

Mass Spectrometry in
Natural Product Research:
Deciphering Microbial Biosyntheses and
Contributions to a Novel Screening Approach

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

von
Michael Hoffmann
Saarbrücken
2017

Tag des Kolloquiums: 25.08.2017

Dekan: Prof. Dr. Guido Kickelbick

Berichtersteller: Prof. Dr. Rolf Müller

Prof. Dr. Uli Kazmaier

Vorsitz: Prof. Dr. Andriy Luzhetskyy

Akad. Mitarbeiter: Dr. Martin Frotscher

Diese Arbeit entstand unter der Anleitung von Prof. Dr. Rolf Müller in der Fachrichtung Pharmazie der Naturwissenschaftlich-Technischen Fakultät der Universität des Saarlandes von April 2013 bis Juni 2017.

Danksagung

Zunächst möchte ich ganz besonders meinen Eltern, Gabi und Dieter, für ihre unablässige Unterstützung während meiner gesamten Ausbildung danken. Euer jahrelanger Einsatz und Schuften hat das Studium in dieser Form erst ermöglicht.

Ganz besonderer Dank gilt auch meinem Bruder Thomas. Er hat mir nicht nur Großteile seiner fachlichen Expertise vermittelt, sondern diente auch indirekt als mein Betreuer. Darüber hinaus agiert er auch privat als wichtiger Angelpunkt meines Lebens. Sein Verständnis für schwierige Situationen und die damit einhergehenden Ermutigungen haben stets neuen Elan hervorgerufen.

An dieser Stelle möchte ich mich auch ganz herzlich bei Aline bedanken, die mich sicher durch die Hochs und Tiefs der letzten Jahre begleitet hat, stets Verständnis hatte und Alles daran setzte mich in schwierigen Zeiten aufzubauen, auch wenn Ihr dafür nicht immer die nötige Dankbarkeit entgegengebracht wurde.

Meinem Doktorvater Prof. Dr. Rolf Müller möchte ich für die herzliche Aufnahme in seiner Arbeitsgruppe, für das entgegengebrachte Vertrauen zur eigenständigen Bearbeitung spannender Themen aber auch für das Übertragen von Verantwortung danken.

Zu guter Letzt möchte ich der ganzen Arbeitsgruppe und insbesondere meinen Bürokollegen/innen für die gute Zusammenarbeit danken. Ganz besonderer Dank gilt aber den „supergeilen Kollegen/innen“ Katrin Jungmann, Lena Keller, David Auerbach, Chantal Bader, Alexander von Tesmar und Tony Abou Fayad. Ihr wart stets bereit tiefgründige, wissenschaftliche Diskussionen zu führen sowie eure Erfahrungen mit mir zu teilen ohne dass zwischendurch der Spaß auf der Strecke blieb. Die Erinnerung an die gemeinsame Zeit mit euch wird mir stets ein Lächeln auf das Gesicht zaubern. Auf dass wir uns nie aus den Augen verlieren und noch viele Freitagsbiere miteinander genießen können. Ach und übrigens, tolle Stifte!

Zusammenfassung

Mikrobielle Naturstoffe und insbesondere Sekundärmetabolite besitzen das Potential als Grundlage für die therapeutische Arzneimittelentwicklung zu dienen. Dementsprechend ist die Entdeckung und Charakterisierung von neuartigen, bisher unbekannten Verbindungen ebenso wichtig wie ein besseres Verständnis der zugrundeliegenden Biosynthesen. Diese Dissertation adressiert beide Themen, indem nicht nur ein neuartiger Ansatz zur Erweiterung des Screenings nach myxobakteriellen Metaboliten beschrieben wird, sondern auch zur Aufklärung zweier Biosynthesewege beigetragen wird.

Tomaymycin, ein Pyrrolo[4,2]benzodiazepin (PBD), ist ein antitumorales Antibiotikum, das unter anderem von *Streptomyces achromogenes* produziert wird. Trotz bereits veröffentlichter Arbeiten zur Biosynthese blieb die Reihenfolge einiger Biosyntheseschritte bisher unklar. Nach einer vollständigen *in vitro* Rekonstitution in Kombination mit einer umfassenden, neu etablierten intakten Protein LC-MS Analyse war es möglich die Reihenfolge der einzelnen Biosyntheseschritte einzuschränken und schließlich den wahrscheinlichsten Biosyntheseweg zu zeigen. Darüber hinaus ermöglichte die LC-MS Methode zum ersten Mal den Nachweis aller biosynthetischen Zwischenschritte auf der Ebene intakter Proteine. Des Weiteren wurde die Biosynthese von Tilivallin, einem Toxin das Antibiotika-assoziierte hämorrhagische Kolitis (AAHC) verursacht, aufgeklärt. Die heterologe Expression des Biosynthese-Genclusters von Tilivalline sowie die anschließende *in vitro* Rekonstitution konnten auch den Ursprung des Indol-Restes erklären, der den Hauptunterschied zwischen Tilivallin und vielen anderen natürlichen PBDs darstellt. Das funktionierende heterologe System diente zusammen mit dem *in vitro*-Biosynthesesystem als wertvolle Screening-Plattform um potenzielle Inhibitoren für die Toxin-Biosynthese zu finden.

Schließlich wurde ein neuartiger Ansatz ausgearbeitet um bisher unbekannte myxobakterielle Metaboliten zu finden. Im Gegensatz zu Screenings, die auf Extrakten von vegetativen Zellen basieren, liegt in dieser Arbeit der Fokus auf Verbindungen, deren Vorkommen mit der Bildung von myxobakteriellen Fruchtkörpern korreliert. Diese Machbarkeitsstudie wurde mit einem der am besten beschriebenen Myxobakterien, *Myxococcus xanthus* DK1622, durchgeführt.

Summary

Natural products, and in particular secondary metabolites of microbes, possess potential to serve as a basis for therapeutic drug development. Along this line, the discovery and characterization of novel, so far unknown compounds is as essential as a better understanding about the underlying biosyntheses. This thesis addresses both topics by not only describing a novel approach to expand the screening for myxobacterial metabolites but also contributing to the elucidation of two biosynthetic pathways.

Tomaymycin, a pyrrolo[4,2]benzodiazepine (PBD), is an antitumor antibiotic produced, among others, by *Streptomyces achromogenes*. Despite already published insights on its biosynthesis, the timing of some steps remained uncertain. Following a complete *in vitro* reconstitution in combination with comprehensive, newly established intact protein LC-MS analysis was instrumental to narrow down timing of individual biosynthetic steps and eventually revealed the most likely biosynthetic pathway. In addition, the LC-MS method allowed the detection of all biosynthetic intermediate steps on intact protein level for the first time. Moreover, the biosynthesis of tilivalline, a causative toxin in antibiotic associated hemorrhagic colitis (AAHC), was elucidated. Heterologous expression of the tilivalline biosynthetic gene cluster and subsequent *in vitro* reconstitution clarified the origin of the indole moiety, which represents the main difference between tilivalline and many other natural PBDs. The functional heterologous expression together with the *in vitro* biosynthetic system served as valuable screening platform aiming to find potential inhibitors of the toxin's biosynthesis.

Finally, a novel approach to find so far unknown myxobacterial metabolites was elaborated. In contrast to screening extracts of vegetative cells, the focus laid on compounds whose appearance is somehow linked to the formation of myxobacterial fruiting bodies. This proof of concept study was performed on one of the best-described myxobacteria, *Myxococcus xanthus* DK1622.

Vorveröffentlichungen zur Dissertation

Teile dieser Arbeit wurden vorab mit Genehmigung der Naturwissenschaftlich-Technischen Fakultät, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen veröffentlicht oder sind derzeit in Vorbereitung zur Veröffentlichung.

Publikationen

von Tesmar A.¹, **Hoffmann M.**¹, Pippel J., Abou Fayad A., Werner S., Bauer A., Blankenfeldt W. & Müller R.: Total biosynthesis of tomaymycin comprehensively monitored on the nonribosomal peptide megasynthetase. *Cell Chem. Biol.* **2017**, in press

von Tesmar A., **Hoffmann M.**, Schmitt V., Abou Fayad A., Herrmann J. & Müller R.: *In vitro* reconstitution and heterologous expression of an enterotoxin produced by *Klebsiella oxytoca*. **2017**, manuscript ready for submission

Hoffmann M.¹, Auerbach D.¹, Panter F., Hoffmann T., Dorrestein P. & Müller R.: Homospermidine Lipids: A compound class specifically formed during fruiting body formation of *Myxococcus xanthus* DK1622. **2017**, manuscript ready for submission

Publikationen, die nicht Teil dieser Arbeit sind

Yin J., **Hoffmann M.**, Bian X., Tu Q., Yan F., Xia L., Ding X., Stewart A. F., Müller R., Fu J. & Zhang Y.: Direct cloning and heterologous expression of the salinomycin biosynthetic gene cluster from *Streptomyces albus* DSM41398 in *Streptomyces coelicolor* A3(2). *Sci. Rep.* **2015**, 5, 15081, DOI: 10.1038/srep15081

Yin J., Zhu H., Xia L., Ding X., Hoffmann T., **Hoffmann M.**, Bian X., Müller R., Fu J., Stewart A. F. & Zhang Y.: A new recombineering system for *Photobacterium* and *Xenorhabdus*. *Nucleic Acids Res.* **2014**, 1–9, DOI: 10.1093/nar/gku1336

Gemperlein K., **Hoffmann M.**, Huo L., Pilak P., Petzke L., Müller R. & Wenzel S. C.: Synthetic biology approaches to establish a heterologous production system for coronatines. *Metab. Eng.* **2017**, submitted

¹ Co-first Authors

Wagner S., Hauck D., **Hoffmann M.**, Sommer R., Joachim I., Müller R., Imberty A., Varrot A. & Titz A.: Covalent lectin inhibition and its application in bacterial biofilm imaging. *Angew. Chemie Int. Ed.* **2017**, submitted

Tang Y., **Hoffmann M.**, Zipf G., Steinmetz H., Andes S., Bernauer H. S., Wenzel S. C. & Müller R.: Synthetic biology approaches to establish and engineer heterologous argyirin production in *Myxococcus xanthus*, manuscript in preparation

Grüter A., **Hoffmann M.**, Wohland T., Müller R., Jung G.: Fluorescence Lifetime Correlation Spectroscopy for Copper(II) sensing, manuscript in preparation

Witzgall F., Depke T., **Hoffmann M.**, Müller R., Brönstrup M. & Blankenfeldt W.: Substrate specificity of PqsBC defines the 2-alkyl-4-quinolone repertoire of *Pseudomonas aeruginosa*, manuscript in preparation

Table of Content

Danksagung	IV
Zusammenfassung	V
Summary	VI
Vorveröffentlichungen zur Dissertation	VII
Table of Content.....	IX
1 Introduction	13
1.1 Natural Product Research.....	13
1.2 Myxobacteria and their potential.....	16
1.3 The Role of Analytics in Natural Product Research	22
1.3.1 Metabolomics.....	22
1.3.2 Proteomics	23
1.4 Outline of this Work	25
1.5 References.....	27
Chapter 2 – Tomaymycin	33
Contributions.....	34
2 Tomaymycin.....	35
2.1 Abstract.....	35
2.2 Introduction.....	35
2.3 Results and Discussion.....	37
2.3.1 <i>In vitro</i> Reconstitution of TomA and TomB.....	37
2.3.2 Intact Protein MS of TomA and TomB.....	39
2.3.3 <i>In vitro</i> Reconstitution and Structural Analysis of TomG.....	44
2.4 Significance	47
2.5 Methods.....	49
2.5.1 Protein Preparation Procedure	49
2.5.2 Activity Assays.....	50
2.5.3 MS-based Analysis	51
2.5.4 Protein Structure Determination and Modeling.....	52
2.5.5 Synthesis.....	54
2.5.6 Data and Software Availability	60
2.5.7 Key Source Table	61

2.6	Supplemental Figures.....	64
2.7	Supplemental Tables	69
2.8	References	71
Chapter 3 – Tilivalline.....		35
	Contributions	78
3	Tilivalline.....	79
3.1	Abstract.....	79
3.2	Introduction	79
3.3	Results.....	80
3.3.1	Clinical Isolates of <i>K. oxytoca</i> Produces Tilivalline	80
3.3.2	Heterologous Expression of the Tilivalline Biosynthetic Gene Cluster Reveals Non-Enzymatic Indole Incorporation	82
3.3.3	<i>In vitro</i> Reconstitution of the Tilivalline NRPS.....	84
3.3.4	Tilivalline and not the Carbinolamine is the Active Species.....	85
3.3.5	Tilivalline NRPS as Heterologous Synthetic Biology Platform	86
3.3.6	Salicylic Acid Based Compounds Inhibit Tilivalline Production in <i>K. oxytoca</i>	87
3.4	Discussion.....	88
3.5	Experimental Procedures	91
3.5.1	Sequencing of Strain <i>K. oxytoca</i> Isolate #6.....	91
3.5.2	Isolation of Tilivalline	91
3.5.3	TAR Assembly of the Tilivalline Biosynthetic Gene Cluster	91
3.5.4	Heterologous Expression of the Tilivalline Biosynthetic Gene Cluster.....	92
3.5.5	Heterologous Expression of the Tilivalline NRPS	92
3.5.6	Protein Purification of NpsA, ThdA, and NpsB	92
3.5.7	Purification of TnaA.....	93
3.5.8	<i>In vitro</i> Reconstitution of Tilivalline NRPS	93
3.5.9	<i>In vitro</i> Reconstitution of TnaA.....	93
3.5.10	<i>In vitro</i> Reconstitution of Tilivalline NRPS and TnaA.....	94
3.5.11	DNA Displacement Assays	94
3.5.12	Inhibition of Tilivalline Production in <i>K. oxytoca</i> Isolate #6.....	94
3.5.13	Analytical Methods	94
3.6	Supplemental Information	96
3.6.1	Synthetic Procedures.....	103

3.7	References.....	106
Chapter 4 – Fruiting Bodies		109
	Contributions.....	110
4	Fruiting Bodies	111
4.1	Abstract.....	111
4.2	Introduction.....	111
4.3	Results and Discussion.....	113
4.3.1	Comparative Analysis of Submersed Cultures and Fruiting Bodies Grown on Agar.....	113
4.3.2	MALDI-MSI of <i>M. xanthus</i> DK1622 Grown on Minimal Medium.....	116
4.3.3	Structure Elucidation of Cmp-552 and Identification of Related Compounds.....	117
4.3.4	Bioactivity and Biosynthesis of Homospermidine-Lipids.....	119
4.4	Conclusion.....	120
4.5	Methods.....	122
4.5.1	Cultivation and Extract Preparation.....	122
4.5.2	LC-hrMS Methods.....	122
4.5.3	MALDI-MSI Method.....	123
4.5.4	MS ⁿ Experiments and Molecular Networking	123
4.5.5	LC Method for Semi-Preparative Purification	124
4.5.6	Bioactivity testing.....	125
4.6	Supplemental Figures & Tables	126
4.7	References.....	137
5	Discussion	141
5.1	<i>Tomaymycin and Tilivalline</i>	143
5.1.1	Pyrrolobenzodiazepines.....	143
5.1.2	<i>In Vitro</i> Reconstitutions: Benefits and Obstacles	144
5.1.3	The Role of Instrumental Analytics in <i>in vitro</i> Experiments	148
5.1.4	Industrial Usage of <i>in vitro</i> Platforms	151
5.2	Alternative Approaches towards the Isolation of Unknown Metabolites	153
5.3	References.....	161

1 Introduction

1.1 Natural Product Research

Natural product research has become a well-established area in science ever since its advent in the early 19th century. Looking back in history, using of, and benefitting from natural products is rather as old as civilization itself. Various diseases threatened human health ever since and it goes without saying that humans have always been eager to treat or prevent all types of threats. Although a precise knowledge about the origins of diseases were missing until the 19th century, people tried to medicate patients or at least alleviate sufferings by using the almost inexhaustible source of natural products. In the first instance, there was no need to know chemical structures or even the mode of action of the compounds that made up an ointment or potion as long as a positive effect for the patients could be observed. The ancient Egyptians already utilized natural products for treatments as proven by two of the oldest texts – known as *Papyrus Ebers* and *Papyrus Edwin Smith*^{1–3}. These writings are dated to the 16th century before Christ and contain a large spectrum of descriptions about different diseases, associated symptoms as well as the suggested treatments and instructions for preparation of medicines. There are, however, many other impressive ancient writings giving evidence of the usage of natural products for treatments like the Chinese *Materia Medica*, *Shennong Herbal* or *Tang Herbal* which are all ascribed prior to the 7th century AD^{1,4}. Thus, the advantages of using natural products and the related diversity of chemical substances to fight diseases were already appreciated by people living more than 3000 years ago.

The reasons for the observed positive effects or even healings remained elusive until the advent of modern medicine and natural sciences. This situation changed in the 19th and 20th century by a tremendous progress in science. Especially with respect to infectious diseases, Edwin Kleb, Jakob Henle, Friedrich Löffler and Robert Koch set a milestone when they first postulated and later proved that infections are often linked to microorganisms which can cause diseases as serious as anthrax and tuberculosis^{5,6}. This very discovery paved the way for targeted treatments of many pathogens as it became clear that a success in healing is directly linked to the treatment of a microbial infection. Finally, the breakthrough came by the discovery and isolation of the first broad-spectrum antibiotic penicillin from the fungus *Penicillium notatum*⁷. The term antibiotic has been defined and penicillin has become a game changer. From this date on, it was obvious that some microorganism can defend themselves against competitors by producing *anti-microbial compounds* and can thus serve as a source for new antibiotics. In the following, the so-called “golden era” of antibiotics revealed many classes of antibiotics, all of which were either directly isolated from different microorganisms or derived from such compounds^{8,9}. As a result of this rocketing

development the problem of infectious diseases was assumed to be solved, the screening for new lead structures stagnated and researchers focused on the optimization of existing compounds¹⁰. Soon it became obvious that this assumption and the resulted strategy failed as resistance developed and its pace surpassed the researchers' expectations.

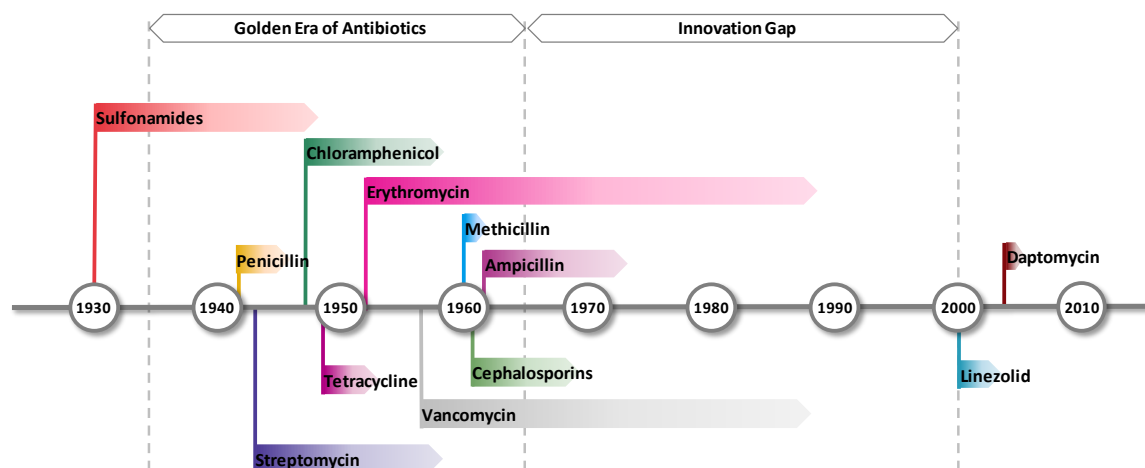


Figure 1.1: Antibiotic discovery and the period until resistance was observed^{9,11}.

Predominantly, the overuse of the new drugs in combination with the short generation time and related mutation rate of many pathogens resulted in the formation of various resistances against almost all known antibiotics within only one century^{12–14}. Fortunately, in most cases the resistance mechanism of the pathogens are limited to a distinct class of antibiotics and can thus be circumvented by using other antibiotics with different modes of action. The more serious issue is the development of pan-resistant pathogens, which cannot be sufficiently treated with any known antibiotic. In this regard, a group of hospital-acquired (nosocomial) infections based on the so-called *ESKAPE bugs* is considered as highly problematic. This group comprises seven pathogens, namely *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumonia*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter species* which cannot be treated appropriately owing to their acquired resistance mechanisms^{15,16}. In this context the WHO recently published a priority pathogens list of twelve bacteria for which new antibiotics are urgently needed. The listed pathogens are subdivided into three groups: critical, high, and medium priority. The first two groups do already comprise nine pathogens, which emphasized the seriousness of the situation¹⁷. Consequently, a need for new antibiotics arose and the search for new antibacterial scaffolds with novel modes of action became a major field of research. Microbes were already revealed as a promising source of antibiotics and constituted thus the focus of scientific efforts to address this issue. Especially, bacteria belonging to the *streptmycetes* became the most popular source of bioactive metabolites¹⁸.

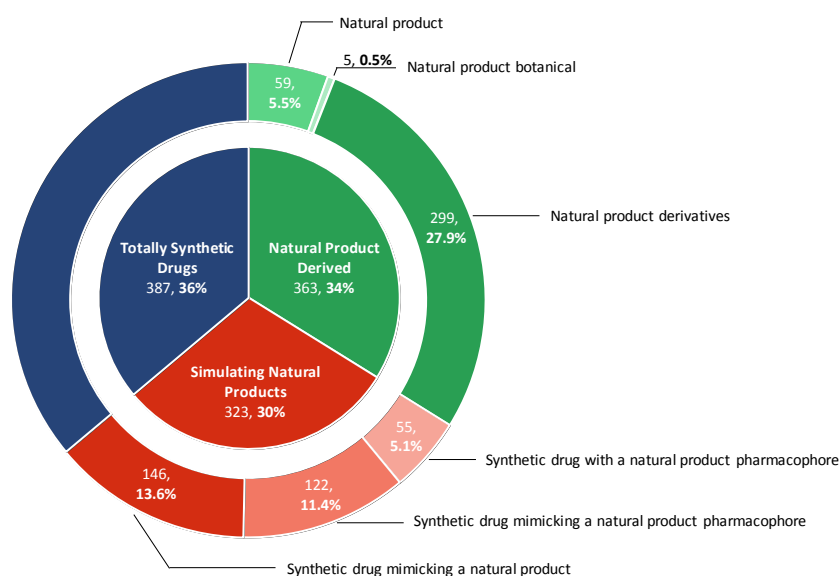


Figure 1.2: Origin of small-molecule approved drugs 1981-2014. Adapted from Newman and Cragg⁸.

Nevertheless, despite the experience and knowledge about the usage of bacteria as a source for new antibiotics it is a challenging endeavor for two basic reasons. Firstly, the low-hanging fruits have been largely harvested leading to the fact that bioactivity screening yields the same structures repeatedly. Hence, considerable effort is required to find novel structures from extensively exploited sources. Secondly, only a very small number of bioactive compounds make it finally to the drug approval and the market¹⁹. Consequently, it is not enough to find only a few new structures with novel modes of action but rather hundreds of new structures to increase the chance towards a new approved drug. Although extensive developments in the field of analytical chemistry have improved the chances of success by high-throughput screening and corresponding data analysis approaches, it is important to choose additional and promising microbial resources. Underexplored and uncommon bacteria like the myxobacteria as well as strains living in biological niches or habitats could represent such sources. Furthermore, alternative cultivation approaches can also lead to the appearance of so far unknown metabolites and thus increase the chance to find new bioactive compounds.

1.2 Myxobacteria and their potential

Myxobacteria are a group of primarily soil-dwelling bacteria belonging to the class of δ -proteobacteria. Roland Thaxter, who originally isolated the first myxobacteria in 1892, already recognized this species as an unusual one with noteworthy characteristics such as their social behavior, gliding motility, and complex life cycle²⁰. Interestingly, most of these features are directly related to the nutrient supply of the bacteria²¹. One of these features is the ability of rod-shaped vegetative cells of myxobacteria to move by gliding over a surface to find new nutrient sources. Although this gliding motility is not fully understood, it is known that two motility systems are responsible for the movement²². Myxobacteria can utilize various sources of nutrients like single amino acids, sugars or bio-macromolecules which were lysed by the excretion of appropriate exoenzymes^{21,23}. Some myxobacterial strains even show the ability to prey competitive microorganisms in their environment like *E. coli*²⁴. For this purpose, they synthesize antimicrobial secondary metabolites to kill the competitor and thus enable the lysis of the cells and the usage as a source of nutrition respectively^{25,26}. In contrast to that not all competitive microbes are fought in order to use them as nutrition but also to defend an ecological niche or decrease the number of food competitors²⁴. Thus, there is a certain chance that myxobacteria are able to produce various bioactive compounds to tackle other microorganisms.

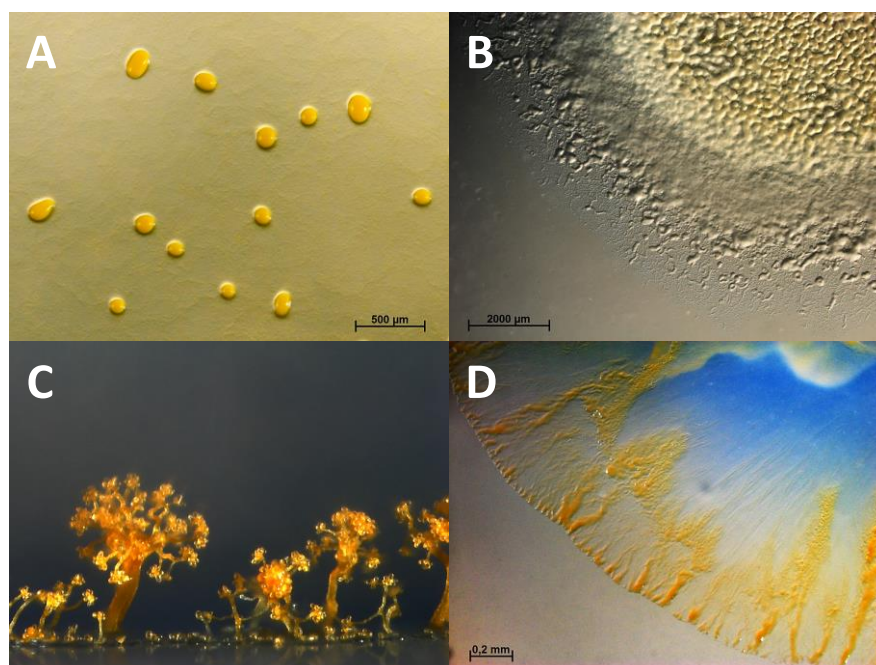


Figure 1.3: **A** *Myxococcus xanthus* DK1622 fruiting bodies on agar. **B** Swarming *Myxococcus xanthus* DK1622. **C** Fruiting bodies of *Chondromyces crocatus* Cmc5 on edge of a Petri dish. **D** *Chondromyces crocatus* Cmc5 swarm colony on agar. (Pictures were taken by Ronald Garcia)

Another distinct feature of the complex myxobacterial life cycle is the among prokaryotes unique ability to form multicellular, spore-containing fruiting bodies. As the shape of such fruiting

bodies and spores show a large variety among the myxobacterial families, the study of their morphology is a helpful tool for fast provisional taxonomic characterization²⁷. The formation of fruiting bodies is primarily based the lack of nutrition or unfavorable environmental conditions and aims to help the bacteria to survive such times. During formation, the vegetative cells aggregate to multicellular mounds in which roughly 90 % of the cells pass a metamorphosis into spherical myxospores and the remaining 10 % of undifferentiated cells enclose the transformed cells as a monolayer. The resulting spores are non-vegetative dormant cells which are characterized by high stress-resistance regarding harsh environmental. Myxobacteria can remain in this status until conditions are more favorable. If nutrients become once again available, the myxospores germinate, eventually retransform into vegetative cells, and enter the normal growth cycle²⁸. While the intra-cellular signaling and translation level during different stages of myxobacterial life cycle has been intensively analyzed, the changes in secondary metabolism received only little attention²⁹. However, it is conceivable that during fruiting body formation not only the physical resistance is increased but also some kinds of chemical defense systems, basically to actively guard the dormant cells. To the best of my knowledge, no scientific paper is published at this time that deals with the secondary metabolism of myxobacteria during fruiting body formation. Therefore, the present work aims to lay the foundation for this particular research question (see chapter xxx).

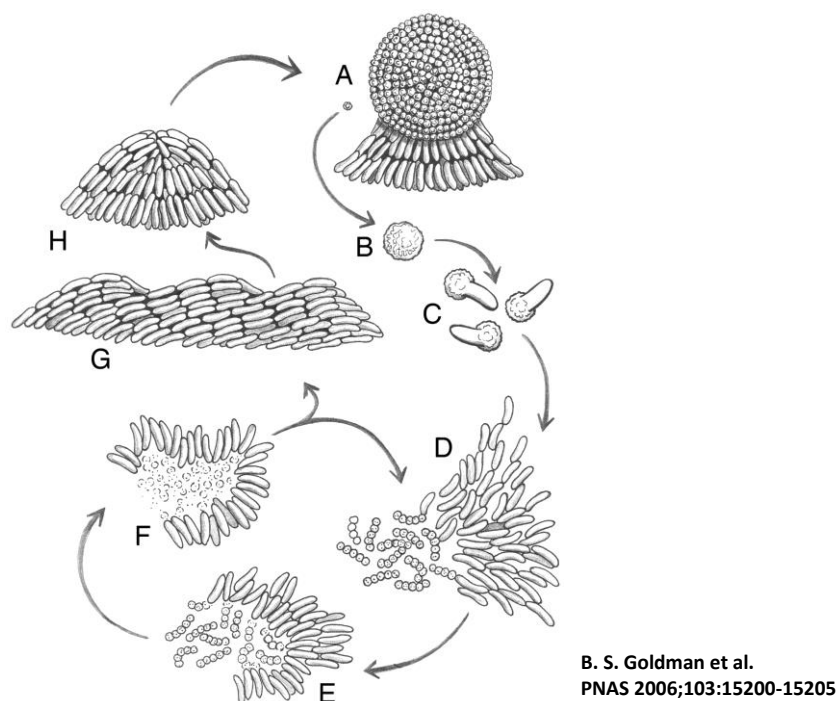


Figure 1.4: Overview of the different myxobacterial life cycle of *Myxococcus xanthus* DK1622. **A** Multicellular fruiting body. **B** Stress-resistant endospore. **C** Germinating spores due to sufficient nutrition supply. **D** thousands of vegetative cells merge together as a swarm. **E and F** Swarm preys competitor microbes by surrounding, contacting, lysing, and using them as a source of nutrition. **G** Starvation conditions lead to a stress response that initiates changed cell movement and aggregation. **H** Cells stream into mounded aggregates.

Regardless to this particular approach, myxobacteria have already proved to be a proficient source of new natural products including bioactive compounds showing so far unknown modes of action³⁰. The potential of myxobacteria as promising source of new structural scaffolds is also reflected in the exceptional genome size. Collected sequence data of many myxobacteria revealed an average genome size of 9-15 Mio. base pairs which thus belong to the largest known bacterial genomes^{31,32}. In this context, it is also noteworthy that myxobacterial DNA shows an uncommon high GC content of 66-72 mol-% which is comparable to the in natural product research well established *actinomycetes*²⁴. Interestingly, up to 10 % of the genetic information seems to be responsible for the biosynthesis of secondary metabolites, which corresponds to an average number of 10-20 encoded clusters per strain^{29,33-35}. These clusters consist of huge protein complexes belonging to the polyketide synthase (PKS) type, the nonribosomal peptide synthetase (NRPS) type or hybrids thereof and are responsible for the stepwise biosynthesis of most of the secondary metabolites. Such mega enzymes may be compared to assembly lines where a complex structure is formed by the attachment of simple building blocks and subsequent modifications. Therefore, gene clusters can usually be split into modules while each module consists of different domains. Each module is responsible for one building block, i.e. the elongation of the molecule including modifications. Depending on the type of enzymes, the distinct modules consist primarily of domains like acyl transferase (AT), ketosynthase (KS), acyl carrier protein (ACP), keto reductase (KR), dehydratase (DH), enoyl reductase (ER) in case of PKS or peptidyl carrier protein (PCP), adenylation (A), condensation (C) and/or epimerization domains (E) in case of NRPS. For a more detailed description of the biosynthesis by PKS and NRPS the reader is referred to textbooks and review articles³⁶⁻³⁸.

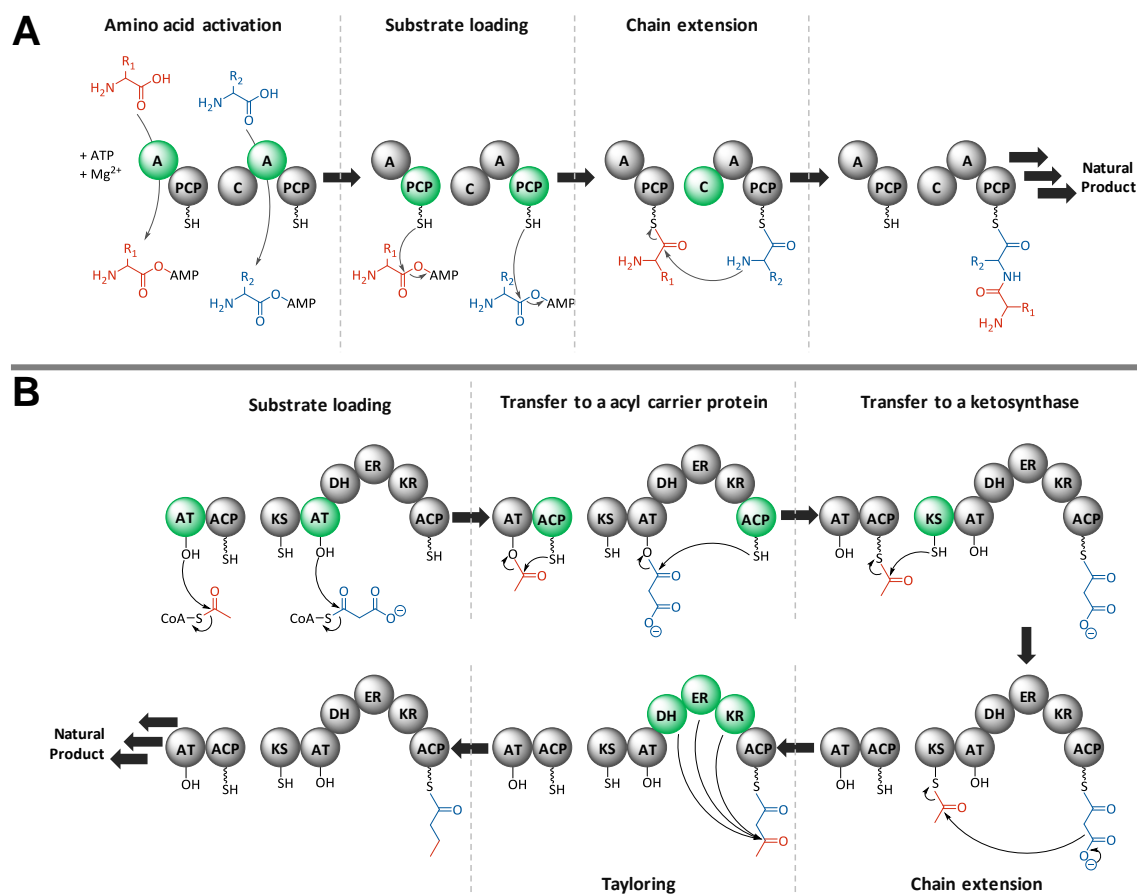


Figure 1.5: Schematic reaction steps of a nonribosomal peptide synthetase (NRPS, **A**) and a polyketide synthase (PKS, **B**). The NRPS assembly line consists of peptidyl carrier protein (PCP), adenylation (A), and condensation (C) domains to obtain chain extension of the natural product intermediate until finally the natural product is released. The shown PKS comprises acyl transferase (AT), ketosynthase (KS), acyl carrier protein (ACP), keto reductase (KR), dehydratase (DH), enoyl reductase (ER) to elongate and modify the growing substrate chain until product release is taking place.

Considering the large number of clusters and the associated combinatory possibilities owing to PKS/NRPS derived biosynthesis a variety of different chemical structures occurs. The diversity of so far isolated structures ranges from rather small molecules with uncommon substituents over macrocycles with synthetically almost inaccessible substructures to large linear molecules exhibiting complex stereochemistry. Examples of such isolated structures are hyaladione³⁹, chlorotonil⁴⁰, vioprolide⁴¹, salimabromide⁴², myxovalargin⁴³ and ambruticin⁴⁴. It is not surprising that many isolated compounds are accompanied by different bioactivities. The majority of secondary metabolites show activity against bacteria or fungi, which are common competitors in the natural habitat of myxobacteria³⁰. Nevertheless, some compounds have, for example, cytotoxic or anti-viral activity, which cannot be directly attributed to an evolutionary advantage for the bacteria towards competitors in the habitat. Some activities simply happen to be useful in an unexpected way. As such, the cytotoxic activity of one myxobacterial compound, epothilone, turned out to be a success story for cancer treatment. Epothilone has similar to taxol antineoplastic

activity and lead after slight structural changes to ixabepilone (Ixempra®) which was approved by the FDA to treat taxol resistant breast cancer^{45–47}. It is likely that such activities are coincidence and the real reasons for the biosynthesis of such compounds remains unclear. Certainly, not all secondary metabolites are produced for defense purposes but even those could show activity against unexpected targets⁴⁸. This underscores the need for broad testing to exploit the real potential of myxobacteria as source of new natural products.

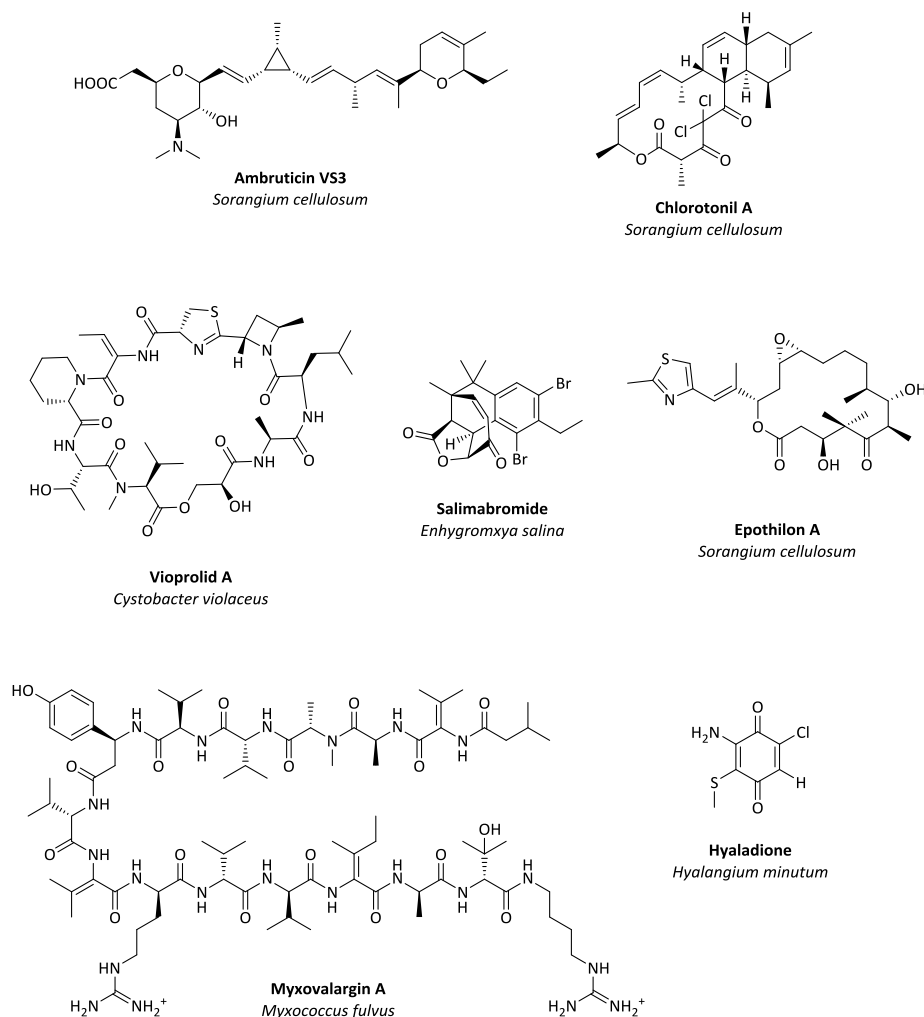


Figure 1.6: Selection of myxobacterial secondary metabolites from various strains reflecting the diversity of chemical scaffolds.

Although sequence data can provide the knowledge about the number of clusters per strain, even for well-studied organisms like *Myxococcus xanthus* DK1622 some clusters cannot be assigned to any isolated compound²⁹. The fact that this problem does not only appear for DK1622 but also for many other intensively studied myxobacteria shows that the pure knowledge about the expected metabolites does not necessarily mean that this compound is detected in bacterial extracts. This observation could be due to non-ideal cultivation conditions, the absence of certain triggers or just caused by very low production levels. Various approaches have been emerged over

the last years helping to address this problem. Affordable genome sequencing led to a plethora of identified gene clusters. At the same time, analytical methods as well as molecular biological methods undergo continuous development and help to understand the functions of the gene clusters. Consequently, a steadily increasing knowledge about the correlation of structural properties and sequence data evolved. This cumulated in rules which aid the prediction of substructures of PKS/NRPS derived compounds just based on genome data. Nevertheless, the *in-silico* prediction of a complete structure remains impossible.

1.3 The Role of Analytics in Natural Product Research

1.3.1 Metabolomics

Modern analytical methods became one of the most important tools in natural product research⁴⁹. In the course of improved knowledge in chemistry and biology during the last centuries, a high demand for corresponding analyses arose. Scientists increased their efforts towards elucidation of the ingredients of complex mixtures such as extracts of any living organism^{50,51}. Besides biosynthetic aspects, compound identification remains one of the key issues in natural product research. New developments in analytical chemistry have been rapidly applied to natural product research to improve this field of research. The first usage of chromatography at the beginning of the 20th century revealed the advantages of analytic methods. When Michail Tswett managed to separate Chlorophyll a and b he introduced a totally new method for natural product purification⁵². The subsequent development of X-ray crystallography and NMR techniques finally facilitated the molecular structure elucidation of isolated compounds⁵³. Based on this progress the chemical structure of penicillin, for instance, could be elucidated in 1945 by X-ray crystallography⁵⁴. Despite major improvements in the field of structure elucidation in the last decades a high-throughput application remains infeasible^{55,56}. Required quantity and quality of compounds for these methods can usually not be offered during initial screening. Fortunately, other analytical methods were increasingly established and further optimized towards the identification of natural products. In this context, the advantages of mass spectrometry became apparent⁵⁷. Besides a broad spectrum of mass spectrometer types, also the possibility to combine them with different separation techniques cumulated into diverse applications⁵⁸. Nowadays, liquid chromatography and gas chromatography coupled with mass spectrometry (LC-MS and GC-MS) are probably the most important analytical setups in natural product research⁵⁹.

Especially for secondary metabolomics, LC-MS systems became an integral part of daily lab work. Applying this technique revealed the complexity of many crude extracts as also low concentrated compounds or those with little or no UV activity were detectable for the first time. Consequently, the spectrum of potential targets for isolation and structure elucidation was expanded; already processed extracts were returned to the workflow and analyzed for non-UV-active substances. Optimization of the instruments and new trends steadily increased the information content per measurement and helped researchers to work more efficiently. The introduction of ultra-high performance liquid chromatography (UHPLC) enabled maximum working pressures of >1000 bar and thus the use of sub-2 μ m particles for separation⁶⁰. Thereby even separations of complex mixtures were achievable in remarkably short times of several minutes^{61,62}. In addition, this development was supported by the advent of high resolution time of flight mass

spectrometer (ToF-MS) which show high mass accuracy and very short duty cycles⁶³. Owing to the high mass accuracy of <5 ppm and the sufficient resolving power to differentiate between the isotope signals sum formula prediction became feasible and could be theoretically performed for each chromatographic peak of a LC-MS measurement. The robustness and high reproducibility of such LC-MS setups demonstrated the advantage in their application in extensive screening projects. Based on standardized analysis in combination with high accurate mass to charge ratios (m/z) and reproducible retention times, databases could be established and helped to focus on new compounds⁶². Probably, the most important advantage of this workflow is the possibility of dereplication and the related prevention of isolating already known compounds. However, databases facilitate additional opportunities for natural product research such as metabolome mining. In this manner, for instance, databases can be used to search for better producers of an already selected target compound and simplifies by that the isolation of sufficient amounts⁶⁴. Furthermore, MS/MS fragmentation of distinct compounds can already give first hints about structural elements or comparison of whole batches of fragmentation spectra can be used to reveal molecular networks, i.e. grouping MS/MS spectra based on similar fragmentation^{65,66}. Consequently, it is possible to assign certain unknown compounds either to known compound classes and predict the structure based on known fragmentation patterns. Along those lines, highlighting yet unknown molecule classes are possible by explicitly looking for MS/MS spectra, which do not fit an existing group. Scientist are usually more interested in the latter as it increases the chance to find new core structures. Another approach to find so far unknown secondary metabolites is based on statistical analysis⁵⁹. Principal component analysis (PCA), t-tests or receiver operating characteristic curve analysis (ROC) can help to find discriminants between distinct samples such as wild type and knock out mutants where results of such analysis can for instance connect genomic and metabolomic data^{67,68}. Despite these improvements, the gain of information per measurement has been steadily growing ever since owing to the advances in analytical hardware and data evaluation.

