.⁴¹ At low concentrations it inhibits tRNA-binding to the A-site of the ribosome, while higher concentrations damage the cell membrane. The observed gyrase inhibition of corramycin with an IC₅₀ of 27 μ M is too weak to explain the strong antibacterial activity reported by Sanofi. Consequently, additional targets with stronger IC₅₀ values are most likely present.

Like other myxobacterial peptides, corramycin contains several unusual moieties, first of all β -hydroxy variants. Myxoprincomid is produced by *Myxococcus xanthus* DK1622 and is made up of nine amino acids, including γ -hydroxy valine, β -lysine and a unusual 2-oxo- β -leucine.²³ The biosynthetic origin of the γ -hydroxy valine moiety could not be identified yet. Comparison of the myxovalgine and the corramycin biosynthetic gene cluster were not successful in identifying an obvious hydroxylase homologue. γ -hydroxy valine is thus likely generally available in the myxobacterial cell and not produced for a distinct natural product. On the contrary, biosynthetic pathways for β -hydroxy histidine, the second building block of corramycin, have been described before. NikQ, a heme protein, facilitates the β -hydroxylation of histidine in the nikkomycin biosynthesis. Interestingly, free histidine is not accepted and has to be loaded onto a PCP-domain by an respective A-domain, before it is recognized by NikQ.⁴² A similar PPant-dependent mechanism is reported for the β -hydroxylation of tyrosine in the novobiocin biosynthesis.^{43, 44} However, cytochrome P450s or other heme proteins are not present in the corramycin biosynthetic gene cluster. One of the present hydroxylase related genes, *orf-4*, *orf-6* or *orf-7*, is thus likely involved in the β -hydroxylation of histidine, although further insights are not possible from *in silico* data alone.

Several modified corramycin derivatives have been identified in this work. Remarkably, all derivatives lose their antibiotic bioactivity, raising the question of the rationale behind their biosynthesis in the first place. Certainly, the generation of multiple variants of a natural product is the core of evolutionary progress and only by the resulting chemical variance novel functions and activities can be achieved. Additionally, the isolation and characterization of novel natural products without obvious bioactivity is a common sight, also in myxobacteria.⁴⁵ Natural products are produced by organisms to gain an evolutionary advantage and not to provide a convenient source for novel antibacterial drugs for human use. Functions as messenger, siderophores, virulence factors or in to date unknown processes are possible as well. Nevertheless, even if the definite reason

for additional hydroxylations and methylations of corramycin can not be elucidated, the genetic basis can be identified. Several hydroxylase related genes as well as an additional methyltransferase, ORF-14, are present in the biosynthetic gene cluster, putatively acting on a fraction of the produced corramycin to generate the observed derivatives. The glycosylation of corramycin derivative 1348 could not be linked to a gene, since no distinct glycosyltransferase is present in the cluster. More likely the attachment of a pentose sugar unit derives from other cell mechanisms, such as protein glycosylation.⁴⁶ Initially thought to only exist in eukaryotes, where about a third of all proteins are *N*- or *O*-glycosylated on specific amino acid side chains,⁴⁷ glycosylation of proteins is also common in bacteria. Best studied is the specific flagella *O*-glycosylation, for example present in *Campylobacter* spp.⁴⁸ Furthermore, general glycosylation systems have been found, for example in *Neisseria* spp., where several *O*-glycosylated membrane proteins were identified. Since *O*-glycosylation of corramycin disables its antibacterial properties it might be a detoxification mechanism of the producer or it is merely incidentally accepted by the glycosylation system.

Another, although putative, derivative is the N-terminal fatty acid linked corramycin as conferred from the module organization and *in vitro* reconstitution of FAAL fatty-acid adenylation activity. FAALs share a high sequence similarity with FACLs (Fatty-acyl CoA ligase), but a short amino acid insertion prevents the second catalytic step, the coupling to CoA. Instead, the activated acyl-chain is transferred to a downstream ACP-domain as a building block for the assembly-line machinery. Lipopeptides like surfactin or the family of iturines, including bacillomycin, mycosubtilin or lichenysin, consist of a circular peptide backbone with an attached fatty acid, that is vital for bioactivity.^{49–51} The genetic organisation of the assembly line modules is similar to corramycin, with a FAAL domain as starter unit that activates and loads the fatty acid.^{52–55} The weak substrate specificity observed for the corramycin FAAL is typical and lipopeptides often occur with various fatty acid chain lengths on the same peptide backbone. The recently described puwainaphycins from *Cyanobacterium cylindrospermum alatosporum* were isolated as C7- and C9-fatty acid linked derivatives in similar quantities.⁵⁶ For corramycin, fatty acid linked derivatives could not be detected. The myxobacterial lipopeptide vioprolide is produced by an assembly line harboring the same first two modules as corramycin, namely FAAL and glycerate incorporating FkbH. Fatty-acid linked vioprolide is a prodrug, that is cleaved and thus activated upon export.⁵⁷ This mechanism of hydrolytic maturation is well known and already described for several natural products including xenocumarin,⁵⁸ heronamide,⁵⁹ paenilamicin,⁶⁰ pyoverdine,⁶¹ colibactin, naphthyridinomycin,⁶² quinocarcin,⁶³ and telomycin.⁶⁴ Peptides like didemin,⁶⁵ saframycin,⁶⁶ and nocardicin A⁶⁷ are matured analogous by proteolytic cleavage. For corramycin, a maturation mechanism by hydrolytic cleavage is likely due to the evidently active FAAL domain in module one and the ultimately fatty-acid free structure. Additionally, two putative ORFs with periplasmatic signal, ORF-0 and ORF-1, encode for a hydrolase and a flipase-like exporter, respectively. Similar to vioprolide, corramycin is likely produced as lipopeptide, that is hydrolyzed to its active form by ORF-0 after ORF-1 mediated export.

4.4 Bibliography

- [1] David J. Newman and Gordon M. Cragg. Natural products as sources of new drugs from 1981 to 2014. *Journal of natural products*, 79(3):629–661, 2016.
- [2] Kozo Ochi and Takeshi Hosaka. New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters. *Applied microbiology and biotechnology*, 97(1):87–98, 2013.
- [3] Tilmann Weber, Kai Blin, Srikanth Duddela, Daniel Krug, Hyun Uk Kim, Robert Brucconeri, Sang Yup Lee, Michael A. Fischbach, Rolf Muller, Wolfgang Wohlleben, Rainer Breitling, Eriko Takano, and Marnix H. Medema. antismash 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic acids research*, 43(W1):W237–43, 2015.
- [4] Peter Cimermancic, Marnix H. Medema, Jan Claesen, Kenji Kurita, Laura C. Wieland Brown, Konstantinos Mavrommatis, Amrita Pati, Paul A. Godfrey, Michael Koehrsen, Jon Clardy, Bruce W. Birren, Eriko Takano, Andrej Sali, Roger G. Linington, and Michael A. Fischbach. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell*, 158(2):412–421, 2014.
- [5] Jose Munoz-Dorado, Francisco J. Marcos-Torres, Elena Garcia-Bravo, Aurelio Moraleda-Munoz, and Juana Perez. Myxobacteria: Moving, killing, feeding, and surviving together. *Frontiers in microbiology*, 7:781, 2016.
- [6] H. Reichenbach. Myxobacteria, producers of novel bioactive substances. *Journal of industrial microbiology & biotechnology*, 27(3):149–156, 2001.
- [7] Silke C. Wenzel, Holger Hoffmann, Jidong Zhang, Laurent Debussche, Sabine Haag-Richter, Michael Kurz, Frederico Nardi, Peer Lukat, Irene Kochems, Heiko Tietgen, Dietmar Schummer, Jean-Paul Nicolas, Loreley Calvet, Valerie Czepczor, Patricia Vrignaud, Agnes Muhlenweg, Stefan Pelzer, Rolf Muller, and Mark Bronstrup. Production of the bengamide class of marine natural products in myxobacteria: Biosynthesis and structure-activity relationships. *Angewandte Chemie (International ed. in English)*, 54(51):15560–15564, 2015.
- [8] B. S. Goldman, W. C. Nierman, D. Kaiser, S. C. Slater, A. S. Durkin, J. A. Eisen, C. M. Ronning, W. B. Barbazuk, M. Blanchard, C. Field, C. Halling, G. Hinkle, O. Iartchuk, H. S. Kim, C. Mackenzie, R. Madupu, N. Miller, A. Shvartsbeyn, S. A. Sullivan, M. Vaudin, R. Wiegand, and H. B. Kaplan. Evolution of sensory complexity recorded in a myxobacterial genome. *Proceedings of the National Academy of Sciences of the United States of America*, 103(41):15200–15205, 2006.
- [9] Haruo Ikeda, Jun Ishikawa, Akiharu Hanamoto, Mayumi Shinose, Hisashi Kikuchi, Tadayoshi Shiba, Yoshiyuki Sakaki, Masahira Hattori, and Satoshi Omura. Complete genome sequence and comparative analysis of the industrial microorganism streptomyces avermitilis. *Nature biotechnology*, 21(5):526–531, 2003.

- [10] Markiyan Oliynyk, Markiyan Samborsky, John B. Lester, Tatiana Mironenko, Nataliya Scott, Shilo Dickens, Stephen F. Haydock, and Peter F. Leadlay. Complete genome sequence of the erythromycin-producing bacterium *saccharopolyspora erythraea* nrrl23338. *Nature biotechnology*, 25(4):447–453, 2007.
- [11] Michael A. Fischbach and Christopher T. Walsh. Antibiotics for emerging pathogens. *Science (New York, N.Y.)*, 325(5944):1089–1093, 2009.
- [12] Christopher T. Walsh and Timothy A. Wencewicz. Prospects for new antibiotics: a molecule-centered perspective. *The Journal of antibiotics*, 67(1):7–22, 2014.
- [13] Christopher Walsh and Timothy Wencewicz. *Antibiotics: Challenges, mechanisms, opportunities*. American Society for Microbiology, 2016.
- [14] Stephane Renard, 2014, Personal Communication.
- [15] Pooja Arora, Aneesh Goyal, Vivek T. Natarajan, Eerappa Rajakumara, Priyanka Verma, Radhika Gupta, Malikmohamed Yousuf, Omita A. Trivedi, Debasisa Mohanty, Anil Tyagi, Rajan Sankaranarayanan, and Rajesh S. Gokhale. Mechanistic and functional insights into fatty acid activation in mycobacterium tuberculosis. *Nature chemical biology*, 5(3):166–173, 2009.
- [16] Fabian Sievers, Andreas Wilm, David Dineen, Toby J. Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, Hamish McWilliam, Michael Remmert, Johannes Soding, Julie D. Thompson, and Desmond G. Higgins. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular systems biology*, 7:539, 2011.
- [17] Grzegorz M. Boratyn, Christiam Camacho, Peter S. Cooper, George Coulouris, Amelia Fong, Ning Ma, Thomas L. Madden, Wayne T. Matten, Scott D. McGinnis, Yuri Merezuk, Yan Raytselis, Eric W. Sayers, Tao Tao, Jian Ye, and Irena Zaretskaya. Blast: a more efficient report with usability improvements. *Nucleic acids research*, 41(Web Server issue):W29–33, 2013.
- [18] Mark Hildebrand, Laura E. Waggoner, Haibin Liu, Sebastian Sudek, Scott Allen, Christine Anderson, David H. Sherman, and Margo Haygood. brya: an unusual modular polyketide synthase gene from the uncultivated bacterial symbiont of the marine bryozoan bugula neritina. *Chemistry & biology*, 11(11):1543–1552, 2004.
- [19] Christine Jahns, Thomas Hoffmann, Stefan Muller, Klaus Gerth, Peter Washausen, Gerhard Hofle, Hans Reichenbach, Markus Kalesse, and Rolf Muller. Pellasoren: structure elucidation, biosynthesis, and total synthesis of a cytotoxic secondary metabolite from *sorangium cellulosum*. *Angewandte Chemie (International ed. in English)*, 51(21):5239–5243, 2012.
- [20] Wenli Li, Jianhua Ju, Scott R. Rajski, Hiroyuki Osada, and Ben Shen. Characterization of the tautomycin biosynthetic gene cluster from *streptomyces spiroverticillatus* unveiling new insights into dialkylmaleic anhydride and polyketide biosynthesis. *The Journal of biological chemistry*, 283(42):28607–28617, 2008.

-
- [21] Marc Rottig, Marnix H. Medema, Kai Blin, Tilmann Weber, Christian Rausch, and Oliver Kohlbacher. Nrpspredictor2—a web server for predicting nrps adenylation domain specificity. *Nucleic acids research*, 39(Web Server issue):W362–7, 2011.
- [22] T. Stachelhaus, H. D. Mootz, and M. A. Marahiel. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry & biology*, 6(8):493–505, 1999.
- [23] Nina Socorro Cortina, Daniel Krug, Alberto Plaza, Ole Revermann, and Rolf Muller. Myxoprincomide: a natural product from *myxococcus xanthus* discovered by comprehensive analysis of the secondary metabolome. *Angewandte Chemie (International ed. in English)*, 51(3):811–816, 2012.
- [24] Brian O. Bachmann and Jacques Ravel. Chapter 8. methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from dna sequence data. *Methods in enzymology*, 458:181–217, 2009.
- [25] Nadine Ziemert, Sheila Podell, Kevin Penn, Jonathan H. Badger, Eric Allen, and Paul R. Jensen. The natural product domain seeker napdos: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PloS one*, 7(3):e34064, 2012.
- [26] H. Motamedi, A. Shafiee, S. J. Cai, S. L. Streicher, B. H. Arison, and R. R. Miller. Characterization of methyltransferase and hydroxylase genes involved in the biosynthesis of the immunosuppressants fk506 and fk520. *Journal of bacteriology*, 178(17):5243–5248, 1996.
- [27] C. T. Walsh, H. Chen, T. A. Keating, B. K. Hubbard, H. C. Losey, L. Luo, C. G. Marshall, D. A. Miller, and H. M. Patel. Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on nrps assembly lines. *Current opinion in chemical biology*, 5(5):525–534, 2001.
- [28] Evan G. Johnson, Stuart B. Krasnoff, Dawn R. D. Bignell, Wen-Chuan Chung, Tao Tao, Ronald J. Parry, Rosemary Loria, and Donna M. Gibson. 4-nitrotryptophan is a substrate for the non-ribosomal peptide synthetase txtB in the thaxtomin A biosynthetic pathway. *Molecular microbiology*, 73(3):409–418, 2009.
- [29] S. Jabbouri, B. Relic, M. Hanin, P. Kamalaprija, U. Burger, D. Prome, J. C. Prome, and W. J. Broughton. nlo and noeI (hsnIII) of *Rhizobium* sp. ngr234 are involved in 3-O-carbamoylation and 2-O-methylation of nod factors. *The Journal of biological chemistry*, 273(20):12047–12055, 1998.
- [30] Hilla Magidovich, Sophie Yurist-Doutsch, Zvia Konrad, Valeria V. Ventura, Anne Dell, Paul G. Hitchen, and Jerry Eichler. Aglp is a S-adenosyl-L-methionine-dependent methyltransferase that participates in the N-glycosylation pathway of *Haloferax volcanii*. *Molecular microbiology*, 76(1):190–199, 2010.
- [31] Thomas J. McQuade, Abbie D. Shallop, Anita Sheoran, James E. Delproposto, Oleg V. Tsodikov, and Sylvie Garneau-Tsodikova. A nonradioactive high-throughput assay for screening and characterization of adenylation domains for nonribosomal peptide combinatorial biosynthesis. *Analytical biochemistry*, 386(2):244–250, 2009.

- [32] Sascha Baumann, Jennifer Herrmann, Ritesh Raju, Heinrich Steinmetz, Kathrin I. Mohr, Stephan Huttel, Kirsten Harmrolfs, Marc Stadler, and Rolf Muller. Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *Angewandte Chemie (International ed. in English)*, 53(52):14605–14609, 2014.
- [33] J. M. Domagala, L. D. Hanna, C. L. Heifetz, M. P. Hutt, T. F. Mich, J. P. Sanchez, and M. Solomon. New structure-activity relationships of the quinolone antibacterials using the target enzyme. the development and application of a dna gyrase assay. *Journal of medicinal chemistry*, 29(3):394–404, 1986.
- [34] Yechiel Shai. Mode of action of membrane active antimicrobial peptides. *Biopolymers*, 66(4):236–248, 2002.
- [35] Michael R. Yeaman and Nannette Y. Yount. Mechanisms of antimicrobial peptide action and resistance. *Pharmacological reviews*, 55(1):27–55, 2003.
- [36] Cheng-Foh Le, Chee-Mun Fang, and Shamala Devi Sekaran. Intracellular targeting mechanisms by antimicrobial peptides. *Antimicrobial agents and chemotherapy*, 61(4), 2017.
- [37] J. L. Vizán, C. Hernandez-Chico, I. del Castillo, and F. Moreno. The peptide antibiotic microcin b17 induces double-strand cleavage of dna mediated by e. coli dna gyrase. *The EMBO journal*, 10(2):467–476, 1991.
- [38] J. Davagnino, M. Herrero, D. Furlong, F. Moreno, and R. Kolter. The dna replication inhibitor microcin b17 is a forty-three-amino-acid protein containing sixty percent glycine. *Proteins*, 1(3):230–238, 1986.
- [39] J. G. Heddle, S. J. Blance, D. B. Zamble, F. Hollfelder, D. A. Miller, L. M. Wentzell, C. T. Walsh, and A. Maxwell. The antibiotic microcin b17 is a dna gyrase poison: characterisation of the mode of inhibition. *Journal of molecular biology*, 307(5):1223–1234, 2001.
- [40] Jayanta Mukhopadhyay, Elena Sineva, Jennifer Knight, Ronald M. Levy, and Richard H. Ebright. Antibacterial peptide microcin j25 inhibits transcription by binding within and obstructing the rna polymerase secondary channel. *Molecular cell*, 14(6):739–751, 2004.
- [41] H. Irschik, K. Gerth, T. Kemmer, H. Steinmetz, and H. Reichenbach. The myxovalargins, new peptide antibiotics from myxococcus fulvus (myxobacterales). i. cultivation, isolation, and some chemical and biological properties. *The Journal of antibiotics*, 36(1):6–12, 1983.
- [42] Huawei Chen, Brian K. Hubbard, Sarah E. O’Connor, and Christopher T. Walsh. Formation of beta-hydroxy histidine in the biosynthesis of nikkomycin antibiotics. *Chemistry & biology*, 9(1):103–112, 2002.
- [43] M. Steffensky, A. Muhlenweg, Z. X. Wang, S. M. Li, and L. Heide. Identification of the novobiocin biosynthetic gene cluster of streptomyces spheroides ncib 11891. *Antimicrobial agents and chemotherapy*, 44(5):1214–1222, 2000.
- [44] H. Chen and C. T. Walsh. Coumarin formation in novobiocin biosynthesis: beta-hydroxylation of the aminoacyl enzyme tyrosyl-s-novh by a cytochrome p450 novi. *Chemistry & biology*, 8(4):301–312, 2001.

-
- [45] J. Herrmann, A. Abou Fayad, and R. Muller. Natural products from myxobacteria: novel metabolites and bioactivities. *Natural product reports*, 2016.
- [46] Harald Nothhaft and Christine M. Szymanski. Protein glycosylation in bacteria: sweeter than ever. *Nature reviews. Microbiology*, 8(11):765–778, 2010.
- [47] R. Apweiler, H. Hermjakob, and N. Sharon. On the frequency of protein glycosylation, as deduced from analysis of the swiss-prot database. *Biochimica et biophysica acta*, 1473(1):4–8, 1999.
- [48] P. Thibault, S. M. Logan, J. F. Kelly, J. R. Brisson, C. P. Ewing, T. J. Trust, and P. Guerry. Identification of the carbohydrate moieties and glycosylation motifs in campylobacter jejuni flagellin. *The Journal of biological chemistry*, 276(37):34862–34870, 2001.
- [49] K. Arima, A. Kakinuma, and G. Tamura. Surfactin, a crystalline peptidelipid surfactant produced by bacillus subtilis: isolation, characterization and its inhibition of fibrin clot formation. *Biochemical and biophysical research communications*, 31(3):488–494, 1968.
- [50] F. Peypoux, M. T. Pommier, D. Marion, M. Ptak, B. C. Das, and G. Michel. Revised structure of mycosubtilin, a peptidolipid antibiotic from bacillus subtilis. *The Journal of antibiotics*, 39(5):636–641, 1986.
- [51] R. Maget-Dana and F. Peypoux. Iturins, a special class of pore-forming lipopeptides: biological and physicochemical properties. *Toxicology*, 87(1-3):151–174, 1994.
- [52] E. H. Duitman, L. W. Hamoen, M. Rembold, G. Venema, H. Seitz, W. Saenger, F. Bernhard, R. Reinhardt, M. Schmidt, C. Ullrich, T. Stein, F. Leenders, and J. Vater. The mycosubtilin synthetase of bacillus subtilis atcc6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. *Proceedings of the National Academy of Sciences of the United States of America*, 96(23):13294–13299, 1999.
- [53] K. Tsuge, T. Akiyama, and M. Shoda. Cloning, sequencing, and characterization of the iturin a operon. *Journal of bacteriology*, 183(21):6265–6273, 2001.
- [54] Shiyi Yao, Xuwen Gao, Norbert Fuchsbaue, Wolfgang Hillen, Joachim Vater, and Jinsheng Wang. Cloning, sequencing, and characterization of the genetic region relevant to biosynthesis of the lipopeptides iturin a and surfactin in bacillus subtilis. *Current microbiology*, 47(4):272–277, 2003.
- [55] Anne-Laure Moyne, Thomas E. Cleveland, and Sadik Tuzun. Molecular characterization and analysis of the operon encoding the antifungal lipopeptide bacillomycin d. *FEMS microbiology letters*, 234(1):43–49, 2004.
- [56] Jan Mares, Jan Hajek, Petra Urajova, Jiri Kopecky, and Pavel Hrouzek. A hybrid non-ribosomal peptide/polyketide synthetase containing fatty-acyl ligase (faal) synthesizes the beta-amino fatty acid lipopeptides puwainaphycins in the cyanobacterium cylindrospermum alatosporum. *PloS one*, 9(11):e111904, 2014.
- [57] David Auerbach, 2017, Personal Communication.

