

Identification of new natural products from nematode-associated bacteria using mass spectrometry

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

vorgelegt beim Fachbereich für Biowissenschaften (15)

der Johann Wolfgang Goethe-Universität

in Frankfurt am Main

von

Moritz Drechsler

aus Darmstadt

Frankfurt am Main 2023

D30

vom Fachbereich für Biowissenschaften (15) der
Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Sven Klimpel

Gutachter: Prof. Dr. Helge B. Bode

Zweitgutachter: Prof. Dr. Martin Grininger

Datum der Disputation: 17.07.2023

Progress is made by trial and failure; the failures are generally a hundred times more numerous than the successes; yet they are usually left unchronicled.

Sir William Ramsay

Nobel Prize in Chemistry 1904

I. TABLE OF CONTENTS

I	Table of contents	I
II	Table of abbreviations	IV
III	Zusammenfassung	VIII
IV	Summary.....	XIII
1	Introduction	1
1.1	Bacteria-host interactions.....	1
1.2	<i>Photorhabdus</i> and <i>Xenorhabdus</i>	2
1.3	<i>Pseudomonas</i>	4
1.4	Natural products.....	5
1.4.1	Natural products as drugs	5
1.4.2	Natural products from <i>Photorhabdus</i> und <i>Xenorhabdus</i>	6
1.4.3	Natural products from <i>Pseudomonas</i>	8
1.4.4	NRPS	10
1.4.5	PKS.....	13
1.4.6	Sfp-type PPTases.....	15
1.5	Analytical approaches and recent advancements.....	17
1.5.1	LC-MS	17
1.5.2	Metabolomics	17
1.5.3	Structural analysis of peptides by LC-MS.....	18
1.6	Aim of this work	20
2	Materials and Methods	22
2.1	Strains	22
2.2	Oligonucleotides	25
2.3	Plasmids	29

2.4	Cultivation	30
2.4.1	Cultivation in 20 L-fermenters	32
2.5	Molecular biology.....	32
2.5.1	Isolation of DNA.....	32
2.5.2	Cloning.....	33
2.5.3	Conjugation.....	34
2.5.4	Promoter exchange.....	35
2.5.5	Deletion.....	35
2.6	Analytical methods.....	35
2.6.1	Sample preparation.....	35
2.6.2	UPLC-MS analysis.....	37
2.6.3	UPLC-HRMS analysis.....	37
2.6.4	Lipidomics analysis.....	37
2.6.5	Network analysis.....	38
2.6.6	Feeding experiments.....	39
2.6.7	Solid phase peptide synthesis.....	39
2.6.8	Bioactivity testing of massetolide E and viscosin.....	39
3	Results.....	41
3.1	Inthraszentin.....	41
3.1.1	Linking Natural Product Families to Biosynthetic gene clusters.....	41
3.1.2	Identification of inthraszentin A from <i>X. szentirmaii</i>	43
3.1.3	Structure elucidation <i>via</i> MS.....	49
3.2	Protective sphinganine from <i>Pseudomonas</i> MYb115.....	59
3.2.1	Bioactivity of MYb115.....	59
3.2.2	Gene cluster identification.....	59
3.2.3	Structure elucidation <i>via</i> MS.....	62

3.2.4	Lipidomics	66
3.2.5	Deletion mutants	68
3.3	Antimicrobial peptides from <i>Pseudomonas</i> MYb11 and MYb12	71
3.3.1	Identification of specific natural products from MYb11 and MYb12.....	71
3.3.2	Structure elucidation and confirmation <i>via</i> MS.....	72
3.3.3	Bioactivity guided fractionation	79
3.4	GameXPeptides from <i>Xenorhabdus miraniensis</i>	82
3.4.1	Identification of GameXPeptides	82
3.4.2	Structure elucidation <i>via</i> MS	84
4	Discussion.....	92
4.1	Inthraszentin.....	92
4.2	Protective Sphinganine from <i>Pseudomonas</i> MYb115	99
4.3	Antimicrobial peptides from <i>Pseudomonas</i> MYb11 and MYb12	106
4.4	GameXPeptides from <i>Xenorhabdus miraniensis</i>	108
5	References	111
6	Attachments	128
7	Curriculum vitae	146
8	Publications	147
9	Record of conferences	147
10	Erklärung	148
11	Versicherung.....	148
12	Danksagung	149

II. TABLE OF ABBREVIATIONS

aa	amino acid
ABC transporter	ATP-binding cassette transporter
A domain	adenylation domain
ACN	acetonitrile
ACP	acyl carrier protein
AMP	Adenosine monophosphate
AmT	aminotransferase
AT	acyl transferase
BGC	biosynthetic gene cluster
bp	base pairs
C domain	condensation domain
Cdual	dual condensation/epimerisation domain
CE	dual condensation/epimerisation domain
CID	collision-induced dissociation
CLF	chain length factor
CLP	cyclic lipopeptide
CmodAA	C domains modifying amino acids
Cstart	starter C domain
CoA	coenzyme A
Cy	Cyclase domain
DH	dehydratase domain

E domain	epimerisation domain
easyPACId	easy Promoter Activated Compound Identification
EIC	extracted ion chromatogram
ER	enoylreductase domain
ESI	electrospray ionisation
F domain	formylation domain
FA	fatty acid
FAS	fatty acid synthase
GNPS	Global Natural Product Social molecular networking
GXP	GameXPeptide
Hal	halogenase domain
HCD	high energy collision dissociation
HPLC	high performance liquid chromatography
HRMS	high-resolution mass spectrometry
IJ	infective juvenile
KR	ketoreductase domain
KS	ketosynthase domain
LB	Lysogeny broth
LC	liquid chromatography
LC-MS	liquid chromatography-mass spectrometry
Lig	Acyl-CoA ligase domain
MeOH	methanol

MRMS	magnetic resonance mass spectrometry
mRNA	messenger RNA
MRSA	methicillin-resistant <i>Staphylococcus aureus</i>
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MT	methyltransferase domain
<i>m/z</i>	mass to charge ratio
NADPH	Nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
NMR	nuclear magnetic resonance
NP	natural product
NPF	natural product family
NRP	non-ribosomal peptide
NRPS	non-ribosomal peptide synthetase
OD600	optical density at 600 nm wavelength
Ox	oxygenase domain
PCP	peptidyl carrier protein
PCR	polymerase chain reaction
PKS	polyketide synthase
Ppant	4'-phosphopantetheine
Ppm	parts per million
PPTase	phosphopantetheinyl transferase

PUFA	polyunsaturated fatty acid
QqQ	triple quadrupole mass spectrometer
qToF	quadrupole/time-of-flight mass spectrometer
rdbe	Rings and double bond equivalent
RiPPs	Ribosomally synthesized and post-translationally modified peptides
R domain	reductase domain
rpm	revolutions per minute
SAT	starter acyl transferase domain
SDR	short-chain dehydrogenase/reductase
sRNA	small bacterial RNA
T domain	thiolation domain
TE	thioesterase domain
UPLC	ultra performance liquid chromatography
WT	wildtype
XINF	Xenoinformycin

III. ZUSAMMENFASSUNG

Diese Arbeit behandelt die Identifizierung neuer Naturstoffe aus nematodenassoziierten Bakterien.

Eine Vielzahl von Bakterien lebt in enger Assoziation mit eukaryotischen Wirtsorganismen. Diese Bakterien können sowohl in als auch auf dem Wirtsorganismus leben. Die Gesamtheit der Mikroorganismen eines Wirts bezeichnet man als Mikrobiom dieses Organismus. Die Interaktionen zwischen den Mikrobiota und dem Wirt können unterschiedlichster Natur sein, mutualistisch, kommensal, pathogen oder parasitär. Um die Interaktionen mit dem Wirt zu beeinflussen produzieren viele Bakterien Naturstoffe. Nematoden bieten ein enormes Potential für die Untersuchung von bakteriellen Naturstoffe, die diesen Interaktionen zugrunde liegen, da Nematoden ungefähr 80 % der Tiere auf der Erde ausmachen.

Zwei wichtige Klassen von Naturstoffen sind Nicht-ribosomale Peptide und Polyketide. Beide werden durch sogenannte Megasyntasen produziert. Megasyntasen sind große, modulare Multienzymsysteme.

Nicht-ribosomale Peptide werden von Nicht-ribosomalen Peptidsynthetasen (NRPS) hergestellt. Wie Ribosomen verknüpfen auch NRPS einzelne Aminosäuren zu Peptiden. Allerdings wird die Aminosäuresequenz von nicht-ribosomalen Peptiden nicht durch Nukleinsäuren sondern durch die Spezifität der Adenylierungsdomänen (A-Domänen) der NRPS bestimmt. A-Domänen aktivieren ihrer Spezifität entsprechend Aminosäuren. Neben A-Domänen besteht ein typisches NRPS-Modul aus einer Thiolierungsdomäne (T-Domäne) und einer Kondensationsdomäne (C-Domäne). Nach der Aktivierung lädt die A-Domäne die entsprechende Aminosäure auf den Phosphopantetheinylarm der T-Domäne. Die C-Domäne katalysiert die Kondensationsreaktion von zwei Aminosäuren. In der Regel wird das fertige Peptid am Terminationsmodul durch eine Thioesterasedomäne (TE-Domäne) freigesetzt. NRPS können strukturell vielfältige Peptide produzieren. So werden z.B. neben proteinogenen Aminosäuren ebenfalls nicht-proteinogene und D-Aminosäuren eingebaut. Des Weiteren können die Peptide methyliert oder acetyliert werden.

Polyketide werden durch PKS synthetisiert, welche ebenfalls modular aufgebaut sind. Ein minimales PKS-Modul besteht aus einer Acyltransferasedomäne (AT-Domäne), einer T-Domäne und einer Ketosynthasedomäne (KS-Domäne). Die AT-Domäne überträgt

Acyleinheiten auf die T-Domäne. Die KS-Domäne verknüpft die einzelnen Acyleinheiten miteinander. Während verschiedene PKS eine Vielzahl unterschiedlicher Starteracyleinheiten verwenden, werden als Verlängerungseinheiten typischerweise Malonyl-CoA oder Methylmalonyl-CoA verwendet. Auch PKS produzieren strukturell sehr diverse Naturstoffe. Die Diversität wird z.B. durch partielle oder vollständige Reduzierung der Ketogruppen, sowie Zyklisierung hervorgerufen.

Häufig sind die Gene, welche für NRPS oder PKS codieren in sogenannten Biosynthesegenclustern (BGCs) organisiert. Die Struktur und Sequenz dieser BGCs erlauben erste Prognosen über die Struktur der produzierten Naturstoffe.

Im Rahmen dieser Arbeit wurden Naturstoffe aus *Photorhabdus*- und *Xenorhabdus*-Bakterien, sowie aus Pseudomonaden untersucht. Bakterien der Gattungen *Photorhabdus* und *Xenorhabdus* sind entomopathogene Bakterien, die in Symbiose mit Nematoden der Gattungen *Heterorhabditis* bzw. *Steinernema* leben. Diese Bakterien produzieren eine Vielzahl an bioaktiven Naturstoffen, die unter anderem das Immunsystem des Insekts supprimieren oder den Insektenkadaver vor Nahrungskonkurrenten schützen. Die hierin untersuchten Bakterienstämme der Gattung *Pseudomonas* sind natürliche Isolate aus *Caenorhabditis elegans*. Alle drei untersuchten Stämme sind in der Lage ihren Wirtsnematoden durch Produktion von Naturstoffen vor bakteriellen Infektionen zu schützen.

Zur Untersuchung der Naturstoffe aus diesen Organismen wurde die Flüssigkeitschromatographie-gekoppelte Massenspektrometrie (LC-MS) eingesetzt. LC-MS eignet sich hervorragend als Technik zur Untersuchung sowohl einzelner Naturstoffe als auch vollständiger Metabolome. Da die Analyte durch die Flüssigchromatographie voneinander separiert werden, verlangt diese Methode nur in seltenen Fällen Vorfraktionierungen. Dies ermöglicht eine zeit- und kosteneffiziente Analyse. Die gewonnen massenspektrometrischen Daten erlauben Rückschlüsse auf die Struktur der Analyte. Eine moderne chemoinformatische Methode namens *GNPS network analysis* ermöglicht darüber hinaus die automatisierte Analyse und Sortierung einzelner Massen anhand ihrer Fragmentierung.

Das erste Projekt im Rahmen dieser Arbeit befasst sich mit der Identifizierung und Strukturaufklärung der Inthrasentine mittels Flüssigchromatographie-gekoppelter Massenspektrometrie (LC-MS).

Zunächst wurde eine *GNPS network analysis*-basierte Methode entwickelt, um Naturstoffe zu detektieren und identifizieren, welche mit herkömmlichen massenspektrometrischen Methoden bisher unentdeckt geblieben sind. Die Anwendung dieser Methode führte zur Identifizierung von Inthrasentin A und dem zugrundeliegenden BGC in *X. szentirmaii*. Durch AntiSMASH-Analysen konnten weitere Stämme gefunden werden, die das Inthrasentin-BGC besitzen, *X. indica*, *Photorhabdus thracensis* und *X. stockiae*. Die gezielte Aktivierung dieser BGCs durch das Einsetzen eines induzierbaren Promotors führte zur Identifizierung von Inthrasentin B-D. Durch Markierung der Inthrasentine mit schweren Kohlenstoff- und Stickstoffisotopen und anschließende LC-MS-Analyse konnten die Summenformeln aufgeklärt werden. Da sich offensichtliche strukturelle Unterschiede zwischen Inthrasentin D und Inthrasentin A-C herauskristallisierten, wurde Inthrasentin D im Umfang dieser Arbeit nicht weiter untersucht. Durch massenspektrometrische Experimente und unter Einbeziehung bioinformatischer Vorhersagen, konnten Strukturvorschläge für Inthrasentin A-C entwickelt werden. Die Struktur von Inthrasentin B konnte durch die Analyse von MS/MS-Spektren bestätigt werden. Des Weiteren konnte gezeigt werden, dass das in früheren Arbeiten beschriebene Xenolindicin nur durch versehentliche Verkürzung des Inthrasentin-BGCs entstanden ist. Außerdem weisen Inthrasentine strukturelle Ähnlichkeiten zu den bekannten Antibiotika Daptomycin und Teixobactin auf.

Das zweite Teilprojekt befasst sich mit der Untersuchung von Naturstoffen aus *Pseudomonas* MYb115. In Zusammenarbeit mit Dr. Kohar A. Kissoyan, Lena Peters und Dr. Katja Dierking von der Christian-Albrechts-Universität zu Kiel konnte gezeigt werden, dass MYb115 seinen Wirtsorganismus *C. elegans* vor bakteriellen Infektionen durch das nematopathogene Bakterium *Bacillus thuringiensis* schützen kann, ohne *B. thuringiensis* zu töten.

Durch AntiSMASH-Analysen des Genoms von MYb115 wurden potentielle Naturstoff-BGCs identifiziert, die für den beobachteten Effekt verantwortlich sein könnten. Die Aktivierung eines BGCs *sgaAB*, welches für eine iterative Typ 1 PKS codiert, durch Einsetzen eines induzierbaren Promotors führte zur Produktion von drei Naturstoffen. Durch Analyse von MS/MS-Spektren und durch Markierung mit schweren Isotopen konnte gezeigt werden, dass es sich bei diesen Naturstoffen um langkettige Sphinganine handelt. Ebenfalls konnte durch Fütterungsexperimente mit markiertem Serin gezeigt werden, dass dieses direkt eingebaut wird. Hierdurch konnten erste Einblicke in die Biosynthese der langkettigen Sphinganine gewonnen werden. In Survival-Assays wurde gezeigt, dass der schützende Effekt von MYb115 mit der Produktion der Sphinganine korreliert, aber von der Menge der produzierten Sphinganine unabhängig zu sein scheint. Sowohl die vollständige, wie auch die teilweise Deletion von *sgaAB* führte zum Verlust des protektiven Effekts.

In einem Lipidomics-Experiment in Kooperation mit Dr. Georgia Angelidou und Dr. Nicole Paczia wurde nachgewiesen, dass neben den Sphinganinen auch neuartige Phosphoglycerol-Sphingolipide produziert werden. Ob die Sphinganine oder die Sphingolipide für den schützenden Effekt verantwortlich sind, konnte im Rahmen dieser Arbeit nicht abschließend geklärt werden. Ein Biosynthesemechanismus für die langkettigen Sphinganine wurde postuliert. Dabei wird eine Palmitoyl-CoA-Startereinheit in drei bis fünf Zyklen durch Malonyl-CoA verlängert. Die β -Ketogruppen der Verlängerungseinheiten werden jeweils vollständig reduziert. Die genomische Umgebung von *sgaAB* in MYb115 und drei anderen Stämmen, die in enger Assoziation mit eukaryotischen Wirtsorganismen leben, legt nahe, dass dieses BGC von Bedeutung für die Synthese der bakteriellen Kapsel ist.

Im dritten Teil dieser Arbeit wird die Identifizierung zweier Lipodepsipeptide der Viscosin-Familie beschrieben, Massetolide E und Viscosin. Diese beiden Naturstoffe werden von den *Pseudomonas*-Stämmen MYb11 und MYb12 produziert. MYb11 und MYb12 zeigten antibiotische Aktivität gegen *B. thuringiensis* *in vivo* und *in vitro*. AntiSMASH-Analysen der Genome beider Stämme bestätigten, dass Myb11 und MYb12 das zweigeteilte Viscosin-BGC in ihrem Genom tragen. Durch MS/MS-Fragmentanalysen und Fütterungsexperimente mit isotonenmarkierten Medien und Aminosäuren wurde die Struktur von Massetolide E und Viscosin bestätigt. Anschließend konnte durch Bioaktivitätstests mit fraktionierten

Extrakten von MYb11 und MYb12 gezeigt werden, dass Massetolide E und nicht Viscosin für die antibiotische Aktivität gegen *B. thuringiensis* verantwortlich ist.

Im vierten und letzten Teilprojekt dieser Arbeit wurden die Unterschiede zwischen zwei sehr ähnlichen NRPSs und deren Produkten identifiziert. Zwischen der bereits charakterisierten GameXPeptid-Synthetase (GxpS) aus *Photorhabdus luminescens* TT01 und der GxpS aus *X. miraniensis*.

Während die GxpS aus TT01 lineare und zyklische Pentapeptide aus Valin, Leucin und Phenylalanin produziert, war das Produktspektrum der GxpS aus *X. miraniensis* gänzlich unbekannt. Frühere bio- und chemoinformatische Analysen deuteten allerdings darauf hin, dass diese NRPS dazu in der Lage ist Xenoinformycine zu produzieren.

Im Rahmen dieser Arbeit wurde zunächst das Produktspektrum der GxpS aus *X. miraniensis* untersucht. Dabei wurden mehrere lineare Tetrapeptide sowie lineare und zyklische Pentapeptide detektiert. Mithilfe von Markierungsexperimenten mit isotoopenmarkierten Medien konnten die Summenformeln aller detektierten Peptide ermittelt werden. Durch inverse Fütterung von Aminosäuren in Kulturen in isotoopenmarkierten Medien und MS/MS-Sequenzanalyse wurde die Aminosäuresequenz der Peptide ermittelt. Um die Strukturen der von *X. miraniensis* produzierten GameXPeptide in Bezug auf die Konfiguration der einzelnen Aminosäuren abschließend zu bestätigen, wurden Koinjektionsexperimente mit synthetischen Peptiden mit definierter Konfiguration und Sequenz durchgeführt.

Es konnte gezeigt werden, dass auch die GxpS aus *X. miraniensis* unpolare Peptide bestehend aus Valin, Leucin und Phenylalanin produziert, jedoch mit unterschiedlicher Sequenz und Konfiguration im Vergleich zu den bekannten GameXPeptiden aus TT01. Es konnte außerdem gezeigt werden, dass die GxpS aus *X. miraniensis* unter den getesteten Bedingungen nicht in der Lage ist Xenoinformycine zu produzieren.

IV. SUMMARY

This work addresses the identification of unknown natural products from nematode-associated bacteria.

Many bacteria and other microorganisms live in close association with eukaryotic host organisms. The entirety of all microorganisms living inside and on a certain organism are characterised as the microbiome of said organism. Bacteria living in close association with a host often produce natural products to facilitate bacteria-host interactions. Those interactions can be of mutualistic, parasitic, pathogenic or commensal nature. Because Nematodes account for roughly 80 % of all individual animals on earth, this taxon offers tremendous potential to study bacteria-host interactions and the natural products involved therein.

Two important classes of natural products that are often involved in bacteria-host interactions are polyketides and non-ribosomal peptides. Non-ribosomal peptides are produced by polyketide synthases (PKS), while non-ribosomal peptides are produced by non-ribosomal peptide synthetases (NRPS). Both, NRPS and PKS, are large modular enzymatic assembly lines in which each module incorporates one building block.

This work addresses NRPS- and PKS-derived natural products from *Photorhabdus* and *Xenorhabdus* species as well as from *Pseudomonas* species, all of which live in close association with nematodes. *Photorhabdus* and *Xenorhabdus* are entomopathogenic bacteria that live in close association with *Heterorhabditis* and *Steinernema* nematodes, respectively. *Photorhabdus* and *Xenorhabdus* bacteria are potent producers of bioactive natural products that influence their host's development, suppress the immune system of their insect prey and protect the insect carcass from food competitors among other functions. The genus *Pseudomonas* contains a number of well-studied model organisms as well as human pathogens. However, this work deals with Pseudomonads that are natural microbiota of the model organism *Caenorhabditis elegans* and were isolated only recently. All three of those *Pseudomonas* strains displayed the ability to protect their host organism from negative repercussions of infections with nematopathogenic bacteria.

Inthraszentins are the first class of natural products identified in this work. Inthraszentins are cyclic lipodepsitradecapeptides. They were identified in an approach that aimed at the identification of natural products that remained obscured in classical LC-MS analysis and where previous attempts to activate silent biosynthetic gene clusters (BGCs) failed. GNPS network analysis of *Xenorhabdus szentirmaii* WT and the silenced promoter exchange strain *P_{BAD_xsze02985}*, revealed a subnetwork representing a hitherto undescribed natural product, inthraszentin A. By using this approach, the natural product was simultaneously identified together with the respective BGC. AntiSMASH analysis revealed additional inthraszentin BGCs that in the genomes of *X. indica*, *P. thracensis* and *X. stockiae*. All four BGCs were activated through promoter exchange although previous attempts in *X. szentirmaii* failed. The structures of inthraszentins A-D were characterised using a combination of bioinformatic and mass spectrometric analysis. Structure proposals for inthraszentins A-C were formulated and b- and y-fragments were assigned to the MS/MS spectra of inthraszentin B. In addition, this work revealed that the *isz* BGC was truncated in a previous approach aiming at the activation of said BGC. The truncation led to the identification of xenolindicins A-C, truncated versions of inthraszentin C. Structural similarities between inthraszentins and known antibiotics hint at a potential bioactivity of those.

The second class of natural products identified in this work, are long chain sphingamines produced by *Pseudomonas* MYb11. The promoter exchange based activation of an iterative type 1 PKS gene cluster led to the production of three different long chain sphingamines. In a collaboration with Lena Peters and Dr. Katja Dierking, we could show that production of these natural products is essential for the protective effect MYb11 has on its host, *C. elegans*. Additionally, we identified a new class of related phosphoglycerol-sphingolipids in a lipidomics experiment together with Dr. Georgia Angelidou and Dr. Nicole Paczia. While it is possible that either the sphingamines or the related sphingolipids trigger the protection, the underlying mechanism remains to be investigated. The proposed biosynthetic mechanism involves the elongation of palmitoyl-CoA with multiple extender units followed by thioesteration and release catalysed by the terminal aminotransferase domain.

The third class of natural products identified in this work are the cyclic lipodepsipeptides viscosin and massetolide E from *Pseudomonas* MYb11 and MYb12. The structures of massetolide E were characterised using feeding experiments and LC-MS analysis.

Bioactivity testing and LC.MS analysis of fractions from MYb11 and MYb12 extracts, revealed that massetolide E inhibits the growth of pathogenic *B. thuringiensis* strains that infect *C. elegans*.

The fourth and last class of natural products characterised in this work are GameXPeptides (GXPs) produced by *X. miraniensis*. Bioinformatic analysis of the responsible GameXPeptide synthetase (GxpS) revealed some differences in comparison to the well-known GameXPeptide synthetase from *Photorhabdus luminescens* TT01. Promoter exchange based activation of the GXP BGC in *X. miraniensis*, which was carried out by my colleague Dr. Zhengyi Qian, led to the production of nine different tetra- and pentapeptides. The peptide sequence of all peptides were analysed using tandem mass spectrometry. Coinjection experiments with synthetic peptides, synthesised by Trinetri Goel, revealed the structure of three pentapeptides and allowed to assess the differences between the GameXPeptide synthetases from *X. miraniensis* and TT01. However, the GXP BGC is not responsible for the production of xenoinformycins, which was suggested in previous studies.

1 INTRODUCTION

1.1 Bacteria-host interactions

For billions of years, bacteria have populated almost every type of surface on earth: From hot to cold, acidic to basic, aerobic to anaerobic. Unknowingly, a huge number of living organisms are and have been hosts to bacteria inhabiting their roots, leaves, hyphae, skin, hair, teeth and gut¹⁻⁴. The nature of these relationships ranges from mutualistic to commensal to parasitic with many of the interactions between bacteria and their respective hosts being essential for survival for either of them^{5,6}.

In recent years, research caught up to the huge amount of effects bacteria can have on their respective host organisms, most notably and prominently of course on humans⁷. Bacteria involved in interactions with humans often produce natural products to facilitate those interactions and affect the host.

As an example, siderophores, iron-chelating natural products, are essential virulence factors in bacteria⁸⁻¹⁰. By producing siderophores thereby being able to assimilate the available iron, bacteria gain a competitive advantage over their host¹¹.

Although many bacteria are known to harm or even kill their host, there is a multitude of positive effects microbiota can have on their respective host, e.g. by inhibiting pathogen colonisation¹² or by facilitating maturation of the host immune system¹³, to name a few.

The human commensal *Staphylococcus lugdunensis*, for instance, produces lugdunin, a non-ribosomally synthesized thiazolidine-containing peptide¹⁴ (Fig. 1). Lugdunin is bioactive against major pathogens, most notably *Staphylococcus aureus*. *S. aureus* is an opportunistic pathogen found in the upper respiratory tract of around one third of the human population¹⁵. *S. aureus* is responsible for a plethora of human bacterial infections worldwide and methicillin-resistant *S. aureus* (MRSA) represents a major threat for human health^{16,17}. Humans carrying lugdunin-producing *S. lugdunensis* in their respiratory tract are less likely to carry *S. aureus*, therefore being less susceptible to infections¹⁴.