1.3.2 Proteomics

Besides metabolomics, proteomics are also extensively studied among natural product research⁶⁹. Once new chemical scaffolds or new bioactive structures are identified and characterized, the elucidation of biosynthesis moves into the center of interest. With regard to an increased need for the compound or the potential production of derivatives by mutasynthesis or synthetic biology the full knowledge of biosynthesis is required⁷⁰⁻⁷². Therefore, biosynthetic gene clusters are initially determined and tried to transfer into a heterologous host which should be thus able to translate the DNA information into the respective biosynthesis protein⁷³. Before any *in vitro*

assay is performed, it is helpful to ensure that the enzymes have the expected characteristics. Although the mass of the proteins can be estimated by gel electrophoresis it is impossible to obtain an exact mass, the protein sequence or to distinguish between *apo* and *holo* forms. Therefore, instrumental analytics are also extensively applied in this field of research⁷⁴⁻⁷⁶. There are in general two approaches for the identification and characterization of proteins: bottom-up and top-down. For the bottom-up approach, proteins are digested by a proteolytic enzyme and the received peptides are analyzed either by LC-MS or by MALDI-MS setups. Following this approach enables the confirmation of wide sequence areas and additional posttranslational modifications (PTMs) if the respective peptide fragment is detected^{77,78}. While measurements based on LC-MS have the advantage of a straightforward analysis with an additional separation step, MALDI is highly sensitive towards peptides and offers a high-throughput approach due to the very short period of time per measurement⁷⁹. Nowadays, MALDI-MS is further improved and can be combined with a previously performed nano-LC separation which is however, rather time consuming and challenging to handle⁸⁰. A fundamental problem of the bottom-up approach is the required digestion of the protein. On the one hand, a 100 % sequence coverage is hardly ever reached as not all of the fragments can be detected. On the other hand, some modifications are labile and do not endure the sample preparation. This is especially a problem for detailed analysis of PKS and NRPS based biosynthesis as the formed thioester coupling of substrate and proteins is labile under digestion condition⁸¹. This problem can be circumvented by applying the top-down approach. Thereby, the proteins are measured intact without any previous proteolytic digestion. The advantage of such analysis is the fast and reliable measurement as, for instance, *in vitro* assays can be measured without any laborious sample preparation. Nevertheless, this method is limited in terms of maximal detectable protein size. Depending on the behavior during the ESI process this limit varies from 200 kDa to 18 MDa^{82,83}. Furthermore, as shown in this work, insufficient purity of proteoforms hampers the measurements as chromatographically overlay of different species surpasses the resolution power of the mass spectrometers. Owing to the progress in this field, the current limits are likely overcome soon, thereby opening up new possibilities. For a more detailed overview of the usage of instrumental analytics in metabolomics and proteomics the reader is referred to respective review articles^{62,84,85}.

1.4 Outline of this Work

This work covers several areas of natural product research and can be divided into three parts. The first part deals with tomaymycin, a NRPS derived natural product initially isolated from *Streptomyces achromogenes* in 1972. Tomaymycin provoked the interest of Arima et al., especially because of the observed antibiotic, antiphage and antitumor activity⁸⁶. Subsequently performed mode of action studies with PBDs indicated the formation of a complex with DNA, which could be confirmed as covalent binding to the minor groove of the DNA in following studies^{87–89}. Based on this knowledge the underlying structure activity relationship has been studied intensively and led to a plethora of PBDs derivatives with increased antitumor activity^{90–93}. Although tomaymycin has been known since over four decades and first biosynthesis studies were performed in 2009 a complete clarification of the biosynthetic pathway remained elusive⁹⁴. Therefore, the aim of this work was to elucidate the complete biosynthetic pathway of tomaymycin with respect to the timing of the single biosynthetic steps. To address this topic, a full *in vitro* reconstitution of the biosynthesis has been established alongside a reliable analytical method to visualize all chemical steps on an intact protein level.

The second part of this work is about the natural product tilivalline. The compound has been identified as the pathogenicity factor of the clinically relevant pathobiont *Klebsiella oxytoca*. The biosynthesis of the tilivalline core structure has been assumed to be identical to the tomaymycin and sibiromycin biosynthesis based on sequence data. The characteristic indole moiety of tilivalline could, however, not be explained based on the data⁹⁵. The aim of this work was to confirm the presence of tilivalline in a clinical isolate of *Klebsiella oxytoca* by activity guided isolation and to compare the mode of action to tomaymycin as well as the elucidation of the tilivalline biosynthesis and in particular the origin of the indole. Hence, the heterologous expression of the gene cluster and the *in vitro* reconstitution of the underlying NRPS has been realized. In addition, a reliable *in vitro* reconstitution system has become useful as a synthetic biology platform to generate derivatives of tilivalline in a convenient approach. In a similar fashion, this platform served as an assay to find potential pathoblocker by interfering with tilivalline's biosynthesis.

The final part covers an approach to discover so far unknown secondary metabolites from myxobacteria. The microbes of the order *myxococcales* are well known as promising source for bioactive secondary metabolites and sequence data suggest that only a fraction of the compounds has been discovered so far⁹⁶. Besides the continuous development in instrumental analytics, alternative approaches regarding cultivation should also be well suited to find so far unknown metabolites. Screening approaches in myxobacterial research are hitherto merely based on cultivation of vegetative cells in liquid media. This raised the question of how the metabolism will

be changed by passing a specific life cycle stage that cannot be achieved in suspension cultures. As myxobacteria are capable of fruiting body formation, the aim of the present thesis was to verify so far unknown metabolites that are linked to the occurrence of such fruiting bodies. For this purpose, a workflow has been established to identify such compounds and provide structure elucidation by utilizing various state-of-the-art techniques in instrumental analytics. The model organism *Myxococcus xanthus* DK 1622 is a fast growing member of the myxobacteria with reproducible fruiting body formation and as such served as the candidate organism to run this proof of concept study.

1.5 References

- [1] Dias, D. A., Urban, S. & Roessner, U.: A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2012**, 2, 303–336, DOI: 10.3390/metabo2020303
- [2] Gerabek, W. E., Haage, B. D., Keil, G. & Wegner, W.: *Enzyklopädie Medizingeschichte*. (2004).
- [3] Joachim, H.: *Papyrus Ebers: Das älteste Buch über Heilkunde*. (1890).
- [4] Ji, H.-F., Li, X.-J. & Zhang, H.-Y.: Natural products and drug discovery. Can thousands of years of ancient medical knowledge lead us to new and powerful drug combinations in the fight against cancer and dementia? *EMBO Rep.* **2009**, 10, 194–200, DOI: 10.1038/embor.2009.12
- [5] Aminov, R. I.: A brief history of the antibiotic era: Lessons learned and challenges for the future. *Front. Microbiol.* **2010**, 1, 134, DOI: 10.3389/fmicb.2010.00134
- [6] Carter, K. C.: Koch's Postulates in Relation To the Work of Jacob Henle and Edwin Klebs. *Pediatr. Infect. Dis. J.* **1986**, 5, 391, DOI: 10.1097/00006454-198605000-00041
- [7] Fleming, A.: on the Antibacterial Action of Cultures of a *Penicillium* , With Special Reference To Their Use in the Isolation of B . *Influenzae*. *Br. J. Exp. Pathol.* **1929**, 10, 226–236,
- [8] Newman, D. J. & Cragg, G. M.: Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **2016**, 79, 629–661, DOI: 10.1021/acs.jnatprod.5b01055
- [9] Clatworthy, A. E., Pierson, E. & Hung, D. T.: Targeting virulence: a new paradigm for antimicrobial therapy. *Nat. Chem. Biol.* **2007**, 3, 541–548, DOI: 10.1038/nchembio.2007.24
- [10] Senn, L., Calandra, T., Bille, J. & Marchetti, O.: Reply to Pasqualotto and Sukiennik. *Clin. Infect. Dis.* **2008**, 47, 293–294, DOI: 10.1086/589572
- [11] Walsh, C. T. & Wencewicz, T. a: Prospects for new antibiotics: a molecule-centered perspective. *J. Antibiot. (Tokyo)*. **2014**, 67, 7–22, DOI: 10.1038/ja.2013.49
- [12] Ventola, C. L.: The antibiotic resistance crisis: part 1: causes and threats. *P T* **2015**, 40, 277–283, DOI: Article
- [13] Bell, B. G., Schellevis, F., Stobberingh, E., Goossens, H. & Pringle, M.: A systematic review and meta-analysis of the effects of antibiotic consumption on antibiotic resistance. *BMC Infect. Dis.* **2014**, 14, 13, DOI: 10.1186/1471-2334-14-13
- [14] Davies, J. & Davies, D.: Origins and Evolution of Antibiotic Resistance. *Microbiol. Mol. Biol. Rev.* **2010**, 74, 417–433, DOI: 10.1128/MMBR.00016-10
- [15] Rice, L. B.: Federal Funding for the Study of Antimicrobial Resistance in Nosocomial Pathogens: No ESKAPE. *J. Infect. Dis.* **2008**, 197, 1079–1081, DOI: 10.1086/533452
- [16] Boucher, H. W. *et al.*: Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clin. Infect. Dis.* **2009**, 48, 1–12, DOI: 10.1086/595011
- [17] WHO | WHO publishes list of bacteria for which new antibiotics are urgently needed. *WHO* **2017**,
- [18] Crommelin, D. J. A., Sindelar, R. D. & Meibohm, B.: *Pharmaceutical biotechnology: Fundamentals and applications, Fourth edition*. *Pharm. Biotechnol. Fundam. Appl. Fourth Ed.* (Wiley-VCH Verlag GmbH & Co. KGaA, **2013**). DOI: 10.1007/978-1-4614-6486-0
- [19] Lipsky, M. S. & Sharp, L. K.: From idea to market: the drug approval process. *J. Am. Board Fam. Pract.* **2001**, 14, 362–7,
- [20] Thaxter, R.: On the Myxobacteriaceae, a New Order of Schizomycetes. *Bot. Gaz.* **1892**, 17, 389, DOI: 10.1086/326866

- [21] Rosenberg, E.: *Myxobacteria : Development and Cell Interactions*. (Springer New York, **1984**).
- [22] Nan, B. & Zusman, D. R.: Uncovering the Mystery of Gliding Motility in the Myxobacteria. *Annu. Rev. Genet.* **2011**, 45, 21–39, DOI: 10.1146/annurev-genet-110410-132547
- [23] Muñoz-Dorado, J., Marcos-Torres, F. J., García-Bravo, E., Moraleda-Muñoz, A. & Pérez, J.: Myxobacteria: Moving, Killing, Feeding, and Surviving Together. *Front. Microbiol.* **2016**, 7, 1–18, DOI: 10.3389/fmicb.2016.00781
- [24] Reichenbach, H.: Myxobacteria, producers of novel bioactive substances. *J. Ind. Microbiol. Biotechnol.* **2001**, 27, 149–156, DOI: 10.1038/sj.jim.7000025
- [25] Xiao, Y., Wei, X., Ebright, R. & Wall, D.: Antibiotic Production by Myxobacteria Plays a Role in Predation. *J. Bacteriol.* **2011**, 193, 4626–4633, DOI: 10.1128/JB.05052-11
- [26] McBride, M. J. & Zusman, D. R.: Behavioral analysis of single cells of *Myxococcus xanthus* in response to prey cells of *Escherichia coli*. *FEMS Microbiol. Lett.* **1996**, 137, 227–231, DOI: 10.1111/j.1574-6968.1996.tb08110.x
- [27] Garcia, R. & Müller, R.: *The Prokaryotes* (Rosenberg, E., DeLong, E. F., Lory, S., Stackebrandt, E. & Thompson, F.) (Springer Berlin Heidelberg, **2014**). 9783642390, 191–212, DOI: 10.1007/978-3-642-39044-9_303
- [28] Mauriello, E. M. F., Mignot, T., Yang, Z. & Zusman, D. R.: Gliding Motility Revisited: How Do the Myxobacteria Move without Flagella? *Microbiol. Mol. Biol. Rev.* **2010**, 74, 229–249, DOI: 10.1128/MMBR.00043-09
- [29] Goldman, B. S. *et al.*: Evolution of sensory complexity recorded in a myxobacterial genome. *Proc. Natl. Acad. Sci.* **2006**, 103, 15200–15205, DOI: 10.1073/pnas.0607335103
- [30] Weissman, K. J. & Müller, R.: Myxobacterial secondary metabolites: bioactivities and modes-of-action. *Nat. Prod. Rep.* **2010**, 27, 1276–95, DOI: 10.1039/c001260m
- [31] Schneiker, S. *et al.*: Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat. Biotechnol.* **2007**, 25, 1281–1289, DOI: 10.1038/nbt1354
- [32] Han, K. *et al.*: Extraordinary expansion of a *Sorangium cellulosum* genome from an alkaline milieu. *Sci. Rep.* **2013**, 3, 1157–1163, DOI: 10.1038/srep02101
- [33] Schneiker, S. *et al.*: Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat. Biotechnol.* **2007**, 25, 1281–1289, DOI: 10.1038/nbt1354
- [34] Ikeda, H. *et al.*: Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat. Biotechnol.* **2003**, 21, 526–531, DOI: 10.1038/nbt820
- [35] Oliynyk, M. *et al.*: Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. *Nat. Biotechnol.* **2007**, 25, 447–453, DOI: 10.1038/nbt1297
- [36] Staunton, J. & Weissman, K. J.: Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **2001**, 18, 380–416, DOI: 10.1039/a909079g
- [37] Marahiel, M. A., Stachelhaus, T. & Mootz, H. D.: Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem. Rev.* **1997**, 97, 2651–2674, DOI: 10.1021/cr960029e
- [38] McMurry, J. & Begley, T. P.: *The organic chemistry of biological pathways*. (Roberts and Co. Publishers, **2005**).
- [39] Okanya, P. W. *et al.*: Hyaladione, an S-methyl cyclohexadiene-dione from *Hyalangium minutum*. *J. Nat. Prod.* **2012**, 75, 768–70, DOI: 10.1021/np200776v

- [40] Gerth, K., Steinmetz, H., Höfle, G. & Jansen, R.: Chlorotonil A, a Macrolide with a Uniquegem-Dichloro-1,3-dione Functionality from *Sorangium cellulosum*, So ce1525. *Angew. Chemie Int. Ed.* **2008**, 47, 600–602, DOI: 10.1002/anie.200703993
- [41] Schummer, D. *et al.*: Antibiotics from Gliding Bacteria, LXXVI. Vioprolides: New Antifungal and Cytotoxic Peptolides from *Cystobacter violaceus*. *Liebigs Ann.* **2006**, 1996, 971–978, DOI: 10.1002/jlac.199619960617
- [42] Felder, S. *et al.*: Salimabromide: Unexpected Chemistry from the Obligate Marine Myxobacterium *Enhygromyxa salina*. *Chem. - A Eur. J.* **2013**, 19, 9319–9324, DOI: 10.1002/chem.201301379
- [43] Irschik, H., Gerth, K., Kemmer, T., Steinmetz, H. & Reichenbach, H.: The myxovalargins, new peptide antibiotics from *Myxococcus fulvus* (Myxobacterales). I. Cultivation, isolation, and some chemical and biological properties. *J. Antibiot. (Tokyo)*. **1983**, 36, 6–12, DOI: 10.7164/antibiotics.36.6
- [44] Höfle, G., Steinmetz, H., Gerth, K. & Reichenbach, H.: Antibiotics from gliding bacteria, XLIV. Ambruticins VS: New members of the antifungal ambruticin family from *Sorangium cellulosum*. *Liebigs Ann. der Chemie* **1991**, 1991, 941–945, DOI: 10.1002/jlac.1991199101161
- [45] GERTH, K., BEDORF, N., HÖFLE, G., IRSCHIK, H. & REICHENBACH, H.: Epothilons A and B: Antifungal and Cytotoxic Compounds from *Sorangium cellulosum* (Myxobacteria). Production, Physico-chemical and Biological Properties. *J. Antibiot. (Tokyo)*. **1996**, 49, 560–563, DOI: 10.7164/antibiotics.49.560
- [46] Bollag, D. M. *et al.*: Epothilones, a new class of microtubule-stabilizing agents with a taxol-like mechanism of action. *Cancer Res.* **1995**, 55, 2325–33,
- [47] Goodin, S., Kane, M. P. & Rubin, E. H.: Epothilones: Mechanism of Action and Biologic Activity. *J. Clin. Oncol.* **2004**, 22, 2015–2025, DOI: 10.1200/JCO.2004.12.001
- [48] Davies, J. & Ryan, K. S.: Introducing the Parvome: Bioactive Compounds in the Microbial World. *ACS Chem. Biol.* **2012**, 7, 252–259, DOI: 10.1021/cb200337h
- [49] Soukup, R. W. & Soukup, K.: (2015). 453–588, DOI: 10.1007/978-3-319-05275-5_5
- [50] Young, A. R. J., Narita, M. & Narita, M.: *Natural Products Isolation. Life Sci.* (Humana Press, **2012**). 864, DOI: 10.1007/978-1-61779-624-1
- [51] Katz, L. & Baltz, R. H.: Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **2016**, 43, 155–176, DOI: 10.1007/s10295-015-1723-5
- [52] Tswett, M.: 60. M. Tswett: Adsorptionsanalyse und chrematographische Methode. Anwendung auf die Chemie des Chlorophylls. *Plant Biol.* **1906**, 24, 384–393, DOI: 10.1111/J.1438-8677.1906.TB06534.X
- [53] Bernal, J. D. & Crowfoot, D.: X-Ray Photographs of Crystalline Pepsin. *Nature* **1934**, 133, 794–795, DOI: 10.1038/133794b0
- [54] Crowfoot, D., Bunn, C. W., Rogers-Low, B. W. & Turner-Jones, A.: *Chem. Penicillin* (Princeton University Press, **1949**). 310–366,
- [55] Fan, T. W.-M. & Lane, A. N.: Applications of NMR spectroscopy to systems biochemistry. *Prog. Nucl. Magn. Reson. Spectrosc.* **2016**, 92–93, 18–53, DOI: 10.1016/j.pnmrs.2016.01.005
- [56] Spek, A. L.: Structure validation in chemical crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2009**, 65, 148–155, DOI: 10.1107/S090744490804362X
- [57] McLafferty, F. W.: A century of progress in molecular mass spectrometry. *Annu. Rev. Anal. Chem. (Palo Alto. Calif.)*. **2011**, 4, 1–22, DOI: 10.1146/annurev-anchem-061010-114018

- [58] de Hoffmann, E. & Stroobant, V.: *Mass Spectrometry: Principles and Applications*. (Wiley, **2013**).
- [59] Cox, D. G., Oh, J., Keasling, A., Colson, K. L. & Hamann, M. T.: The utility of metabolomics in natural product and biomarker characterization. *Biochim. Biophys. Acta - Gen. Subj.* **2014**, 1840, 3460–3474, DOI: 10.1016/j.bbagen.2014.08.007
- [60] Swartz, M. E. & Swartz, M. E.: UPLCTM: An Introduction and Review. *J. Liq. Chromatogr. Relat. Technol.* **2005**, 28, 1253–1263, DOI: 10.1081/JLC-200053046
- [61] Kaufmann, A.: Combining UHPLC and high-resolution MS: A viable approach for the analysis of complex samples? *TrAC - Trends Anal. Chem.* **2014**, 63, 113–128, DOI: 10.1016/j.trac.2014.06.025
- [62] Krug, D. & Müller, R.: Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products. *Nat. Prod. Rep.* **2014**, 31, 768, DOI: 10.1039/c3np70127a
- [63] Krug, D., Zurek, G., Schneider, B., Garcia, R. & Müller, R.: Efficient mining of myxobacterial metabolite profiles enabled by liquid chromatography-electrospray ionisation-time-of-flight mass spectrometry and compound-based principal component analysis. *Anal. Chim. Acta* **2008**, 624, 97–106, DOI: 10.1016/j.aca.2008.06.036
- [64] Cortina, N. S., Krug, D., Plaza, A., Revermann, O. & Müller, R.: Myxoprincomide: A Natural Product from *Myxococcus xanthus* Discovered by Comprehensive Analysis of the Secondary Metabolome. *Angew. Chemie Int. Ed.* **2012**, 51, 811–816, DOI: 10.1002/anie.201106305
- [65] Yang, J. Y. *et al.*: Molecular Networking as a Dereplication Strategy. *J. Nat. Prod.* **2013**, 76, 1686–1699, DOI: 10.1021/np400413s
- [66] Watrous, J. *et al.*: Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci.* **2012**, 109, E1743–E1752, DOI: 10.1073/pnas.1203689109
- [67] Hou, Y. *et al.*: Microbial Strain Prioritization Using Metabolomics Tools for the Discovery of Natural Products. *Anal. Chem.* **2012**, 84, 4277–4283, DOI: 10.1021/ac202623g
- [68] Gromski, P. S. *et al.*: A tutorial review: Metabolomics and partial least squares-discriminant analysis – a marriage of convenience or a shotgun wedding. *Anal. Chim. Acta* **2015**, 879, 10–23, DOI: 10.1016/j.aca.2015.02.012
- [69] Leung, H. E.: *Integrative Proteomics*. (InTech, **2012**). DOI: 10.5772/2473
- [70] Fu, C. *et al.*: Solving the Puzzle of One-Carbon Loss in Ripostatin Biosynthesis. *Angew. Chemie Int. Ed.* **2017**, 56, 2192–2197, DOI: 10.1002/anie.201609950
- [71] Sahner, J. H. *et al.*: Advanced Mutasynthesis Studies on the Natural α -Pyrone Antibiotic Myxopyronin from *Myxococcus fulvus*. *ChemBioChem* **2015**, 16, 946–953, DOI: 10.1002/cbic.201402666
- [72] Burgard, C. *et al.*: Genomics-Guided Exploitation of Lipopeptide Diversity in Myxobacteria. *ACS Chem. Biol.* **2017**, 12, 779–786, DOI: 10.1021/acscchembio.6b00953
- [73] Ongley, S. E., Bian, X., Neilan, B. A. & Müller, R.: Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat. Prod. Rep.* **2013**, 30, 1121, DOI: 10.1039/c3np70034h
- [74] Meehan, M. J. *et al.*: FT-ICR-MS characterization of intermediates in the biosynthesis of the α -methylbutyrate side chain of lovastatin by the 277 kDa polyketide synthase LovF. *Biochemistry* **2011**, 50, 287–99, DOI: 10.1021/bi1014776
- [75] Meluzzi, D., Zheng, W. H., Hensler, M., Nizet, V. & Dorrestein, P. C.: Top-down mass spectrometry on low-resolution instruments: Characterization of phosphopantetheinylated

- carrier domains in polyketide and non-ribosomal biosynthetic pathways. *Bioorganic Med. Chem. Lett.* **2008**, 18, 3107–3111, DOI: 10.1016/j.bmcl.2007.10.104
- [76] Hicks, L. M. *et al.*: Investigating Nonribosomal Peptide and Polyketide Biosynthesis by Direct Detection of Intermediates on >70 kDa Polypeptides by Using Fourier-Transform Mass Spectrometry. *ChemBioChem* **2006**, 7, 904–907, DOI: 10.1002/cbic.200500416
- [77] Nesvizhskii, A. I.: Interpretation of Shotgun Proteomic Data: The Protein Inference Problem. *Mol. Cell. Proteomics* **2005**, 4, 1419–1440, DOI: 10.1074/mcp.R500012-MCP200
- [78] Chalkley, R. J. & Clauser, K. R.: Modification Site Localization Scoring: Strategies and Performance. *Mol. Cell. Proteomics* **2012**, 11, 3–14, DOI: 10.1074/mcp.R111.015305
- [79] Lim, H. *et al.*: Identification of 2D-gel proteins: A comparison of MALDI/TOF peptide mass mapping to μ LC-ESI tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2003**, 14, 957–970, DOI: 10.1016/S1044-0305(03)00144-2
- [80] Yang, Y. *et al.*: A comparison of nLC-ESI-MS/MS and nLC-MALDI-MS/MS for GeLC-based protein identification and iTRAQ-based shotgun quantitative proteomics. *J. Biomol. Tech.* **2007**, 18, 226–37,
- [81] Fenton, S. S. & Fahey, R. C.: Analysis of biological thiols: Determination of thiol components of disulfides and thioesters. *Anal. Biochem.* **1986**, 154, 34–42, DOI: 10.1016/0003-2697(86)90492-6
- [82] Snijder, J., Rose, R. J., Veessler, D., Johnson, J. E. & Heck, A. J. R.: Studying 18 MDa virus assemblies with native mass spectrometry. *Angew. Chemie - Int. Ed.* **2013**, 52, 4020–4023, DOI: 10.1002/anie.201210197
- [83] van de Waterbeemd, M. *et al.*: High-fidelity mass analysis unveils heterogeneity in intact ribosomal particles. *Nat. Methods* **2017**, 14, 283–286, DOI: 10.1038/nmeth.4147
- [84] Toby, T. K., Fornelli, L. & Kelleher, N. L.: Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu. Rev. Anal. Chem.* **2016**, 9, 499–519, DOI: 10.1146/annurev-anchem-071015-041550
- [85] Gillet, L. C., Leitner, A. & Aebersold, R.: Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annu. Rev. Anal. Chem.* **2016**, 9, 449–472, DOI: 10.1146/annurev-anchem-071015-041535
- [86] Arima, K., Kosaka, M., Tamura, G., Imanaka, H. & Sakai, H.: Studies on tomaymycin, a new antibiotic. I. Isolation and properties of tomaymycin. *J. Antibiot. (Tokyo)*. **1972**, 25, 437–44, DOI: 10.7164/antibiotics.25.437
- [87] Nishioka, Y., Beppu, T., Kosaka, M. & Arima, K.: Mode of action of tomaymycin. *J. Antibiot. (Tokyo)*. **1972**, 25, 660–667,
- [88] Puvvada, M. S. *et al.*: Inhibition of bacteriophage T7 RNA polymerase in vitro transcription by DNA-binding pyrrolo[2,1-c][1,4]benzodiazepines. *Biochemistry* **1997**, 36, 2478–2484, DOI: 10.1021/bi952490r
- [89] Hurley, L. H. *et al.*: Pyrrolo[1,4]benzodiazepine antitumor antibiotics: relationship of DNA alkylation and sequence specificity to the biological activity of natural and synthetic compounds. *Chem Res Toxicol* **1988**, 1, 258–268, DOI: 10.1021/tx00005a002
- [90] Gerratana, B.: Biosynthesis, synthesis, and biological activities of pyrrolobenzodiazepines. *Med. Res. Rev.* **2012**, 32, 254–93, DOI: 10.1002/med.20212
- [91] Thurston, D. E. *et al.*: Effect of a-ring modifications on the DNA-binding behavior and cytotoxicity of pyrrolo[2,1-c][1,4]benzodiazepines. *J. Med. Chem.* **1999**, 42, 1951–1964, DOI: 10.1021/jm981117p

-
- [92] Gregson, S. J. *et al.*: Effect of C2-exo unsaturation on the cytotoxicity and DNA-binding reactivity of pyrrolo[2,1-c][1,4]benzodiazepines. *Bioorganic Med. Chem. Lett.* **2000**, 10, 1845–1847, DOI: 10.1016/S0960-894X(00)00351-6
- [93] Kumar, R. & Lown, J.: Recent Developments in Novel Pyrrolo[2,1-c][1,4]Benzodiazepine Conjugates: Synthesis and Biological Evaluation. *Mini-Reviews Med. Chem.* **2003**, 3, 323–339, DOI: 10.2174/1389557033488097
- [94] Li, W., Chou, S., Khullar, A. & Gerratana, B.: Cloning and characterization of the biosynthetic gene cluster for tomaymycin, an sjg-136 monomeric analog. *Appl. Environ. Microbiol.* **2009**, 75, 2958–2963, DOI: 10.1128/AEM.02325-08
- [95] Schneditz, G. *et al.*: Enterotoxigenicity of a nonribosomal peptide causes antibiotic-associated colitis. *Proc. Natl. Acad. Sci.* **2014**, 111, 13181–13186, DOI: 10.1073/pnas.1403274111
- [96] Wenzel, S. C. & Müller, R.: *Compr. Nat. Prod. II* (Elsevier, **2010**). 2, 189–222, DOI: 10.1016/B978-008045382-8.00645-6

Chapter 2 – Tomaymycin

Total biosynthesis of tomaymycin comprehensively monitored on the nonribosomal peptide megasynthetase

Alexander von Tesmar^{1,2,3}, **Michael Hoffmann**^{1,2,3}, Jan Pippel⁵, Antoine Abou Fayad^{2,3},
Stefan Werner², Armin Bauer⁷, Wulf Blankenfeldt^{5,6}, Rolf Müller^{2,3}

Submitted to *CELL Chemical Biology* (05.2017)

¹These authors contributed equally to this work

²Department of Microbial Natural Products (MINS), Helmholtz Institute for Pharmaceutical Research Saarland (HIPS) - Helmholtz Centre for Infection Research (HZI) and Institute for Pharmaceutical Biotechnology, Saarland University, 66123 Saarbrücken, Germany

³German Center for Infection Research (DZIF), Partner site Hannover-Braunschweig, 38124 Braunschweig, Germany

⁵Structure and Function of Proteins, Helmholtz Centre for Infection Research, Inhoffenstr. 7, 38124 Braunschweig, Germany

⁶Institute for Biochemistry, Biotechnology and Bioinformatics, Technische Universität Braunschweig, Spielmannstr. 7, 38106 Braunschweig, Germany

⁷Sanofi-Aventis Deutschland GmbH, R&D Therapeutic Area Infectious Diseases, Industriepark Höchst G878, 65926 Frankfurt am Main, Germany

Contributions

Author's effort:

The author significantly contributed to the conception of the study, designed and performed experiments, evaluated and interpreted resulting data. The author 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 the author. Furthermore, the author contributed equally to conceiving and writing of the manuscript.

Contribution by Alexander von Tesmar:

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.

Contribution by others:

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.

2 Tomaymycin

2.1 Abstract

In vitro reconstitution and biochemical analysis of natural product biosynthetic pathways remains a challenging endeavor, especially if megaenzymes of the nonribosomal peptide synthetase (NRPS) type are involved. In theory, all biosynthetic steps may be deciphered using MS 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. Proteoforms were analyzed by LC-MS to 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 nonribosomally synthesized natural products. 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.

2.2 Introduction

Natural products account for a large part of the known therapeutics and thus remain to be a major topic of drug discovery¹. 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 patterns³. 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 polymerase 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 revealed common genes encoding a bi-modular nonribosomal peptide synthetase (NRPS) as the basis of the biosynthetic pipeline for PBD production^{9–11}.

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'-phosphopanthetein 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 epimerization¹². 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}, pacidamycin²⁰, antimycin²¹, beauvericin²² and the siderophores enterobactin²³, yersinibactin²⁴, pyochelin²⁵, vibriobactin²⁶, myxochelin^{27,28} and pseudomonine¹⁵. 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 form²⁹. 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 analytics³⁰. 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.

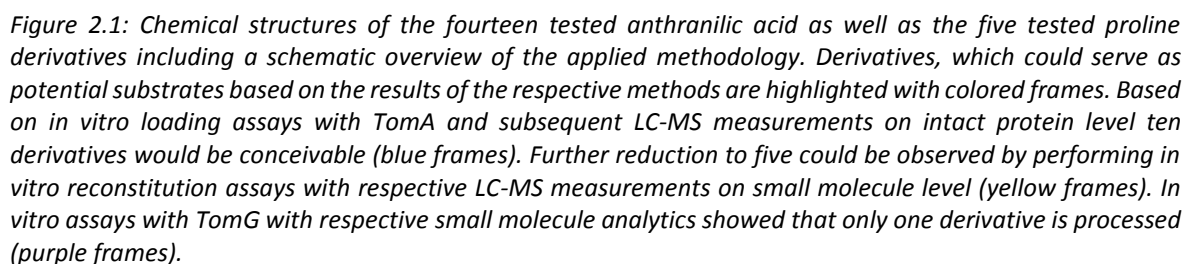
2.3 Results and Discussion

2.3.1 *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 structure³⁴. 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 MtaA²⁷. 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 system³⁵. 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 2.3). 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 2.2). 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 2.4). 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 detected³⁴. Further, incorporation 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 process 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 β -lactam rings³⁸ or the recruitment of *in trans* acting tailoring enzymes³⁹. 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 NRPS⁴².

2.3.2 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^{43–48}. In many of these studies, the PPant-ejection assay⁴⁹ was used. Its applicability 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^{50–52}. Nevertheless, this approach led to the successful identification of several intermediates on the fungal iterative polyketide synthases responsible for the biosynthesis of aflatoxin B1⁵³. 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 proteoform⁵⁴ or the visualization of inhibitor binding⁵⁵. 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 quality⁵⁶. 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 2.1). 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 & Figure 2.2). This again indicates that downstream domains restrict the tolerance of tomaymycin biosynthesis towards carrier protein loaded anthranilic acid derivatives.

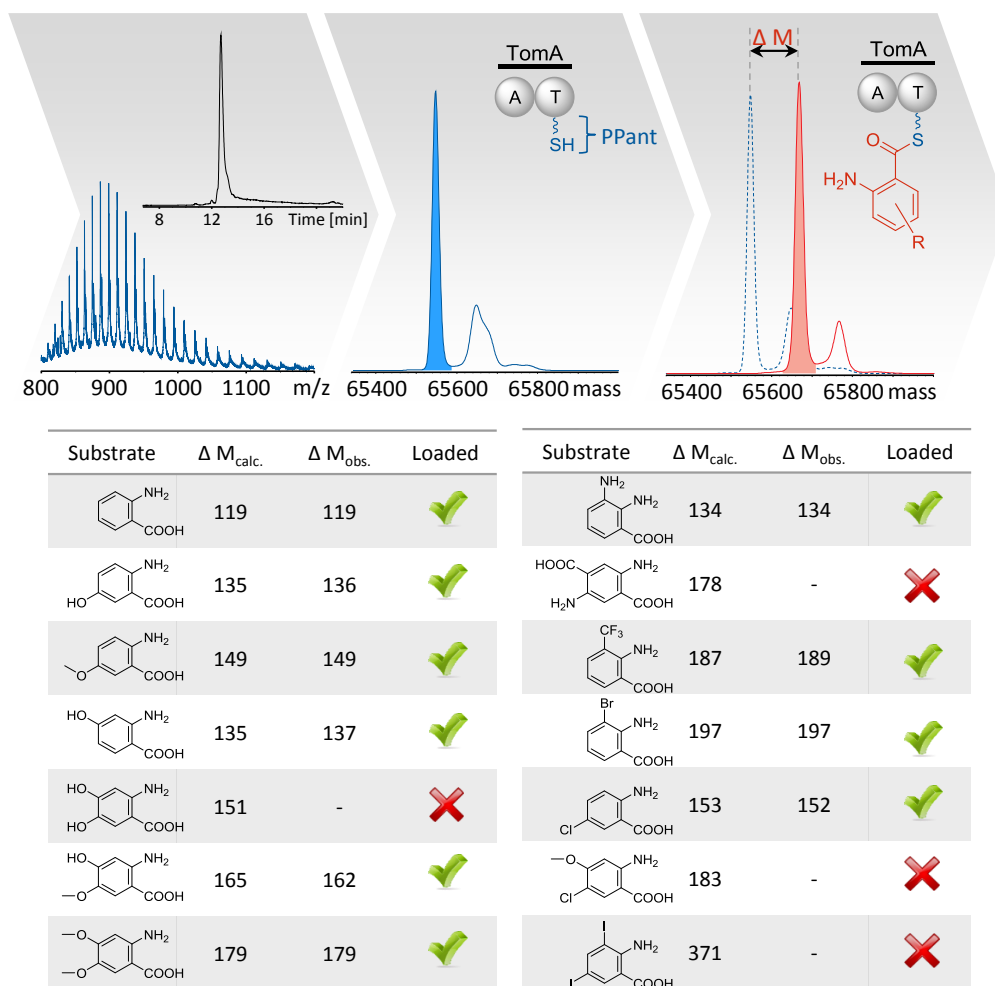


Figure 2.2: Scheme showing the workflow of TomA loading assays including the results of loading attempts with different, available anthranilic acid derivatives: The top panel shows exemplary LC-MS results obtained for loading assays of TomA. Charge state distributions were typically found to be in the 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.

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.

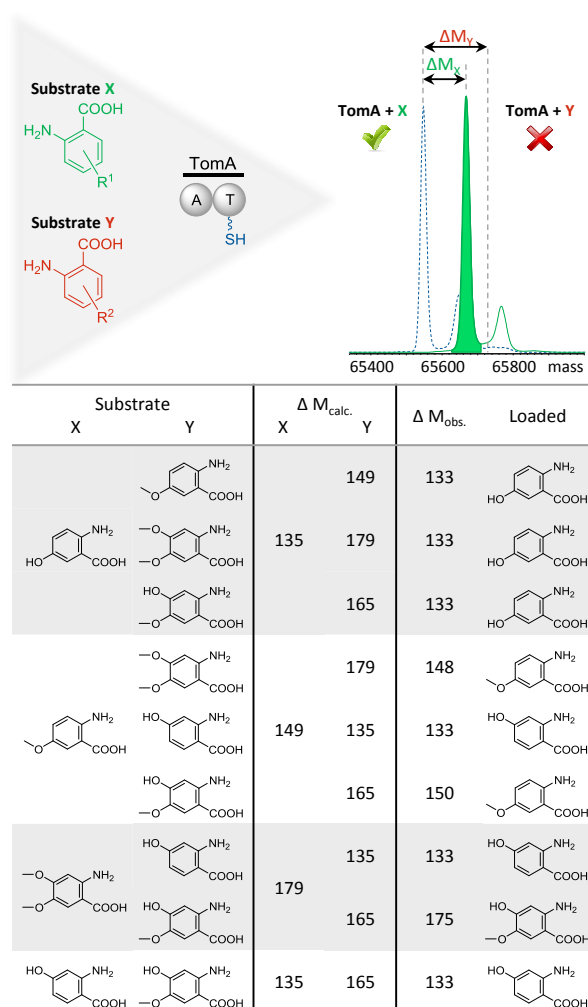


Figure 2.3: Approaches towards the analysis of substrate specificity of TomA by competitive loading assays with pairs of anthranilic acid derivatives in equimolar ratio. Detected delta masses are related to holo TomA and showed which substrate is preferred.

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 S 2.4). 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.

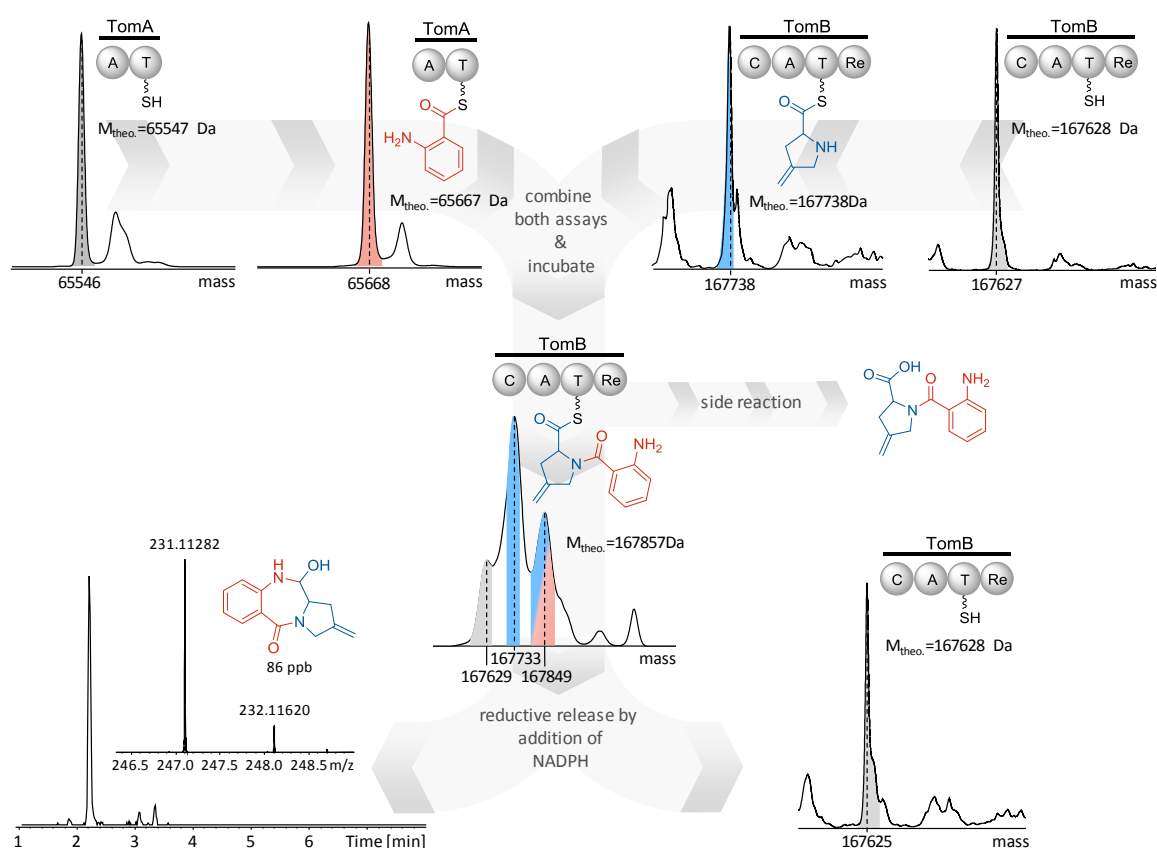


Figure 2.4: Composited steps of the tomaymycin *in vitro* reconstitution including the respective deconvoluted mass spectra for each protein species to confirm the substrate binding and thus monitoring the complete biosynthesis. The observed mass deviation in case of the dipeptide-loaded TomB is attributable 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.

2.3.3 *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 S_N2 reaction. This mechanism involves an Mg^{2+} ion which positions the respective target group by coordinating the hydroxyl group(s) of the substrate^{57–60}. In both chains of the TomG crystal structure, additional electron density corresponding to SAM could be identified whereas no Mg^{2+} ion was detected in the active site.

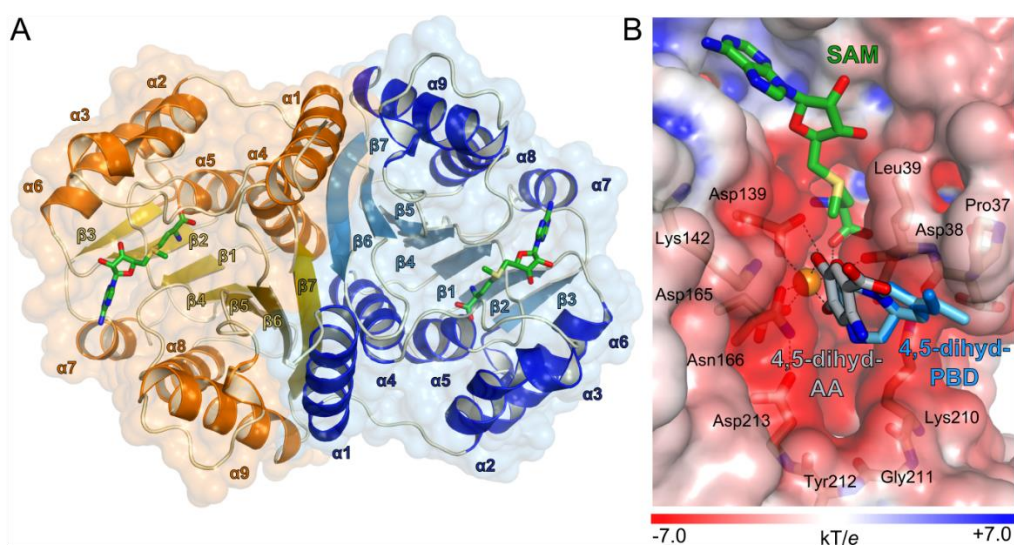


Figure 2.5: **A** Overview of the dimeric crystal structure of TomG (monomer A: yellow and orange; monomer B: blue) with the bound cofactor *S*-adenosyl methionine (SAM) (stick presentation in green); **B** Proposed binding modes of 4,5-dihydroxy anthranilic acid (4,5-dihyd-AA; in light grey) and 4,5-dihydroxy-PBD (4,5-dihyd-PBD; two binding modes in stick presentation in light and dark blue) to the TomG monomer (Mg^{2+} 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 webserver⁶¹. Electrostatic potential was calculated using the APBS 2.1 plugin⁶² under PYMOL and the PDB2PQR webserver (http://nbc-222.ucsd.edu/pdb2pqr_2.0.0/; Dolinsky et al.⁶³).

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 2.1), likely because the missing second hydroxy group in *ortho*-position is essential for the reaction as also described for OMTs in the catechol metabolism^{64,65}. 4-hydroxylation of the tomaymycin anthranilic acid moiety is most likely accomplished by TomE and TomF as already supposed by Li et al.³⁴ and the high homology of these proteins to the 4-nitrophenol oxidation system from *Rhodococcus* sp.⁶⁶ 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 biosynthesis⁶⁷. 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, pdb code: 3duw, Z-score = 34.2, sequence identity = 48 %, Cho et al.⁶⁸, Figure S 2.6). BcOMT2 shows promiscuous substrate tolerance for flavonoids^{68,69}, 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 2.5B). 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: 3bwm⁷⁰, 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^{9,65,71}. 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 (Figure 2.6). 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 2.6), which may explain why TomG does not methylate PBDs *in vitro*.

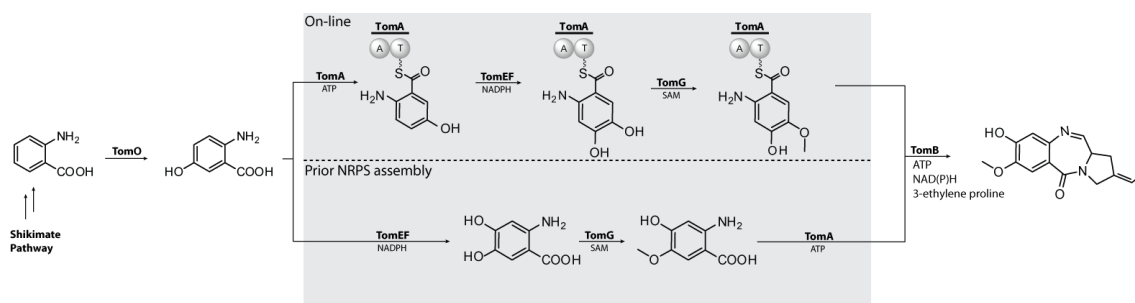


Figure 2.6: Tomaymycin biosynthesis comprising hydroxylation of 5-OH anthranilic acid by TomE and TomF, methylation by TomG and subsequent PBD assembly by TomAB. According to our data tailoring of anthranilic acid likely takes place either on-line while tethered to TomA or prior NRPS assembly.

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 (Figure 2.6). 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*.

2.4 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 assembly⁹. 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.

2.5 Methods

2.5.1 Protein Preparation Procedure

2.5.1.1 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 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. ⁷²)).

2.5.1.2 Expression and Purification of TomG for biochemical experiments

The respective gene was inserted into pET28b with an N-terminal 6xHis tag using according primers (see Key Source Table). 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.