- [58] Daniela Reimer, Klaas M. Pos, Marco Thines, Peter Grun, and Helge B. Bode. A natural prodrug activation mechanism in nonribosomal peptide synthesis. *Nature chemical biology*, 7(12):888–890, 2011.
- [59] Yiguang Zhu, Wenjun Zhang, Yaolong Chen, Chengshan Yuan, Haibo Zhang, Guangtao Zhang, Liang Ma, Qingbo Zhang, Xinpeng Tian, Si Zhang, and Changsheng Zhang. Characterization of heronamide biosynthesis reveals a tailoring hydroxylase and indicates migrated double bonds. *Chembiochem : a European journal of chemical biology*, 16(14):2086–2093, 2015.
- [60] Sebastian Muller, Eva Garcia-Gonzalez, Andi Mainz, Gillian Hertlein, Nina C. Heid, Eva Mosker, Hans van den Elst, Herman S. Overkleeft, Elke Genersch, and Roderich D. Sussmuth. Paenilamicin: structure and biosynthesis of a hybrid nonribosomal peptide/polyketide antibiotic from the bee pathogen *paenibacillus* larvae. *Angewandte Chemie (International ed. in English)*, 53(40):10821–10825, 2014.
- [61] Isabelle J. Schalk and Laurent Guillon. Pyoverdine biosynthesis and secretion in *pseudomonas aeruginosa*: implications for metal homeostasis. *Environmental microbiology*, 15(6):1661–1673, 2013.
- [62] Jin-Yue Pu, Chao Peng, Man-Cheng Tang, Yue Zhang, Jian-Ping Guo, Li-Qiang Song, Qiang Hua, and Gong-Li Tang. Naphthyridinomycin biosynthesis revealing the use of leader peptide to guide nonribosomal peptide assembly. *Organic letters*, 15(14):3674–3677, 2013.
- [63] Tomoshige Hiratsuka, Kento Koketsu, Atsushi Minami, Shunsuke Kaneko, Chiaki Yamazaki, Kenji Watanabe, Hiroki Oguri, and Hideaki Oikawa. Core assembly mechanism of quinocarcin/sf-1739: bimodular complex nonribosomal peptide synthetases for sequential mannich-type reactions. *Chemistry & biology*, 20(12):1523–1535, 2013.
- [64] Chengzhang Fu, Lena Keller, Armin Bauer, Mark Bronstrup, Alexandre Froidbise, Peter Hammann, Jennifer Herrmann, Guillaume Mondesert, Michael Kurz, Matthias Schiell, Dietmar Schummer, Luigi Toti, Joachim Wink, and Rolf Muller. Biosynthetic studies of telomycin reveal new lipopeptides with enhanced activity. *Journal of the American Chemical Society*, 137(24):7692–7705, 2015.
- [65] Ying Xu, Roland D. Kersten, Sang-Jip Nam, Liang Lu, Abdulaziz M. Al-Suwailem, Hua-jun Zheng, William Fenical, Pieter C. Dorrestein, Bradley S. Moore, and Pei-Yuan Qian. Bacterial biosynthesis and maturation of the didemnins anti-cancer agents. *Journal of the American Chemical Society*, 134(20):8625–8632, 2012.
- [66] Kento Koketsu, Kenji Watanabe, Haruna Suda, Hiroki Oguri, and Hideaki Oikawa. Reconstruction of the saframycin core scaffold defines dual pictet-spengler mechanisms. *Nature chemical biology*, 6(6):408–410, 2010.
- [67] Jeanne M. Davidsen, David M. Bartley, and Craig A. Townsend. Non-ribosomal propeptide precursor in nocardicin a biosynthesis predicted from adenylation domain specificity dependent on the mbth family protein nci. *Journal of the American Chemical Society*, 135(5):1749–1759, 2013.

4.5 Experimental Procedures

Isolation of genomic DNA from *Corallococcus coralloides*

The described procedure was used for both strains, *Corallococcus coralloide* ST201330 and MCy10984. The strains were grown and maintained in 50 mL AMB-media (0.5 % (w/v) soluble starch, 0.25 % (w/v) casitone (bacto), 0.05 % (w/v) $\text{MgSO}_4 \cdot \text{H}_2\text{O}$, 0.025 % (w/v) K_2HPO_4 , 10 mM HEPES, adjusted to pH 7.0 with KOH before autoclaving) for 5 days at 30 °C and 180 rpm. The culture was pelleted, resuspended in 5 mL 10 mM Tris-HCl (pH 7.5) and 300 μL Proteinase K solution (10 mg/mL) and 600 μL SDS-solution (10 %) were added. The mixture was incubated at 55 °C for 2h. Subsequent extraction using Phenol:Chloroform:Isoamylalcohol (25:24:1) was conducted three times as follows: One volume of P:C:I was added and the tube was swung for 60 min at 5 rpm. After centrifugation at 8000 rpm for 10 min the upper phase was transferred to a new tube and again extracted. After three extraction steps the upper phase was extracted with Chloroform:Isoamylalcohol (24:1) in the same way. 1/10 volume of 3 M Na-acetate solution (pH 5.5) was added to the upper phase and mixed. 2.5 volume of ice-cold ethanol (100 %) were added and the tube inverted several times. The white DNA was subsequently transferred to 1 mL of 70 % ethanol with a glass pasteur-pipette. gDNA was spun down at 1000 rpm for 1 min. After removal of the supernatant the DNA was air-dried overnight. The DNA-pellet was resuspended in 500 μL of 10 mM Tris-HCl buffer (pH 7.5).

Illumina sequencing

Draft genome sequence of *Corallococcus coralloides* ST201330 (MCy10705) was determined using Illumina sequencing technology in cooperation with the SEQ-IT GmbH & Co. KG (Kaiserlautern, Germany). Raw sequencing data obtained from the MiSeq platform comprised 26,054,366 paired-end reads and 14,483,850 mate-pair reads, for both libraries the read length was 250 bp. This data was assembled into contigs with Abyss-pe assembler software version 1.3.6, resulting in 26 sequences with total length of 10,243,981 bp.

Draft genome sequence of *Corallococcus* sp. MCy10984 was determined using Illumina sequencing technology in cooperation with the Genome Analytics group at the Helmholtz Centre for Infection Research (Braunschweig, Germany). Raw sequencing data obtained from the MiSeq platform comprised 9,509,890 paired-end reads with the length of 250 bp each. This data was assembled into contigs with Abyss-pe assembler software version 1.3.6. resulting in 59 sequences with total length of 11,107,395 bp.

Sequence analysis

Routine DNA and protein sequence analysis tasks were carried out using the Geneious v10 Software. Other tools used are indicated in the respective result section.

Fermentation of *Corallococcus coralloides* MCy10984

Corallococcus coralloides was grown in 50 mL Amb-media supplemented with 0.00005 % vitamin B12 and 2 % Amperlite XAD-16 resin (Sigma-Aldrich) after autoclaving. Cultures were allowed to grow for 96-120 h at 200 rpm and 30 °C in 250-mL Erlenmeyer flasks. The cell and XAD resin

pellets were harvested by centrifugation (20 min, 5000 rpm, 20 °C). The pellet was extracted with 2 x 30 mL methanol. Extracts were dried in vacua and resuspended in 1 mL methanol.

All measurements were performed on a Dionex Ultimate 3000 RSLC system using a BEH C18, 100 x 2.1 mm, 1.7 μ m dp column (Waters, Germany). Separation of 1 μ l sample was achieved by a linear gradient from (A) H₂O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 600 μ L/min and 45 °C. The gradient was initiated by a 0.5 min isocratic step at 5 % B, followed by an increase to 95 % B in 18 min to end up with a 2 min step at 95 % B before reequilibration under the initial conditions. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 75 μ L/min before entering the maXis 4G hr-ToF mass spectrometer (Bruker Daltonics, Germany) using the Apollo ESI source. Mass spectra were acquired in centroid mode ranging from 150–2500 m/z at a 2 Hz scan rate.

All described MS/MS data was generated as follows: Minimum precursor intensity is set to 10000. Full scan spectra are acquired at 2 Hz followed by MS/MS spectra acquisition at variable scan speed ranging from 1 to 3 Hz, as a function of precursor intensity. CID energy varies linearly from 30, 35, 45, to 55 eV with respect to the precursor m/z from 300, 600, 1000, to 2000 m/z . The collision cell is set to ramp collision energy (80-120 % of the set value with equal weights of both values), collision RF (700 to 1000 Vpp with equal weights of both values) and transfer time (90-110 μ s) for every MS/MS scan. The number of precursor was set to 2 and precursors were moved to an exclusion list for 0.2 min after two spectra were measured (typical chromatographic peak width was 0.10-0.15 min). Precursors were reconsidered if their intensity changed fivefold.

Purification and structural determination of corramycin and glycosylated corramycin (1348)

Large scale production of corramycin and derivative 1348 was carried out in E-media supplemented with 0.00005 % vitamin B12 after autoclaving. Cultures were allowed to grow for 96-120 h at 200 rpm and 30 °C in 5-L Erlenmeyer flasks. For corramycin 2 % Amperlite XAD-16 resin (Sigma-Aldrich) was supplemented and the compound purified from the cell and XAD pellet. 1348 was purified from the filtered supernatant without supplementation of XAD. Methanol extracts of the pellet/dried supernatant were subjected to preparative HPLC.

Preparative HPLC step was performed with a Waters Autopurifier System (APS) equipped with a Phenomenex Biphenyl 250 x 4.6 mm, 5 μ m dp analytical column for method development at 1.0 mL/min flow rate. Preparative separations conducted a Biphenyl 250 x 22 mm, 5 μ m dp column at 25 mL/min flow rate. Both used (A) H₂O + 0.1 % FA and (B) ACN + 0.1 % FA as solvent system. Separation of an 800 μ l sample injection started with a 1 min isocratic step at 25 % B, followed by a gradient to 29 % B in 22 min, a steep increase to 95 % B in 1 min to end up with a plateau for 4 min prior to return and reequilibration with the initial conditions. Fraction collection was triggered by a mass trigger set to 1184.5 and 1348.5 m/z , respectively. Solvent was removed from the collected fraction and resolved in a small amount of methanol for further analysis.

NMR spectra were recorded on a 700 MHz Avance III (Ascend) spectrometer by Bruker BioSpin GmbH, equipped with a 5 mm TXI cryoprobe, at 298 K. Chemical shift values of ¹H- and ¹³C-NMR spectra are reported in ppm relative to the residual solvent signal given as an internal standard. ¹³C-signals were assigned via 2D-CH and CCH correlations (HSQC and HMBC) (see

Table S 4.4 and 4.5).

Supplementation of labeled precursors

Shake flask production cultures in 10 mL AMB-media containing 2 % XAD-16 of *C. coralloides* MCy10984 were supplemented with 1mM of labeled amino acids. The following amino acids were used: L-leucine-5,5,5-d3 (Deutero); L-serine-2,3,3-d3 (CIL); L-valine-d8 (Deutero); L-threonine-13C4-15N (Isotec), L-phenylalanine-d5 (euriso-top); L-methionine (methyl-d3) (Campro Scientific); sodium 1-13C-acetate (CIL); sodium 2-13C-acetate (CIL); 1,2-13C-acetate, L-aspartate-13C4-15N; L-glycine-2,2-d2, hydroxy-valine-d6 and glycerol-13C3. Cultures were grown for 70 hours at 30 °C , 200 rpm. The cells and XAD pellet were separated from the medium by centrifugation and were extracted with two times 15 ml methanol. The extracts were concentrated to 100 μ L and subsequently analyzed using LC-hrMS.

All measurements were performed on a Dionex Ultimate 3000 RSLC system using a Waters BEH C18, 50 x 2.1 mm, 1.7 μ m dp column by injection of two μ L methanolic sample. Separation was achieved by a linear gradient with (A) H₂O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 600 μ L/min and 45 °C. The gradient was initiated by a 0.33 min isocratic step at 5 % B, followed by an increase to 95 % B in 9 min to end up with a 1 min flush step at 95 % B before reequilibration under the initial conditions. UV and MS detection were performed simultaneously. Coupling the HPLC to the MS was supported by an Advion Triversa Nanomate nano-ESI system attached to a Thermo Scientific Orbitrap. Mass spectra were acquired in centroid mode ranging from 200–2000 m/z at a resolution of $R = 30000$.

Protein purification of FAAL

The respective gene was inserted into a petM44 expression vector with an N-terminal 6xHis-MBP tag using according primers (restriction site in bold):

FAAL-FP AAAAAA**CATGTT**GGACCTCGCTCGGA

FAAL-RP AAAAAAA**GCTTCT**ACGCCGGCGAG

The construct was expressed in *E. coli* BL21 (DE3) cells (500 mL LB, 0.1 mM IPTG, 16 °C, 16h). The cell pellet was resuspended in lysis buffer (150 mM NaCl, 25 mM Tris, 40 mM Imidazole, pH 7.5), sonicated and centrifuged. The supernatant was loaded onto a gravity flow column containing NiNTA loaded sepharose, washed with lysis buffer and subsequently eluted in one step (250 mM Imidazole). The tag was cleaved using HRV3C protease during overnight dialysis against SEC buffer (150 NaCl, 25 Tris, pH 7.5) at 4 °C, and removed by a second Ni-NTA chromatography step. After passing through a Superdex 200 16/60 pg column (GE Healthcare Life Sciences) the protein was concentrated using a 30 kDa cutoff filter and stored at -80°C in 10 % glycerol. Protein purity was determined by SDS-PAGE. Protein concentration was determined spectrophotometrically upon determining the respective extinction coefficient from the amino acid sequence using the PROTPARAM webserver (<http://web.expasy.org/protparam/>; Gasteiger et al. 2005).

Malachite green assay of FAAL

The substrate specificity was assayed as described previously.³¹ Briefly, 4 μ M FAAL was incubated with 2 μ M ATP and 2 μ M linear fatty acid (C6-C13) in reaction buffer (150 mM NaCl, 25 mM Tris-HCl, pH 7.5) containing inorganic phosphatase at 37 °C for 1h. The reaction was stopped by adding 20 μ L malachite green solution and the absorption at 620 nm was measured after 10 min. Fold increase of absorption was calculated based on a negative control without substrate. All reactions were performed in triplicates.

Gyrase inhibition assay

To test the anti-gyrase activity of corramycin, commercial *Escherichia coli* gyrase supercoiling kits (Inspiralis, Norwich, UK) were used. For standard reactions 0.5 μ g relaxed plasmid were mixed with 1 unit (20.5 nM) gyrase in 1x reaction buffer (30 μ L final volume, see kit manual) and incubated for 30 minutes at 37 °C. The reactions were quenched by the addition of DNA gel loading buffer containing 10 % (w/v) SDS. The samples were separated on 0.8 % (w/v) agarose gels and DNA was visualized using ethidium bromide. Corramycin stock solution and dilution were prepared in DMSO and added to the supercoiling reactions giving a final DMSO concentration of 5 % (v/v). Control reactions were: no enzyme and a standard reaction in presence of 5 % (v/v) DMSO. All reaction samples were equilibrated for 15 minutes at room temperature in the absence of DNA. Then the relaxed plasmid was added to start the reaction. All reactions were performed in triplicates.

4.6 Supporting Information

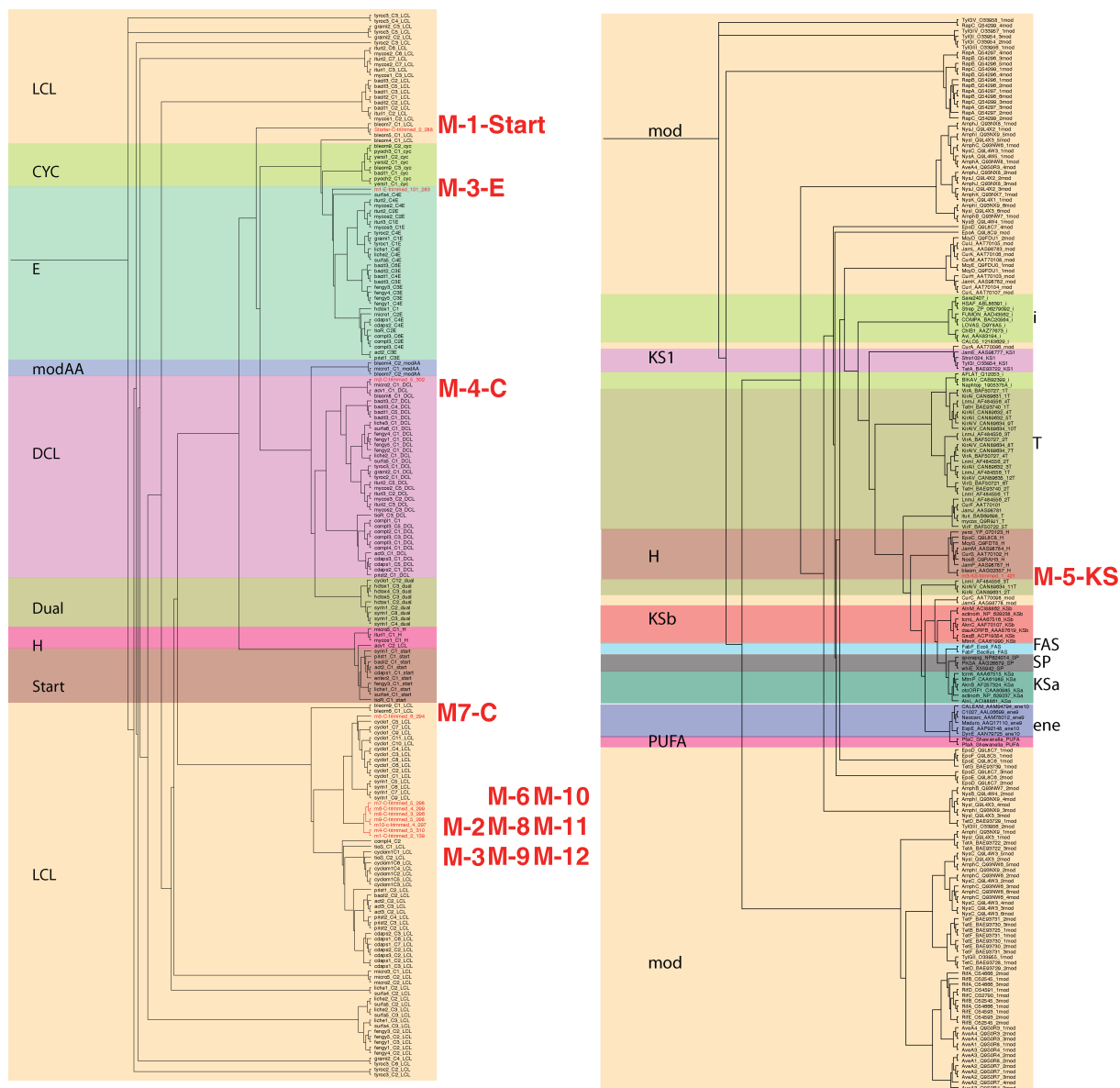


Figure 4.8: Phylogenetic tree of the corramycin assembly line C-domains (right) and the KS-domain (left) generated by the NapDos online server.²⁵

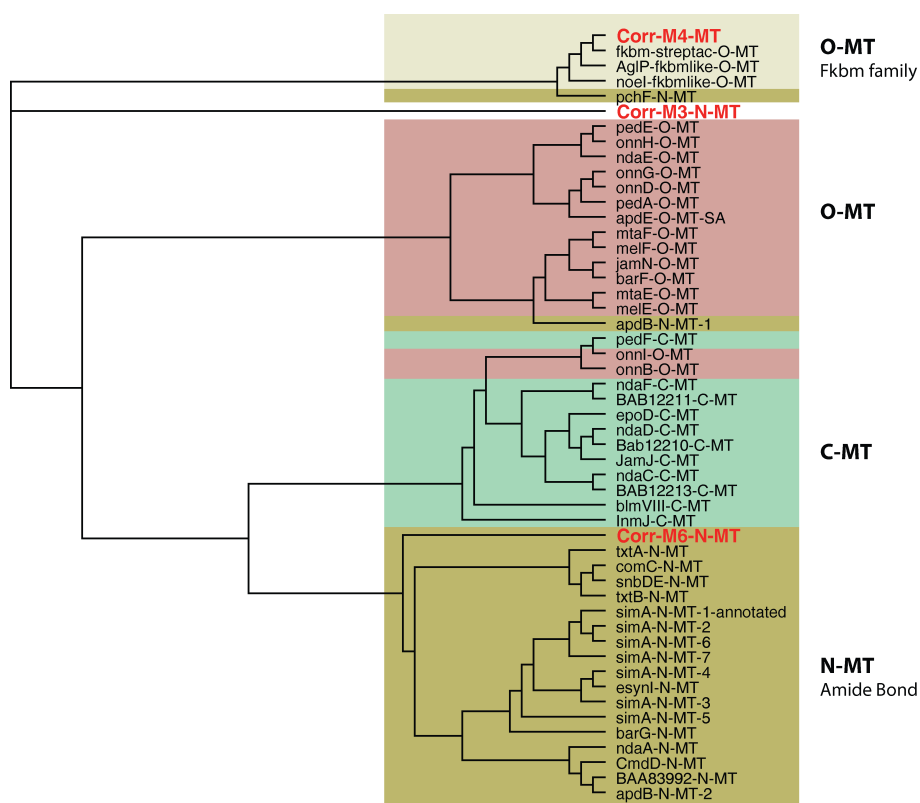


Figure 4.9: Phylogenetic tree of the three assembly line methyltransferases in the corrAMYCIN biosynthesis. The module 3 methyltransferase does not group with any published one, hence constituting a new clade.

4 Biosynthesis of Corramycin



Figure 4.10: Alignment of the FAAL starter module showing the insertion (blue) that distinguishes this class from the FACs.

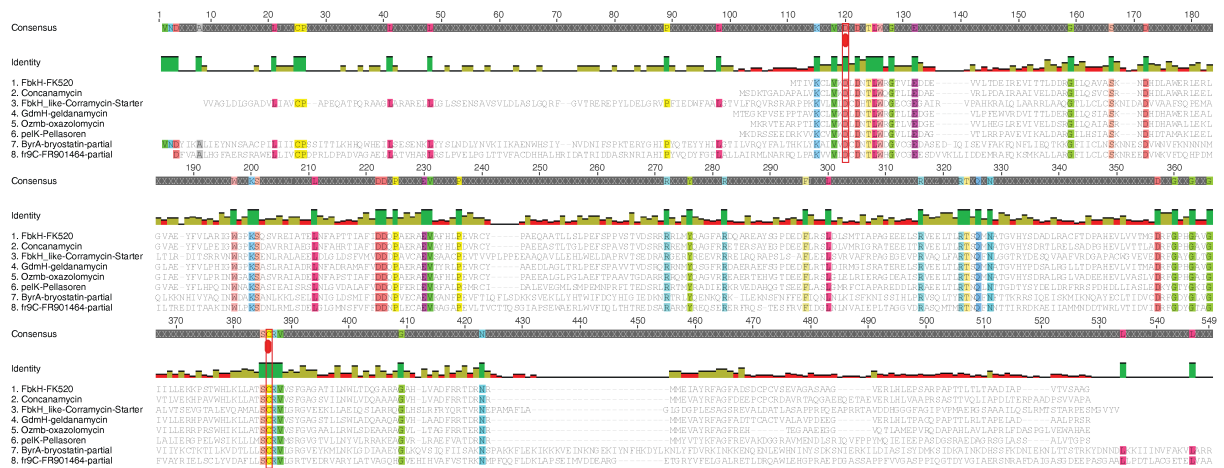


Figure 4.11: Alignment of the FbkH-like module 2 with the published sequences reveals the presence of all active centers necessary to incorporate glycerate.

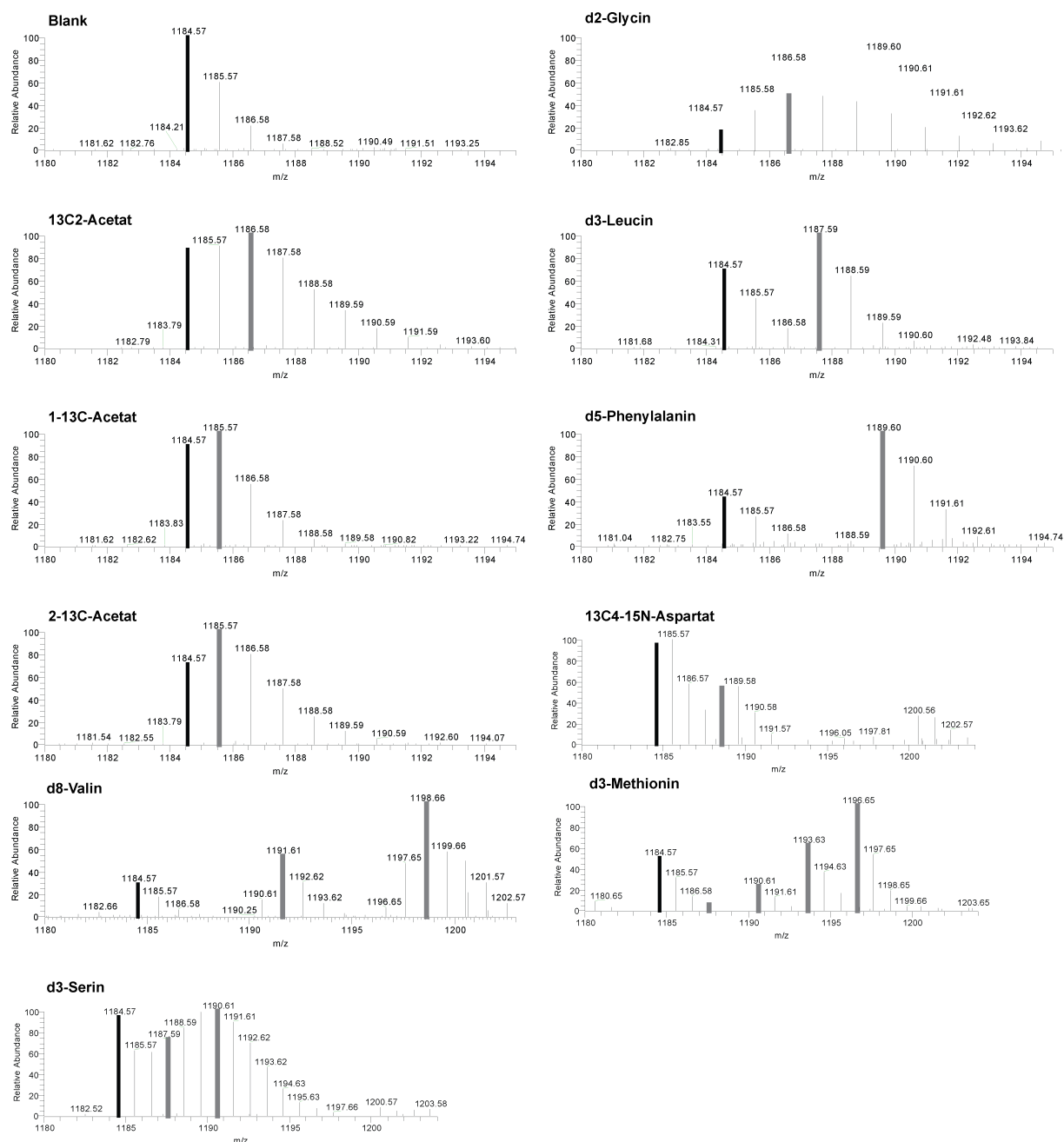


Figure 4.12: Shifts in the observed m/z value after supplementation of isotope labeled precursors to the fermentation broth. The respective precursor including the isotope labeling is stated.