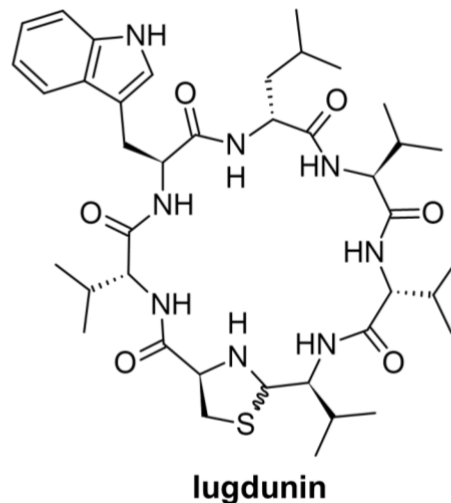


Figure 1. Structure of Lugdunin.

The above example found in humans points at nature's arsenal of natural products and suggests that many more, with either beneficial or harmful effects on the host, lie hidden in unexplored bacteria-host interactions¹⁸.

Nematodes are one taxon that offers tremendous prospects for potential interactions with bacteria. Interestingly, the phylum Nematoda accounts for 80 % of all individual animals on earth¹⁹ and is estimated to contain up to 100 million species²⁰. Presumably, there are bacteria associated with every animal on earth²¹, hence there are at least 100 million bacteria-nematode interactions and their underlying mechanisms to study.

In this work, I identified and investigated natural products produced by nematode-associated bacteria of the genera *Photorhabdus*, *Xenorhabdus* and *Pseudomonas*.

1.2 *Photorhabdus* and *Xenorhabdus*

Photorhabdus and *Xenorhabdus* are species of Gram-negative, rod-shaped, facultatively anaerobic bacteria. Phylogenetically, they belong to the family Morganellaceae within the class γ -Proteobacteria in the phylum Pseudomonadota²². *Photorhabdus* species live in association with *Heterorhabditis* nematodes while *Xenorhabdus* bacteria are associated with nematodes of the genus *Steinernema*²³⁻²⁵. Those bacteria are closely associated with their

nematode host and have not been isolated so far as free-living organisms²⁶. *Photorhabdus* and *Xenorhabdus* bacteria display a highly similar and complex life cycle^{27,28} (Fig. 2).

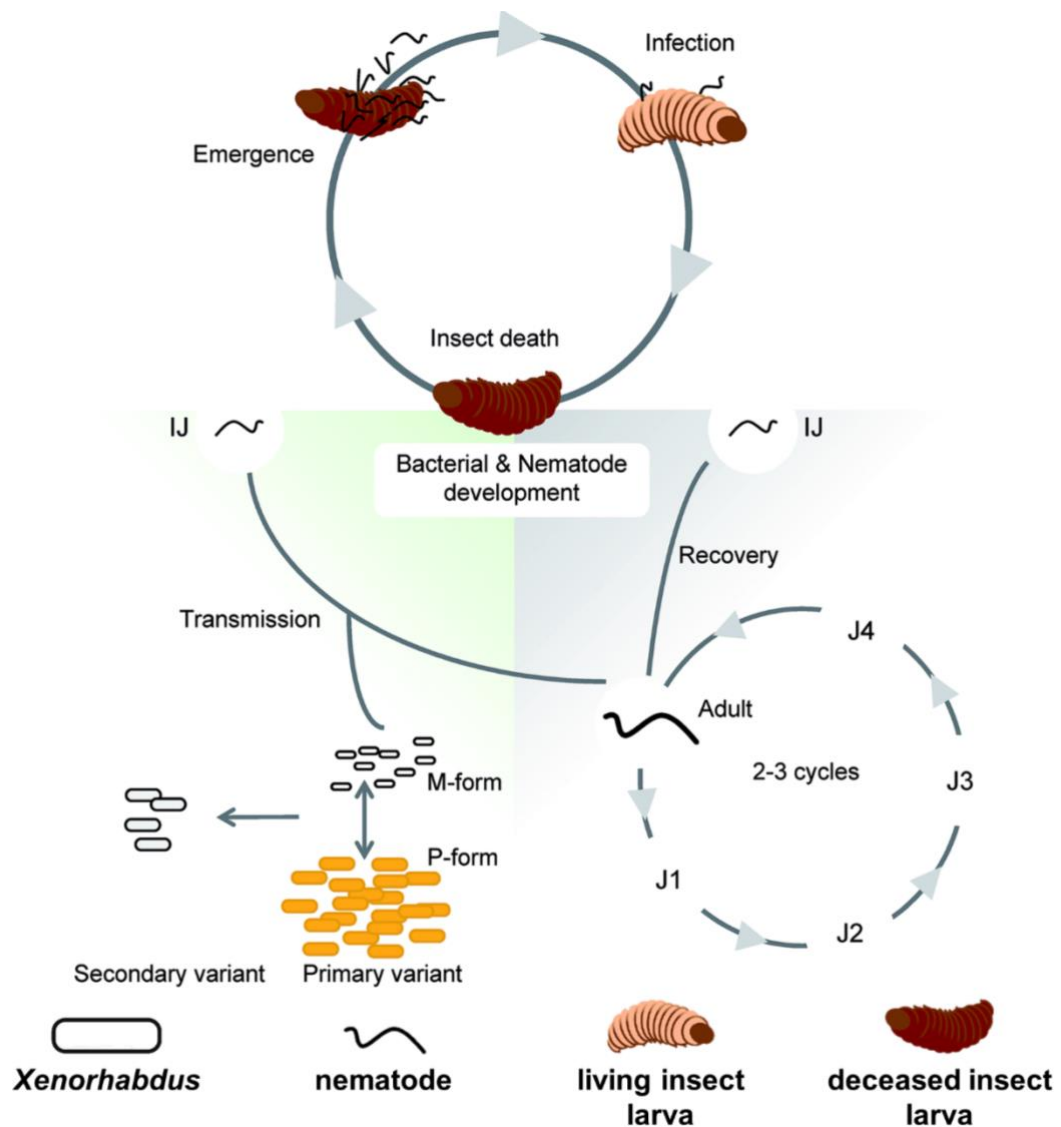


Figure 2. Life cycle of *Photorhabdus* and *Steinernema* summarised by Shi and Bode²⁹.

During their non-feeding infective juvenile (IJ) stage³⁰, the soil living nematodes roam their habitat actively searching for insect larvae³¹. During their search for prey, they can be attracted by chemical signalling compounds, which are released by plant roots damaged by insect larvae³². When the IJ nematodes enter the insect larvae through natural openings in the cuticle, they release the bacteria into the haemolymph³³. Upon entering the haemolymph, the bacteria produce protein toxins and a myriad of natural products to suppress the insect's

immune system and to kill the prey^{34,35}. Subsequently, the bacteria and nematodes feed of the carcass³⁶. After two to three reproduction cycles, when nutrients in the carcass are becoming scarce, a new generation of infective juveniles equipped with their symbiotic bacteria starts its search for prey^{36,37}. While both, male and female, *Steinernema* nematodes are required for the development of a new generation, female, hermaphroditic *Heterorhabditis* individuals are sufficient for reproduction³⁷.

Photorhabdus and *Xenorhabdus* display two distinct phenotypical variants^{38,39}. The primary variant, which can be isolated from the nematode host, can switch to the secondary form under laboratory conditions⁴⁰. The secondary variant produces less pigments and antibiotics, is less bioluminescent, is deficient in protease as well as lipase activity and differs in colony morphology²⁷. It is speculated that the reason behind phenotypical variation of *Photorhabdus luminescens* is a change of lifestyle. Since the secondary variant of *P. luminescens* is able to sense plant signals, Eckstein *et al.* suggest that it lives within the rhizosphere of plants feeding on plant-derived compounds⁴¹. While the secondary variant of *P. luminescens* is not able to revert to its primary form⁴¹, *X. nematophila* is switching between a mutualistic form and a form that is virulent to insects⁴².

Unsurprisingly, this life cycle warrants a myriad of reactions acting in concert, many of which are facilitated by natural products produced by the mutualistic bacteria^{29,35}. *Photorhabdus* and *Xenorhabdus* exhibit the potential to produce relatively high numbers of different natural products displaying a wide range of bioactivities^{43,44}, which is detailed in chapter 1.4.2.

1.3 *Pseudomonas*

The genus *Pseudomonas* belongs to the family Pseudomonaceae, inside the phylum Pseudomonadota and the class γ -Proteobacteria. *Pseudomonas sp.* are Gram-negative, rod-shaped, motile bacteria. This genus includes a number of important and well-studied organisms including the opportunistic human pathogen *P. aeruginosa*⁴⁵, which causes lethal infections especially in patients dealing with cystic fibrosis⁴⁶. Other members of this genus, including some the recently discovered isolates from *C. elegans*, support their host's development or protect it from infections^{47,48}. Nevertheless, Pseudomonads are not limited

to animal hosts. In fact, they are found predominantly in aquatic and soil environments, where they follow a free-living or host-associated lifestyle⁴⁹⁻⁵¹. It shares the soil environment with a plethora of different organisms from all domains of life⁵². Among those is the soil living nematode *Caenorhabditis elegans*⁵³. *C. elegans* is one of the most commonly deployed model organisms. However, the native microbiome has only recently been described⁴⁷. It consists mainly of unclassified Enterobacteriaceae as well as members of the genera *Ochrobacterium*, *Stenotrophomonas*, *Sphingomonas* and *Pseudomonas*⁴⁷. Since these natural *Pseudomonas* isolates from *C. elegans* displayed the ability to protect the nematode from infections, I chose to investigate some of these isolates and their natural products in detail.

In addition, *Pseudomonas sp.* are potent natural product producers⁵⁴, which is detailed in chapter 1.4.3.

1.4 Natural products

1.4.1 Natural products as drugs

Natural products (NPs) are canonically defined as small molecules produced by a living organism. While products of the primary metabolism are included in this definition, it is most often used in regard to secondary or specialised metabolites. While NPs have been used by humankind for millennia to treat diseases, wounds or to ease pain⁵⁵, research on natural products started in the 19th century with the isolation and characterisation of compounds like caffeine⁵⁶, morphine⁵⁷ and atropine⁵⁸, to name but a few. The discovery and use of natural products as pharmaceutical drugs was jumpstarted by the characterisation of penicillin by Sir Alexander Fleming⁵⁹. The role of natural products as drugs or as inspiration for drug design increased rapidly. Of all newly approved drugs between January of 1981 and September of 2019, approximately 65 % are natural products, natural product derivatives, mimics of natural products or contain active moieties from natural products⁶⁰. Roughly one sixth of all annotated natural products are of bacterial origin⁶¹, making bacteria an important resource for natural product research. Many of the natural products produced by bacteria are derived from either non-ribosomal peptide synthetases (NRPS), polyketide synthases (PKS) or NRPS-PKS-hybrids, which are described later in this chapter.

1.4.2 Natural products from *Photorhabdus* und *Xenorhabdus*

Bacteria of the genera *Photorhabdus* and *Xenorhabdus* are potent natural product producers²⁹. 6.5 % of the genome of *P. luminescens* contain biosynthetic gene cluster (BGCs) that are responsible for natural product production⁶². The potential for natural production of *Xenorhabdus* and *Photorhabdus* is in line with the potential of *Streptomyces* bacteria, which are the most potent bacterial genus for the discovery of natural products^{63,64}. Recently, there is a pressing urge for the discovery of novel bioactive natural products due to the rise of antibiotic resistances⁶⁵. Because they use natural products to fight off bacterial, fungal, protozoal and nematodal food competitors, *Photorhabdus* and *Xenorhabdus* bacteria make excellent candidates for the quest to find natural products to fight infections⁶⁶⁻⁶⁹. The regulation of natural product biosynthesis in *Xenorhabdus* and *Photorhabdus* is relatively well studied.

For instance, the deletion of the post-transcriptional regulator Hfq, that mediates the binding of sRNAs to their respective target mRNAs, leads to the abolishment of natural product production in *Photorhabdus luminescens*⁷⁰. Subsequent deletion of the repressor HexA in the same strain restored the production of natural products⁷⁰. Furthermore, Neubacher *et al.* showed that the Hfq dependant small regulatory RNA *ArcZ* is essential for natural product production⁷¹. *ArcZ* directly binds to the *HexA* mRNA and represses the production of HexA.

The knowledge of the regulation of natural product production in *Photorhabdus* and *Xenorhabdus* enabled the development of genetic tools for the targeted activation of natural product BGCs that were hitherto inactive under laboratory conditions. In the easyPACId method, the native promoter in front of a target BGC is replaced by an inducible promoter in a *hfq* deletion mutant⁷².

This method reduces the noise created by the plethora of natural products in *Xenorhabdus*, *Photorhabdus* and even *Pseudomonas* and simplifies the analysis, purification and bioactivity testing^{72,73}.

1.4.2.1 Bioactive natural products

In their natural environment, *Xenorhabdus* and *Photorhabdus* produce a multitude of natural products to fight off food competitors coveting the insect carcass²⁹. Among those,

fabclavines, rhabdopeptides and xenocoumacins displayed the ability to kill nematodes⁶⁹ (Fig. 3).

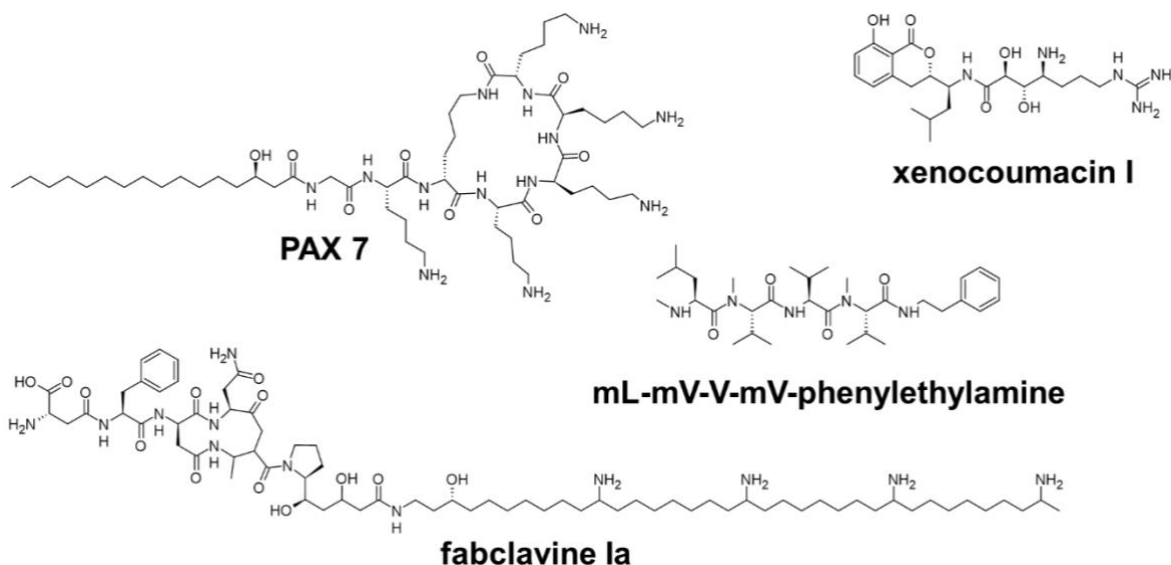


Figure 3. Structures of PAX 7, xenocoumacin I, fabclavine Ia and mL-mV-V-mV-phenylethylamine, a rhabdopeptide.

Rhabdopeptides are a large group of non-ribosomally synthesised peptides (NRPs). They are produced through flexible and iterative use of monomodular NRPS by *Photorhabdus* and *Xenorhabdus* bacteria⁷⁴. Xenocoumacins, however are produced by a NRPS/PKS hybrid⁷⁵. First, precursor molecules called prexenicoumacins are synthesised, that are then cleaved by a membrane bound peptidase to form xenocoumacins⁷⁶. Fabclavines are also produced by a NRPS/PKS hybrid but are also connected to a polyamine moiety that is produced by fatty acid/polyketide synthase resembling polyunsaturated fatty acid (PUFA) producing enzymes⁷⁷.

Furthermore, fabclavines showed additional activity against fungi and oomycetes^{78,79}.

The non-ribosomally produced PAX peptides another class of antifungal compounds produced by *X. nematophila* and other *Xenorhabdus* strains^{43,80,81}.

In addition, fabclavines, PAX peptides, xenocoumacins and xenorhabdins were identified as antiprotozoal⁷³. These natural products displayed the ability to kill human pathogenic protozoa in a microdilution assay⁷³.

Moreover, *Photorhabdus* and *Xenorhabdus* bacteria produce several compounds with antibiotic properties²⁹ (Fig. 4).

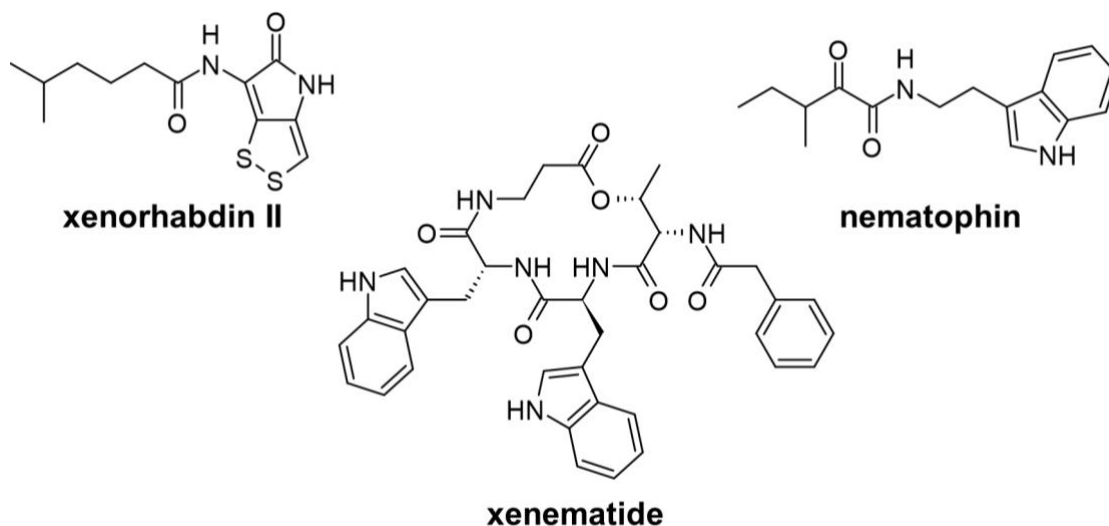


Figure 4. Structures of xenorhabdin II, xenematide and nematophin.

Among the identified natural products that display the ability to kill bacteria are the xenematides, nematophins as well as the aforementioned xenorhabdins and xenocoumacins^{29,82–85}. Xenematide, a non-ribosomally synthesised cyclic lipodepsipeptide, showed antibiotic activity against both Gram-positive and Gram-negative bacteria⁸⁵. The amide nematophin, that is derived from a monomodular NRPS, selectively kills the human pathogen *Staphylococcus aureus*^{86–88}.

The ability by *Xenorhabdus* and *Photorhabdus* bacteria to produce bioactive natural products only affirms the need for more research to uncover the full potential of these genera in regards to natural products. In this work, I aimed to identify hitherto undescribed natural products from *Photorhabdus* and *Xenorhabdus*.

1.4.3 Natural products from *Pseudomonas*

Over three thousand *Pseudomonas* draft genomes have been deposited publicly by 2016⁸⁹. Analysis of *Pseudomonas* genomes revealed that they are relatively large and contain a giant

number of BGCs which encode for natural product producing enzymes⁵⁴. In fact, 5.7 % of the genome of *Pseudomonas aeruginosa* are estimated to be dedicated to natural product biosynthesis, almost rivalling the natural product potential of the aforementioned *Xenorhabdus* and *Photorhabdus*^{62,90}.

For example, *Pseudomonas* spp. are potent producers of NRPs and other amino acid derived natural products that display a wide range of bioactivities^{54,91,92}. Among the most widespread NRPS-derived natural products from Pseudomonads are non-ribosomal lipopeptides (NRLPs)^{54,93}. NRLPs are N-terminally acetylated NRPs. During non-ribosomal peptide synthesis, the acetylation step is located at the beginning of the biosynthesis, where the first amino acid is acetylated by a Starter C domain (C_{start}). The mechanistic details of non-ribosomal peptide synthesis are further explained in chapter 1.4.4. Some NRLPs contain intramolecular ester- or amide-bonds forming heterocyclic compounds⁹³. Cyclic NRLPs are also called cyclic lipopeptides (CLPs), while CLPs that contain an ester are often characterised as cyclic lipodepsipeptides.

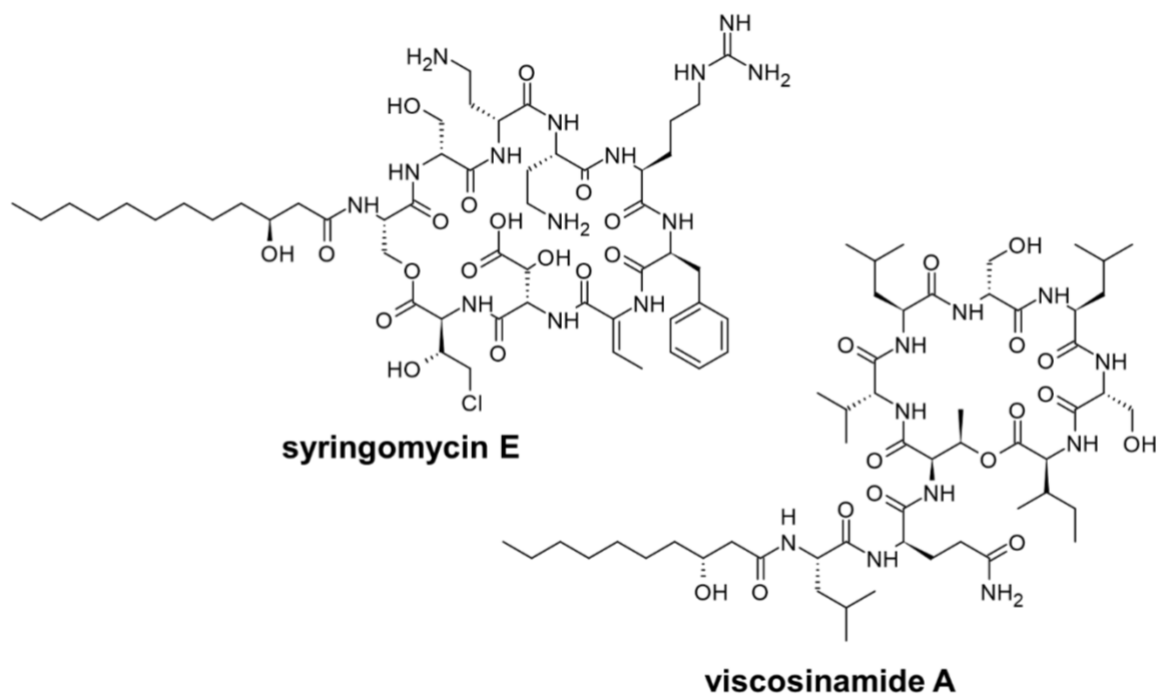


Figure 5. Structures of cyclic lipodepsipeptides syringomycin E and viscosinamide A from *Pseudomonas* spp.

Two prominent cyclic lipodepsipeptides produced by *Pseudomonas* spp. are syringomycin E and viscosinamide A⁹³ (Fig. 5). Syringomycin E displays ion channel-forming ability in lipid bilayers and biomimetic membranes^{94,95}. Syringomycins are produced by the plant pathogen *P. syringae*, which causes diseases in over 50 different host plants including many agricultural crops⁹⁶. Viscosinamides are produced by *P. fluorescens* DR54 and act as biosurfactants and antifungal agents⁹⁷.

However, many other CLPs from *Pseudomonas* spp. also display antifungal, cytotoxic, phytotoxic and bactericidal activity⁹⁸. Like syringomycin E, the biological activity of CLPs is most often attributed to their ability to form pores in biological membranes⁹⁸. Yet, Reder-Christ *et al.* showed that some CLPs are not able to disrupt biological membranes and still display potent bactericidal activity⁹⁹. Although *Pseudomonas* spp. exhibit a huge variety of CLPs, this genus is not limited to this class of natural products¹⁰⁰. Recent genome mining studies indicated that pseudomonads not only offer a great natural product spectrum in terms of numbers, but also great diversity, including arylpolyenes, β -lactams and β -lactones, RiPPs, terpenes, thiopeptides, phenazines, PKS among others¹⁰⁰.

The capability of *Pseudomonas* spp. to synthesise a large number of bioactive natural products along with its genetic accessibility make these bacteria perfect candidates for the identification of novel bioactive natural products^{54,72,101}:

In this work, I investigated the bioactivity of three different *Pseudomonas* isolated from *C. elegans* and linked the observed bioactivities to specific natural products.

1.4.4 NRPS

Peptides were long believed to be synthesised exclusively through ribosomal peptide synthesis as the central dogma of molecular biology postulates¹⁰². During the 1960s several groups could show that the peptide antibiotics gramicidin S and tyrocidine are synthesised in a ribosome independent manner, opening the field of non-ribosomal peptides (NRPs)^{103,104}. NRPs are synthesised by modular megasynthases called non-ribosomal peptide synthetases (NRPS)¹⁰⁵.

Typically, each module of a NRPS assembly line incorporates one building block, which is commonly known as the collinearity rule¹⁰⁶. Hence, a NRPS consisting of five modules

normally synthesizes a NRP containing five amino acids, though stuttering and skipping of single modules occurs as well^{107–109}.

Each module consists of at least three domains: an adenylation domain (A domain), a peptidyl-carrier-protein also referred to as thiolation domain (T domain) and a condensation domain (C domain)¹⁰⁷ (Fig. 6). The first and last module of each NRPS are exceptions to this rule because the first module often contains just an A and a T domain, though starter C domains that incorporate fatty acids are quite common¹⁰⁷. The last module, however, consists of an A and a T domain coupled with either a terminal C domain, a thioesterase domain (TE domain) or a reductase domain (R domain)¹⁰⁷. Terminal C or TE domain catalyse the release of either linear or cyclic peptides¹⁰⁷.