2.5.1.3 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 (see Key Source Table), 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) (2xSOC-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 ÄKTA 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.

2.5.2 Activity Assays

2.5.2.1 Hydroxamate-MesG Assay

Adenylation activity of TomA A-domain was tested using the hydroxamate-MesG assay as described previously (Wilson and Aldrich, 2010). Assays in 100 µL 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.

2.5.2.2 *In vitro* Reconstitution Assay of TomA and TomB

TomA and TomB (0.5 µM 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 µL 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 µL water:acetonitrile (1:1), centrifuged and submitted to LC-MS analysis.

2.5.2.3 *In vitro* Reconstitution Assay of TomG

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

2.5.3 MS-based Analysis

2.5.3.1 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 μm dp column (Waters, Germany). Separation of 5 μL sample was achieved by a linear gradient from (A) H_2O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 600 $\mu\text{L}/\text{min}$ and 45 $^\circ\text{C}$. 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 $\mu\text{L}/\text{min}$ 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 $^\circ\text{C}$, 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.

2.5.3.2 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 μm dp column. Separation of 1 μL sample was achieved by a linear gradient with (A) H_2O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 600 $\mu\text{L}/\text{min}$ and 45 $^\circ\text{C}$. 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 $\mu\text{L}/\text{min}$ 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 $^\circ\text{C}$, 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.

2.5.3.3 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.).

2.5.3.4 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.

2.5.4 Protein Structure Determination and Modeling

2.5.4.1 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, Burkhardt et al.⁷³). The diffraction data were indexed, integrated and scaled with XDS⁷⁴ and AIMLESS⁷⁵ of the CCP4 suite⁷⁶.

The structure of TomG was solved by molecular replacement with PHASER⁷⁷ by using the coordinates of residues 44-220 of a methyltransferase from *Cyanobacterium Synechocystis* (pdb entry: 3CBG; Kopycki et al.⁷⁸ as search model (32 % sequence identity). The model was further improved by iterative cycles of translation liberation screw-motion (TLS; Painter and Merritt⁷⁹ refinement using PHENIX.REFINE⁸⁰ and manual rebuilding with COOT⁸¹. Final structure building and interpretation was aided by MOLPROBITY⁸² as well as by feature enhanced maps calculated with PHENIX^{80,83}. 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 proteins⁸⁴. Figures were prepared using PyMOL (The PyMOL Molecular Graphics System, Version 1.8.2.3 Schrödinger, LLC).

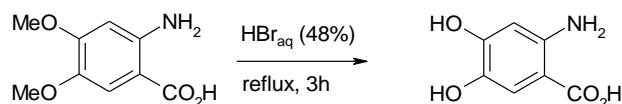
2.5.4.2 *In-silico* Modelling

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; pdb code: 3bwm, ref: Rutherford et al.⁷⁰ using only the metal-coordinating residues with the program SUPERPOSE under CCP4⁷⁶. 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, Mg^{2+} and the hydroxyl oxygens of the ligands were kept fixed to ensure a productive binding mode.

2.5.5 Synthesis

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

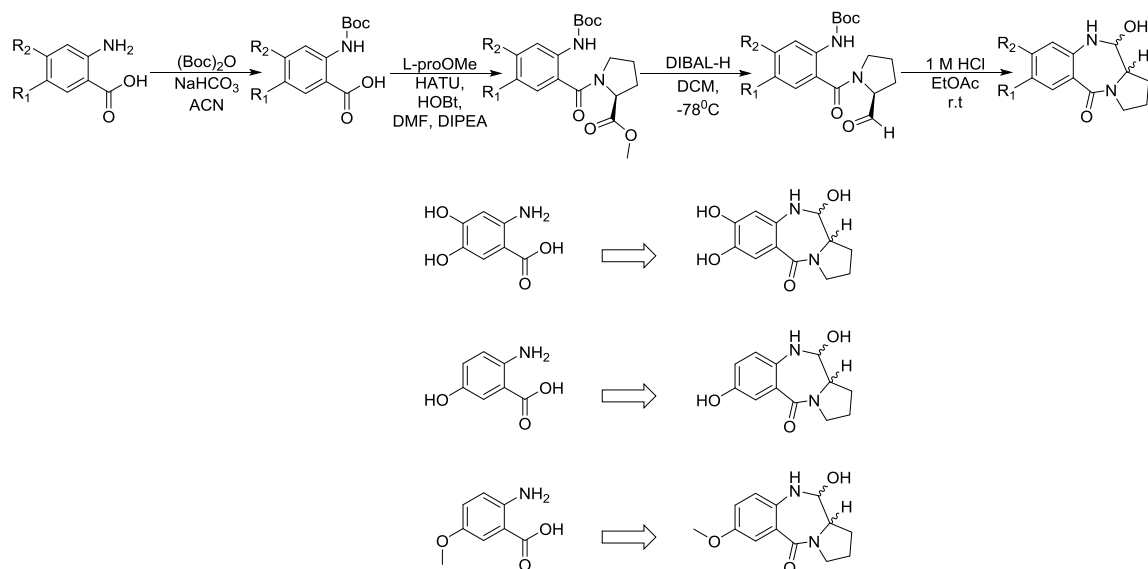


2-Amino-4,5-dihydroxy-benzoic acid was prepared by ether cleavage from commercially available 2-amino-4,5-dimethoxy-benzoic acid under acidic conditions. A suspension of 2-amino-4,5-dimethoxy-benzoic acid (5.0 g, 24.85 mmol) in aqueous hydro-bromic acid (48%, 127mL, 1.12 mol) was heated under reflux 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 NaHCO_3 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.

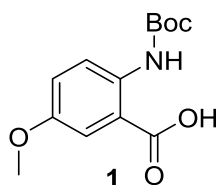
^1H NMR (600MHz, DMSO-d_6) δ 9.39 (s, 1H), 8.20 (s, 1H), 7.96 (br s, 3H), 7.06 (s, 1H), 6.12 (s, 1H); ^{13}C -NMR (150MHz, DMSO-d_6): δ 169.08, 152.05, 146.87, 135.48, 116.28, 102.09, 100.72; MS (ESI-) $\text{C}_7\text{H}_6\text{NO}_4$ $[\text{M-H}]^-$ calculated for 168.0, found 168.0.

2.5.5.2 Synthesis of the tomaymycin derivatives

Scheme S 2.1: Synthesis of the tomaymycin derivatives

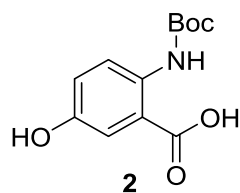


- 2-*N*-boc-amino-5-methoxybenzoic acid (**1**):



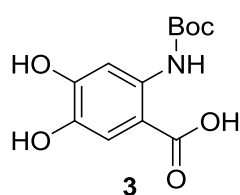
To a stirred solution of 2-amino-5-methoxybenzoic acid (50 mg, 0.3 mmol) in acetonitrile (20 ml) at room temperature, a mixed solution of di-*tert*-butyl dicarbonate (98 mg, 0.45 mmol), saturated solution of sodium bicarbonate (NaHCO_3) (19.6 ml) and acetonitrile (10 ml) was added. Following addition, 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 :: 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) $\text{C}_{13}\text{H}_{17}\text{NO}_5$ $[\text{M}+\text{H}]^+$ calculated for 268.1179, found 268.1168.

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



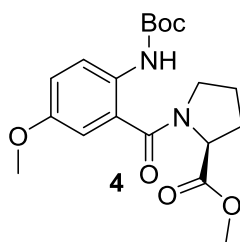
2-*N*-boc-amino-5-hydroxy benzoic acid (**2**) was prepared following the same protocol as for compound (**1**) and was used without any further purification. HRMS (ESI, +ve) $C_{12}H_{15}NO_5$ $[M+H]^+$ calculated for 254.1023, found 254.1027.

- 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) $C_{12}H_{15}NO_6$ $[M+H]^+$ calculated for 270.0972, found 270.0967.

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

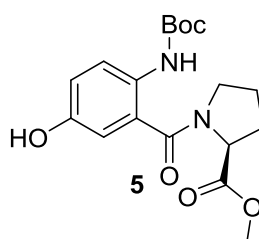


2-*N*-boc-amino-5-methoxy benzoic acid (**1**) (50 mg, 0.18 mmol), HATU (68.4 mg, 0.18 mmol) and HOBt (5 mg, 0.03 mmol) was dissolved in anhydrous DMF (1 ml) under nitrogen atmosphere. DIPEA (61 μ L, 0.36 mmol) was added to the mixture and the solution was left stirring at room temperature for 5 minutes 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, CDCl_3) δ 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),; ^{13}C NMR (126 MHz, CDCl_3) δ 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) $\text{C}_{19}\text{H}_{26}\text{N}_2\text{O}_6$ $[\text{M}+\text{H}]^+$ calculated for 379.1864, found 379.1860.

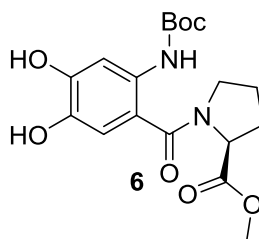
- 2-*N*-*boc*-amino-5-hydroxybenzoyl- L-proline methyl ester (**5**):



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, CDCl_3) δ 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),; ^{13}C NMR (126 MHz, CDCl_3) δ 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) $\text{C}_{18}\text{H}_{24}\text{N}_2\text{O}_6$ $[\text{M}+\text{H}]^+$ calculated for 365.1707, found 365.1699.

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

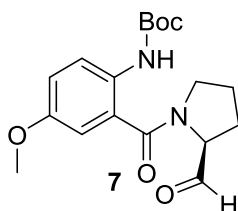


2-*N*-*boc*-amino-4,5-dihydroxybenzoyl- L-proline methyl ester (**6**) was prepared following the same protocol as for compound (**4**).

^1H NMR (500 MHz, CDCl_3) δ 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) δ 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) $C_{18}H_{24}N_2O_7$ $[M+H]^+$ calculated for 381.1656, found 381.1652.

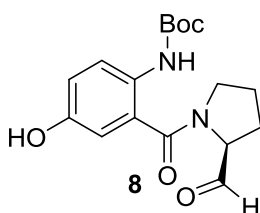
- S-1- (2'-N-boc-amino-5'-methoxybenzoyl)pyrrolidine-2-carbaldehyde (**7**):



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) δ 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) δ 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 $C_{18}H_{24}N_2O_5$ m/z $[M+H]^+$: 349.1758; found: 349.1746.

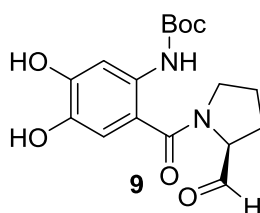
- S-1- (2'-N-boc-amino-5'-hydroxybenzoyl)pyrrolidine-2-carbaldehyde (**8**):



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, CDCl_3) δ 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); ^{13}C NMR (126 MHz, CDCl_3) δ 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) $\text{C}_{17}\text{H}_{22}\text{N}_2\text{O}_5$ $[\text{M}+\text{H}]^+$ calculated for 335.1605, found 335.1609.

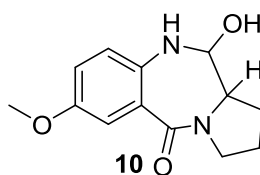
- S-1- (2'-*N*-boc-amino-4',5'-dihydroxybenzoyl)pyrrolidine-2-carbaldehyde (**9**):



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, CDCl_3) δ 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); ^{13}C NMR (126 MHz, CDCl_3) δ 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) $\text{C}_{17}\text{H}_{22}\text{N}_2\text{O}_6$ $[\text{M}+\text{H}]^+$ calculated for 351.1551, found 351.1546.

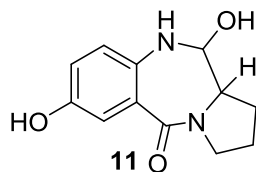
- Carbinolamine from 5-methoxy anthranilic acid (**10**):



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, CD_3OD) δ 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); ^{13}C NMR (126 MHz, CD_3OD) δ 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 $\text{C}_{13}\text{H}_{16}\text{N}_2\text{O}_3$ m/z $[\text{M}+\text{H}]^+$: 249.1234; found: 249.1239.

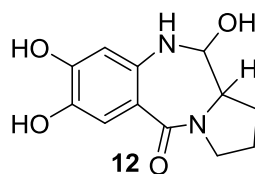
- Carbinolamine from 5-hydroxy anthranilic acid (**11**):



Carbinolamin X (**11**) was prepared following the same protocol as for compound (**10**).

^1H NMR (500 MHz, CD_3OD) δ 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); ^{13}C NMR (126 MHz, CD_3OD) δ 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 $\text{C}_{12}\text{H}_{14}\text{N}_2\text{O}_2$ m/z $[\text{M}+\text{H}]^+$: 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, CD_3OD) δ 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); ^{13}C NMR (126 MHz, CD_3OD) δ 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 $\text{C}_{12}\text{H}_{14}\text{N}_2\text{O}_4$ m/z $[\text{M}+\text{H}]^+$: 251.1026; found: 251.1018.

2.5.6 Data and Software Availability

The software used in this study is listed in the Key Source Table. Coordinates of the TomG-SAM co-crystal structure are available at the PDB (<http://rcsb.org/>; Berman et al.⁸⁵ under accession code 5N5D.

2.5.7 Key Source Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
Escherichia coli BL21 (DE3)	Agilent Technologies	200131
Escherichia coli Dh10b	Thermo Scientific	18297010
Chemicals, Peptides, and Recombinant Proteins		
TomA	This study	N/A
TomB	This study	N/A
TomA A-domain	This study	N/A
TomG	This study	N/A
TomG-crystal	This study	N/A
HRV3C Protease	Thermo Scientific	88946
Kanamycin	Sigma	60615
Chloramphenicol	Sigma	C0378
Ni-NTA Agarose	Qiagen	30210
IPTG	Sigma	10724815
TCEP	Sigma	C4706
ATP	Thermo	R0441
NADPH	ACROS	AC328740010
Purine nucleoside phosphorylase	Sigma	53113
Inorganic pyrophosphatase	Thermo	EF0221
MesG	Sigma	
Anthranilic acid	Sigma	A89855
5-hydroxy anthranilic acid	Sanofi	N/A
5-methoxy anthranilic acid	Sanofi	N/A
4-hydroxy-5-methoxy anthranilic acid	Sanofi	N/A
4-hydroxy anthranilic acid	Sanofi	N/A
3-bromo anthranilic acid	Sanofi	N/A
4,5-dimethoxy anthranilic acid	Sanofi	N/A
5-chloro anthranilic acid	Sanofi	N/A
3-CF ₃ anthranilic acid	Sanofi	N/A
3-amino anthranilic acid	Sanofi	N/A
4-methoxy-5-chloro anthranilic acid	Sanofi	N/A
4,5-dihydroxy anthranilic acid	This study	See Synthesis
3,5-diiodo anthranilic acid	Sanofi	N/A
4-carboxy-5-amino anthranilic acid	Sanofi	N/A
proline	Sigma	
3-methyl proline	Sanofi	N/A
3-methylene BOC-proline	Sanofi	N/A
2,3-dehydro proline	Sanofi	N/A
3-difluoro proline	Sanofi	N/A
(2R,3R)-(-)-2,3-Butanediol	Sigma-Aldrich	Cat#237639
S-adenosyl methionine (SAM) chloride dihydrochloride	Sigma-Aldrich	Cat#A7007
See Chemical Synthesis for additional compounds	This study	N/A
Deposited Data		
TomG-SAM co-crystal structure	This paper	PDB: 5N5D

Experimental Models: Organisms/Strains		
gDNA <i>Streptomyces</i> sp. FH6421	Lattemann et al. ⁸⁶	N/A
Oligonucleotides		
TomA-FP AAAAAAGCTTCTAGTGACCTTTCTTCGGATCGC G	Sigma	N/A
TomA-RP AAAAAAGCTTCTACTGGGTCTGTGCGGCG	Sigma	N/A
TomA-RP AAAAAAGCTTCTACTGGGTCTGTGCGGCG	Sigma	N/A
TomB-FP AAAAACATGTTAATGAACTCCCCCTCCGAAC	Sigma	N/A
TomB-RP AAAAAAGCTTTCTAGTCGTGCGCCAGTTCTCC	Sigma	N/A
TomG-FP AAAAAACATATGATGGAACAAGAGCGATGGAA	Sigma	N/A
TomG-RP AAAAAAGGATCCCTAGCCGGTGACCAG	Sigma	N/A
TomGcrystal-FP CGCCATATGATGGAACAAGAGCG	Sigma	N/A
TomGcrystal-RP attctcgagCTAGCCGGTGACC	Sigma	N/A
Recombinant DNA		
pET28b	Novagen	69865
pETM44	Dummler et al. ⁸⁷	N/A
pSU-MtaA	Gaitatzis et al. ⁸⁸	N/A
pET-19b-6xHis-T7-Lysozyme-TomG	This paper	N/A
pET28b-TomG	This paper	N/A
pETM44-TomA	This paper	N/A
pETM44-TomA-A	This paper	N/A
pETM44-TomB	This paper	N/A
Software and Algorithms		
OriginPro 2016	OriginLab Corporation	N/A
ChemDraw Professional 15.1	Perkin Elmer	N/A
DATAANALYSIS 4.4	Bruker	N/A
Maximum Entropy Deconvolution Algorithm	Bruker	N/A
XDS	Kabsch ⁷⁴	http://xds.mpimf-heidelberg.mpg.de/
AIMLESS	Evans and Murshudov ⁷⁵	http://www.ccp4.ac.uk/html/aimless.html
CCP4	⁷⁶	http://www.ccp4.ac.uk/
COOT	⁸¹	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
ESPRIT 3.0 webserver	⁸⁹	http://espript.ibcp.fr/ESPrpt/ESPrpt/
MOE	MOE, 2013.08; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2017	http://www.chemcomp.com/MOE-Molecular_Operating_Environment.htm
PHASER	⁷⁷	http://phenix-online.org
PHENIX.REFINE	⁸⁰	http://phenix-online.org
PHENIX	^{83,90}	http://phenix-online.org

PYMOL	The PyMOL Molecular Graphics System, Version 1.8.2.3 Schrödinger, LLC	https://www.pymol.org/
DALI webserver	84	http://ekhidna.biocenter.helsinki.fi/dali_server/start
MOLPROBITY webserver	82	http://molprobity.biochem.duke.edu/
Other		
Pur-A-Lyzer™ Maxi Dialysis Kit	Sigma-Aldrich	PURX60100-1KT
HisTrap HP 5 mL column	GE Healthcare Life Sciences	Cat#17-5248-02
HiLoad 16/600 Superdex 75 pg column	GE Healthcare Life Sciences	Cat#28-9893-33
Vivaspin 6 MWCO 10 000	GE Healthcare Life Sciences	Cat#28-9322-96

2.6 Supplemental Figures

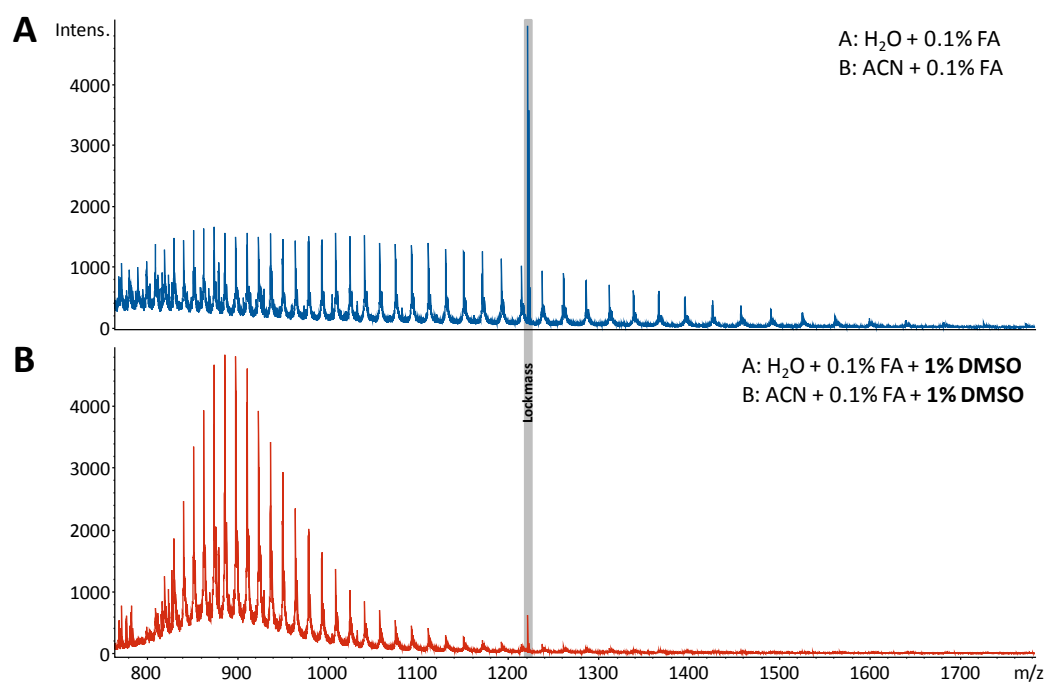


Figure S 2.1: Comparative mass spectra of TomA show the influence on spectra quality owing to the addition of DMSO (B) to the standard LC eluents (A). Reduction of charge state signals leads to an intensity increase for the remaining signals by a factor of approx. 3.

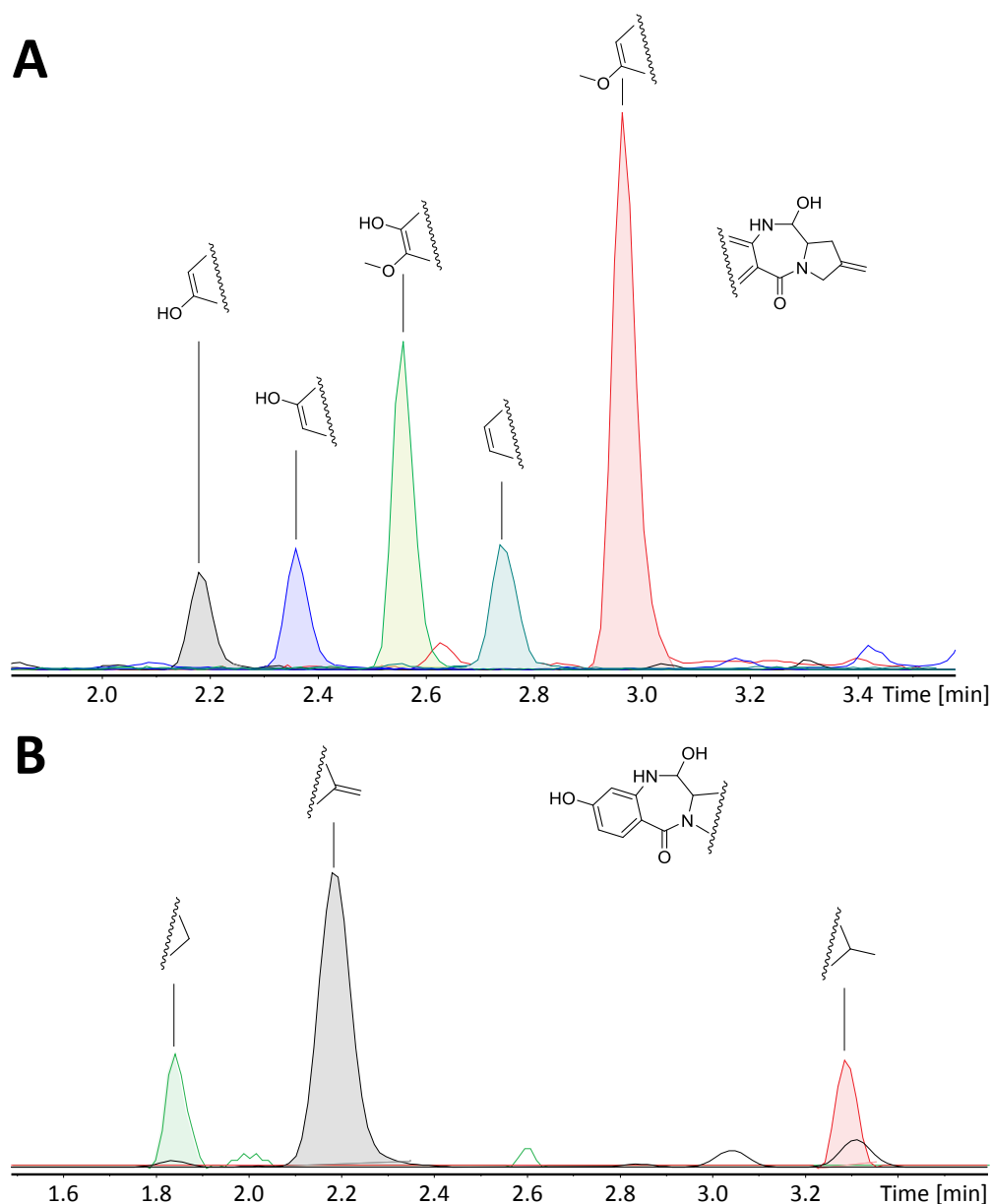


Figure S 2.2, related to Figure 2.1: **A** LC-MS data of tomaymycin derivatives derived from *in vitro* reconstitution of TomA and TomB with different anthranilic acid derivatives combined with methylene proline. Overlaid EIC traces of the respective derivatives with a m/z width of 0.002. **B** LC-MS 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.

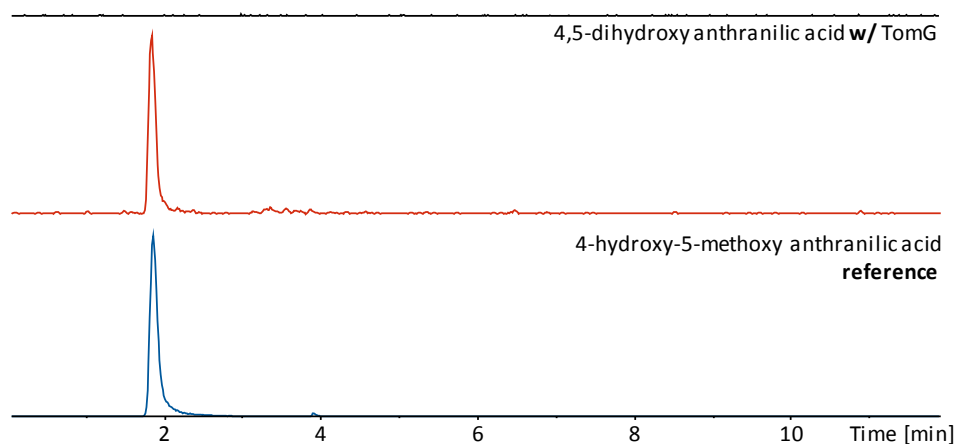
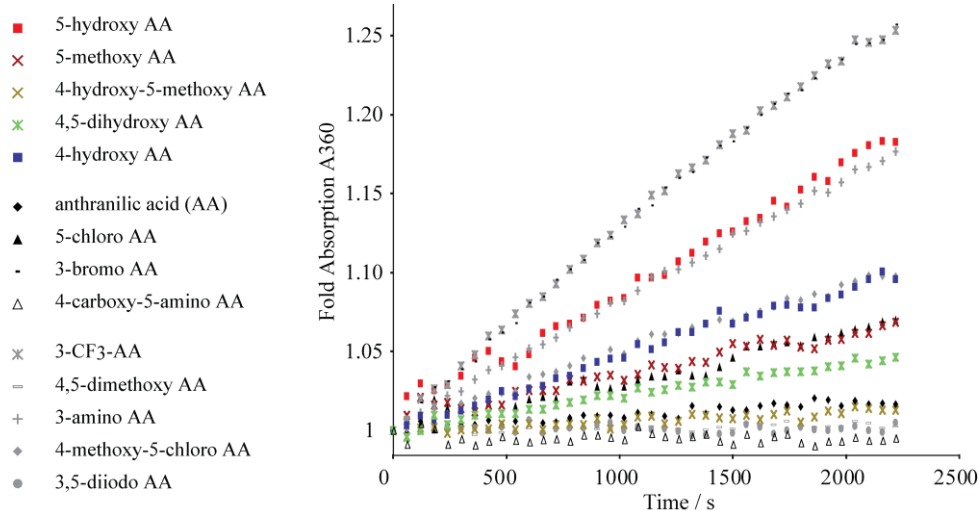
A4,5-dihydroxy anthranilic acid **w/o** TomG**B**

Figure S 2.3, related to Figure 2.1: **A** EIC traces of 4-hydroxy-5-methoxy anthranilic acid (m/z 184.06043 \pm 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.

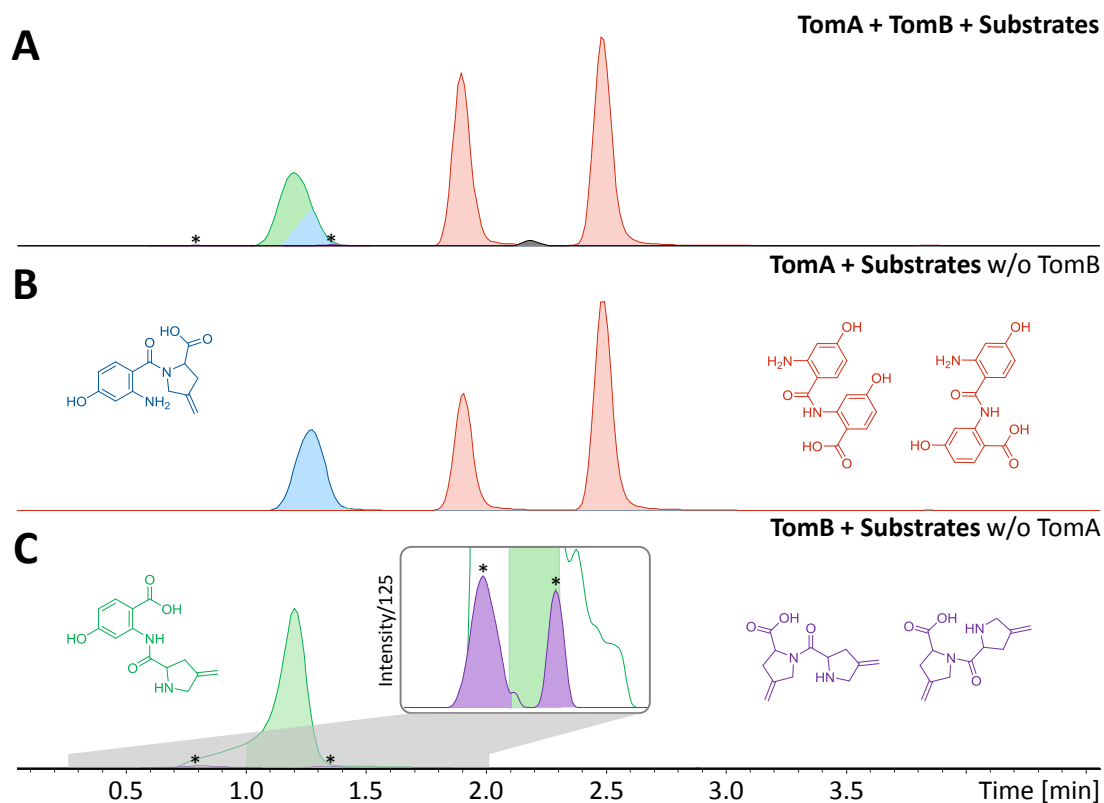


Figure S 2.4: Overview of detected dimeric side products in the *in vitro* reconstitution assays. Overlaid EIC traces of the hetero-dimeric peptides with proposed different linkage (m/z 263.10263 ± 0.002), proposed homo-dimeric side product conformers of two anthranilic acids (m/z 289.08190 ± 0.002) and two prolines (m/z 237.12337 ± 0.002). **A** Assay with TomA and B including 5-hydroxy anthranilic acid, methylene proline and the respective co-factors. **B** Same mixture but lacking TomB. **C** Same mixture as in A but lacking TomA.

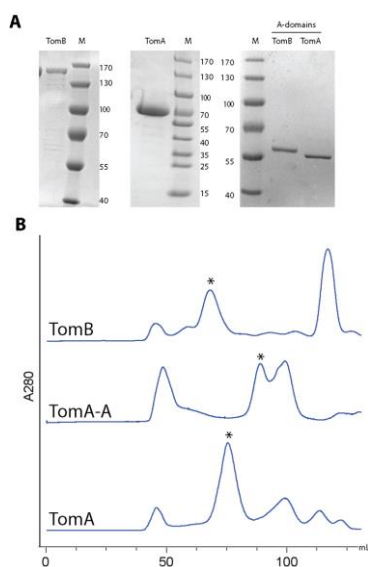


Figure S 2.5: **A** SDS-PAGE analysis of SEC purified proteins TomA (65 kDa), TomB (167 kDa) and the TomA A-domain (55 kDa). **B** SEC traces of TomA, TomA-A (A-domain) and TomB.

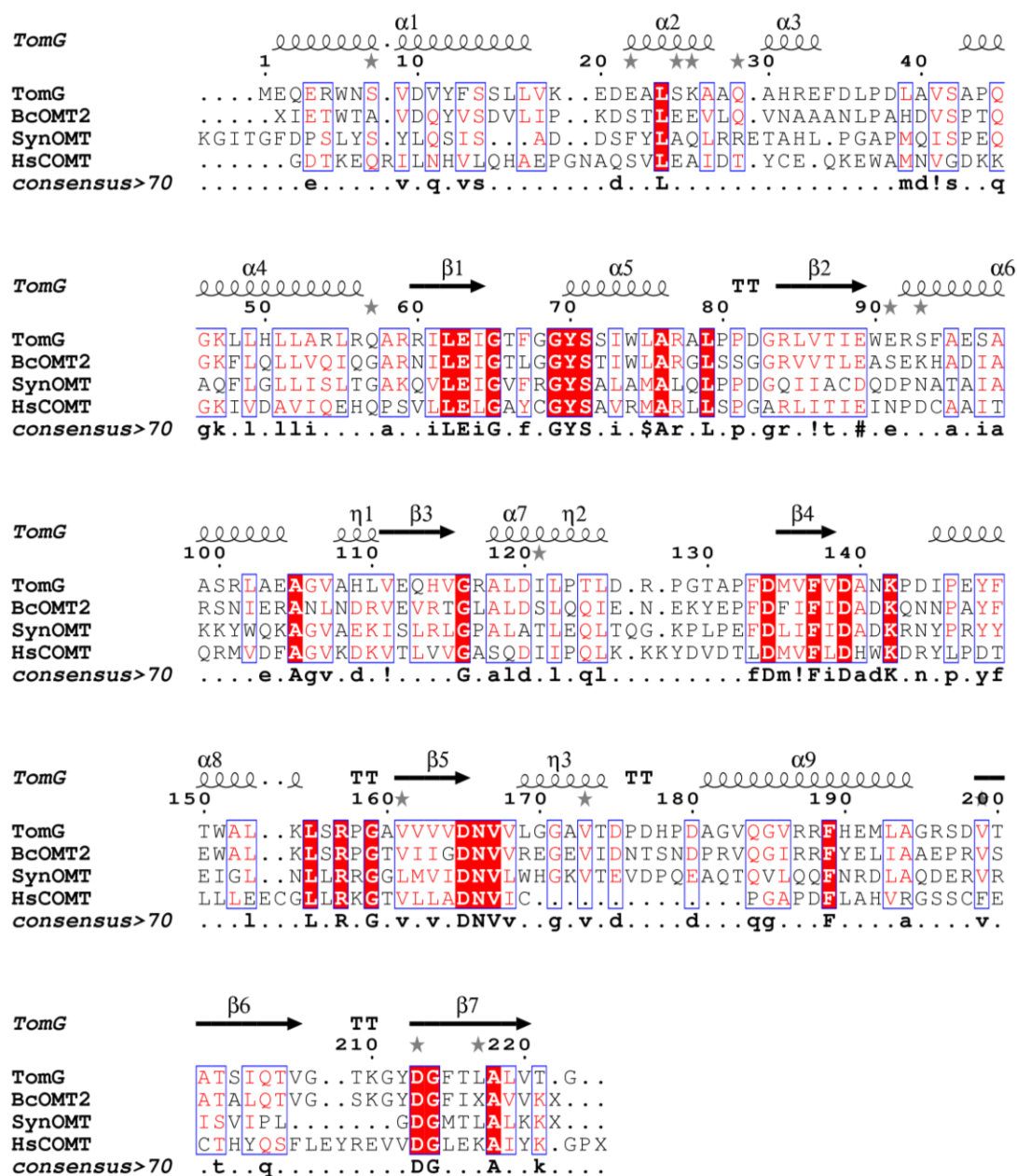


Figure S 2.6: Multiple structural alignment and consensus sequence of diverse OMTs. Secondary structure assignment and sequence numbering corresponds to the structure of TomG. Sequences were aligned using the DALI webserver. The alignment was manually edited and visualized using the ESPrpt 3.0 webserver. Abbreviations used: BcOMT2 - O-methyltransferase from *Bacillus cereus*; SynOMT - O-methyltransferase from *Cyanobacterium synechocystis*; HsCOMT - catechol-O-methyltransferase from *Homo sapiens*.

2.7 Supplemental Tables

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

	TomG
Crystallization and data Collection	
Crystallization buffer	2.1 M $\text{NH}_4(\text{SO}_4)_2$, 0.1 M Li_2SO_4 , 0.1 M Hepes pH 7.3
Crystallization drop	0.1 μL crystallization buffer + 0.1 μL 5 mg/mL protein in 50 mM Bicine pH 8.0, 200 mM NaCl, 1 mM CaCl_2 , 10 % Glycerol, 0.6 mM S-Adenosyl methionine
Wavelength (Å)	1.033
Resolution range (Å) ¹	50.47 – 1.55 (1.58– 1.55)
Space group	P2 ₁ 2 ₁ 2 ₁
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
	PSB12282
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

2.8 References

- [1] Newman, D. J. & Cragg, G. M.: Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **2016**, 79, 629–661, DOI: 10.1021/acs.jnatprod.5b01055
- [2] Arima, K., Kosaka, M., Tamura, G., Imanaka, H. & Sakai, H.: Studies on tomaymycin, a new antibiotic. I. Isolation and properties of tomaymycin. *J. Antibiot. (Tokyo)*. **1972**, 25, 437–44, DOI: 10.7164/antibiotics.25.437
- [3] Gerratana, B.: Biosynthesis, synthesis, and biological activities of pyrrolobenzodiazepines. *Med. Res. Rev.* **2012**, 32, 254–293, DOI: 10.1002/med.20212
- [4] Puvvada, M. S. *et al.*: Inhibition of bacteriophage T7 RNA polymerase in vitro transcription by DNA-binding pyrrolo[2,1-c][1,4]benzodiazepines. *Biochemistry* **1997**, 36, 2478–2484, DOI: 10.1021/bi952490r
- [5] Hurley, L. H. *et al.*: Pyrrolo[1,4]benzodiazepine antitumor antibiotics: relationship of DNA alkylation and sequence specificity to the biological activity of natural and synthetic compounds. *Chem Res Toxicol* **1988**, 1, 258–268, DOI: 10.1021/tx00005a002
- [6] Thurston, D. E. *et al.*: Effect of a-ring modifications on the DNA-binding behavior and cytotoxicity of pyrrolo[2,1-c][1,4]benzodiazepines. *J. Med. Chem.* **1999**, 42, 1951–1964, DOI: 10.1021/jm981117p
- [7] Gregson, S. J. *et al.*: Effect of C2-exo unsaturation on the cytotoxicity and DNA-binding reactivity of pyrrolo[2,1-c][1,4]benzodiazepines. *Bioorganic Med. Chem. Lett.* **2000**, 10, 1845–1847, DOI: 10.1016/S0960-894X(00)00351-6
- [8] Kumar, R. & Lown, J.: Recent Developments in Novel Pyrrolo[2,1-c][1,4]Benzodiazepine Conjugates: Synthesis and Biological Evaluation. *Mini-Reviews Med. Chem.* **2003**, 3, 323–339, DOI: 10.2174/1389557033488097
- [9] Li, W., Chou, S., Khullar, A. & Gerratana, B.: Cloning and characterization of the biosynthetic gene cluster for tomaymycin, an sjg-136 monomeric analog. *Appl. Environ. Microbiol.* **2009**, 75, 2958–2963, DOI: 10.1128/AEM.02325-08
- [10] Li, W., Khullar, A., Chou, S., Sacramo, A. & Gerratana, B.: Biosynthesis of sibiromycin, a potent antitumor antibiotic. *Appl. Environ. Microbiol.* **2009**, 75, 2869–2878, DOI: 10.1128/AEM.02326-08
- [11] Hu, Y. *et al.*: Benzodiazepine Biosynthesis in *Streptomyces refuineus*. *Chem. Biol.* **2007**, 14, 691–701, DOI: 10.1016/j.chembiol.2007.05.009
- [12] Hur, G. H., Vickery, C. R. & Burkart, M. D.: Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat. Prod. Rep.* **2012**, 29, 1074, DOI: 10.1039/c2np20025b
- [13] Finking, R. & Marahiel, M. A.: Biosynthesis of Nonribosomal Peptides. *Annu. Rev. Microbiol.* **2004**, 58, 453–488, DOI: 10.1146/annurev.micro.58.030603.123615
- [14] Fischbach, M. A. & Walsh, C. T.: Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic machinery, and mechanisms. *Chem. Rev.* **2006**, 106, 3468–3496, DOI: 10.1021/cr0503097
- [15] Sattely, E. S. & Walsh, C. T.: A latent oxazoline electrophile for N-O-C bond formation in pseudomonine biosynthesis. *J. Am. Chem. Soc.* **2008**, 130, 12282–12284, DOI: 10.1021/ja804499r
- [16] Wyatt, M. A., Mok, M. C. Y., Junop, M. & Magarvey, N. A.: Heterologous Expression and Structural Characterisation of a Pyrazinone Natural Product Assembly Line. *ChemBioChem* **2012**, 13, 2408–2415, DOI: 10.1002/cbic.201200340

- [17] Wilson, D. J., Shi, C., Teitelbaum, A. M., Gulick, A. M. & Aldrich, C. C.: Characterization of AusA: A Dimodular Nonribosomal Peptide Synthetase Responsible for the Production of Aureusimine Pyrazinones. *Biochemistry* **2013**, 52, 926–937, DOI: 10.1021/bi301330q
- [18] Feifel, S. C. *et al.*: In vitro synthesis of new enniatins: Probing the α -D-hydroxy carboxylic acid binding pocket of the multienzyme enniatin synthetase. *ChemBioChem* **2007**, 8, 1767–1770, DOI: 10.1002/cbic.200700377
- [19] Müller, J., Feifel, S. C., Schmiederer, T., Zocher, R. & Süssmuth, R. D.: In vitro synthesis of new cyclodepsipeptides of the PF1022-Type: Probing the α -D-hydroxy acid tolerance of PF1022 synthetase. *ChemBioChem* **2009**, 10, 323–328, DOI: 10.1002/cbic.200800539
- [20] Zhang, W. *et al.*: Nine enzymes are required for assembly of the pacidamycin group of peptidyl nucleoside antibiotics. *J. Am. Chem. Soc.* **2011**, 133, 5240–5243, DOI: 10.1021/ja2011109
- [21] Sandy, M., Rui, Z., Gallagher, J. & Zhang, W.: Enzymatic synthesis of dilactone scaffold of antimycins. *ACS Chem. Biol.* **2012**, 7, 1956–1961, DOI: 10.1021/cb300416w
- [22] Si, S. I. *et al.*: *ChemComm.* **2012**,
- [23] Gehring, A. M., Mori, I. & Walsh, C. T.: Reconstitution and characterization of the Escherichia coli enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry* **1998**, 37, 2648–2659, DOI: 10.1021/bi9726584
- [24] Miller, D. A., Luo, L., Hillson, N., Keating, T. A. & Walsh, C. T.: Yersiniabactin synthetase: A four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of Yersinia pestis. *Chem. Biol.* **2002**, 9, 333–344, DOI: 10.1016/S1074-5521(02)00115-1
- [25] Patel, H. M. & Walsh, C. T.: In vitro reconstitution of the Pseudomonas aeruginosa nonribosomal peptide synthesis of pyochelin: Characterization of backbone tailoring thiazoline reductase and N-methyltransferase activities. *Biochemistry* **2001**, 40, 9023–9031, DOI: 10.1021/bi010519n
- [26] Keating, T. A., Marshall, C. G. & Walsh, C. T.: Reconstitution and characterization of the Vibrio cholerae vibriobactin synthetase from VibB, VibE, VibF, and VibH. *Biochemistry* **2000**, 39, 15522–15530, DOI: 10.1021/bi0016523
- [27] Gaitatzis, N., Hans, a, Müller, R. & Beyer, S.: 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. *J. Biochem.* **2001**, 129, 119–24,
- [28] Li, Y., Weissman, K. J. & Müller, R.: Myxochelin biosynthesis: Direct evidence for two- and four-electron reduction of a carrier protein-bound thioester. *J. Am. Chem. Soc.* **2008**, 130, 7554–7555, DOI: 10.1021/ja8025278
- [29] Jaitzig, J., Li, J., Süssmuth, R. D. & Neubauer, P.: Reconstituted biosynthesis of the nonribosomal macrolactone antibiotic valinomycin in Escherichia coli. *ACS Synth. Biol.* **2014**, 3, 432–438, DOI: 10.1021/sb400082j
- [30] McLafferty, F. W.: A Century of Progress in Molecular Mass Spectrometry. *Annu. Rev. Anal. Chem.* **2011**, 4, 1–22, DOI: 10.1146/annurev-anchem-061010-114018
- [31] Snijder, J., Rose, R. J., Veisler, D., Johnson, J. E. & Heck, A. J. R.: Studying 18 MDa virus assemblies with native mass spectrometry. *Angew. Chemie - Int. Ed.* **2013**, 52, 4020–4023, DOI: 10.1002/anie.201210197
- [32] Thompson, N. J., Rosati, S., Rose, R. J. & Heck, A. J.: The impact of mass spectrometry on the study of intact antibodies: from post-translational modifications to structural analysis.