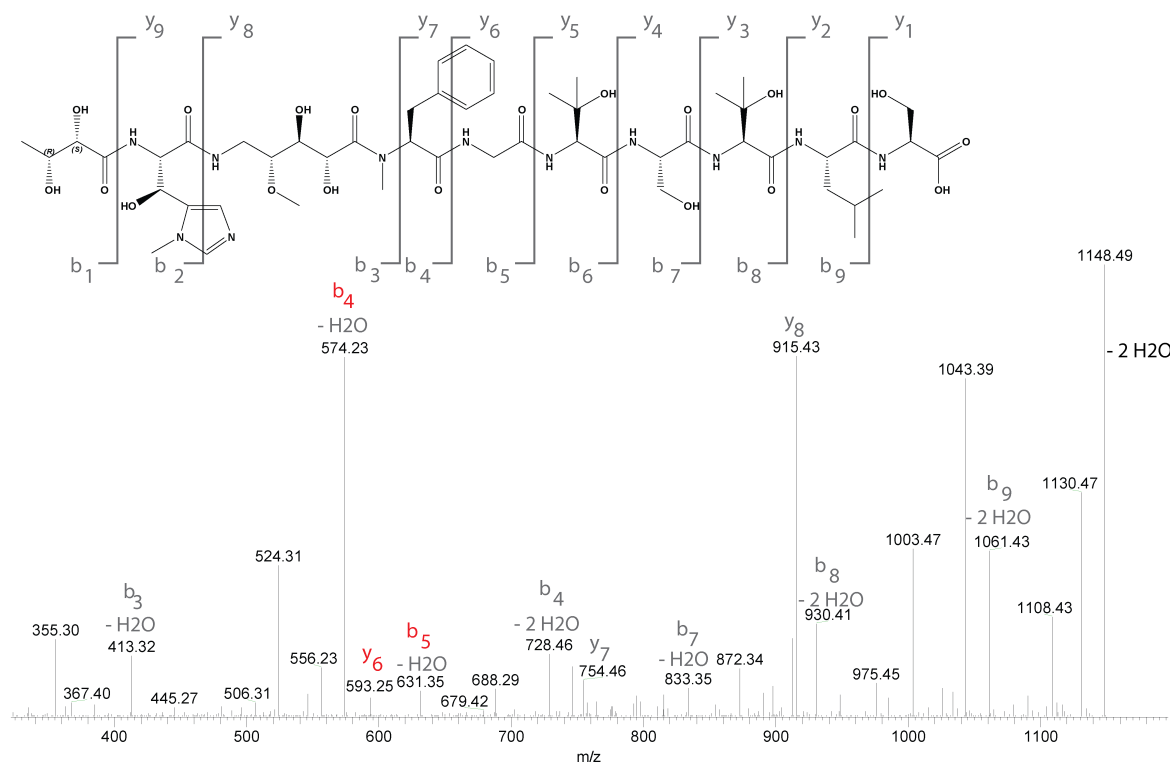
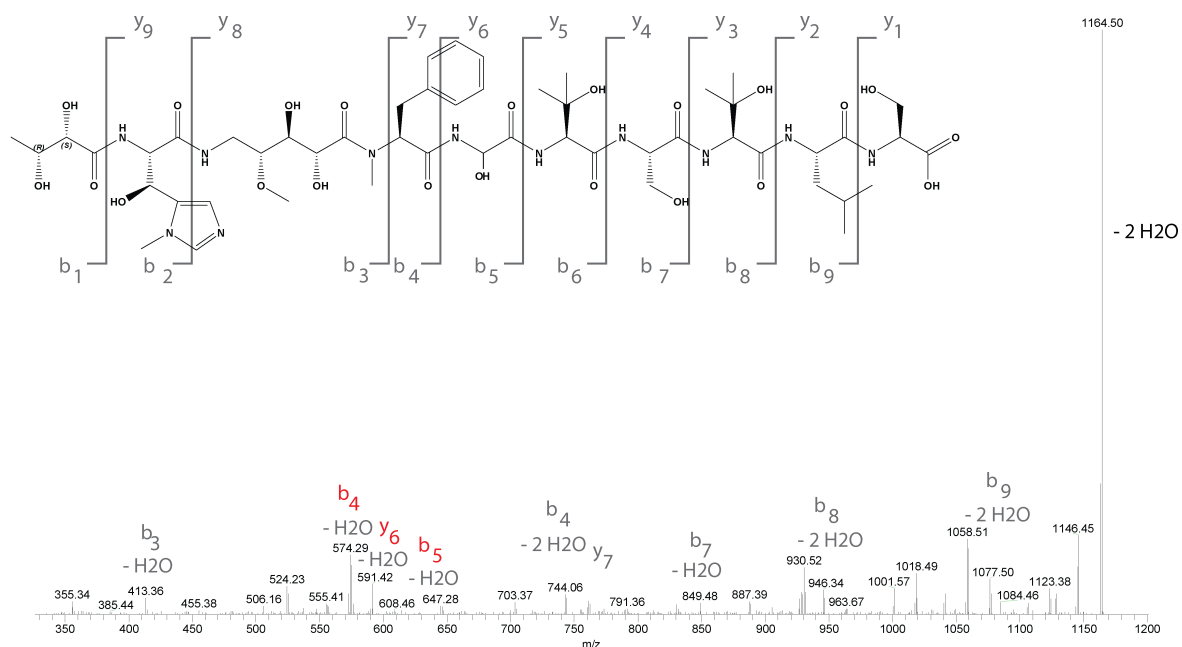
A**B**

Figure 4.13: ITMS³ spectra acquired for **A** corramycin (m/z 1184.6 \rightarrow 1166.6 \rightarrow) and **B** OH-corramycin (m/z 1200.6 \rightarrow 1182.6 \rightarrow). Fragments were assigned according to the Biemann modification of the Roepstorff nomenclature. Sequence ions marked in red are distinct between the two compounds and indicate the presence of α -hydroxyglycine instead of glycine in OH-corramycin. Experimental details can be found in the materials and methods section.

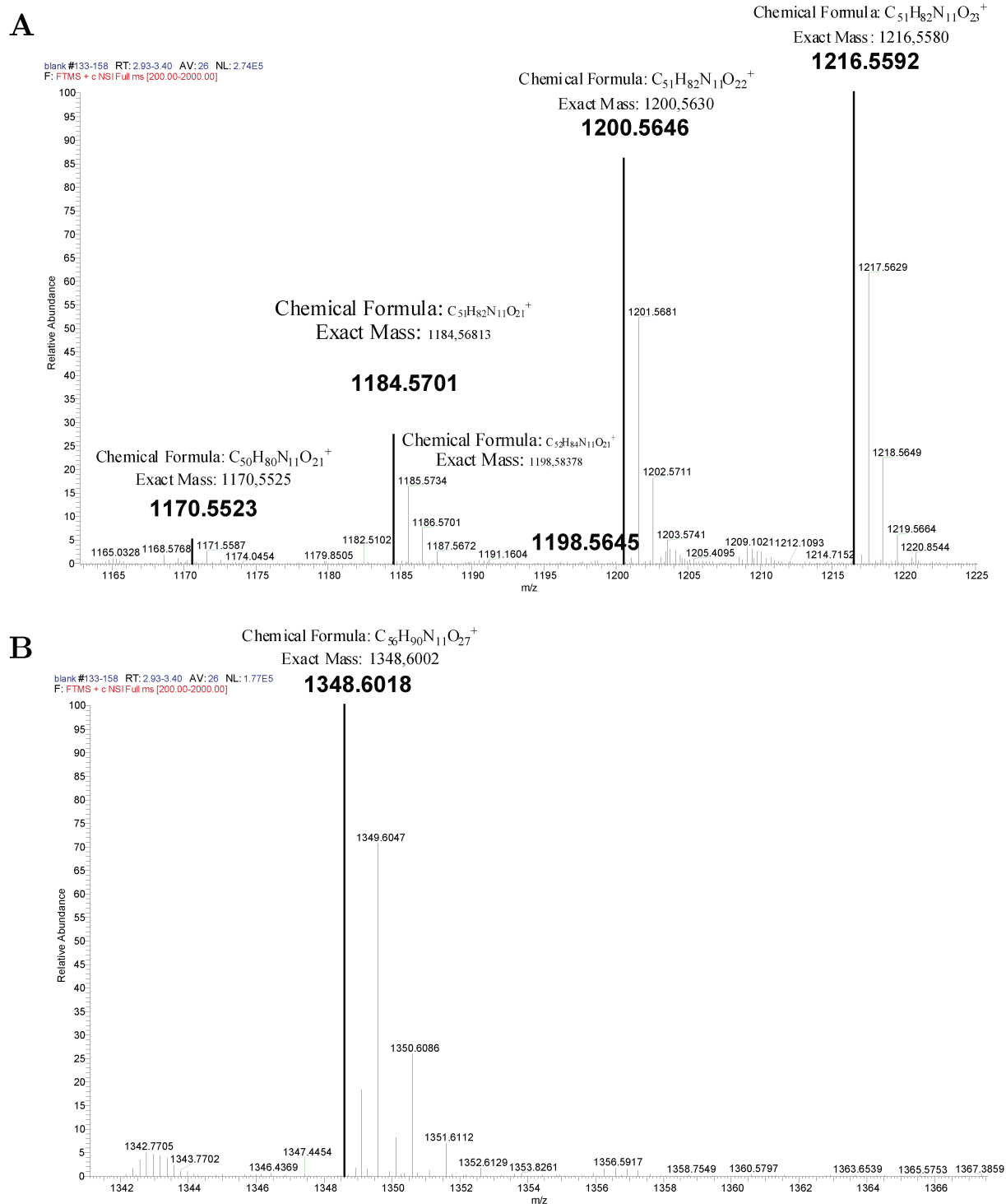


Figure 4.14: High-resolution mass spectra for all observed corramycin derivatives. Calculated exact mass and molecular formula for $[M+H]^+$ is given. **A** Corramycin 1170, 1184, 1198, 1200 and 1216 **B** Corramycin 1348

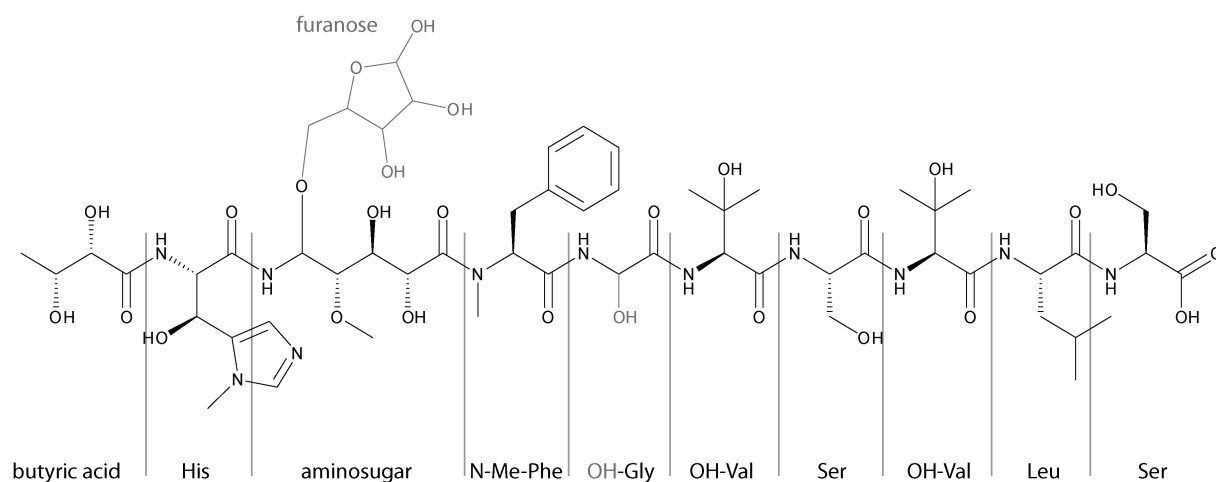


Figure 4.16: NMR position labeling for corramycin (black) and corramycin 1348 (grey). Numbering in NMR tables always starts at carbonyl carbon.

Table 4.4: NMR data of corramycin in D₂O. For structure reference see Figure S 4.16.

Position	Group	δ_C	δ_H
Ser			
1	C	175.7	
2	CH	57.0	4.29
3	CH ₂	61.9	3.85
Leu			
1	C	173.7	
2	CH	52.2	4.50
3	CH ₂	39.4	1.69
4	CH	24.1	1.68
5	CH ₃	22.0	0.95
6	CH ₃	20.4	0.89
β -OH Val			
1	C	171.3	
2	CH	60.6	4.44
3	C	71.6	
4	CH ₃	25.8	1.32
5	CH ₃	25.0	1.32
Ser			
1	C	171.0	
2	CH	55.3	4.61
3	CH ₂	60.8	3.88
β -OH Val			
1	C	171.3	
2	CH	60.52	4.56
3	C	71.8	
4	CH ₃	25.8	1.32
5	CH ₃	25.0	1.29
Gly			
1	C	170.71	
2	CH ₂	42.4	4.10
			4.01
N-Me-Phe			
1	C	172.1	
2	CH	59.6	5.37
3	CH ₂	32.8	3.09
			3.45
4	C	136.8	
5	CH	128.6	7.39
6	CH	128.9	7.34
7	CH	126.7	7.32
8	CH ₃	32.1	3.04
aminosugar			
1	C	173.7	
2	CH	68.2	4.58
3	CH	70.2	4.08
4	CH	78.4	3.58
5	CH ₂	38.16	3.58
6	CH ₃	56.9	3.30
β -OH- δ -Me-His			
1	C	171.05	
2	CH	55.9	4.96
3	CH	63.9	5.16
4	C	130.4	
5	CH	125.0	7.14
6	CH	139.1	7.90
7	CH ₃	32.1	3.81
butyric acid			
1	C	173.99	
2	CH	74.7	3.95
3	CH	67.9	3.44
4	CH ₃	17.7	1.11

Table 4.5: NMR data of corramycin 1348 in D₂O. For structure reference see Figure S 4.16.

Position	Group	δ_C	δ_H
Ser			
1	C	175.7	
2	CH	57.2	4.28
3	CH ₂	62.1	3.84
Leu			
1	C	173.7	
2	CH	52.4	4.47
3	CH ₂	39.4	1.68
4	CH	24.4	1.67
5	CH ₃	20.6	0.88
6	CH ₃	22.3	0.94
β -OH Val			
1	C	172.5	
2	CH	60.6	4.44
3	C	71.8	
4	CH ₃	25.9	1.31
5	CH ₃	25.2	1.27
Ser			
1	C	171.1	
2	CH	55.4	4.65
3	CH ₂	61.2	3.84
β -OH Val			
1	C	170.8	
2	CH	59.9	4.56
3	C	72.1	
4	CH ₃	26.7	1.32
5	CH ₃	25.2	1.27
α -OH-Gly			
1	C	169.6	
2	CH	72.1	5.86
N-Me-Phe			
1	C	172.5	
2	CH	58.4	5.63
3	CH ₂	32.1	3.00
			3.50
4	C	137.0	
5	CH	128.9	7.30
6	CH	128.6	7.38
7	CH	126.9	7.31
8	CH ₃	30.2	3.00
aminosugar			
1	C	173.5	
2	CH	69.7	4.47
3	CH	69.0	4.07
4	CH	79.3	3.75
5	CH	76.4	5.91
6	CH ₃	60.0	3.37
β -OH- δ -Me-His			
1	C	172.0	
2	CH	55.2	5.42
3	CH	64.7	5.07
4	C	131.1	
5	CH	130.5	7.15
6	CH	n.d.	n.d.
7	CH ₃	33.0	3.84
butyric acid			
1	C	173.5	
2	CH	74.9	3.95
3	CH	68.2	3.95
4	CH ₃	18.0	1.10
furanose			
1	CH	104.1	5.02
2	CH	81.6	4.04
3	CH	77.2	3.93
4	CH	84.3	4.10
5	CH ₂	61.1	3.73

Chapter 5

Corramycin self-resistance mechanism of *Corallococcus* *coralloides*

ALEXANDER VON TESMAR, DAVID AUERBACH, KIRSTEN HARMROLFS, STEPHANE RENARD, ROLF MÜLLER*

*To whom correspondence should be addressed.

Author's efforts

The author was responsible for the conception of the study, designed and performed experiments, evaluated and interpreted resulting data. The author designed and performed all experiments for the identification of resistance genes in the corramycin biosynthetic gene cluster. The author cloned, expressed and purified all proteins to prove the mode-of-action of corramycin inactivation by ORF-5. Furthermore, the author conducted all in vitro work with ORF-2. The author conceived and wrote the manuscript.

Contributions by others

David Auerbach purified phosphate-corramycin. Kirsten Harmrolfs aided the structural elucidation of phosphate-corramycin by NMR. Stephane Renard (Sanofi) provided corramycin and background information. The project was supervised by Rolf Müller.

Abstract

Corramycin, a potent antibacterial linear peptide, is produced by several *Corallococcus coraloides* species. Resistance development studies in *E. coli* were employed to identify two genes in the biosynthetic gene cluster, ORF-2 and ORF-5, that confer resistance against corramycin. ORF-5 inactivates corramycin by phosphorylation, a similar resistance mechanisms already known for several glycoside antibiotics, including kanamycin. ORF-5 kinase activity was reconstituted *in vitro* and kinetic parameters determined. Phosphate-corramycin was purified from the *in vitro* assay and the structure determined by NMR. Bioactivity testing furthermore confirmed the complete loss of activity towards *E. coli*. ORF-2 is a protein of unknown function and is part of a second, independent resistance mechanism in the corramycin producer. ORF-2 has weak ATPase activity. Hit-1, the closest homologue of ORF-2 in the producer genome, as well conferred corramycin resistance to *E. coli* upon expression and therefore is most likely involved in the mode-of-action of the antibiotic. Although the underlying molecular mechanisms remain elusive, ORF-2 most likely acts as a corramycin insensitive substitute for Hit-1 and thus enables resistance.

5.1 Introduction

Antibiotics are one of the main constituents of modern medicine. Not only are they among the most prescribed drugs but are also widely used in agriculture as prophylactic agents in stock breeding.¹ The widespread use of antibiotics like sulfonamide, penicillin and streptomycin in the 1930s and 1940s initiated the golden age of antibiotics. In the upcoming three decades the discovery of most of the important antibiotic classes in use today took place. Since the 1960s the development of novel antibiotics was discontinued due to the general impression that infectious diseases have been defeated and the momentous undervaluation of the rampant emerging of resistance.²⁻⁴ In 2001 the world health organization mentioned the resistance problematic for the first time in its annual report on public health, ten years later it was already the main topic on the WHO world health day.⁵ Recently, the world economic forum categorized antibiotic resistance of pathogens as a major threat in the annual global risk report. Consequently the development of novel antibiotics and the successful combat against the spread of resistance is inevitable to preserve the current level of health-care.

The understanding of underlying mechanisms of bacterial resistance against antibiotics is vital to develop successful counter-strategies. Bacteria achieve resistance by three general mechanisms: First, the production of enzymes that inactivate antibiotics like the penicillin hydrolysing β -lactamases. Second, a mutation in the binding pocket of the molecular target of the antibiotic that inhibits binding. This can either occur by a mutation in the respective gene or the expression of a second, altered version of the target. Third, efficient efflux systems that transport the antibiotic out of the cell.⁶ As many antibiotics are toxic to their producing strain, resistance conferring genes are often found within the biosynthetic gene cluster.⁴ These genes can aid pathogens to acquire antibiotic resistance, when they are transferred horizontally.⁷ Identification of self-resistance mechanism is thus valuable for both, the identification of the mode-of-action and the prediction of possible resistance pathways in pathogens.

Corramycin, a peptide with Gram-negative antibacterial activity, is produced by several *Corallococcus coralloides* species. The biosynthesis is facilitated by a NRPS/PKS assembly line that is encoded in a 55 kb gene cluster. Here, we present two genes that confer corramycin resistance in the producing strain *Corallococcus coralloides* STST201330.

5.2 Results

Heterologous expression in *E. coli* identifies ORF-2 and ORF-5 as corramycin resistance proteins.

The strong antibiotic activity of corramycin against Gram-negative bacteria enabled the development of a rapid 96-well plate based method using *E. coli* BL21 (DE3) to examine genes with a putative resistance conferring function. Respective genes were cloned tag-less into pET-28b and subsequently expressed in *E. coli* BL21 (DE3) while different concentrations of corramycin were supplemented. Increasing viability in presence of corramycin was observed upon expression of resistance genes. Resulting changes in the IC₅₀-values were quantified by monitoring the OD₆₀₀-values. Coexpression of two genes was facilitated using pET-28b and the compatible vector pAvT1, that was generated from pET-28b and pACYC177. The multiple cloning site (MCS)

including promoter region and *lacI* repressor gene from pET-28b were combined with the p15a origin of replication (ori) and the ampicillin resistance cassette from pACYC177. Incubation was carried out in 96-well plates with increasing concentrations of corramycin.

Corramycin exhibited an IC_{50} of 15 nM on *E. coli* BL21 (DE3) transformed with an empty vector as negative control. A 10-fold increase in resistance could be observed upon transcription of RecF-mRNA, that was missing a ribosomal binding site due to an error in primer design and was thus not translated into the respective protein. Expression of ORF-1, the putative exporter of the fatty acid linked corramycin precursor, or the aspartate decarboxylase ORF-3 as additional negative control led to a similar degree of resistance. Consequently, the basal IC_{50} value to whom all further experiments were compared is 250 nM and is calculated as the mean value from the combined data of resistance development assay of ORF-1, ORF-2 and RecF-mRNA. Expression of ORF-2, a gene of unknown function, increased the corramycin resistance of *E. coli* BL21 (DE3) by a factor of 10^3 with a potential IC_{50} -value larger than 100 μ M, the highest concentration used. The addition of an N-terminal SUMO-tag to ORF-2 did hamper the resistance development, but a 10-fold increase remained. The expression of ORF-5, a protein with the conserved domain architecture of the aminoglycoside kinase (APH) family, led to a degree of resistance similar to the expression of ORF-2. The co-expression of both genes, ORF-2 and ORF-5, led to insensitivity up to a corramycin concentration of 100 μ M (Figure 5.1 C). ORF-2 as a protein of unknown function was intensively analyzed *in silico* (see 5.2). Two BLAST hits were subsequently heterologously expressed to assess a possible resistance development. The expression of the first BLAST hit of ORF-2 in the UniProtKB/Swiss-Prot database, *E. coli* RecF, did not increase resistance. The first BLAST hit of ORF-2 in the producer genome (*Corallococcus coralloides* STST201330), named Hit-1, did increase the resistance by a factor of around 10.