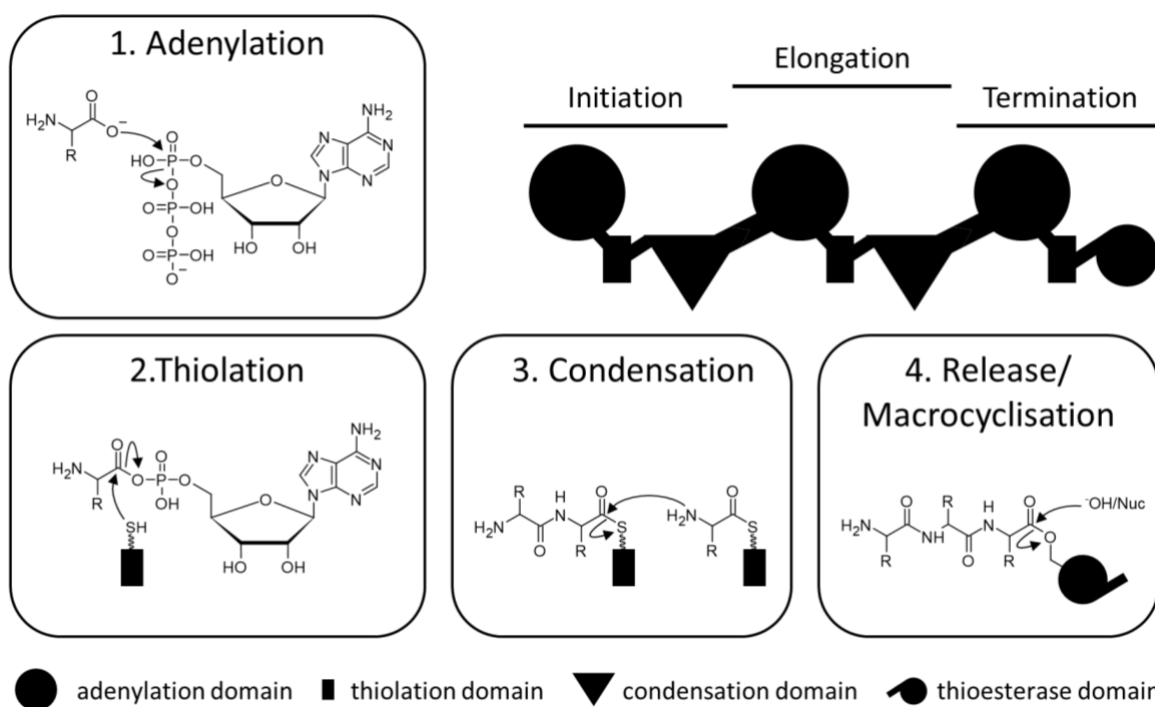


Figure 6. Reactions involved in non-ribosomal peptide biosynthesis and schematic domain organisation of a multimodular NRPS consisting of initiation, elongation and termination modules. Adapted from Süßmuth & Mainz (2017)¹¹⁰.

The first step necessary for NRP synthesis is the ATP dependent activation of amino acid into amino acyl-AMPs¹⁰⁷ (Fig. 6).

Each A domain selectively activates one or a small spectrum of amino acids¹⁰⁷. Even though there have been multiple attempts and models to predict the specificity of a domains through biochemical characterization and analyses of closely related A domains, specificity prediction still needs to be treated with caution^{111,112}.

T domains exist in two distinct forms: an *apo*- and a *holo*-form. The inactive *apo*-form is post-translationally modified to the active *holo*-form, when a phosphopantetheinyl (Ppant) moiety from Coenzyme A is transferred onto a hydroxyl sidechain of a conserved serine residue of the T domain¹¹³. For NRPS T domains, this reaction is catalyzed by an phosphopantetheinyltransferase (PPTase) family called Sfp-type PPTases, named after the PPTase involved in surfactin biosynthesis¹¹³.

After activation of the respective amino acid by the A domain, the amino acid moiety of the amino acyl-AMP is then transferred onto the Ppant moiety of the T domain^{107,110}.

Thereupon, the C domain facilitates the condensation between the intermediates bound at neighbouring T domains. The amino group of the amino acid attached to the downstream T domain acts as a nucleophile in this reaction to form the amide bond with the Ppant bound carbonyl carbon of the upstream T domain¹⁰⁷. Some NRPS contain C domains with an additional function. Condensation/epimerization domains (CE domains) epimerize the Ppant bound amino acid of the upstream T domain¹¹⁴.

Once the end of the assembly line is reached, a terminal thioesterase domain (TE domain) cleaves of the final product¹⁰⁵. A conserved serine residue in the TE domain attacks as a nucleophile the peptidyl-thioester bound to the T domain of the final module to form a covalently bound intermediate¹⁰⁵. This intermediate is either released via hydrolysis to form a linear peptide or via a nucleophilic attack from a nucleophile within the NRP to form a cyclic peptide¹⁰⁵. Alternative peptide release mechanisms include cyclisation by terminal C domains or the release through a terminal reductase domain (R domain), which can lead to a multitude of different structures^{105,115}. Further modifications of peptides are caused by either tailoring enzymes acting *in trans* or additional modifying domains within the assembly line, e.g. formylation (F), epimerization (E), cyclase (Cy), reductase (R), oxygenase (Ox) or methyltransferase (MT) domains¹¹⁰.

Type B and C NRPS introduce even more structural diversity by deviating from the collinearity rule, which only applies to type A NRPS¹¹⁶. Type B NRPS distinguish themselves from type A NRPS by the iterative use of single modules, whereas type C NRPS display a nonlinear domain architecture¹¹⁶.

Overall, NRPS can produce a plethora of natural products, diverse in size, structure and function. In addition, the potential biological activity of natural and engineered NRPs makes them important research subjects. In this work, I studied different type A NRPS from *Photorhabdus*, *Xenorhabdus* and *Pseudomonas* spp.

1.4.5 PKS

Polyketide synthesis is in many aspects analogous to NRP biosynthesis on one hand and to fatty acid (FA) biosynthesis on the other^{107,117}. As for NRPS, there are also three different types of PKS¹¹⁸. For type I PKS as well as for type I fatty acid synthases (FAS), usually found in eukaryotes, the assembly line metaphor applies, as each of them contains different domains with distinct functions in single proteins¹⁰⁷. Moreover, two kinds of type I PKS exist. In fungi, type I PKS typically consist of single modules that are iteratively used to produce polyketides, while bacterial type I PKS are usually multimodular and act in a directional manner^{118,119}. Since exceptions prove the rule, some bacteria deploy monomodular iterative type I PKS for the production of small aromatic compounds and polyenes¹¹⁸.

In type II PKS, as well as in type II FASs, an acyl chain attached to a single T domain is successively elongated by a ketosynthase (KS) domain and an acyltransferase (AT) domain¹⁰⁷. In such systems, the chain length of the final product is determined by a so-called chain length factor (CLF)¹²⁰. Type II FASs are the rule in prokaryotes and type II PKS are almost exclusive to actinomycetes and few other bacteria¹¹⁸. The anthraquinone producing type II PKS from *Photorhabdus luminescens* is one of the few examples from Gram-negative bacteria^{121,122}.

Type III PKS form homodimers in which each active site catalyzes all reactions to form the final product^{123,124}. They are most prominently found in plants but also in bacteria and fungi^{123,125–127}.

Typically, polyketides are synthesized through a repetitive cycle of Claisen condensations of thioesters¹¹⁷ (Fig. 7). This process involves the KS, AT and T domains¹⁰⁷. PKS T domains are also referred to as acyl carrier proteins (ACP). A starter unit, often acetyl-CoA or malonyl-CoA, is elongated through multiple steps until the final product is formed and released¹¹⁷. The AT domain selects an acyl elongation unit and transfers the CoA-bound elongation unit to the Ppant arm of the T domain¹⁰⁷. Type III PKS typically use CoA-bound substrates without transfer to a T domain¹²⁸. The elongation unit is then transferred onto the acyl acceptor at the downstream T domain in a decarboxylative Claisen condensation reaction^{107,117}. The keto group at the β carbon of the acyl chain can be modified by additional PKS domains (Fig. 7). Prototypical modifying domains that are also found in FASs include ketoreductase (KR) domains, dehydratase (DH) domains and enoylreductase (ER) domains¹¹⁷. A full cycle including these three domains results in a fully reduced unsaturated acyl chain which is also typical for type I FASs in animals¹¹⁷.

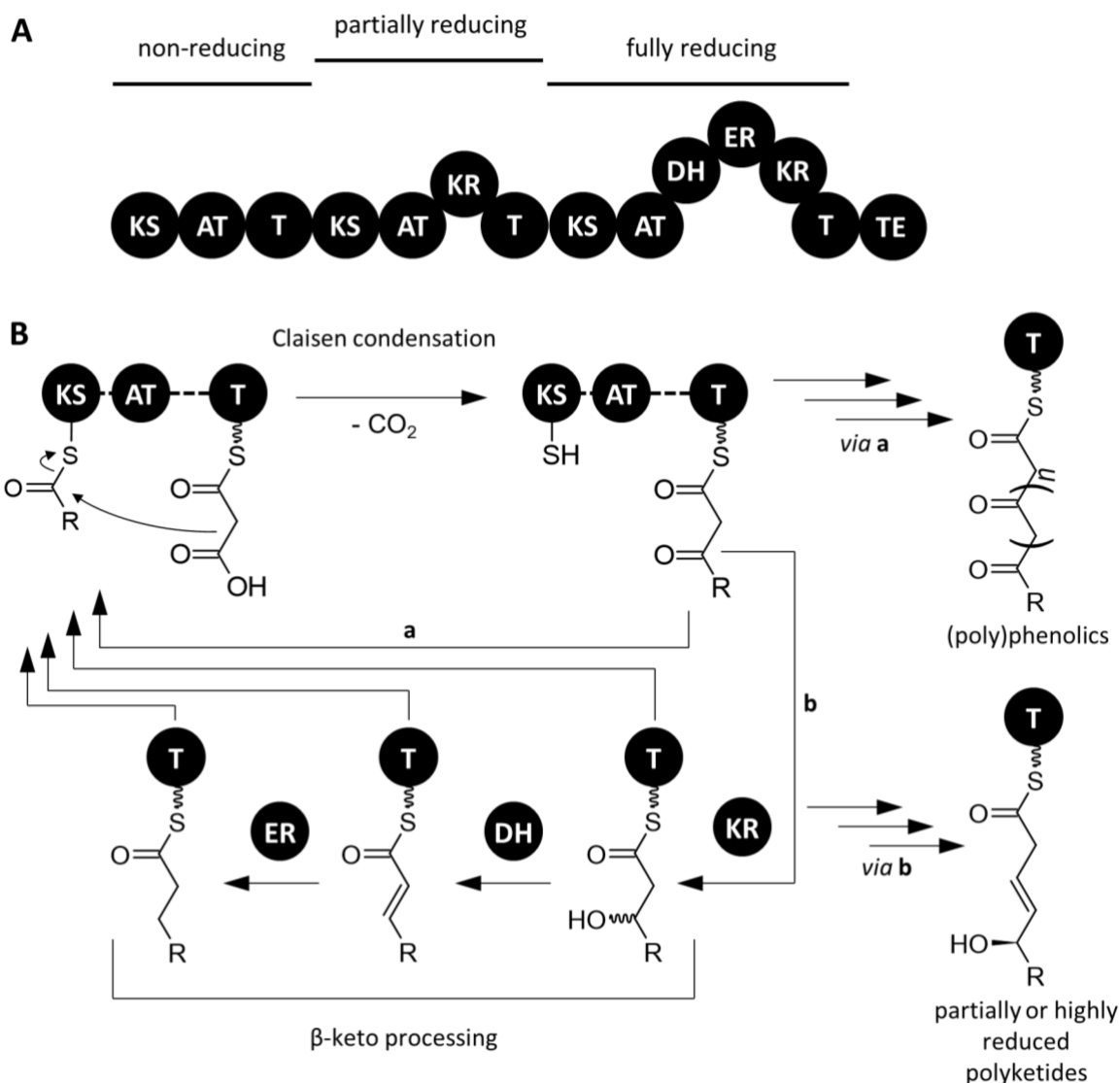


Figure 7. Principles of polyketide biosynthesis. **A:** Ball scheme of a multimodular PKS containing non-reducing, partially reducing and fully reducing modules. **B:** Mechanisms involved in polyketide biosynthesis. Adapted from Hertweck (2009)¹¹⁸.

Additional modifying domains include methyl- (MT) and aminotransferase (AT) domains and halogenase (Hal) domains among others¹⁰⁷. Starter acyl transferase (SAT) domains and Acyl-CoA ligase (Lig) domains are able to provide alternative starter units¹¹⁸.

1.4.6 Sfp-type PPTases

NRPS, PKS and FAS share one important similarity: a PPTase converts the T domain from its inactive *apo*- to its biosynthetically active *holo*-form¹¹³. EntD and Sfp from *Escherichia*

coli and *Bacillus subtilis*, respectively, were among the first characterized PPTases^{129,130}. The *entD* gene is required for the synthesis of the siderophore enterobactin, a known virulence factor of pathogenic *E. coli* strains^{129,131}. Enterobactin is a macrolactone synthesized by a type I/type II NRPS hybrid¹³².

Sfp is essential for surfactin biosynthesis in *B. subtilis*¹³⁰. Surfactin is a non-ribosomally synthesized lipodepsipeptide^{133–135}. Both enzymes are members of the Sfp-type PPTase family^{113,136}. Most notably, enzymes of this family transfer the Ppant moiety from coenzyme A to a conserved serine residue in T domains of NRPS and PKS, converting the *apo*-form to the *holo*-form (Fig. 8), thereby granting the megasynthases their activity¹³⁶.

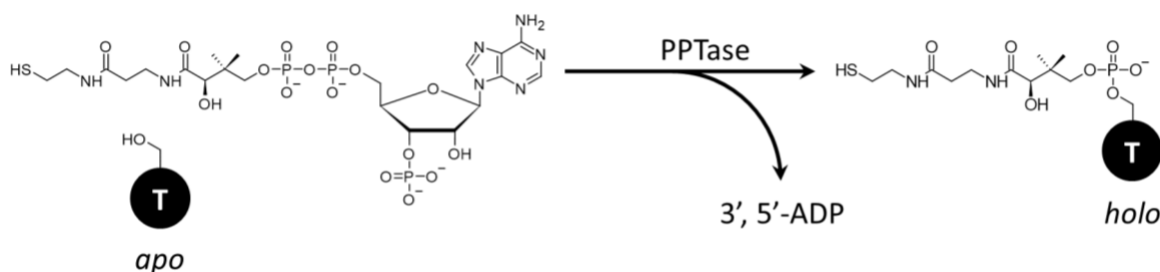


Figure 8. General reaction scheme for posttranslational transfer of the 4'-phosphopantetheinyl group from coenzyme A to a conserved serine residue in NRPS and PKS T domains. Through this reaction the phosphopantetheinyl transferase (PPTase) converts the T domain from its *apo*-form to the *holo*-form.

While AcpS-type PPTases are essential for fatty acid synthesis and therefore for the viability of an organism, Sfp-type PPTases are not¹³⁶. A deletion of the *ngrA* gene, which encodes for a Sfp-type PPTase in *Photorhabdus luminescens*, resulted in the inability of the bacterium to support growth and development of its host nematode and to produce siderophores and antibiotics¹³⁷. Interestingly, three PPTases are encoded in the genome of *P. luminescens*: NgrA, an AcpS-type PPTase necessary for FA biosynthesis and AntB, which is solely dedicated to anthraquinone production^{122,138}. The deletion of *ngrA* in *Photorhabdus luminescens* and related *Xenorhabdus* strains allows for the analysis of the entirety of natural products produced by these strains under defined conditions¹³⁹. In contrast, many *Pseudomonas* species express only one PPTase, which is essential for primary and secondary metabolism^{136,140}.

1.5 Analytical approaches and recent advancements

1.5.1 LC-MS

Liquid chromatography-mass spectrometry (LC-MS) is the workhorse technique to study the chemistry of natural products or the composition of metabolomes¹⁴¹. LC-MS does rarely require sample prepurification, though optional prepurification can be helpful, nor does it require volatility of the analyte outdoing various other analytical techniques concerning time- and cost-efficiency as well as analyte spectrum¹⁴².

The standard setup includes four parts: (i) a liquid chromatograph, where the sample is separated, (ii) an ion source, where the target molecules are ionized and the solvent is evaporated, (iii) a mass analyzer, where the ions are separated by mass to charge ratio (m/z) and (iv) a detector, where the separated ions are detected and their relative abundance is measured¹⁴³. In tandem mass spectrometry, also called MS/MS, two mass analyzers are combined¹⁴³. MS/MS analysis consists of three steps: precursor ion selection followed by collision-induced dissociation (CID), followed by product ion analysis¹⁴³. Tandem mass spectrometers can either be separated by space, like in a triple quadrupole (QqQ) instrument, where one step is performed in each quadrupole, or separated by time, like in an ion trap instrument, where all three steps happen successively in the same device¹⁴⁴. Tandem-in-space instruments can also consist of two different mass analyzers, e.g. in a quadrupole/time-of-flight (qToF) instrument¹⁴⁵. The high selectivity through two separation mechanisms, paired with fast data acquisition speed and high sensitivity make LC-MS a great tool to study metabolomes¹⁴¹.

1.5.2 Metabolomics

The analysis of all metabolites, primary and secondary, in a given biological system was named metabolomics¹⁴⁶. Typically, an experimental metabolomics approach is either of targeted or untargeted nature¹⁴¹. Targeted approaches aim to identify and quantify certain metabolites or groups of metabolites, e.g. quantification of certain disease biomarkers in a patient¹⁴¹. However, untargeted approaches are real global metabolome analysis leading to new hypotheses e.g. finding biomarkers by comparing the metabolomes of healthy and ill individuals^{141,147}.

In August 2016, a landmark paper by Wang *et al.* introduced an easy-to-use, crowd curated, open access platform for analysing and sharing MS data of natural products, called Global Natural Product Social Molecular Networking (GNPS)¹⁴⁸. GNPS enables a network-based analysis of raw MS data, where each precursor ion is represented by a node and each subnetwork of nodes represents a natural product family (NPF), which clusters together based on fragment ion similarities (Fig. 9). A continually growing spectral library, which can be searched for each network analysis, aids identification of known natural products and minimizes rediscovery of *known unknowns*¹⁴⁹.

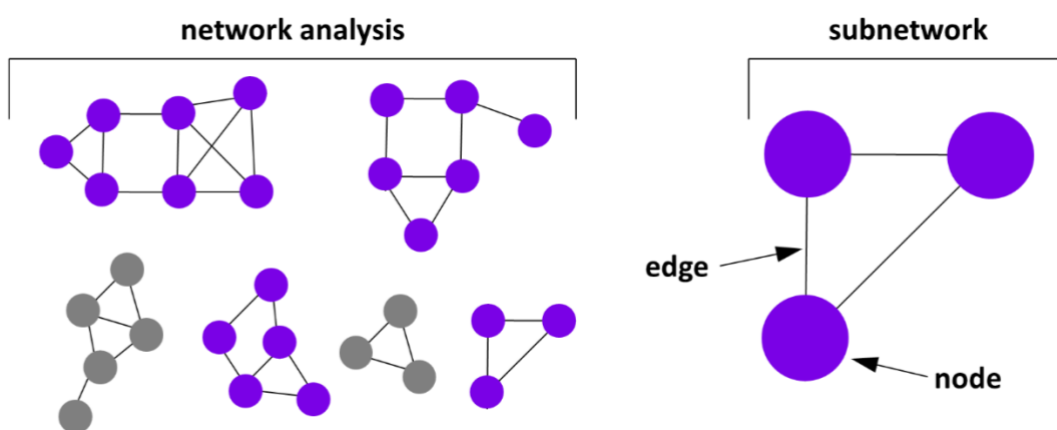


Figure 9. Terminology of a GNPS network analysis and a subnetwork.

Tobias *et al.* utilized this platform in a comparative genomics/metabolomics study of the natural product diversity associated with *Photorhabdus* and *Xenorhabdus* species⁴³. However, they linked NPFs to biosynthetic gene clusters (BGCs) known for the production of certain natural products. Yet, BGCs that are responsible for the production of unknown, potentially bioactive, NPFs remain to be investigated.

1.5.3 Structural analysis of peptides by LC-MS

The elucidation of the chemical structure of peptides and other small molecules by LC-MS includes multiple steps¹⁵⁰. Nowadays, the first step is the exclusion of rediscovery of *known unknowns* by searching spectral libraries. If the target peptide is identified as an *unknown unknown*, the sum formula has to be determined¹⁵⁰. A list of candidate formulas is generated by calculating possible compositions based on high-resolution MS data¹⁵¹. For all

candidates, the number of rings or double bonds in a molecule, termed double bond equivalents (dbe) can already be calculated from their sum formula^{152,153}. MS instruments with extreme high resolution are able to determine the elemental composition of molecules without additional experiments¹⁵⁴. However, these instruments are not available in most laboratories and stable isotope labelling paired with MS analysis offers a cheap and effective alternative¹⁵⁵. In this approach, the organisms producing the target peptide grow in ¹³C- and ¹⁵N-labeled media, respectively. The quantity of carbon and nitrogen atoms in the peptide is then determined by comparison of the MS¹ spectra of the labeled and unlabeled natural product, leading to the correct sum formula from the candidate list.

Peptides and proteins display primarily backbone fragmentation in ion dissociation MS² experiments^{150,156} (Fig. 10). Hence, Analysis of the characteristic peptide fragmentation pattern allows for peptide sequencing^{150,157}.

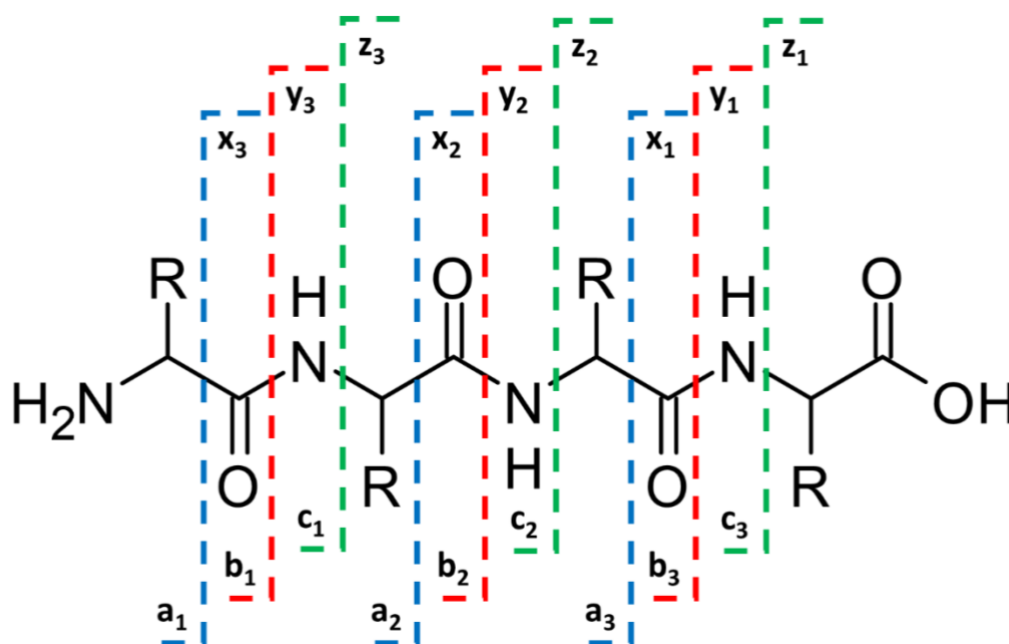


Figure 10. Nomenclature for fragments arising from backbone cleavage of peptides. N-terminal fragments are annotated as a, b, c, while the respective C-terminal fragments are annotated as x, y, z. Adapted from Kool and Niessen¹⁵⁰.

Additionally, this approach can be paired with inverse feeding experiments, in which non-labeled amino acids (or other building blocks) are added to isotopically labeled culture medium¹⁵⁵. Although the absolute configuration can be predicted by analysis of C-, CE- and

E-domains in the BGC of interest (if known), it has to be experimentally determined¹⁵⁵. If a pure or enriched sample of the peptide of interest is obtainable, the advanced Marfey's method remains the best option^{158,159}. Since purification or enrichment are not always easy to achieve, other methods merit consideration. In transaminase knockout strains, amino acids that carry a deuterium label at the α position can be fed to a culture¹⁵⁵. If the amino acid is epimerized, the label is lost, which can be detected *via* MS. Moreover, the peptide of interest can be compared to chemically synthesized candidate peptides using LC-MS/MS¹⁵⁵. If the all main features, retention time, m/z and fragmentation pattern, are identical, the defined structure of the synthesized peptide and the structure of the peptide of interest are identical.

1.6 Aim of this work

This work aims to find unknown natural products produced by bacteria, that live in close association with nematodes and to elucidate their structure by using mass spectrometry.

The first chapter of this work is dedicated to the detection of hitherto unknown natural products by using a metabolomics approach and subsequent structure elucidation of said compounds. This chapter includes metabolomics analysis of *Xenorhabdus szentirmaii* wild type and knockout mutants, overproduction of the target compound, identification of derivatives from other strains and MS based structure elucidation.

The second and third chapters are about natural products that protect *C. elegans* from *B. thuringiensis* infections.

The second chapter deals with natural products that protect the nematode host without killing the pathogen. I deployed molecular biology methods to generate deletion and overproduction strains of a target compound, identified it *via* LC-MS/MS analysis and used LC-MS/MS and lipidomics to analyse the chemical properties of the active compound.

The third chapter aims at finding natural products, which are produced by *Pseudomonas* strains MYb11 and MYb12, respectively. These natural products display the ability to protect *C. elegans* by killing *B. thuringiensis*. I identified said compounds *via* fractionation and subsequent bioactivity testing. After identification, I generated production strains of the target compounds and elucidated the structure of the bioactive derivative.

The last chapter deals with the structure elucidation of peptides produced by an unusual GameXPeptide synthetase in *Xenorhabdus miraniensis*. I analysed producer strains of GameXPeptides using LC-MS and elucidated the structural differences between the known GameXPeptides, produced by *P. luminescens* TT01, and the unusual ones produced by *X. miraniensis*.