- Chem. Commun.* **2013**, 49, 538–548, DOI: 10.1039/c2cc36755f
- [33] Toby, T. K., Fornelli, L. & Kelleher, N. L.: Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu. Rev. Anal. Chem.* **2016**, 9, 499–519, DOI: 10.1146/annurev-anchem-071015-041550
- [34] Li, W., Chou, S., Khullar, A. & Gerratana, B.: Cloning and characterization of the biosynthetic gene cluster for tomaymycin, an sjg-136 monomeric analog. *Appl. Environ. Microbiol.* **2009**, 75, 2958–2963, DOI: 10.1128/AEM.02325-08
- [35] Wilson, D. J. & Aldrich, C. C.: A continuous kinetic assay for adenylation enzyme activity and inhibition. *Anal. Biochem.* **2010**, 404, 56–63, DOI: 10.1016/j.ab.2010.04.033
- [36] Balibar, C. J., Vaillancourt, F. H. & Walsh, C. T.: Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. *Chem. Biol.* **2005**, 12, 1189–1200, DOI: 10.1016/j.chembiol.2005.08.010
- [37] Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D. H.: Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.* **2007**, 7, 78, DOI: 10.1186/1471-2148-7-78
- [38] Gaudelli, N. M., Long, D. H. & Townsend, C. A.: β -Lactam formation by a non-ribosomal peptide synthetase during antibiotic biosynthesis. *Nature* **2015**, 520, 383–387, DOI: 10.1038/nature14100
- [39] Haslinger, K., Peschke, M., Brieke, C., Maximowitsch, E. & Cryle, M. J.: X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis. *Nature* **2015**, 521, 105–109, DOI: 10.1038/nature14141
- [40] Stachelhaus, T., Mootz, H. D., Bergendahl, V. & Marahiel, M. a: Peptide Bond Formation in Nonribosomal Peptide Biosynthesis. *J. Biol. Chem.* **1998**, 273, 22773–22781, DOI: 10.1074/jbc.273.35.22773
- [41] Belshaw, P. J.: Aminoacyl-CoAs as Probes of Condensation Domain Selectivity in Nonribosomal Peptide Synthesis. *Science (80-.)*. **1999**, 284, 486–489, DOI: 10.1126/science.284.5413.486
- [42] Meyer, S. *et al.*: Biochemical Dissection of the Natural Diversification of Microcystin Provides Lessons for Synthetic Biology of NRPS. *Cell Chem. Biol.* **2016**, 23, 462–471, DOI: 10.1016/j.chembiol.2016.03.011
- [43] Mazur, M. T., Walsh, C. T. & Kelleher, N. L.: Site-Specific Observation of Acyl Intermediate Processing in Thiotemplate Biosynthesis by Fourier Transform Mass Spectrometry: The Polyketide Module of Yersiniabactin Synthetase. *Biochemistry* **2003**, 42, 13393–13400, DOI: 10.1021/bi035585z
- [44] Hicks, L. M., O'Connor, S. E., Mazur, M. T., Walsh, C. T. & Kelleher, N. L.: Mass spectrometric interrogation of thioester-bound intermediates in the initial stages of epothilone biosynthesis. *Chem. Biol.* **2004**, 11, 327–335, DOI: 10.1016/j.chembiol.2004.02.021
- [45] McLoughlin, S. M. & Kelleher, N. L.: Kinetic and regiospecific interrogation of covalent intermediates in the nonribosomal peptide synthesis of yersiniabactin. *J. Am. Chem. Soc.* **2004**, 126, 13265–13275, DOI: 10.1021/ja0470867
- [46] Garneau, S., Dorrestein, P. C., Kelleher, N. L. & Walsh, C. T.: Characterization of the formation of the pyrrole moiety during clorobiocin and coumermycin A1 biosynthesis. *Biochemistry* **2005**, 44, 2770–2780, DOI: 10.1021/bi0476329
- [47] Miller, L. M., Mazur, M. T., McLoughlin, S. M. & Kelleher, N. L.: Parallel interrogation of covalent intermediates in the biosynthesis of gramicidin S using high-resolution mass spectrometry. *Protein Sci.* **2005**, 14, 2702–2712, DOI: 10.1110/ps.051553705

- [48] Meier, J. L. *et al.*: Practical 4'-Phosphopantetheine active site discovery from proteomic samples. *J. Proteome Res.* **2011**, 10, 320–329, DOI: 10.1021/pr100953b
- [49] Meluzzi, D., Zheng, W. H., Hensler, M., Nizet, V. & Dorrestein, P. C.: Top-down mass spectrometry on low-resolution instruments: Characterization of phosphopantetheinylated carrier domains in polyketide and non-ribosomal biosynthetic pathways. *Bioorganic Med. Chem. Lett.* **2008**, 18, 3107–3111, DOI: 10.1016/j.bmcl.2007.10.104
- [50] Mitchell Wells, J. & McLuckey, S. A.: *Methods Enzymol.* (Elsevier, **2005**). 402, 148–185, DOI: 10.1016/S0076-6879(05)02005-7
- [51] Han, X. *et al.*: Sheathless capillary electrophoresis-tandem mass spectrometry for top-down characterization of pyrococcus furiosus proteins on a proteome scale. *Anal. Chem.* **2014**, 86, 11006–11012, DOI: 10.1021/ac503439n
- [52] Brodbelt, J. S.: Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **2016**, 88, 30–51, DOI: 10.1021/acs.analchem.5b04563
- [53] Crawford, J. M. *et al.*: Deconstruction of Iterative Multidomain Polyketide Synthase Function. *Science (80-.)*. **2008**, 320, 243–246, DOI: 10.1126/science.1154711
- [54] Sun, X., Li, H., Alfermann, J., Mootz, H. D. & Yang, H.: Kinetics profiling of gramicidin S synthetase A, a member of nonribosomal peptide synthetases. *Biochemistry* **2014**, 53, 7983–7989, DOI: 10.1021/bi501156m
- [55] Tarry, M. J., Haque, A. S., Bui, K. H. & Schmeing, T. M.: X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. *Structure* **2017**, 25, 783–793.e4, DOI: 10.1016/j.str.2017.03.014
- [56] Hahne, H. *et al.*: DMSO enhances electrospray response, boosting sensitivity of proteomic experiments. *Nat. Methods* **2013**, 10, 989–991, DOI: 10.1038/nmeth.2610
- [57] Schluckebier, G., O'Gara, M., Saenger, W. & Cheng, X.: Universal Catalytic Domain Structure of AdoMet-dependent Methyltransferases. *J. Mol. Biol.* **1995**, 247, 16–20, DOI: 10.1006/jmbi.1994.0117
- [58] Martin, J. L. & McMillan, F. M.: SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol.* **2002**, 12, 783–93, DOI: 10.1016/S0959-440X(02)00391-3
- [59] Liscombe, D. K., Louie, G. V. & Noel, J. P.: Architectures, mechanisms and molecular evolution of natural product methyltransferases. *Nat. Prod. Rep.* **2012**, 29, 1238, DOI: 10.1039/c2np20029e
- [60] Asano, Y., Woodard, R. W., Houck, D. R. & Floss, H. G.: Stereochemical course of the transmethylation catalyzed by histamine N-methyltransferase. *Arch. Biochem. Biophys.* **1984**, 231, 253–256, DOI: 10.1016/0003-9861(84)90385-0
- [61] Frishman, D. & Argos, P.: Knowledge-based protein secondary structure assignment. *Proteins Struct. Funct. Bioinforma.* **1995**, 23, 566–579, DOI: 10.1002/prot.340230412
- [62] Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A.: Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci.* **2001**, 98, 10037–10041, DOI: 10.1073/pnas.181342398
- [63] Dolinsky, T. J. *et al.*: PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **2007**, 35, W522-5, DOI: 10.1093/nar/gkm276
- [64] Axelrod, J. & Tomchick, R.: Enzymatic O-Methylation of Epinephrine and Other Catechols. *J. Biol. Chem.* **1958**, 233, 702–705,

- [65] Männistö, P. T. & Kaakkola, S.: Catechol-O-methyltransferase (COMT): biochemistry, molecular biology, pharmacology, and clinical efficacy of the new selective COMT inhibitors. *Pharmacol. Rev.* **1999**, 51, 593–628,
- [66] Takeo, M. *et al.*: Mechanism of 4-nitrophenol oxidation in *Rhodococcus* sp. strain PN1: Characterization of the two-component 4-nitrophenol hydroxylase and regulation of its expression. *J. Bacteriol.* **2008**, 190, 7367–7374, DOI: 10.1128/JB.00742-08
- [67] Giessen, T. W., Kraas, F. I. & Marahiel, M. A.: A four-enzyme pathway for 3,5-dihydroxy-4-methylanthranilic acid formation and incorporation into the antitumor antibiotic sibiromycin. *Biochemistry* **2011**, 50, 5680–5692, DOI: 10.1021/bi2006114
- [68] Cho, J.-H., Park, Y., Ahn, J.-H., Lim, Y. & Rhee, S.: Structural and Functional Insights into O-Methyltransferase from *Bacillus cereus*. *J. Mol. Biol.* **2008**, 382, 987–997, DOI: 10.1016/j.jmb.2008.07.080
- [69] Lee, H. J., Bong Gyu, K. & Joong-Hoon, A.: Molecular cloning and characterization of *Bacillus cereus* O-methyltransferase. *J. Microbiol. Biotechnol.* **2006**, 16, 619–622,
- [70] Rutherford, K., Le Trong, I., Stenkamp, R. E. & Parson, W. W.: Crystal Structures of Human 108V and 108M Catechol O-Methyltransferase. *J. Mol. Biol.* **2008**, 380, 120–130, DOI: 10.1016/j.jmb.2008.04.040
- [71] Lautala, P., Ulmanen, I. & Taskinen, J.: Molecular mechanisms controlling the rate and specificity of catechol O-methylation by human soluble catechol O-methyltransferase. *Mol. Pharmacol.* **2001**, 59, 393–402, DOI: 10.1124/mol.59.2.393
- [72] Gasteiger, E. *et al.*: *Proteomics Protoc. Handb.* (Walker, J. M.) (Humana Press, **2005**). 112, 571–607, DOI: 10.1385/1-59259-890-0:571
- [73] Burkhardt, A. *et al.*: Status of the crystallography beamlines at PETRA III. *Eur. Phys. J. Plus* **2016**, 131, DOI: 10.1140/epjp/i2016-16056-0
- [74] Kabsch, W.: XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, 66, 125–132, DOI: 10.1107/S0907444909047337
- [75] Evans, P. R. & Murshudov, G. N.: How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, 69, 1204–1214, DOI: 10.1107/S0907444913000061
- [76] Winn, M. D. *et al.*: Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, 67, 235–242, DOI: 10.1107/S0907444910045749
- [77] McCoy, A. J. *et al.*: Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, 40, 658–674, DOI: 10.1107/S0021889807021206
- [78] Kopycki, J. G. *et al.*: Functional and structural characterization of a cation-dependent O-methyltransferase from the cyanobacterium *Synechocystis* sp. strain PCC 6803. *J. Biol. Chem.* **2008**, 283, 20888–20896, DOI: 10.1074/jbc.M801943200
- [79] Painter, J. & Merritt, E. A.: Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2006**, 62, 439–450, DOI: 10.1107/S0907444906005270
- [80] Afonine, P. V. *et al.*: Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2012**, 68, 352–367, DOI: 10.1107/S0907444912001308
- [81] Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K.: Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, 66, 486–501, DOI: 10.1107/S0907444910007493
- [82] Chen, V. B. *et al.*: MolProbity: All-atom structure validation for macromolecular

- crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, 66, 12–21, DOI: 10.1107/S0907444909042073
- [83] Adams, P. D. *et al.*: PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, 66, 213–221, DOI: 10.1107/S0907444909052925
- [84] Holm, L. & Rosenström, P.: Dali server: Conservation mapping in 3D. *Nucleic Acids Res.* **2010**, 38, W545–9, DOI: 10.1093/nar/gkq366
- [85] Berman, H. *et al.*: The protein data bank. *Nucleic Acids Res.* **2000**, 28, 235–242,
- [86] Lattemann, C., Broenstrup, M., Werner, S., Müller, R. & Harmrolfs, K.: Method for Producing Recombinant 11-De-O-Methyltomaymycin. **2012**,
- [87] Dümmler, A., Lawrence, A.-M. & de Marco, A.: Simplified screening for the detection of soluble fusion constructs expressed in E. coli using a modular set of vectors. *Microb. Cell Fact.* **2005**, 4, 34, DOI: 10.1186/1475-2859-4-34
- [88] Gaitatzis, N., Hans, A., Müller, R. & Beyer, S.: 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. *J. Biochem.* **2001**, 129, 119–124,
- [89] Robert, X. & Gouet, P.: Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **2014**, 42, W320–4, DOI: 10.1093/nar/gku316
- [90] Afonine, P. V. *et al.*: FEM: Feature-enhanced map. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2015**, 71, 646–666, DOI: 10.1107/S1399004714028132

Chapter 3 – Tilivalline

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

Alexander von Tesmar^{1,2}, **Michael Hoffmann**^{1,2}, Viktoria Schmitt^{1,2}, Antoine Abou Fayad^{1,2},
Jennifer Herrmann^{1,2}, Rolf Müller^{1,2}

Manuscript ready for submission

¹Department of Microbial Natural Products (MINS), Helmholtz Institute for Pharmaceutical Research
Saarland (HIPS) - Helmholtz Centre for Infection Research (HZI) and Institute for Pharmaceutical
Biotechnology, Saarland University, 66123 Saarbrücken, Germany

²German Center for Infection Research (DZIF), Partner site Hannover-Braunschweig, 38124 Braunschweig,
Germany

Contributions

Author's effort:

The author contributed partly to the conception of the study, performed experiments, evaluated and interpreted resulting data. The author 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 the author. The author conducted also the experiments on the analytical scale to prove the spontaneous indole incorporation. Furthermore, the author contributed to the writing and editing of the manuscript.

Contribution by others:

Alexander von Tesmar designed and performed most of the experiments, described, evaluated and interpreted the resulting data. He contributed substantially to the conception of the study and wrote 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.

3 Tilivalline

3.1 Abstract

Tilivalline 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.

3.2 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 activity³.

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.² 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 4'phosphopanthien 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 products¹⁵.

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.

3.3 Results

3.3.1 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).

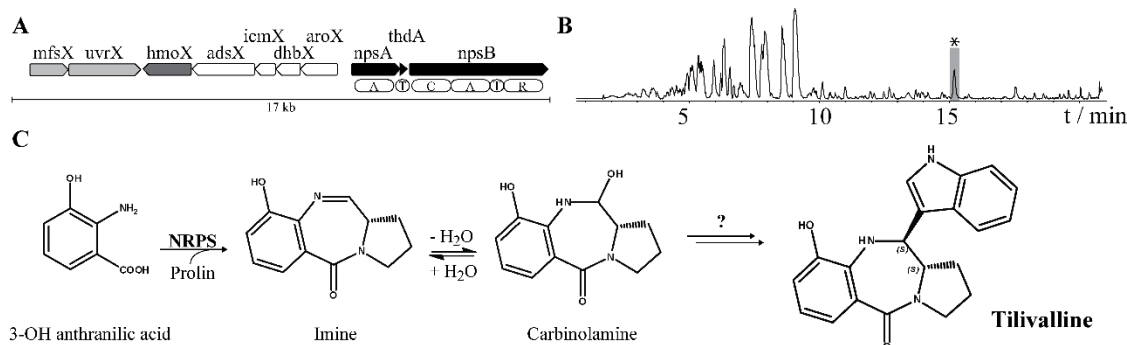


Figure 3.1: A Biosynthetic gene cluster of tilivalline from *K. oxytoca* isolate #6 spans 17 kb. It contains ten open reading frames constituting a two modular NRPS (black), the putative 3-hydroxylase (dark grey), genes for providing anthranilic acid (white) and regulators (light grey). B UV (200-600 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.

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.² 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 3.1).

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*¹¹. 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 moiety¹⁷. 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).

3.3.2 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 P_{tet} and P_{BAD} as well as the vector backbone were PCR amplified while suitable homologous regions at the 5'-end of the primers were introduced¹⁸. 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).

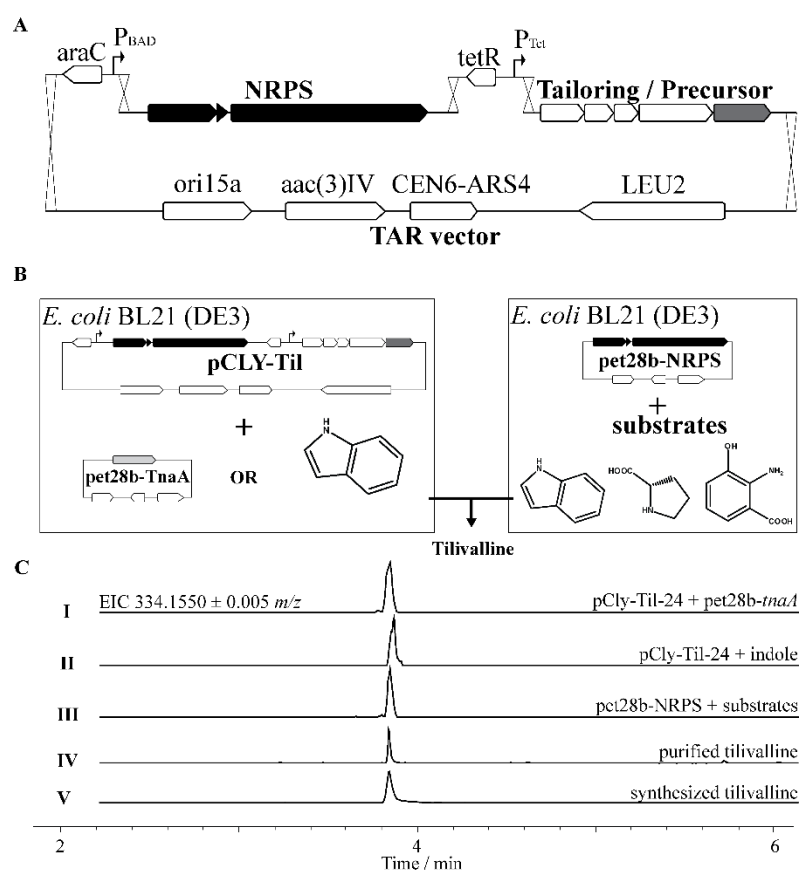


Figure 3.2: **A** Schematic display of the TAR assembly of the five PCR generated DNA fragments to yield the heterologous expression construct pCLY-Til. **B** Two different heterologous expression systems for tilivalline in *E. coli* BL21 (DE3) have been generated. Simultaneous expression of pCLY-Til and pet28b-tnaA or indole supplementation led to successful restoration of tilivalline production. Expression of the construct pet28b-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 pet28b-tnaA (I) or supplemented in the culture media (II). *E. coli* BL21 (DE3) pet28b-NRPS produces tilivalline when the substrates are supplemented (III). Synthesized (V) and purified tilivalline from *K. oxytoca* isolate #6 (IV) were used as standard.

Upon induction of *E. coli* BL21 (DE3) pCLY-Til-24 in M9 minimal medium containing tetracycline and L-arabinose, tilivalline was 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 media¹⁹, 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 pet28b expression vector. Upon induction of an *E. coli* transformant harboring both constructs, pCLY-Til-24 and pet28b-*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 pet28b, 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* pet28b-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).

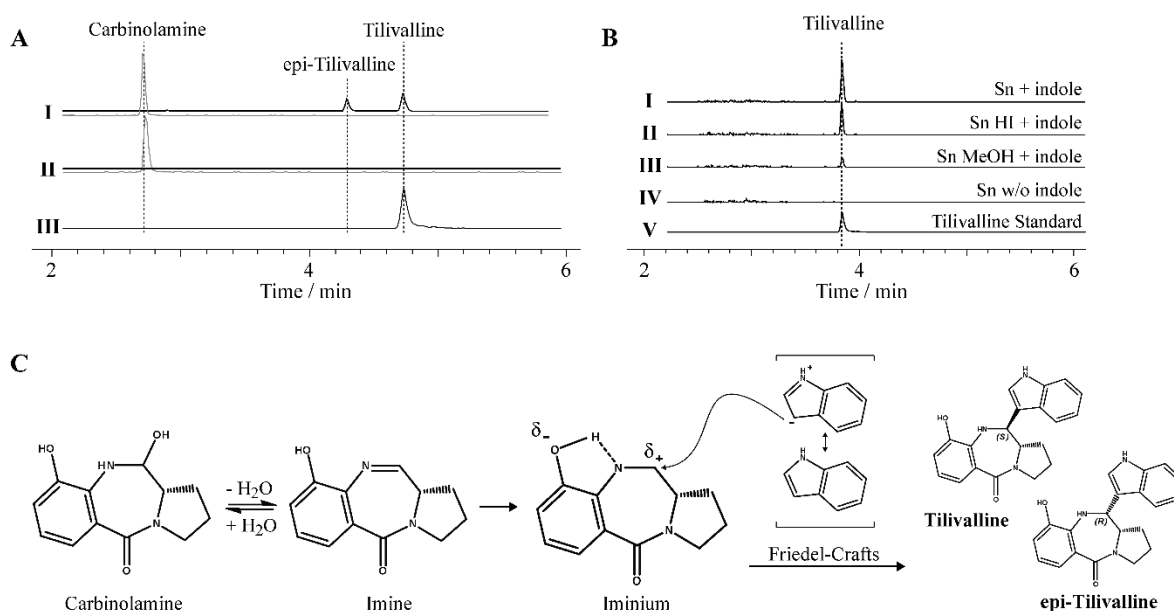


Figure 3.3: A Formation of tilivalline and epi-tilivalline can be detected upon incubation of indole with the purified PBD precursor in the carbinolamine form (I). PBD-precursor was purified from the *in vitro* reconstitution of NpsA, ThdA and NpsB without indole supplementation (II). Synthesized tilivalline 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) *pet28b*-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.

3.3.3 *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 HRV3C digestion to remove the tags, and a polishing gel filtration step (Figure S 3.3). 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 3.4).

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*²⁰ (Figure S 3.4). 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.

3.3.4 Tilivalline and not the Carbinolamine is the Active Species

Schneditz et al.² reported an IC_{50} of 10 μM 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 μM 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 μM 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 μM . We speculate that the substitution pattern found in tomaymycin allows for stability of the imine which is not the case in the hypothetical tilivalline 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 μM 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.

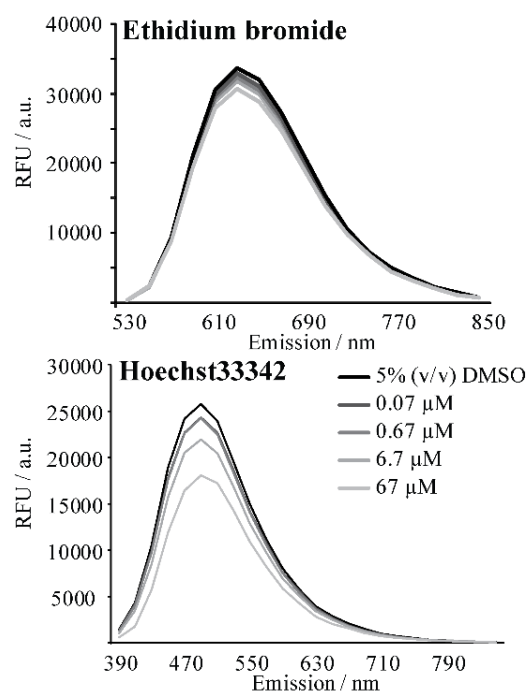
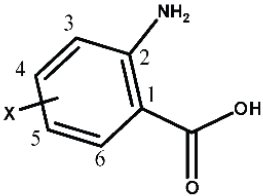
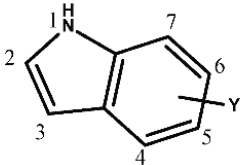


Figure 3.4: Tilivalline does not displace ethidium bromide but displaces Hoechst33342 at concentrations higher than 6.7 μ M. This indicates a weak binding of tilivalline to the minor groove of the DNA but excludes an intercalation.

3.3.5 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 (Table 3.1). 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 3.1). 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 3.2). The halogenated as well as the hydroxylated and methoxylated anthranilic acid derivatives were readily accepted with product signal intensities comparable to the native 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.

Table 3.1: Mutasynthesis of tilivalline derivatives by supplementing *E. coli* BL21 (DE3) *pet28b*-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).

			
anthranilic acid		indole	
Derivative	Production	Derivative	Production
AA	+++	4-OH	+++
4-OH	+++	5-OH	+++
3-Me	++	6-OH	+++
3-OMe	+++	4-NO ₂	+
5-OMe	++	5-NO ₂	++
3-F	++	6-NO ₂	++
5-F	+++	7-NO ₂	+
3-Cl	++	8-F	+++
4-Cl	+++	9-F	+++
5-Cl	+++	10-F	+++
5-Br	++	5-Br	+
3-CF ₃	tr.	5-I	tr.

3.3.6 Salicylic Acid Based Compounds Inhibit Tilivalline Production in *K. oxytoca*

Tilivalline production in *K. oxytoca* isolate #6 was evaluated for different fermentation conditions (see table), 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 3.5).

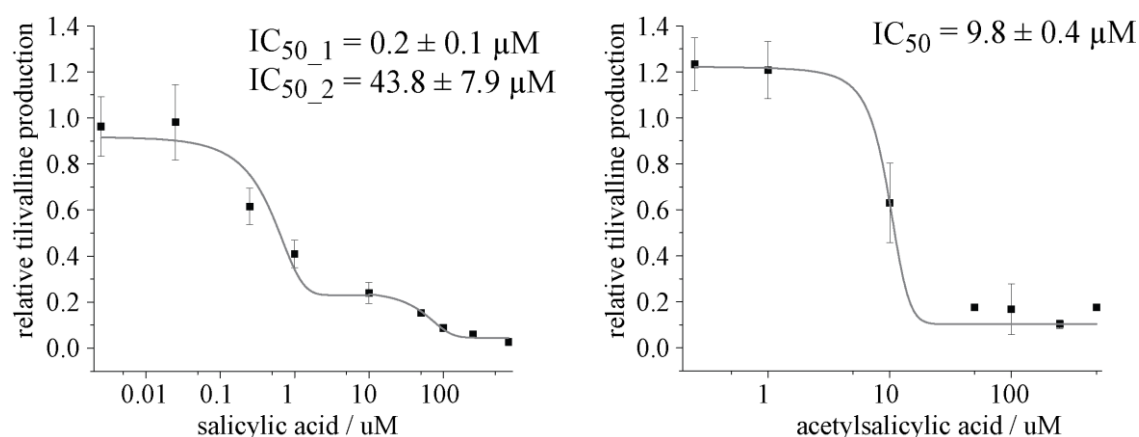


Figure 3.5: Salicylic acid and acetylsalicylic acid inhibit tilivalline biosynthesis in *K. oxytoca* isolate #6 liquid culture. Experiments were carried out in triplicates. Mean values and standard deviations are given. A two-step and one-step dose-response model was used to fit salicylic acid and acetylsalicylic acid data, respectively.

3.4 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 regime²⁷. *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 efforts². 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. Since *K. oxytoca* is indole-positive the required free indole is provided by degrading tryptophan²⁹.

Other PBDs like tomaymycin, sibiromycin, and anthramycine 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 DNA³². 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 constitutes 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 adenyating enzymes³⁵. Despite the established application in *in vitro* experiments, the application in liquid culture is hampered due to poor cell-permeability³⁶. 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 conditions³⁷. 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 prodrug³⁸. 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.

3.5 Experimental Procedures

3.5.1 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 software³⁹, 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: K1535631) in Geneious software⁴⁰, version 9.0.5. The resulting draft genome circular scaffold of 5,720,468 bp was used for the downstream analysis.

3.5.2 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 µm 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 µm 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 µL 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.

3.5.3 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 previously¹⁸. Colony PCR targeting fragment borders was used to preselect yeast colonies (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.

3.5.4 Heterologous Expression of the Tilivalline Biosynthetic Gene Cluster

E. coli BL21 (DE3) pCly-Til-24 (or 28) 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) 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 pet28b 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.

3.5.5 Heterologous Expression of the Tilivalline NRPS

The operon harboring the tilivalline NRPS was cloned into pet28b using suitable primers allowing for tag-free expression. *E. coli* BL21 (DE3) pet28b-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 µL volume using a 96-well flat bottom plate.

3.5.6 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 (restriction site in bold):

NpsA-FP AAAA**CCATGG**CAACGCATTGAGCATA

NpsA -RP AAAAA**AAGCTTTT**TACACCTGCTCCAGTAAAGAATTT

ThdA-FP AAAA**CCATGG**CAGACAACGTTGAGCAA

ThdA -RP AAAAA**AAGCTTTT**TAGAGCTTTTGCCGTTGCC

NpsB-FP AAAA**CCATGG**CACCATCATTCAATAGCG

NpsB-RP AAAA**GGATCC**CTAGACAATTTCTTCTGCCGCT

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 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 (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/>; Gasteiger et al.⁴¹).

3.5.7 Purification of TnaA

The respective gene was inserted into pet28b with an N-terminal 6xHis tag for purification and without tag for heterologous expression using according primers (restriction site in bold):

TnaA-noHis-FP	AAAAA CCATGG CACGTATCCCTGAGCCGT
TnaA-RP	AAAAA GGATCCT TATTTCACCGGTTTCAGTCTG
TnaA-FP	AAAAA AGCT AGCATGAAACGTATCCCTGAGC

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.

3.5.8 *In vitro* Reconstitution of Tilivalline NRPS

NpsA, ThdA and NpsB (0.4 µM each) were incubated with the respective cofactors (1 mM ATP, 0.5 mM NADPH) and substrates (0.5 mM Proline, 3-OH-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 µL at room temperature for 4h. The mixture was dried in a vacuum condenser and subsequently resuspended in 20 µL water:acetonitrile (1:1), centrifuged and submitted to LC-MS analysis.

3.5.9 *In vitro* Reconstitution of TnaA

4 µM 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 µL for 2h at 37°C. The mixture was dried in a vacuum condenser and subsequently resuspended in 20 µL water:acetonitrile (1:1), centrifuged and submitted to LC analysis. Equimolar standards of indole and tryptophan were used to identify the UV-peaks.

3.5.10 *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.

3.5.11 DNA Displacement Assays

DNA displacement assays using either ethidium bromide or Hoechst33342 was conducted as described previously⁴².

3.5.12 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.

3.5.13 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 using a Phenomenex Luna PFP, 100 x 2.0 mm, 3 μ m dp column and applying the following gradient: 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.

3.6 Supplemental Information

Table S 3.1: Sequence homologies of tilivalline biosynthetic gene cluster found in *K. oxytoca* isolate #6 compared to the Schneditz et al². cluster and conferred from BLAST.

Gene	Identity* / %	Putative Function	Size / bp	Homolog	Identity / %	BLAST	
						E-value	Score
<i>mfsX</i>	100.0	Uncharacterized MFS-type transporter	410	ydgK	34	1,00E-69	590
<i>uvrX</i>	96.7	UvrABC system protein A	740	uvrA	39	2,00E-137	1117
<i>hmoX</i>	99.8	4-nitrophenol-monooxygenase, oxy. component	505	NphA1	62	0	654
<i>adsX</i>	98.8	Anthranilate synthase, phenazine specific	649	phzE	42	1,00E-149	1169
<i>icmX</i>	100.0	2,3 dihydro-2,3 dihydroxybenzoate synthase	210	phzD	50	1,00E-63	515
<i>dhbX</i>	100.0	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase	261	dhbA	46	6,00E-68	533
<i>aroX</i>	99.5	phospho-2-dehydro-3-deoxyheptonate aldolase	389	phzC	42	5,00E-90	725
<i>npsA</i>	98.8	NRPS [A]	506		99		
<i>thdA</i>	97.4	NRPS [T]	76		97		
<i>npsB</i>	99.5	NRPS [C-A-T-R]	1456		99		

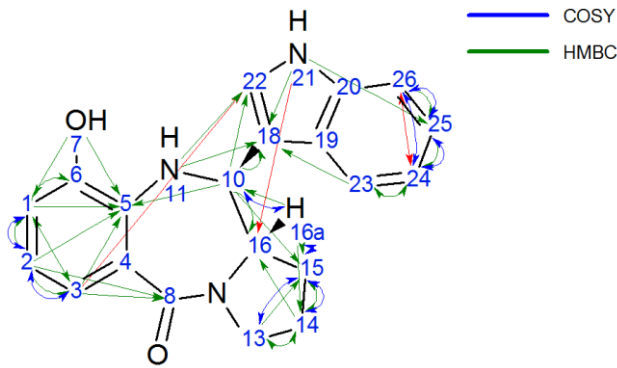
* Pairwise identity with Schneditz et al. Cluster

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

Primer	Sequence
TAR Fragments	
TILTAR-capture-Para-RP	AAAATCATACCTGACCTCCATAGCAGAAAGTCAAAGCCTTCGCCCTTGACGTTGGAGT
TILTAR-NRPS -RP	CCAGCTTATCCGATGACTTC
TILTAR-NRPS-FP	ATGACGCATTGAGCATATGT
TILTAR-Para-FP	AGGCTTTTGACTTTCTGCTA
TILTAR-Para-OL-NRPS-RP	GGACAGCCTTTAACTGATAGACATATGCTGAATGCGTCATGGGTATATCTCCTTCTTAAA
TILTAR-Para-RP	GGGTATATCTCCTTCTTAAA
TILTAR-Ptet-FP	TACGAACGGTATTAAGACCC
TILTAR-Ptet-OL-NRPS-FP	TATATTTGCGCAGTGATGGAGAAGTCATCGGATAAGCTGGTACGAACGGTATTAAGACCC
TILTAR-Ptet-OL-Tail-RP	GCTGTTGCCAGTTGGACTTATTGCCAGGAAATGATGTCATTTAGTGCCTCTTCTCTATCA
TILTAR-Ptet-RP	TTAGTGCCTCTTCTCTATCA
TILTAR-Tail-noP-FP	ATGACATCATTTCTGGCAA
TILTAR-Tail-RP	TTATTCGTCTAAAGGCCAGT
Colony PCR	
Check-KProm-ara-NRPS-FP	TTGCATCAGACATTGCCGTCAC
Check-KProm-ara-NRPS-FP	AATTCCTGCTTCTTTGTTGCGT
Check-KProm-Ende-FP	TTGCGATGCCAAAGATTTGCT
Check-KProm-Ende-RP	TTTCGCCACCACTGATTTGAGC
Check-KProm-Start-FP	ACGATACCTGAGTATTTCCACAGT
Check-KProm-Start-RP	GCTGCTGGCGATAAATCTGCTT
Check-KProm-Tet-Tail-FP	AGCACATCTAAAACCTTTAGCGT
Check-KProm-Tet-Tail-RP	CTTCGGGCCGGTTGACTGAATC

Table S 3.3: NMR data for tilivalline

C				H				2D			
#	Atom	Label	Shift	XHn	Label	Shift	H Multiplicity	COSY	H HMBC	C HMBC	
1	14	C12	21.647	CH2	H 13	1.887	m	15, 14, 15	15, 14, 15, 13, 16a	15, 13, 16	
2	15, 14	C12	21.647	CH2	H 14	1.700	m	14	15, 13, 16a	15	
3	15	C13	29.842	CH2	H 15	1.702	br s	14, 13, 16a	14, 13, 16a, 10	14	
4		C13	29.842	CH2	H 16	1.528	m	15, 14, 16a	15, 14, 13, 16a, 10	14, 13, 16	
5	13	C11	47.109	CH2	H 12	3.556	m	13	14	14, 15	
6	13	C11	47.109	CH2	H 11	3.693	m	15, 13	14		
7	16a	C10	58.950	CH	H 10	4.187	m	15, 10	14, 10, 21	14, 15, 10	
8	10	C9	60.394	CH	H 9	4.734	d (9.17)	16a	16a	15, 16, 2, 18, 22, 5	
9	23	C1	111.370	CH	H 3	7.391	br s	24	24	2, 18, 24	
10	1	C7	114.560	CH	H 7	6.809	dd (7.61, 1.47)	2	2, 3, 7	3, 5, 6	
11	2, 18	C8	115.962	C, CH	H 8	6.571	m	1, 3	10, 23, 21, 11	1, 3, 5, 8	
12	26	C6	118.555	CH	H 2	7.403	m	24, 25		25	
13	24	C2	118.555	CH	H 6	6.938	m	25, 26	25, 23	23	
14		C20	119.758	C				2	2		
15	25	C5	121.001	CH	H 5	7.092	m	24, 26	26, 21	24	
16	3	C4	121.244	CH	H 4	7.207	m	2	2, 1	1, 22, 5, 8	
17	22	C3	123.298	CH	H 1	7.391	m		10, 3, 11		
18		C16	124.827	C					24, 23, 11		
19	5	C17	134.301	C					10, 2, 1, 3, 7		
20		C15	136.099	C					25, 23, 11		
21		C19	144.734	C					2, 21, 7		
22	6	C18	145.334	C					1		
23	8	C14	165.960	C					2, 3		
24	7			OH	H 20	9.710	s			1, 5	
25					H 18	3.694	m	14		14, 15	
26	21			NH	H 19	4.991	s			16, 2, 18, 25	
27	11			NH	H 17	11.123	br s			2, 18, 22	



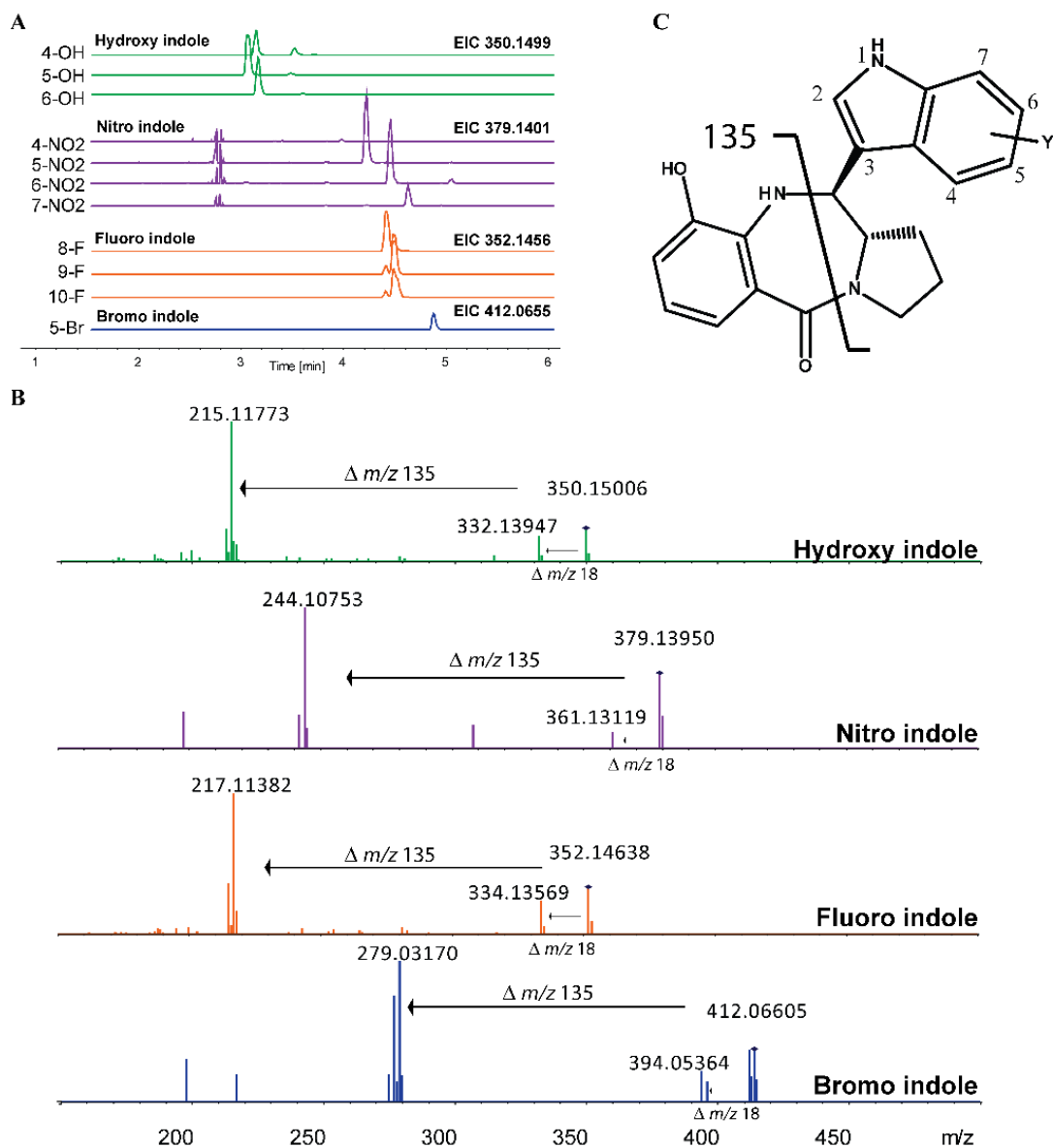


Figure S 3.1: LC-MS/MS data of tilivalline derivatives derived by supplementing indole derivatives to the heterologous expression system. A Listed EICs of the expected $[M+H]^+$ -ions with a m/z width of 0.005 to prove the incorporation of the respective indole derivatives. B Exemplary fragmentation 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.

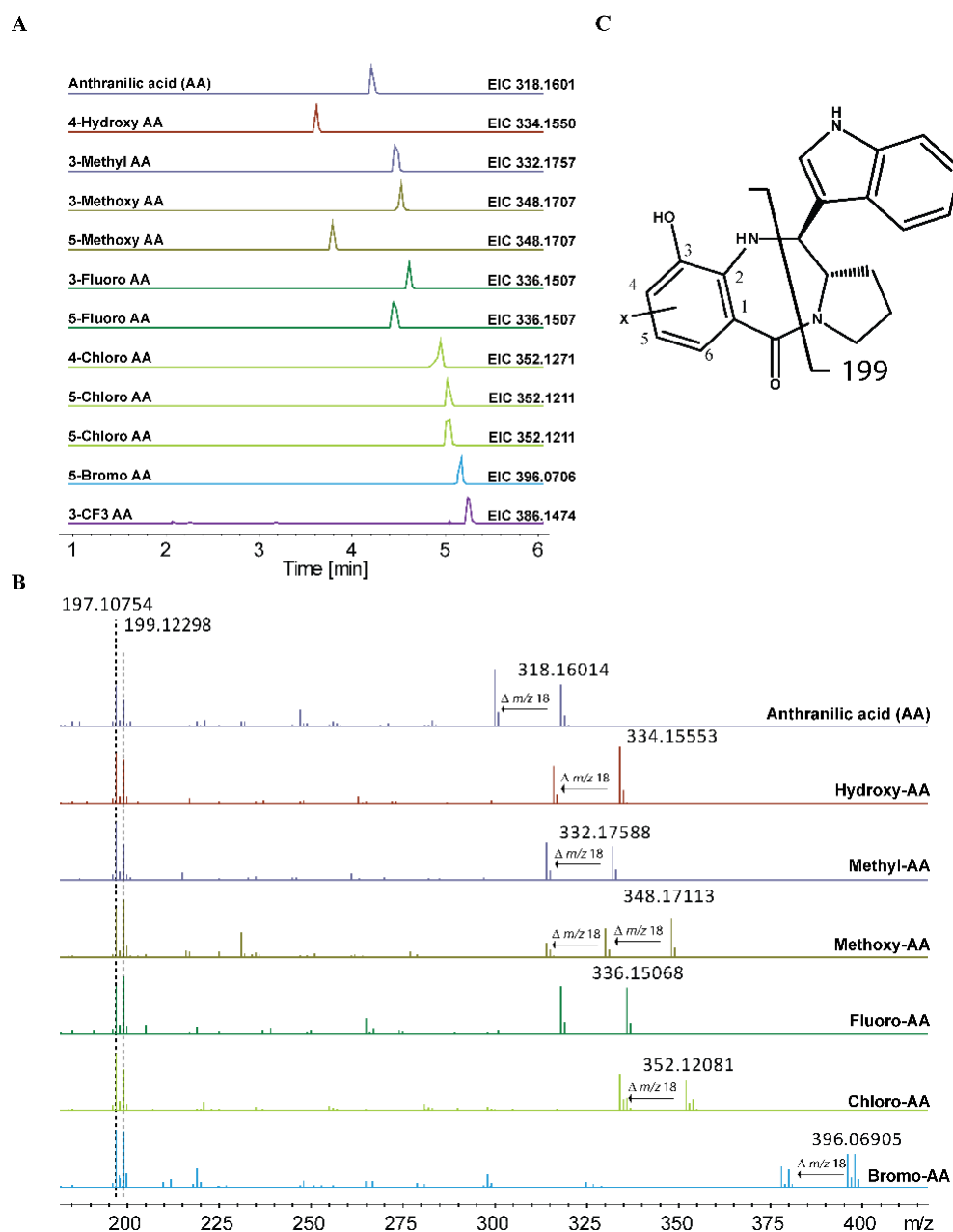


Figure S 3.2: LC-MS/MS data of tilivalline derivatives derived by supplementing anthranilic acid derivatives to the heterologous expression system. **A** Listed EICs of the expected $[M+H]^+$ -ions with a m/z width of 0.005 to prove the incorporation of the respective anthranilic acid derivatives. **B** 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.

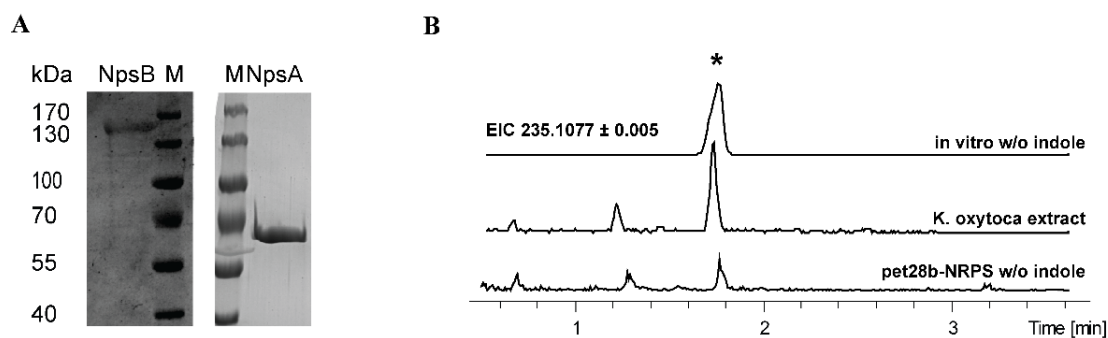


Figure S 3.3: **A** SDS-PAGE analysis of NpsA (163 kDa) and NpsB (56 kDa). **B** PBD-precursor formation can be detected in the in vitro reconstitution of NpsA, ThdA and NpsB when indole is omitted. Low amounts can also be detected in the *K. oxytoca* isolate #6 extract and the *E. coli* BL21 (DE3) pet28b-NRPS fermentation in M9 minimal media without indole.