Table 5.1: Corramycin IC_{50} values for *E. coli* BL21 (DE3) transformed with different inserts.

Insert	Vector	IC_{50}
-	pET-28b	15 nM
-	pAvT1	13 nM
ORF-1	pET-28b	106 nM
ORF-3	pET-28b	102 nM
mRNA	pET-28b	126 nM
RecF	pET-28b	108 nM
ORF-5	pET-28b	> 100 μ M
ORF-2	pET-28b	> 100 μ M
ORF-5/ORF-2	pET-28b/pAvT1	> 100 μ M
Hit-1	pET-28b	7.8 μ M
SUMO-ORF-2	petM44	10.4 μ M

ORF-5 inactivates corramycin by phosphorylation

SCOP domain search using the hidden Markov model identify ORF-5 as part of the protein-kinase-like (PK-like) superfamily (e-value $2.03 \cdot 10^{-33}$) and further as part of the aminoglycoside kinase (APH) family (e-value 0.046), although the low e-value indicates a weak prediction. The APH family of enzymes is responsible for bacterial self-resistance against different aminoglycoside antibiotics like kanamycin and hygromycin by phosphorylation of the compound. Heterologous expression of ORF-5 was employed to analyze the underlying mechanism of corramycin resistance development. ORF-5 was cloned into pET-28b, expressed in *E. coli* BL21 (DE3) and the respective 6xHis-tagged protein was purified to homogeneity in a two-step protocol comprising nickel-affinity and size exclusion chromatography. Co-purification of the interfering *E. coli* chaperon GroEL could be avoided by addition of ATP to the Ni-NTA-washbuffer. Upon incubation of ORF-5 and ATP with corramycin the appearance of a novel m/z value with a mass shift of 80 Da could be observed. MS/MS analysis of the m/z signal 1264.637 indicated the presence of phosphorylated corramycin (P-corramycin). The retention time was not affected by the additional phosphate group. A 96-well plate based assay was used to determine the Michaelis-Menten kinetic parameters by quantifying the initial velocity at five different substrate concentrations using EIC peak areas of singly and doubly-protonated species. (Figure 5.2).

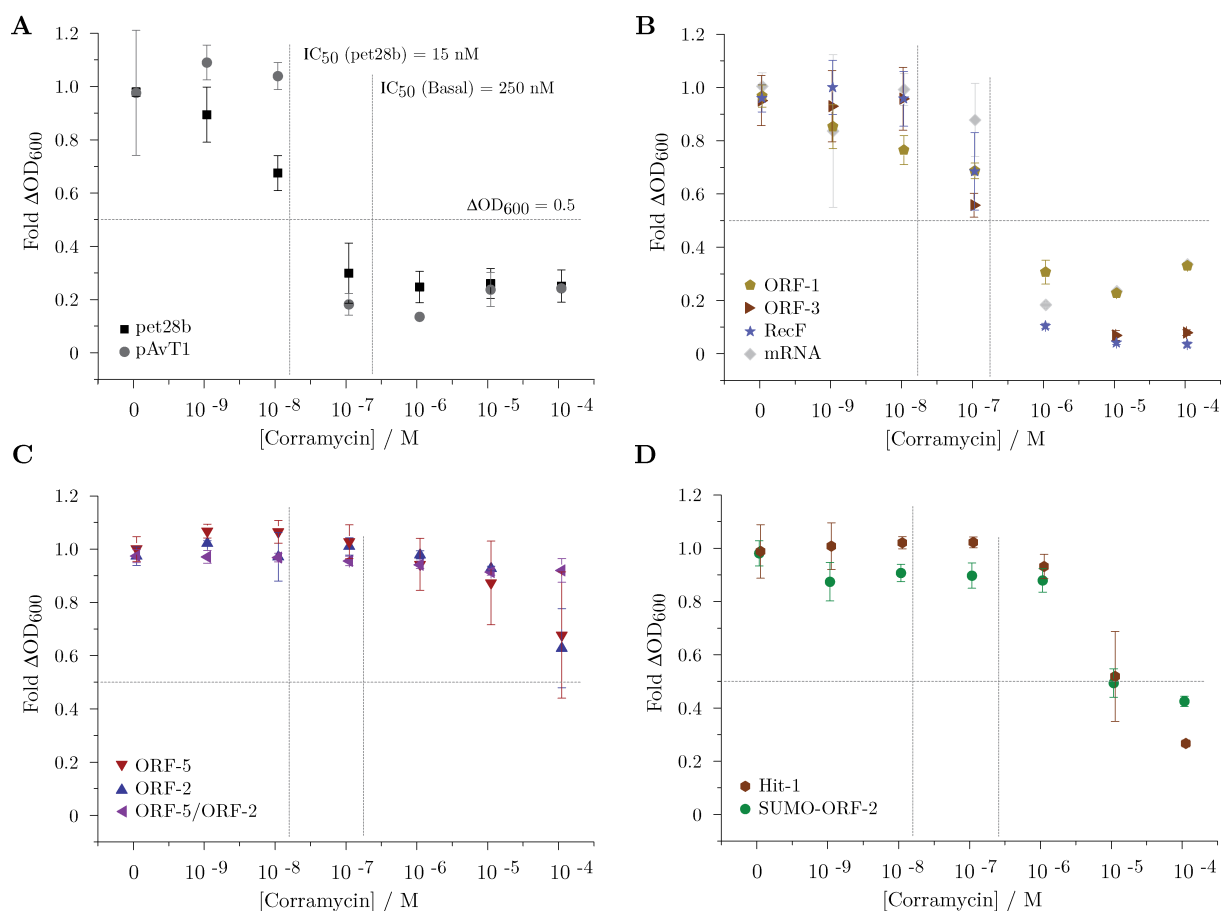


Figure 5.1: Results of the resistance development experiment in *E. coli* BL21 (DE3). **A** Empty vector control **B** Basal corramycin resistance while expression of unrelated genes like ORF-1, ORF-3 or RecF or the transcription of non-coding mRNA. **C** Corramycin resistance increased by three magnitudes upon expression of ORF-2 or ORF-5. **D** Resistance increased by the factor 10 upon expression of Hit-1 or the N-terminal tagged SUMO-ORF-2.

To elucidate the location of the phosphate group and assess the antibacterial properties, purification of P-corramycin was conducted. Since P-corramycin could not be detected in the extract of the two *Coralloloccus coralloides* producer strains, it was purified from an upscaled *in vitro* assay. Superimposing of the HMBC-spectra of corramycin and phosphate corramycin identified the characteristic ^1H -shift of 0.3 ppm and ^{13}C -shift of about 3 ppm on the C_3 -atom of the β -alanine-PKS moiety. The neighboring atoms C_2 and C_3 underwent smaller shifts in the observed ppm values. The phosphate group introduced by ORF-5 is located at the C_3 -hydroxy group originating from the carboxyl-group of β -alanine (see figure 5.3).

Resistance development assay with *E. coli* BL21 (DE3) and P-corramycin showed no antibacterial activity, clearly proving the inactivation of corramycin via phosphorylation by ORF-5.

ORF-2 is a corramycin resistance protein with unknown function

ORF-2 clearly confers resistance to corramycin in the *E. coli* based screening platform. Nevertheless, the molecular function as well as the mechanism of resistance remain elusive. SCOP

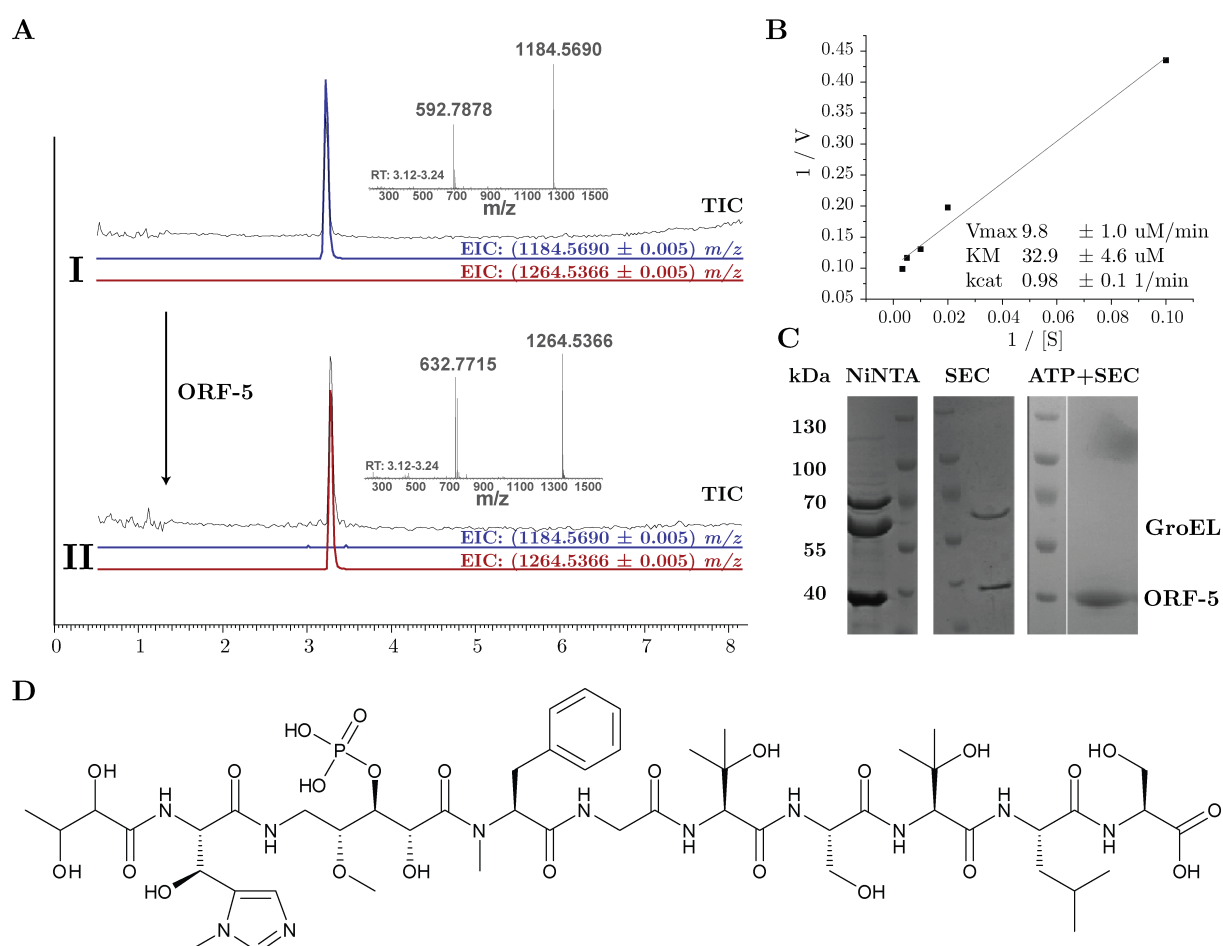


Figure 5.2: **A** LC-MS analysis of the *in vitro* incubation of ORF-5, ATP and corramycin indicates the formation of phosphate-corramycin. **B** Lineweaver-Burk plot of the phosphate-corramycin formation reaction generated by quantification of EIC signal areas of P-corramycin. Linear fit was used to calculate kinetic parameters. **C** SDS-PAGE analysis of Ni-NTA and SEC purified ORF-5 indicates the presence of co-purified GroEL. Addition of ATP to the washing buffer removed the impurity. **D** Structure of P-corramycin.

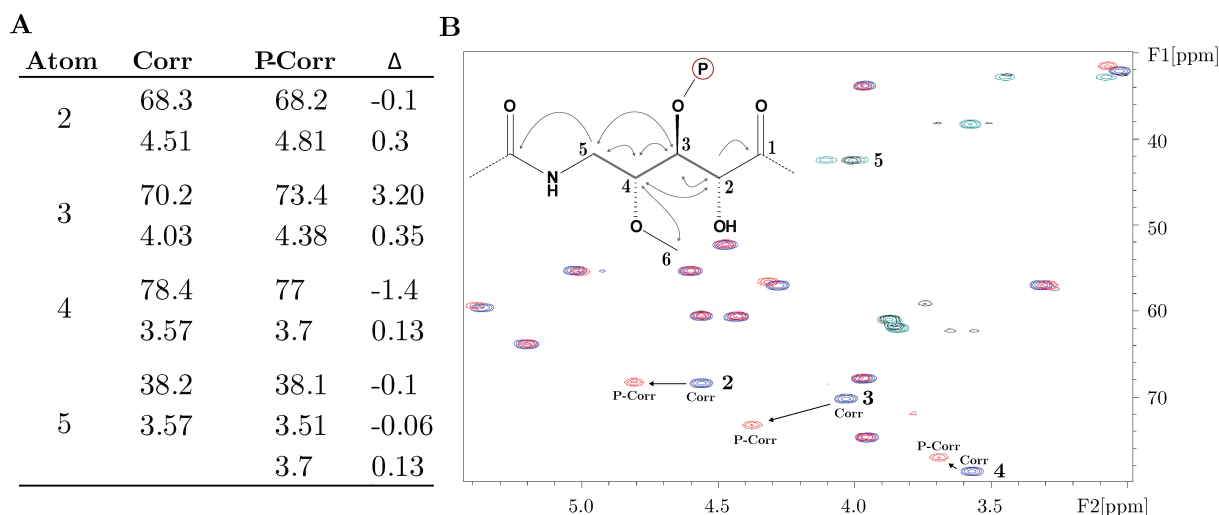


Figure 5.3: **A** Table of the observed ppm-shifts in the β -alanine-PKS moiety. **B** Superimposed HMBC-spectra of corramycin (blue/dark green) and phosphate-corramycin (red/green) localize the phosphate-group at atom C₃.

domain search classify ORF-2 as member of the superfamily of P-loop containing nucleoside triphosphate hydrolases (P-loop_NTPase) (e-value $2.96 \cdot 10^{-22}$). Furthermore, a weak classification as part of the ABC transporter ATPase domain-like (ABC) family (e-value 0.039) was made. BLAST analysis using the non-redundant database (nr) only identified likewise hypothetical proteins with no experimental data available. Using the curated swissprot database, the highest ranked hit with score of 10^{-3} is RecF, a protein involved in DNA repair. Direct sequence alignment with ORF-2 indicated that only a small fragment of ORF-2 and RecF exhibit significant sequence homologies and no overall similarity exists, making the analysis unsuitable to assess ORF-2 activity.

One possible mechanism of bacterial resistance is surrogation of the target with a second, antibiotic insensitive variant. Thus, BLAST analysis using the genome of the producer *Coralloccoccus coralloides* STST201330 as database was employed to search for a potential gene involved in the corramycin mode-of-action, that is surrogated by ORF-2 to establish resistance. Only one gene encoding a protein of unknown function with a significant E-value of 10^{-115} was retrieved and subsequently termed Hit-1. Hit-1 conferred a 10-fold increased resistance to corramycin in the *E. coli* based screening, indicating an involvement in the mode-of-action. The gene coding for 367 amino acids has a sequence-identity of 58.1 % to the 383 amino acids of ORF-2. Conserved domain search furthermore identified a similar domain organization. As expected, BLAST analysis of Hit-1 yielded RecF like proteins in the non-redundant databases with small regions of homology. This small stretch of homology in the N-terminus harbors a conserved Walker motif, revealed by alignment of ORF-2, Hit-1 and the RecF from *Vibrio fischeri*. The motif GXXXXGK(T/S) is present in all three proteins and is known to facilitate nucleotide binding.⁸ This N-terminal domain is responsible for retrieving RecF in all conducted BLAST analyses. Nevertheless, the large C-terminal part of ORF-2 seems to contain parts of another ATP binding motif, namely Walker B, indicated by the NCBI conserved domain search. The corresponding consensus sequence (R/K)XXXXGXXXXLhhhhD is albeit not complete in ORF-2 and Hit-1. Although the WalkerB motif sequence is somehow flexible, the main criteria, a charged amino acid (e.g. aspartate) following a stretch of hydrophobic amino acids, is not fulfilled (RecF LL-

LLD; Hit-1 LEPLG; ORF-2 IAPLA).^{8–10} If this second motif is binding nucleotides and what function the C-terminal domain exhibits can not be predicted with *in silico* analysis alone, thus heterologous expression of ORF-2 was pursued.

Purification of heterologously expressed ORF-2 to conduct *in vitro* experiments was albeit hampered by poor solubility. The 6xHis tagged ORF-2 using the pET-28b vector showed sufficient overexpression, but eluted as precipitate from the Ni-NTA column. Assuming DNA binding is a possible requisite for stability, shared gDNA was added to the elution buffer, but without success. Eventually, N-terminal fusion with a SUMO-tag enabled the purification of soluble protein using a two-step protocol comprising Ni-NTA and SEC. The handling remained challenging and cleavage of the tag induced protein precipitation. The observed impurity of *E. coli* ArnA, a known Ni-NTA co-purification, could not be removed.

Ni-NTA based pull-down experiments were employed using purified SUMO-ORF-2 and *E. coli* BL21 (DE3) lysate to identify potential binding partners. SUMO-ORF-2 still confers resistance to corramycin, hence putative binding partners are still thought to bind. Nevertheless, the pull-down experiment led to the accumulation of the known impurity ArnA, while no other additional band could be identified in the SDS-PAGE analysis.

Based on the presence of motifs related to nucleotide binding, ATPase activity of SUMO-ORF-2 was assessed. Indeed, a weak ATPase activity could be detected in a malachite-green based assay. Addition of shared gDNA or corramycin as possible substrates to increase activity had no impact (Figure 5.4).

As a consequence of the poor solubility and difficult handling of purified ORF-2, an alternative *in vivo* approach was developed. A pET-28b vector harboring an N-terminal GFP-fusion of ORF-2 was sent to the lab of Theresia Stradal in Braunschweig to set up localization assays in *E. coli* and *Shigella*. ORF-2 function in *E. coli* has already been proven in the presented resistance development assay, hence potential interactions with localized partners will likely be visible.

Summarized, ORF-2 was successfully identified as resistance protein. Nevertheless, the molecular function and mechanism of resistance of ORF-2 and Hit-1 could not be elucidated.

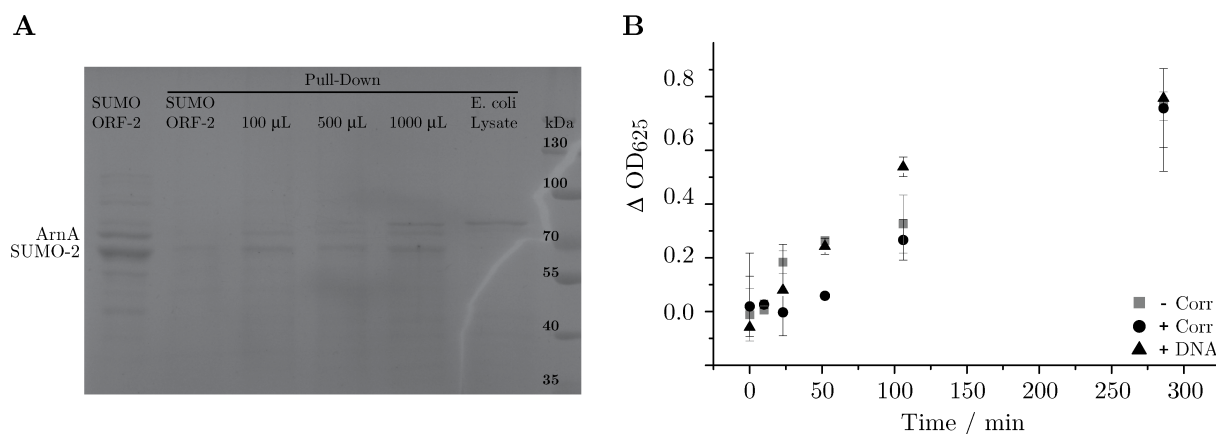


Figure 5.4: **A** SDS-PAGE analysis of purified SUMO-ORF-2 and the pull-down experiment. ArnA is a known impurity that could be removed and also accumulated during the pull-down experiment. **B** Basal ATPase activity of ORF-2 visualized using a malachite-green based assay.

5.3 Discussion

Producer of antibacterial compounds usually possess an intrinsic resistance mechanism as a safeguard. In the corramycin producing *Coralloccoccus coralloides* STST201330 several redundant components ensure self-resistance. The first one is encoded in the biosynthetic machinery. Analysis of the NRPS/PKS assembly line responsible for corramycin production revealed a fatty-acid activating first module, similar to the one found in vioprolide.¹¹ This fatty-acid linked corramycin is thought to be exported to the periplasma by ORF-1 and afterwards activated through cleavage by ORF-0. The production of a defused version and the subsequent export is a first intrinsic resistance mechanism. The presented study reveals two additional genes that confer self-resistance: one functions via chemical inactivation and a second by an unknown mechanism, presumably involving target replacement. Hence, all three categories of antibiotic resistance mechanisms in bacteria, namely export, modification and surrogation, are likely present in the corramycin producer.

ORF-5

ORF-5 inactivates an antibiotic by chemical modification, a common self-resistance strategy in bacteria. β -lactamases or aminoglycoside kinases (APH) are widespread among multi-resistant pathogens and render whole classes of antibiotics inactive. APHs transfer the γ -phosphate group from an nucleotide triphosphate to a hydroxy-group of an aminoglycoside, that subsequently loses its antibacterial activity.¹² As a phosphate donor ATP is most common, but recent studies showed the ability of some APHs to use GTP instead or in addition.^{13–15} ORF-5 uses the same mechanism, the transfer of a phosphate group to a hydroxy group, to inactivate the myxobacterial produced antibiotic corramycin. SCOPE domain analysis indicated a weak prediction for the APH family. The alignment of the amino-acid sequence of ORF-5 and two representative APHs indeed showed significant similarities (Figure 5.5). Over 40 crystal structures of APHs have been solved and submitted to the PDB, revealing a common architecture. APHs have two distinct domains, a N-terminal and a C-terminal one, that are connected by a hinge region. The N-terminal part consists of five antiparallel β -strands and two α -helices. The C-terminal domain can be further divided into two subdomains, the core- and the helical-domain. The core-subdomain and the N-terminal domain together harbor the conserved features for nucleoside triphosphate binding and the catalytic center. This catalytic domain is highly similar to the eukaryotic protein kinases. The helical subdomain on the C-terminus is responsible for aminoglycoside recognition and therefore more diverse among APHs.¹⁶

Although ORF-5 does not group well into the APH family, the key features are present. The N-terminal domain of ORF-5 contains similar secondary structure elements as APHs, as predicted by the EMBOSS algorithm.¹⁷ Most of the conserved residues for nucleotide triphosphate binding, like G30 and E72, are present. The hinge region is also followed by a long α -helix labeled as α 3 in APHs. The catalytic region in the core-subdomain of the C-terminus is also highly similar, especially around the catalytic loop. Unsurprisingly the highest diversity is found within the helical subdomain of the C-terminus, that is responsible for substrate recognition. Two large insertions around α C illustrate the difference in substrates, a large peptide for ORF-5 and small, highly cationic aminoglycosides for APHs. Steady-state kinetic parameters established for ORF-

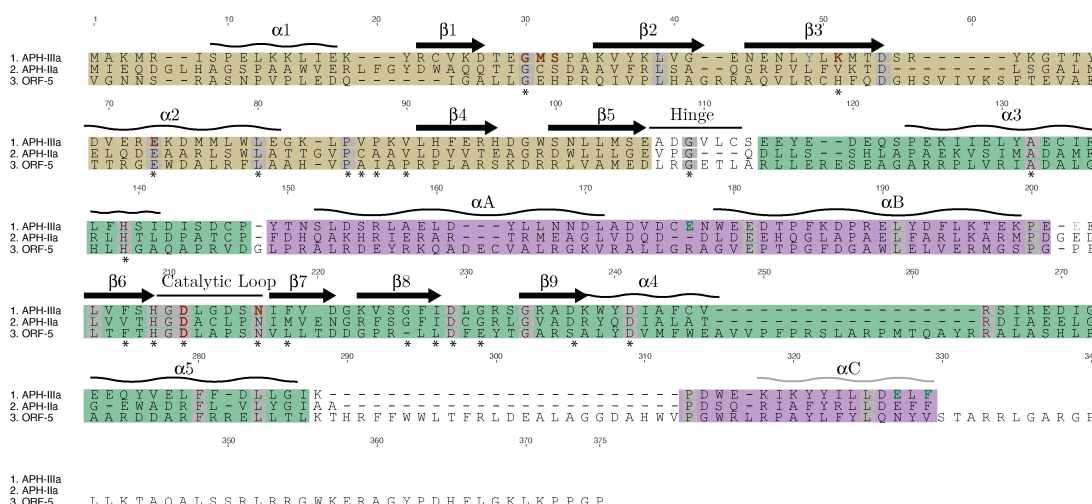


Figure 5.5: Amino acid sequence alignment of ORF-5 with two representative members of the APH-family, Ila and IIIa. The secondary structure elements for IIIa are indicated in black, if they are also present in ORF-5, otherwise in grey. Conserved amino acids are indicated with an asterisk according to alignments of the APH-family.¹⁵ N-terminal domain is colored brown, the C-terminal core-subdomain green and the helical-domain purple. Amino acids involved in nucleoside-triphosphate binding of IIIa are marked brown, the catalytic center in red. Green residues are involved in the catalytic pocket.