2 MATERIALS AND METHODS

2.1 Strains

Table 1. Bacterial strains used in this work

Strain	Genotype	Reference
<i>E. coli</i> ST18 DSM 22074	<i>pro thi hsdR</i> ⁺ T _p ^r Sm ^r ; chromosome::RP4-2 Tc::Mu- Kan::Tn7/λpir Δ <i>hemA</i>	¹⁶⁰
MD13	<i>E. coli</i> ST18 + pCEP_kan_xszeV2_10685 (<i>iszA</i>) (pCEP_MD16)	This work
MD14	<i>E. coli</i> ST18 + pCEP_kan_xszeUSV2_12255 (<i>iszA</i>) (pCEP_MD17)	This work
MD15	<i>E. coli</i> ST18 + pCEP_kan_xindV2_04300 (<i>iszA</i>) (pCEP_MD22)	This work
MD16	<i>E. coli</i> ST18 + pCEP_kan_ptthrac_19260 (<i>iszA</i>) (pCEP_MD15)	This work
MD17	<i>E. coli</i> ST18 + pCEP_kan_xstoV2_10830 (<i>iszA</i>) (pCEP_MD24)	This work
MD10	<i>E. coli</i> ST18 + pCEP_kan_sgaA (pCEP_MD11)	This work
MD24	<i>E. coli</i> ST18 + pEB17_kan_Δ <i>hfq</i> _Ptt (pEB17_MD22)	This work
MD23	<i>E. coli</i> ST18 + pEB17_kan_Δ <i>hfq</i> _Xind (pEB17_MD23)	This work

MD32	<i>E. coli</i> ST18 + pEB17_kan_ΔsgaA (pEB17_MD24)	This work
MD33	<i>E. coli</i> ST18 + pEB17_kan_ΔsgaB (pEB17_MD25)	This work
MD34	<i>E. coli</i> ST18 + pEB17_kan_ΔsgaAB (pEB17_MD26)	This work
<i>Xenorhabdus szentirmaii</i> DSM 16338	wildtype	161
<i>X. szentirmaii</i> Δhfq	Δhfq	Nick L. V. Neubacher
<i>X. szentirmaii</i> ΔngrA	ΔngrA (xsze01772)	Sebastian L. Wenski
<i>X. szentirmaii</i> P _{BAD} _02985	P _{BAD} _xsze02985	Jürgen Breitenbach
MD18 <i>X. szentirmaii</i> P _{BAD} _iszA	<i>X. szentirmaii</i> + pCEP_kan_ xszeV2_10685	This work
MD27 <i>X. szentirmaii</i> Δhfq P _{BAD} _iszA	<i>X. szentirmaii</i> Δhfq + pCEP_kan_ xszeV2_10685	This work
<i>Xenorhabdus szentirmaii</i> US	wildtype	161
<i>X. szentirmaii</i> US Δhfq	Δhfq	Nick L. V. Neubacher
MD19 <i>X. szentirmaii</i> US P _{BAD} _iszA	<i>X. szentirmaii</i> US + pCEP_kan_ xszeUSV2_12255	This work
MD28 <i>X. szentirmaii</i> US Δhfq P _{BAD} _iszA	<i>X. szentirmaii</i> US Δhfq + pCEP_kan_ xszeUSV2_12255	This work
<i>Xenorhabdus indica</i> DSM 17382	wildtype	162
MD25 <i>X. indica</i> Δhfq	Δhfq	This work & Karin Münch
MD20 <i>X. indica</i> P _{BAD} _iszA	<i>X. indica</i> + pCEP_kan_xindV2_04300	This work

MD29 <i>X. indica</i> Δhfq <i>P_{BAD}_iszA</i>	<i>X. indica</i> Δhfq pCEP_kan_xindV2_04300	+	This work & Karin Munch
<i>Photorhabdus thracensis</i> DSM 15199	wildtype		163
MD26 <i>P. thracensis</i> Δhfq	Δhfq		This work & Karin Munch
MD21 <i>P. thracensis</i> <i>P_{BAD}_iszA</i>	<i>P. thracensis</i> pCEP_kan_ptthrac_19260	+	This work
MD30 <i>P. thracensis</i> Δhfq <i>P_{BAD}_iszA</i>	<i>P. thracensis</i> Δhfq pCEP_kan_ptthrac_19260	+	This work & Karin Munch
<i>Xenorhabdus stockiae</i> DSM 17904	wildtype		23
<i>X. stockiae</i> Δhfq	Δhfq		Elizabeth Ransone
MD22 <i>X. stockiae</i> <i>P_{BAD}_iszA</i>	<i>X. stockiae</i> pCEP_kan_xstoV2_10830	+	This work & Karin Munch
MD31 <i>X. stockiae</i> Δhfq <i>P_{BAD}_iszA</i>	<i>X. stockiae</i> Δhfq pCEP_kan_xstoV2_10830	+	This work & Karin Munch
<i>Pseudomonas</i> MYb115	wildtype		164
MD12 MYb115 <i>P_{BAD}_sgaA</i>	MYb115 + pCEP_kan_sgaA		This work
MD35 MYb115 $\Delta sgaA$	MYb115 + pEB17_kan_ $\Delta sgaA$		This work
MD36 MYb115 $\Delta sgaB$	MYb115 + pEB17_kan_ $\Delta sgaB$		This work
MD37 MYb115 $\Delta sgaAB$	MYb115 + pEB17_kan_ $\Delta sgaAB$		This work
<i>Pseudomonas</i> MYb11	wildtype		164
<i>Pseudomonas</i> MYb12	wildtype		164
<i>Xenorhabdus miraniensis</i> DSM 17902	wildtype		23
<i>X. miraniensis</i> <i>P_{BAD}_gxps</i>	<i>P_{BAD}_gxps</i>		Zhengyi Qian
<i>Photorhabdus luminescens</i> subsp. <i>laumondii</i> TT01	wildtype, <i>rifR</i>		165
TT01 <i>P_{BAD}_gxps</i>	<i>P_{BAD}_gxps</i>		Zhengyi Qian

2.2 Oligonucleotides

Table 2. Oligonucleotides used in this work.

Purpose/target	Name	Sequence (5' to 3')
Upstream region of <i>sgaA</i> for deletion	MDp89	CGATCCTCTAGAGTCGACCTGCAGCA CAACAGATTCTCCTCATTCC
Starting region of <i>sgaA</i> for promoter exchange	MDp105	GGCTAACAGGAGGCTAGCATATGTTG ACAAAGCGTAGACAGG
Starting region of <i>sgaA</i> for promoter exchange	MDp106	CCGTTTAAACATTTAAATCTGCAGGC AATACTTGAGGTGTTGC
Starting region of <i>iszA</i> for promoter exchange in <i>P. thracensis</i>	MDp142	CCGTTTAAACATTTAAATCTGCAGGC AACGTCTGCTAATGTATCC
Starting region of <i>iszA</i> for promoter exchange in <i>P. thracensis</i>	MDp145	GGCTAACAGGAGGCTAGCATATGAA TACGCCACATAGCCAC
Starting region of <i>iszA</i> for promoter exchange in <i>X. szentirmaii</i>	MDp146	GGCTAACAGGAGGCTAGCATATGAA TTCGCCACATAGCCTC
Starting region of <i>iszA</i> for promoter exchange in <i>X. szentirmaii</i>	MDp147	CCGTTTAAACATTTAAATCTGCAGCC AAGCAATTGTAATTATCTGC
Starting region of <i>iszA</i> for promoter exchange in <i>X. indica</i>	MDp149	GGCTAACAGGAGGCTAGCATATGAA TATGACACGTAACCATACATCC
Starting region of <i>iszA</i> for promoter exchange in <i>X. indica</i>	MDp150	CCGTTTAAACATTTAAATCTGCAGCT GCAAGATCTTTGATAACCAG
Verification of the promoter exchange in front of <i>iszA</i> in <i>P. thracensis</i>	MDp151	GCGGAATTAGGGCTGGAGGC

Verification of the promoter exchange in front of <i>iszA</i> in <i>X. szentirmaii</i>	MDp152	CGTGTAATTTAGAGTCAAGATAGGCCG CG
Verification of the promoter exchange in front of <i>iszA</i> in <i>X. indica</i>	MDp153	GCAGCTTCTGCCAGTGCTTCC
Starting region of <i>iszA</i> for promoter exchange in <i>X. stockiae</i>	MDp163	GCTAACAGGAGGCTAGCATATGAAT ACACAGCGTAACCACAATC
Starting region of <i>iszA</i> for promoter exchange in <i>X. stockiae</i>	MDp164	CGTTTAAACATTTAAATCTGCAGCTG TAGTGTTCTAAAAGCGCCTTG
Upstream region of <i>hfq</i> in for deletion in <i>P. thracensis</i>	MDp166	CGATCCTCTAGAGTCGACCTGCAGCA TCGTTTGATCGATATTCTTG
Upstream region of <i>hfq</i> in for deletion in <i>P. thracensis</i>	MDp167	CTTCATTTTGTACCGTTCTCAAGCTC TATATTTTCCTATTTTGTGTTTTTA AC
Downstream region of <i>hfq</i> in for deletion in <i>P. thracensis</i>	MDp168	CAACAAAATAAGGAAAATATAGAGC TTGAGAACGGTAACAAAATG
Downstream region of <i>hfq</i> in for deletion in <i>P. thracensis</i>	MDp169	GCTCAGATCTACGCGTTTCATATGCG GTATATCCGCCTTATTCC
Verification of the deletion of <i>hfq</i> in <i>P. thracensis</i>	MDp170	CTACGATGAAATGGTTTATCGC
Verification of the deletion of <i>hfq</i> in <i>P. thracensis</i>	MDp171	CCTTCTCCGACAAAATATTTCG
Upstream region of <i>hfq</i> in for deletion in <i>X. indica</i>	MDp172	CGATCCTCTAGAGTCGACCTGCAGCC ACAGAAGAACAGGTACAGG
Upstream region of <i>hfq</i> in for deletion in <i>X. indica</i>	MDp173	CATCAGGCACTATCACTGATTCTATA TTTTCTTATTTTGTGTTTTTAAC

Downstream region of <i>hfq</i> in for deletion in <i>X. indica</i>	MDp174	CAACAAAATAAGGAAAATATAGAAT CAGTGATAGTGCCTGATGTAA
Downstream region of <i>hfq</i> in for deletion in <i>X. indica</i>	MDp175	GCTCAGATCTACGCGTTTCATATGGC TGATCAGCAGCATATAACC
Verification of the deletion of <i>hfq</i> in <i>X. indica</i>	MDp176	GGGTTTGAAGATGAAGTTAAAGC
Verification of the deletion of <i>hfq</i> in <i>X. indica</i>	MDp177	CTACCTGTCACTATCTGCACAGG
Verification of the promoter exchange in front of <i>iszA</i> in <i>X. stockiae</i>	MDp178	CTTCAGTAATAAAAATCAAGATAAGAA GGAGC
Forward primer for verification of pCEP plasmids	MDp180	GCTATGCCATAGCATT TTTATCCATA AG
Upstream region of <i>sgaB</i> for deletion in MYb115	MDp181	CGATCCTCTAGAGTCGACCTGCAGGG TGTTCGAACGTATCGC
Upstream region of <i>sgaB</i> for deletion in MYb115	MDp182	CACAATCGCCTGCACATTTAGATCAC GAACTATCACTCAATCAAACG
Downstream region of <i>sgaB</i> for deletion in MYb115	MDp183	CCTAACCGTTTGATTGAGTGATAGTT CGTGATCTAAATGTGCAGGCG
Downstream region of <i>sgaB</i> for deletion in MYb115	MDp184	GGAGAGCTCAGATCTACGCGTTTCAT ATGCGAACTCGGTGAACTGC
Forward primer for verification of pEB17 plasmids	MDp185	GATCGATCCTCTAGAGTCGACCT
Reverse primer for verification of pEB17 plasmids	MDp186	ACATGTGGAATTGTGAGCGG
Upstream region of <i>sgaA</i> for deletion in MYb115	MDp191	CCGGTGGGCGTTTTTACAGTCATGTC TACGCTTTGTCAAACCC

Downstream region of <i>sgaA</i> for deletion in MYb115	MDp192	GGAATGGGTTTTGACAAAGCGTAGAC ATGACTGTAAAAACGCCAC
Downstream region of <i>sgaA</i> for deletion in MYb115	MDp193	CCCGGGAGAGCTCAGATCTACGCGTT TCATATGCCCATCCAGATATCCACG
Upstream region of <i>sgaA</i> for deletion of <i>sgaAB</i> in MYb115	MDp194	CCACACAATCGCCTGCACATTTAGAT CAGTCTACGCTTTGTCAAACCC
Downstream region of <i>sgaA</i> for deletion of <i>sgaAB</i> in MYb115	MDp195	GGAATGGGTTTTGACAAAGCGTAGAC TGATCTAAATGTGCAGGCG
Verification of the deletion of <i>sgaB</i>	MDp196	CTGTTGCGCTATGCGC
Verification of the deletion of <i>sgaB</i>	MDp197	GCGATTTCTACCAGAACACCC
Verification of the deletion of <i>sgaAB</i>	MDp204	GCACCTCTTACTCATGTTATGCC
Verification of the deletion of <i>sgaAB</i>	MDp205	GCGGCAGACCACAAAATCTG
Verification of the deletion of <i>sgaA</i>	MDp208	CGCGAAAACTTCCGGACAC
Verification of the deletion of <i>sgaA</i>	MDp209	GAACCGATAGAACGCTTCCG
check Deletion pks fw mit 211	MDp210	CGGCATCCATCGCTTAGC
check Deletion pks rev mit 210	MDp211	TAGAGGATCGGCTGGGCG

2.3 Plasmids

Table 3. Plasmids used in this work.

Name	Description	Reference
pCEP_kan	Based on pDS132, R6K γ , <i>oriT</i> , Km ^R , <i>araC</i> , P _{BAD}	¹⁶⁶
pEB17_kan	Based on pDS132, R6K γ , <i>oriT</i> , Km ^R , <i>cipB</i> , <i>sacB</i>	⁷²
pCEP_kan_xszeV2_10685 (<i>iszA</i>)	pCEP_kan with first 517 bp of <i>iszA</i> from <i>X. szentirmaii</i> DSM (<i>xszeV2_10685</i>)	This work
pCEP_kan_xszeUSV2_12255 (<i>iszA</i>)	pCEP_kan with first 517 bp of <i>iszA</i> from <i>X. szentirmaii</i> US (<i>xszeUSV2_12255</i>)	This work
pCEP_kan_xindV2_04300 (<i>iszA</i>)	pCEP_kan with first 499 bp of <i>iszA</i> from <i>X. indica</i> DSM (<i>xindV2_04300</i>)	This work
pCEP_kan_ptthrac_19260 (<i>iszA</i>)	pCEP_kan with first 545 bp of <i>iszA</i> from <i>P. thracensis</i> DSM (<i>ptthrac_19260</i>)	This work
pCEP_kan_xstoV2_10830 (<i>iszA</i>)	pCEP_kan with first 491 bp of <i>iszA</i> from <i>X. stockiae</i> DSM (<i>xstoV2_10830</i>)	This work
pCEP_kan_sgaA	pCEP_kan with first 457 bp of <i>sgaA</i> from <i>Pseudomonas</i> MYb115 (<i>ctg18_3</i>)	This work
pEB17_kan_Δ <i>hfq</i> _Ptt	pEB17_kan with 854 bp upstream and 691 bp downstream of <i>hfq</i> from <i>P. thracensis</i> DSM (<i>ptthrac_21455</i>)	This work
pEB17_kan_Δ <i>hfq</i> _Xind	pEB17_kan with 922 bp upstream and 824 bp downstream of <i>hfq</i> from <i>indica</i> DSM (<i>xindV2_18045</i>)	This work
pEB17_kan_Δ <i>sgaA</i>	pEB17_kan with 898 bp upstream and 884 bp downstream of <i>sgaA</i> (<i>ctg18_3</i>) from <i>Pseudomonas</i> MYb115	This work
pEB17_kan_Δ <i>sgaB</i>	pEB17_kan with 1013 bp upstream and 1010 bp downstream of <i>sgaB</i> (<i>ctg18_4</i>) from <i>Pseudomonas</i> MYb115	This work

pEB17_kan_ΔsgaAB	pEB17_kan with 922 bp upstream of sgaA (ctg18_3) and 1010 bp downstream of sgaB (ctg18_4) from <i>Pseudomonas</i> MYb115	This work
------------------	--------------------------------------------------------------------------------------------------------------------------	-----------

2.4 Cultivation

All *E. coli*, *Photorhabdus* and *Xenorhabdus* strains were cultivated in Lysogeny broth (LB) medium (Tab. 4).

Table 4. Composition of Lysogeny broth (LB) medium.

Component	Concentration
Tryptone	10 g/L
Yeast extract	5 g/L
NaCl	5 g/L

For cultivation of agar plates 1.5 % (w/v) agar was added prior to autoclaving.

All liquid *Pseudomonas* cultures were cultivated in either LB or tryptic soy broth (TSB, Merck). Cultivation of *Pseudomonas* strains on agar plates was done on either LB agar or tryptic soy agar (TSA, Merck).

For natural product production, cultures either LB, SF900 (Gibco™) or XPP (Tab 5) were used as described in the respective chapter.

Table 5. Composition of Xenorhabdus-Photorhabdus-Production (XPP) medium.

Component	Concentration
glycerol	10 g/L
sodium pyruvate	1 g/L
M9 salt A	2 % (v/v)
M9 salt B	2 % (v/v)
proteinogenic L-amino acids	0.1 g/L of each
vitamin solution	0.2 % (v/v)
Trace element solution	0.1 % (v/v)
M9 salt A	
K ₂ HPO ₄	350 g/L
KH ₂ PO ₄	100 g/L
M9 salt B	
Sodium citrate	29.4 g/L
Ammonium sulphate	50 g/L
Magnesium sulphate	5 g/L
Vitamin solution	
Folic acid	10 mg/L
biotin	6 mg/L
<i>para</i> -aminobenzoic acid	200 mg/L
Thiamine hydrochloride	1 g/L
Panthenic acid	1.2 g/L
Riboflavin	100 mg/L
Nicotinic acid	2.3 g/L
Pyridoxine hydrochloride	12 g/L
Vitamin B12	20 mg/L
Trace element solution	

ZnCl ₂	40 mg/L
FeCl ₃ x 6 H ₂ O	200 mg/L
CuCl ₂ x 2 H ₂ O	10 mg/L
MnCl ₂ x 4 H ₂ O	10 mg/L
Na ₂ B ₄ O ₇ x 10 H ₂ O	10 mg/L
(NH ₄) ₆ Mo ₇ O ₂₄ x 4 H ₂ O	10 mg/L

If the respective strains carried an antibiotic resistance gene, antibiotics were added to the culture medium in the following concentrations (Tab. 6).

Table 6. Antibiotics

antibiotic	final concentration
chloramphenicol	34 µg/mL
kanamycin	50 µg/mL

2.4.1 Cultivation in 20 L-fermenters

Large scale fermentations of in 20 L-fermenters were carried out by Dr. Anja Schüffler (IBWF gGmbH, Mainz). For intraszentin production the oxygen flow was set at 2 L/min and the stirrer was set to 150 rpm. For sphinganine production the oxygen flow was set at 2 L/min and the stirrer was set to 120 rpm.

2.5 Molecular biology

2.5.1 Isolation of DNA

2.5.1.1 Isolation of genomic DNA

Genomic DNA was isolated using the Gentra Puregene Yeast/Bact. Kit (Qiagen) according to the manufacturer's instructions.

2.5.1.2 Isolation of Plasmid DNA

Plasmidic DNA was isolated using either the Invisorb® Plasmid Spin Mini *Two* kit (Strattec) according to the manufacturer's instructions.

2.5.2 Cloning

2.5.2.1 PCR

Polymerase Chain reaction (PCR) was used to amplify DNA fragments¹⁶⁷. 0.2 units of either Q5® High-Fidelity DNA Polymerase, Phusion® High-Fidelity DNA Polymerase (both New England Biolabs), S7 Fusion Polymerase™ (MobiDiag) or in-house purified Taq-Polymerase were used for 20 µL of PCR reaction according to the manufacturer's instructions (if available).

2.5.2.2 Purification of PCR products

PCR products were purified from the reaction mix using the MSB® Spin PCRapace Kit (Strattec) according to the manufacturer's instructions.

2.5.2.3 Enzymatic Digest of Plasmid Vectors

For insertion of DNA fragments into plasmid vectors *via* Hot Fusion cloning¹⁶⁸ Plasmid vectors pCEP and pEB17 were digested with NEB restriction enzymes PstI-HF® and NdeI according to the manufacturer's instructions and subsequently purified.

2.5.2.4 Purification of DNA from Agarose Gels

After Agarose gel electrophoresis, digested plasmid vectors were purified using the Invisorb® Spin DNA extraction kit (Strattec) or the Monarch® DNA Gel Extraction Kit (New England Biolabs) according to the respective manufacturer's instructions.

2.5.2.5 Hot Fusion cloning

Plasmid assembly was performed by using either Hot Fusion Cloning¹⁶⁸ or the NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs) was used according to the manufacturer's instructions.

2.5.2.6 Transformation into *E. coli*

Assembled plasmids were transformed into electrocompetent *E. coli* cells *via* electroporation¹⁶⁹. First a shaking flask containing LB medium was inoculated 1:100 with an overnight culture. The culture was grown to OD₅₉₅ = 0.6. The cells were cooled on ice for 30 min. The cells were harvested *via* centrifugation at 2000 x g for 10 min at 4 °C and resuspended in one culture volume ice-cold ddH₂O and centrifuged again. After resuspension in one culture volume ice-cold ddH₂O, the cells were incubated at 4 °C for 30 min before they were harvested by centrifugation (4 °C, 15 min, 2000 x g). Subsequently the

cells were resuspended in 1/10 culture volume of 10 % (v/v) ice-cold glycerol and once more incubated at 4 °C for 30 min. After centrifugation (4 °C, 15 min, 1500 x g) the cells were resuspended in 1/250 culture volume of 10 % (v/v) ice-cold glycerol. Aliquots of 50 µL were stored at -80 °C.

Electrocompetent cells were thawed on ice ahead of electroporation. 50 to 100 ng of plasmid DNA was added and the mixture was transferred into a precooled 1 mm electroporation cuvette. Electroporation was performed at 1250 V, 25 µF and 200 Ω using the GenePulser® Xcell™ (Bio-Rad Laboratories). Afterwards, the cells were incubated in 1 mL LB at 37 °C shaking for 1 h. Thereafter, the cells were plated on LB agar plates containing the respective antibiotics and incubated overnight at 37 °C.

2.5.3 Conjugation

2.5.3.1 Conjugation of *Photorhabdus* and *Xenorhabdus*

E. coli ST18 cells containing conjugatable plasmids containing either the pCEP- or pEB17-backbone were used for conjugation of *Photorhabdus* and *Xenorhabdus* strains. First 5 mL overnight cultures of the donor strain, *E. coli* ST18, were grown in liquid LB medium containing ALA and the respective antibiotic at 37 °C, while 5 mL overnight cultures of the acceptor strain, *Photorhabdus* or *Xenorhabdus*, were grown in LB medium at 30 °C. On the day of the conjugation liquid cultures of donor and acceptor strains were inoculated and grown to OD₆₀₀ = 0.8 to 1 (donor) and OD₆₀₀ ≥ 1 (acceptor) respectively. 1 mL of donor culture was harvested via centrifugation (17.000 x g, 1 min) and subsequently washed three times with LB. 5 mL of acceptor culture was also harvested via centrifugation (17.000 x g, 1 min). Cell pellets of the donor and acceptor cultures were resuspended in 50 µL each LB and mixed. Thereafter 90 µL of the mixture was dropped on a LB agar plate and incubated at 30 °C. On the next day the cells were scraped off the plate, resuspended in 1 mL LB and plated in different dilutions onto fresh agar plates containing the respective selection marker.

2.5.3.2 Conjugation of *Pseudomonas MYb115*

E. coli ST18 cells containing conjugatable plasmids containing either the pCEP- or pEB17-backbone were used for conjugation of *Pseudomonas* strains. First 5 mL overnight cultures of the donor strain, *E. coli* ST18, were grown in liquid LB medium containing ALA and the respective antibiotic at 37 °C, while 5 mL overnight cultures of the acceptor strain,

Pseudomonas, were grown in LB or TSB medium at 30 °C. The OD₆₀₀ of all cultures was adjusted to 6 and the donor strain culture was washed twice with LB. Subsequently 1 mL donor strain culture and 100 µL of acceptor strain culture were harvested via centrifugation (17.000 x g, 1 min), resuspended in 50 µL LB each, mixed and dropped on a agar plate. The plate was then incubated at 30 °C for six hours. The plaque was then scraped off the plate, resuspended in 1 mL LB and plated in different dilutions onto fresh agar plates containing the respective selection marker.

2.5.4 Promoter exchange

For promoter exchanges for targeted gene cluster activation, a plasmid with P_{BAD} was integrated upstream of the target gene as previously described (10.1002/cbic.201500094.). First, a fragment containing of the first 400 to 800 bp of the target gene was amplified by PCR. Using this fragment and a previously prepared vector, that was linearised *via* restriction digest with *NdeI* and *PstI*-HF (NEB) according to the manufacturer's instructions, the promoter exchange plasmid was assembled *via* HotFusion cloning (2.5.2.5). The plasmid was subsequently transformed into *E. coli* ST18. Conjugation into *Photorhabdus*, *Xenorhabdus* or *Pseudomonas* was carried out as described above (2.5.3).

2.5.5 Deletion

Deletions were carried out following a previously established protocol based on conjugation and homologous recombination¹²². First, fragments of 700 to 1000 bp directly upstream and downstream of the target gene were amplified by PCR. These fragments were assembled into a plasmid *via* HotFusion cloning (2.5.2.5) using the previously prepared and linearised vector pEB17. The assembled deletion plasmid was transformed into *E. coli* ST18. Conjugation into *Pseudomonas* was carried out as described above (2.5.3). Deletion strains were selected on LB agar plates containing 6 % (w/v) sucrose.

2.6 Analytical methods

2.6.1 Sample preparation

2.6.1.1 Extraction with XAD

For extraction of hydrophobic compounds and most peptides Amberlite® XAD-16 resin (Merck) was added to the growing culture. At a particular time point the cell culture was

separated from the resin via decantation. The resin was washed in one culture volume ddH₂O, which was discarded as well. The resin was then incubated in one culture volume methanol for 15 to 30 minutes. The resulting extract was diluted 1:10 with methanol, centrifuged (17,000 x g, 30 min), transferred to 0.3 mL PP short thread micro vials (Fisher Scientific) and submitted for measurement.

2.6.1.2 Extraction from culture

For whole culture extracts, especially used for network analysis in this work, 500 μ L cell culture were diluted with 500 μ L methanol and incubated for 15 to 30 minutes at 30 °C. The extract was diluted 1:5 with methanol and subsequently centrifuged (17,000 x g, 30 min) to sediment the resulting cell debris. Thereupon the extract was transferred to 0.3 mL PP short thread micro vials (Fisher Scientific) and submitted for measurement.

2.6.1.3 Extraction from cell pellet

Cells from 1 mL of cell culture were harvested via centrifugation (17,000 x g, 3 min). Subsequently the supernatant was discarded and the cells were resuspended in 1 mL methanol and incubated for 15 to 30 minutes at 30 °C. The resulting extract was diluted 1:10 with methanol and centrifuged (17,000 x g, 30 min) to sediment the resulting cell debris. Thereafter the extract was transferred to 0.3 mL PP short thread micro vials (Fisher Scientific) and submitted for measurement.