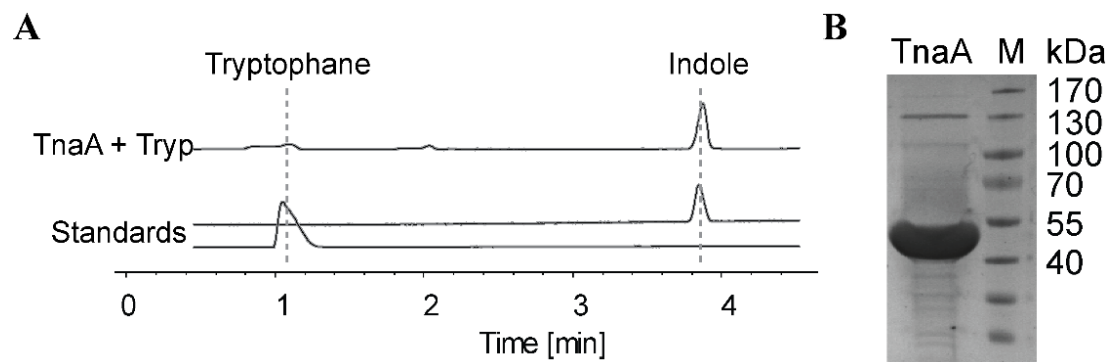


Figure S 3.4: **A** In vitro conversion of L-tryptophane to indole by the *K. oxytoca* isolate #6 tryptophanase TnaA is indicated by LC analysis (200-600nm). L-Tryptophane and indole were used as a standard.

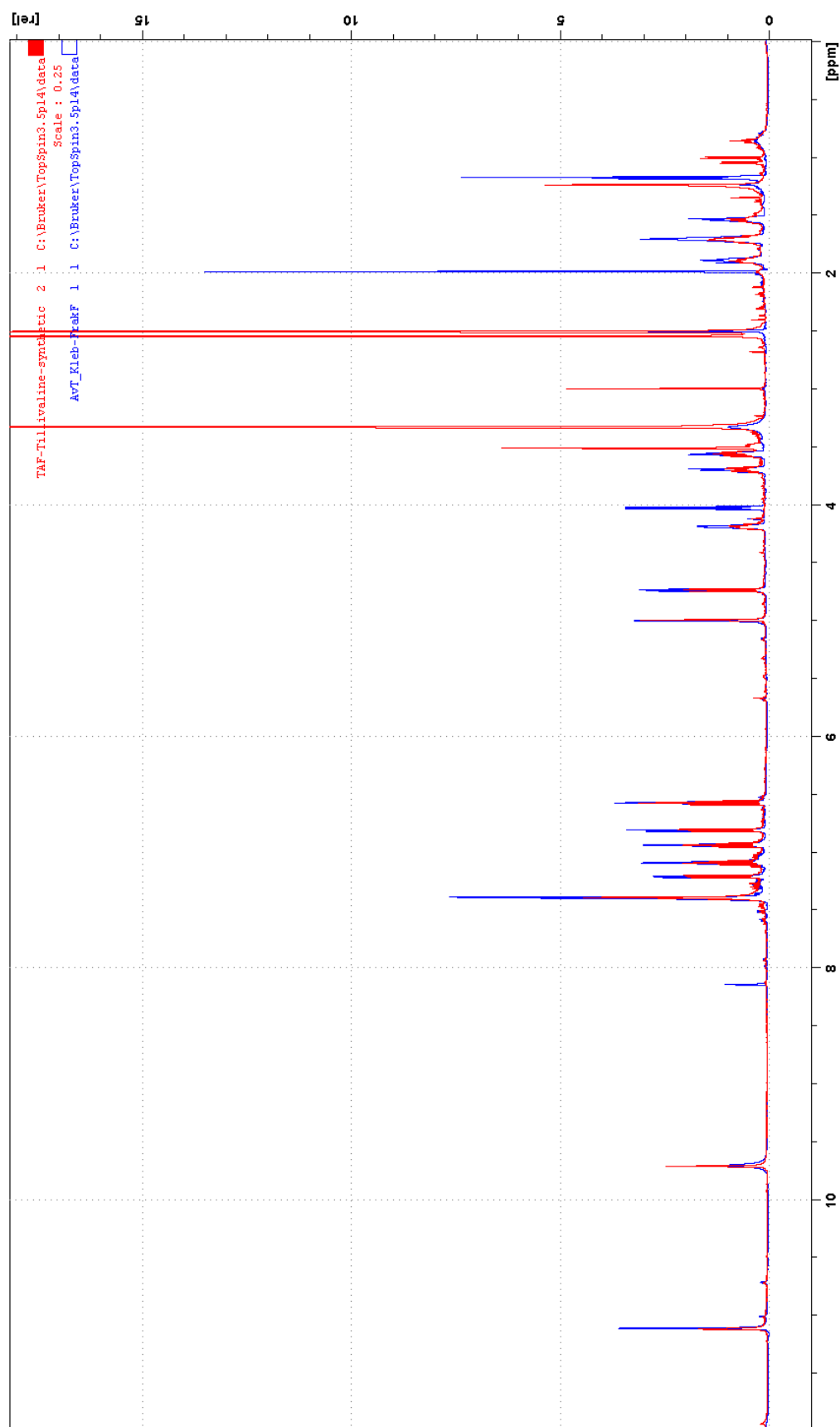
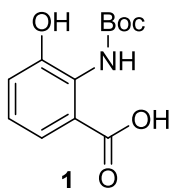


Figure S 3.5: Comparison of ^1H NMR (500 MHz, DMSO) shifts of synthetic and isolated tilivalline to confirm the stereochemistry.

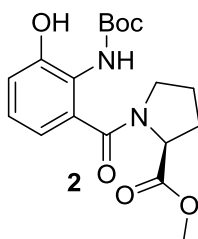
3.6.1 Synthetic Procedures

- 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 (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 : formic acid :: 1:1:0.1 R_f 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) C₁₂H₁₅NO₅ [M+H]⁺calculated for 254.1023, found 254.1027..

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

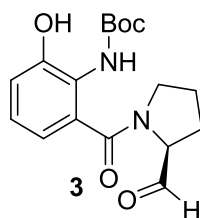


2-*N*-*boc*-amino-3-hydroxybenzoic acid (**1**) (50 mg, 0.18 mmol), HATU (68.4 mg, 0.18 mmol) and HOBt (5 mg, 0.03 mmol) was dissolved in anhydrous DMF (1 ml) under nitrogen atmosphere. DIPEA (61 μL, 0.36 mmol) was added to the mixture and the solution was left stirring at room temperature for 5 minutes under nitrogen atmosphere. A mixture of L-proline methyl ester hydrochloride (30 mg, 0.18 mmol) and DIPEA (61 μL, 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) δ 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) δ 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.

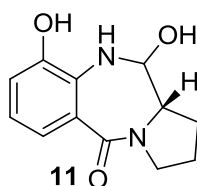
▪ S-1- (2'-*N*-*boc*-amino-3'-hydroxybenzoyl)pyrrolidine-2-carbaldehyde (**7**):



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 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-3'-hydroxybenzoyl)pyrrolidine-2-carbaldehyde (**7**).

^1H NMR (500 MHz, CDCl_3) δ 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) δ 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 $C_{17}H_{22}N_2O_5$ $[M+H]^+$ calculated for 235.10775, found 235.10774.

▪ Tillivalline:

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; see methods chapter).

HRMS (ESI, +ve) calculated for $C_{20}H_{20}N_3O_2$ $[M+H]^+$ calculated for 334.1550, found 334.15493

3.7 References

- [1] Mohr, N. & Budzikiewicz, H.: Tilivalline, a new pyrrolo[2, 1-c][1,4] benzodiazepine metabolite from klebsiella. *Tetrahedron* **1982**, 38, 147–152, DOI: 10.1016/0040-4020(82)85058-8
- [2] Schneditz, G. *et al.*: Enterotoxicity of a nonribosomal peptide causes antibiotic-associated colitis. *Proc. Natl. Acad. Sci.* **2014**, 111, 13181–13186, DOI: 10.1073/pnas.1403274111
- [3] Gerratana, B.: Biosynthesis, synthesis, and biological activities of pyrrolobenzodiazepines. *Med. Res. Rev.* **2012**, 32, 254–293, DOI: 10.1002/med.20212
- [4] Shioiri, T. *et al.*: Structure-cytotoxicity relationship of tilivalline derivatives. *Anticancer. Drug Des.* **1995**, 10, 167–76,
- [5] Minami, J., Okabe, A., Shiode, J. & Hayashi, H.: Production of a unique cytotoxin by *Klebsiella oxytoca*. *Microb. Pathog.* **1989**, 7, 203–211, DOI: 10.1016/0882-4010(89)90056-9
- [6] Higaki, M., Chida, T., Takano, H. & Nakaya, R.: Cytotoxic component(s) of *Klebsiella oxytoca* on HEP-2 cells. *Microbiol. Immunol.* **1990**, 34, 147–51,
- [7] Darby, A. *et al.*: Cytotoxic and pathogenic properties of *Klebsiella oxytoca* isolated from laboratory animals. *PLoS One* **2014**, 9, e100542, DOI: 10.1371/journal.pone.0100542
- [8] Beaugerie, L. *et al.*: *Klebsiella oxytoca* as an agent of antibiotic-associated hemorrhagic colitis. *Clin. Gastroenterol. Hepatol.* **2003**, 1, 370–6, DOI: 10.1053/S1542-3565(03)00183-6
- [9] Högenauer, C. *et al.*: *Klebsiella oxytoca* as a causative organism of antibiotic-associated hemorrhagic colitis. *N. Engl. J. Med.* **2006**, 355, 2418–26, DOI: 10.1056/NEJMoa054765
- [10] Chow, J., Tang, H. & Mazmanian, S. K.: Pathobionts of the gastrointestinal microbiota and inflammatory disease. *Curr. Opin. Immunol.* **2011**, 23, 473–480, DOI: 10.1016/j.coi.2011.07.010
- [11] Li, W., Chou, S., Khullar, A. & Gerratana, B.: Cloning and characterization of the biosynthetic gene cluster for tomaymycin, an sjg-136 monomeric analog. *Appl. Environ. Microbiol.* **2009**, 75, 2958–2963, DOI: 10.1128/AEM.02325-08
- [12] Li, W., Khullar, A., Chou, S., Sacramo, A. & Gerratana, B.: Biosynthesis of sibiromycin, a potent antitumor antibiotic. *Appl. Environ. Microbiol.* **2009**, 75, 2869–2878, DOI: 10.1128/AEM.02326-08
- [13] Finking, R. & Marahiel, M. A.: Biosynthesis of Nonribosomal Peptides. *Annu. Rev. Microbiol.* **2004**, 58, 453–488, DOI: 10.1146/annurev.micro.58.030603.123615
- [14] Fischbach, M. A. & Walsh, C. T.: Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic machinery, and mechanisms. *Chem. Rev.* **2006**, 106, 3468–3496, DOI: 10.1021/cr0503097
- [15] Hur, G. H., Vickery, C. R. & Burkart, M. D.: Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat. Prod. Rep.* **2012**, 29, 1074–1098, DOI: 10.1039/c2np20025b
- [16] Weber, T. *et al.*: antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **2015**, 43, W237-43, DOI: 10.1093/nar/gkv437

- [17] Takeo, M. *et al.*: Mechanism of 4-nitrophenol oxidation in *Rhodococcus* sp. Strain PN1: characterization of the two-component 4-nitrophenol hydroxylase and regulation of its expression. *J. Bacteriol.* **2008**, 190, 7367–7374, DOI: 10.1128/JB.00742-08
- [18] Bilyk, O., Sekurova, O. N., Zotchev, S. B. & Luzhetskyy, A.: Cloning and Heterologous Expression of the Grecoacycline Biosynthetic Gene Cluster. *PLoS One* **2016**, 11, e0158682, DOI: 10.1371/journal.pone.0158682
- [19] Han, T. H., Lee, J.-H., Cho, M. H., Wood, T. K. & Lee, J.: Environmental factors affecting indole production in *Escherichia coli*. *Res. Microbiol.* **2011**, 162, 108–116, DOI: 10.1016/j.resmic.2010.11.005
- [20] Ku, S. Y., Yip, P. & Howell, P. L.: Structure of *Escherichia coli* tryptophanase. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2006**, 62, 814–823, DOI: 10.1107/S0907444906019895
- [21] Liu, G. Y. *et al.*: *Staphylococcus aureus* golden pigment impairs neutrophil killing and promotes virulence through its antioxidant activity. *J. Exp. Med.* **2005**, 202, 209–15, DOI: 10.1084/jem.20050846
- [22] Chang, C.-I., Chelliah, Y., Borek, D., Mengin-Lecreulx, D. & Deisenhofer, J.: Structure of tracheal cytotoxin in complex with a heterodimeric pattern-recognition receptor. *Science* **2006**, 311, 1761–4, DOI: 10.1126/science.1123056
- [23] Holden, V. I. & Bachman, M. A.: Diverging roles of bacterial siderophores during infection. *Metallomics* **2015**, 7, 986–95, DOI: 10.1039/c4mt00333k
- [24] Neumann, W., Gulati, A. & Nolan, E. M.: Metal homeostasis in infectious disease: recent advances in bacterial metallophores and the human metal-withholding response. *Curr. Opin. Chem. Biol.* **2017**, 37, 10–18, DOI: 10.1016/j.cbpa.2016.09.012
- [25] Song, Y. *et al.*: Inhibition of staphyloxanthin virulence factor biosynthesis in *Staphylococcus aureus*: In vitro, in vivo, and crystallographic results. *J. Med. Chem.* **2009**, 52, 3869–3880, DOI: 10.1021/jm9001764
- [26] Lamb, A. L.: Breaking a pathogen's iron will: Inhibiting siderophore production as an antimicrobial strategy. *Biochim. Biophys. Acta* **2015**, 1854, 1054–70, DOI: 10.1016/j.bbapap.2015.05.001
- [27] Maura, D., Ballok, A. E. & Rahme, L. G.: Considerations and caveats in anti-virulence drug development. *Curr. Opin. Microbiol.* **2016**, 33, 41–46, DOI: 10.1016/j.mib.2016.06.001
- [28] Zollner-Schwetz, I. *et al.*: The Toxin-Producing Pathobiont *Klebsiella oxytoca* Is Not Associated with Flares of Inflammatory Bowel Diseases. *Dig. Dis. Sci.* **2015**, 60, 3393–3398, DOI: 10.1007/s10620-015-3765-y
- [29] Podschun, R. & Ullmann, U.: *Klebsiella* spp. as nosocomial pathogens: Epidemiology, taxonomy, typing methods, and pathogenicity factors. *Clin. Microbiol. Rev.* **1998**, 11, 589–603, DOI: 0893-8512/98/\$04.00?0
- [30] Kumar, R. & Lown, J. W.: Recent developments in novel pyrrolo[2,1-c][1,4]benzodiazepine conjugates: synthesis and biological evaluation. *Mini Rev. Med. Chem.* **2003**, 3, 323–39, DOI: 10.2174/1389557033488097
- [31] Mantaj, J., Jackson, P. J. M., Rahman, K. M. & Thurston, D. E.: From Anthramycin to

- Pyrrolobenzodiazepine (PBD)-Containing Antibody-Drug Conjugates (ADCs). *Angew. Chem. Int. Ed. Engl.* **2017**, 56, 462–488, DOI: 10.1002/anie.201510610
- [32] Puvvada, M. S. *et al.*: Inhibition of bacteriophage T7 RNA polymerase in vitro transcription by DNA-binding pyrrolo[2,1-c][1,4]benzodiazepines. *Biochemistry* **1997**, 36, 2478–2484, DOI: 10.1021/bi952490r
- [33] Gaytán, M. O., Martínez-Santos, V. I., Soto, E. & González-Pedrajo, B.: Type Three Secretion System in Attaching and Effacing Pathogens. *Front. Cell. Infect. Microbiol.* **2016**, 6, 129, DOI: 10.3389/fcimb.2016.00129
- [34] Green, E. R. & Meccas, J.: *Virulence Mech. Bact. Pathog. Fifth Ed.* (American Society of Microbiology, **2016**). 4, 215–239, DOI: 10.1128/microbiolspec.VMBF-0012-2015
- [35] Finking, R. *et al.*: Aminoacyl adenylate substrate analogues for the inhibition of adenylation domains of nonribosomal peptide synthetases. *ChemBioChem* **2003**, 4, 903–906, DOI: 10.1002/cbic.200300666
- [36] Davis, T. D. *et al.*: Design, synthesis, and biological evaluation of α -hydroxyacyl-AMS inhibitors of amino acid adenylation enzymes. *Bioorg. Med. Chem. Lett.* **2016**, 26, 5340–5345, DOI: 10.1016/j.bmcl.2016.09.027
- [37] Ferreras, J. A., Ryu, J.-S., Di Lello, F., Tan, D. S. & Quadri, L. E. N.: Small-molecule inhibition of siderophore biosynthesis in *Mycobacterium tuberculosis* and *Yersinia pestis*. *Nat. Chem. Biol.* **2005**, 1, 29–32, DOI: 10.1038/nchembio706
- [38] Edwards, L. J.: The hydrolysis of aspirin. A determination of the thermodynamic dissociation constant and a study of the reaction kinetics by ultra-violet spectrophotometry. *Trans. Faraday Soc.* **1950**, 46, 723, DOI: 10.1039/tf9504600723
- [39] Simpson, J. T. *et al.*: ABySS: A parallel assembler for short read sequence data. *Genome Res.* **2009**, 19, 1117–1123, DOI: 10.1101/gr.089532.108
- [40] Kearse, M. *et al.*: Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, 28, 1647–1649, DOI: 10.1093/bioinformatics/bts199
- [41] Gasteiger, E. *et al.*: *Proteomics Protoc. Handb.* (Walker, J. M.) (Humana Press, **2005**). 112, 571–607, DOI: 10.1385/1-59259-890-0:571
- [42] Kunwar, A. *et al.*: Interaction of a curcumin analogue dimethoxycurcumin with DNA. *Chem. Biol. Drug Des.* **2011**, 77, 281–7, DOI: 10.1111/j.1747-0285.2011.01083.x

Chapter 4 – Fruiting Bodies

Homospermidine Lipids: A compound class specifically formed during fruiting body formation of *Myxococcus xanthus* DK1622

Michael Hoffmann^{1,2}, David Auerbach^{1,2}, Fabian Panter², Thomas Hoffmann²,
Pieter C. Dorrestein³, Rolf Müller²

Manuscript ready for submission

¹These authors contributed equally to this work

²Department of Microbial Natural Products (MINS), Helmholtz Institute for Pharmaceutical Research Saarland (HIPS) - Helmholtz Centre for Infection Research (HZI) and Institute for Pharmaceutical Biotechnology, Saarland University, 66123 Saarbrücken, Germany

³Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, CA 92093, USA

Contributions

Author's effort:

The author significantly contributed to the conception of the study, designed and performed experiments, evaluated and interpreted resulting data. The author contributed substantially to the development of a feasible workflow to determine discriminating metabolites depending on the type of cultivation. Cultures on agar were prepared, harvested and extracted by the author. LC-MS analyses and evaluation of the resulting data as well as structure elucidation of Cmp-552 by NMR analysis were conducted by the author. Furthermore, the author contributed equally to conceiving and writing of the manuscript.

Contribution by David Auerbach:

The author significantly contributed to the conception of the study, designed and performed experiments, evaluated and interpreted resulting data. The author contributed substantially to the development of a feasible workflow to determine discriminating metabolites depending on the type of cultivation. Cultures on agar were prepared, harvested and extracted by the author. Dereplication the observed LC-MS data, semi-preparative separation and purification of Cmp-552 was performed by the author. Furthermore, the author contributed equally to conceiving and writing of the manuscript.

Contribution by others:

Fabian Panter performed the isolation and structure elucidation of the derivatives from *Myxococcus* MCy9290. In addition, he performed the feeding studies and analyzed the genomic data to elucidate the biosynthesis. All MALDI-MSI experiments described were performed and evaluated by Thomas Hoffmann. The project was supervised by Pieter C. Dorrestein and Rolf Müller, who also contributed to conceiving and proofreading of the manuscript.

4 Fruiting Bodies

4.1 Abstract

The fascinating ability of myxobacteria to form multicellular spore filled fruiting bodies under starvation condition was widely studied as a model for cooperative microbial behavior. The potential of a life cycle induced changes of secondary metabolism as a means to discover novel natural compounds remains largely underexplored. We therefore studied the model organism *Myxococcus xanthus* DK1622 under submersed and solid cultivation conditions to find putatively life-cycle related compounds by applying statistical analysis on analytical data. Utilizing the advantageous characteristics of LC-MS, LC-MS/MS and MALDI-MSI allowed identifying compounds unambiguously associated to myxobacterial fruiting bodies. Our screening effort resulted in the purification and structure elucidation of a novel compound from cultures that had undergone sporulation. A combination of molecular networking and targeted LC-MS/MS in conjunction with our in-house metabolomics database subsequently revealed alternative producers of the respective compound as well as a number of compounds belonging to the same structural class. Three further members of this compound class were isolated from an alternative producer and structurally elucidated by NMR. Insights into the biosynthesis of this novel compound class was gained by feeding of isotopically labeled substrates.

4.2 Introduction

Microbial secondary metabolism is a long-standing source of novel structural scaffolds used in medicinal drug discovery, especially concerning antibiotic agents¹. Among microbes, Myxobacteria are remarkably well suited to provide a variety of secondary metabolites². Details concerning these outstanding characteristics can be found in the literature³. Briefly, myxobacteria have an exceptional biosynthetic potential due to their large genomes harboring dozens of secondary metabolite encoding gene clusters⁴. The advances in genome sequencing technologies and bioinformatics tools allowed to assign biosynthetic gene clusters in a large number of complete genome sequences⁵⁻⁷ and the number of these clusters is often contrasted by a far lower quantity of known secondary metabolites^{5,8}. A common explanation for the “absence” of secondary metabolites is that gene clusters might be tightly regulated under laboratory growth conditions leading to low or zero basal expression levels due to missing environmental signals required for activation. A pronounced change in secondary metabolism in response to culturing conditions could be shown for a range of marine actinomycetes in terms of detected molecular networks⁹. For the myxobacterial model species *Myxococcus xanthus* DK1622, 5 compound families and 18 presumed

PKS/NRPS biosynthetic gene clusters are known, but combined transcriptomics and proteomics efforts indicated that 17 out of the 18 clusters were transcriptionally active. Detection of transcription of these clusters with unknown products (on the transcriptome level) is another clear indication, that the tight regulation of BGCs require a stimulus to raise the otherwise low basal levels of transcripts¹⁰. Even though a gene cluster might be functionally expressed, the detection and identification of the corresponding metabolite remains challenging. It is common sense and well documented in the literature, that instrumental setup as well as sample preparation have a major impact on the coverage of compounds^{9,11,12}. To harness the biosynthetic potential of bacteria, it is thus mandatory to have a sound analytical platform that allows identifying target compounds from metabolome data with the important additional demand of dereplication to avoid re-discovery of known compounds. More information regarding dereplication of metabolome data can be found in recent reviews^{13–17}. As argued above, discovery of novel natural products essentially requires an easy and reliable cultivation method. Over the course of the past decades, cultivation processes in shaking flask using liquid media with added adsorbents has become the method of choice for myxobacteria. It remains a matter of debate, though, whether this cultivation approach is ideal for soil-dwelling bacteria, which pass through various life cycle stages in their natural environment¹⁸. This is supported by the fact that many isolated myxobacterial species could not yet been successfully cultivated in liquid media. It can easily be imagined, that external stimuli required to activate certain biosynthetic gene clusters (e.g. interspecies cell-cell interaction) are missing in shake flask cultivation. Only limited data of studies aiming at the comparison of liquid and solid cultivation on a larger scale is available. It could be shown however, that a substantial difference between submerged and solid cultivation can be observed for a range of 30 *actinomycetes*, with only a minimal overlap of 6.8 %⁹. It is thus clear, that any study aiming to gain insights into the fruiting body formation on a metabolome level, requires an analysis technique that connects phenotypic changes to the metabolome to avoid false positive hits caused by differences in cultivation. Hence, we applied mass spectrometry imaging (MSI) as a tool to correlate the appearance of fruiting body specific metabolites for cultures grown on agar in a time dependent manner. In particular, we were interested in changes in metabolism during fruiting body formation to find presently unknown natural products as a follow-up study that addressed semi-quantitative changes of known metabolites¹⁹. We decided to use the well-studied *M. xanthus* DK1622 which shows favorably fast growth alongside reproducible fruiting body formation. We studied changes in the metabolome based on extracts of cultures after fruiting body formation grown on agar and cultures grown in suspension. Imaging mass spectrometry served as a tool to correlate the appearance of fruiting body specific metabolites for cultures grown on agar in a time dependent manner, to filter out false positives.

4.3 Results and Discussion

Fruiting body formation in myxobacteria was extensively studied on genome and transcriptome level, to gain a better understanding of the underlying regulatory and signaling mechanisms leading to cell aggregation and sporulation^{20–23}. Effects of fruiting body formation on the metabolome level to increase the structural diversity of experimentally accessible structures however is underexplored. We thus designed an experiment to assess the prospects of fruiting body formation for the discovery of novel secondary metabolites. As a first hypothesis-generating step, cultures grown in suspension were compared to sporulated cultures grown on a solid medium. Possible novel target compounds putatively connected to the myxobacterial life cycle were obtained by statistical comparison of five biological replicates grown in liquid medium and five biological replicates grown on agar (each biological replicate consisted of five agar plates). The resulting distinguishing features can be both attributed to the growth condition as well as to the life cycle, since the presence of vegetative cells on the extracted agar plates cannot be fully excluded. A MALDI mass spectrometry imaging (MALDI-MSI) approach allows achieving this distinction by correlation of putative candidate m/z -values with fruiting body formation observed on light microscopic images over a time course covering the development from vegetative cells to fruiting body formation.

4.3.1 Comparative Analysis of Submersed Cultures and Fruiting Bodies Grown on Agar

Comparison of both cultivation conditions first required a methodology for extraction of cultures grown on agar plates. Submersed cultures were treated following the well-established approach of cultivation in presence of an adsorber resin^{24–26}. To promote fruiting body formation, *M. xanthus* DK1622 was cultivated on a minimal agar medium as well as on cellulose membranes placed on top of agar plates containing minimal medium. Liquid cultures used to inoculate plates were washed with minimal medium prior to inoculation. Cultivation on membranes placed on the agar surface was dropped at an early stage due to poor growth characteristics and the lack of fruiting body formation. The presence of adsorber resin randomly spread over agar plates to increase the production of secondary metabolites to avoid feedback inhibition of production and self-toxicity was found to be non-beneficial. Small amounts of XAD16 beads on the agar surface favored the extraction of media compounds while signals related to bacterial growth were underrepresented as compared to plates without adsorber resin (Figure S 4.1). Apparently, production and extraction efficiency of secondary metabolites cannot be improved for solid cultivation by this approach. The workup of solid cultures was performed by scraping off cells and fruiting bodies using a glass microscope slide (Figure 4.1). Extraction of complete cultures including

agar media turned out to be inadequate, as the concentration of extracted media compounds is orders of magnitude higher than the compounds related to bacterial growth (data not shown).

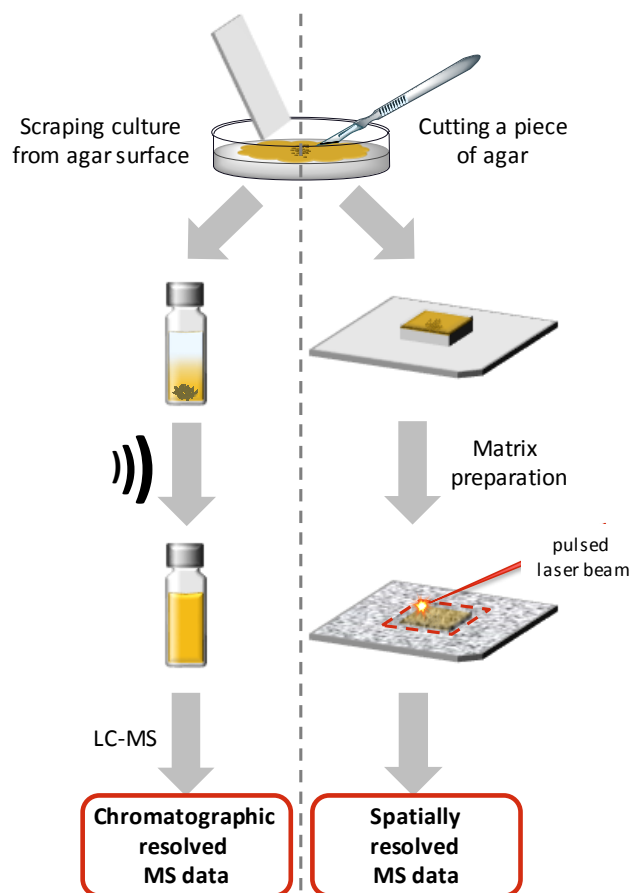


Figure 4.1: Overview of the two workflows applied for detection and verification of fruiting body related metabolites. To obtain chromatographic resolved MS data vegetative cells and fruiting bodies were scraped from the agar surface using a microscope slide, extracted with methanol, and measured using LC-MS. Spatially resolved MS data were received by following a MALDI-IMS approach. An agar piece including cells and fruiting bodies were transferred to a MALDI target, dried, matrix applied, and measured using by MALDI-ToF MS.

Extracts obtained from five biological replicates for both conditions (solid and submersed) were subsequently screened as two technical replicates by LC-MS to generate a set of m/z values that is putatively connected to changes in metabolism. The highly reduced background observed for extracts from agar plates allowed to highlight many signals exclusive to fruiting body samples by mere visual comparison of the base peak chromatograms (BPC) (Figure 4.2). A statistically sound list of candidates was generated using principal component analysis (PCA) in conjunction with t -tests to reveal significant differences in the metabolic profiles. Compounds found in at least 4 out of 5 biological replicates were considered as a significant hit, when no appearance could be detected in media blanks. A total of 62 discriminating molecular features (RT, m/z , intensity) were detected by this means (Figure 4.3).

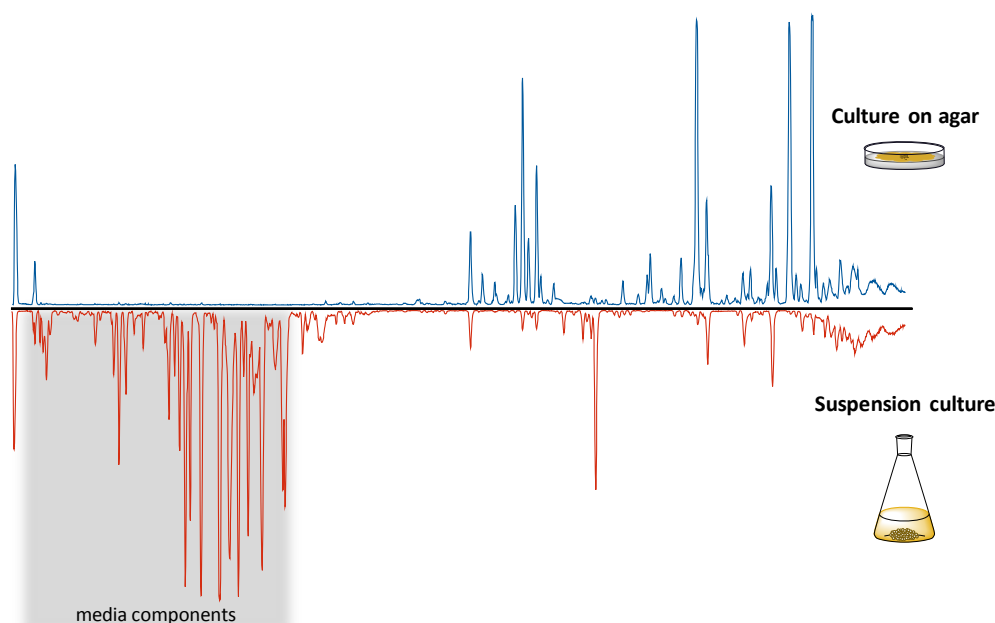


Figure 4.2: Comparative base peak chromatograms (150-2500 m/z) of an extract derived from *M. xanthus* DK1622 cultivation on agar plate (blue) and one derived from cultivation in liquid medium with adsorber resin (red). Time range with predominantly media component is marked in grey (based on Figure S 4.2).

Although cultivation of *M. xanthus* DK1622 on minimal agar over 5-7 days mainly results in the formation of fruiting bodies, the presence of vegetative cells cannot be excluded²⁷. A comparison to vegetative cells grown on agar would be required to further specify the obtained data, which can hardly be achieved using minimal medium, as the presence of fruiting bodies can also not be fully excluded. We thus regarded the set of 62 features as a hypothetical dataset that needed to be tested by a suitable additional analysis technique.

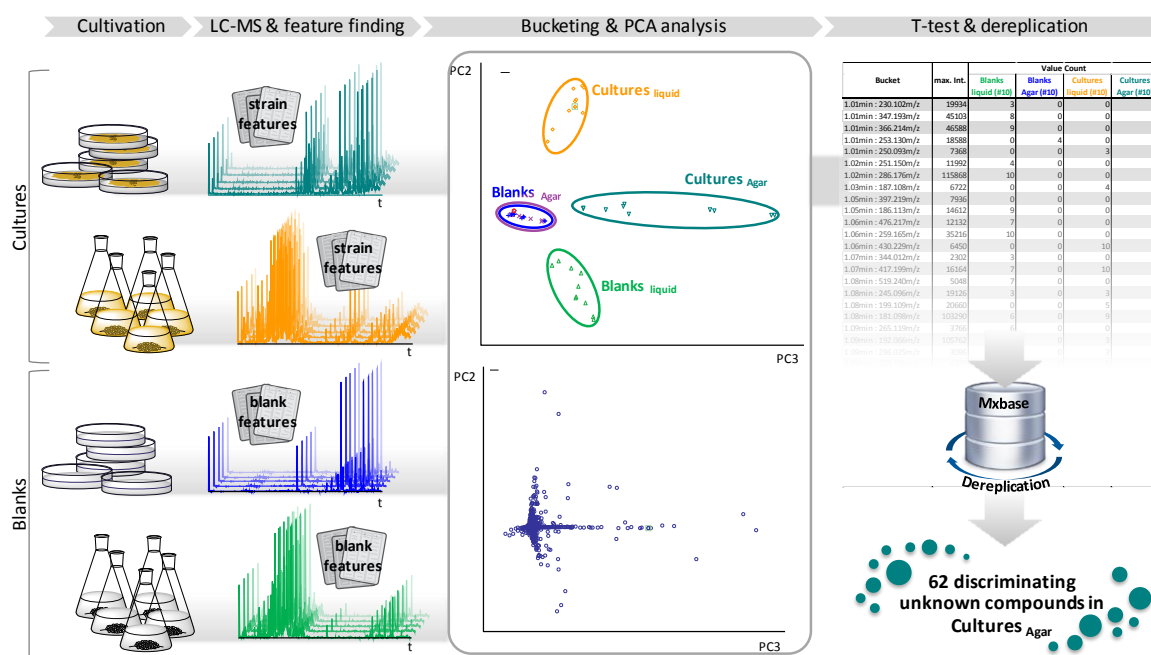


Figure 4.3: Workflow towards the detection of discriminating compounds, which are verifiably due to cultivation on solid media, possibly also fruiting body formation. Five replicates of culture as well as 5

replicates of blanks were prepared for both cultivation methods. Subsequently, all extracts were injected into the LC-MS twice and the molecular features are determined. Based on the molecular features a principle component analysis and a t-test was performed to reveal the discriminating peaks for cultures grown on agar. Dereplication of these data led to 62 compounds, which are directly related to cultivation on agar.

4.3.2 MALDI-MSI of *M. xanthus* DK1622 Grown on Minimal Medium

Mass spectrometry imaging of bacterial cultures became well accepted by the research community due to its inherent strength to provide complementary information to other well established molecular analysis workflows^{28,29}. Imaging mass spectrometry excels at providing spatially resolved molecular information in combination with light microscopic images allowing to correlate spatially resolved m/z values with visually observed phenotypic changes. MALDI-MSI was thus the method of choice to relate the 62 features obtained from LC-MS measurements with the development of fruiting bodies. The applied workflow is illustrated in Figure 4.1 while details regarding the method can be found in the literature³⁰.

Four time points covering different stages of the life cycle, including the fruiting body formation of *Myxococcus xanthus* DK1622 were analyzed for the presence of the 62 features in question. Mass spectral images were analyzed towards the presence of the 62 features as $[M+H]^+$, $[M+Na]^+$, $[M+NH_4]^+$, and $[M+K]^+$ ions. Surprisingly, only two out of the 62 features could be detected in mass spectral images as sodium and potassium adducts. One of the substances was specifically found in areas covered by fruiting bodies (590.4669 m/z) while the other was characteristic for the surrounding cells (724.3979 m/z). The mass spectral image of m/z 724.3979 shown in Figure 4.4 illustrates the necessity of an analysis technique that connects phenotypic changes and molecular information, as this compound is clearly specific for vegetative cells grown on agar. The sodiated species with m/z 590.4669 corresponds to the compound eluting at 15.9 min with a mass to charge ratio of 552.5104 (Cmp-552) in LC-MS runs. The predominance of cationization observed for these two compounds is likely owed to the presence of salts in the agar medium – a phenomenon that is well known for biological samples of all sorts^{31,32}. Possible causes for the inability to detect the remaining 60 compounds found by LC-MS are numerous but can only be speculated about. Some of the analytes might simply be present in amounts below the detection limit, which can be additionally impaired by ion suppression as often encountered in MALDI measurements of biological samples^{33,34}. Another important influence on the outcome of MALDI-MSI is the efficiency of analyte extraction, as low efficiency results in a lack of analyte-doped matrix crystals. Several of the distinguishing m/z values found by LC-MS are likely specific for fruiting bodies as well, since they share similar MS/MS patterns and belong to the same molecular network (for details see below). Unambiguous evidence, however, could not be obtained in these cases.

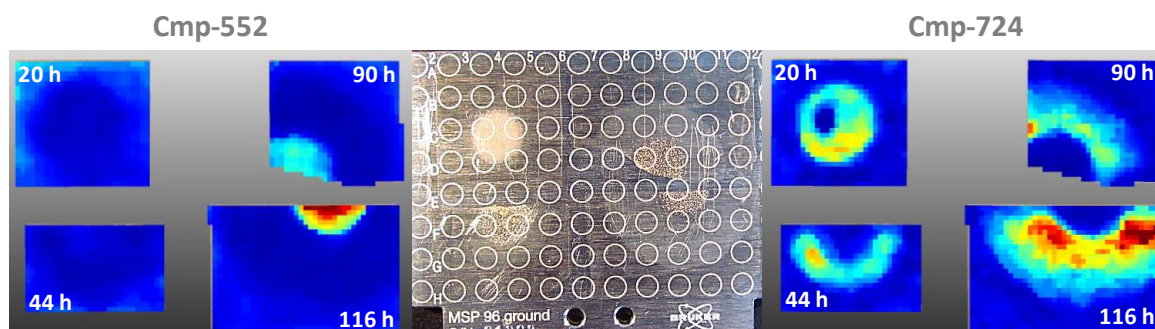


Figure 4.4: MALDI-MSI data of Cmp-552 and Cmp-686 (both detected as $[M+K]^+$) in dependency of cultivation time. The picture in the middle shows the respective samples on a MALDI target prior to matrix application.

Correlation of the MALDI images of m/z 590.4669 and the respective light microscopic images revealed that the biosynthesis of Cmp-552 starts at a later stage of fruiting body formation since the signal intensity was the highest after 116 hours cultivation although fruiting bodies could already be observed after 44 hours (Figure 4.4). We next set out to purify Cmp-552 as the only fruiting body specific compound that could be confirmed by MALDI-imaging.

4.3.3 Structure Elucidation of Cmp-552 and Identification of Related Compounds

Purification and structure elucidation of microbial compounds discovered by imaging mass spectrometry is a challenging task. Only one metabolite obtained from MSI data of solid cultivation could be successfully purified and structurally confirmed by NMR^{28,35,36}. Owing to the limited biomass per agar plate, a substantial amount of petri dishes was necessary to provide sufficient crude extract for subsequent purification attempts.

We could isolate 300 μg of Cmp-552 by semi-preparative LC from a methanolic extract of fruiting bodies of *Myxococcus xanthus* DK1622 grown on roughly 500 Petri dishes. The structure of Cmp-552 was elucidated by NMR and turned out as a sym-homospermidine core structure with fatty acyl residues attached to the secondary and the two primary amine groups (Figure 4.5). Since the compound reminds of a lipid, it was named as homospermidine lipid 552. As a descriptive shorthand notation based on the lipidmaps nomenclature the following form is suggested: HS(iso5:0/5:0/iso15:0), where HS stands for homospermidine and the three substituents are given as fatty acyl moieties on the two primary amines of sym-homospermidine followed by the substituent on the secondary amine. According to the lipidmaps classification system, the compound belongs to the subclass of N-acyl amines (FA0802)³⁷.

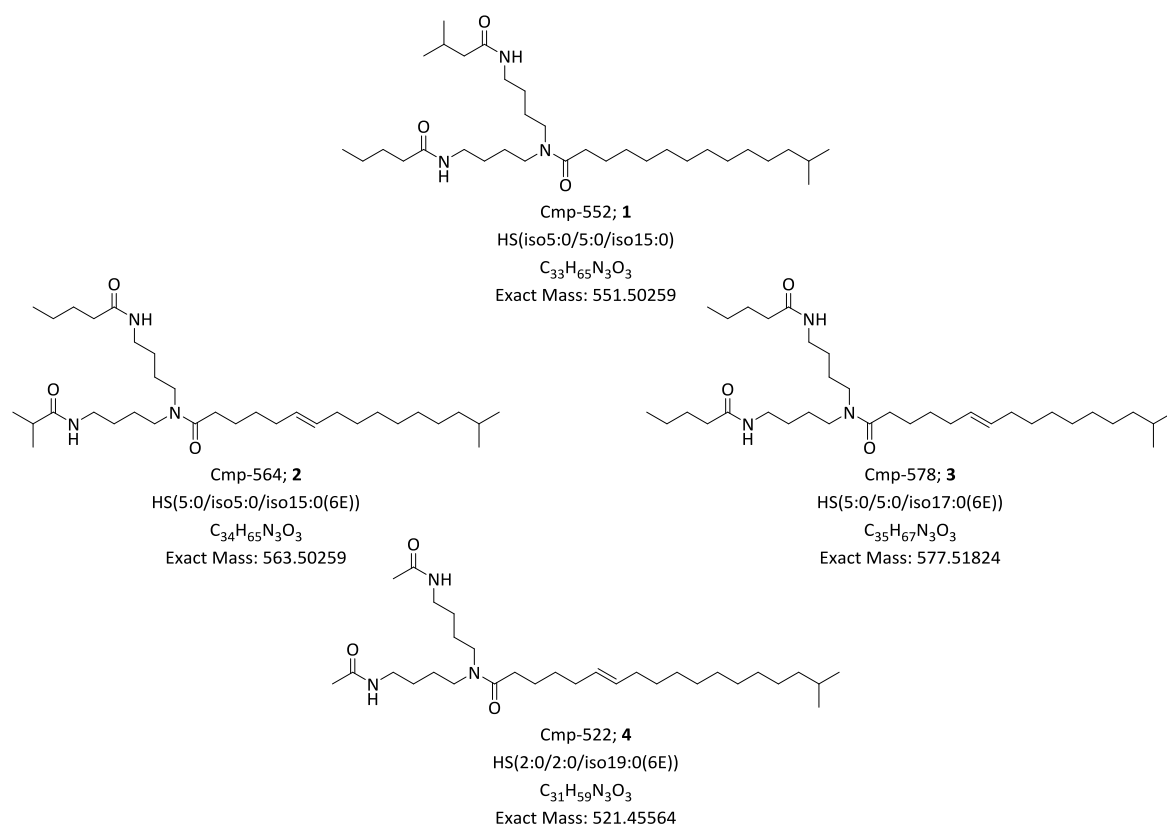


Figure 4.5: Chemical structures of the homospermidine lipids (HS). Compounds were named based on the lipidmaps guidelines.

Use of our in-house metabolome database allowed to identify *Myxococcus* MCy9290 as a strain that seemed to form the homospermidine lipid in liquid cultures¹⁴. Since the results shown above proof the relation between the biosynthesis of Cmp-552 and the fruiting body formation of *M. xanthus* DK1622, we assume that fundamental differences in regulation lead to an activation of the respective gene cluster during cultivation of MCy9290 in liquid medium, especially as no other strain belongig to the genus *Myxococcus* was identified as producer in liquid culture. The lipid like structure of HS(iso5:0/5:0/iso15:0) led us to the expectation, that further derivatives might be found among the 64 putatively fruiting body related compounds as well as in extracts of MCy9290. A powerful and very well received tool to identify such structurally related compounds based on MS/MS similarity is the molecular networking workflow provided within the Global Natural Product Social Molecular Networking (GNPS) platform³⁸. Molecular networking is perfectly suited to identify compound classes and is increasingly implemented in dereplication and discovery campaigns³⁹.

Extracts of five biological replicates of MCy9290 and fruiting bodies of *M. xanthus* DK1622 were obtained from liquid and solid cultivation, respectively. Each sample was measured twice and results were subjected to PCA analysis with subsequent t-tests as described above. The resulting culture specific features were used for the acquisition of targeted MS/MS data as an input for spectral networking. The resulting MS/MS networks are shown in Figure 4.6.

As liquid cultures are more accessible to prepare larger amounts of extract, we cultivated *Myxococcus* MCy9290 in a larger scale to isolate further derivatives of the novel compound class. We were able to isolate three additional derivatives differing mainly by alkyl chain length while showing the same bioactivities (Figure 4.5 & Table 4.1).