5 are comparable to APH(2'')-IVa,¹⁸ further approving the corramycin inactivation process as closely related to the ones for aminoglycoside antibiotics.

The overall amino acid sequence similarity of the PK superfamily is very low and the common motifs are mostly only recognizable on a structural level. The obvious similarities between ORF-5 and the APH-family, the two-domain organization, secondary structure and the few conserved residues, are also the common features of the whole PK superfamily. Unfortunately, the sequence diversity makes further phylogenetic analysis solely based on amino acid information useless. A successful study has to take structural information into account that are not yet available for ORF-5. This is a known problem and studies focusing on the structural evolution for the PK superfamily, like the work of Scheeff et al., could only be conducted by combining sequence and structural information. The low e-value for the APH affiliation of ORF-5 and the different substrate class might indicate the presence of a new family, responsible for resistance against peptide antibiotics through phosphate transfer, of that no member has been characterized yet.

ORF-2

The experimental and *in silico* results for ORF-2 are insufficient to elucidate the mechanism of corramycin resistance, but they indicate the involvement of an yet unknown DNA-related process in *Coralloccoccus coralloides*. The confident SCOP classification of ORF-2 in the P-loop_NTPase superfamily is in accordance with the observed ATPase activity *in vitro*. The affiliation to the ABC family has a low e-value of 0.039 and is thus speculative. The ABC family is one of the largest superfamilies known and exists in all three kingdoms of life. The underlying ATP-binding cassette (ABC) has a highly conserved primary sequence and consists of three short motifs: The WalkerA and B motif constitute the nucleotide binding site and the ABC-signature motif.^{8–10} Sequence alignment of Hit-1, ORF-2 and RecF as a member of the ABC-family not involved into transport reveals the incomplete conservation of the WalkerB motif as well as the lack of the

Table 5.2: Phyre2 results for ORF-2 and Hit-1. For each template the sequence identity (Ident.) and the confidence (Conf.) is given. The first hit was used to generate a model of the query and the resulting coverage (Cov.) is stated

Template	Function	Hit-1 / %			ORF-2 / %		
		Ident.	Cov.	Conf.	Ident.	Cov.	Conf.
SMC	Chromosome partition protein	20	88	100	17	81	100
SMC-3	Structural maintenance of chromosomes	12	-	100	12	-	100

ABC signature. Most members of the ABC-family are transmembrane transporters, but some are also involved in DNA repair or translation.¹⁹ LmrA is an ABC-transporter from *Lactococcus lactis* and confers multi-drug resistance (MDR) by facilitating ATP-dependent efflux of several antibiotics like daunomycin, doxorubicin or viblastine from the cell.²⁰ ORF-2 and Hit-1 however are omitting any transmembrane region or putative interaction partners, hence a function as transporter is unlikely.²¹ A growing part of the ABC-family is involved in mechanisms outside transporting. The UVR-type proteins contain the ABC motif and are engaged in DNA repair (for example RecF²²), but also resistance against antibiotics.¹⁹ DrrC is an ABC-protein from *Streptomyces peuceii* that is involved in resistance against the DNA-intercalator daunorubicin. It is proposed that DrrC uses the energy of ATP hydrolysis to remove the agent from the DNA and thus conferring resistance.^{23, 24} RecF is part of the RecFOR pathway, which repairs single and also doublestrand breaks in DNA via homologous recombination.²⁵ Nevertheless, RecF-like function of ORF-2 is unlikely due to two reasons: First, the alignment shows low overall similarity and the BLAST-analysis in the non-redundant database only results in hypothetical, mostly myxobacterial, proteins, that exhibit similar low alignment scores with RecF. Second, a confident BLAST hit in the producer genome, Hit-1, is present additional to a RecF copy. Therefore RecF is merely the closed relative that has been characterized and not the real homologue. More likely ORF-2 is a corramycin resistant version of Hit-1. The increasing resistance upon over-expressing Hit-1 as well as the failure of RecF to do so support this hypothesis.

Methods solely relying on primary sequence proved to be not suitable for predicting the function of ORF-2 and Hit-1. Therefore structural alignments using the Phyre2 webserver were employed and resulted in the same hit table for both query proteins, headed by the chromosome partitioning protein smc from *Geobacillus stearothermophilus* and SMC-3 from yeast.²⁶ The signature member of the SMC family is the eukaryotic cohesin complex. It is vital for several DNA associated tasks like sister chromatid cohesion, repair of double-strand breaks and the modulation of interphase chromatin structure and transcription.^{27, 28} The cohesin complex is a three-membered ring made from SMC-1, the closely related SMC-3 and an α -kleisin unit.²⁹ SMC-3 has two distinct parts, a linear coiled-coil region and the ATPase-domain containing head.²⁶ Hit-1 exhibits significant structural homologies to the SMC-3 head. The N- and C-terminal ATPase heads of SMC proteins, that are separated by the large, 50 nm long coiled-coil, come together to form a functional ATPase. Hit-1 and ORF-2 resemble this head ATPase domain while omitting the coiled-coil region in the center. SMCs as well as the bacterial homologue MukB use ATP hydrolysis for conformational change to encircle DNA.³⁰ The low ATPase activity found for ORF-2 could therefore be not the result of a missing substrate, but of a conformational change that is only possible once per protein. Nevertheless, SMC/MukB like function is impossible due to the

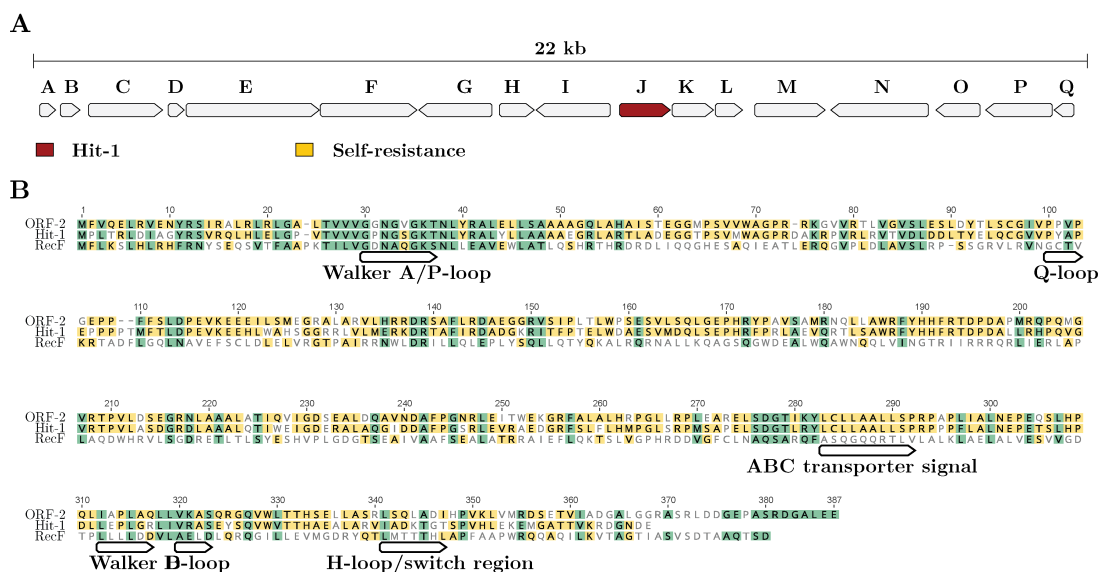


Figure 5.6: A 22-kb range genetic surrounding of Hit-1 (J). Detailed BLAST analysis is listed in table 5.4 **B** Alignment of ORF-2, Hit-1 and RecF with highlighted signature motifs of the ABC-superfamily (based on RecF sequence). Similar residues on all three chains is marked in green while similarities on two chains is marked yellow.

lacking coiled-coil region, but an involvement in one of the many processes involving the MukB complex is highly likely.

Hit-1 is located in a genetic environment that is rich in regulatory genes. 16 open-reading frames surrounding Hit-1 in a span of 22 kb were analyzed by BLAST using both, the non-redundant (nr) and the curated (Uniprot/SwissProt) database (Table 5.4). Non-redundant search queries resulted exclusively in hits from the producer species *Coralloccoccus coralloides* with no experimental data available. The locus of Hit-1 has a high prevalence in other myxobacteria and seems independent of specific biosynthetic gene clusters, but can not be found in *E. coli*. The putative functions were derived from conserved domain architecture and are mostly of regulatory function. Using the curated database (swissprot) yielded in low confidence values for most queries, but confirmed the overall regulatory nature of the locus. The only non-regulatory gene with a significant e-value ($3.76 \cdot 10^{-42}$) is the tRNA pseudouridine synthase (K) directly downstream of Hit-1. This class of enzymes is responsible for conversion of tRNA-bound uridine to pseudouridine. This modification in the anticodon stem loop is critical for the proper function of biogenesis and the ribosome.^{31, 32} If the tRNA pseudouridine synthase interacts with Hit-1 and how it influences the corramycin mode-of-action remains elusive. Furthermore, the locus of the tRNA-pseudouridine synthase in *E. coli* differs significantly from the one in myxobacteria and instead of a Hit-1 homologue a well described ribosomal protein is located downstream. Summarized, the most likely scenario for ORF-2 mediated corramycin resistance is as follows: ORF-2 is a corramycin resistant surrogate for Hit-1. The locus of Hit-1 is conserved among myxobacteria and contains several regulatory genes as well as a tRNA pseudouridine synthase. Hit-1 and ORF-2 have conserved domain architecture for ATPase activity, that could be experimental proven for ORF-2. Structural modeling as well as in-depth sequence analysis revealed repeating evidence for involvement in a DNA-related processes like chromosome organization or DNA-repair. The process in question is most likely not yet elucidated, what makes a definite statement about ORF-2 and Hit-1 function impossible.

5.4 Bibliography

- [1] Marc Lipsitch and Matthew H. Samore. Antimicrobial use and antimicrobial resistance: a population perspective. *Emerging infectious diseases*, 8(4):347–354, 2002.
- [2] Michael A. Fischbach and Christopher T. Walsh. Antibiotics for emerging pathogens. *Science (New York, N.Y.)*, 325(5944):1089–1093, 2009.
- [3] Christopher T. Walsh and Timothy A. Wencewicz. Prospects for new antibiotics: a molecule-centered perspective. *The Journal of antibiotics*, 67(1):7–22, 2014.
- [4] Christopher Walsh and Timothy Wencewicz. *Antibiotics: Challenges, mechanisms, opportunities*. American Society for Microbiology, 2016.
- [5] World Health Organisation. World health day - combat drug resistance, 2011.
- [6] Fred C. Tenover. Mechanisms of antimicrobial resistance in bacteria. *The American journal of medicine*, 119(6 Suppl 1):S3–10; discussion S62–70, 2006.
- [7] Gerard D. Wright. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nature reviews. Microbiology*, 5(3):175–186, 2007.
- [8] W. Saurin, M. Hofnung, and E. Dassa. Getting in or out: early segregation between importers and exporters in the evolution of atp-binding cassette (abc) transporters. *Journal of molecular evolution*, 48(1):22–41, 1999.
- [9] I. B. Holland and M. A. Blight. Abc-atpases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans. *Journal of molecular biology*, 293(2):381–399, 1999.
- [10] E. Schneider and S. Hunke. Atp-binding-cassette (abc) transport systems: functional and structural aspects of the atp-hydrolyzing subunits/domains. *FEMS microbiology reviews*, 22(1):1–20, 1998.
- [11] David Auerbach, 2017, Personal Communication.
- [12] G. A. McKay, P. R. Thompson, and G. D. Wright. Broad spectrum aminoglycoside phosphotransferase type iii from enterococcus: overexpression, purification, and substrate specificity. *Biochemistry*, 33(22):6936–6944, 1994.
- [13] Tushar Shakya and Gerard D. Wright. Nucleotide selectivity of antibiotic kinases. *Antimicrobial agents and chemotherapy*, 54(5):1909–1913, 2010.
- [14] Hilary Frase, Marta Toth, and Sergei B. Vakulenko. Revisiting the nucleotide and aminoglycoside substrate specificity of the bifunctional aminoglycoside acetyltransferase(6′)-ie/aminoglycoside phosphotransferase(2′′)-ia enzyme. *The Journal of biological chemistry*, 287(52):43262–43269, 2012.
- [15] Kun Shi and Albert M. Berghuis. Structural basis for dual nucleotide selectivity of aminoglycoside 2′′-phosphotransferase iva provides insight on determinants of nucleotide specificity of aminoglycoside kinases. *The Journal of biological chemistry*, 287(16):13094–13102, 2012.

-
- [16] Kun Shi, Shane J. Caldwell, Desiree H. Fong, and Albert M. Berghuis. Prospects for circumventing aminoglycoside kinase mediated antibiotic resistance. *Frontiers in cellular and infection microbiology*, 3:22, 2013.
- [17] J. Garnier, D. J. Osguthorpe, and B. Robson. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *Journal of molecular biology*, 120(1):97–120, 1978.
- [18] Marta Toth, Hilary Frase, Nuno Tiago Antunes, Clyde A. Smith, and Sergei B. Vakulenko. Crystal structure and kinetic mechanism of aminoglycoside phosphotransferase-2^{''}-iva. *Protein science : a publication of the Protein Society*, 19(8):1565–1576, 2010.
- [19] N. Goosen and G. F. Moolenaar. Role of atp hydrolysis by uvra and uvr_b during nucleotide excision repair. *Research in microbiology*, 152(3-4):401–409, 2001.
- [20] H. W. van Veen, K. Venema, H. Bolhuis, I. Oussenko, J. Kok, B. Poolman, A. J. Driessen, and W. N. Konings. Multidrug resistance mediated by a bacterial homolog of the human multidrug transporter mdr1. *Proceedings of the National Academy of Sciences of the United States of America*, 93(20):10668–10672, 1996.
- [21] Esther Biemans-Oldehinkel, Mark K. Doeven, and Bert Poolman. Abc transporter architecture and regulatory roles of accessory domains. *FEBS letters*, 580(4):1023–1035, 2006.
- [22] A. E. Gorbalenya and E. V. Koonin. Superfamily of uvra-related ntp-binding proteins. implications for rational classification of recombination/repair systems. *Journal of molecular biology*, 213(4):583–591, 1990.
- [23] K. Furuya and C. R. Hutchinson. The drrc protein of streptomyces peucetius, a uvra-like protein, is a dna-binding protein whose gene is induced by daunorubicin. *FEMS microbiology letters*, 168(2):243–249, 1998.
- [24] N. Lomovskaya, S. K. Hong, S. U. Kim, L. Fonstein, K. Furuya, and R. C. Hutchinson. The streptomyces peucetius drrc gene encodes a uvra-like protein involved in daunorubicin resistance and production. *Journal of bacteriology*, 178(11):3238–3245, 1996.
- [25] Benedicte Michel, Hasna Boubakri, Zeynep Baharoglu, Marie LeMasson, and Roxane Lestini. Recombination proteins and rescue of arrested replication forks. *DNA repair*, 6(7):967–980, 2007.
- [26] Thomas G. Gligoris, Johanna C. Scheinost, Frank Burmann, Naomi Petela, Kok-Lung Chan, Pelin Uluocak, Frederic Beckouet, Stephan Gruber, Kim Nasmyth, and Jan Lowe. Closing the cohesin ring: structure and function of its smc3-kleisin interface. *Science (New York, N.Y.)*, 346(6212):963–967, 2014.
- [27] Jan-Michael Peters, Antonio Tedeschi, and Julia Schmitz. The cohesin complex and its roles in chromosome biology. *Genes & development*, 22(22):3089–3114, 2008.
- [28] Kim Nasmyth. Cohesin: a catenase with separate entry and exit gates? *Nature cell biology*, 13(10):1170–1177, 2011.

- [29] Stephan Gruber, Christian H. Haering, and Kim Nasmyth. Chromosomal cohesin forms a ring. *Cell*, 112(6):765–777, 2003.
- [30] Sophie Nolivos and David Sherratt. The bacterial chromosome: architecture and action of bacterial smc and smc-like complexes. *FEMS microbiology reviews*, 38(3):380–392, 2014.
- [31] Nancy S. Gutgsell, Murray P. Deutscher, and James Ofengand. The pseudouridine synthase rlud is required for normal ribosome assembly and function in escherichia coli. *RNA (New York, N.Y.)*, 11(7):1141–1152, 2005.
- [32] Yong-Heng Yang, Su-Zhen Huang, Yu-Lin Han, Hai-Yan Yuan, Chun-Sun Gu, and Yan-Hai Zhao. Base substitution mutations in uridinediphosphate-dependent glycosyltransferase 76g1 gene of stevia rebaudiana causes the low levels of rebaudioside a: mutations in ugt76g1, a key gene of steviol glycosides synthesis. *Plant physiology and biochemistry : PPB*, 80:220–225, 2014.

5.5 Experimental Procedures

Strains

Cloning was carried out in *E. coli* Dh10b or Hs916. Protein expression was conducted in *E. coli* BL21 (DE3). Genomic DNA of *Coralloccoccus coralloides* ST201330 was used as template for PCR amplification.

Resistance screening in *E. coli*

Respective genes were cloned into pET-28b using according primers. Vector pAvT was ligated from two PCR fragments originating from pET-28b and pACYC-177 (see table 5.1). Overnight cultures were diluted 1:100 and protein expression induced by addition of 0.1 mM IPTG. 150 μ L culture were suspended in a 96-flat-bottom well-plate containing increasing corramycin concentration (1 nM to 100 μ M final concentration) and incubated overnight (16h) on a shaker at 30 °C. OD₆₀₀ was quantified to assess resistance development.

Protein purification of ORF-5

Orf-5 was inserted into pET-28b with an N-terminal 6xHis tag using according primers. The construct was expressed in *E. coli* BL21 (DE3) (500 mL LB, 0.1 mM IPTG, 16 °C, 16h). The cell pellet was resuspended in lysis buffer (150 mM NaCl, 20 mM imidazole, 25 mM Tris-HCl, pH 7.5), sonicated and centrifuged. The supernatant was loaded onto a gravity flow column containing Ni-NTA loaded sepharose, washed and incubated for 60 min with lysis buffer containing additional 5 mM ATP/MgCl₂ and subsequently eluted in one step (250 mM Imidazole). 6xHis-fusion protein was passed through a Superdex 200 16/60 pg column in SEC buffer (150 mM NaCl, 25 mM Tris, pH 7.5), concentrated using a 10 kDa cutoff filter and stored at -80°C in 10 % glycerol. Protein purity was determined by SDS-PAGE. Protein concentration was determined spectrophotometrically upon determining the respective extinction coefficient from the amino acid sequence using the PROTPARAM webserver (<http://web.expasy.org/protparam/>; Gasteiger et al. 2005).

ORF-5 *in vitro* assay

10 μ M ORF-5, 1 mM corramycin and 1 mM ATP were incubated in reaction buffer (150 NaCl, 10 mM MgCl₂, 25 mM Tris-HCl, pH 7.5) were incubated in 30 uL final volume at room temperature (22 °C) for 2h. Protein was precipitated by addition of 20 uL MeOH and the supernatant subjected to LC-MS analysis. The assay was scaled to 50 mL volume and incubation time increased to over-night, lyophilised and subjected to semi-preparative HPLC.

Kinetic parameter were determined by incubating 15 μ ORF-5, 4 mM ATP and various corramycin concentrations (10, 50, 100, 200, 300 μ M) for different time-spans (0, 1, 5, 10, 30, 45, 60 min) at 30 °C in a total volume of 15 μ L. Reaction was stopped by adding tri-chloro acetic acid to a final concentration of 5 % (w/v). All experiments were carried out in triplicates. Supernatant was subjected to LC-MS analysis.

All measurements were performed on a Dionex Ultimate 3000 RSLC system using a BEH C18, 50 x 2.1 mm, 1.7 μ m dp column (Waters, Germany). Separation was achieved by a linear gradient

from (A) H₂O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 600 μ L/min and 45 °C. The gradient was initiated by a 0.5 min isocratic step at 2 % B, followed by an increase to 95 % B in 6 min to end up with a 2 min step at 95 % B before reequilibration under the initial conditions. UV spectra were recorded by a DAD in the range from 200 to 600 nm. MS data was acquired with an Amazon Speed 3D ion trap mass spectrometer (Bruker Daltonics, Germany) using the Apollo ESI source. Mass spectra were acquired in centroid mode ranging from 150–1500 m/z . Corramycin conversion was quantified using Bruker QuantAnalysis software integrating the m/z values for both, single and double protonated species.

Purification and structural determination of phosphate-corramycin

Freeze-dried 50-mL *in vitro* assay was resuspended in methanol and purified using semi-preparative HPLC. Purification conducted a Dionex HPLC system (Famos autosampler, P680 pump, TCC100 thermostat, and PDA100 detector) equipped with a Phenomenex Luna C18, 250 x 10 mm, 80 μ m dp column. Separation was achieved by a linear gradient using (A) H₂O + 0.1 % FA and (B) ACN + 0.1 % FA at a flow rate of 5 mL/min and 30 °C. The gradient started at 10 % B with a 3 min hold and increased to 70 % B in 15 min (4 % B/min). Subsequent B was increased to 95 % in 1 min with a 1 min hold. UV data was acquired at 200–600 nm. The sample was injected by μ L-pick-up technology with a water/methanol (50:50 v/v) mixture as supporting solvent. A maximum of 50 μ l (6 μ g/ μ L) of the sample was injected before manual fraction collection. Purity of fractions was tested using the above described LC-MS method.

NMR spectra were recorded on a 700 MHz Avance III (Ascend) spectrometer by Bruker BioSpin GmbH, equipped with a 5 mm TXI cryoprobe, at 298 K. Chemical shift values of ¹H- and ¹³C-NMR spectra are reported in ppm relative to the residual solvent signal given as an internal standard. ¹³C-signals were assigned via 2D-CH and CCH correlations (HSQC and HMBC) (see Table S 5.5).

Protein purification of SUMO-ORF-2

Two constructs (N-terminal 6xHis and 6xHis-SUMO) for *orf-2* were designed using the same primers. The constructs were expressed in *E. coli* BL21 (DE3) (500 mL LB, 0.1 mM IPTG, °C, 16h). The cell pellet was resuspended in lysis buffer (150 mM NaCl, 0.1 % Triton X-100 (v/v), 20 mM imidazole, 20 mM Bis-Tris, pH 6.8), sonicated and centrifuged. The supernatant was loaded onto a gravity flow column containing Ni-NTA loaded sepharose, washed with lysis buffer and subsequently eluted in one step (250 mM Imidazole). 6xHis-fusion protein did elute as precipitate. SUMO-ORF-2 was recovered as soluble protein and passed through a Superdex 200 16/60 pg column in SEC buffer (150 mM NaCl, 25 mM Bis-Tris, pH 6.8), concentrated using a 10 kDa cutoff filter and stored at -80°C in 10 % glycerol. Protein purity was determined by SDS-PAGE. Protein concentration was determined spectrophotometrically as described above.