2.6.1.4 Extraction for Lipidomics

MYb115 wild type and MYb115 P_{BADSGa} were grown in 5 mL XPP medium to limit background noises. For each strain either 0.05 % (w/v) glucose or 0.02 % (w/v) arabinose were added, to either repress or enable gene expression in MYb115 P_{BADSGa}. Moreover both sugars were also added separately to the wild type strain serving as control for additional effects caused by addition of either glucose or arabinose. For each of these four conditions I did biological replicates and technical replicates for each biological replicate.

The samples were prepared using a modified protocol of Brown *et al.* (2019)¹⁷⁰. All samples were extracted from the equivalent of 1 mL OD=5. First, cells were harvested and the supernatant was discarded. Then the pellet was resuspended in 0.4 mL H₂O and transferred into a 4 mL glass vial. 0.5 mL CHCl₃ and 1 mL MeOH as well as 10 μ L internal standard were added. The samples were extracted on a shaking incubator for 18 h. 0.5 mL H₂O and

0.5 mL CHCl₃ were added. The lower phase was dried and stored at -20 °C until the measurement.

2.6.2 UPLC-MS analysis

UPLC-MS analyses were performed on a Dionex Ultimate 3000 (Thermo Fisher Scientific) coupled to an amaZon X Ion trap mass spectrometer (Bruker Daltonics). 5 µL sample were injected and a multistep gradient from 5 to 95 % ACN with 0.1 % formic acid in water with 0.1 % formic acid over 16 minutes with a flow rate of 0.4 mL/min was run (0-2 min 5 % ACN; 2-14 min 5-95 % ACN; 14-15 min 95 % ACN; 15-16 min 5% ACN) on a Acquity UPLC BEH C₁₈ 1.7 µm column (Waters). MS data acquisition took place between minutes 1 and 15 of the multistep LC gradient. The mass spectrometer was set to positive or alternating polarity mode with a capillary voltage of 4.5 kV and a nitrogen flow rate of 8 L/min.

2.6.3 UPLC-HRMS analysis

UPLC-HRMS analyses for determination of exact masses and sum formulas of analytes as well as for data acquisition for GNPS network analysis were performed on a Dionex Ultimate 3000 (Thermo Fisher Scientific) coupled to an Impact II qToF mass spectrometer (Bruker Daltonics). 5 µL sample were injected and a multistep gradient from 5 to 95 % ACN with 0.1 % formic acid in water with 0.1 % formic acid over 16 minutes with a flow rate of 0.4 mL/min was run (0-2 min 5 % ACN; 2-14 min 5-95 % ACN; 14-15 min 95 % ACN; 15-16 min 5% ACN) on a Acquity UPLC BEH C₁₈ 1.7 µm column (Waters). MS data acquisition took place between minutes 1.5 and 15 of the multistep LC gradient. The mass spectrometer was set to positive polarity mode with a capillary voltage of 2.5 kV and a nitrogen flow rate of 8 L/min. The mass spectrometer was calibrated using 10 mM sodium formate before data acquisition. The MS method used for data acquisition also included an internal calibrant window before the data acquisition of each biological sample where 10 mM sodium formate were injected. The internal calibrant was used by Bruker DataAnalysis to correct the acquired mass data.

2.6.4 Lipidomics analysis

Lipidomics experiments were performed by Dr. Georgia Angelidou and Dr. Nicole Paczia.

The extracts were resuspended and subjected to measurement. 5 µL sample were injected and a multistep gradient from 37 to 98 % B (IPA with 10 % (v/v) ACN; A: 60 % ACN in

water with 10 mM ammonium formate) over 25 minutes with a flow rate of 0.3 mL/min was run. (Tab. 7). An Acquity Premier CSH C18 column (2.1 × 100 mm, 1.7 μm particle size, VanGuard) was used. The column oven was set to 40 °C.

A Thermo Scientific ID-X Orbitrap mass spectrometer was used. Ionisation was performed using a high temperature electro spray ion source at a static spray voltage of 3500 V (positive) and a static spray voltage of 2800 V (negative), sheath gas at 50 (Arb), auxiliary Gas at 10 (Arb), and ion transfer tube and vaporizer at 325 °C and 300 °C, respectively.

Data dependent MS² measurements were conducted applying an orbitrap mass resolution of 120 000 using quadrupole isolation in a mass range of 200 – 2000 and combining it with a high energy collision dissociation (HCD). HCD was performed on the ten most abundant ions per scan with a relative collision energy of 25 %. Fragments were detected using the orbitrap mass analyser at a predefined mass resolution of 15 000. Dynamic exclusion with an exclusion duration of 5 sec after one scan with a mass tolerance of 10 ppm was used to increase coverage.

Table 7. LC gradient for lipidomics.

t [min]	c(B)
0 - 1.5	37 %
1.5 - 4	37 - 45 %
4 - 5	45 - 52 %
5 - 8	52 - 58 %
8 - 11	58 - 66 %
11 - 14	66 - 70 %
14 - 18	70 - 75 %
18 - 20	75 - 98 %
20 - 25	98 %
25 -30	37 %

2.6.5 Network analysis

A molecular network was created using GNPS¹⁴⁸. The precursor ion mass tolerance was set to 0.05 Da and a MS/MS fragment ion tolerance of 0.01 Da. A network was then created

where edges were filtered to have a cosine score above 0.7 and more than 6 matched peaks. Further, edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 7 most similar nodes. Finally, the maximum size of a molecular family was set to 100, and the lowest scoring edges were removed from molecular families until the molecular family size was below this threshold. The spectra in the network were then searched against GNPS' spectral libraries. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 6 matched peaks.

2.6.6 Feeding experiments

2.6.6.1 Stable isotope labelling

ISOGRO®-¹³C, -¹⁵N or ¹³C¹⁵N (Merck) was prepared according to the manufacturer's instructions. Before inoculating the production cultures, precultures were washed three times in 500 µL of the respective ISOGRO®-medium. Bacterial cultures were grown at 30 °C.

2.6.6.2 Amino acid feeding experiments

Inverse feeding experiments with unlabelled amino acids were carried out in ISOGRO®-¹³C as previously described¹⁵⁵.

For experiments carried out to determine the incorporation of serine into sphinganine compounds, either unlabeled or ¹³C₃¹⁵N-labelled serine was added to XPP medium lacking serine.

2.6.7 Solid phase peptide synthesis

Solid phase peptide synthesis for coinjection experiments to determine the stereochemistry of GameXPeptides was carried out by Trinetri Goel.

2.6.8 Bioactivity testing of massetolide E and viscosin

2.6.8.1 Preparative HPLC

Methanolic XAD extracts from 1 L cultures of MYb11 and MYb12 in XPP medium were used for fractionation. Preparative HPLC was performed using an Agilent LC 1260 Infinity II Preparative LC/MSD System with an Agilent prep C18 column (10 µm, 30 x 250 mm) with a flow of 40 mL/min and a gradient from 5% to 100% ACN in 18 min. The ACN concentration remained at 100% for 3 min (MYb11) or 5,5 min (MYb12), respectively. 0.8 min-fractions were collected over time from 1.5 min to the end of the respective run. The

fractions were dried and stored at -20 °C. 50 mg of each fraction were used in the disc diffusion assay.

2.6.8.2 *Disc diffusion assay*

For the disc diffusion assay, Mueller Hinton agar (Merck) plates were inoculated with *Bacillus thuringiensis* strains BT247 and BT679 using cotton swabs. Immediately thereafter, sterile 6 mm Whatman discs (GE healthcare), each containing 50 mg of one of the HPLC fractions, were placed on the agar plates. Whatman discs containing only methanol or 750 µg ampicillin were used as controls. The plates were incubated at 30 °C for 48 h.

3 RESULTS

3.1 Intraszentin

3.1.1 Linking Natural Product Families to Biosynthetic gene clusters

In order to investigate the complete NRPS- and PKS-metabolome of *Xenorhabdus szentirmaii*, I compared the wildtype strain to a *ngrA* deletion mutant as shown in my Master thesis¹³⁹. This past work was revisited and expanded during this study, in which I compared the wildtype strain *Xenorhabdus szentirmaii* DSM 16338 to the *ngrA* deletion mutant as well as to mutants in which the promotor in front of single biosynthetic gene clusters (BGCs) were exchanged, but the strains were cultivated without induction. I analysed the metabolome by using MS/MS-based GNPS network analysis¹⁴⁸.

The linkage between BGCs and natural product families (NPFs) was achieved by comparison of extracts of *X. szentirmaii* wildtype to $\Delta ngrA$ and promoter exchange strains. The latter were taken from a strain collection containing single promoter exchanges in front of most of the BGCs identified by antiSMASH 3.0¹⁷¹ based upon the most complete *X. szentirmaii* genome accessible at the time (accession no. GCA_002632585). The MS/MS data obtained from each promoter exchange strain was individually compared to wildtype and $\Delta ngrA$ data.

By applying this method the natural products (NPs) that are missing in the samples of the inactivated promoter exchange strain, should be distinguishable from all other NPs. This allows linkage of NPFs, represented by subnetworks, to BGCs. Furthermore, this method delivers a snapshot of all derivatives of a NPF produced under certain conditions. The promoter exchange in front of the BGC *xsze00346*, which is responsible for GameXPepide (GXP) production in *X. szentirmaii* DSM 16338, served as a proof of principle for this method (Fig. S 1). In this case one known GXP (GXP C m/z 552.409 [M+H]⁺) and two hitherto undescribed GXPs (m/z 566.425 [M+H]⁺ and m/z 597.466 [M+H]⁺) represented the GXP NPF. The presence of GXP C in the subnetwork confirmed the identity of this NPF and approved the functionality of this method.

An initial analysis, revealed only one subnetwork consisting of only three nodes that are abolished when the BGC *xsze02985*-*xsze02984* is inactivated (m/z 872.426 [M+2H]²⁺; m/z

872.928 [M+1+2H]²⁺; *m/z* 873.433 [M+2+2H]²⁺, data not shown). Samples for the initial analysis were taken only after 72 hours. However, in the analysis displayed below, the promoter exchange of the BGC *xsze02985*-*xsze02984* without induction led to the identification of a subnetwork consisting of *m/z* 872.426 [M+2H]²⁺ its isotopologue peaks (*m/z* 872.928 [M+1+2H]²⁺; *m/z* 873.433 [M+2+2H]²⁺; *m/z* 873.935 [M+3+2H]²⁺; *m/z* 873.936 [M+3+2H]²⁺) along with a sodium adduct (*m/z* 884.423 [M+H+Na]²⁺) and its isotopologue peak (*m/z* 884.928 [M+1+H+Na]²⁺) (Fig. 11 A and B). In this experiment, samples were taken after 24 h and 72 h and pooled together.

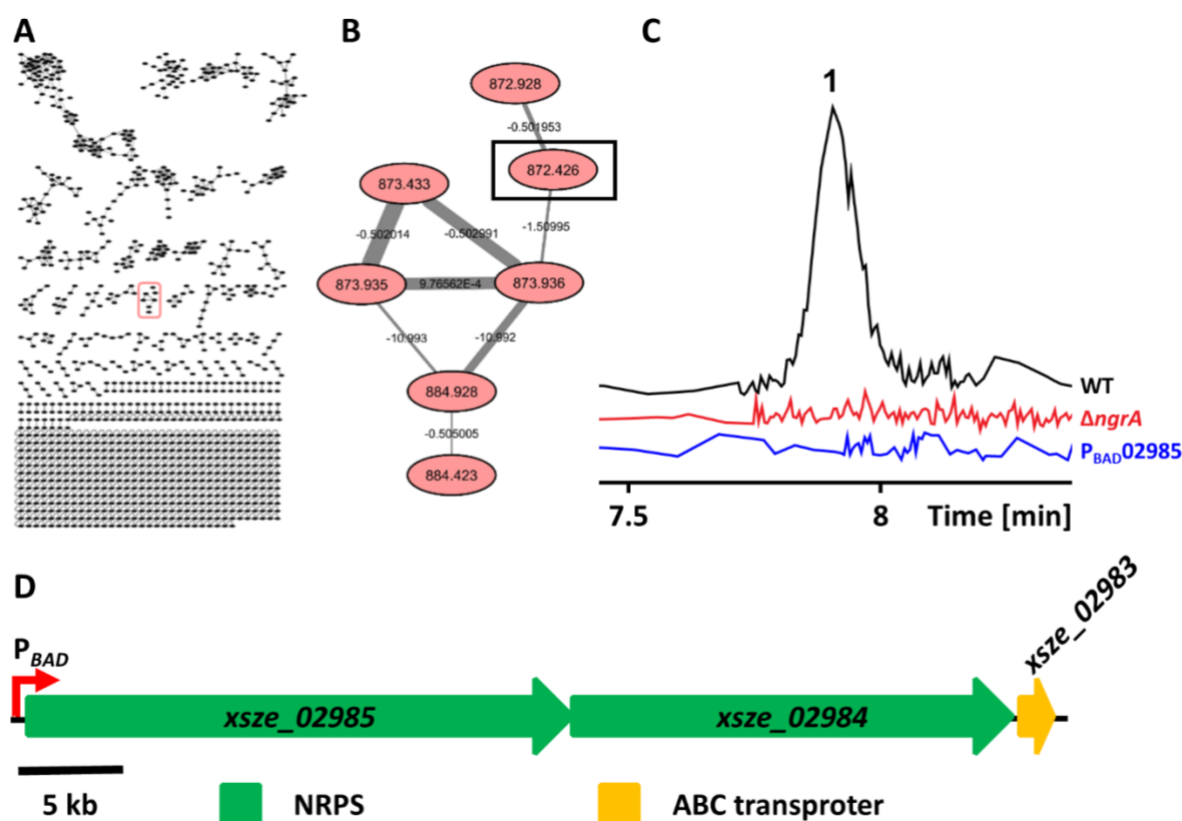


Figure 11. Network-guided identification of **1** and the corresponding BGC. A: Overview of the GNPS network analysis comparing *X. szentirmaii* DSM wild type, $\Delta ngrA$ and the silenced promoter exchange strain of BGC *xsze02985*-*xsze02984*. For detailed network analysis, see Figure S 2. B: Subnetwork of NPF containing **1** along with its isotopologues and sodium adducts. C: Extracted ion chromatograms (EICs) of **1** displaying production only in the wild type. D: Schematic display of the promoter exchange in front of BGC *xsze02985*-*xsze02984*.

This result was confirmed by individual analysis of the data underlying the network analysis (Fig. 11 C). **1** was only detectable in wild type samples, while production of **1** was abolished in both the non-induced promoter exchange and $\Delta ngrA$. The BGC responsible for production

of **1** consists of two NRPS encoding genes (*xsze_02985* and *xsze_02984*) and one gene encoding for an ABC transporter-like protein (*xsze_02983*) (Fig. 11 D). I named **1** intraszentin A and the BGC *isz*, consisting of the three genes *iszA* (*xsze_02985*), *iszB* (*xsze_02984*) and *iszC* (*xsze_02983*).

In addition to the subnetwork with **1**, the analysis displayed above, revealed two additional subnetworks that were influenced by the inactivation of *xsze02985-xsze02984* (Fig. S2). However, because of the results from the initial network analysis, **1** was regarded as the most interesting candidate.

The chemical identity of **1** was further investigated as it represented a hitherto undescribed NP.

3.1.2 Identification of intraszentin A from *X. szentirmaii*

The *isz* BGC remained silent in the previously used promoter exchange strains even upon induction (unpublished work by Svenja Simonyi and Jürgen Breitenbach). Since experiments aimed at the structure elucidation of **1** without overproduction, e.g. purification from wildtype cultures, inverse feeding experiments and MS/MS sequence analysis failed in this context, the decision was made to retry overproduction of **1** *via* promoter exchange at different positions in front of *iszA*. This was aided by the fact that the complete genome of *X. szentirmaii* DSM 16338 was resequenced during that time (unpublished sequence, Bode group).

3.1.2.1 Bioinformatic analysis

Bioinformatic analyses using antiSMASH v5.2.1¹⁷² led to the identification of *iszA-C* in the new genome sequence of *X. szentirmaii* DSM 16338. Furthermore, highly similar gene clusters have been found in the resequenced genomes of *X. szentirmaii* US, *X. indica* DSM 17382 and *P. thracensis* DSM 15199 (formerly *P. temperata subsp. thracensis*) (Fig. 12). A different version has been found in the genome sequence of *X. stockiae* DSM 17904 (Fig. 12 and personal communication; Dr. Yi-Ming Shi). All five of those BGCs consist of two genes encoding a NRPS and a gene encoding a putative ABC transporter (Fig. 12). While the NRPS encoding genes *iszA* and *iszB* of *X. szentirmaii*, *X. indica* and *P. thracensis* are 26 kb and 21 kb in length, the genetic makeup of the BGC in *X. stockiae* is different. In *X. stockiae* both NRPS-encoding genes are 13 kb and 34 kb, respectively. Unlike the transporter

gene in *X. stockiae*, which is encoded directly downstream of *iszB*, the transporter gene in the other strains starts 162 bp downstream of the *iszB*. Albeit equaling the size of the NRPS BGC in the four other strains, the genetic architecture displayed by the *X. stockiae* NRPS BGC hints at the production of a different NRP by this strain. While the BGCs of *X. szentirmaii* US, *X. indica* and *P. thracensis* all share at least 77.1 % identity with the BGC of *X. szentirmaii* DSM, the BGC of *X. stockiae* only shares 50.9 % identity with this cluster (Fig. 12).

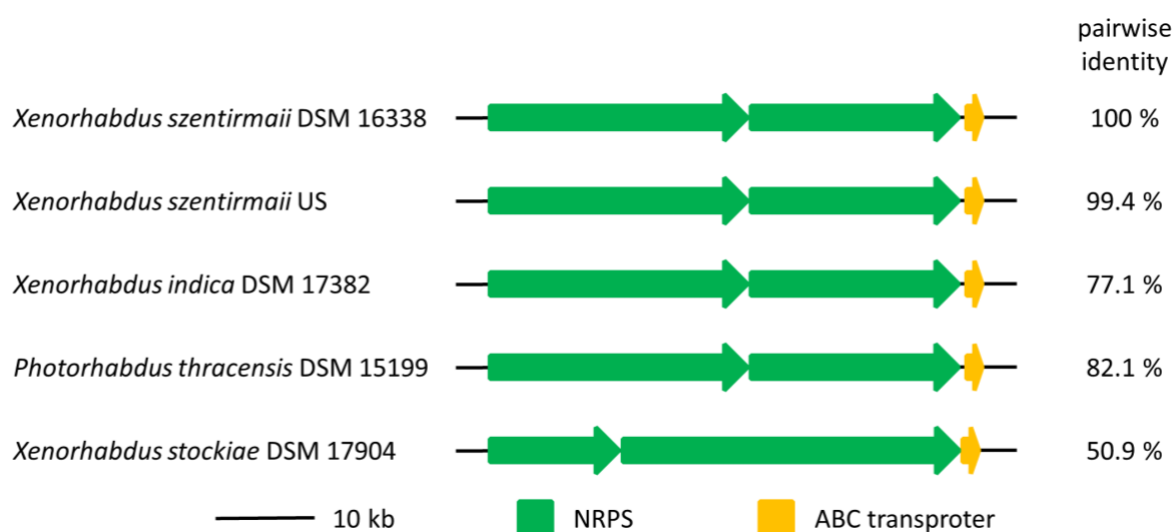


Figure 12. Comparison of *isz* BGCs found in *Xenorhabdus szentirmaii* DSM 16338, *X. szentirmaii* US, *X. indica* DSM 17382, *Photorhabdus thracensis* DSM15199 and *X. stockiae* DSM 17904 based on genetic makeup and pairwise identity compared to the *isz* BGC from *X. szentirmaii* DSM 16338. Pairwise identity of single genes is displayed in tables S1-S6.

To gather information what peptides the NRPS encoded in this BGCs might produce, the BGCs were further analysed using antiSMASH (Fig. 13).

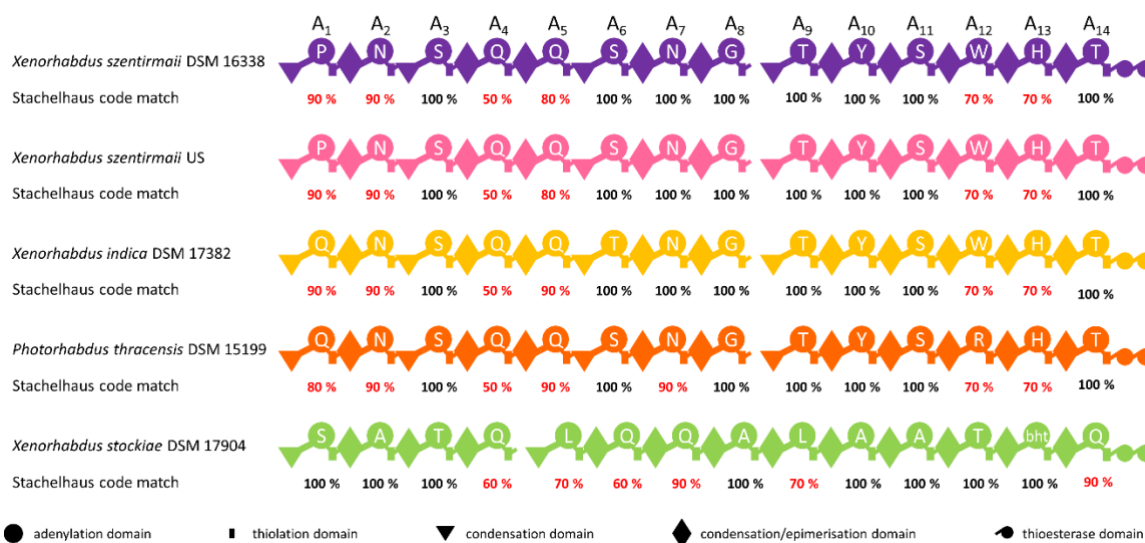


Figure 13. Comparison of Isz NRPS found in *Xenorhabdus szentirmaii* DSM 16338, *X. szentirmaii* US, *X. indica* DSM 17382, *Photorhabdus thracensis* DSM15199 and *X. stockiae* DSM 17904 based on predicted A domain specificity. Confidence of predicted A domain specificity is based on matches of the specific Stachelhaus code of each A domain to the Stachelhaus code used by antiSMASH to predict A domain specificity.

The predicted NRPS consist of fourteen modules with a starter C domain (C_{starter}) and a double TE domain (Fig. 13), indicating that they are closely related. NRPS with double TE domains, C_{starter} domains that incorporate at least one amino acid containing a hydroxyl group (Ser, Thr, Tyr), typically produce cyclic Lipodepsipeptides^{173–175}. Interestingly, some NRPS encoded by BGCs, that also encode for ABC transporters, have been found to produce peptides with antibiotic properties^{174,175}. The NRPS from *X. szentirmaii*, *X. indica* and *P. thracensis* share many predicted domain specificities. While the predicted specificities based on the Stachelhaus code are not always correct, similar predictions can be an indicator for similar specificities.

Nine of the fourteen adenylation (A) domains share a prediction based on a Stachelhaus code match of moderate to strong ($\geq 90\%$). If weaker predictions ($\leq 80\%$) are taken into account, eleven of the fourteen domains share the same prediction, indicating that the peptides produced by these NRPS are also highly similar to each other. On the contrary, the NRPS from *X. stockiae* shares zero predictions of the moderate to strong range ($\geq 90\%$) with the other NRPS and only one at the A4 domain in the weak range ($\leq 80\%$).

Moreover, every Isz NRPS except for the one from *X. stockiae* share the exact same C-C/E domain pattern, which is another indicator of their close relatedness (Fig. 13).

Based on the above analysis *isz* BGCs from both *X. szentirmaii* strains as well as from *X. indica* and *P. thracensis* can be grouped into a closely related group. The *isz* BGC from *X. stockiae* most likely shares a common ancestor with the other four, but more changes have taken place since. Nevertheless, all five BGCs were activated *via* insertion of the arabinose inducible promoter P_{BAD} in an approach that was shown to be effective for the targeted activation of BGCs¹⁶⁶.

3.1.2.2 Promoter exchange based gene cluster activation

A broad range promoter exchange approach was initiated to achieve overproduction of **1** and its homologues. For this approach, P_{BAD} was inserted directly in front of the respective start codons. This resulted in production of **1** in *X. szentirmaii* DSM 16338 (Fig. 14).

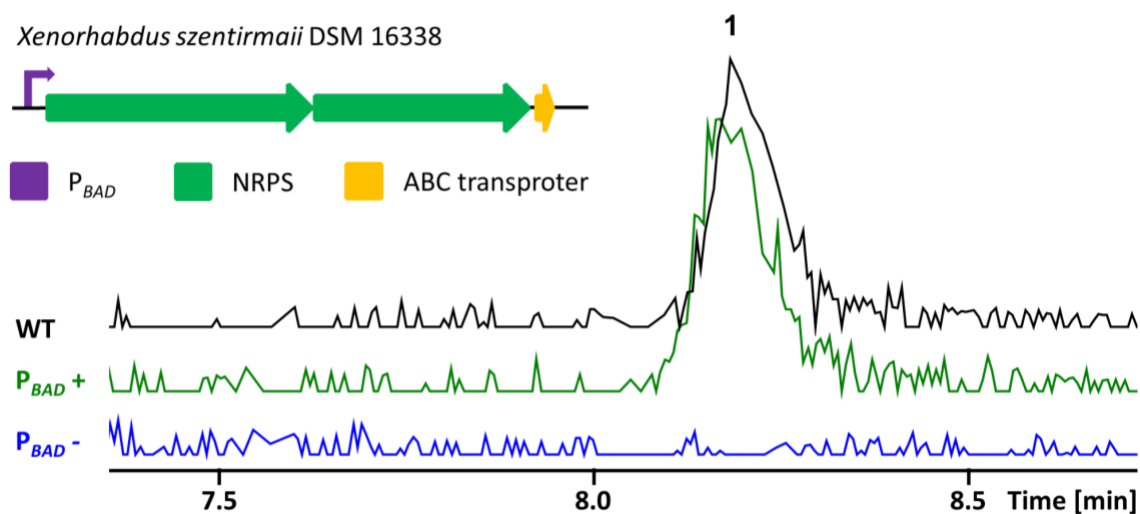


Figure 14. Promoter exchange based activation of *iszABC* in *X. szentirmaii* DSM 16338 led to the production of **1**.

Figure 14 displays the arabinose induced production of **1** in the promoter exchange strain of *X. szentirmaii* DSM 16338 in comparison to the respective strain without arabinose and the wild type strain. While production of **1** could be induced through promoter exchange, the production level did not exceed wild type levels (Fig. 14), nor was it visible in the base peak chromatogram (BPC). This again underlines the importance of the network analysis based

method presented in this chapter, since the target mass would still be unknown without that method.

To identify the NPs produced by the different strains and to circle out the best producing strain for subsequent isolation and structure elucidation, the promoter exchange strains of *X. szentirmaii* US, *X. indica*, *P. thracensis* and *X. stockiae* were analysed (Fig. 15).