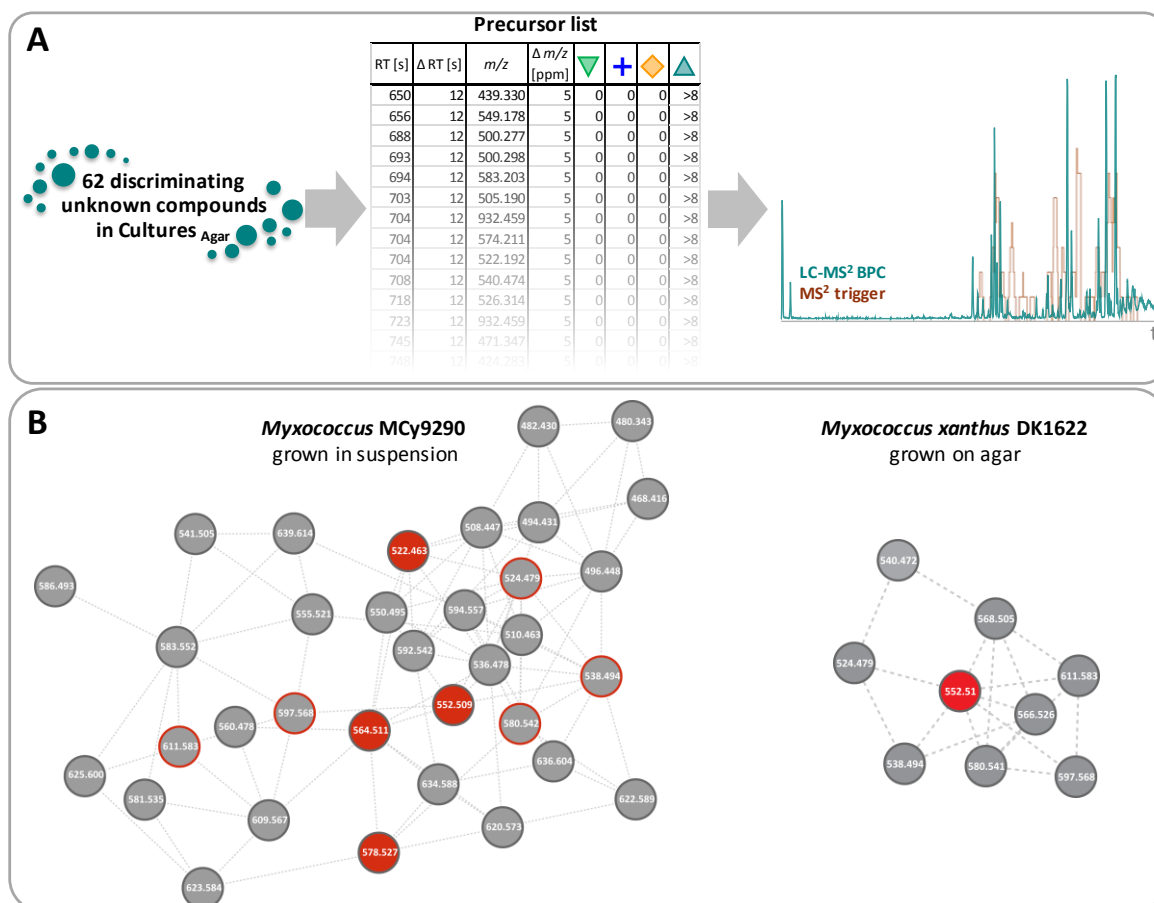


Figure 4.6: **A** Precursor directed MS/MS based on PCA and t-test results exemplified by the 62 discriminating unknown compounds in agar grown cultures. **B** Cmp-552 including MS/MS networks derived from extracts of *Myxococcus* MCy9290 (grown in suspension) and *Myxococcus xanthus* DK1622 (grown on agar). Red coloured nodes indicating derivatives, which were isolated, and structures elucidated. Nodes bordered in red marking derivatives of Cmp-552 which were also found in extracts of *Myxococcus xanthus* DK1622.

4.3.4 Bioactivity and Biosynthesis of Homospermidine-Lipids

Testing compounds 2-4 for bioactivity against a panel of microorganisms and cell lines revealed cytotoxicity values in the range of 20 μ M and moderate activity against some Gram-positive bacteria ranging from 14-116 μ M (Table 4.1). The purpose of fruiting body formation is to endure times of low nutrient supply and it seems reasonable that the organism tries to defend the fruiting body against any competitors or potential threats and by that increases chances of survival. Lipids might also serve as carbon storage and additional potential biological functions are possible.

Table 4.1: Minimum inhibitory concentration of homospermidine lipids **1-4**. As observed activities do not differ for the individual derivatives, values are not listed separately.

Test organism	MIC / IC ₅₀
<i>Staphylococcus aureus</i>	16 µg/ml
<i>Bacillus subtilis</i>	16 µg/ml
<i>Micrococcus luteus</i>	8 µg/ml
<i>Mucor hiemalis</i>	64 µg/ml
Chinese hamster ovary CHO-K1	10-15 µg/ml

The biosynthesis of the homospermidine lipids was elucidated based on the assumption that the homospermidine-lipids derive from sym-homospermidine and thereby originate from putrescine and arginine, respectively (Figure S 4.8). (Kuttan and Radhakrishnan, 1972) For this reason, [¹³C₆,¹⁵N₄]-L-arginine was fed to a liquid culture of *Myxococcus* MCy9290 and the obtained extracts were checked for the corresponding mass shift of the compounds. LC-MS analysis confirmed the incorporation of two arginine moieties, substantiating the homospermidine origin of this substance class (Figure S 4.6). Since the genomic sequence data for *M. xanthus* DK1622 is available, we searched for the responsible biosynthetic gene cluster. A genetic island containing the homospermidine synthase, a *N*-acyltransferase and a fatty acid acyltransferase was identified by searching for known homospermidine synthases using BLAST. (Altschul, 1990) More details regarding the potential biosynthesis can be found in the supplementary material.

4.4 Conclusion

In this work we present the first secondary metabolomics study devoted to the influence of fruiting body formation in a myxobacterium. Application of a broad scope of measurement and analysis techniques allowed to unambiguously identify, purify and structurally elucidate a novel compound that is directly connected with fruiting body formation of *M. xanthus* DK1622. We present one of the very few cases, where imaging mass spectrometry leads to the isolation of a compound obtained from cultures grown on solid medium. The data obtained from mass spectral images impressively show the necessity of applying a molecular imaging technique that allows to connect spatial information with visually observed phenotypic changes to establish a clear-cut evidence for the localization of a candidate *m/z* value on fruiting bodies of *M. xanthus* DK1622. Further database and spectral networking based analysis resulted in the identification of a new compound family and allowed us to isolate three more derivatives from an alternative and deregulated producer capable to form homospermidine lipids in liquid culture. Data obtained from spectral networking indicates that other members related to HS(iso5:0/5:0/iso15:0) are also specific for fruiting bodies, but confirmation by MALDI MSI was not possible due to the inability to detect the respective *m/z* values.

Owing to the observed cytotoxicity and activity against several Gram-positive bacteria these compounds might be part of the defense strategy of the bacterium or play a role in carbon storage. As the activities of the compounds were rather nonselective, it can be assumed that they are part of a more general defense against competitors. We believe that we have shown a very promising workflow that will not only lead to the identification of novel natural compounds but also offers unique opportunities to gain fundamental insights into molecular processes taking place during fruiting body formation.

4.5 Methods

4.5.1 Cultivation and Extract Preparation

4.5.1.1 Cultures grown in suspension

Pre-cultures of *Myxococcus xanthus* DK1622 and *Myxococcus* MCy9290 were grown at 30 °C for 3 days in 25-50 mL CTT medium (2 g/L $\text{MgSO}_4 \times 7\text{H}_2\text{O}$, 10 g/L casitone, 1.2 g/L TRIS, 136 mg/L K_2HPO_4 ; pH 7.6) and VY/2 medium (5 g/L bakers yeast, 50 mg/L $\text{CaCl}_2 \times 2\text{H}_2\text{O}$, 1.2 g/L HEPES, 10 g/L starch; pH 7.0), respectively. Main cultures were inoculated by the well-grown pre-cultures (2 %) and cultivated at 30 °C for 7 days. Two days after inoculation, XAD-16 absorber resin was added to the culture (1 %). Extracts were prepared by extracting the resin including cells with methanol, evaporation of solvent and re-dissolving in methanol. 1 μL of crude extract was analyzed by LC-hrMS. For purification purposes of *Myxococcus* MCy9290, a total of 7.5 L suspension culture including XAD-16 was grown and extracted in accordance to the method described above.

4.5.1.2 Cultures grown on agar

All cultures were cultivated in standard Petri dishes on HS agar medium (1 g/L $\text{MgSO}_4 \times 7\text{H}_2\text{O}$, 1 g/L KNO_3 , 1.5 g/L casiton, 2 g/L TRIS, 4 g/L Glucose, 8 mg/L Fe-EDTA, 75 mg/L $\text{CaCl}_2 \times 2\text{H}_2\text{O}$, 6.25 mg/L K_2HPO_4 , 2 % agar; pH 7.2) at 30 °C for 7 days. Pre-cultures, cultivated in liquid CTT medium, were washed with CF medium (2 g/L $\text{MgSO}_4 \times 7\text{H}_2\text{O}$, 150 mg/L casitone, 1.2 g/L TRIS, 174 mg/L K_2HPO_4 , 0.2 mg/L $(\text{NH}_4)\text{SO}_4$, 1 g/L sodium pyruvate, 2 g/L sodium citrate; pH 7.6) and 10-fold concentrated. Petri dishes were inoculated with 150 μL /plate of the concentrated pre-culture. Extracts were prepared by scraping off the cells and fruiting bodies from the agar surface, extraction with methanol, evaporation and re-dissolving in methanol.

In total, approximately 500 Petri dishes were prepared, harvested, and extracted to obtain sufficient amounts of Cmp-552.

4.5.2 LC-hrMS Methods

All measurements of crude extracts were performed on a Dionex Ultimate 3000 RSLC comprising a high pressure gradient pump (HPG-3400RS) with a 150 μL mixing chamber. All LC connections are realized by 0.13 μm stainless steel capillaries before the column. The method is based on separation with BEH C18, 100 x 2.1 mm, 1.7 μm dp column (Waters, Eschborn, Germany). One μL sample is separated by a linear gradient from (A) H_2O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 600 $\mu\text{L}/\text{min}$ and 45 °C. The gradient is 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 are recorded by a DAD in the range from

200 to 600 nm with 2 nm width. The LC flow is split to 75 $\mu\text{L}/\text{min}$ before entering the maXis 4G hr-ToF mass spectrometer (Bruker Daltonics, Germany) using the Apollo ESI source. A fused silica (f.s.) capillary (20 cm, 100 μm ID, 360 μm OD) coming from the mass spectrometer's switching valve is connected to an Upchurch PEEK microtight T-junction. Another f.s. capillary of 20 cm and 75 μm ID is connected to the inlet of the mass spectrometer whereas an f.s. capillary of 10 cm and 100 μm ID is directed to waste. The split is approximately 1:8 based solely on the f.s. capillary ratio. The real split is even more than 1:8 as the effect of the ESI needle is not taken into account here. The ion source parameters are: capillary, 4000 V; endplate offset, -500 V; nebulizer, 1 bar; dry gas, 5 L/min; dry gas temperature, 200 °C. Ion transfer parameters were: funnelRF, 350 Vpp; multipoleRF, 400 Vpp; quadrupole ion energy, 5 eV @ low m/z 120. Collision cell is set to 5 eV with a collisionRF of 1100 Vpp; transfer time, 110 μs ; pre puls storage, 5 μs in full scan mode. Mass spectra are acquired in centroid mode ranging from 150 – 2500 m/z at a 2 Hz scan rate in full scan positive ESI mode. Each measurement is started with the injection of a 20 μL plug of basic sodium formate solution, which is introduced by a loop that is connected to the system's 6-port switching valve. The resulting peak is used for automatic internal m/z calibration. In addition, three lock masses (622, 922, 1222 m/z) are present in the ion source all the time and allows recalibration of each spectrum.

Fractions obtained by semi-preparative purification were measured on the same setup using a 50 mm instead of a 100 mm column of the same series. Hence, only the gradient was slightly adjusted. Initiation 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.

Statistical evaluation of the data was performed using *ProfileAnalysis* V2.1 (Bruker Daltonics, Germany)

4.5.3 MALDI-MSI Method

Experimental details regarding strain cultivations, sample preparation, and instrumental parameters can be found in literature³⁰. All MS imaging data were evaluated and visualized with *SCiLS Lab* 2015 (Bruker Daltonics, Germany).

4.5.4 MSⁿ Experiments and Molecular Networking

All LC-MS/MS measurements were based on the already described 18 min gradient separation using the U3000-maXis 4G setup. Fragmentation spectra were achieved either by auto-MS/MS measurements or by targeted precursor selection. Selected precursor list (SPL) were generated by taking into account the dereplication and PCA results respectively. Fixed MS/MS parameters were an m/z -adjusted precursor isolation width and fragmentation energy. Values are interpolated in between the set points [300 m/z / 4 m/z / 30 eV], [600 m/z / 6 m/z / 35 eV], [1000 m/z / 8 m/z / 45

eV], and [2000 m/z / 10 m/z / 55 eV]. The precursor intensity threshold is always set to 5000. All methods are set up with an active precursor exclusion, which puts precursors after two spectra on an exclusion list. Each precursor m/z remains there for 0.2 minutes.

Molecular networking was performed using the GNPS (Global Natural Products Social Molecular Networking) portal and data were visualized with *Cytoscape* V 3.5.0^{38,40}.

MSⁿ fragmentation experiments were conducted by direct infusion MS using a solarix XR (7T) FT-ICR mass spectrometer (Bruker Daltonics, Germany) equipped with an 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. While MS/MS spectra were obtained by collision-induced dissociation (CID) in the collision cell, all further fragmentations were obtained by isolating the respective precursor in the ICR cell and applying sustained off resonance irradiation (SORI) CID. Parameters were individually adjusted for each measurement.

4.5.5 LC Method for Semi-Preparative Purification

4.5.5.1 Homospermidine-lipid HS(iso5:0/5:0/iso15:0) – *Myxococcus xanthus* DK1622

Purification was carried out on a Thermo Dionex (Germering, Germany) Ultimate 3000 low pressure gradient system, equipped with SR3000 solvent rack, LPG-3400SD pump module, WPS-3000TSL autosampler, Knauer (Berlin, Germany) Jetstream column oven, DAD-3000 photodiode array detector with a cell volume of 10 μ L and an AFC 3000 fraction collector. LC separations were performed using 50 μ L injections on a Phenomenex Kinetex Biphenyl column (10 x 250 mm, 5 μ m) at 40°C and a flow rate of 5 mL min⁻¹ with the following gradient conditions: 0-55 min, 5-95 % B; 55-59 min, 95 % B; 59-59.5 min, 95-5 % B; 59.5-70 min, 5 % B. Mobile phases were (A) water + 0.1 % formic acid and (B) acetonitrile + 0.1 % formic acid. Fractions of 18 s width were collected from 20-55 min. The resulting fractions were combined, evaporated to dryness and reconstituted in 1 mL of methanol.

4.5.5.2 Homospermidine-lipid-derivatives - *Myxococcus* MCy9290

First separation was done on a Sephadex LH20 gravitation column and follow-up separation was done on Waters preparative HPLC using a Waters X-Bridge C-18, 150 x 19 mm, 5 μ m dp column. Separation was performed using (A) water + 0.1% formic acid and (B) acetonitrile + 0.1% formic acid and a flow rate of 5 mL min⁻¹ with the following gradient: 0-3 min, 5 % B; 3-7 min, 5-55 % B; 7-32 min, 55-98 % B; 32-35 min, 98 % B; 35-36 min, 98-5 % B; 36-40 min, 5 % B.

Residual impurities were removed by an additional semi-preparative chromatography step on a Phenomenex Synergi Fusion-RP, 250 x 10 mm, 4 μ m dp column. Separation was performed using

the same eluent as described above and a flow rate of 5 mL min⁻¹ by using the following gradient: 0-3 min, 5 % B; 3-7 min, 5-65 % B; 7-27 min, 55-85 % B; 27-28 min, 85-98 % B; 28-31 min, 98 % B; 31-32 min, 98-5 % B; 32-36 min, 5 % B.

4.5.6 Bioactivity testing

4.5.6.1 Antimicrobial Assay

All microorganisms were handled according to standard procedures and were obtained from the German Collection of Microorganisms and Cell Cultures (*Deutsche Sammlung für Mikroorganismen und Zellkulturen*, DSMZ) or were part of our internal strain collection. For microdilution assays, overnight cultures of Gram-positive bacteria in Müller-Hinton broth (0.2 % beef infusion, 0.15 % corn starch, 1.75 % casein peptone; pH 7.4) were diluted in the growth medium to achieve a final inoculum of ca. 10⁶ cfu/ml. Serial dilutions of homospermidine lipids were prepared from MeOH stocks in sterile 96-well plates. The cell suspension was added and microorganisms were grown on a microplate shaker (750 rpm, 37°C, 16 h). Growth inhibition was assessed by visual inspection and given MIC (minimum inhibitory concentration) values are the lowest concentration of antibiotic at which no visible growth was observed.

4.5.6.2 Cytotoxic activity

Cell lines were obtained from the German Collection of Microorganisms and Cell Cultures (*Deutsche Sammlung für Mikroorganismen und Zellkulturen*, DSMZ) or were part of our internal collection and were cultured under conditions recommended by the depositor. Cells were seeded at 6 x 10³ cells per well of 96-well plates in 180 µl complete medium and treated with homospermidine lipids in serial dilution after 2 h of equilibration. Each compound was tested in duplicate as well as the internal solvent control. After 5 d incubation, 20 µl of 5 mg/ml MTT (thiazolyl blue tetrazolium bromide) in PBS was added per well and it was further incubated for 2 h at 37°C. The medium was then discarded and cells were washed with 100 µl PBS before adding 100 µl 2-propanol/10 N HCl (250:1) in order to dissolve formazan granules. The absorbance at 570 nm was measured using a microplate reader (Tecan Infinite M200Pro), and cell viability was expressed as percentage relative to the respective methanol control. IC₅₀ values were determined by sigmoidal curve fitting.

4.6 Supplemental Figures & Tables

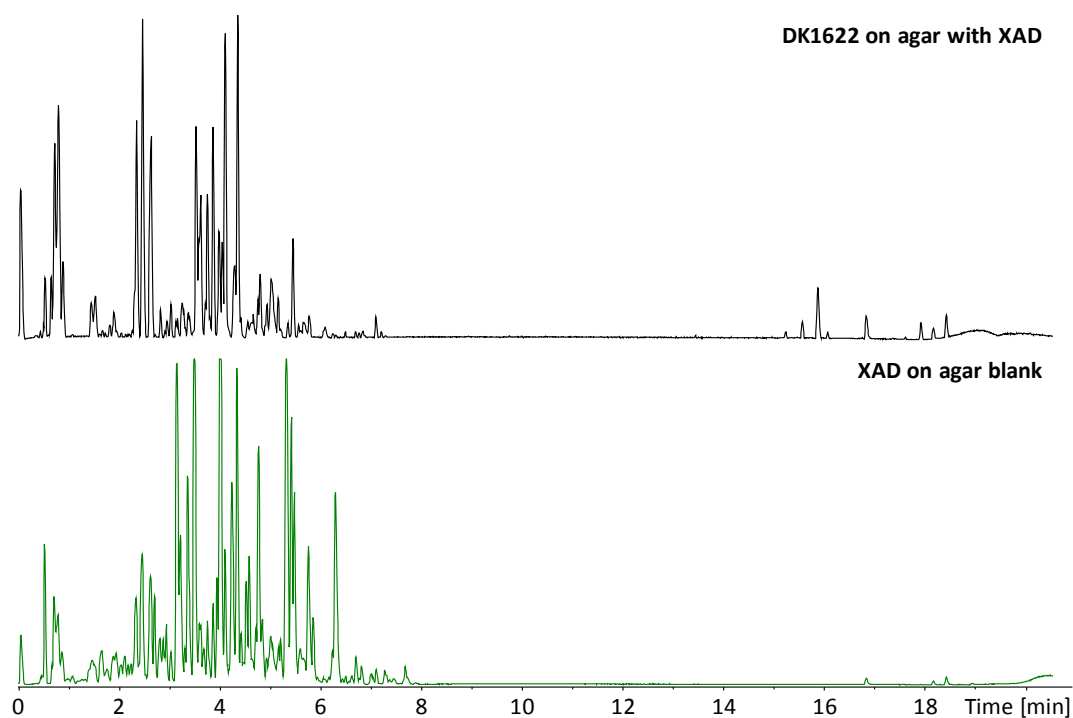


Figure S 4.1: Base peak chromatograms (150-2500 m/z) showing the effect of XAD addition regarding the cultivations on agar.

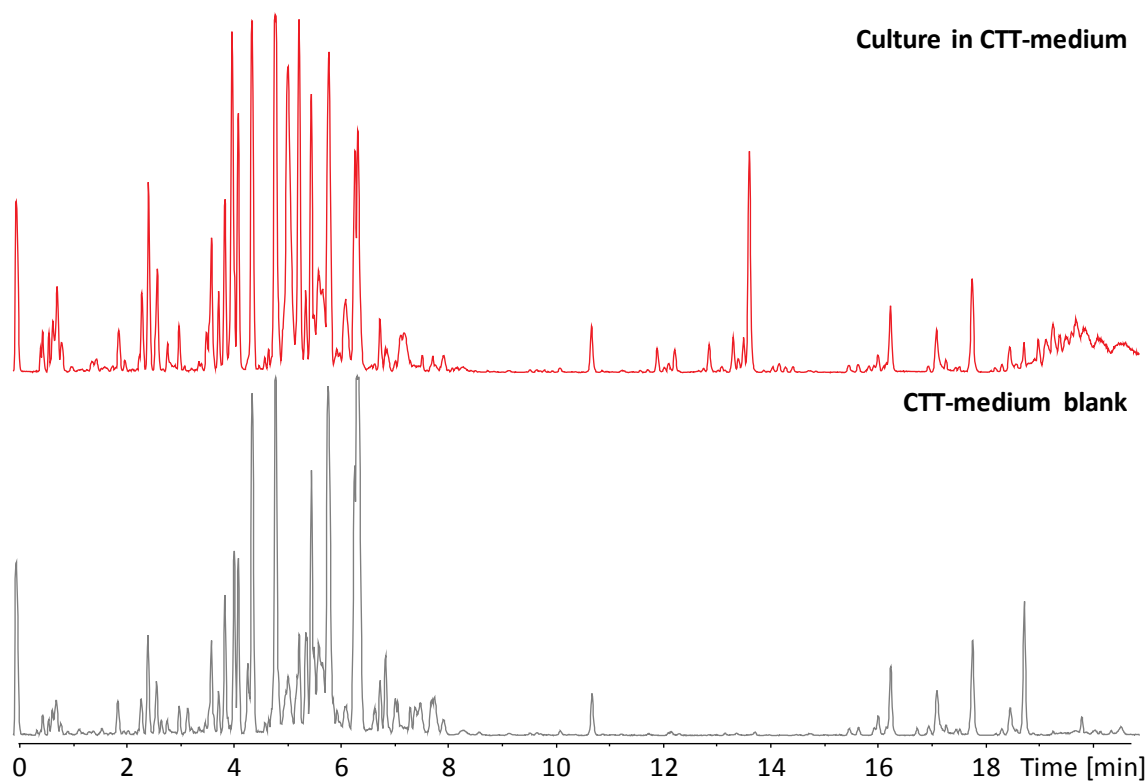


Figure S 4.2: Comparative base peak chromatograms (150 – 2500 m/z) of a *M. xanthus* DK1622 extract originating from suspension culture in CTT medium and the respective blank.

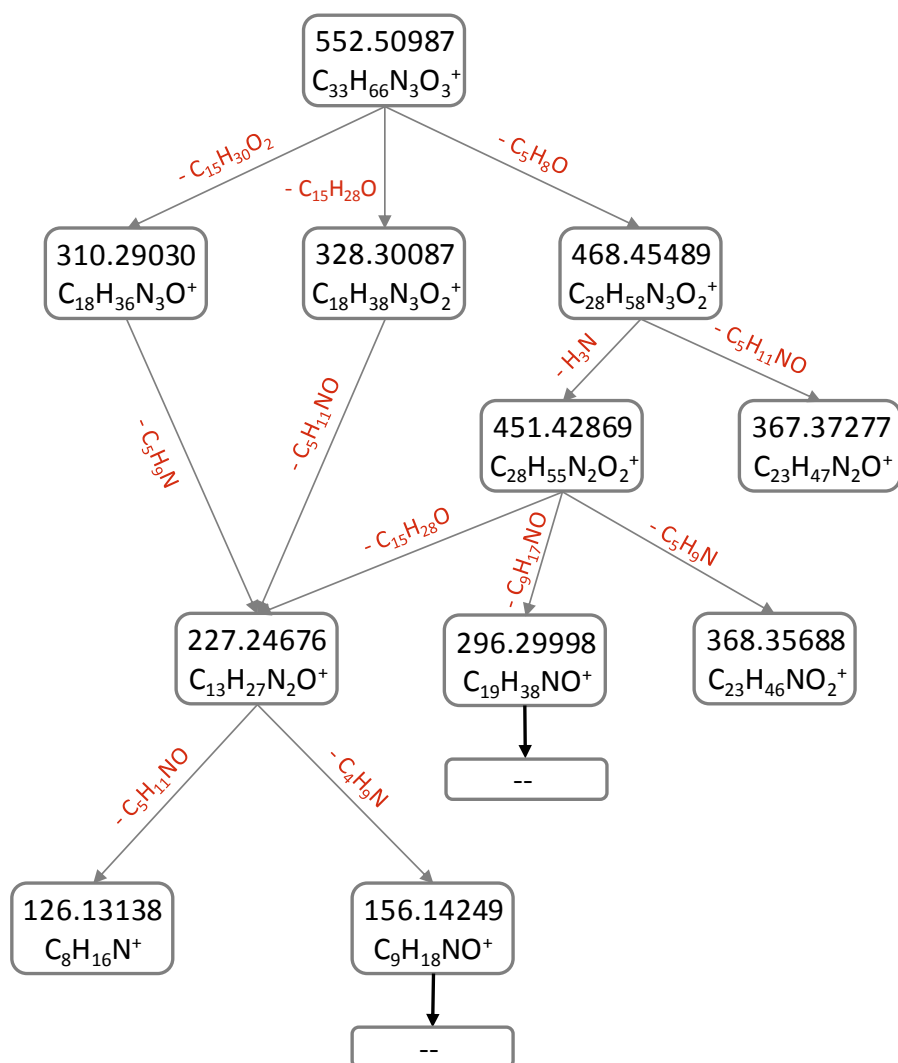


Figure S 4.3: MS fragmentation tree of Cmp-552 based on MSⁿ experiments performed on FT-ICR.

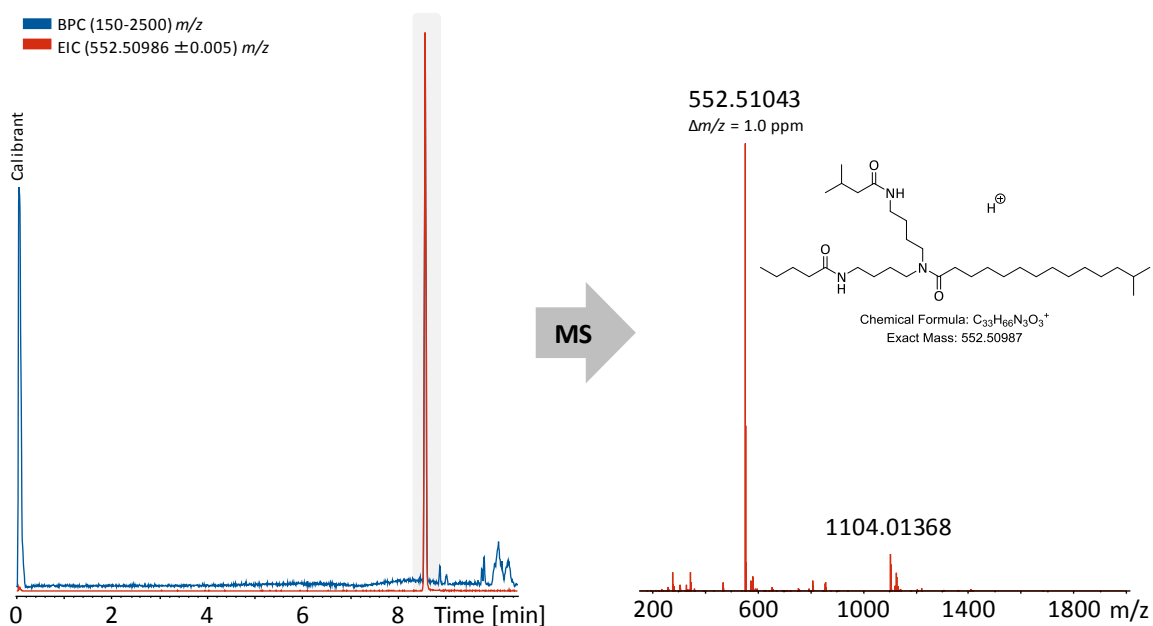


Figure S 4.4: Overlaid base peak chromatogram and extracted ion chromatogram (552.50986 ± 0.005 m/z) of the Cmp-552 containing fraction indicated sufficient purity for NMR analysis. The respective fraction was received by semi-preparative isolation effort of *M. xanthus* DK1622 extracts from cultivation on agar.

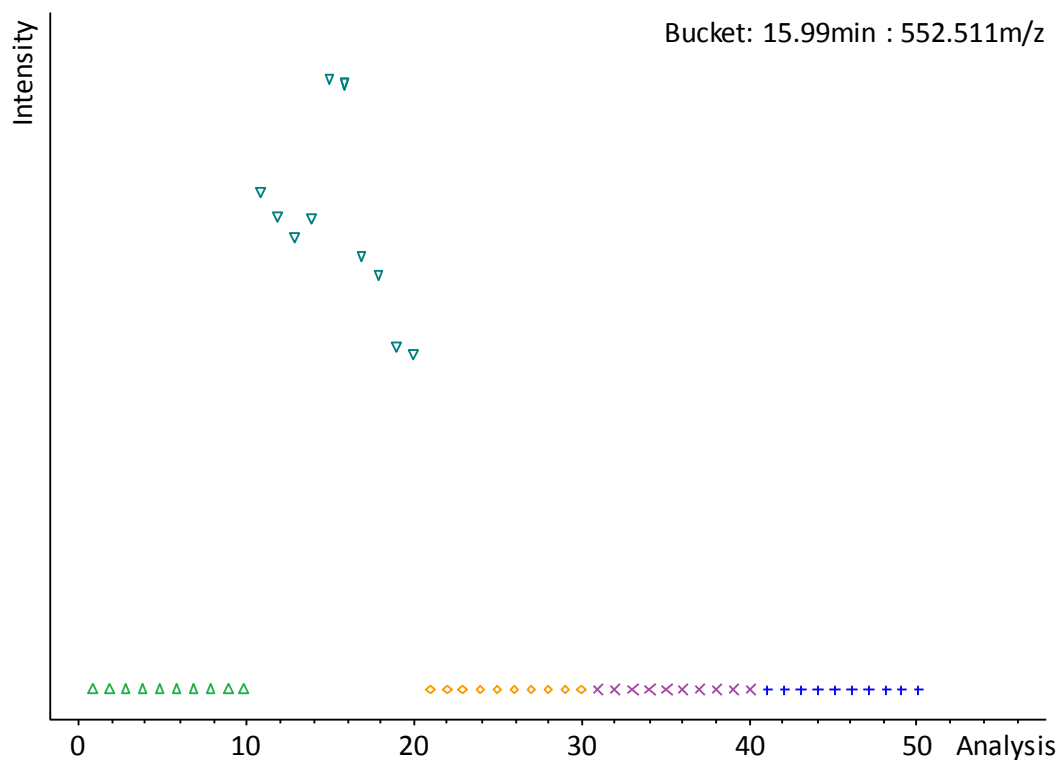


Figure S 4.5: Bucket statistic for Cmp-552 containing bucket. Color code is explained in figure 2

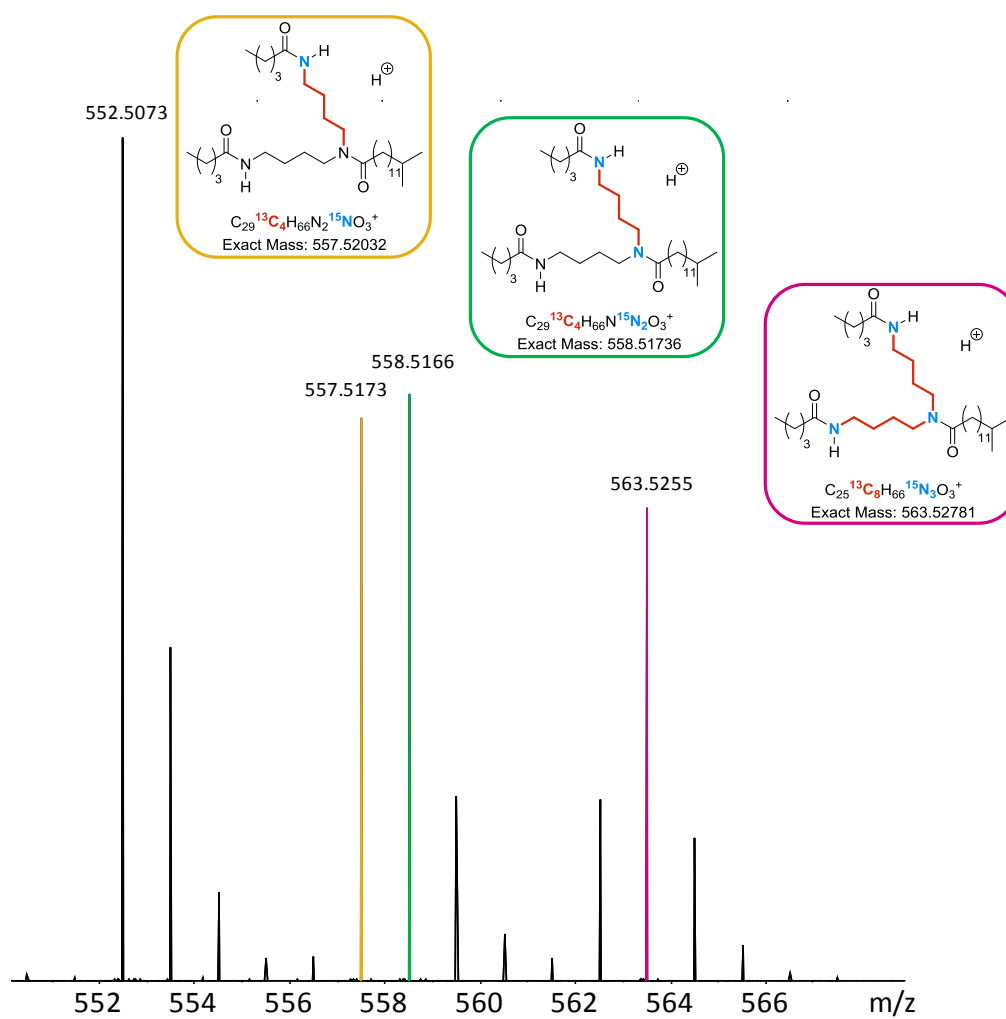


Figure S 4.6: MS spectrum of feeding experiment with L-Arginine- $^{13}C_6$, $^{15}N_4$ confirms the incorporation of up to two arginine building blocks in Cmp-552.

Table S 4.1: Table of all open reading frames (orfs) belonging to the putative homospermidine lipid biosynthesis cluster with proposed function and closest homologue according to a blastp search against the non-redundant protein database and the position of the orfs on the published M. xanthus DK1622 genome (NCBI reference sequence NC_008095.1)

Name	Length [bp]	Proposed function	Closest homologue	Identity [%] and length of alignment [AA]	Accession of closest homologue	Locus tag on the DK1622 genome
hslA	1023	Homospermidine synthesis	Homospermidine synthase (Myxococcus virescens)	98,5 / 339	SDF06447	MXAN_1678
hslB	1896	fatty acyl chain dehydrogenation	oxidase – FAD/FMN dependent (Myxococcus fulvus)	97,2 / 631	WP_043712221	MXAN_1676
hslC	2607	fatty acyl chain transfer	glycerol-3-phosphate acyl transferase (Myxococcus virescens)	99.5 / 868	SDF06388	MXAN_1675
hslD	990	unknown	NAD dependent epimerase (Myxococcus virescens)	98.5 / 329	SDF06360	MXAN_1674
hslE	1542	unknown	hypothetical adventurous gliding motility protein N (Myxococcus xanthus)	98.8 / 344	AAO66306	MXAN_1673
hslF	1005	unknown	hypothetical Protein (Stigmatella aurantiaca)	71.6 / 334	WP_075004684	MXAN_1687
hslG	897	N-acylation of the homospermidine moiety	Acetyltransferase GNAT family (Myxococcus virescens)	93.6 / 298	SDF06648	MXAN_1686
hslH	174	unknown	hypothetical protein (Myxococcus virescens)	92.9 / 56	SDF06632	MXAN_1685
hslI	558	regulation	histidine kinase (Myxococcus fulvus)	75.9 / 191	WP_013938788	MXAN_1684
hslK	2280	unknown	Dipeptidyl aminopeptidase (Myxococcus virescens)	98.3 / 747	SDF06517	MXAN_1682
hslL	798	unknown	Dienelactone hydrolase (Myxococcus virescens)	97.7 / 265	SDF06504	MXAN_1681
hslM	675	regulation	DNA binding response regulator	99.6 / 224	SDF06483	MXAN_1680
hslN	1167	regulation	signal transduction histidine kinase	96.6 / 338	SDF06456	MXAN_1679
hslO	2610	compound sensing	tonB like protein (Myxococcus virescens)	96.8 / 869	SDF06711	MXAN_1688
hslP	1269	unknown	hypothetical protein (Myxococcus virescens)	98.7 / 422	SDF06729	MXAN_1689
hslQ	945	unknown	esterase (Myxococcus fulvus)	79.9 / 313	WP_013935160	MXAN_1690
hslR1	525	regulation	TetR family regulator (Myxococcus virescens)	99.4 / 174	SDF06422	MXAN_1677
hslR2	906	regulation	LysR family regulator (Myxococcus virescens)	99.3 / 301	SDF06540	MXAN_1683



Figure S 4.7: Gene organization of the putative homospermidine lipid biosynthesis cluster. Biosynthesis genes depicted in blue; regulatory elements depicted in red; genes of unknown function or genes not involved in homospermidine lipid biosynthesis depicted in gray.

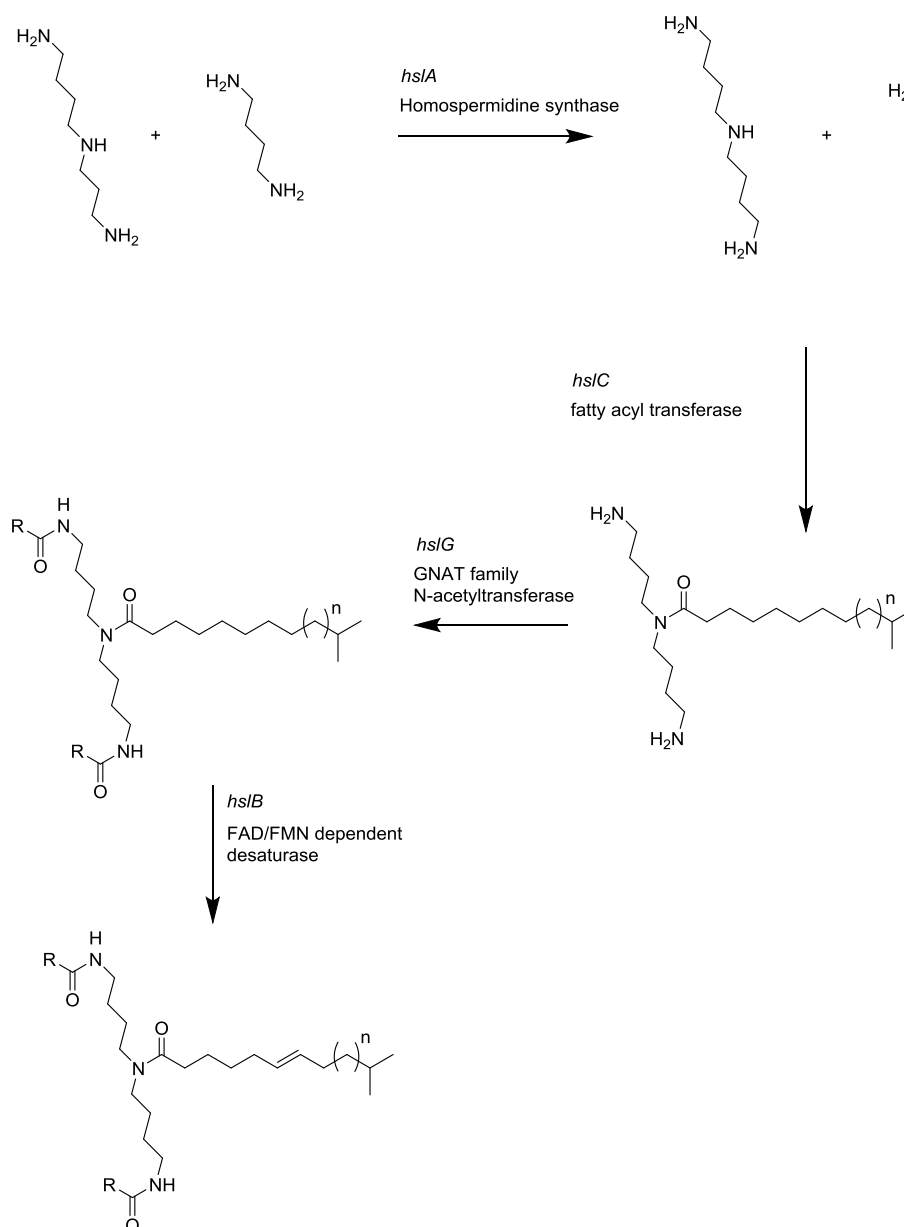


Figure S 4.8: Putative biosynthesis of the homospermidine lipids according to *in silico* analysis

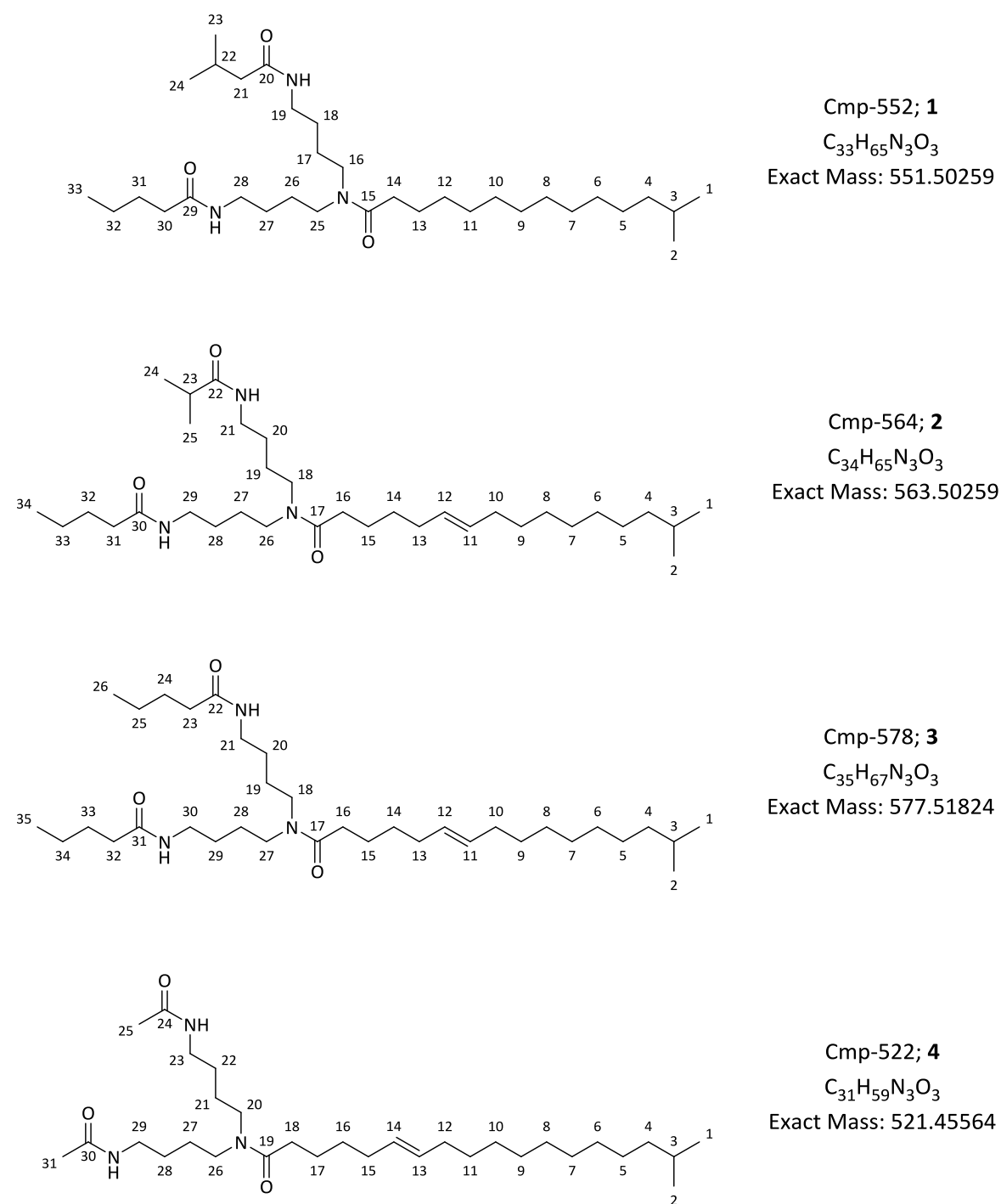


Figure S 4.9: Chemical structures of homospermidine-lipid derivatives with numbered carbon atoms corresponding to the NMR-data tables (Table S 4.2 Table S 4.5)

Table S 4.2: NMR data for Cmp-552, 1; HS(iso5:0/5:0/iso15:0); (^{a-c} HSQC signals overlapping)