Generation of GFP-ORF-2 construct

N-terminal EGFP-ORF-2 was generated using the pET-28b-ORF-2 construct. EGFP was amplified from pEGFP-C1 and cloned into pET-28b-ORF-2 using the according primers. Test expression in *E. coli* BL21 (DE3) (100 mL LB, 0.1 mM IPTG, °C, 16h) and subsequent SDS-PAGE

analysis did indicated the presence of EGFP-ORF-2. The construct was sent to the lab of Theresa Stradal.

Pull-down assay SUMO-ORF-2

150 μ g purified SUMO-ORF-2 were incubated for 30 min with 50 μ L pre-equilibrated (150 mM NaCl, 20 mM imidazol, 25 mM Bis-Tris, pH 6.8) magnetic Ni-NTA beads (Thermo Scientific) at 4 °C. Subsequently different ammounts of E. coli lysate were added and the mixture incubated for 2.5 h at 4 °C. The lysate was generated from E. coli BL21 (DE3) pET-28b with the procedure described for SUMO-ORF-2. After incubation the beads were washed three times with 400 μ L buffer before the proteins were eluted with 50 μ L elution buffer (250 mM imidazole). 10 μ L of the eluted fraction were analyzed by SDS-PAGE.

Malachite-green assay of SUMO-ORF-2

The ATPase activity of SUMO-ORF-2 was assayed as described previously (McQuade et al., 2009). Briefly, 10 μ M SUMO-ORF-2 was incubated with 1 mM ATP and optional with either 0.75 mM corramycin or 40 μ g of μ M *Coralloccoccus coralloides* STST201330 gDNA (sheared by 30 s vortexing) in reaction buffer (150 mM NaCl, 25 mM Bis-Tris, pH 6.8) containing inorganic phosphatase at 37 C. The reaction was stopped after different time-points up to 300 min by adding 20 μ L malachite green solution. The absorption at 620 nm was measured after 10 min. Fold increase of absorption was calculated based on a negative control without enzyme.

Table 5.3: Sequence of used oligonucleotides.

Primer	Sequence	Vector	Tag	Origin
pAvT1-177-FP	AAAAAA CCCGGG CATCAGAAGGGCACTGGTGC	pAvT1	-	pACYC-177
pAvT1-177-RP	AAAAAAC ACACGT GTCGCTCACTGACTCGCT		-	pET-28b
pAvT1-28-FP	AAAAAA ACACGT TGTGCGCCAATCCGGATA		-	pET-28b
pAvT1-28-RP	AAAAAA CCCGGG GAAACGTTTGGTGGCG		-	pET-28b
ORF-1-nt-FP	AAAAAA ACATG TTGGTCCTGGTGCTG	pET-28b	-	<i>C. c.</i> STST201330
ORF-1-RP	AAAAAA GGATC CTCACTGGCACCG	pET-28b	6xHis(N)	<i>C. c.</i> STST201330
ADC-FP	AAAAAA CATATG AGACGCATTGTCTTCAAGTC	pET-28b	-	<i>E. coli</i> BL21 (DE3)
ADC-RP	AAAAAA GAATT CTCACGCATTCCGGC		-	<i>E. coli</i> BL21 (DE3)
RecF-mRNA-FP	AAAAAATCTAGATGTCCCTCACCCGCTT		-	<i>E. coli</i> BL21 (DE3)
RecF-FP	AAAAAA CATATG TCCCTCACCCGCTTGT		-	<i>E. coli</i> BL21 (DE3)
RecF-RP	AAAAAA GAATT CCTTAATCCGTTATTTTACCCT	pET-28b	-	<i>C. c.</i> STST201330
ORF-2-nt-FP	AAAAAA CCATG GCATGTTTCGTTCAAGAGC	pET-28b/pAvT1	-	<i>C. c.</i> STST201330
ORF-2-nt-RP	AAAAAAA GAATT CCTACTCCTCCAAAGCGCCGTC		-	<i>C. c.</i> STST201330
ORF-5-nt-FP	AAAAAA ACATG TTAGTGGGTAACAACCTCGCGTG		-	<i>C. c.</i> STST201330
ORF-5-nt-RP	AAAAAAA GAATT CTCAAGGACCCGGCG		-	<i>C. c.</i> STST201330
Hit-1-FP	AAAAAA ACATG TTAATGCCCTCACGC	pET-28b	-	<i>C. c.</i> MCy10984
Hit-1-RP	AAAAAA GAATT CTCATTCGTCGTTCCCATTC		-	<i>C. c.</i> MCy10984
ORF-2-FP	AAAAAA ACATATG TTCGTTCAAGAGCTCCGTG		6xHis(N)	<i>C. c.</i> STST201330
ORF-2-RP	AAAAAAA GAATT CCTACTCCTCCAAAGCGCCGTC		SUMO(N)	<i>C. c.</i> STST201330
ORF-5-FP	AAAAAA ACATATG GGTAACAACCTCGCGTGC	pET-28b	6xHis(N)	<i>C. c.</i> STST201330
ORF-5-RP	AAAAAAA GAATT CTCAAGGACCCGGCG		6xHis(N)	<i>C. c.</i> STST201330
EGFP-ORF2-RP	AAAAAATGGTGAGCAAGGGCGA		6xHis(N)	pEGFP-C2
EGFP-ORF2-RP	AAAAAA CATATG TCTAGATCCGGTGG		6xHis(N)	pEGFP-C2

5.6 Supporting Information

Table 5.4: BLAST analysis of the genetic locus surrounding Hit-1 (here marked as J, see Figure 5.6 for reference) using the non-redundant (nr) and the curated swissprot database.

Gene	Putative function	Source	Ident / %	e-value	Accession
non-redundant (nr) database					
A	Transcriptional regulator	<i>C. coralloides</i>	99	$3.03 \cdot 10^{79}$	WP_014400725.1
B	DN A binding protein	<i>C. coralloides</i>	89	$1.41 \cdot 10^{58}$	WP_014400724.1
C	Diguanylate cyclase	<i>C. coralloides</i>	96	0.0	WP_014400723.1
D	Response regulator	<i>C. coralloides</i>	100	$4.46 \cdot 10^{83}$	WP_014400722.1
E	Response regulator Histidine kinase	<i>C. coralloides</i>	95	0.0	WP_014400721.1
F	Response regulator GGDEF domain	<i>C. coralloides</i>	95	0.0	WP_014400720.1
G	Hypothetical protein	<i>C. coralloides</i>	80	0.0	WP_014400719.1
H	Hypothetical protein	<i>C. coralloides</i>	92	$3.19 \cdot 10^{167}$	WP_014400718.1
I	Hypothetical protein	<i>C. coralloides</i>	54	$5.36 \cdot 10^{173}$	WP_052313000.1
J	RecF	<i>C. coralloides</i>	95	0.0	WP_014400717.1
K	tRNA pseudouridine synthase B	<i>C. coralloides</i>	96	0.0	WP_014400716.1
L	N-acetyltransferase GCN5 CDS	<i>C. coralloides</i>	96	$3.81 \cdot 10^{121}$	WP_014400715.1
M	Hypothetical protein	<i>C. coralloides</i>	90	0.0	WP_014400714.1
N	Transketolase	<i>C. coralloides</i>	98	0.0	WP_043322454.1
O	Arginase	<i>C. coralloides</i>	59	$2.58 \cdot 10^{93}$	WP_036113570.1
P	Hypothetical protein	<i>C. coralloides</i>	27	$2.23 \cdot 10^{10}$	KYG03003.1
Q	Hypothetical protein	<i>C. coralloides</i>	27	$2.47 \cdot 10^{10}$	WP_012240504.1
curated (swissprot) database					
A	HTH-type transcriptional regulator	<i>A. fulgidus</i>	41	4.3	Y1787_ARCFU
B	-				
C	Sensor histidine kinase ResE	<i>B. subtilis</i>	34	$5.83 \cdot 10^{26}$	RESE_BACSU
D	Transcriptional regulatory protein	<i>B. subtilis</i>	45	$5.40 \cdot 10^{20}$	PHOP_BACSU
E	Signal histidine kinase J	<i>D. discoideum</i>	29	$1.89 \cdot 10^{39}$	DHKJ.DICDI
F	Response regulator RcaC	<i>M. diplosiphon</i>	32	$1.04 \cdot 10^{18}$	RCAC_MICDP
G	Type IV pilus biogenesis factor PilY1	<i>P. aeruginosa</i>	37	5.0	PILY1.PSEAW
H	Alpha-tubulin N-acetyltransferase 2	<i>C. elegans</i>	50	5.4	ATAT2.CAEEL
I	-				
J	DNA replication and repair RecF	<i>Marinomonas sp.</i>	52	0.018	RECF_MARMS
K	tRNA pseudouridine synthase	<i>B. campestris</i>	41	$3.76 \cdot 10^{42}$	TRUB_XANCP
L	Kynureninase 2	<i>C. globosum</i>	28	2.3	KYNU2.CHAGB
M	Pre-mRNA-splicing factor SYF2	<i>A. fulgidus</i>	42	7.3	Y1225_ARCFU
N	Transketolase	<i>G. stearotherm.</i>	51	0.0	TKT_GEOSE
O	Gamma-glutamyl phosphate reductase	<i>B. velezensis</i>	37	0.35	PROA_BACVZ
P	-				
Q	RNA polymerase subunit beta	<i>C. demersum</i>	33	3.9	RPOC2_CERDE

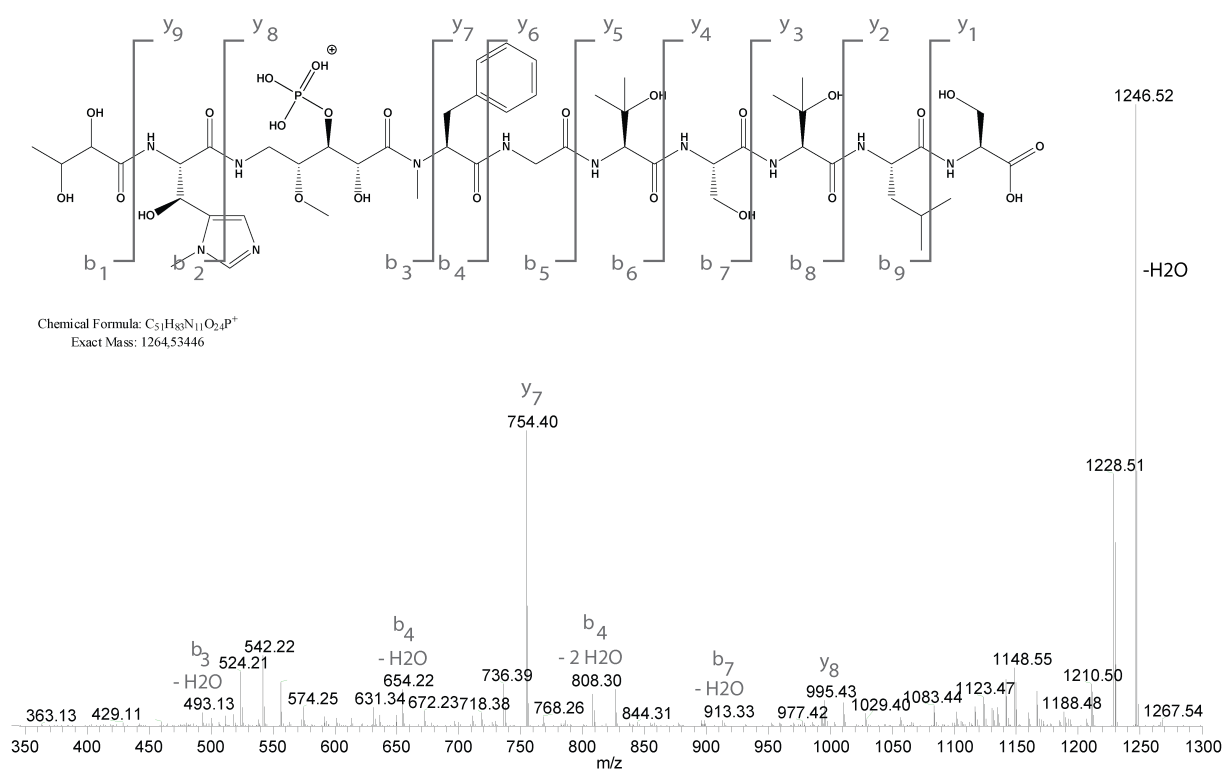


Figure 5.7: ITMS² spectra acquired for phosphate corramycin (m/z 1264.6). Fragments were assigned according to the Biemann modification of the Roepstorff nomenclature.

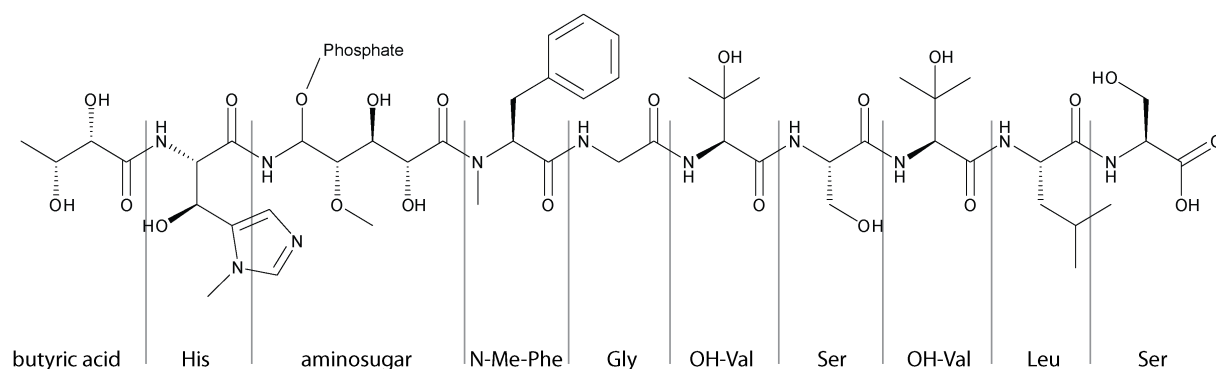


Figure 5.8: NMR position labeling for phosphate-corramycin. Numbering in NMR table always starts at carbonyl carbon.

Table 5.5: NMR data of phosphate-corramycin in D₂O. For structure reference see Figure S 5.8

Position	Group	δ_C	δ_H
Ser			
1	C	175.7	
2	CH	57.0	4.29
3	CH ₂	61.9	3.85
Leu			
1	C	173.7	
2	CH	52.2	4.50
3	CH ₂	39.4	1.69
4	CH	24.1	1.68
5	CH ₃	22.0	0.95
6	CH ₃	20.4	0.89
β -OH Val			
1	C	171.3	
2	CH	60.6	4.44
3	C	71.6	
4	CH ₃	25.8	1.32
5	CH ₃	25.0	1.32
Ser			
1	C	171.0	
2	CH	55.3	4.61
3	CH ₂	60.8	3.88
β -OH Val			
1	C	171.3	
2	CH	60.52	4.56
3	C	71.8	
4	CH ₃	25.8	1.32
5	CH ₃	25.0	1.29
Gly			
1	C	170.71	
2	CH ₂	42.4	4.10
			4.01
N-Me-Phe			
1	C	172.1	
2	CH	59.6	5.37
3	CH ₂	32.8	3.09
			3.45
4	C	136.8	
5	CH	128.6	7.39
6	CH	128.9	7.34
7	CH	126.7	7.32
8	CH ₃	32.1	3.04
aminosugar			
1	C	173.7	
2	CH	68.2	4.58
3	CH	70.2	4.08
4	CH	78.4	3.58
5	CH ₂	38.16	3.58
6	CH ₃	56.9	3.30
β -OH- δ -Me-His			
1	C	171.05	
2	CH	55.9	4.96
3	CH	63.9	5.16
4	C	130.4	
5	CH	125.0	7.14
6	CH	139.1	7.90
7	CH ₃	32.1	3.81
butyric acid			
1	C	173.99	
2	CH	74.7	3.95
3	CH	67.9	3.44
4	CH ₃	17.7	1.11

Chapter 6

Discussion

6.1 Summary

Bacterial secondary metabolite pathways and their role in pharmaceutical research is the central topic of this thesis. Pervading all chapters, *in vitro* reconstitution of enzyme activity and heterologous expression were the key methods to address various tasks such as the elucidation of biosynthesis pathways, inhibitor screening, generation of derivatives or identification of resistance mechanisms.

Chapter 2 and 3 describe the elucidation of the biosynthesis of the two pyrrolo[4,2]benzodiazepines (PBD) tomaymycin and tilivalline. The tomaymycin NRPS TomA and TomB were reconstituted *in vitro* and intact protein mass spectrometry was employed to visualize each biochemical step on the protein itself. Additional biochemical efforts as well as the structural analysis of the associated methyltransferase TomG contributed to a clearer picture of the tailoring sequence of the anthranilic acid moiety in the tomaymycin biosynthesis. The broad substrate tolerance of TomA enabled the production of several tomaymycin derivatives by feeding the respective precursors, extending the tool kit for the generation of PBD derivatives, an important compound class in the endeavor for novel cancer treatments. Chapter 3 revealed the biosynthesis of the PBD tilivalline, the causative toxin in *K. oxytoca* related antibiotic associated hemorrhagic colitis (AAHC). Several tilivalline producing platforms were established, including heterologous expression and *in vitro* reconstitution of the biosynthetic pathway, enabling the elucidation of a non-enzymatic indole incorporation. The gained knowledge of the tilivalline biosynthesis enabled the identification of inhibitors, capable of impairing the toxin production in *K. oxytoca* liquid culture. Thus, this work contributed to a potential anti-virulence drug in the treatment of AAHC by providing proof-of-concept as well as heterologous and *in vitro* platforms that can be used for further inhibitor screening efforts.

Chapter 4 and 5 deal with the biosynthesis and self-resistance mechanism of corramycin, a peptide antibiotic from *Corallococcus coralloides*. *In silico* analysis of two producer genomes together with feeding studies established a consistent picture of the underlying biosynthesis. Biochemical characterization of the first module furthermore indicate the presence of a predrug mechanism comprising an acyl-corramycin intermediate. Analysis of the biosynthetic gene cluster and subsequent biochemical experiments led to the identification of two independent resistance mechanisms. ORF-5 is a kinase that inactivates corramycin by the transfer of a phosphate group. ORF-2 is a protein of unknown function, that probably acts as corramycin insensitive surrogate of the molecular target of the antibiotic. Elucidation of self-resistance mechanism is highly valuable for the development of antibiotics in general, as it provides evidence for possible resistance pathways in pathogens as well as insights into the mode-of-action and thus the molecular target.

Summarized, this work contributes to the field of anti-infectives research in two ways: Corramycin is a potent antibiotic currently investigated by Sanofi. The elucidation of its biosynthesis and the identification of self-resistance mechanisms in its producer will provide valuable information for the further development process as an antibiotic. The presented work on the biosynthesis of the toxin tilivalline paved the way for an anti-virulence strategy in *K. oxytoca* related AAHC.

6.2 Obstacles in antibiotic development

Introducing a new antibiotic to the market is the finishing line of a process that takes up to 15 years and costs a substantial amount of money. Two phases in the development have the highest associated risks of failure and thus are main bottlenecks: The first obstacle is located in the discovery process that tries to identify novel compounds that can enter the development pipeline. Only one-third of positive signals from the initial screening such as high-throughput screenings (HTS) can be validated and are advanced as hits. Again, only one-third of the identified hits is further advanced as lead and subsequently developed towards clinical testing. The second bottleneck is the starting phase of clinical trials, where only 25 % of antibacterials succeed, after around eight years of costly development.¹

To overcome the high failure rates in the initial discovery process, two prerequisites are important: Firstly, the screening system used has to be validated and be predictive of important aspects such as cell-permeability and toxicity, to avoid constipation of the pipeline with ultimately useless compounds. Secondly, the source of compounds to be screened has to be sufficient in both, quality and quantity.² One of the research focuses of the Helmholtz Institute for Pharmaceutical Research Saarland (HIPS) is to provide a potent library of compounds derived from myxobacterial extracts, that are available for screening. The opening of novel and untapped compound sources is mandatory for the success in antibiotic discovery and development, since previous strategies failed and resulted in the starvation of the development pipeline we see today.^{2, 3}

Historically, whole-cell screening approaches of extracts derived from soil dwelling actinomycetes was the main pillar of the "golden age of antibiotics",⁴ resulting in the discovery of numerous antibacterials and the development of all important antibiotic classes like macrolides, β -lactams, aminoglycosides, glycopeptides, lipopeptides, ansamycins or tetracyclines.^{5, 6} Overmining and the increasing problem of rediscovery of known compounds led to a decreasing output of this approach: While between 1960 and 1980 75 new classes of antibacterial compounds were discovered from actinomycetes, only 20 were found between 1980 and 2000.⁷ Massive effort in the derivatization of known antibiotics could maintain the supply for some time, but eventually the pipeline ran out of novel structures to work on. The upcoming rise of the genomic age and the lure of combinatorial chemistry prompted the industry to switch from the well validated and successful platform of whole-cell screens of microbial extracts to a target-based HTS approach using huge compound libraries generated by combinatorial chemistry. This strategic decision proved disastrous for antibiotic development and not a single drug emerged from this pipeline and entered the market, albeit the huge financial investment by big pharma companies like GlaxoSmithKline.¹ This failure has several reasons: Targets identified by genomics regularly turned out to be not drugable or too prone to develop resistance.⁸ For example, a potent inhibitor of the bacterial deformylase showed good results in animal testings, but failed because of too high resistance development frequency.⁹ In contrast, validated target from the well established antibiotic classes like the ribosome or the cell wall biosynthesis are not suitable for HTS and were therefore neglected, although their complex underlying structure causes a lower probability of resistance development than single enzyme targets. One of the largest drawbacks

was the missing selection for cell permeability compared to whole-cell screening. Thus, most of the hits from the HTS had to be discontinued because they did not reach their respective target *in vivo*.² Furthermore, the combinatorial chemistry derived libraries used were not designed to enter the prokaryotic cell, but to be orally available. Guidelines like Lipinski's rule of five proved not to be beneficial for antibiotic development, but were nevertheless the foundation of the used combinatorial chemistry.^{3, 10} Summarized, the application of genomics driven HTS and combinatorial chemistry derived libraries for antibiotic discovery were overoptimistic and together with the abandoning of the successful strategy of whole-cell screenings of microbial extracts led to the observed gap in the development pipeline.^{2, 3, 11} While taking the step back to whole-cell screening is easy to do, finding novel sources to generate compounds that can enter the pipeline is more challenging. Nevertheless, several promising approaches are within reach.

6.3 Microbial sources of compounds in antibiotic drug discovery

Historically, actinomycetes were in the center of a successful antibiotic drug discovery for decades. Consequently, reviving this source for novel efforts is logical, although severe problems arise. Totally, over 10 million actinomycetes have been cultured and over 2000 antibacterial compounds were identified.^{5, 7, 12, 13} It is estimated that 10 million more strains have to be screened to find one antibiotic that can enter the market.² Obviously, following up on old routines is not enough for this strategy to succeed. Nevertheless, actinomycetes are far from being totally grazed for drug discovery. Although the number of known antibiotics from this genus rises steadily, the respective frequency drops exponentially: 200 known antibiotics from actinomycetes have a frequency of one in $4 \cdot 10^{-7}$ strains, 800 antibiotics occur in one of $2 \cdot 10^{-7}$ strains and 1000 antibiotics are found in only one of $1 \cdot 10^{-7}$ strains.⁵ This proves that still a lot of potent antibiotics can be found in actinomycetes, but they are buried under a thick layer of known compounds that hamper discovery by false positive hits. A powerful dereplication strategy using LC-HRMS or reporter strains that are resistant to known antibiotics together with a massive screening effort is therefore necessary to revive actinomycetes as source for antibiotics. This approach is thus only feasible in the industrial context and not for academic institutions like the HIPS that focuses on myxobacteria instead.