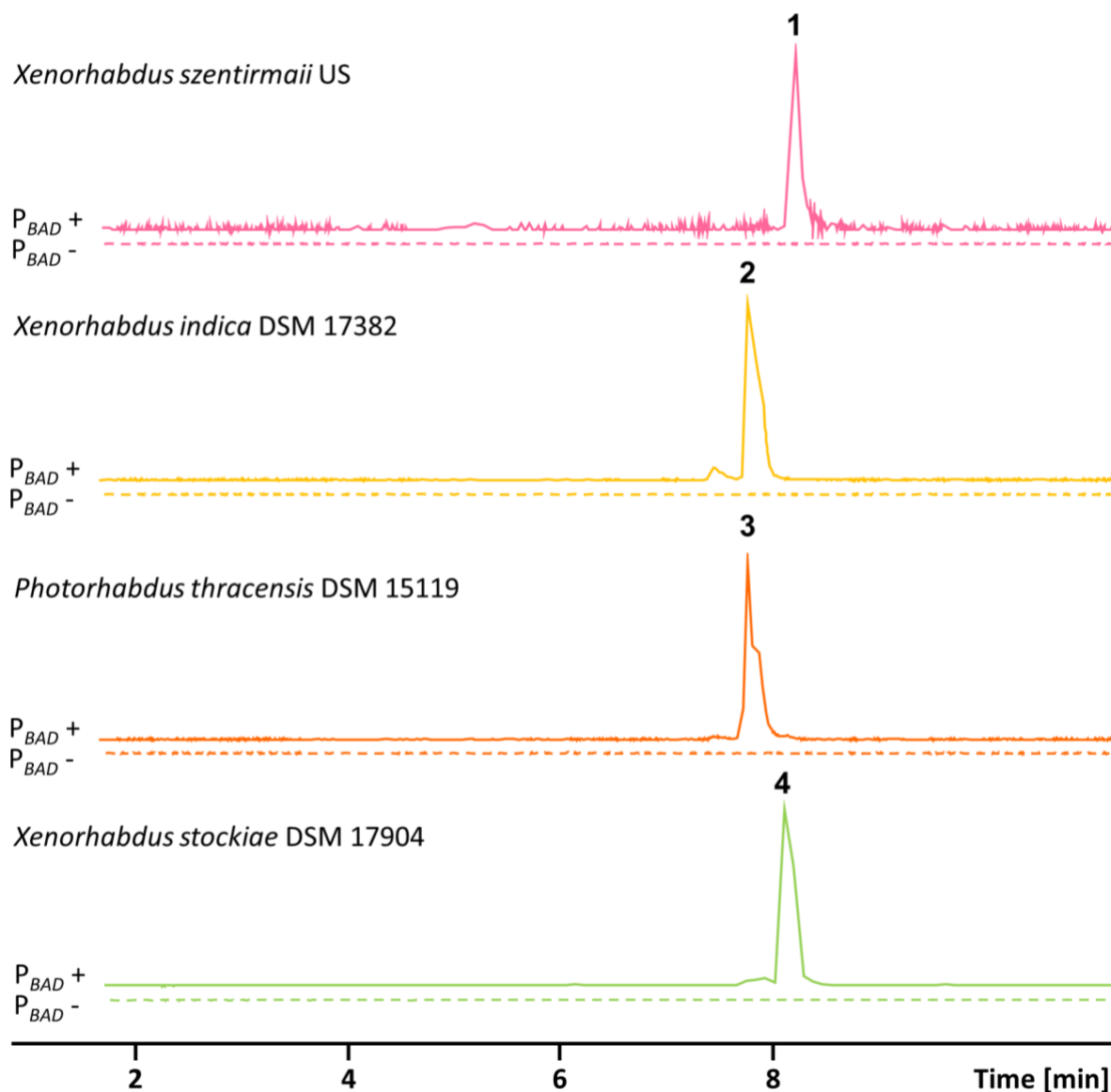


Figure 15. Promoter exchange based activation of *iszABC* in *X. szentirmaii* US, *X. indica* DSM 17382, *P. thracensis* DSM 15199 and *X. stockiae* DSM 17904 led to the production of **1**, **2**, **3** and **4**, respectively.

Promoter exchange based activation led to the identification of three additional NPs, **2** (m/z 904.95 $[M+2H]^{2+}$), **3** (m/z 888.94 $[M+2H]^{2+}$) and **4** (m/z 852.52 $[M+2H]^{2+}$) (Fig. 15). NPs **1** to **4** were given the names intraszentin A to D. To minimize the background for further investigation of intraszentins, the promoter exchanges were repeated in *hfq* deletion mutants of each strain by Karin Münch at the MPI for terrestrial microbiology in Marburg or me (Tab. 1). Hfq is a global molecular regulator for NP biosynthesis⁷⁰. Deletion mutants of *hfq* can be exploited for the production of NPs⁷².

The LC-MS based analysis of these strains led to the identification of the best producing promoter exchange strain for each NP (Fig. 16).

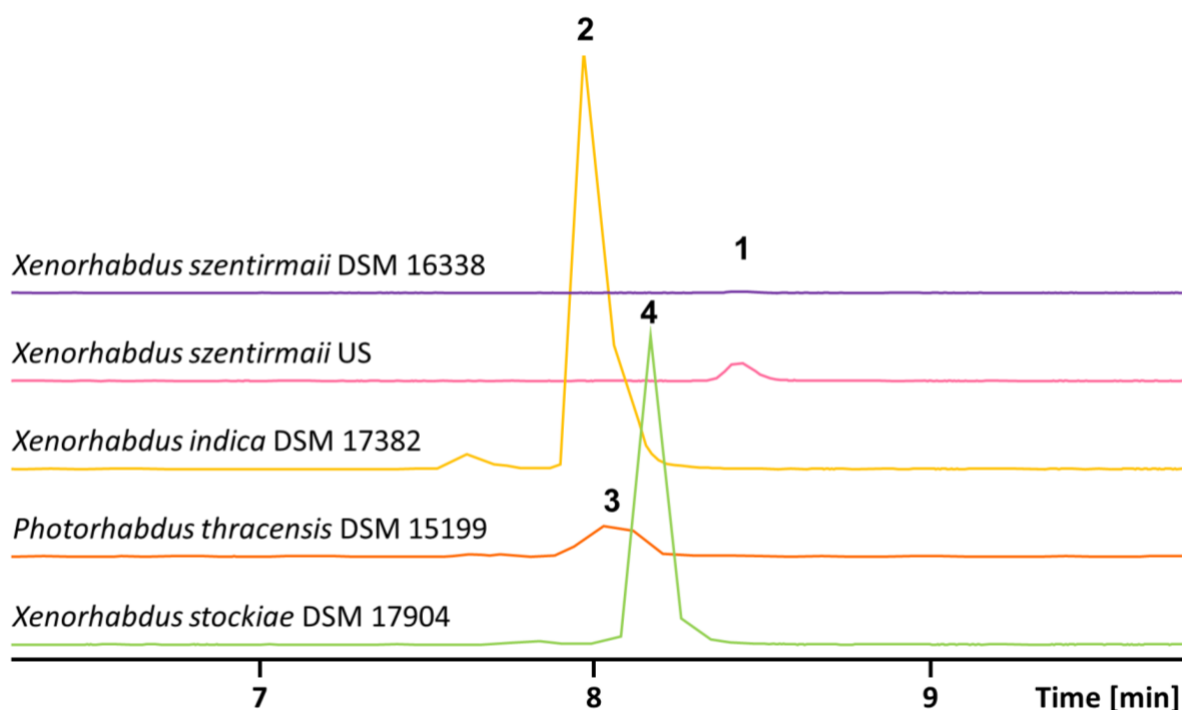


Figure 16. Production levels of **1-4** in $\Delta hfq P_{BAD_iszA}$ strains of *X.szentirmaii*, *X. indica*, *P. thracensis* and *X. stockiae*. Peak intensities were normalised.

While the production exceeded wildtype levels in every strain (data not shown), *X. indica* $\Delta hfq P_{BADxindV2_04300}$ was identified as the best producer (Fig. 16). This was ascertained by peak intensity only and concerns regarding possible differences in ionisation of **1-4** were put aside, because of the expected similarity of these compounds.

3.1.3 Structure elucidation via MS

3.1.3.1 Isotopic labelling

The best producers of **1-4** were grown in ^{13}C -ISGROTM and ^{15}N -ISOGROTM media as well as regular LB. Thereafter XAD extracts of these cultures were measured by LC-MS (Fig. 17). Since the molecules were detected as $[\text{M}+2\text{H}]^{2+}$, i.e. doubly charged ions, a mass shift of 0.5 correlates with the incorporation of 1 atom.

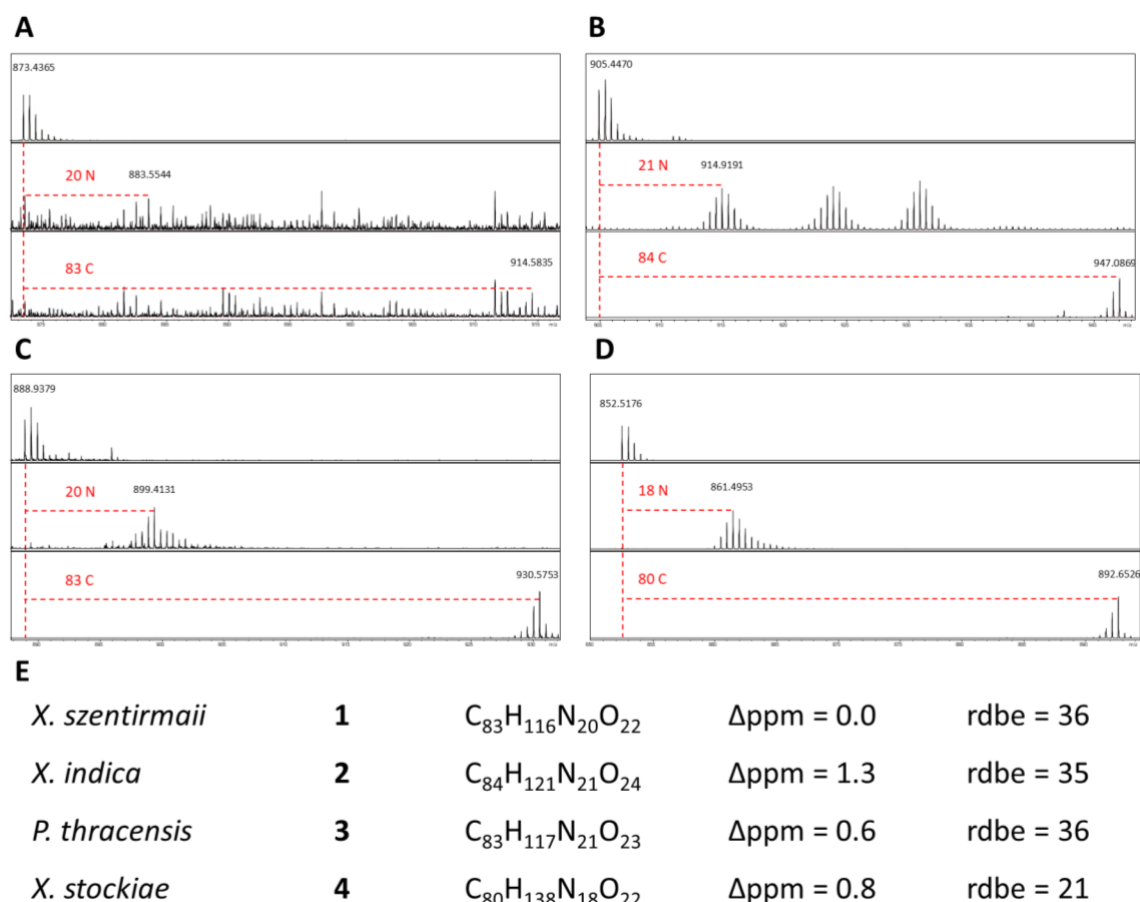


Figure 17. Sum formula determination of compounds **1-4** using isotopic labelling and LC-MS. **A-D:** Mass shifts of ^{13}C - and ^{15}N -labeled compounds compared to unlabeled compounds from LB cultivation. Dashed red lines represent the number of carbon and nitrogen atoms incorporated. **E:** Sum formulae of compounds **1-4**, as well as Δppm values and calculated rdbs.

Inthraszentin A (**1**) displayed a mass shift of 41.5 in ^{13}C -ISGROTM and 10 in ^{15}N -ISOGROTM (Fig. 17 A). Inthraszentin A therefore contains 83 carbon atoms and 20 nitrogen atoms. Thus, the sum formula of inthraszentin A was derived to be $\text{C}_{83}\text{H}_{116}\text{N}_{20}\text{O}_{22}$. A corresponding

structure would contain 36 double bond equivalents. Since the signal intensity for intraszentin A feeding experiments is rather low, even after identifying the best producer, the feeding data was backed up with MRMS data from Bruker (Fig. S 4).

Inthraszentin B (**2**) displayed a mass shift of 42 in ^{13}C -ISGROTM and 10.5 in ^{15}N -ISOGROTM (Fig. 17 B). Inthraszentin B therefore contains 84 carbon atoms and 21 nitrogen atoms. Thus, the sum formula of intraszentin B was derived to be $\text{C}_{84}\text{H}_{121}\text{N}_{21}\text{O}_{24}$. A corresponding structure would contain 35 double bond equivalents.

Inthraszentin C (**3**) displayed a mass shift of 41.5 in ^{13}C -ISGROTM and 10.5 in ^{15}N -ISOGROTM (Fig. 17 C). Inthraszentin C therefore contains 83 carbon atoms and 21 nitrogen atoms. Thus, the sum formula of intraszentin C was derived to be $\text{C}_{83}\text{H}_{117}\text{N}_{21}\text{O}_{23}$. A corresponding structure would contain 36 double bond equivalents.

Inthraszentin D (**4**) displayed a mass shift of 40 in ^{13}C -ISGROTM and 9 in ^{15}N -ISOGROTM (Fig. 17 D). Inthraszentin D therefore contains 80 carbon atoms and 18 nitrogen atoms. Thus, the sum formula of intraszentin D was derived to be $\text{C}_{80}\text{H}_{138}\text{N}_{18}\text{O}_{22}$. A corresponding structure would contain 20 double bond equivalents. As a result of this set of experiments, it became obvious that the structure of intraszentin D (**4**) deviates more strongly from those of intraszentins A-C (**1-3**) and therefore was not further investigated.

3.1.3.2 Inverse amino acid feeding

An inverse feeding approach¹⁵⁵ was applied to determine the amino acid composition. All strains were incubated in ^{13}C -ISGROTM and 22 amino acids, 20 proteinogenic plus β -alanine and ornithine, were added to the cultures, one per culture. The extracts were analyzed by LC-MS. Through comparison of the mass shifts, the different amino acids that are incorporated by the NRPS can be identified. Because doubly charged ions were analysed, a mass shift of 5 in a culture where glutamate was added, correlates to the incorporation of two glutamate molecules into the peptide.

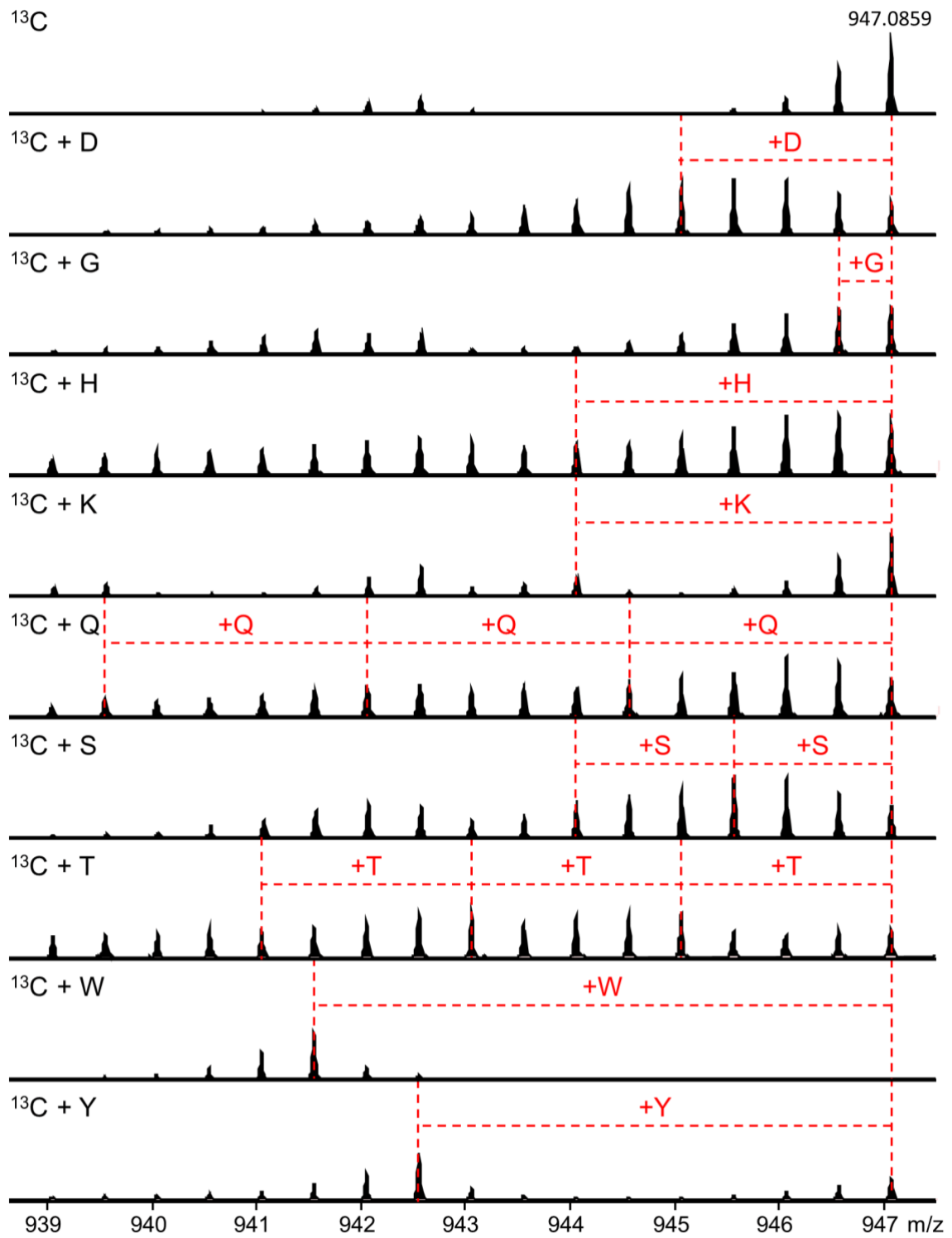


Figure 18. Mass shifts resulting from inverse feeding of amino acids to cultures grown in ^{13}C -labeled medium. Only samples in which a shift was observed are shown.

Inthraszentin B (**2**) consists of one aspartate residue, one glycine residue, one histidine residue, one lysine residue, three glutamine residues, two serine residues, three threonine residues, one tryptophan residue and one tyrosine residue (Fig. 18).

For, intraszentin C (**3**), however, one aspartate residue, one glycine residue, one histidine residue, one lysine residue, three glutamine residues, three serine residues, two threonine residues and one tryptophan residue could be confirmed (Fig. S 5). The culture, to which tyrosine was added, did not grow.

These results have to be treated with caution, since amino acids are part of the primary metabolism and some can be converted to another one with as little as one reaction, as I will reflect on this later in this chapter.

The signal intensities for intraszentin A (**1**) were too low to analyse the data confidently. For the structure predictions, the missing amino acid residues for **1** and **3** were derived from the similarities in A domain specificities, compared to IszAB from *X. indica*.

3.1.3.3 Structure predictions

Considering only the previous feeding experiments, I observed a disparity between the experimentally determined sum formulae and the added sum formulae of all incorporated amino acids (Tab. 8).

Table 8. Comparison of sum formulae of intraszentins A-C (**1-3**). Sum_{exp}: Experimentally determined sum formula; Sum_{aa}: Sum formula of a theoretical peptide consisting of all amino acids that were shown to be incorporated; Sum_{aa+dha}: Sum formula of a theoretical peptide consisting of all amino acids that were shown to be incorporated where every serine residue was regarded as dehydro alanine residue.

Compound	Sum _{exp}	Sum _{aa}	Sum _{aa+dha}
inthraszentin A	C ₈₃ H ₁₁₆ N ₂₀ O ₂₂	C ₇₀ H ₁₀₀ N ₂₀ O ₂₅	C ₇₀ H ₉₄ N ₂₀ O ₂₂
inthraszentin B	C ₈₄ H ₁₂₁ N ₂₁ O ₂₄	C ₇₁ H ₁₀₃ N ₂₁ O ₂₆	C ₇₁ H ₉₉ N ₂₁ O ₂₄
inthraszentin C	C ₈₃ H ₁₁₇ N ₂₁ O ₂₃	C ₇₀ H ₁₀₁ N ₂₁ O ₂₆	C ₇₀ H ₉₅ N ₂₁ O ₂₃

This disparity could not be explained by just considering a fatty acid being incorporated by the C_{starter}-domain. The added sum formulae of all amino acids, that are incorporated, contained more oxygen atoms than the experimentally determined sum formulae. This

stirred the thought, that some of the amino acid residues could be dehydro amino acids, since some known NRPs contain dehydroalanine or dehydrobutyrine. Dehydroalanine was considered, because the odd oxygen atoms correlated with the number of serine residues in each peptide, three for intraszentins A and C and two for intraszentin B. Another disparity was observed when comparing the rings and double bond equivalents (rdbe) (Tab. 9).

Table 9. Comparison of rdbes of intraszentins A-C (1-3). $rdbe_{exp}$: Experimentally determined rdbes; $rdbe_{aa}$: rdbe of a theoretical peptide consisting of all amino acids that were shown to be incorporated; $rdbe_{aa+dha}$: rdbe of a theoretical peptide consisting of all amino acids that were shown to be incorporated where every serine residue was regarded as dehydro alanine residue.

Compound	$rdbe_{exp}$	$rdbe_{aa}$	$rdbe_{aa+dha}$
intraszentin A	36	31	34
intraszentin B	35	31	33
intraszentin C	36	31	34

The number of the rdbe determined through the sum formulae that were obtained by feeding experiments (36 for intraszentins A and C and 35 for intraszentin B) differs from the ones obtained by adding up the rdbe for all amino acids that were shown to be incorporated (31). However, after taking possible dehydroalanine residues into account, only two rdbe for each derivative had to be determined. One of those is accounted for by a carbonyl group of a fatty acid incorporated by the $C_{starter}$ domain. The second can be explained by an intramolecular esterification between the hydroxyl group of one of the threonine residues or the tyrosine residue and the carboxyl group of the C-terminal amino acid. To fit all the data together, a C_{13} saturated fatty acid has to be added. Since most fatty acids with an odd number of carbon atoms are iso-branched¹⁷⁶, isotridecanoic acid was the most likely candidate. Iso-branched fatty acids are also incorporated into other NRPs in *Xenorhabdus* sp.^{29,76,82}.

After determination of the sum formulae and the likely building blocks, the amino acid sequence and stereochemistry were the last pieces missing in order to formulate educated structure proposals for intraszentins A to C. Both of which were derived from bioinformatic data provided through antiSMASH analyses.

I predicted the configuration of each amino acid by assessing the C-C/E-domain pattern. Considering the epimerization part of each one of the C/E domains is active, the

configuration directly correlates with the C-C/E-domain pattern. This approach leads to the pattern of D-L-D-D-L-D-D-L-D-D-D-D-D for intraszentins A-C (Fig. 13).

Since intraszentins A-C are similar, intraszentin B (2) was chosen for the assignment of the amino acid sequence (Fig. 19). Subsequently, the sequences of intraszentins A (1) and C (3) were derived from that by comparing the similarity of the predicted A domain specificities.

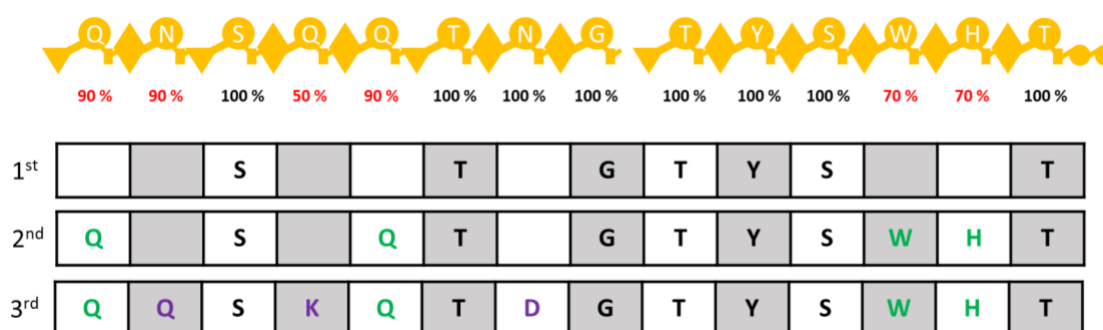


Figure 19. Deduction of the amino acid sequence of intraszentin A (1) by combination of predicted A domain specificities and results from feeding experiments.

First, all amino acids that matched a Stachelhaus code identity of 100 % in the antiSMASH analysis were assigned to their respective domains (Fig. 19). Except for the prediction of asparagine at A₇, I was able to find a match for every A domain with a Stachelhaus code identity of 100 %. Asparagine was not shown to be incorporated (Fig. 18). In a second round, I assigned amino acids shown to be incorporated to their respective A domains with a Stachelhaus code identity of 70 % or higher, if the prediction matched the respective amino acids. Those assignments were glutamine to A₁ and A₅, tryptophan to A₁₂ and histidine to A₁₃ (Fig. 19). This left one glutamine, one aspartate, one lysine residue to be assigned. By comparison of the Stachelhaus code identities and the other prediction tools provided by antiSMASH, I assigned glutamate to A₂, lysine to A₄ and aspartate to A₇ (Fig 19).

This assignment also matched perfectly with the differences between the predictions (Fig. 13) and the amino acid feeding experiments (Fig. 18 and Fig. S5) for intraszentins B and C. In comparison with intraszentin B, antiSMASH predicted the incorporation of threonine

instead of serine at A₆ for both intraszentin A and C. Amino acid feeding experiments suggested that intraszentin B has two threonine and three serine residues, while intraszentin C has three threonine and two serine residues (Fig. 18 and Fig. S5). Thus, the differences between intraszentin B and C were solved. For intraszentin A the only other difference to intraszentins B and C was the prediction of proline in contrast to glutamate at A₁. Since the amino acid feeding experiments hinted at the incorporation of only two glutamate residues for intraszentin A, the amino acid sequences of intraszentins A to C were assigned by combining bioinformatic analysis with labelling approaches analysed *via* LC-MS. The combination of all results from this chapter led to the preliminary structure proposals displayed in figure 20.