Carbon number	$\delta^1\text{H}$ [ppm]	Multiplicity, J [Hz] and proton number	$\delta^{13}\text{C}$ [ppm]	COSY correlations [ppm]	HMBC correlations [ppm]
1/2	0.84	d, 6.73, 6	23.0	3	1/2, 3, 4
3 ^a	1.49	-	28.2	1/2, 4	1/2, 4
4 ^b	1.13	-	39.4	3, 5, 6	1/2, 4, 5, 6
5/6/7 ^c	1.23	-	30.1	4, 8/9/10	4, 8/9/10
8/9/10 ^a	1.58	-	26.0	5/6/7, 8/9/10, 11/12	5/6/7, 8/9/10, 11/12
11/12 ^c	1.29	-	30.1	8/9/10, 13	8/9/10, 11/12, 13, 14
13	1.59	-	26.1	12, 14	11, 12, 14, 15
14	2.25	t, 7.59, 2	33.6	12, 13	12, 13, 15
15	-	-	173.5	-	14, 16, 25
16	3.30	-	45.5	17	15, 17, 25
17 ^a	1.52	-	25.1	17, 18	16, 18, 19
18 ^a	1.47	-	26.7	17, 19	16, 17, 19
19	3.27	-	39.0	18	17, 18, 20
20	-	-	177.7	-	19, 21, 22
21	2.35	m	33.8	22	22, 23/24
22	2.35	-	36.0	21, 23/24	20, 21, 23/24
23/24 ^b	1.13	t, 6.69, 6	20.0	21, 22	20, 21, 22, 23/24
25	3.23	t, 7.59, 2	48.0	26	15, 16, 26/27
26/27 ^a	1.54	-	26.1	25, 27	25, 26/27, 28
28	3.27	-	39.0	26/27	25, 26/27, 29
29	-	-	173.7	-	28, 30, 31
30	2.14	t, 7.64, 2	39.1	31	31, 32, 33
31	1.64	m	19.5	30, 32, 33	30, 32, 33
32 ^c	1.29	-	25.9	31, 33	30, 31, 33
33	m	-	14.3	31, 32	31, 33

Table S 4.3: NMR data for Cmp-564, 2; HS(5:0/iso5:0/iso15:0(6E))

Carbon number	$\delta^1\text{H}$ [ppm]	Multiplicity, J [Hz] and proton number	$\delta^{13}\text{C}$ [ppm]	COSY correlations [ppm]	HMBC correlations [ppm]
1 + 2	0.91	d 6.7 Hz 2H	23.58	1.55	29.50, 40.69
3	1.55	m 1H	29.7	0.91, 1.21	23.57, 40.32
4	1.21	m 2H	40	1.55, 1.31	23.4, 29.6, 32.0
5	1.31	m 2H	32	1.2, 1.3	31.0, 30
6	ca 1.3	m	ca. 30	ca 1.3	ca. 30
7	ca 1.3	m	ca. 30	ca 1.3	ca. 30
8	ca 1.3	m	ca. 30	ca 1.3	ca. 30
9	1.37	m 2H	30.9	2.0, 1.34	132.0, 33.8
10	2	m 2H	33.8	1.37, 5.43	132.0, 30.9
11	5.43	m 1H	132	2.0, 5.37	-
12	5.37	m 1H	131.1	2.07, 5.43	28.8
13	2.07	m 2H	28.8	1.42, 5.34	131.1, 31.3
14	1.42	m 2H	31.3	2.07, 1.65	131.1, 33.4, 26.8, 28.1
15	1.62	m 2H	26.9	1.42, 2.26	31.3, 175.0
16	2.27	t 2H	36.9	1.62	26.9, 31.3, 179.5
17	-	-	176.5	-	-
18	3.34	m 2H	48.07	1.58, 3.2	26.5, 176.5
19	1.58	m 2H	26.5	1.508, 3.34	48.1, 28.5
20	1.508	m 2H	28.5	1.58, 3.34	26.5, 40.3
21	3.21	t 2H 7.19 Hz	40.3	1.51, 1.58	28.5, 180.3
22			180.3		
23	2.454	dq 1H	36.9	1.133	20.3, 180.3
24+25	1.133	CH3 2 mal?	20.36	2.467	36.9, 20.28, 180.0
26	3.35	m 2H	47.26	1.58, 3.20	26.5, 176.5
27	1.58	m 2H	26.5	1.51, 3.35	47.3, 28.5
28	1.507	m 2H	28.5	1.58, 3.20	26.5, 36.9
29	3.22	t 2H 7.19 Hz	36.9	1.50, 1.58	28.5, 176
30			175.93		
31	2.378	t 2H	34.45	1.6442,	26.7, 30.9
32	1.6442	m 2H	26.7	1.347, 3.36	
33	1.35	m 2H	30.1		
34	0.934	1 3H 6.7 Hz	14.8	1.347	

Table S 4.4: NMR data for *Cmp-578, 3; HS(5:0/5:0/iso17:0(6E))*

Carbon number	$\delta^1\text{H}$ [ppm]	Multiplicity, J [Hz] and proton number	$\delta^{13}\text{C}$ [ppm]	COSY correlations [ppm]	HMBC correlations [ppm]
1 + 2	0.9	d 6.7 Hz 6H	23.67	1.54	29.08, 40.8
3	1.55	m 1H	29.67	0.9, 1.2	23.37, 28.11, 40.71
4	1.2	m 2H	40.72	1.54	23.6, 29.6, 33.75
5	1.31	m 2H	33.75	1.2, 1.3	40.7, ~30
6	ca 1.3	m	ca. 30	ca 1.3	ca. 30
7	ca 1.3	m	ca. 30	ca 1.3	ca. 30
8	ca 1.3	m	ca. 30	ca 1.3	ca. 30
9	1.38	m 2H	30.3	1.3, 2.0	34.1, 132.0
10	2	m 2H	34.1	1.38, 5.45, 5.34	30.3, 131.1, 132.0
11	5.45	m 1H	132	2.0, 5.34, 2.07	34.1, 132.0, 33
12	5.375	m 1H	131.1	2.07, 2.2, 5.45	31.0, 131.1, 132.0
13	2.07	m 2H	31	1.42, 5.375	30.9, 131.5, 132.0
14	1.42	m 2H	30.9	2.07, 1.62,	131.1, 31, 26.89, 33.78
15	1.6	m 2H	26.8	1.42, 2.31	30.9, 33.68, 176.3
16	2.375	t, 2H, 7.2Hz	33.7	1.63	26.8, 31.0, 176.3
17	-	-	176.3	-	-
18	3.36	m 2H	47.04	1.58, 3.2	26.6, 176.5
19	1.57	m 2H	26.65	1.508, 3.34	47.0, 28.4
20	1.5	m 2H	28.38	1.58, 3.34	26.7, 40.2
21	3.21	t 2H 7.19 Hz	40.19	1.51, 1.58	28.3, 178.5
22			178.3		
23	2.3	t 2H	35.62	1.62	28.8, 31, 178.5
24	1.62		26.8	1.34	35.46
25	1.34		33.4	0.93	35.4, 14.7
26	0.93		14.7	1.34	33.4
27	3.36	m 2H	47.04	1.58, 3.20	26.6, 176.5
28	1.57	m 2H	26.65	1.51, 3.35	47.0, 28.4
29	1.5	m 2H	28.38	1.58, 3.20	26.7, 40.2
30	3.21	t 2H 7.19 Hz	40.19	1.50, 1.58	28.3, 178.5
31			178.3		
32	2.3	t 2H	35.62	1.62	28.8, 31, 178.5
33	1.62	m 2H	26.8	1.34	35.46
34	1.34	m 2H	33.4	0.93	35.4, 14.7
35	0.93	1 3H	14.7	1.34	33.4

Table S 4.5: NMR data for Cmp-522, 4; HS(2:0/2:0/iso19:0(6E))

Carbon number	δ 1H [ppm]	Multiplicity, J [Hz] and proton number	δ 13C [ppm]	COSY correlations [ppm]	HMBC correlations [ppm]
1 + 2	0.88	d, 6.7 Hz, 6H	22.9	1.51	29.0, 39.8
3	1.51	m, 1H	29	0.88, 1.18	39.8, 22.9, 28.3
4	1.18	m, 2H	39.8	1.29, 1.51	22.9, 29.0, 28.3
5	1.29	m, 2H	28.3	1.18, 1.3	30, 39.8
6	ca. 1.3	m, 2H	ca. 30	ca. 1.3	ca. 30
7	ca. 1.3	m, 2H	ca. 30	ca. 1.3	ca. 30
8	ca. 1.3	m, 2H	ca. 30	ca. 1.3	ca. 30
9	ca. 1.3	m, 2H	ca. 30	ca. 1.3	ca. 30
10	ca. 1.3	m, 2H	ca. 30	ca. 1.3	ca. 30
11	1.35	m, 2H	30.4	1.98, ca.1.30	131.6, 33.3, 30
12	1.98	m, 2H	33.3	1.35, 5.45, 5.39	30.4, 131.6, 130.6
13	5.45	m, 1H	131.6	1.98, 2.01, 5.39	30.4, 33.3, 130.6
14	5.39	m, 1H	130.6	2.01, 1.98, 5.45	33.0, 30.0, 131.6
15	2.01	m, 2H	33	1.40, 5.39	30.2, 33.3, 130.6, 131.6
16	1.4	m, 2H	30.2	1.59, 2.35, 2.01	33.0, 26.0
17	1.59	m, 2H	26	1.40, 2.35	33.7, 30.2 175.2
18	2.35	t, 7.6 Hz, 2H	33.7	1.59, 1.40, 3.32	30.2, 26.0, 175.2
19	-	-	175.2	-	-
20	3.32	m, 2H	46.3	1.56, 3.35, 2.35	25.9, 48.5, 175.2
21	1.56	m, 2H	25.9	3.32, 1.48	46.3, 27.1
22	1.48	m, 2H	27.1	3.17, 1.56	25.9, 46.3, 39.6
23	3.17	t, 6.2 Hz, 2H	39.6	1.48, 1.92	27.1, 25.9, 173.3
24	-	-	173.3	-	-
25	1.92	s, 3H	22.1	3.17	39.6, 173.3
26	3.35	m, 2H	48.5	1.59, 3.32, 2.35	27.0, 46.3, 175.2
27	1.59	m, 2H	27	3.35, 1.52	48.5, 27.5
28	1.52	m, 2H	27.5	3.20, 1.59	48.5, 27.0, 39.4
29	3.2	t, 6.2 Hz, 2H	39.4	1.52, 1.94	27.5, 27.0, 173.5
30	-	-	173.5	-	-
31	1.94	s, 3H	22.1	3.2	39.4, 173.5

4.7 References

- [1] Newman, D. J. & Cragg, G. M.: Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **2016**, 79, 629–661, DOI: 10.1021/acs.jnatprod.5b01055
- [2] Herrmann, J., Fayad, A. A. & Müller, R.: Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **2017**, 34, 135–160, DOI: 10.1039/c6np00106h
- [3] Wenzel, S. C. & Muller, R.: Myxobacteria--'microbial factories' for the production of bioactive secondary metabolites. *Mol. Biosyst.* **2009**, 5, 567–574, DOI: 10.1039/b901287g
- [4] Zaburannyi, N. *et al.*: Genome Analysis of the Fruiting Body-Forming Myxobacterium *Chondromyces crocatus* Reveals High Potential for Natural Product Biosynthesis. *Appl. Environ. Microbiol.* **2016**, 82, 1945–1957, DOI: 10.1128/AEM.03011-15
- [5] Wenzel, S. C. & Müller, R.: The impact of genomics on the exploitation of the myxobacterial secondary metabolome. *Nat. Prod. Rep.* **2009**, 26, 1385–407, DOI: 10.1039/b817073h
- [6] Medema, M. H. & Fischbach, M. A.: Computational approaches to natural product discovery. *Nat. Chem. Biol.* **2015**, 11, 639–48, DOI: 10.1038/nchembio.1884
- [7] Ziemert, N., Alanjary, M. & Weber, T.: The evolution of genome mining in microbes - a review. *Nat. Prod. Rep.* **2016**, 33, 988–1005, DOI: 10.1039/c6np00025h
- [8] Jensen, P. R.: Natural Products and the Gene Cluster Revolution. *Trends Microbiol.* **2016**, 24, 968–977, DOI: 10.1016/j.tim.2016.07.006
- [9] Crüsemann, M. *et al.*: Prioritizing Natural Product Diversity in a Collection of 146 Bacterial Strains Based on Growth and Extraction Protocols. *J. Nat. Prod.* **2017**, 80, 588–597, DOI: 10.1021/acs.jnatprod.6b00722
- [10] Rutledge, P. J. & Challis, G. L.: Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **2015**, 13, 509–23, DOI: 10.1038/nrmicro3496
- [11] Wolfender, J.-L. L.: HPLC in natural product analysis: the detection issue. *Planta Med.* **2009**, 75, 719–734, DOI: 10.1055/s-0028-1088393
- [12] Barkal, L. J. *et al.*: Microbial metabolomics in open microscale platforms. *Nat. Commun.* **2016**, 7, 10610, DOI: 10.1038/ncomms10610
- [13] Doroghazi, J. R. *et al.*: A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat. Chem. Biol.* **2014**, 10, 963–8, DOI: 10.1038/nchembio.1659
- [14] Krug, D., Muller, R. & Müller, R.: Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products. *Nat. Prod. Rep.* **2014**, 31, 768–783, DOI: 10.1039/c3np70127a
- [15] Fisch, K. M., Schaberle, T. F. & Schäberle, T. F.: Toolbox for Antibiotics Discovery from Microorganisms. *Arch. Pharm. (Weinheim)*. **2016**, 349, 683–691, DOI: 10.1002/ardp.201600064
- [16] Allard, P.-M., Genta-Jouve, G. & Wolfender, J.-L.: Deep metabolome annotation in natural products research: towards a virtuous cycle in metabolite identification. *Curr. Opin. Chem.*

- Biol.* **2017**, 36, 40–49, DOI: 10.1016/j.cbpa.2016.12.022
- [17] Covington, B. C., McLean, J. A. & Bachmann, B. O.: Comparative mass spectrometry-based metabolomics strategies for the investigation of microbial secondary metabolites. *Nat. Prod. Rep.* **2017**, 34, 6–24, DOI: 10.1039/c6np00048g
- [18] Garcia, R. O., Krug, D. & Müller, R.: Chapter 3. Discovering natural products from myxobacteria with emphasis on rare producer strains in combination with improved analytical methods. *Methods Enzymol.* **2009**, 458, 59–91, DOI: 10.1016/S0076-6879(09)04803-4
- [19] Volz, C., Kegler, C. & Müller, R.: Enhancer binding proteins act as hetero-oligomers and link secondary metabolite production to myxococcal development, motility, and predation. *Chem. Biol.* **2012**, 19, 1447–59, DOI: 10.1016/j.chembiol.2012.09.010
- [20] Huntley, S. *et al.*: Comparative genomic analysis of fruiting body formation in Myxococcales. *Mol. Biol. Evol.* **2011**, 28, 1083–1097, DOI: 10.1093/molbev/msq292
- [21] Marcos-Torres, F. J. *et al.*: In depth analysis of the mechanism of action of metal-dependent sigma factors: characterization of CorE2 from *Myxococcus xanthus*. *Nucleic Acids Res.* **2016**, 44, 5571–5584, DOI: 10.1093/nar/gkw150
- [22] Kroos, L.: Highly Signal-Responsive Gene Regulatory Network Governing *Myxococcus* Development. *Trends Genet.* **2017**, 33, 3–15, DOI: 10.1016/j.tig.2016.10.006
- [23] Lloyd, D. G. & Whitworth, D. E.: The Myxobacterium *Myxococcus xanthus* Can Sense and Respond to the Quorum Signals Secreted by Potential Prey Organisms. *Front. Microbiol.* **2017**, 8, 439, DOI: 10.3389/fmicb.2017.00439
- [24] Gerth, K., Bedorf, N., Höfle, G., Irschik, H. & Reichenbach, H.: Epothilons A and B: antifungal and cytotoxic compounds from *Sorangium cellulosum* (Myxobacteria). Production, physico-chemical and biological properties. *J. Antibiot. (Tokyo)*. **1996**, 49, 560–3, DOI: 10.7164/antibiotics.49.560
- [25] Jahns, C. *et al.*: Pellasoren: Struktur, Biosynthese und Totalsynthese eines zytotoxischen Sekundärmetaboliten aus *Sorangium cellulosum*. *Angew. Chemie* **2012**, 124, 5330–5334, DOI: 10.1002/ange.201200327
- [26] Keller, L. *et al.*: Macyranones: Structure, Biosynthesis, and Binding Mode of an Unprecedented Epoxyketone that Targets the 20S Proteasome. *J. Am. Chem. Soc.* **2015**, 137, 8121–8130, DOI: 10.1021/jacs.5b03833
- [27] Kraemer, S. A., Touns, M. A. & Velicer, G. J.: Natural variation in developmental life-history traits of the bacterium *Myxococcus xanthus*. *FEMS Microbiol. Ecol.* **2010**, 73, 226–233, DOI: 10.1111/j.1574-6941.2010.00888.x
- [28] Luzzatto-Knaan, T., Melnik, A. V. & Dorrestein, P. C.: Mass spectrometry tools and workflows for revealing microbial chemistry. *Analyst* **2015**, 140, 4949–4966, DOI: 10.1039/c5an00171d
- [29] Dunham, S. J. B., Ellis, J. F., Li, B. & Sweedler, J. V.: Mass Spectrometry Imaging of Complex Microbial Communities. *Acc. Chem. Res.* **2017**, 50, 96–104, DOI: 10.1021/acs.accounts.6b00503
- [30] Hoffmann, T. & Dorrestein, P. C.: Homogeneous matrix deposition on dried agar for MALDI

- imaging mass spectrometry of microbial cultures. *J. Am. Soc. Mass Spectrom.* **2015**, 26, 1959–1962, DOI: 10.1007/s13361-015-1241-8
- [31] Monroe, E. B., Koszczuk, B. A., Losh, J. L., Jurchen, J. C. & Sweedler, J. V.: Measuring salty samples without adducts with MALDI MS. *Int. J. Mass Spectrom.* **2007**, 260, 237–242, DOI: 10.1016/j.ijms.2006.08.019
- [32] Gode, D. & Volmer, D. A.: Lipid imaging by mass spectrometry - a review. *Analyst* **2013**, 138, 1289–1315, DOI: 10.1039/c2an36337b
- [33] Petkovic, M. *et al.*: Detection of individual phospholipids in lipid mixtures by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry: phosphatidylcholine prevents the detection of further species. *Anal. Biochem.* **2001**, 289, 202–16, DOI: 10.1006/abio.2000.4926
- [34] Knochenmuss, R.: The Coupled Chemical and Physical Dynamics Model of MALDI. *Annu. Rev. Anal. Chem.* **2016**, 9, 365–385, DOI: 10.1146/annurev-anchem-071015-041750
- [35] Liu, W.-T. W.-T. *et al.*: Imaging mass spectrometry of intraspecies metabolic exchange revealed the cannibalistic factors of *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, 107, 16286–16290, DOI: 10.1073/pnas.1008368107
- [36] Shih, C.-J., Chen, P.-Y., Liaw, C.-C., Lai, Y.-M. & Yang, Y.-L.: Bringing microbial interactions to light using imaging mass spectrometry. *Nat. Prod. Rep.* **2014**, 31, 739–755, DOI: 10.1039/c3np70091g
- [37] Fahy, E. *et al.*: Update of the LIPID MAPS comprehensive classification system for lipids. *J. Lipid Res.* **2009**, 50 Suppl, S9-14, DOI: 10.1194/jlr.R800095-JLR200
- [38] Wang, M. *et al.*: Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, 34, 828–837, DOI: 10.1038/nbt.3597
- [39] Allard, P.-M. *et al.*: Integration of Molecular Networking and In-Silico MS/MS Fragmentation for Natural Products Dereplication. *Anal. Chem.* **2016**, 88, 3317–23, DOI: 10.1021/acs.analchem.5b04804
- [40] Shannon, P. *et al.*: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, 13, 2498–2504, DOI: 10.1101/gr.1239303

5 Discussion

In summary, the work presented in the previous chapters covers several aspects of natural product research. The second and third chapter are dealing with two similar natural products, tomaymycin and tilivalline. These chapters are describing comprehensive analysis and elucidation of biosynthetic pathways, successfully performed *in vitro* reconstitutions as well as the development of an advanced LC-MS method to study the relevant biosynthetic proteins involved in biosynthesis of the respective compounds. Chapter four describes a novel approach to discover new myxobacterial secondary metabolites. As an alternative to the standardized approach based on cultivation in liquid media, the here introduced workflow covers the metabolomics changes due to fruiting body formation.

In detail, in chapter two the timing possibilities of the tomaymycin biosynthesis has been finally narrowed down to two scenarios by combining results from various different experiments. It was possible to reconstitute detectable amounts tomaymycin by full *in vitro* reconstitution and to observe first mutasynthesis products by using this system. The LC-MS method developed to analyze intact proteins with masses up to 200 kDa allowed comprehensive insights into the individual steps of the NRPS derived biosynthesis of tomaymycin. The usage of this method enabled to study substrate specificities but also to track each individual step on the assembly line. Finally, the inconsistency of results derived from different methodologies revealed also the risk of hypotheses, which are based on results from a single analytical method. Combinatory conclusion of various results led to the final elucidation of the synthetic pathway of tomaymycin.

Chapter three is about tilivalline, which was identified unambiguously in a clinical isolate of *Klebsiella oxytoca*. Despite similarity of the biosynthetic gene clusters of tomaymycin and tilivalline, the origin of the indole moiety remained obscure. Heterologous expression as well as a functional *in vitro* reconstitution yielded a truncated tilivalline structure lacking the indole moiety. It finally turned out, that the indole attachment results from a spontaneous Friedel-Craft-like alkylation of free indole and the respective carbinolamine derived from the assembly line. The established heterologous expression system was subsequently used as synthetic biology platform to obtain several derivatives. In addition, it served as a screening platform to find potential tilivalline production inhibitors. These experiments revealed that salicylic acid based compounds seem to inhibit the production of tilivalline. Consequently, efforts towards this project will be intensified to confirm these promising results in a mouse model.

Chapter four covers yet another aspect of natural product research. The respective work represents an alternative approach towards detection and isolation of yet unknown myxobacterial

metabolites. In this context, the attempts were focused on the metabolism of the unique myxobacterial fruiting bodies of the model strain *Myxococcus xanthus* DK1622. In order to explore the metabolomic potential of this myxobacterial vegetative state, different analytical workflows aiming to gain high quality data with optimized information content and to track down yet unknown compound classes. Finally, a new compound class with antibiotic activity has been identified and its production could undoubtedly be linked to the fruiting body formation of *Myxococcus xanthus* DK1622. The metabolome mining studies performed in this context facilitated the isolation of additional derivatives from a related myxobacterial strain *Myxococcus* MCy9290. This clearly demonstrated the potential of this novel workflow, which can serve as another helpful tool when searching for novel myxobacterial metabolites with promising chemical scaffolds and bioactivities.

5.1 Tomaymycin and Tilivalline

5.1.1 Pyrrolobenzodiazepines

Tomaymycin and tilivalline are both representatives of the compound class of pyrrolobenzodiazepines (PBDs). Despite the identical core structure of the compounds, their origin, and characteristics in terms of bioactivity are fundamentally different. While tomaymycin is a compound produced by *Streptomyces achromogenes* showing antibacterial as well as good antineoplastic activity by selective DNA alkylation, tilivalline is synthesized by *Klebsiella oxytoca* and has been identified as a pathogenicity factor in antibiotic associated hemorrhagic colitis¹⁻⁵.

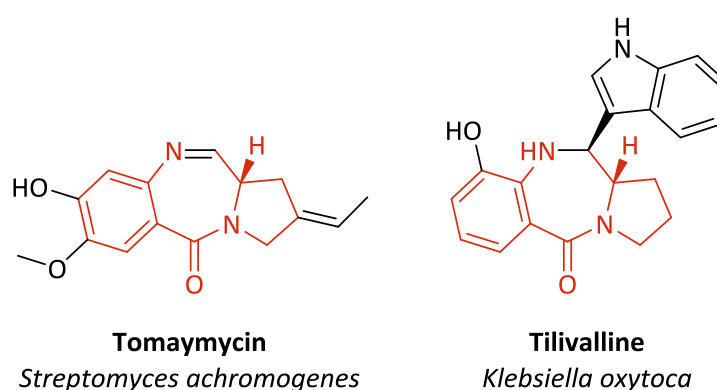


Figure 5.1: Chemical structures of tomaymycin and tilivalline. Identical core structure is highlighted in red.

PBD-related research efforts mainly focused on their optimization towards potential drugs for cancer treatment⁶. Nevertheless, there is still a need to perform *in vitro* experiments to elucidate biosynthetic pathways, to test mutasynthesis approaches and to reveal substrate specificities of involved enzymes. The elucidation of the underlying biosynthesis is of special interest with respect to virulence factors as it enables the development of inhibition strategies but operational *in vitro* systems in general can be utilized as synthetic biology platform providing an easy access to new derivatives. In addition to already performed biosynthesis studies about tomaymycin, anthramycin or sibiromycin, knowledge regarding the mode of action and the influence of distinct moieties was gained by synthesizing roughly 60 not naturally occurring PBDs in the last decades. As a result, potential lead structures could be further improved by increasing specificity, bioavailability and decreasing side effects⁶. In contrast to research projects supporting drug development based on tomaymycin and similar molecules the role of tilivalline in human diseases has been widely neglected. Instead of optimizing the tilivalline structure towards maximum bioactivity, the inhibition of biosynthesis and the mode of action was of high interest in this work. Heterologous expression system and *in vitro* reconstitution platform allowed searching for compounds, which lead to an inhibitory effect of the tilivalline production by blocking the assembly line. The inability of tilivalline to intercalate into DNA could finally be explained by the missing imine substructure,

and indicated a different mode of action causing the toxicity. Considering the similar structures of tomaymycin and tilivalline, it is astonishing that they obviously do not share the same mode of action. While the results of the individual projects were already extensively discussed in the respective chapters, this section is generally dealing with the usage of *in vitro* assays in research and industry.

5.1.2 *In Vitro* Reconstitutions: Benefits and Obstacles

Despite tremendous technical developments and increasing knowledge in the field of biochemistry, complete *in vitro* reconstitutions of biosynthetic pathways remain challenging. In particular, the reconstitution of complex biosynthetic pathways comprising multimodular enzyme systems like NRPS or PKS is extremely difficult. However, there is a continued effort towards *in vitro* reconstitution of biosynthetic pathways, as these platforms allow for improved insights in how nature generates this plethora of different chemical structures. This valuable knowledge can further contribute to the in-depth investigation of natural products and can support, for instance, the modification of complex chemical scaffold or can inspire the development of biomimetic chemistry⁷.

A major benefit of using *in vitro* but also *in vivo* reconstitution is the possibility to produce complex natural products from simple precursors without complex multistage total syntheses approaches. Especially, structures with various stereocenters are often highly problematic and difficult to access by chemical synthesis. An example for this challenge were the attempts to obtain the antibiotic erythromycin, which was initially isolated from *Saccharopolyspora erythraea*, by total synthesis approaches. It took three decades after the discovery to establish a total synthesis route that finally yielded the correct stereochemistry for the 10 stereospecific carbons as well as for the distinct substitutions^{8,9}. Formation of the erythromycin precursor 6-deoxyerythronolide, including most of the stereocenters represented the most difficult step in this approach. The chemical synthesis of the 14-membered macrolide already requires at least 20 steps and thus associated with a highly intensified workload¹⁰. In contrast to that, production of erythromycin by fermentation is preferred due to outcompeting efficiency¹¹. Khosla et al. managed, however, to reconstitute this precursor completely *in vitro* with turnover rates above 1 min⁻¹ by incubating the five biosynthetic proteins involved, with the required precursor molecules propionyl-CoA, methylmalonyl-CoA, and NADPH. After two decades of research a system with acceptable production efficiency has been achieved and paved the way for further improvements towards efficient *in vitro* reconstitutions of complex natural products¹². In line with an efficient production by *in vitro* reconstitution would be the availability of inexpensive precursors and the possibility to recycle co-factors as shown by Greunke et al.¹³. While the full reconstitution of erythromycin was not achievable *in vitro*, the most

difficult part of the synthesis could be managed and missing substituents subsequently attached by a semi-synthetic approach. A similar success story is the full reconstitution of norsolorinic acid, a precursor of aflatoxin B¹⁴.

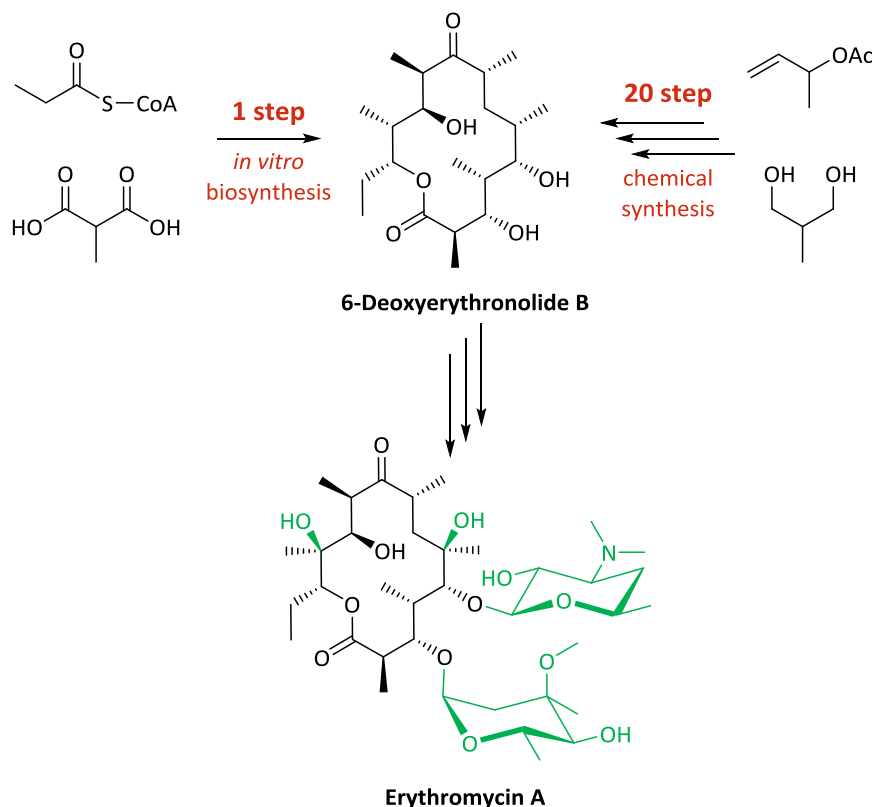


Figure 5.2: Comparison of two methods to obtain 6-deoxyerythronolide B, an erythromycin precursor. While the precursor can be achieved in a single step by a complete *in vitro* reconstitution, the respective total synthesis approach needs at least 20 steps.

Besides the synthesis itself, the relatively simple reaction mixtures of *in vitro* reconstitutions in comparison to organic synthesis and heterologous production systems are a big advantage, in particular with regard to analyzing and subsequent product isolation. Frequently used LC-MS analyses can be performed either on the crude incubation mixture or after removal of the enzymes, depending on the proteins sizes. Also in terms of compound isolation, *in vitro* production systems excel classic organic synthesis, since the sought product can easily be separated from the enzymes, precursors and co-factors used. This is by far less time consuming than the purification cascades usually followed after organic synthesis.

In vitro systems are hitherto mainly been used for biosynthesis studies because of the low efficiency but thereby already promote the optimization of *in vivo* production systems. Efficient *in vitro* assays would, however, be a powerful tool providing an easy way to obtain derivatives of interesting natural products, and thereby a possibility of optimization by means of medicinal chemistry. In the best case, such derivatives can be produced by simply supplementing non-natural

precursors¹⁵. If the intended modification is not directly achievable, there is still a chance to introduce more readily accessible moieties for semi-synthetic follow-up work¹⁶. The spectrum of derivatives that could principally be obtained can be determined through preliminary small scale testing as described in this work for tomaymycin and tilivalline. Therefore, the investigations about substrate specificity presented here are key factors regarding such workflows. If an efficiently working *in vitro* assay can be set up, it can allow for production of a plethora of non-naturally observed derivatives that could be synthesized in a well-plate format and immediately tested for bioactivity, solubility, or other features of interest. In contrast, a total synthesis approach is, in most cases, rather elaborate regarding time consumption and resources but also flexibility of such systems is usually very limited when it comes to exchange of educts¹⁶. Consequently, this field of research should be further supported and investigated as it represents a great potential for natural product based drug developments.

Another application for *in vitro* experiments is a detailed analysis of reaction kinetics. Not only single enzymes but also the complete biosynthetic assembly lines can be studied towards kinetic parameters like turnover rates. Limiting factors can be ascertained and corresponding incubation temperature, enzyme, substrate, or co-factor concentrations are adjusted to increase the product titers¹⁷. However, experiments with opposite intentions can also be performed as using functioning *in vitro* systems enables to test inhibitors which shall decrease the natural product formation^{18,19}. The major advantage of these *in vitro* approaches is the isolated analysis of the system. The analytical results of such experiments will not be distorted due to the occurrence of secondary effects like possible metabolization or formation of substrates via other biochemical routes. This is a clear advantage compared to *in vivo* studies and underlines *in vitro* experiments as an excellent test platform for preliminary screenings and pre-selection of potential inhibitors. Cell-free experiments do also contribute to an easier analysis and more reliable data sets compared to very complex mixtures derived from *in vivo* studies.

In vitro reconstitutions represent not least the method of choice to elucidate the biosynthetic pathway of a natural product in detail⁷. Experiments with the isolated biosynthesis proteins are often more reliable to get better insights in the functionality of NRPS/PKS megaenzymes compared to well established to well established DNA modifications that block distinct protein functions. While gene inactivation experiments represent a solid method to associate distinct biosynthetic gene clusters with the production of a natural product, this method does not allow for detailed analysis of certain enzymatic reactions^{20,21}. Reasons are again the commonly emerging secondary effects by *in vivo* studies, which hamper the analysis. From the above discussions, one can be concluded that *in vitro* experiments are the method of choice for comprehensive studies regarding biosynthesis. Therefore, this approach was also followed in case of tomaymycin and tilivalline. The

application of *in vitro* experiments under defined conditions with pure proteins and subsequently performed analysis finally resulted in the full elucidation of the tomaymycin and tilivalline biosynthesis.

Despite all advantages of *in vitro* experiments in biochemistry, it cannot be neglected that the establishment of a functional and robust systems involves a lot of work. In particular, natural products derived by PKS/NRPS assembly lines are very difficult to reconstitute artificially. This fact is also reflected in the very limited number of successfully established full reconstitutions of natural products. In the last decades only a dozen publications reported fully functional systems that were able to achieve detectable amounts of product⁷. The reasons for this are attributable to various problems. As already indicated by the designation, megasynthetases and megasynthases are often characterized by the size of the enzyme complex. The largest known PKS (17019 aa) has nine modules which can be divided into 47 domains. The largest NRPS (16367 aa) shows a similar size with 15 modules constituting 46 domains whereas the largest hybrid thereof (14274 aa) was found with 2 modules consisting of 39 domains²². Although these values represent the maxima, it is not surprising that the size of such protein complexes frequently exceeds several hundred kDa²³. These enormous sizes often hamper protein expression in heterologous hosts like *Escherichia coli* as they typically surpass the size of their own proteins. The largest known *E. coli* protein is roughly 170 kDa in size and already significantly smaller than the comparably small two modular NRPSs synthesizing tomaymycin and tilivalline²⁴. Hence, heterologous hosts face difficulties with the complete biosynthesis of large proteins, which frequently lead to truncated proteins. Even if a complete expression in sufficient amount could be achieved, the handling of large sized proteins is challenging. Poor solubility, effects like agglomeration or biochemical inactivity due to mis-folded proteins are common issues when working with large protein complexes²⁵. Consequently, protein complexes are often divided into the respective modules and expressed separately. That way, the expression is often more efficient and in addition issues regarding protein folding and solubility can be circumvented. However, other complications may occur because of the resulting interruptions of the assembly line. To ensure the complete assembly of a natural product it is required that the particular modules form a complex in proper orientation which enables a channeling of the growing substrate until it is finally released from the assembly line. Presumably, these properly aligned modules are difficult to achieve by simply mixing separately expressed and purified proteins and therefore obstruct high product titers. Another point that can complicate the successful implementation of an *in vitro* reconstitution is the lack of required co-factors or co-enzymes. It is not always obvious, which additional compounds or proteins in what ratio are needed to obtain a functional *in vitro* system. In some instances, biosynthetic proteins even have to be co-expressed with partner proteins, which are evidently lacking any enzymatic function, to enable their own

functionality or solubility in an assay. Such observations are known for MbtH-like proteins and were also found during comprehensive *in vitro* studies about HHQ and PQS, two *Pseudomonas aeruginosa* quorum-sensing molecules (unpublished date)^{26,27}. In view of the mentioned obstacles that can appear during working on *in vitro* systems the low number of known full *in vitro* reconstitutions seems comprehensible. It is furthermore not surprising that most of the functional systems known so far are synthesizing comparatively small natural products⁷.

5.1.3 The Role of Instrumental Analytics in *in vitro* Experiments

Modern analytical chemistry is an integral and indispensable part of comprehensive natural product biosynthesis studies *in vitro*. Fast and reliable analytical methods help scientists to interpret experiments and enable to acquire a deeper understanding and knowledge regarding complex biosynthesis processes²⁸. Analysis can be subdivided into two main fields: the analysis of small molecules (< 2000 Da) to which usually the educts, by-products, as well as products belong and the analysis of large molecules (> 2000 Da) like the involved proteins or protein complexes. While the analysis of large molecules is often still based on rather unprecise techniques like SDS-PAGE, small molecules are usually analyzed on highly sensitive and accurate LC-MS or GC-MS setups^{29,30}. The difference is certainly caused by the fact that analyzing large molecules is much more difficult compared to small molecules. However, measurements of *in vitro* experiments on small molecule level do initially provide a wealth of information, which can already answer many unresolved issues about a biosynthesis.

First, the analysis of an *in vitro* reconstitution assay towards the expected product give clear indication for the functionality of the system. Even a very low product titer can be unambiguously detected as shown for tomaymycin in this work. Targeted search for by-products or truncated products additionally offer first insights into the biosynthesis route or, for instance, reveal that some tailoring enzymes are missing. Once a biosynthesis hypothesis is proposed based on comprehensive analysis of genetic data and observed side products, it is required to prove the assumptions. Analysis of feeding experiments with isotopically labeled precursors or building blocks is a well established approach providing evidence for the single steps of a biosynthetic pathway^{31–33}. Although such feeding experiments are mainly used in *in vivo* systems, they can also be helpful in *in vitro* systems, especially to elucidate the biosynthesis of unusual substructures. In this context, there are some reasons in terms of analytics why these experiments should preferably be performed *in vitro*, provided a functional system is available. Exclusive incorporation of the labeled precursor will only take place if the natural precursor will not be supplemented. As no mixtures of unlabeled and labeled product will occur, the obtained products are analytically easier to interpret. In an *in vivo* setting, this is nearly impossible since building blocks like amino acids are already found

in the media components or derived by the primary metabolism and can thus not be prevented. To obtain sufficient incorporation and enable determination including MS/MS experiments, the added quantity of labelled precursor has to be increased. Another problem are the other metabolic pathways of an organism, which can hamper the analysis. Nature uses only a few building block to produce a plethora of compounds³⁴. Hence, there is a high chance that the added labelled precursor is also consumed by many other pathways, which again requires an increased addition of labelled compound. These issues can easily be circumvented by using an *in vitro* test platform for feeding studies. Concerning possible mutasynthesis approaches, the same limitations apply, but in addition, there is a certain chance that the tested precursors are not able to pass the cell membranes. In this case, no incorporation of the respective precursor can be observed which could lead to wrong conclusions. Instrumental analytics are also important tools to study kinetics *in vitro*^{35–37}. In particular, for quantification purposes, *in vitro* experiments facilitate the analysis significantly. Laborious sample preparation becomes unnecessary since only proteins are removed prior to the measurements and furthermore, matrix effects due to growth media as well as other substances derived from the organism are negligible compared to *in vivo* experiments. Another conceivable application of operational *in vitro* systems is the access to isotopically labeled standards of high isotopic purity for analytical purposes that would otherwise require complex organic synthesis. The challenges regarding *in vitro* system should, however, not be underestimated. As the proteins for such systems are usually overexpressed in heterologous hosts it can be assumed that the proteins are partly misfolded and consequently not functional. It is also necessary to ensure that possible post translational modification by the native producer are not required to obtain fully operational proteins.

Besides the analysis on small molecule level, the analysis on protein level, the so called “top down proteomics”, is of increased interest in natural product research³⁸. Such measurements are more complicated compared to the standard measurements applied for small molecules but have nevertheless an important role for basic research. Even though they are partly applicable for *in vivo* studies, samples derived from *in vitro* studies are preferred³⁹. Based on such experiments, hypotheses about the detailed process of biosynthesis by NRPS or PKS can be evaluated because it is, for example, possible to confirm substrate or intermediate species attached to the assembly line^{40,41}. Intact protein analysis can also be used to test substrate specificities of distinct modules or domains and help to reveal the bottlenecks of biosynthetic assembly lines. Combined results based on analyzing *in vitro* experiments on small molecule as well as on intact protein level help to shed light on the complex process of natural product biosynthesis. The development of a fast and reliable analytical method allowed to address such questions and is one major aspect of this work.

Many publications from the last decade deal with intact protein measurements from *in vitro* assays, whereby these analyses frequently reach the limits of modern analytical setups⁴². Despite the availability of FT-ICR mass spectrometers with resolving powers of several millions, which theoretically should allow for detection of large NRPS/PKS proteins complexes, this cannot be accomplished easily⁴³. Even with an ideal sample, these resolving values are challenging due to physical processes in the ICR-cell. The immense number of different ion species that are simultaneously present in the cell, results in mutual interference of the cyclotron motion and thus to a loss of resolution by signal broadening⁴⁴. In addition, the missing solubility in buffer free solvents and inhomogeneity of the enzymes are some of the major reasons that hamper a straightforward detection of large NRPS/PKS enzymes by such setups. Proteins are generally obtained in several proteoforms, differing only in slight structural changes. These changes can be caused by diverse posttranslational modifications, *holo/apo* forms, translation dysfunction resulting in amino acid exchanges, or incomplete substrate loading³⁸. The resulting number of different species with minor mass shifts does not only result in very close MS-signals for the distinct charge states but also in a loss of intensities because the determined protein concentration is actually distributed to many different proteoforms. Furthermore, under these circumstances the above-mentioned dephasing effect of ions in the ICR-cell is further compounded. Sufficient chromatographic separation prior to MS detection can circumvent the simultaneous analysis of several proteoforms and help to exploit the full potential of high resolution MS. However, taking into account that the obtained modifications are negligible regarding the sizes of the enzymes, a chromatographic separation is a challenging endeavor requiring tremendous effort. This may be partly caused by the predominantly underlying on-off mechanism of large proteins on RP-columns⁴⁵. Separations by capillary electrophoresis (CE) are a promising alternative to RP-LC and seem to be more suitable, but especially the required coupling to a mass spectrometer remains challenging in everyday laboratory practice^{46,47}.

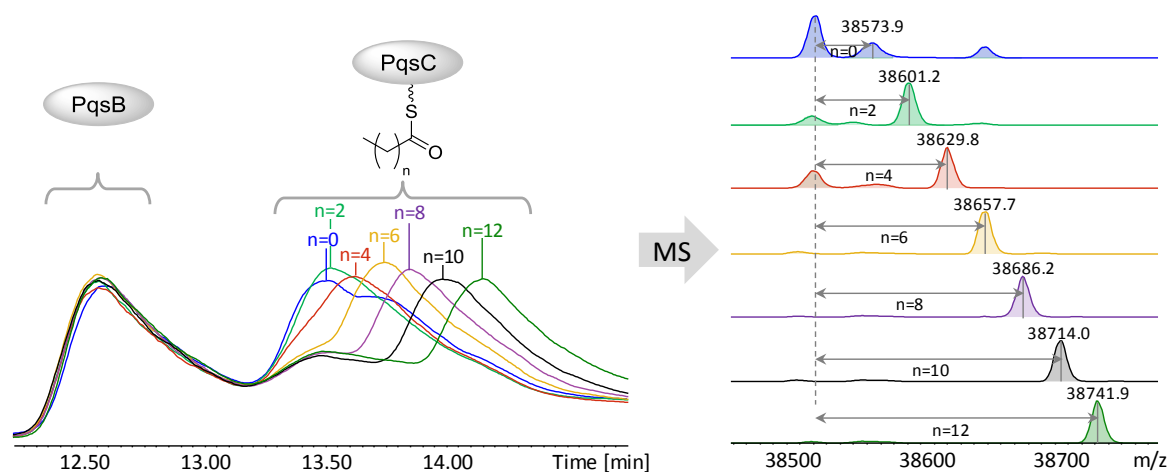


Figure 5.3: Overlaid total ion chromatograms of PqsBC loading experiments with different acyl-CoAs (C2-C14) showing shifted retention times for PqsC and the unchanged retention times for PqsB. Deconvoluted MS spectra confirmed the respective loaded species of PqsC.

Interestingly, in some cases even minimal modification appear to be sufficient for significant changes in retention times on RP-columns. The biosynthesis enzyme PqsC, which is involved in the previously mentioned HHQ and PQS synthesis, can load acyl moieties with different alkyl chain lengths. Loading experiments with a series of acyl-CoAs (C2-C14) revealed significantly increasing retention times with increasing chain length. While only one such case of shifting retention times is known at our institute so far it emphasizes that required separation effort is an individual task depending on the target protein. Regarding the analysis of tomaymycin and tilivalline biosynthesis proteins, the chromatographic separation of both modules was achievable, whereas it was not possible to sufficiently separate loaded and unloaded protein. It is thus comprehensible that difficulties will increase tremendously by dealing with *in vitro* systems containing more than just two modules including different loading stages. However, this work revealed the potential of comprehensive intact protein analysis in *in vitro* studies towards the elucidation of biosynthetic pathways.