6.3.1 Myxobacteria as industrial source for antibiotics

To switch from actinomycetes to another order of bacteria displays a promising possibility to generate novel compounds. Myxobacteria are long known to be prolific producers of novel natural products¹⁴ and furthermore harbor unusual large genomes with numerous biosynthetic gene clusters, most of them not yet linked to the corresponding small molecule.¹¹ The HIPS method of using long-standing techniques like whole-cell screenings in combination with myxobacterial extracts and bioactivity guided purification reflects the traditional way of antibiotic research. Corramycin, an antibacterial compound produced by *C. coralloides*, was likewise identified by Sanofi and is currently developed as a lead structure, proving that this old-fashioned approach is still relevant today. Nevertheless, fermentation conditions as well as the corresponding time-frame is far less suitable for HTS approaches and automation, what is preferred by industry. While single spores of actinomycetes can be readily fermented in small scale, myxobacteria are

not always cultivable on agar plates and even more work is needed to facilitate fermentation in liquid media.^{15, 16} A generalized approach to identify, culture and screen myxobacteria seems therefore unfeasible. Nevertheless, many more novel compound classes and potential antibiotics are certainly hidden in myxobacteria, although different methodologies than for industrial scale actinomycetes screenings are necessary to harvest them. One strategy involves the careful selection of strains to exploit the biodiversity. The dogma that phylogenetic distance correlates with chemical diversity can be applied by isolating myxobacteria from different habitats and subsequently provide rigid classification, so that the available work force is not wasted on too similar strains. This microbiology based approach also includes the screening of different media and culture conditions, to eventually trigger production of novel natural products.¹¹ Especially the media composition has a strong influence on the produced metabolome, what also holds true for the corramycin producing *C. coralloides*. Optimization of production media increased the corramycin titer to around 7 mg/mL, more than one magnitude higher than in standard media used at HIPS. Furthermore, production in *C. coralloides* ST201330, isolated by Sanofi, proved to be unstable and could not be reproduced with certainty at HIPS, making experiments like feeding of isotope labeled precursors difficult. In contrast, *C. coralloides* MCy10984 showed reliable production of corramycin. Additionally, several derivatives of corramycin could be identified, that do not occur in the extracts of *C. coralloides* ST201330. This highlights again the importance of media and fermentation conditions in general as well as the availability of alternative producers.

Keeping the large genome with its untapped potential for novel natural products in mind, this becomes particular important. *C. coralloides* MCy10984 was fermented under standard conditions and the extract subsequently screened for bioactivity. Corramycin was identified by strong activity against *E. coli* Δ TolC and subsequent analysis by mass spectrometry, confirming the previous results from Sanofi. LC-MS based dereplication processing in our in-house database revealed the presence of 10 additional known natural product classes. Nevertheless, genome sequencing and subsequent analysis on the antismash platform¹⁷ revealed the presence of 52 putative biosynthetic gene clusters (Figure 6.1). Only two terpene linked clusters have sufficient significant sequence identity (> 90 %) to link them to a known natural product. Interestingly, 27 biosynthetic clusters of the NRPS and PKS type were identified, while only corramycin was linked to its cluster by this work. Taking into account that corramycin is a novel compound as well, *C. coralloides* MCy10984 alone harbors the genetic potential to produce almost 30 novel peptide- and/or polyketide natural products.

6.3.2 Activation of silent gene clusters

One strain alone harbors a much larger potential for natural product discovery than what is actually accessed in a fixed fermentation and screening setup. Given that natural products play a specific part in the habitat of their producer, it appears logical that the corresponding biosynthesis is induced by specific environmental factors and is not turned on unconditional. Several of these inducers have been identified and aided the isolation of novel compounds, for example DMSO, rare metals or co-cultivation with other microbes.^{18–20}

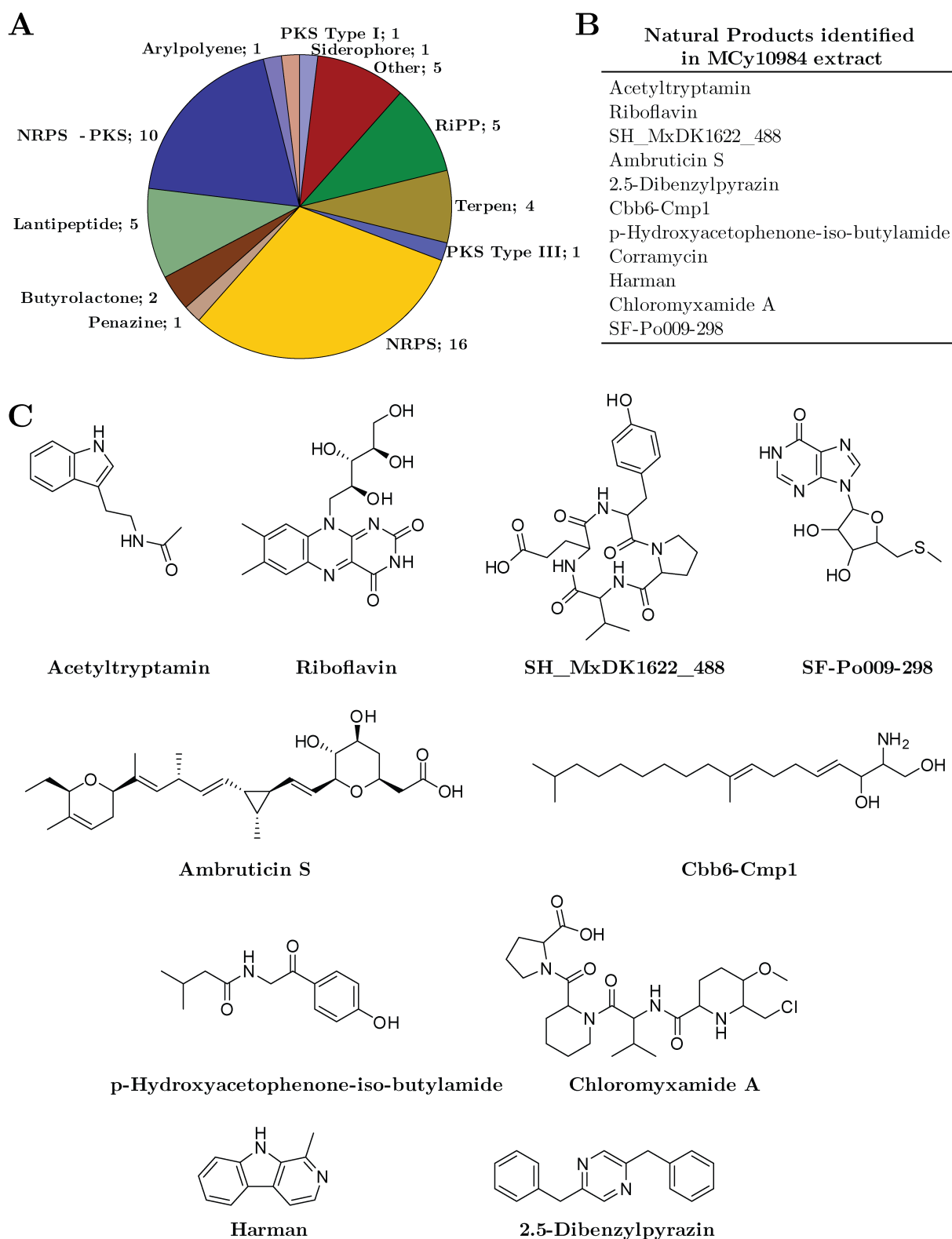


Figure 6.1: **A** Classification of the 52 biosynthetic gene clusters identified by Antismash in the genome of *C. coralloides* MCy10984. Only two terpene clusters were assigned to known biosynthetic pathways based on sequence identity (> 90 %). **B** Result of the in-house dereplication database mxbase reveals the presence of 11 known natural product classes, including corramycin. **C** Structures of the known natural products identified in the LC-MS analysis of the *C. coralloides* MCy10984 extract.

Co-cultivation with selected myxobacterial strains has also been conducted in course of this thesis, albeit without success. Two myxobacterial strains, MCy10585 and MCy10649, were selected because of their great number of unknown biosynthetic gene clusters (> 30) but missing antibacterial activity against *E. coli* wild type under standard screening conditions. Repeated co-cultivation on agar plate together with *E. coli* was conducted to trigger production of a compound with Gram-negative activity. Screening of extracts originating from different rounds of incubation did not show any emerging activity against the used *E. coli*. Nevertheless, this does neither mean that no compound with Gram-negative activity is encoded in the genome nor that no silent gene cluster was induced, since the screening system containing only one indicator strain constrained the possible hits significantly. Summarized, activation of silent gene clusters by randomly changing external factors is laborious and probably not efficient, especially in the genomic age, where the underlying sequence data of potential biosynthetic gene clusters is already available. Direct access to the silent gene clusters by means of molecular biology is thus a far more promising foundation for the strategic exploitation of untapped natural product potential, not only for myxobacteria but general.

6.3.3 Heterologous expression

Direct access to biosynthetic gene clusters identified by genome mining using heterologous expression in a host organism is a promising strategy, because it bypasses one of the main source of problems in natural product discovery from bacteria: the producer strain. One key problem is the cultivability of exotic strains, an important source for novel compounds, as described above. It is estimated that only a fraction of the existing strains can be cultivated in a lab environment, and for some species laborious and time-consuming techniques have to be employed. If the strain is cultivable and even produces the desired compound, low production yield and challenging fermentation conditions often render it useless for industrial application. Molecular biology tools to manipulate the producer strain are also often not available, preventing for example the introduction of artificial promoters to enhance yield. Heterologous expression of the desired biosynthetic gene cluster in a host strain with well established fermentation conditions and working protocols for genetic manipulation is thus a viable alternative.²¹

The great strength in heterologous expression is the independence from the source organism. Even genomic DNA isolated directly from soil samples has been used to successfully restore heterologous production of novel compounds, without even isolating the corresponding strain.²² Also several myxobacterial biosynthetic gene clusters have been expressed using different host strains. In this study, heterologous expression was mainly intended as a tool for studies of the biosynthesis of the corresponding natural product and not for the discovery of novel entities. Both *C. coralloides* strains that produce corramycin are not susceptible for genetic manipulation and thus heterologous expression of the biosynthetic gene cluster was aspired. The corramycin cluster was split in two parts to facilitate the TAR-based assembly from PCR-derived fragments. TAR cloning is based on homologous recombination taking place in *S. cerevisiae* upon transformation with suitable DNA fragments. The operon harboring the corramycin assembly line as well as several tailoring enzymes was integrated into a vector backbone designed for integration

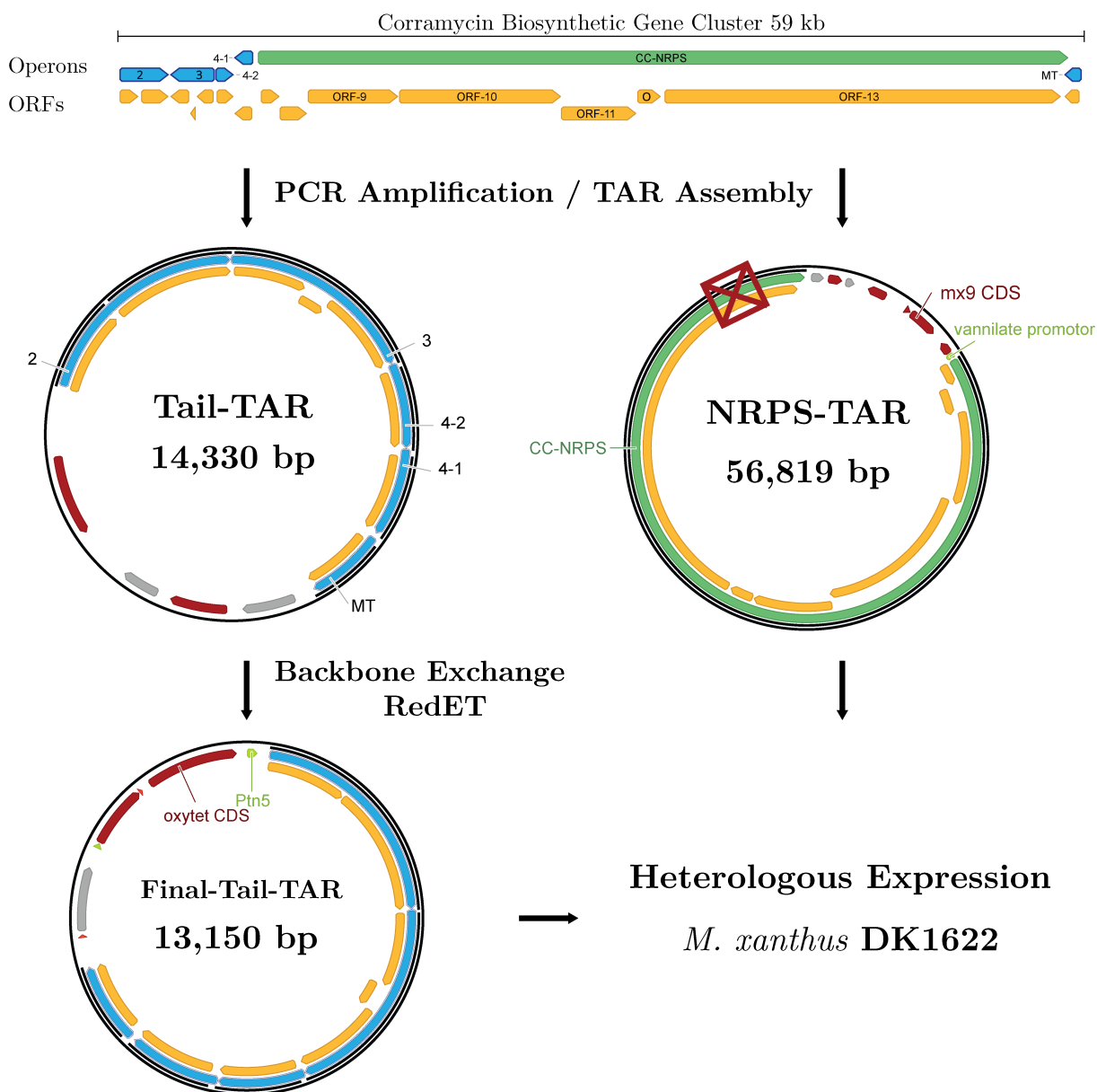


Figure 6.2: Strategy of heterologous expression of the corramycin biosynthetic gene clusters. The assembly line operon is inserted into a plasmid already suitable for integration in the *M. xanthus* DK1622 mx9 phage-attachement site. During assembly module 9 was lost (red rectangle). Remaining genes are rearranged into one operon and inserted and assembled via TAR. Plasmid backbone is subsequently exchanged using RedET to allow integration by homologous recombination using a previously inserted oxytetracycline resistance cassette. *M. xanthus* transformed with both constructs did not produce any corramycin related compounds.

into the *M. xanthus* genome using the mx9 phage-attachement site. The remaining resistance, transport and tailoring genes were rearranged and assembled using a backbone for *M. xanthus* integration via homologous recombination in a previously inserted tetracycline resistance cassette. While Illumina based sequencing proved the tailoring construct to be correct, no complete NRPS-harboring construct could be obtained. Sequence homologies due to repeating module specificities in the last third of the assembly line probably lead to homologous recombination inside the cluster itself, hampering correct assembly: Module 9 was missing in the final construct. Interestingly, the deletion preserved the reading frame and the module was excised completely, creating a putatively functional assembly line. It is probably no coincidence that Illumina based sequencing of the complete *C. coralloides* MCy10984 genome also failed at this exact position due to too short reads and resequencing of PCR amplified fragments had to be employed to ultimately resolve the sequence information. Repeating sequence information is a common problem for NRPS and especially PKS systems due to their modular structure, often responsible for the incorporation of the same or at least highly similar building blocks. Assembly strategies based on homologous recombination therefore also are associated with a risk of mis-assembly due to homologous regions outside the intended ones. The screening of all obtained clones for the corramycin assembly line construct did not yield a correct clone and although sequencing revealed that no frameshift was introduced and the missing module was cut out completely, no production of a corresponding shortened corramycin could be observed.

Failure of heterologous expression can also be host-related. The chosen heterologous host can be unable to facilitate the translation of the large genes, relevant for long clusters like corramycin (55 kb). Furthermore, relevant enzymes for post-translational modification like PPant transferases or exporters are missing, preventing the successful production of the compound. *M. xanthus* DK1622 has been previously used to heterologously produce several myxobacterial compounds,²³ including myxothiazol,²⁴ tubulysins,²⁵ myxochromide S,²⁶ disorazol²⁷ and epothilone.²⁶ Therefore it was chosen as the host strain for corramycin. Tilivalline, naturally produced by *K. oxytoca*, was heterologously produced in *E. coli* BL21, the well studied workhorse of bacterial research altogether and member of the same family of enterobacteriaceae as *K. oxytoca*. Beside the host, sequence-related problems may arise when employing heterologous expression. For corramycin, the deletion of module 9 was complete and under preservation of the reading frame, enabling successful production in theory. Nevertheless, the missing module is likely to affect the downstream processing and may prevent product formation of a shortened corramycin altogether. Furthermore, the tailoring enzymes harbored on a separate plasmid have been heavily rearranged to build one artificial operon. The design is solely based on *in silico* analysis and therefore might be incorrect. Additionally, promoters were introduced to gain control over the expression of the biosynthetic gene cluster. Since the used tn5 and vanilic acid inducible promoters have been successfully used before in *M. xanthus* DK1622, they are probably not responsible for the failure. Nevertheless, several potential issues with the corramycin constructs remain that require intensive work to be identified and addressed.

The problems encountered during the assembly of the corramycin biosynthetic gene cluster highlight the general pitfalls of heterologous expression. Although several different methods are

available to assemble the respective genetic information, either directly from genetic DNA or from DNA fragments, the whole process remains laborious and produces a high number of false constructs, which have to be identified reliably.²⁸ The overall rate of success correlates furthermore with the size of the biosynthetic gene cluster. The assembly and heterologous expression of the tilivalline cluster with a size of under 20 kb was accomplished rapidly and without major hurdles, enabling in-depth studies on the biosynthesis. The intended cloning of the 55 kb corramycin cluster ultimately failed, although considerate efforts were made. To make heterologous expression of large assembly lines more efficient, other strategies have to be employed. The use of synthetic DNA fragments to assemble the cluster is a feasible alternative to genomic DNA as the primary source of sequence information. Here, the sequence can be optimized and rearranged without limitations and thus enabling a faster assembly, for example by removing repeating DNA sequence that hampered TAR assembly of the corramycin cluster. Furthermore, GC-content, promoters, restriction sites and codon-usage can be optimized according to the needs of the assembly method and the intended host. The approach of using the genomic DNA of the producer as the primary sequence source for heterologous production of corramycin has therefore been terminated. Instead, optimized synthetic DNA fragments will be employed, what is however not part of the herein presented thesis.

Summarized, heterologous expression of biosynthetic pathways gains more and more importance for both, biosynthetic studies and the discovery of novel compounds linked to silent gene clusters. Several different methods for DNA assembly or direct cloning as well as affordable possibilities for synthetic DNA exist today, greatly expanding the tool kit. At this time, every biosynthetic gene cluster demands intensive work on all stages of design, assembly and heterologous expression and is therefore not yet suitable to replace the HTS efforts of microbial extracts in the industrial context. Nevertheless, the development of a pipeline that automates the steps from identification to expression of a biosynthetic gene cluster certainly has the potential to change the face of natural product discovery.

6.4 Anti-virulence drugs

6.4.1 Development of anti-virulence drugs

The antagonist of antibiotic development is the inevitable emergence of resistance, already a major threat to public health after little short of a century of systematic antibiotic use. The underlying reasons for the emergence and spread of resistant pathogens are diverse and some, e.g. bad stewardship of existing antibiotics, lack of hygiene in hospitals or global traffic, can realistically be addressed. Nevertheless, the fundamental problems are outside the realm of human control. Resistance mechanisms against antibacterial compounds existed long before the introduction of antibiotic regimes and are an intrinsic part of nature.^{29, 30} The biosynthetic gene cluster of corramycin already contains two independent resistance mechanisms to protect the producer from its own product. One of the discovered strategies is the chemical inactivation of the compound by a kinase, a mechanism well known for glycoside antibiotics like kanamycin. Horizontal gene transfer is a possible pathway for the cross-species propagation of the underlying gene,^{31, 32} setting the stage for the spread of corramycin resistance, before even one patient has

been treated.^{31, 32} Apart from existing mechanisms the high number of bacterial life on earth generates a frequency of random mutations that builds up a huge potential for a successful counterstrike by evolution to whatever novel antibiotic is developed. The very definition of an antibiotic, to kill bacterial life, exhibits the strongest of all evolutionary pressures and unlocks this intrinsic potential, making the emergence of resistance to a respective drug a matter of time: Recent reports of pathogens that exhibit significant resistance to the last resort antibiotics colistin^{33, 34} and vancomycin^{35, 36} are the proof.

The question arises what countermeasures are necessary to prevent a public health relapse in the pre-antibiotic age. Obviously, the constant development of novel antibiotics and vaccines is vital. Additionally, alternative strategies for the fight against infectious diseases like anti-virulence drugs merit a closer examination, as they can aid or even substitute antibiotic regimes, at least in some cases.³⁷ A non-killing antibacterial compound has to target a non-vital cell function, that is nevertheless important for pathogenity. The different parts of the pathogen-host interaction are an appropriate source of novel targets, distinct from the classical targets used in screening efforts for novel antibiotics. During infection pathogens have to sense and adapt to the host environment and have to overcome physical barriers, secretion mechanisms like mucus and the active immune response. Furthermore, the host microbiota has to be either exploited or displaced to allow for an successful infection.³⁸ Virulence factors are employed by the pathogen to overcome these hurdles and are a viable target for impairing the course of the respective disease. Publications of anti-virulence approaches are rising exponentially, proof of principle studies have been published and some compounds have entered clinical trials or are already approved by the FDA.^{39, 40}

Targeting the virulence of a pathogen has advantages over the use of antibiotics. Since the vital cell-functions are not affected and bacterial growth is not impaired, the evolutionary pressure for the development of resistance is thought to be much smaller or even non existing compared to antibiotics.^{41, 42} Furthermore, occurrence of a specific virulence factor is often restricted to one species, making the spread of resistance via horizontal gene transfer impossible. Likewise, anti-virulence drugs do not alter the host microbiota, one major side effect of broad spectrum antibiotics.^{43, 44} A major hurdle in the development of antibiotics is to ensure that the drug enters the pathogen cell and remains there long enough to hit its target. Several virulence factors act extracellularly and can therefore be targeted by drugs that are not restricted in their chemistry or size by requirements for cell permeability, e.g. antibodies against the anthrax virulence factor.⁴⁵

Beside this advantages of anti-virulence drugs over antibiotics, several intrinsic disadvantages exist. First of all, the promise of no resistance development may fall short of expectations. If the successful manifestation of the pathogenity is vital for host change or other means of survival for the pathogen, an evolutionary pressure does exist after all, with all its implications on resistance development.⁴⁶ Furthermore, the pathogen is not killed and per definition not impaired in its growth, leaving the work to the host immune system. Therefore an anti-virulence drug can only aid the antibiotic regime during immunosuppression, e.g. in HIV-positive patients or during chemotherapy in cancer treatment.⁴¹ Additionally, the remaining pathogen population

in the host can have severe side effects like septic shocks. Next to this systematic disadvantages, several practical difficulties occur in the development of anti-virulence drugs. High-throughput screenings in antibiotic development can be realized using a culture of the respective pathogen and detect growth, e.g. by optical density measurements. Knowledge of the underlying mode-of-action is not relevant at this stage of development and hit structures can be identified without it. The screening for anti-virulence drugs is more expensive and requires in-depth knowledge of the respective virulence factors to develop a clear strategy compared to the screening procedure for antibiotics, that can be realized with no more than liquid culture of a pathogen and a compound library. Most importantly, development of anti-virulence drugs demands the determination of a target prior to screening. Next, a suitable disease model, *in vitro* or *in vivo*, has to be developed that allows for rigid detection of hits. This linear strategy was also employed in this work that led to the identification of an tilivalline biosynthesis inhibitor. Tilivalline was previously identified as causative toxin for *K. oxytoca* mediated AAHC⁴⁷ and thus its biosynthesis qualified as viable target. Our work initially focused on the elucidation of the biosynthesis and was not only valuable from the biochemistry perspective, but also the foundation for the subsequent screening efforts for identifying inhibitors. The development of a suitable assay system to easily screen for tilivalline biosynthesis inhibition proved difficult. Three different models of tilivalline biosynthesis were established during this thesis: First of all, a tilivalline positive clinical isolate of *K. oxytoca* was available. Second, the tilivalline NRPS was reconstituted and *in vitro* formation of tilivalline from its precursors, L-proline, 3-hydroxy anthranilic acid, and indole, was established. Third, heterologous expression systems of the complete biosynthetic gene cluster and the NRPS alone were designed.