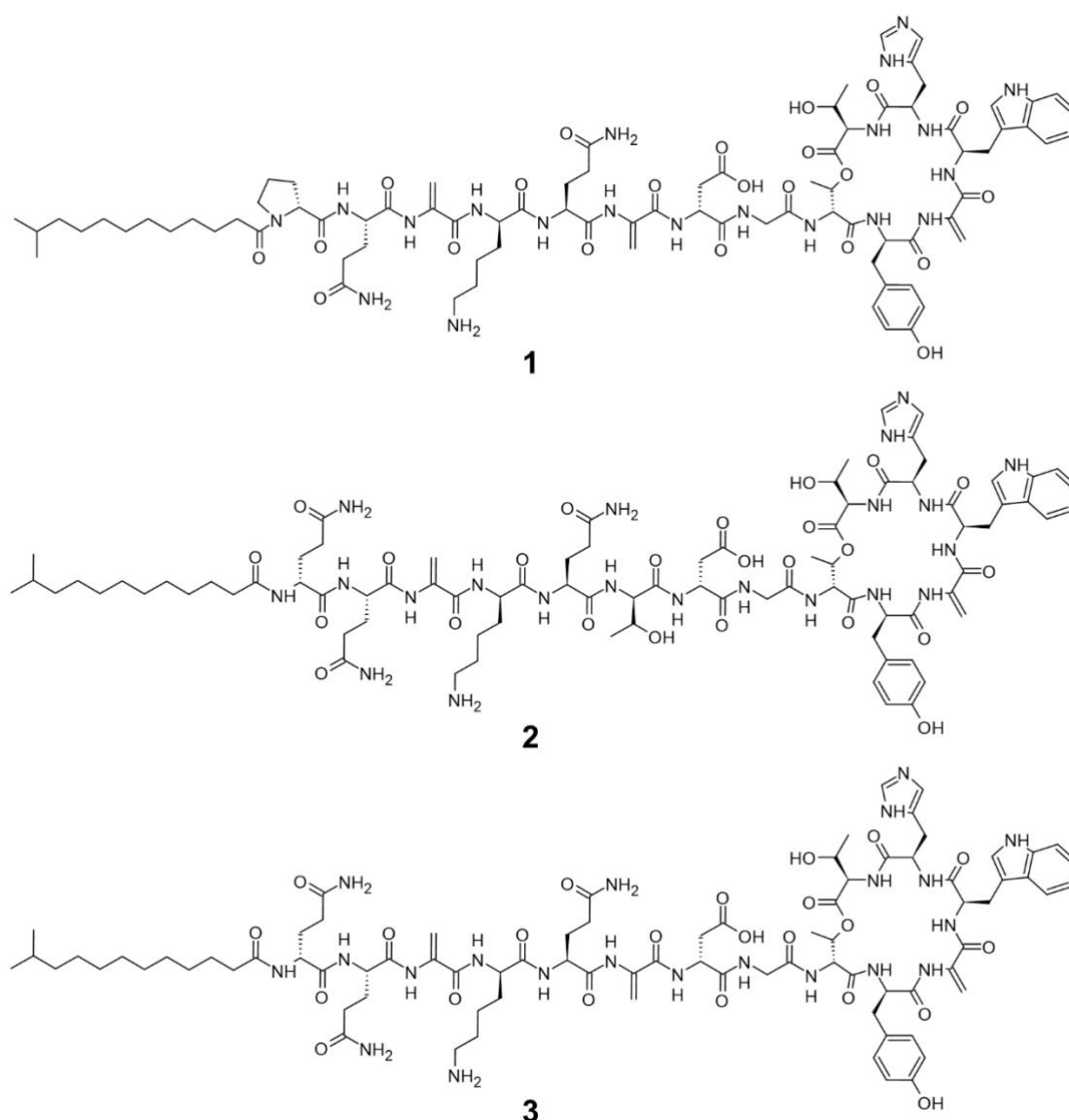


Figure 20. Proposed structures for intraszentins A (1), B (2) and (3).

In theory three residues, that are the same in all three peptides, are able to form an ester bond with the carboxyl group at the C terminus of the peptide, the threonine residue at position 9, the tyrosine residue at position 10 and the threonine residue at position 14. The threonine residue at position 14 was considered too close to the C terminus, leaving only two options. Of those two, only the fragment with m/z 758.33 $[M+H]^+$ representing the y_6 fragment containing only the ring between the C terminus and the threonine at position 9 could be detected (Fig. 21).

Since I only predicted the structures and had no proof, if my model actually works, I tried to confirm my structure predictions by analysing the MS² fragmentation patterns. I was only able to record high quality MS² spectra for intraszentin B (**2**). The fragmentation analysis showed that, in contrast to the previous amino acid feeding experiment, asparagine instead of aspartate is incorporated by the A₇ domain (Fig. 21).

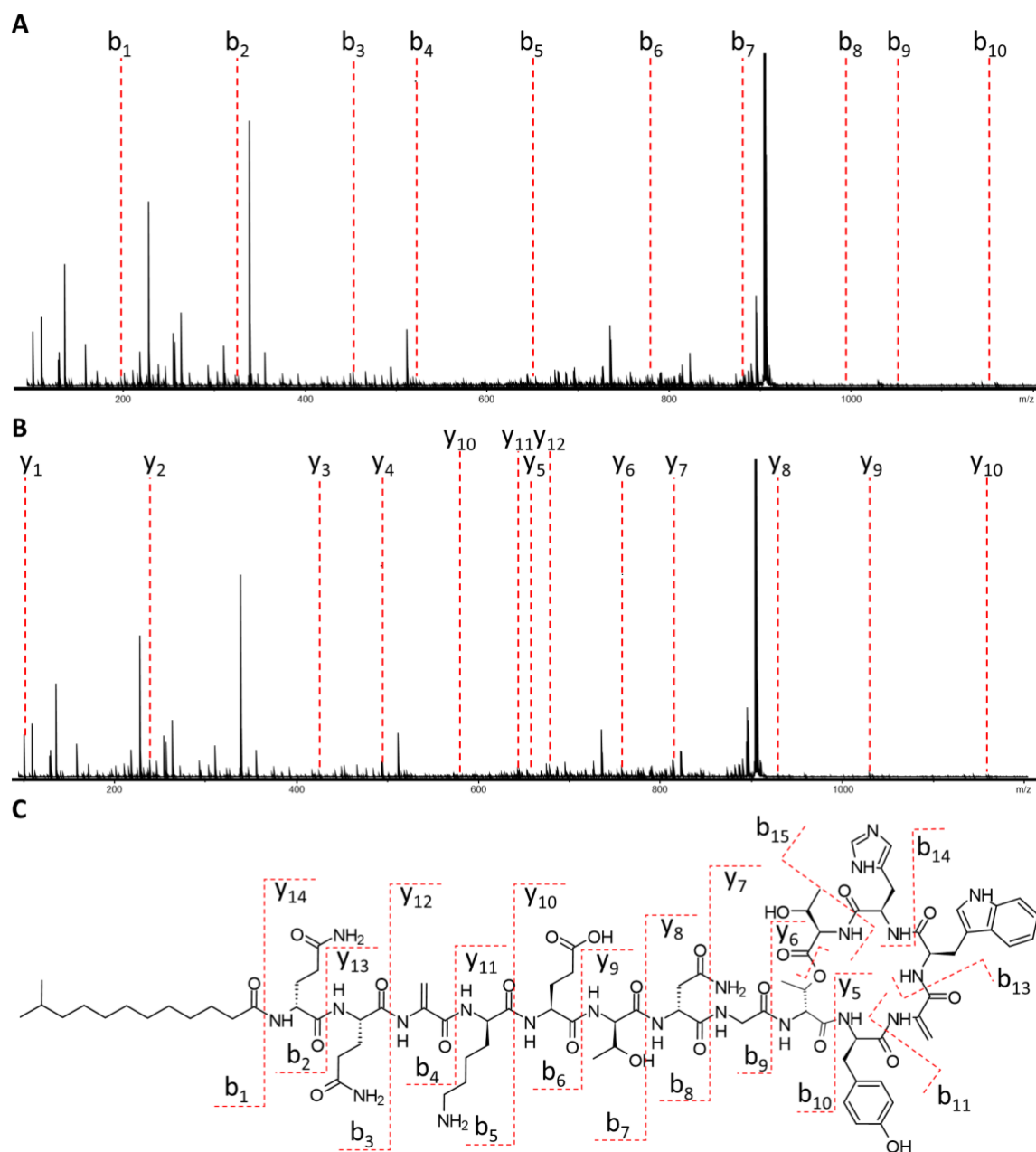


Figure 21. Assignment of MS² fragments to the structure of intraszentin B (**2**). **A**: b-fragments; **B**: y-fragments; **C**: Structure of intraszentin B (**2**). All theoretical b- and y-fragments were assigned. The corresponding y-fragments of b₁₅, b₁₄, b₁₃ and b₁₁ are y₁, y₂, y₃ and y₄, respectively.

Despite the conflicting results, the difference can be explained. Amino acids are a key part of the primary metabolism. If the amino acid homeostasis is altered by adding one amino acid in excess, as I did in the amino acid feeding experiment, it can be used by the organism to produce energy or can easily be converted into another amino acid. In the case of aspartate/asparagine and glutamate/glutamine, only one reaction is necessary to convert one into the other, which can lead to the incorporation of asparagine, that was initially fed as

3.2 Protective sphinganes from *Pseudomonas* MYb115

3.2.1 Bioactivity of MYb115

My Collaborator Dr. Kohar A. Kissoyan from the Christian-Albrechts-Universität zu Kiel observed that three natural microbial isolates from *Caenorhabditis elegans* protect their host from harming effects that are typically caused by nematopathogenic *Bacillus thuringiensis* strains¹⁶⁴. Those three isolates are *Pseudomonas* strains MYb11, MYb12 and MYb115. While the principle of protection caused by MYb11 and MYb12 is described in chapter 3.3 of this work, this chapter deals with the effect caused by MYb115. While MYb11 and MYb12 kill the pathogenic *B. thuringiensis* strain, MYb115 protects the host nematode without killing the pathogen¹⁶⁴. In this chapter, I describe the measures I took in order to identify the causing agent for this effect.

3.2.2 Gene cluster identification

First, I identified all BGCs that may be responsible for natural product production using antiSMASH¹⁷⁷. Ten BGCs were identified (data not shown). A promoter exchange in front of the type I PKS BGC revealed that this cluster is essential for the bioactivity of MYb115.

3.2.2.1 Promoter exchange based gene cluster activation

The promoter exchange in front of the *pks* gene in antiSMASH cluster 25.1 led to the identification of compounds **5**, **6** and **7** through LC-MS analysis (Fig. 23).

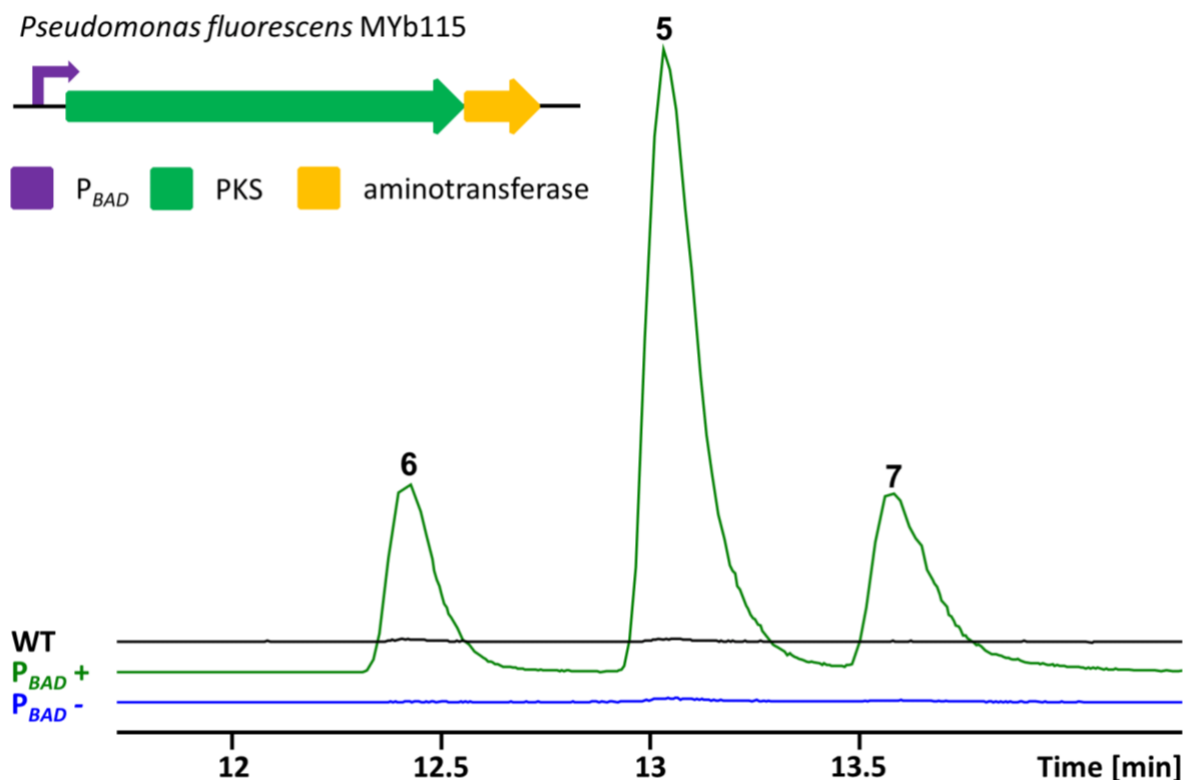


Figure 23. Schematic illustration of the promoter exchange in front of the *pks* gene and UPLC-MS base peak chromatograms of the MYb115 wild type (black) and promoter exchange mutant without arabinose (blue) and with arabinose for induction (green). Production levels of compounds **5**, **6** and **7** are greatly increased in the induced promoter exchange compared to the wild type.

HRMS analysis revealed that the masses of compounds **5** (m/z 414.4292 $[M+H]^+$, RT 13.0 min), **6** (m/z 386.3982 $[M+H]^+$, RT 12.4 min) and **7** (m/z 442.4608 $[M+H]^+$, RT 13.6 min) only differ by the mass of an ethyl group (28.03 Da), while the RT shifts by 0.6 min from **6** to **5** and from **5** to **7** (Fig. 23). The production drastically increased in the induced promoter exchange strain compared to the wildtype.

3.2.2.2 Bioactivity of promoter exchange mutants

To prove that the bioactivity of MYb115 inside the host organism is due to production of compounds **5**, **6** and **7**, my collaborator Lena Peters from the Christian-Albrechts-Universität zu Kiel performed a survival assay with *C. elegans* nematodes that were exposed to a single bacterial strain, mixed with either pathogenic or non-pathogenic *B. thuringiensis* (Fig. 24).

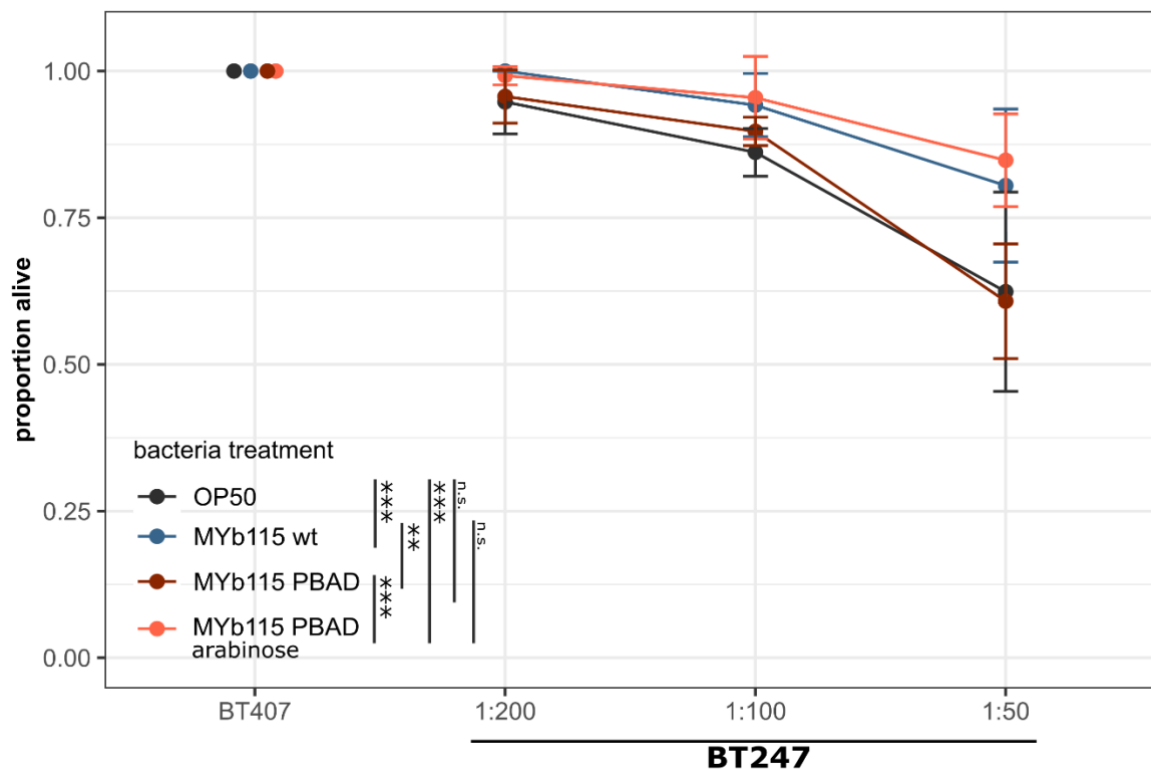


Figure 24. Survival of *C. elegans* N2 on different concentrations of pathogenic BT247 mixed with either *E. coli* OP50 (black), MYb115 wild type (blue), MYb115 promoter exchange without arabinose (red) and MYb115 promoter exchange with arabinose (orange). Non-pathogenic BT407 served as control. This experiment was conducted and the graphic was created by Lena Peters.

The nematodes were treated with *E. coli* OP50 as negative control, MYb115 wild type as positive control, as well as MYb115 P_{BAD} *pks* with and without addition of arabinose. Different dilutions of pathogenic *B. thuringiensis* BT247 as well as the non-pathogenic control BT407 were added. On one side, the nematodes treated with MYb115 P_{BAD} *pks* without the addition of arabinose showed the same survival rate as the nematodes treated with OP50. On the other side the nematodes that were treated with MYb115 P_{BAD} *pks* with the addition of arabinose restored the effect that Kissoyan *et al.*¹⁶⁴ observed from the MYb115 wild type, strongly suggesting that the protective effect is indeed triggered by the induction of the type I PKS BGC. In following experiments, the structures of NPs **5**, **6** and **7** were investigated.

3.2.3 Structure elucidation via MS

3.2.3.1 Isotopic labelling

First, the sum formulae of compounds **5**, **6** and **7** were determined. Therefore, MYb115 *P_{BAD} pks* was cultivated in ¹³C-ISGRO™ and ¹⁵N-ISOGRO™ media as well as regular LB. Arabinose was added for induction. A separate culture in LB without arabinose was used as negative control. Methanolic extracts were taken from the cell pellet. The extracts were then analysed via LC-MS (Fig. 25).

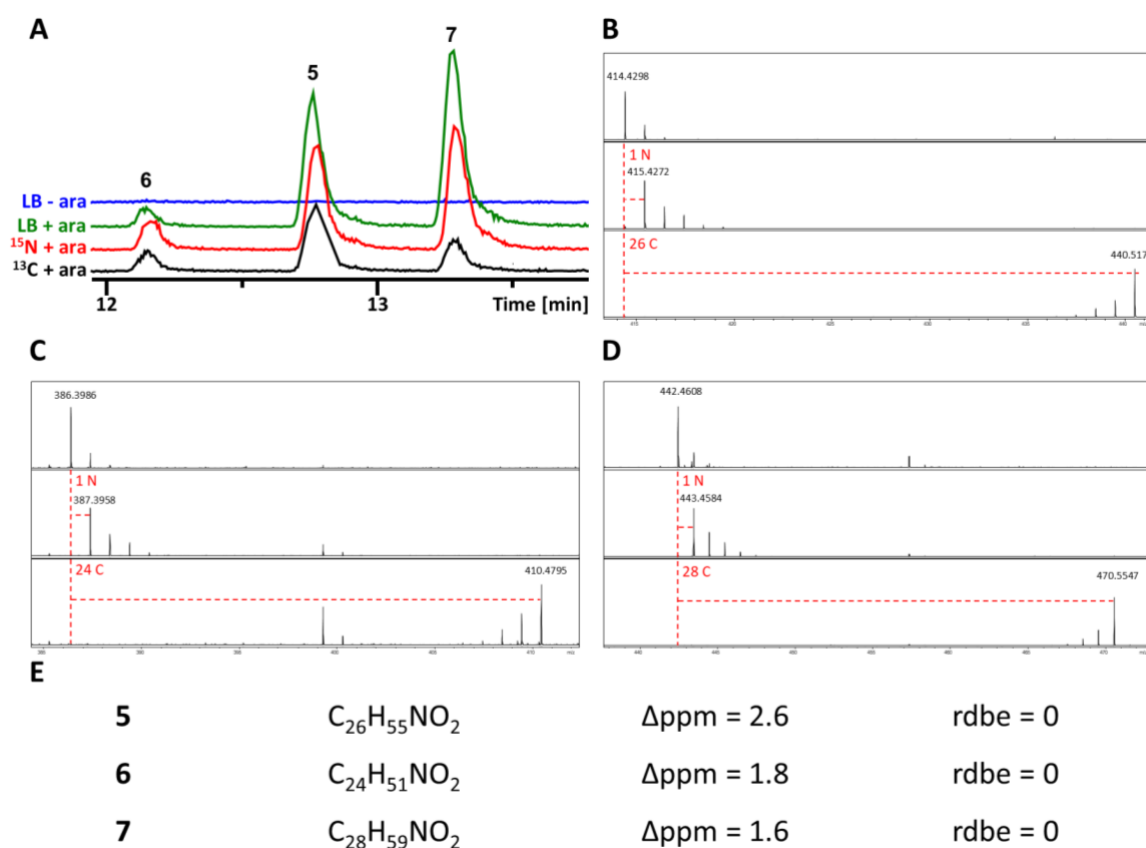


Figure 25. Sum formula determination of compounds **5**, **6** and **7** using isotopic labelling and LC-MS. **A:** EICs of compounds **5**, **6** and **7** in unlabelled samples with (green) and without arabinose (blue), as well as ¹⁵N- (red) and ¹³C-labelled (black) samples. **B-D:** Mass shifts compared to LB cultivation (dashed red lines) represent the number of carbon and nitrogen atoms incorporated. **E:** Sum formulae of compounds **5**, **6** and **7**, as well as Δppm values and calculated rdbes.

There was no production of compounds **5** to **7** in the negative control, whereas **5**, **6** and **7** were produced in all three samples where arabinose was added (Fig. 25 A). **5** consists of one nitrogen and 26 carbon atoms (Fig. 25 B). In combination with the HR mass, the sum formula C₂₆H₅₅NO₂ was derived for compound **5**. The calculated and the measured *m/z* of this

molecule only differ by 2.6 ppm (Fig. 25 E). **6** consists of one nitrogen and 24 carbon atoms (Fig. 25 C). In combination with the HR mass, the sum formula $C_{24}H_{51}NO_2$ was derived for compound **6**. The calculated and the measured m/z of this molecule only differ by 1.8 ppm (Fig. 25 E). **7** consists of one nitrogen and 28 carbon atoms (Fig. 25 D). In combination with the HR mass, the sum formula $C_{28}H_{59}NO_2$ was derived for compound **7**. The calculated and the measured m/z of this molecule only differ by 1.6 ppm (Fig. 25 E). Hence, compounds **5-7** were shown to contain zero rings or double bonds (Fig. 25 E).

The sum formulae determined through this experiment, combined with interpretation of the BGC structure, hinted at compounds **5**, **6** and **7** consisting of a long acyl chain with a head group containing nitrogen and oxygen. The NPs were compared to known structures using LC-MS/MS in the following step.

3.2.3.2 Comparison with known structures

By searching literature for structures with similar properties to the ones elucidated in the previous step, I found that NPs **5**, **6** and **7** have many parallels to sphingosines and even more so fully saturated sphinganine. After comparing the fragmentation pattern of compounds **5** to **7** to published sphingosine fragmentation patterns by Li *et al.* (2017)¹⁷⁸, I compared the MS² data of compound **5**, **6** and **7**, to those of C₁₈- and C₂₀-sphinganine (Fig. 26). Sphinganine with C₂₄, C₂₆ and C₂₈ acyl chains were not commercially available.

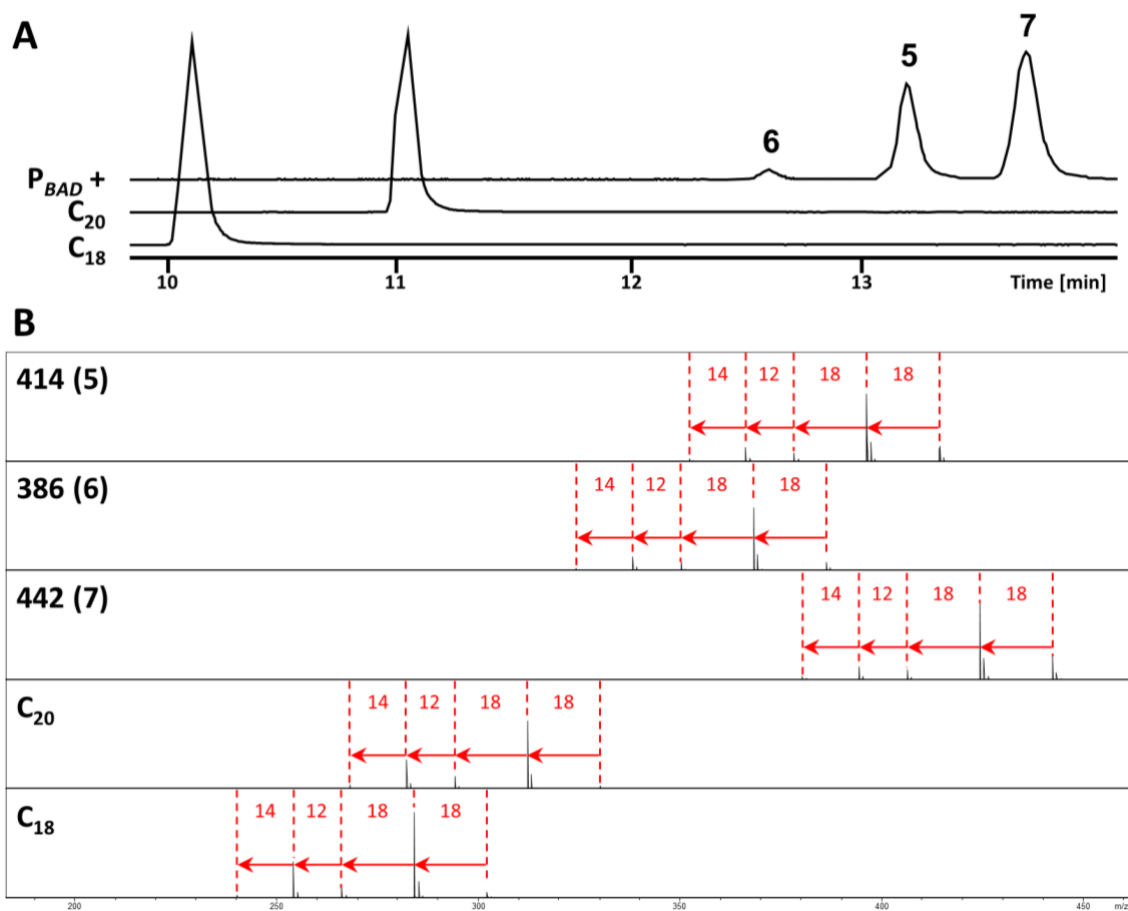


Figure 26. Comparison of natural products **5**, **6** and **7** and commercially available sphinganine standards. **A:** EICs of compounds **5**, **6** and **7**, as well as C₁₈- and C₂₀-sphinganine standards. **B:** Fragmentation patterns of **5**, **6** and **7**, as well as C₁₈- and C₂₀-sphinganine standards.