5.1.4 Industrial Usage of *in vitro* Platforms

Considering the above-mentioned advantages of *in vitro* systems, large-scale application of these systems is also of high interest for industrial biotechnology. Fundamentally, enzymes are macromolecular biocatalysts, which are able to perform a plethora of chemical reactions. More precisely, Mother Nature outcompetes even the most skilled organic chemists when it comes to distinct substrate-, chemo-, regio-, and stereoselectivity⁴⁸. Even seemingly easy intentions like targeted introduction of stereo centers, the addition of distinct groups, the reduction at specific positions, or separation of a racemate is a challenging task for synthesis chemists⁴⁹. Consequently, not only in academia but also in industry, scientists are trying to take advantage of the tools provided by nature. Two methods can be followed for this purpose: either an *in vivo* or an *in vitro*

approach whereas both methods have their pros and cons. But in both cases, the variety of publications addressing this topic reveal that most of the known applications are limited to single biotransformations so far⁴⁸. Industry uses, for instances, enzymatic oxygenation reactions to convert cheap raw materials, usually derived from petroleum industry, into more valuable products. Typical examples are the conversion of p-xylene to terephthalic acid, ethylene to ethylene oxide, ethylene to acetaldehyde, and ethylene/acetic acid to vinyl acetate⁴⁸. Another application in industry is based on immobilized enzymes, which can be used in flow-through reactors to obtain diverse biotransformations. The most prominent example of such processes is probably the isomerization of glucose to fructose by the enzyme class of glucose isomerases. Thus, high fructose corn syrup is obtained from corn syrup in a product scale of 10^7 tons per year⁵⁰. Beside such fully established systems the jet fuel farnesene, could be a promising candidate for the first application of large-scale *in vitro* reconstitution. Zhu et al. managed to reconstitute farnesene from acetyl-CoA in decent amounts, after optimizing the ratios of proteins, substrates, and co-factors⁵¹. Nevertheless, the real potential of Mother Nature's synthesis skills is widely underexploited. It is conceivable that in the future, complete synthesis of natural products or derivatives thereof will be performed in bioreactors. Complex precursors or substructures will be presumably achieved first by *in vitro* reconstitutions and will help to facilitate semi-synthetic approaches. Hence, the supply of complicated chemical structures for various applications will be improved. Unfortunately, despite the promising perspectives it is still a long road ahead until such full *in vitro* reconstitution systems can be applied easily for large-scale purposes. The work presented here and the related publications describe the immense potential of *in vitro* reconstitution systems whilst also taking into account the upcoming challenges when such systems shall be established. It is highly complicated to obtain sufficient amounts of pure proteins, to preserve the functionality, to regenerate activated substrates or co-factors, and to realize high turnover rates for the proteins, just to mention a few obstacles. A large-scale application in industry will remain in the distant future as long as these problems cannot be addressed and the efforts regarding this are costlier compared to classical synthetic or isolation approaches.

5.2 Alternative Approaches towards the Isolation of Unknown Metabolites

Myxobacteria already confirmed their potential as source for new natural products in the last decades. Since the first research efforts by Höfle et al. and Reichenbach et al. in the 1980s^{52–55}, these distinct bacteria were investigated systematically, which led to the discovery of many previously unknown natural products and, in particular, bioactive compounds⁵⁶. An extensive screening project was initiated where an increasing number of myxobacterial strains were isolated, cultivated and subsequently analyzed towards bioactive secondary metabolites. Over the years, this screening approach yielded roughly 900 new compounds out of 140 compound classes, which were isolated and characterized. Considering that only a rather small number of < 12 000 strains has been screened, the number of isolated compounds is substantial (data from in-house database). However, it is noteworthy that the success of this approach is certainly attributed to the tremendous development in the field of analytical chemistry. Increasing sensitivity, resolution, mass accuracy, and separation efficiency allowed for improved analyses of the highly complex crude extracts, which frequently contain up to 3000 individual compounds. Nevertheless, it is easily comprehensible that besides the discovery of new compounds, the identification and especially the purification of one single compound out of a mixture of thousands remains a challenging endeavor.

Despite the continuous development of methods and instrumentation, the isolation of new compounds from myxobacterial extracts is getting more challenging. For most of the myxobacterial families, the highly abundant compounds are already elucidated and many others are repeatedly detected within the extracts of one genus. Even if the increasing number of sequenced strains still reveals the yet unexploited biosynthetic potential which can be deduced from the number of unassigned biosynthetic gene clusters, the corresponding compounds remain undiscovered^{57–59}. It is certainly conceivable that some of the missing secondary metabolites are just produced in trace amounts and therefore escape the detection by standardized measurements. While the limit of detection can be addressed by extremely sensitive MS setups like FT-ICR-MS which are already able to detect 1000 ions, the efforts to isolate sufficient amounts of such low concentrated compounds is still tremendous^{60,61}. In some cases, like in the here presented work, a metabolome mining approach can help to overcome low production yields by finding producers with higher titers of the sought compound. A successful application of this approach, however, requires a metabolite, which is not restricted to a certain strain. Besides the low abundance, it is often more likely that the responsible biosynthetic gene cluster is not activated under cultivation conditions applied. The bacteria might need very specific signaling molecules or environmental conditions to activate some of the so called silent gene clusters⁶². Therefore, expanding screening efforts in various directions

are followed to increase the chance to find the missing metabolites. One possibility to search compounds of presumably inactive gene clusters is based on the so-called genome mining approach. Once a gene cluster for a secondary metabolite is located, there is the possibility to search for it in sequence data of other strains. Following this approach provides a certain probability to find a strain in which this gene section is translated into the respective megaenzymes that produce the missing compound⁶³. However, such a procedure requires sequence data from a large number of different strains. As so far only a negligible number of all isolated myxobacterial strains is fully sequenced, scientist can unfortunately not exploit the maximum potential of this workflow. Another approach attempts the activation or upregulation of such silent gene clusters by genetic modifications of the DNA and a particular promoter replacement. Compared to other bacteria, this is a challenging task in myxobacteria as genetic modifications cannot readily be introduced into myxobacterial genomes⁶⁴. Nevertheless, this methodology is heretofore underexploited for myxobacteria but seems to offer promising potential as it was confirmed by initial experiments at the Helmholtz Institute. Following this workflow led to the detection of three so far unknown compounds while one was already isolated and the chemical structure elucidated (Fabian Panter, personal communication, February 2017). In addition to the described methods, co-cultivation based search is another option in natural products research which is underexploited so far⁶⁵. The idea behind this concept is, that the biosynthesis of anti-microbial substances is induced by the presence of a competitor. Some competitors might excrete distinct molecules, which are recognized by the myxobacteria as an indicator for their existence in the immediate vicinity and cause the activation of defense or even predatory mechanisms. Although such consideration is comprehensible, it is difficult to prove such dependence which is also reflected in the limited number of publication⁶⁶. Various approaches were already tested at the Helmholtz Institute, whereas for most of the methods an implementation remains challenging. A promising workflow is the co-cultivation of two myxobacterial strains on agar with subsequently performed MALDI-imaging measurements on a FT-ICR-MS platform. Based on the immense quantity of data, it was possible to detect a variety of discriminating signals and reveal compounds that are directly correlated to a co-cultivation. However, the following isolation and structure elucidation approaches were not met with success due to the very limited quantity of compound per plate. The extractable surface area is even smaller than in the described fruiting body approach because the compounds of interest are predominantly limited to the interaction area between both organisms.

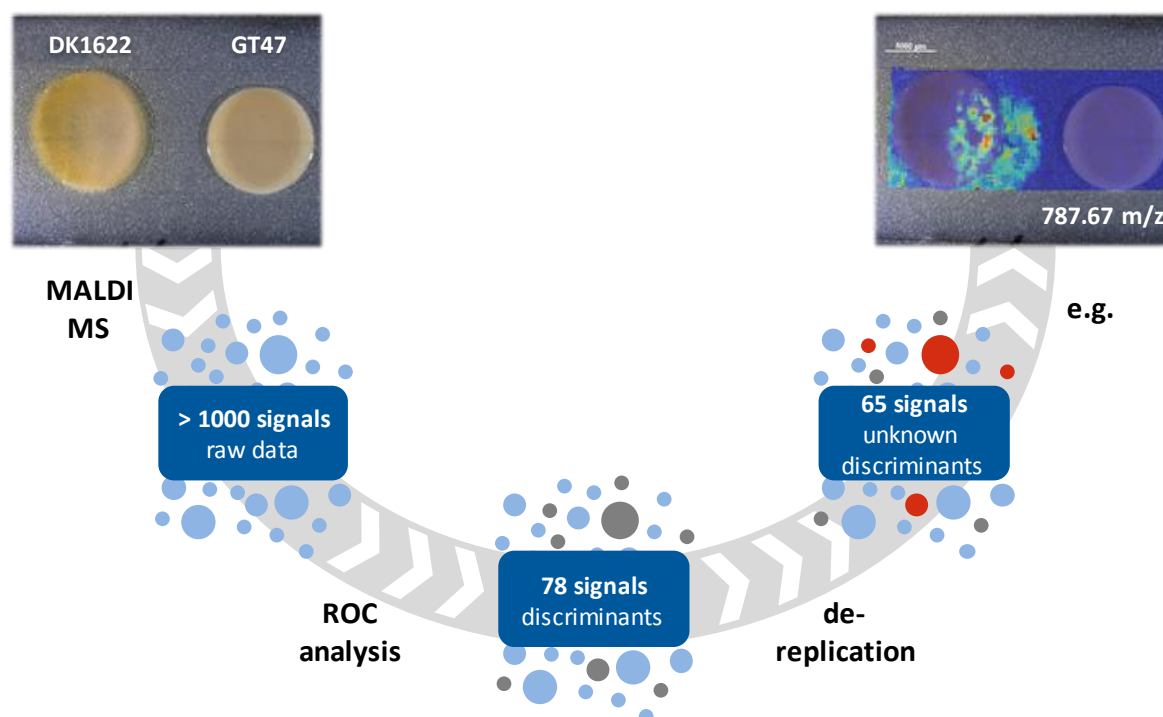


Figure 5.4: Exemplary workflow for MALDI-MSI data evaluation. Complex raw data from MALDI-MSI measurements of *Myxococcus xanthus* DK1622/ *Sorangium cellulosum* So ceGT47 co-cultivation were first reduced by receiver operating characteristic (ROC) analysis and subsequently by dereplication to reveal unknown compounds which are directly related to the co-cultivation.

Nevertheless, the impressive results of the co-cultivation experiments finally led to the herein described fruiting body approach. While co-cultivation is rather complicated to realize in large replicates, the cultivation of single agar plates is fairly simple. Hence, the question arose whether the secondary metabolites of myxobacteria growing on agar and forming fruiting bodies differ from those metabolites produced in liquid culture of the same myxobacterial strain. The screening project of myxobacteria at the Helmholtz Institute is so far based on cultivation of bacterial strains in liquid media in the presence of adsorber resins. This approach is due to several reasons: transformations (electroporation, conjugation, transduction) to insert genetic modifications requires suspension cultures; supplementation of precursor for feeding studies is difficult on agar; scale-up is usually based on suspension cultures in bio fermenters; cell viability can easily be determined by microscopic examination of liquid culture. Thus, in contrast to the exploration of the cell cultures growing on agar plates, the workflow for the preparation of extracts is well established for liquid media based cultivations⁶⁷. A workflow for the preparation of extracts from cell cultures growing on agar had to be established first but once a feasible workflow was elaborated, the initial analyses were astonishing. Contrary to the expectations, the base peak chromatograms of fruiting body extracts and liquid culture extract of *Myxococcus xanthus* DK1622 as well as *Chondromyces crocatus* Cm c5 showed fundamental differences.

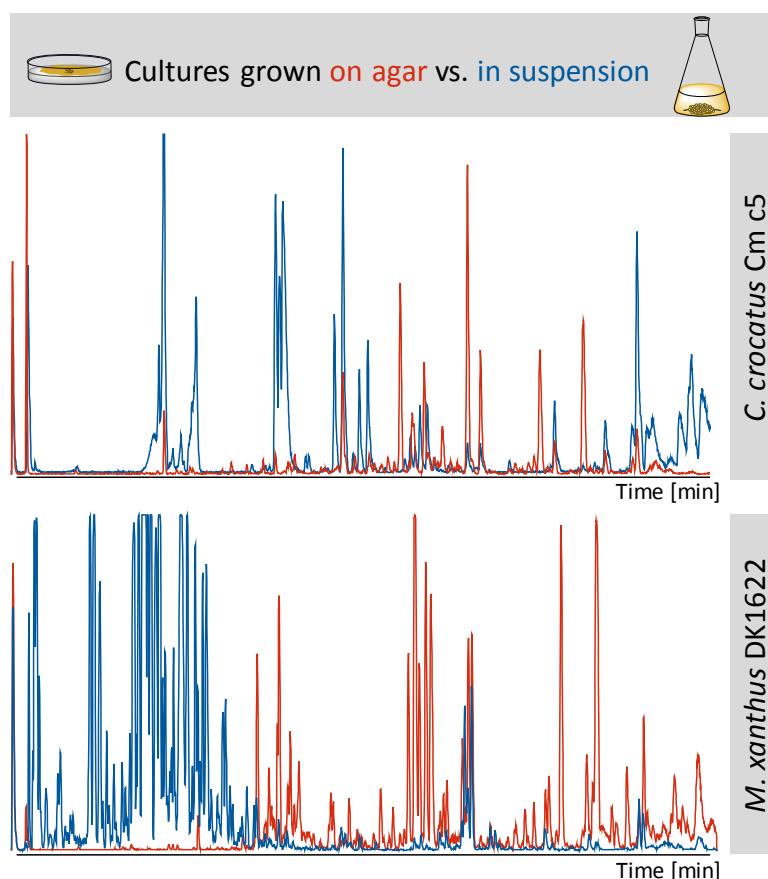


Figure 5.5: Data of initial experiment revealed differences in metabolism due to different cultivation approaches. Comparative BPCs of *M. xanthus* DK1622 and *C. crocatus* Cm c5 extracts originated from cultures grown on agar (red) and grown in suspension (blue), respectively.

Certainly, not all differences can be referred to the presence or absence of respective secondary metabolites but it became obvious that myxobacterial life cycles have a significant influence on the natural products produced at the specific lifecycle stage. A valid and reliable evaluation by manual comparison of such data sets is, however, impractical, owing to the already mentioned complexity. Based on high performance analytical setups, statistical tools like PCA, t-tests or ROC analysis can facilitate such intentions. The usage of these techniques is not limited to the questions described in this work, but is already rather established in natural product research. Thus, media components are easily distinguishable from organism derived products by respective design of experiment; this helps to focus on actual new compounds⁶⁸. Recently, the usage of PCA results was even further improved to acquire precursor directed MS/MS measurements of myxobacterial extracts, which increase the information content for distinct compounds and avoid the fragmentation of already known substances or media components⁶⁹. Acquiring such data sets is highly beneficial in particular to build databases with the information about thousands of different strains, as it will become easier to select the promising candidates in the plethora of signals. In combination with the development of molecular networking techniques like GNPS, the advanced analysis even enables the grouping of various signals to a presumable compound class⁷⁰.

Consequently, it is often not mandatory to isolate and elucidate all selected unknowns, as the information about the networking and the fragmentation pattern could already reveal the chemical structures of some derivatives (unpublished results about argyirin derivatives by Tang et al.). Even if the data does not unambiguously prove the predicted structures, feeding experiments can be performed to confirm the structures nearly without any doubt. Especially for derivatives in very low concentrations, this approach is a powerful tool. While in this work a comprehensive application of the described techniques helped to trace 62 unknown compounds for only one strain (*Myxococcus xanthus* DK1622), the clear allocation to vegetative cells and/or fruiting body formation was elaborated. Based on the results of these measurements it was not possible to ensure that those compounds were just related to the growth on solid medium. This issue was addressed by applying MALDI imaging experiments at different time points of the life cycle and analyzing the data towards the spatial distribution of selected signals. Hence, the potential linkage between fruiting body formation and the appearance of distinct compounds like Cmp-552 could successfully be confirmed. MALDI-MSI has become a popular method in various fields of research and showed its potential not only in proteomics but also in metabolomics^{71–73}. However, it is noteworthy that in most cases application of this technique aims at tissue imaging, which seems much easier as imaging of microbes growing on agar^{74–76}. During the last years, continuous improvement of techniques, in particular, sample preparation and matrix application, helped to overcome many initial obstacles and enabled a reasonable use in the detection of metabolites⁷⁷. MALDI is known to be highly sensitive, especially concerning lipids and peptides⁷⁸, which are certainly present in bacterial samples, but not in the center of interest for the here presented research. But it is especially the high sensitivity regarding lipids and peptides hampering the detection of here sought metabolites. Lipid concentrations are usually surpassing the concentrations of NRPS/PKS derived metabolites and results in ion suppression effects which than hamper the detection of the low abundant metabolites^{79–81}. In addition, substantial amounts of complex media components and salts in the cultivation agar are leading to the same effect and makes the detection even more complicated. This issue can also be shown by a detailed analysis of the here presented data. Many of the verified 62 discriminating signals based on LC-MS measurements could not be detected in the respective MALDI-MSI data set, though plausible adducts have been taken into account. This observation applies, in particular, to the low intense signals detected by LC-MS, although it is worth mentioning that the extracts for LC-MS measurements were derived by combining the extracts of five single plates which led to higher metabolite concentrations. The sensitivity of this method can certainly be improved by using a FT-ICR-MS instead of a ToF-MS, but this will probably not completely circumvent this issue. Nevertheless, the work showed that a combination of high performance LC-MS and MALDI-MSI can increase the information content and help to get better

insights in myxobacterial life cycles. For example, not only compounds, which are directly related to fruiting body formation, were detected but also molecules, which are demonstrably showing an inverse correlation. This includes the compound with 724.398 m/z , which has exclusively been observed in the area of vegetative cells while it completely disappeared in areas where fruiting body formation took place.

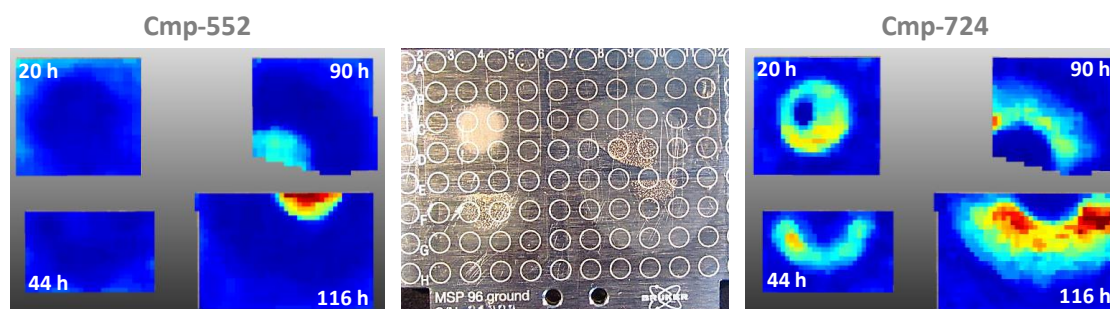


Figure 5.6: MALDI-MSI data for Cmp-552 and Cmp-724 in dependency on different cultivation times of *Myxococcus xanthus* DK1622 showing an inverse appearance. The photo in the middle shows the respective samples on a MALDI target prior to matrix application.

Once so much effort was invested to find unknown compounds, which are directly related to fruiting body formation, the isolation of sufficient amounts for NMR analysis and bioactivity testing remained challenging. It seemed obvious that there were no alternatives to the preparation of multitudes of cultures on agar plates. Although this proceeding led to sufficient amounts of the target compound Cmp-552, the isolation of some lower intense compounds remained difficult. Currently, this is the bottleneck of the presented approach but could probably be eliminated by a high-throughput preparation and extraction of the respective agar plates. As already indicated, a metabolome mining approach could also be applied to circumvent the low product yields. In the here described work, not all derivatives of Cmp-552 could be found in other extracts but MS/MS data allows a prediction of the underlying chemical structures. Targeted MS/MS experiments with the extracts from a potential alternative producer including subsequently performed networking analysis revealed further a mass shift of two Dalton for many of the detected derivatives. This indicates that the metabolites produced by *Myxococcus* MCy9290 predominantly exhibit double bonds, while derivatives of Cmp-552 in DK1622 are lacking this structural element. Probably, the differences are due to the variations of the attached fatty acids, which presumably originate from the primary metabolism. Such observations are not surprising, as the link between production of certain fatty acids and lipids and fruiting body formation has been studied intensively^{82,83}. Thus, changes in the fatty acid and lipid profiles are expected when myxobacteria pass through their various lifecycle stages. Different compositions of fatty acids also seem to be unique for different myxobacterial genera and can be utilized for phylogeny studies⁸⁴.

Bioactivity testing for Cmp-552 and all isolated derivatives showed moderate cytotoxicity as well as antibiotic activity against some Gram-positive pathogens. Cytotoxic and microbial activity is a commonly observed characteristic among natural products from myxobacteria and many compounds with outstanding bioactivity values were isolated within the last years. Not only epothilone but also compounds like disorazol, argyirin, tubulysin moved to the center of interest due to their cytotoxicity. With regard to antibacterial effects, disciformycin, chlorotonil and coralopyronin are exemplary representatives⁵⁶. As already discussed in the introduction, some of the observed bioactivities are unexpected as the respective target organism like, for instance, *plasmodium* or retroviruses does definitely not appear in the natural habitat of myxobacteria. Observed activities regarding these targets are rarely the only detected bioactivity and may be due to comparable active sites structures and are thus probably a result of coincidental protein similarity. In these cases, the real function of the compound just remains obscure. It is however obvious those myxobacteria synthesize secondary metabolites for a certain purpose, as the production is an energy intensive process. The presence of respective gene clusters has consequently an evolutionary advantage for the myxobacteria. For instance, antibacterial or antifungal activity are comprehensible as myxobacteria are mainly surrounded by other bacterial or fungal competitors in their natural habitat⁶⁶. Thus, there is a certain reason either to defend their settlement area or to use the surrounding competitors as source of nutrition by the secretion of antibacterial compounds, especially during starvation and/or harsh environmental conditions. However, despite the increasing knowledge about microbes it is still a long way to fully understand nature's evolutionary developed approaches. The reasons for the biosynthesis of compounds remain, in many cases, obscure and it can only be speculated about required triggers to activate silent gene clusters.

In summary, the novel approach described in this work demonstrates a high potential regarding the detection of yet unknown myxobacterial natural products and could help to increase the number of characterized new chemical scaffolds in the coming years. Furthermore, the workflow confirmed once again the importance of a rational design of experiments, the targeted application of diverse analytical setups as well as a detailed evaluation of acquired data to receive high quality information, which can settle outstanding issues. The combination of new arising methodologies, optimized workflows, better understanding of backgrounds, and longstanding experience created a magnificent toolbox for the research of myxobacteria within the last years. These tools open up new and as yet unknown paths, which provide new insight to myxobacteria and help to enable the interdisciplinary networking of various research areas. The here presented work formed only the foundation for the build-up of a new approach, finding so far unknown myxobacterial metabolites. The isolated compound could not be allocated to one of the silent gene

clusters of *Myxococcus xanthus* DK1622 but the workflow has been proven to be feasible and a new compound class has been identified. Despite the fairly simple structures, the compounds revealed moderate bioactivity values and emphasized once again that chemical structures can lead to false impressions regarding their bioactivity. Structure complexity does not correlate with bioactivity, as underlined by a view on the currently approved drugs. Many rather simple molecules show striking bioactivity and are widely used for treatments. The probably most prominent representative is acetylsalicylic acid, first marketed under the name Aspirin®, which is even listed in the WHO Model List of Essential Medicines (EML)⁸⁵. Thus, isolation efforts should not be restricted to compounds with presumably extraordinary structure elements. The presented data revealed a variety of still unknown compounds for only one single strain, which could be further pursued. It is conceivable to include our new agar plate based approach for investigation of the myxobacterial metabolite profile into the standard setup of myxobacterial characterization since production of myxobacterial cultures on solid media is in any case performed routinely. The comparison of metabolism must not necessarily be restricted to fruiting bodies and cells growing in liquid media, but can basically include cells from solid based cultivations and liquid based cultivations. For some myxobacteria, investigation of the fruiting body metabolome would even not be achievable, due to their inability to form fruiting bodies. The only limitation of the here described workflow would be myxobacteria belonging to the agar degrading suborder as *Sorangineae* and *Nannocystineae*, while in these limited cases an extraction of the complete agar plate could be performed. Even if the distinct workflow for these cases is more elaborated it would be surpassed by the possibility to explore myxobacterial strains which cannot be cultured in liquid medium so far. Besides the almost complete absence of media components in the extracts and the thus resulting “clear view” on the metabolism, this would be the big advantage of such an extension of screening. The missing cultivation and harvesting automation to enable a high-throughput screening and simplify the isolation of sufficient compound amounts is currently the bottleneck and the main challenge, which should be tackled to implement the workflow into standardized screening projects. The approach will certainly help to get further insights toward the almost inexhaustible source of natural products from myxobacteria.

5.3 References

- [1] Tozuka, Z. & Takaya, T.: Studies on tomaymycin. I. The structure determination of tomaymycin on the basis of NMR spectra. *J. Antibiot. (Tokyo)*. **1983**, 36, 142–6, DOI: 10.7164/antibiotics.36.142
- [2] Arima, K., Kosaka, M., Tamura, G., Imanaka, H. & Sakai, H.: Studies on tomaymycin, a new antibiotic. I. Isolation and properties of tomaymycin. *J. Antibiot. (Tokyo)*. **1972**, 25, 437–44, DOI: 10.7164/antibiotics.25.437
- [3] Darby, A. *et al.*: Cytotoxic and pathogenic properties of *Klebsiella oxytoca* isolated from laboratory animals. *PLoS One* **2014**, 9, DOI: 10.1371/journal.pone.0100542
- [4] Schneditz, G. *et al.*: Enterotoxicity of a nonribosomal peptide causes antibiotic-associated colitis. *Proc. Natl. Acad. Sci.* **2014**, 111, 13181–13186, DOI: 10.1073/pnas.1403274111
- [5] Mohr, N. & Budzikiewicz, H.: Tilivalline, a new pyrrolo[2, 1-c][1,4] benzodiazepine metabolite from *klebsiella*. *Tetrahedron* **1982**, 38, 147–152, DOI: 10.1016/0040-4020(82)85058-8
- [6] Gerratana, B.: Biosynthesis, synthesis, and biological activities of pyrrolobenzodiazepines. *Med. Res. Rev.* **2012**, 32, 254–293, DOI: 10.1002/med.20212
- [7] Lowry, B., Walsh, C. T. & Khosla, C.: In Vitro reconstitution of metabolic pathways: Insights into nature's chemical logic. *Synlett* **2015**, 26, 1008–1025, DOI: 10.1055/s-0034-1380264
- [8] Lund, E.: ERYTHROMYCIN. *Acta Pathol. Microbiol. Scand.* **2009**, 33, 393–400, DOI: 10.1111/j.1699-0463.1953.tb01535.x
- [9] Woodward, R. B. *et al.*: Asymmetric total synthesis of erythromycin. 1. Synthesis of an erythronolide A secoacid derivative via asymmetric induction. *J. Am. Chem. Soc.* **1981**, 103, 3210–3213, DOI: 10.1021/ja00401a049
- [10] Gao, X., Woo, S. K. & Krische, M. J.: Total Synthesis of 6-Deoxyerythronolide B via C–C Bond-Forming Transfer Hydrogenation. *J. Am. Chem. Soc.* **2013**, 135, 4223–4226, DOI: 10.1021/ja4008722
- [11] Minas, W., Brunker, P., Kallio, P. T. & Bailey, J. E.: Improved Erythromycin Production in a Genetically Engineered Industrial Strain of *Saccharopolyspora erythraea*. *Biotechnol. Prog.* **1998**, 14, 561–566, DOI: 10.1021/bp980055t
- [12] Lowry, B. *et al.*: In vitro reconstitution and analysis of the 6-deoxyerythronolide B synthase. *J. Am. Chem. Soc.* **2013**, 135, 16809–16812, DOI: 10.1021/ja409048k
- [13] Greunke, C., Glöckle, A., Antosch, J. & Gulder, T. A. M.: Biokatalytische Totalsynthese von Ikarugamycin. *Angew. Chemie* **2017**, 129, 4416–4420, DOI: 10.1002/ange.201611063
- [14] Crawford, J. M. *et al.*: Deconstruction of iterative multidomain polyketide synthase function. *Science* **2008**, 320, 243–246, DOI: 10.1126/science.1154711
- [15] Levengood, M. R., Knerr, P. J., Oman, T. J. & Van Der Donk, W. A.: In vitro mutasynthesis of lantibiotic analogues containing nonproteinogenic amino acids. *J. Am. Chem. Soc.* **2009**, 131, 12024–12025, DOI: 10.1021/ja903239s
- [16] Goss, R. J. M., Shankar, S. & Fayad, A. A.: The generation of 'unNatural' products: Synthetic biology meets synthetic chemistry. *Nat. Prod. Rep.* **2012**, 29, 870, DOI: 10.1039/c2np00001f

- [17] Lowry, B. *et al.*: In Vitro Reconstitution and Analysis of the 6-Deoxyerythronolide B Synthase. *J. Am. Chem. Soc.* **2013**, 135, 16809–16812, DOI: 10.1021/ja409048k
- [18] Rossi, A. *et al.*: The 5-lipoxygenase inhibitor, zileuton, suppresses prostaglandin biosynthesis by inhibition of arachidonic acid release in macrophages. *Br. J. Pharmacol.* **2010**, 161, 555–570, DOI: 10.1111/j.1476-5381.2010.00930.x
- [19] Laufer, S., Greim, C. & Bertsche, T.: An in-vitro screening assay for the detection of inhibitors of proinflammatory cytokine synthesis: a useful tool for the development of new antiarthritic and disease modifying drugs. *Osteoarthr. Cartil.* **2002**, 10, 961–967, DOI: 10.1053/joca.2002.0851
- [20] Hoffmann, T., Müller, S., Nadmid, S., Garcia, R. & Müller, R.: Microsclerodermins from terrestrial myxobacteria: an intriguing biosynthesis likely connected to a sponge symbiont. *J. Am. Chem. Soc.* **2013**, 135, 16904–11, DOI: 10.1021/ja4054509
- [21] Jahns, C. *et al.*: Pellasoren: Struktur, Biosynthese und Totalsynthese eines zytotoxischen Sekundärmetaboliten aus *Sorangium cellulosum*. *Angew. Chemie* **2012**, 124, 5330–5334, DOI: 10.1002/ange.201200327
- [22] Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L. & Sivonen, K.: Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, 111, 9259–9264, DOI: 10.1073/pnas.1401734111
- [23] Wenzel, S. C. & Müller, R.: Myxobacterial natural product assembly lines: fascinating examples of curious biochemistry. *Nat. Prod. Rep.* **2007**, 24, 1211, DOI: 10.1039/b706416k
- [24] Reuven, N. B., Koonin, E. V., Rudd, K. E. & Deutscher, M. P.: The gene for the longest known *Escherichia coli* protein is a member of helicase superfamily II. *J. Bacteriol.* **1995**, 177, 5393–400, DOI: 0021-9193/95/\$04.00+0
- [25] Fakruddin, M., Mohammad Mazumdar, R., Bin Mannan, K. S., Chowdhury, A. & Hossain, M. N.: Critical Factors Affecting the Success of Cloning, Expression, and Mass Production of Enzymes by Recombinant *E. coli*. *ISRN Biotechnol.* **2013**, 2013, 590587, DOI: 10.5402/2013/590587
- [26] Herbst, D. A., Boll, B., Zocher, G., Stehle, T. & Heide, L.: Structural Basis of the Interaction of MbtH-like Proteins, Putative Regulators of Nonribosomal Peptide Biosynthesis, with Adenylating Enzymes. *J. Biol. Chem.* **2013**, 288, 1991–2003, DOI: 10.1074/jbc.M112.420182
- [27] Felnagle, E. A. *et al.*: MbtH-Like Proteins as Integral Components of Bacterial Nonribosomal Peptide Synthetases. *Biochemistry* **2010**, 49, 8815–8817, DOI: 10.1021/bi1012854
- [28] Krug, D. & Müller, R.: Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products. *Nat. Prod. Rep.* **2014**, 31, 768, DOI: 10.1039/c3np70127a
- [29] Krug, D., Zurek, G., Schneider, B., Garcia, R. & Müller, R.: Efficient mining of myxobacterial metabolite profiles enabled by liquid chromatography-electrospray ionisation-time-of-flight mass spectrometry and compound-based principal component analysis. *Anal. Chim. Acta* **2008**, 624, 97–106, DOI: 10.1016/j.aca.2008.06.036
- [30] Schauer, N. *et al.*: GC-MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Lett.* **2005**, 579, 1332–1337, DOI: 10.1016/j.febslet.2005.01.029
- [31] Wenzel, S. C. *et al.*: Production of the Bengamide Class of Marine Natural Products in

- Myxobacteria: Biosynthesis and Structure-Activity Relationships. *Angew. Chemie Int. Ed.* **2015**, 54, 15560–15564, DOI: 10.1002/anie.201508277
- [32] Jungmann, K. *et al.*: Two of a Kind-The Biosynthetic Pathways of Chlorotoni and Anthracimycin. *ACS Chem. Biol.* **2015**, 10, DOI: 10.1021/acscchembio.5b00523
- [33] Keller, L. *et al.*: Macyranonones: Structure, Biosynthesis, and Binding Mode of an Unprecedented Epoxyketone that Targets the 20S Proteasome. *J. Am. Chem. Soc.* **2015**, 137, 8121–8130, DOI: 10.1021/jacs.5b03833
- [34] Dewick, P. M.: *Medicinal Natural Products. J. Biosci. Bioeng.* (John Wiley & Sons, Ltd, **2009**). 110, DOI: 10.1002/9780470742761
- [35] Sun, X., Li, H., Alfermann, J., Mootz, H. D. & Yang, H.: Kinetics profiling of gramicidin S synthetase A, a member of nonribosomal peptide synthetases. *Biochemistry* **2014**, 53, 7983–7989, DOI: 10.1021/bi501156m
- [36] Gode, D., Schmitt, C., Engel, M. & Volmer, D. A.: Screening Dyrk1A inhibitors by MALDI-QqQ mass spectrometry: systematic comparison to established radiometric, luminescence, and LC–UV–MS assays. *Anal. Bioanal. Chem.* **2014**, 406, 2841–2852, DOI: 10.1007/s00216-014-7703-1
- [37] Simithy, J., Gill, G., Wang, Y., Goodwin, D. C. & Calderón, A. I.: Development of an ESI-LC-MS-based assay for kinetic evaluation of mycobacterium tuberculosis shikimate kinase activity and inhibition. *Anal. Chem.* **2015**, 87, 2129–2136, DOI: 10.1021/ac503210n
- [38] Toby, T. K., Fornelli, L. & Kelleher, N. L.: Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu. Rev. Anal. Chem.* **2016**, 9, 499–519, DOI: 10.1146/annurev-anchem-071015-041550
- [39] Kelleher, N. L.: Peer Reviewed: Top-Down Proteomics. *Anal. Chem.* **2004**, 76, 196 A-203 A, DOI: 10.1021/ac0415657
- [40] Belecki, K. & Townsend, C. a: Biochemical Determination of Enzyme-Bound Metabolites: Preferential Accumulation of a Programmed Octaketide on the Eneidyne Polyketide Synthase CalE8. *J. Am. Chem. Soc.* **2013**, 135, 14339–48, DOI: 10.1021/ja406697t
- [41] Hicks, L. M. *et al.*: Investigating Nonribosomal Peptide and Polyketide Biosynthesis by Direct Detection of Intermediates on >70 kDa Polypeptides by Using Fourier-Transform Mass Spectrometry. *ChemBioChem* **2006**, 7, 904–907, DOI: 10.1002/cbic.200500416
- [42] Catherman, A. D., Skinner, O. S. & Kelleher, N. L.: Top Down proteomics: Facts and perspectives. *Biochem. Biophys. Res. Commun.* **2014**, 445, 683–693, DOI: 10.1016/j.bbrc.2014.02.041
- [43] Popov, I. a, Nagornov, K., Vladimirov, G. N., Kostyukevich, Y. I. & Nikolaev, E. N.: Twelve million resolving power on 4.7 T Fourier transform ion cyclotron resonance instrument with dynamically harmonized cell-observation of fine structure in peptide mass spectra. *J. Am. Soc. Mass Spectrom.* **2014**, 25, 790–9, DOI: 10.1007/s13361-014-0846-7
- [44] Nikolaev, E. N., Kostyukevich, Y. I. & Vladimirov, G. N.: Fourier transform ion cyclotron resonance (FT ICR) mass spectrometry: Theory and simulations. *Mass Spectrom. Rev.* **2016**, 35, 219–258, DOI: 10.1002/mas.21422
- [45] Kastner, M.: *Protein Liquid Chromatography.* (Elsevier Science, **1999**).
- [46] Bush, D. R., Zang, L., Belov, A. M., Ivanov, A. R. & Karger, B. L.: High Resolution CZE-MS

- Quantitative Characterization of Intact Biopharmaceutical Proteins: Proteoforms of Interferon- γ . *Anal. Chem.* **2016**, 88, 1138–1146, DOI: 10.1021/acs.analchem.5b03218
- [47] Han, X., Wang, Y., Aslanian, A., Bern, M. & Iii, J. R. Y.: Sheathless Capillary Electrophoresis - Tandem Mass Spectrometry for Top - Down Characterization of *Pyrococcus furiosus* Proteins on a Proteome Scale. *Anal. Chem.* **2014**, 1, 1–7, DOI: dx.doi.org/10.1021/ac503439n
- [48] Faber, K.: *Biotransformations in Organic Chemistry. Igarss 2014* (Springer Berlin Heidelberg, **2014**). DOI: 10.1007/s13398-014-0173-7.2
- [49] Tobergte, D. R. & Curtis, S.: *From Biosynthesis to Total Synthesis*. (John Wiley & Sons, Inc, **2013**). 53, DOI: 10.1017/CBO9781107415324.004
- [50] DiCosimo, R., McAuliffe, J., Poulouse, A. J. & Bohlmann, G.: Industrial use of immobilized enzymes. *Chem. Soc. Rev.* **2013**, 42, 6437, DOI: 10.1039/c3cs35506c
- [51] Zhu, F. *et al.*: In vitro reconstitution of mevalonate pathway and targeted engineering of farnesene overproduction in *Escherichia coli*. *Biotechnol. Bioeng.* **2014**, 111, 1396–1405, DOI: 10.1002/bit.25198
- [52] Kunze, B., Reichenbach, H., Augustiniak, H. & Höfle, G.: Isolation and identification of althiomycin from *Cystobacter fuscus* (myxobacterales). *J. Antibiot. (Tokyo)*. **1982**, 35, 635–6,
- [53] Gerth, K. *et al.*: Pyrrolnitrin from *Myxococcus fulvus* (Myxobacterales). *J. Antibiot. (Tokyo)*. **1982**, 35, 1101–3,
- [54] Gerth, K. & Reichenbach, H.: Induction of myxospore formation in *Stigmatella aurantiaca* (Myxobacterales). *Arch. Microbiol.* **1978**, 117, 173–182, DOI: 10.1007/BF00402305
- [55] Kleinig, H. & Reichenbach, H.: A new type of carotenoid pigment isolated from myxobacteria. *Naturwissenschaften* **1970**, 57, 92–93, DOI: 10.1007/BF00590700
- [56] Herrmann, J., Fayad, A. A. & Müller, R.: Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **2017**, 34, 135–160, DOI: 10.1039/C6NP00106H
- [57] Hoffmann, T.: Screening for myxobacterial natural products : new structures, biosynthesis, and contributions to a comprehensive screening workflow. **2014**,
- [58] Goldman, B. S. *et al.*: Evolution of sensory complexity recorded in a myxobacterial genome. *Proc. Natl. Acad. Sci.* **2006**, 103, 15200–15205, DOI: 10.1073/pnas.0607335103
- [59] Schneiker, S. *et al.*: Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat. Biotechnol.* **2007**, 25, 1281–1289, DOI: 10.1038/nbt1354
- [60] Nikolaev, E. N., Heeren, R. M. A., Popov, A. M., Pozdneev, A. V. & Chingin, K. S.: Realistic modeling of ion cloud motion in a Fourier transform ion cyclotron resonance cell by use of a particle-in-cell approach. *Rapid Commun. Mass Spectrom.* **2007**, 21, 3527–3546, DOI: 10.1002/rcm.3234
- [61] N. Nikolaev, E., N. Vladimirov, G., Jertz, R. & Baykut, G.: From Supercomputer Modeling to Highest Mass Resolution in FT-ICR. *Mass Spectrom.* **2013**, 2, S0010–S0010, DOI: 10.5702/massspectrometry.S0010
- [62] Rutledge, P. J. & Challis, G. L.: Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **2015**, 13, 509–523, DOI:

- 10.1038/nrmicro3496
- [63] Feng Qiu, J.B. McAlpine, E.C. Krause, Shao-nong Chen, and G.F. Pauli N. Gaboriaud-Kolar, S. Nam, and A.-L. S. M. I. M. N.: *Progress in the Chemistry of Organic Natural Products* 99. (Springer International Publishing, **2014**). 99, DOI: 10.1007/978-3-319-04900-7
- [64] Yang, Z., Rodriguez, E., Julien, B. & Black, W. P.: *Man. Ind. Microbiol. Biotechnol. Third Ed.* (American Society of Microbiology, **2010**). 15, 262–272, DOI: 10.1128/9781555816827.ch18
- [65] Bertrand, S. *et al.*: Metabolite induction via microorganism co-culture: A potential way to enhance chemical diversity for drug discovery. *Biotechnol. Adv.* **2014**, 32, 1180–1204, DOI: 10.1016/j.biotechadv.2014.03.001
- [66] Xiao, Y., Wei, X., Ebright, R. & Wall, D.: Antibiotic Production by Myxobacteria Plays a Role in Predation. *J. Bacteriol.* **2011**, 193, 4626–4633, DOI: 10.1128/JB.05052-11
- [67] Plaza, A. & Müller, R.: *Nat. Prod.* (John Wiley & Sons, Inc., **2014**). 103–124, DOI: 10.1002/9781118794623.ch6
- [68] Cortina, N. S., Krug, D., Plaza, A., Revermann, O. & Müller, R.: Myxoprincomide: A Natural Product from *Myxococcus xanthus* Discovered by Comprehensive Analysis of the Secondary Metabolome. *Angew. Chemie Int. Ed.* **2012**, 51, 811–816, DOI: 10.1002/anie.201106305
- [69] Hoffmann, T., Krug, D., Hüttel, S. & Müller, R.: Improving Natural Products Identification through Targeted LC-MS/MS in an Untargeted Secondary Metabolomics Workflow. *Anal. Chem.* **2014**, DOI: 10.1021/ac502805w
- [70] Wang, M. *et al.*: Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, 34, 828–837, DOI: 10.1038/nbt.3597
- [71] Svatos, A.: Mass spectrometric imaging of small molecules. *Trends Biotechnol.* **2010**, 28, 425–34, DOI: 10.1016/j.tibtech.2010.05.005
- [72] Watrous, J. D. & Dorrestein, P. C.: Imaging mass spectrometry in microbiology. *Nat. Rev. Microbiol.* **2011**, 9, 683–694, DOI: 10.1038/nrmicro2634
- [73] Bjarnholt, N., Li, B., D’Alvise, J. & Janfelt, C.: Mass spectrometry imaging of plant metabolites – principles and possibilities. *Nat. Prod. Rep.* **2014**, 31, 818–837, DOI: 10.1039/C3NP70100J
- [74] Goodwin, R. J. A.: Sample preparation for mass spectrometry imaging: Small mistakes can lead to big consequences. *J. Proteomics* **2012**, 75, 4893–4911, DOI: 10.1016/j.jprot.2012.04.012
- [75] Puolitaival, S. M., Burnum, K. E., Cornett, D. S. & Caprioli, R. M.: Solvent-Free Matrix Dry-Coating for MALDI Imaging of Phospholipids. *J. Am. Soc. Mass Spectrom.* **2008**, 19, 882–886, DOI: 10.1016/j.jasms.2008.02.013
- [76] Goodwin, R. J., MacIntyre, L., Watson, D. G., Scullion, S. P. & Pitt, A. R.: A solvent-free matrix application method for matrix-assisted laser desorption/ionization imaging of small molecules. *Rapid Commun. Mass Spectrom.* **2010**, 24, 1682–1686, DOI: 10.1002/rcm.4567
- [77] Hoffmann, T. & Dorrestein, P. C.: Homogeneous matrix deposition on dried agar for MALDI imaging mass spectrometry of microbial cultures. *J. Am. Soc. Mass Spectrom.* **2015**, 26, 1959–1962, DOI: 10.1007/s13361-015-1241-8
- [78] Fuchs, B., Süß, R. & Schiller, J.: An update of MALDI-TOF mass spectrometry in lipid research.

- Prog. Lipid Res.* **2010**, 49, 450–475, DOI: 10.1016/j.plipres.2010.07.001
- [79] Knochenmuss, R.: A bipolar rate equation model of MALDI primary and secondary ionization processes, with application to positive/negative analyte ion ratios and suppression effects. *Int. J. Mass Spectrom.* **2009**, 285, 105–113, DOI: 10.1016/j.ijms.2009.05.002
- [80] Knochenmuss, R.: A Quantitative Model of Ultraviolet Matrix-Assisted Laser Desorption/Ionization Including Analyte Ion Generation. *Anal. Chem.* **2003**, 75, 2199–2207, DOI: 10.1021/ac034032r
- [81] Mallet, C. R., Lu, Z. & Mazzeo, J. R.: A study of ion suppression effects in electrospray ionization from mobile phase additives and solid-phase extracts. *Rapid Commun. Mass Spectrom.* **2004**, 18, 49–58, DOI: 10.1002/rcm.1276
- [82] Bhat, S., Ahrendt, T., Dauth, C., Bode, H. B. & Shimkets, L. J.: Two lipid signals guide fruiting body development of *Myxococcus xanthus*. *MBio* **2014**, 5, e00939-13, DOI: 10.1128/mBio.00939-13
- [83] Bhat, S., Boynton, T. O., Pham, D. & Shimkets, L. J.: Fatty Acids from Membrane Lipids Become Incorporated into Lipid Bodies during *Myxococcus xanthus* Differentiation. *PLoS One* **2014**, 9, e99622, DOI: 10.1371/journal.pone.0099622
- [84] Garcia, R., Pistorius, D., Stadler, M. & Muller, R.: Fatty Acid-Related Phylogeny of Myxobacteria as an Approach to Discover Polyunsaturated Omega-3/6 Fatty Acids. *J. Bacteriol.* **2011**, 193, 1930–1942, DOI: 10.1128/JB.01091-10
- [85] WHO | WHO Model Lists of Essential Medicines. *WHO* **2016**,