Screening of several inhibitors in the wild-type was not possible due to safety restrictions and laborious fermentation in 50 mL flasks. Furthermore, handling a clinical isolate does not satisfy the demands for a rigid and easy to use screening system. Therefore *K. oxytoca* was only used to validate the inhibitor salicylic acid and its precursor acetylsalicylic acid. *In vitro* reconstitution of the tilivalline NRPS proteins was unproblematic and vital for the elucidation of the biosynthetic pathway. Nevertheless, laborious expression and purification of three proteins and the use of expensive chemicals like precursors or cofactors discouraged the application as a screening platform. Eventually, the heterologous expression system of the tilivalline NRPS in *E. coli* was chosen to be developed as a screening platform for inhibitors. It was preferred over the system harboring the full biosynthetic gene cluster, that proved to produce tilivalline unreliable and in to small quantities. *E. coli* belongs to the same family (enterobacteriaceae) as *K. oxytoca* and therefore provides a suitable host that does not only expresses the relevant biosynthetic pathway, but also mimics the cell of the pathogen, ruling out the identification of potent inhibitors that can not access the cell. Nevertheless, timing, strength of induction and inhibitor concentrations have still to be optimized to provide reproducible hits. Furthermore, the non-enzymatic indole addition proved to be non-quantitative under the conditions used, impairing a rigid quantification. Therefore, indole is omitted in further optimization efforts and the carbinolamine precursor is detected instead. In general, detection and quantification of production of tilivalline or its precursors was in all cases realized by LC-MS analysis and subsequent integration of the EIC signal area. Although 96-well plates were employed, LC-MS seems too costly and tedious to be effective for its application in HTS approaches. Alternative starting points for developing

a rapid and cost-efficient method for quantification can possibly exploit the UV-activity or the cytotoxic properties.

The application of a tailored assay is characteristic for anti-virulence drugs, that are always developed with a specific model for a predefined virulence factor. Consequently, no broad-spectrum application will be achieved. In praxis, this makes empirical therapy difficult and requires costly identification of the present pathogen, for example with molecular biology tools, to subsequently choose the right anti-virulence drug. Tilivalline could prove as a viable biomarker for *K. oxytoca* caused AAHC, allowing for rapid identification. Nonetheless, in general more than one drug will likely be needed to treat a patient to cover several possible pathogens.³⁷ Concerning this practical disadvantages, an application for anti-virulence compounds as stand-alone drug in the treatment of infectious diseases seems unlikely. This holds also true for tilivalline and the associated AAHC. Although a successful developed biosynthesis inhibitor would be able to contain the symptoms in the colon, the underlying reason of the disease, a monoculture of *K. oxytoca*, is not addressed. Nonetheless, combined with other means of therapy tilivalline inhibition could improve the life quality of patients suffering from AAHC significantly, particularly because of the associated long time period of this disease.

6.4.2 Blocking the NRPS based biosynthesis of virulence factors

One viable target for anti-virulence drugs is the NRPS based biosynthesis of peptidic virulence factors like siderophores, that help pathogens to scavenge iron in the host. Past efforts to identify NRPS inhibitors were mostly focused on two key reactions: The adenylation of amino acid building blocks by the A-domain and the introduction of the obligatory PPant-modification by the PPant-transferase. While A-domains in general exhibit significant sequence differences depending on the respective substrate, PPant transferases are rather homologous and often active on several PCP-domains. *In vitro* based assay systems employing tritiated CoA or FRET have been established to allow HTS using compound libraries.^{48–50} Several compounds have been identified that also retained activity in the pathogen. ML267 is an allosteric inhibitor of Sfp, the PPant transferase of the model organism *B. subtilis*.^{51, 52} Calmidazolium and sanguinarine were both shown to inhibit the PPant transferase from *M. tuberculosis* with low micromolar IC₅₀. Although targeting the PPant transferase would presumably lead to a certain degree of broad-spectrum activity, none of the initial hits has been developed further.

Far more research has been done on inhibitors targeting the NRPS itself, more precisely the adenylation activity of the A-domain. Biosynthesis of salicylate-capped siderophores like yersiniabactin (*Y. pestis*), mycobactin (*M. tuberculosis*) or pyochelin (*P. aeruginosa*) have been successfully inhibited using salicylate-AMS derivatives, an non-hydrolyzable mimic of salicylate-adenylate. Salicyl-AMS was active in a mouse model, but further development was stopped due to toxic off-target effects.^{53, 54} AMS derivatives were also employed for inhibition of domains activating dihydroxybenzoate, another common building block for pathogen siderophores like enterobactin (*E. coli*), vibriobactin (*V. cholera*) and bacillibactin (*B. subtilis*). Expanding the strategy to amino acid building blocks is problematic, since human aminoacyl-tRNA synthetase are evolutionary related and perform similar chemistry, thus toxic side effects are likely. In general, the strategy of non-hydrolysable reaction intermediates for blocking NRPS systems

is not new. NRPS proteins are highly flexible and undergo a variety of conformational changes during the catalytic cycle. AMS derivatives like vinylsulfonamides bind covalently to the PCP domain and lock the system in a defined state. They display a group of highly valuable probes to facilitate biochemical assays or structural experiments by X-ray or cryo-EM.⁵⁵

Inhibition of the tilivalline NRPS with salicylic acid derivatives differs from the above described use of non-hydrolysable compounds. Focused on the adenylation domain NpsA, salicylic acid is a competitive substrate to the native 3-hydroxy anthranilic acid and does not affect the overall activity. The broad substrate tolerance of NpsA gave us the notion to employ this strategy. Adenylated salicylic acid subsequently acts as competitive inhibitor for ThdA, because once loaded onto the PPant arm, downstream processing is not possible due to the missing amino function. Generally speaking, this approach can face serious hurdles as a lot of organisms harbor type II thioesterases that detect and remove mis-loaded building blocks.⁵⁶ Furthermore, the thioester bond is not as stable as the covalent bonds formed by AMS derivatives and hydrolysis can occur without enzymatic catalysis. Therefore a complete abolishment of natural product biosynthesis is presumably hard to obtain, as reflected by our results where a final tilivalline production of 20 % remained even after the inhibition plateau was reached. Specifically for tilivalline this partial inhibition seems unproblematic, since high quantities of the toxin are likely produced and also a non-quantitative reduction will have beneficial effects for the patient. Nevertheless, if expansion of the competitive substrate inhibition strategy to other toxins is reasonable, will therefore be dependent on the individual case.

6.5 Conclusion

Natural products were, are and will be an indispensable part of pharmaceutical research. This work highlights several distinct ways in which natural product research contributes to the overall drug development process. First, natural products are still an important source for novel scaffolds: Corramycin is currently developed as antibiotic by the industry and the here presented studies on the biosynthesis and self-resistance mechanism can significantly aid the process of derivatization, what is necessary to advance to the clinical phase. Furthermore, every successful *in vitro* reconstitution of biosynthetic enzymes broadens the available tool kit for synthetic biology and thus helps to establish a more predictive generation of derivatives outside of classic organic chemistry. In this study, several PBD derivatives based on tomaymycin and tilivalline were produced in a convenient and fast way, enlarging the chemical diversity available for screening in a drug development environment. Additionally, the knowledge of biosynthetic pathways of virulence factors like tilivalline does not only help to exploit the activity of the toxin for pharmaceutical needs, but also paves the road for anti-virulence drugs, an alternative strategy to antibiotics that is urgently needed in times of rising resistance among pathogens. Summarized, natural product research is anything but obsolete and will continue to play a significant role in pharmaceutical research.

6.6 Bibliography

- [1] David J. Payne, Michael N. Gwynn, David J. Holmes, and David L. Pompliano. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nature reviews. Drug discovery*, 6(1):29–40, 2007.
- [2] Richard H. Baltz. Marcel faber roundtable: is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? *Journal of industrial microbiology & biotechnology*, 33(7):507–513, 2006.
- [3] Kim Lewis. New approaches to antimicrobial discovery. *Biochemical pharmacology*, 134:87–98, 2017.
- [4] A. Schatz, E. Bugie, and S. A. Waksman. Selman abraham waksman, ph.d. 22 july 1888–16 august 1973. streptomycin reported. *Annals of internal medicine*, 79(5):678, 1973.
- [5] Richard H. Baltz. Natural product discovery and development at eli lilly and company: one scientist’s view. *SIM News*, 55:5–16, 2005.
- [6] Richard H. Baltz. Renaissance in antibacterial discovery from actinomycetes. *Current opinion in pharmacology*, 8(5):557–563, 2008.
- [7] M. G. Watve, R. Tickoo, M. M. Jog, and B. D. Bhole. How many antibiotics are produced by the genus streptomyces? *Archives of microbiology*, 176(5):386–390, 2001.
- [8] J. Tao, P. Wendler, G. Connelly, A. Lim, J. Zhang, M. King, T. Li, J. A. Silverman, P. R. Schimmel, and F. P. Tally. Drug target validation: lethal infection blocked by inducible peptide. *Proceedings of the National Academy of Sciences of the United States of America*, 97(2):783–786, 2000.
- [9] P. S. Margolis, C. J. Hackbarth, D. C. Young, W. Wang, D. Chen, Z. Yuan, R. White, and J. Trias. Peptide deformylase in staphylococcus aureus: resistance to inhibition is mediated by mutations in the formyltransferase gene. *Antimicrobial agents and chemotherapy*, 44(7):1825–1831, 2000.
- [10] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, 46(1-3):3–26, 2001.
- [11] Rolf Muller and Joachim Wink. Future potential for anti-infectives from bacteria - how to exploit biodiversity and genomic potential. *International journal of medical microbiology : IJMM*, 304(1):3–13, 2014.
- [12] T. Arai. Actinomycetes: the boundary microorganisms. *Toppan Company Limited, Tokyo*, pages 561–651, 1976.
- [13] Janos Berdy. Bioactive microbial metabolites. *The Journal of antibiotics*, 58(1):1–26, 2005.
- [14] J. Herrmann, A. Abou Fayad, and R. Muller. Natural products from myxobacteria: novel metabolites and bioactivities. *Natural product reports*, 2016.

-
- [15] Ronald O. Garcia, Daniel Krug, and Rolf Muller. Chapter 3. discovering natural products from myxobacteria with emphasis on rare producer strains in combination with improved analytical methods. *Methods in enzymology*, 458:59–91, 2009.
- [16] Ronald Garcia, Klaus Gerth, Marc Stadler, Irineo J. Dogma, JR, and Rolf Muller. Expanded phylogeny of myxobacteria and evidence for cultivation of the 'unculturables'. *Molecular phylogenetics and evolution*, 57(2):878–887, 2010.
- [17] Tilmann Weber, Kai Blin, Srikanth Duddela, Daniel Krug, Hyun Uk Kim, Robert Brucconeri, Sang Yup Lee, Michael A. Fischbach, Rolf Muller, Wolfgang Wohlleben, Rainer Breitling, Eriko Takano, and Marnix H. Medema. antismash 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic acids research*, 43(W1):W237–43, 2015.
- [18] G. Chen, G. Y. Wang, X. Li, B. Waters, and J. Davies. Enhanced production of microbial metabolites in the presence of dimethyl sulfoxide. *The Journal of antibiotics*, 53(10):1145–1153, 2000.
- [19] Yukinori Tanaka, Takeshi Hosaka, and Kozo Ochi. Rare earth elements activate the secondary metabolite-biosynthetic gene clusters in streptomyces coelicolor a3(2). *The Journal of antibiotics*, 63(8):477–481, 2010.
- [20] Kirstin Scherlach and Christian Hertweck. Triggering cryptic natural product biosynthesis in microorganisms. *Organic & biomolecular chemistry*, 7(9):1753–1760, 2009.
- [21] Yunzi Luo, Bing-Zhi Li, Duo Liu, Lu Zhang, Yan Chen, Bin Jia, Bo-Xuan Zeng, Huimin Zhao, and Ying-Jin Yuan. Engineered biosynthesis of natural products in heterologous hosts. *Chemical Society reviews*, 44(15):5265–5290, 2015.
- [22] Jeffrey H. Kim, Zhiyang Feng, John D. Bauer, Dimitris Kallifidas, Paula Y. Calle, and Sean F. Brady. Cloning large natural product gene clusters from the environment: piecing environmental dna gene clusters back together with tar. *Biopolymers*, 93(9):833–844, 2010.
- [23] Silke C. Wenzel and Rolf Muller. Myxobacteria-'microbial factories' for the production of bioactive secondary metabolites. *Molecular bioSystems*, 5(6):567–574, 2009.
- [24] Olena Perlova, Jun Fu, Silvia Kuhlmann, Daniel Krug, A. Francis Stewart, Youming Zhang, and Rolf Muller. Reconstitution of the myxothiazol biosynthetic gene cluster by red/et recombination and heterologous expression in myxococcus xanthus. *Applied and environmental microbiology*, 72(12):7485–7494, 2006.
- [25] Yi Chai, Shiping Shan, Kira J. Weissman, Shengbiao Hu, Youming Zhang, and Rolf Muller. Heterologous expression and genetic engineering of the tubulysin biosynthetic gene cluster using red/et recombineering and inactivation mutagenesis. *Chemistry & biology*, 19(3):361–371, 2012.
- [26] Jun Fu, Silke C. Wenzel, Olena Perlova, Junping Wang, Frank Gross, Zhiru Tang, Yulong Yin, A. Francis Stewart, Rolf Muller, and Youming Zhang. Efficient transfer of two

- large secondary metabolite pathway gene clusters into heterologous hosts by transposition. *Nucleic acids research*, 36(17):e113, 2008.
- [27] Qiang Tu, Jennifer Herrmann, Shengbiao Hu, Ritesh Raju, Xiaoying Bian, Youming Zhang, and Rolf Muller. Genetic engineering and heterologous expression of the disorazol biosynthetic gene cluster via red/et recombineering. *Scientific reports*, 6:21066, 2016.
- [28] Yunzi Luo, Behnam Enghiad, and Huimin Zhao. New tools for reconstruction and heterologous expression of natural product biosynthetic gene clusters. *Natural product reports*, 33(2):174–182, 2016.
- [29] Julian Davies and Dorothy Davies. Origins and evolution of antibiotic resistance. *Microbiology and molecular biology reviews : MMBR*, 74(3):417–433, 2010.
- [30] Vanessa M. D’Costa, Christine E. King, Lindsay Kalan, Mariya Morar, Wilson W. L. Sung, Carsten Schwarz, Duane Froese, Grant Zazula, Fabrice Calmels, Regis Debruyne, G. Brian Golding, Hendrik N. Poinar, and Gerard D. Wright. Antibiotic resistance is ancient. *Nature*, 477(7365):457–461, 2011.
- [31] Peter M. Hawkey and Annie M. Jones. The changing epidemiology of resistance. *The Journal of antimicrobial chemotherapy*, 64 Suppl 1:i3–10, 2009.
- [32] Erica C. Pehrsson, Pablo Tsukayama, Sanket Patel, Melissa Mejia-Bautista, Giordano Sosa-Soto, Karla M. Navarrete, Maritza Calderon, Lilia Cabrera, William Hoyos-Arango, M. Teresita Bertoli, Douglas E. Berg, Robert H. Gilman, and Gautam Dantas. Interconnected microbiomes and resistomes in low-income human habitats. *Nature*, 533(7602):212–216, 2016.
- [33] Yi-Yun Liu, Yang Wang, Timothy R. Walsh, Ling-Xian Yi, Rong Zhang, James Spencer, Yohei Doi, Guobao Tian, Baolei Dong, Xianhui Huang, Lin-Feng Yu, Danxia Gu, Hongwei Ren, Xiaojie Chen, Luchao Lv, Dandan He, Hongwei Zhou, Zisen Liang, Jian-Hua Liu, and Jianzhong Shen. Emergence of plasmid-mediated colistin resistance mechanism mcr-1 in animals and human beings in china: a microbiological and molecular biological study. *The Lancet. Infectious diseases*, 16(2):161–168, 2016.
- [34] Henrik Hasman, Anette M. Hammerum, Frank Hansen, Rene S. Hendriksen, Bente Olesen, Yvonne Agerso, Ea Zankari, Pimlapas Leekitcharoenphon, Marc Stegger, Rolf S. Kaas, Lina M. Cavaco, Dennis S. Hansen, Frank M. Aarestrup, and Robert L. Skov. Detection of mcr-1 encoding plasmid-mediated colistin-resistant escherichia coli isolates from human bloodstream infection and imported chicken meat, denmark 2015. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*, 20(49), 2015.
- [35] Dawn M. Sievert, James T. Rudrik, Jean B. Patel, L. Clifford McDonald, Melinda J. Wilkins, and Jeffrey C. Hageman. Vancomycin-resistant staphylococcus aureus in the united states, 2002-2006. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 46(5):668–674, 2008.

-
- [36] Mohit Kumar. Multidrug-resistant staphylococcus aureus, india, 2013-2015. *Emerging infectious diseases*, 22(9):1666–1667, 2016.
- [37] Lloyd Czaplewski, Richard Bax, Martha Clokie, Mike Dawson, Heather Fairhead, Vincent A. Fischetti, Simon Foster, Brendan F. Gilmore, Robert E. W. Hancock, David Harper, Ian R. Henderson, Kai Hilpert, Brian V. Jones, Aras Kadioglu, David Knowles, Sigriethur Olafsdottir, David Payne, Steve Projan, Sunil Shaunak, Jared Silverman, Christopher M. Thomas, Trevor J. Trust, Peter Warn, and John H. Rex. Alternatives to antibiotics-a pipeline portfolio review. *The Lancet. Infectious diseases*, 16(2):239–251, 2016.
- [38] A. Casadevall and L. A. Pirofski. Host-pathogen interactions: redefining the basic concepts of virulence and pathogenicity. *Infection and immunity*, 67(8):3703–3713, 1999.
- [39] Benjamin K. Johnson and Robert B. Abramovitch. Small molecules that sabotage bacterial virulence. *Trends in pharmacological sciences*, 38(4):339–362, 2017.
- [40] Seth W. Dickey, Gordon Y. C. Cheung, and Michael Otto. Different drugs for bad bugs: antivirulence strategies in the age of antibiotic resistance. *Nature reviews. Drug discovery*, 2017.
- [41] Anne E. Clatworthy, Emily Pierson, and Deborah T. Hung. Targeting virulence: a new paradigm for antimicrobial therapy. *Nature chemical biology*, 3(9):541–548, 2007.
- [42] Richard C. Allen, Roman Popat, Stephen P. Diggle, and Sam P. Brown. Targeting virulence: can we make evolution-proof drugs? *Nature reviews. Microbiology*, 12(4):300–308, 2014.
- [43] Ilseung Cho and Martin J. Blaser. The human microbiome: at the interface of health and disease. *Nature reviews. Genetics*, 13(4):260–270, 2012.
- [44] Michael S. Gilmore, Marcus Rauch, Matthew M. Ramsey, Paul R. Himes, Sriram Varahan, Janet M. Manson, Francois Lebreton, and Lynn Ernest Hancock. Pheromone killing of multidrug-resistant enterococcus faecalis v583 by native commensal strains. *Proceedings of the National Academy of Sciences of the United States of America*, 112(23):7273–7278, 2015.
- [45] Thi-Sau Migone, G. Mani Subramanian, John Zhong, Letha M. Healey, Al Corey, Matt Devalaraja, Larry Lo, Stephen Ullrich, Janelle Zimmerman, Andrew Chen, Maggie Lewis, Gabriel Meister, Karen Gillum, Daniel Sanford, Jason Mott, and Sally D. Bolmer. Rax-ibacumab for the treatment of inhalational anthrax. *The New England journal of medicine*, 361(2):135–144, 2009.
- [46] S. Alizon, A. Hurford, N. Mideo, and M. van Baalen. Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *Journal of evolutionary biology*, 22(2):245–259, 2009.
- [47] Georg Schneditz, Jana Rentner, Sandro Roier, Jakob Pletz, Kathrin A. T. Herzog, Roland Bückler, Hanno Troeger, Stefan Schild, Hansjörg Weber, Rolf Breinbauer, Gregor Gorkiewicz, Christoph Högenauer, and Ellen L. Zechner. Enterotoxicity of a nonribosomal peptide causes antibiotic-associated colitis. *Proceedings of the National Academy of Sciences of the United States of America*, 111(36):13181–13186, 2014.

- [48] Cecile Leblanc, Thomas Prudhomme, Guillaume Tabouret, Aurelie Ray, Sophie Burbaud, Stephanie Cabantous, Lionel Mourey, Christophe Guilhot, and Christian Chalut. 4'-phosphopantetheinyl transferase pptt, a new drug target required for mycobacterium tuberculosis growth and persistence in vivo. *PLoS pathogens*, 8(12):e1003097, 2012.
- [49] Timothy L. Foley and Michael D. Burkart. A homogeneous resonance energy transfer assay for phosphopantetheinyl transferase. *Analytical biochemistry*, 394(1):39–47, 2009.
- [50] Adam Yasgar, Timothy L. Foley, Ajit Jadhav, James Inglese, Michael D. Burkart, and Anton Simeonov. A strategy to discover inhibitors of bacillus subtilis surfactin-type phosphopantetheinyl transferase. *Molecular bioSystems*, 6(2):365–375, 2010.
- [51] T. L. Foley, G. Rai, A. Yasgar, M. S. Attene-Ramos, M. D. Burkart, A. Simeonov, A. Jadhav, and D. J. Maloney. Discovery of ml 267 as a novel inhibitor of pathogenic sfp phosphopantetheinyl transferase (pptase), 2010.
- [52] Timothy L. Foley, Ganesha Rai, Adam Yasgar, Thomas Daniel, Heather L. Baker, Matias Attene-Ramos, Nicolas M. Kosa, William Leister, Michael D. Burkart, Ajit Jadhav, Anton Simeonov, and David J. Maloney. 4-(3-chloro-5-(trifluoromethyl)pyridin-2-yl)-n-(4-methoxypyridin-2-yl)piperazine- 1-carbothioamide (ml267), a potent inhibitor of bacterial phosphopantetheinyl transferase that attenuates secondary metabolism and thwarts bacterial growth. *Journal of medicinal chemistry*, 57(3):1063–1078, 2014.
- [53] Julian A. Ferreras, Jae-Sang Ryu, Federico Di Lello, Derek S. Tan, and Luis E. N. Quadri. Small-molecule inhibition of siderophore biosynthesis in mycobacterium tuberculosis and yersinia pestis. *Nature chemical biology*, 1(1):29–32, 2005.
- [54] Shichun Lun, Haidan Guo, John Adamson, Justin S. Cisar, Tony D. Davis, Sivagami Sundaram Chavadi, J. David Warren, Luis E. N. Quadri, Derek S. Tan, and William R. Bishai. Pharmacokinetic and in vivo efficacy studies of the mycobactin biosynthesis inhibitor salicylams in mice. *Antimicrobial agents and chemotherapy*, 57(10):5138–5140, 2013.
- [55] Alexander Koglin and Christopher T. Walsh. Structural insights into nonribosomal peptide enzymatic assembly lines. *Natural product reports*, 26(8):987–1000, 2009.
- [56] Magdalena Kotowska and Krzysztof Pawlik. Roles of type ii thioesterases and their application for secondary metabolite yield improvement. *Applied microbiology and biotechnology*, 98(18):7735–7746, 2014.