Sphinganine C₁₈ (m/z 302.3048 [M+H]⁺) and C₂₀ (m/z 330.3357 [M+H]⁺) eluted after 10.1 min and 11.1 min, respectively (Fig. 26 A). The MS² fragmentation ion pattern of the two standards matched to those of **5**, **6** and **7**, supporting the notion that those are long-chain sphinganine standards (Fig. 26 B). The fragment ions were checked for their most likely molecular composition and rdbes, resulting in a fragmentation tree (Fig. S 6).

3.2.3.3 Serine feeding experiments

The common route for sphinganine biosynthesis includes the condensation of serine with a CoA-bound fatty acid¹⁷⁹.

However, simple feeding experiments with serine are not suited to investigate this matter, because it can easily be converted to other building blocks. If serine is added in excess to a living cell, it will be degraded to acetyl-CoA in two reactions. Acetyl-CoA is then inserted into the TCA cycle and used for every biosynthetic pathway, thus, diluting the signal of labelled compounds. Hence, an experimental setup was needed, in which serine molecules, that contain their original amino moieties, are distinguishable from acetyl-CoA derived building blocks

Therefore, MYb115 $P_{BAD}pk_s$ cultures in synthetic XYP medium lacking serine were fed with either $^{13}\text{C}_3^{15}\text{N}$ -labelled serine or with regular serine displaying the usual isotopic abundances, mostly consisting of $^{12}\text{C}_2^{14}\text{N}$ -serine. This should result in the production of two isotopologues of **5**. With addition of $^{13}\text{C}_3^{15}\text{N}$ -labelled serine, the isotopologue that is $m_{\text{monoisotopic}}+3$ should be labelled with two ^{13}C isotopes and one ^{15}N isotope, since one carbon atom is lost through the elimination of CO_2 during the condensation. In XYP plus regular serine, the isotopologue that is $m_{\text{monoisotopic}}+3$ should be labelled with three ^{13}C isotopes because of the higher natural abundance of ^{13}C compared (1.1 %) to ^{15}N (0.4 %). The two isotopologues, $^{13}\text{C}_3$ and $^{13}\text{C}_2^{15}\text{N}$, are distinguishable by their respective masses.

After extraction, the exact masses of the distinctly labelled isotopologues of compound **5** were then detected and identified using HRMS (m_{exp}). In addition, the theoretical masses of each isotopologues were calculated (m_{theo}). Comparison of m_{exp} from each sample with m_{theo} of both isotopologues enabled the calculation of Δppm values. The lower Δppm value helped identify the isotopologue that was detected *via* HRMS. Table 10 shows the comparison of the theoretical masses of both isotopologues with the experimentally determined mass of the isotopologue of **5** detected in the sample to which $^{13}\text{C}_2^{15}\text{N}$ -labelled serine was added. The comparison with the experimental data of the sample from the culture with $^{12}\text{C}_2^{14}\text{N}$ -serine is displayed in Table 11.

Table 10. Comparison of theoretical masses (m_{theo}) of two distinct isotopologes of **5** with the experimentally determined mass (m_{exp}). During the experiment $^{13}\text{C}_3^{15}\text{N}$ -labelled serine was fed.

	$^{13}\text{C}_2^{15}\text{N}$	$^{13}\text{C}_3$
m_{theo}	417.4348	417.4411
m_{exp}	417.4370	
Δppm	5.1984	-9.9176

The data for the sample where $^{13}\text{C}_3^{15}\text{N}$ -labelled serine was added, displayed in table 10, show a Δppm value of 5.2 for $^{13}\text{C}_2^{15}\text{N}$ -labelled compound **5** compared to -9.9 for the $^{13}\text{C}_3$ -labelled isotopologue. Showing that $^{13}\text{C}_3^{15}\text{N}$ -labelled serine is incorporated and proving that serine is used once for the biosynthesis of sphinganine-like compounds **5**, **6** and **7**. The control, where regular serine was added to the culture, shows reverse results with a Δppm of 10.2 the $^{13}\text{C}_2^{15}\text{N}$ -labelled isotopologue compared to -4.9 for the $^{13}\text{C}_3$ -labelled version (Tab. 11).

Table 11. Comparison of theoretical masses (m_{theo}) of two distinct isotopologes of **5** with the experimentally determined mass (m_{exp}). During the experiment unlabelled serine was fed.

	$^{13}\text{C}_2^{15}\text{N}$	$^{13}\text{C}_3$
m_{theo}	417.4348	417.4411
m_{exp}	417.4391	
Δppm	10.2291	-4.8869

3.2.4 Lipidomics

To assess if the herein described sphinganine species exist as free compounds or are part of i.e. lipids, I performed lipidomics experiments in collaboration with Dr. Georgia Angelidou and Dr. Nicole Paczia from the Metabolomics core facility at the MPI for terrestrial microbiology in Marburg. I designed the experiment together with Dr. Nicole Paczia and I prepared the samples, Dr. Georgia Angelidou and Dr. Nicole Paczia performed the measurements, Dr. Georgia Angelidou analyzed the data and Dr. Georgia Angelidou and I evaluated the data.

In addition to compounds **5** to **7**, we identified compound **8** (m/z 358.36868 $[M+H]^+$, RT 7.058 min) with a similar fragmentation pattern in P_{BADpks} upon induction, a shorter sphinganine with the molecular formula $C_{22}H_{47}NO_2$ (Figure S 7).

Furthermore, we identified compounds that displayed fragments corresponding to ions of **5**, **6**, **7** and **8**. Those masses were 154 Da heavier than those of the sphinganine NPs (**9**: m/z 568.43420 $[M+H]^+$, RT 8.45 min; **10**: m/z 540.40295 $[M+H]^+$, RT 6.73 min; **11**: m/z 596.46564 $[M+H]^+$, RT 10.234 min; **12**: m/z 512.37189 $[M+H]^+$, RT 4,675 min) (Tab. 12). The masses of **9** to **12** did not correspond to any known lipid species contained in the data base (Georgia Angelidou, personal communication). Hence, we used the exact mass and tried different lipid headgroups to come up with the structure proposals for compounds **9** to **12** (Tab. 12). We concluded that compounds **9** to **12** are most likely phosphoglycerolsphingolipids. The identification of the corresponding fragment ion of **9** lacking the phosphoglycerol (PG) head group supported this conclusion (Figure S 8). Additionally, the neutral loss of 154 matches with the mass of the PG head group.

Table 12. Properties and structure proposals of phosphoglycerol sphingolipids **9-12**.

#	n	sum formula $[M+H]^+$	RT [min]	m/z_{theo}	m/z_{exp}	Δppm
9	22	$C_{29}H_{63}NO_7P$	8.45	568.43367	568.43420	0.93
10	20	$C_{27}H_{59}NO_7P$	6.72	540.40237	540.40295	1.07
11	24	$C_{31}H_{67}NO_7P$	10.23	596.46497	596.46564	1.12
12	18	$C_{25}H_{55}NO_7P$	4.68	512.37107	512.37189	1.60

Moreover, we analysed the relative quantity of sphinganines and PG-sphinganines between the tested conditions. Therefore, we compared the relative abundance of sphinganines **5**, **6**, **7** and **8**, along with C_{18} (m/z 302.30536 $[M+H]^+$, RT 2.184 min) and C_{20} (m/z 330.33666

[M+H]⁺, RT 3.341 min) sphinganine, that were found in wildtype samples, between the four conditions (Fig. S 13). C₁₈ and C₂₀ sphinganine were only detected in wild type samples (Fig. S 13 A-B).

Sphinganine **5**, **6**, **7** and **8** were far more abundant in induced *P_{BAD}pks* samples, compared to wild type samples (Fig. S 13 C-F). Nevertheless, the abundancies of PG-sphingolipids **9** and **10** were similar between wild type samples and the samples from induced *P_{BAD}pks* (Fig. S 14). This indicates that increase in sphinganine production does not lead to increase in PG-sphingolipid production.

3.2.5 Deletion mutants

To further analyse the biological role of compounds **5** to **7**, deletion mutants were constructed as described (2.5.5). This yielded three strains with different genotypes: One strain in which only the *pks* gene (*sgaA*) was deleted, one strain in which only the *amt* gene (*sgaB*) was deleted and one strain in which the whole BGC (*sgaAB*) was deleted.

3.2.5.1 Sphinganine production in deletion mutants

All three deletion strains were checked for sphinganine production (Fig. 27). The deletion mutants were compared to the wild type strain.

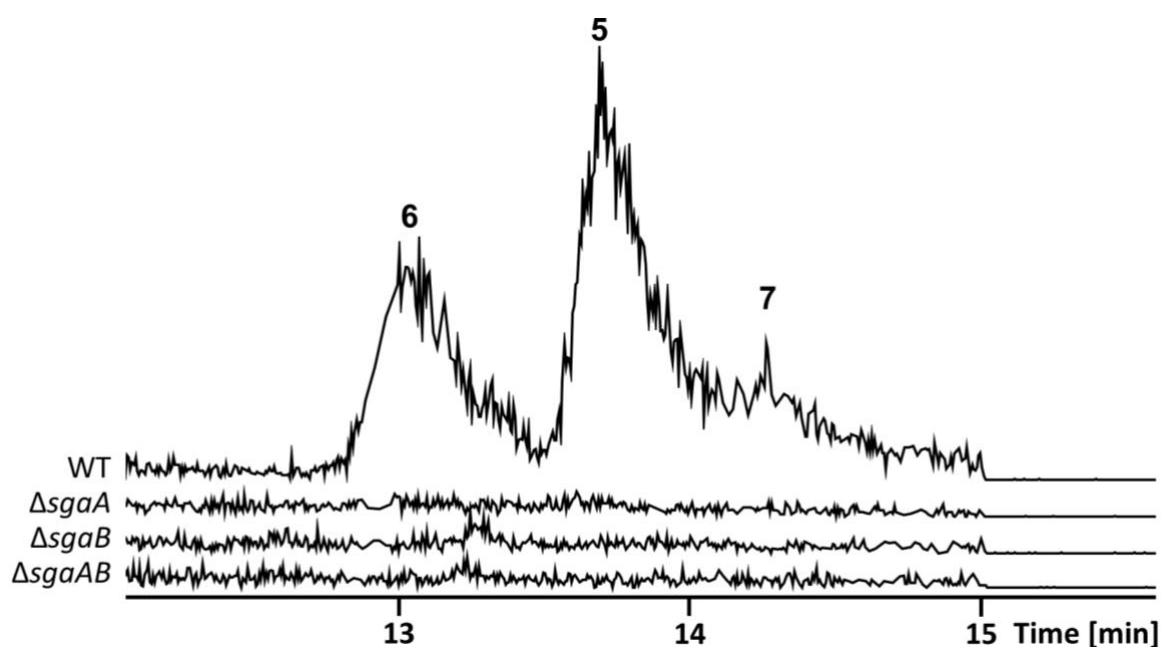


Figure 27. Sphinganine production was abolished in all three mutant strains. Combined EIC of **5**, **6** and **7** is shown.

No sphinganine production was detected in $\Delta sgaA$, $\Delta sgaB$ and $\Delta sgaAB$ (Fig. 27). Compounds **5-7** could only be detected in the wild type but not in the deletion mutants. Hence, the deletion was successful. To gain insights into the sphinganine biosynthesis in MYb115, I compared the BPCs of the samples of the mutant strains $\Delta sgaA$ and $\Delta sgaB$ to the BPCs of the wild type sample and $P_{BAD}pks$ to find possible shunt products of the disrupted biosynthesis, but no candidates could be identified (data not shown). The slight shift in RT of compounds **5-7** in comparison to earlier measurements can be attributed to deterioration of the precolumn in the LC system.

3.2.5.2 Bioactivity of deletion mutants

I sent the deletion mutants $\Delta sgaA$, $\Delta sgaB$ and $\Delta sgaAB$ to my collaborator Lena Peters for bioactivity testing. She performed a survival assay according to a standardised protocol (DOI 10.1186/s12864-016-2603-8). In this assay, she compared the survival of *C. elegans* nematodes infected with *B. thuringiensis* BT247 and either *E. coli* OP50, MYb115 wild type or one of the three deletion mutants, $\Delta sgaA$, $\Delta sgaB$ and $\Delta sgaAB$. Non-pathogenic *B. thuringiensis* BT407 served as a control. The survivability of nematodes infected with each of the deletion mutant strains was significantly lower than the survivability of the nematodes infected with MYb115 wild type (Fig. 28).

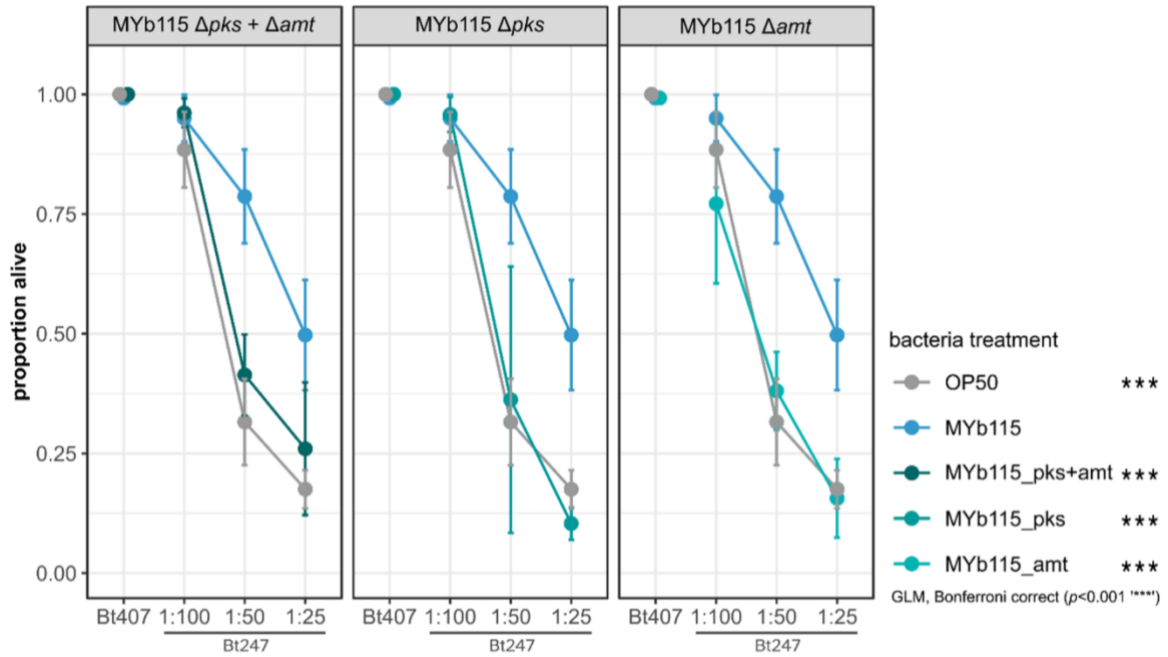


Figure 28. Survival of *C. elegans* N2 on different concentrations of pathogenic BT247 mixed with either *E. coli* OP50 (grey), MYb115 wild type (light blue) or one of the three deletion mutant strains, $\Delta pks+amt$ (dark green), Δpks (green) or Δamt (light green) Non-pathogenic BT407 served as control. This experiment was conducted and the graphic was created by Lena Peters.

The results prove that the natural protection caused by MYb115 is disrupted by deletion of either gene of the sphingosine BGC or the whole cluster. Hence, sphinganine 5, 6 and 7 play a key role in the protection mechanism, which we have yet to elucidate.

3.3 Antimicrobial peptides from *Pseudomonas* MYb11 and MYb12

Because of the aforementioned bioactivity shown by *Pseudomonas* isolates MYb11 and MYb12¹⁶⁴, I tried to identify the antibiotic natural products produced by those strains. Parts of this work were already published¹⁶⁴.

3.3.1 Identification of specific natural products from MYb11 and MYb12

LC-MS analysis of methanolic XAD extracts of all bacterial isolates that were identified by Dr. Kohar A. Kissoyan¹⁶⁴, led to the identification of compounds **13** (m/z 1126.68 [M+H]⁺, RT 12.1 min) and **14** (m/z 1112.67 [M+H]⁺, RT 11.8 min) that are only produced by the strains that show antibiotic activity against *B. thurigiensis*, MYb11 and MYb12. Library search using GNPS revealed **13** as a member of the viscosin family.

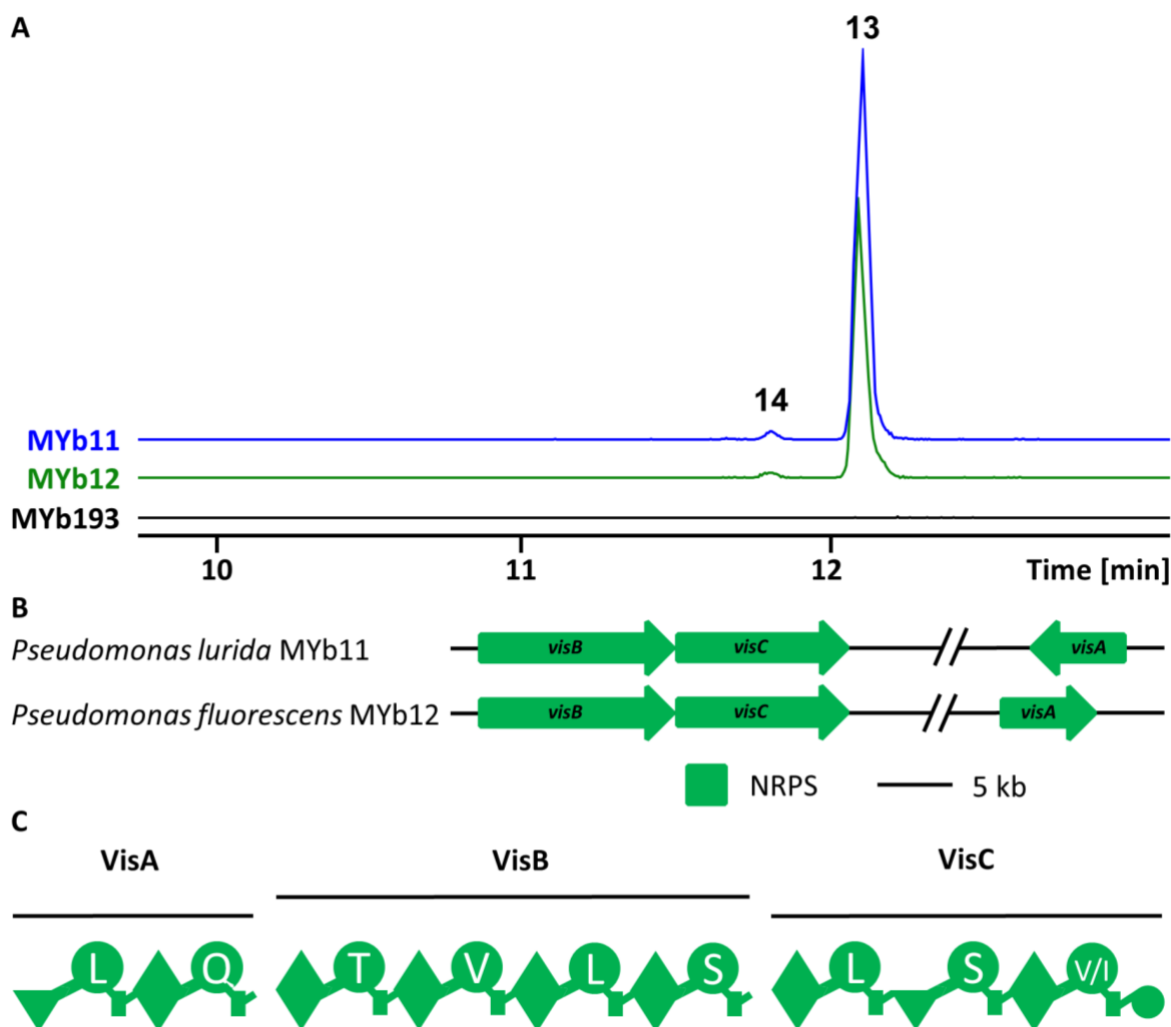


Figure 29. **13** and **14** are only produced by MYb11 and MYb12 but not by MYb193. **A:** Production of **13** and **14** in MYb11 and MYb12. No production was observed in MYb193. Combined EIC of **13** and **14** is shown. **B:** Viscosin BGC in MYb11 and MYb12. **C:** NRPS domain organisation of VisABC with A domain specificities.

NRPs of the viscosin family have shown antibiotic properties against Gram-positive bacteria in previous studies¹⁸⁰. AntiSMASH analyses of the genome sequences of MYb11 and MYb12 showed the presence of the viscosin BGC (Fig. 29 C). As described in previous studies, the viscosin BGC is typically encoded at two different loci in the genome¹⁸¹. This is also the case for MYb11 and MYb12 where *visA* is encoded at a different locus than *visBC*. In addition, the viscosin BGC is also present in the MYb193 genome sequence. MYb193 did not produce any NPs of the viscosin NPF under the tested conditions (Fig. 29) nor did it display antibiotic activity against *B. thuringiensis* in a disc diffusion assay¹⁶⁴.

3.3.2 Structure elucidation and confirmation via MS

I investigated the structure of **13** and **14** by using MS/MS sequence analysis and stable isotopic labelling. Comparative MS analysis of extracts from MYb11 cultured in LB, ¹³C ISOGROTM and ¹⁵N ISOGROTM revealed the sum formula of **13** as C₅₄H₉₅N₉O₁₆ and **14** as C₅₃H₉₃N₉O₁₆ (Fig. 30 A).

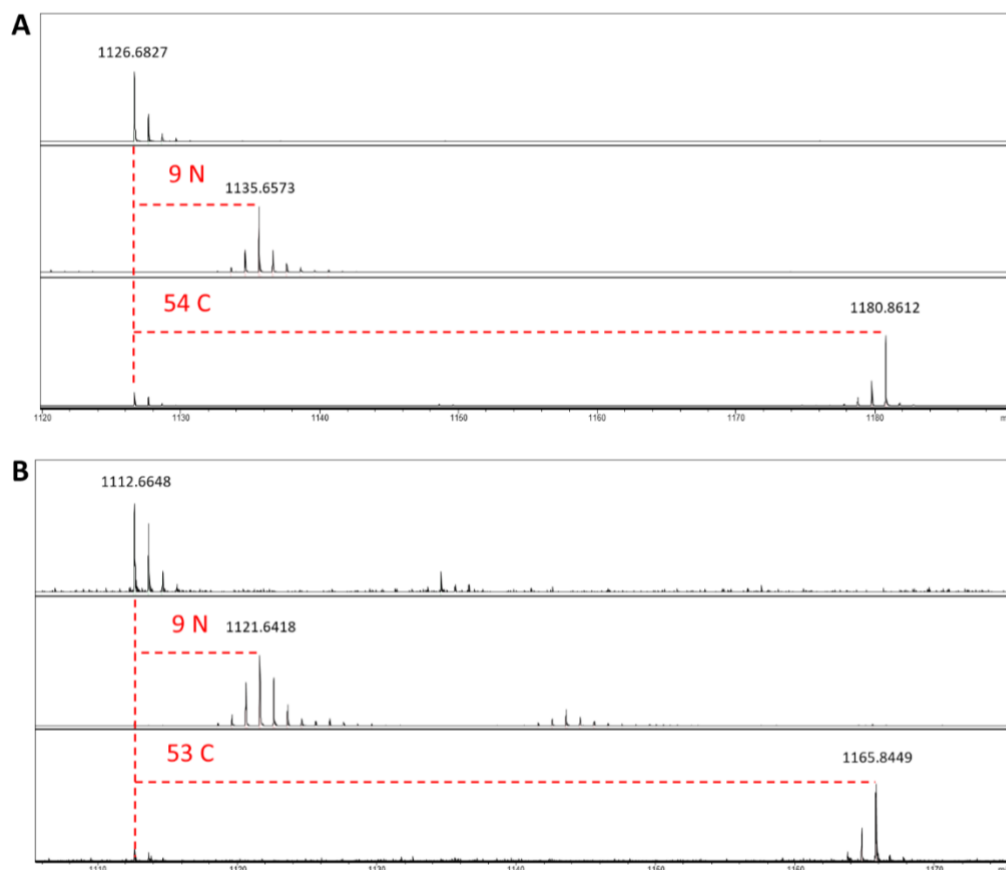


Figure 30. Sum formula determination of compounds **13** and **14** using isotopic labelling and LC-MS. **A:** Mass shifts compared to LB cultivation (dashed red lines) represent the number of carbon and nitrogen atoms incorporated in **13**. **B:** Mass shifts compared to LB cultivation (dashed red lines) represent the number of carbon and nitrogen atoms incorporated in **14**.

For **13**, cultivation in ^{13}C ISOGROTM resulted in a shift of 54 representing the incorporation of 54 carbon atoms, while cultivation in ^{15}N ISOGROTM resulted in a shift of 9 representing the incorporation of nine nitrogen atoms into the molecule (Fig.30 A). For **14**, cultivation in ^{13}C ISOGROTM resulted in a shift of 53 representing the incorporation of 53 carbon atoms, while cultivation in ^{15}N ISOGROTM resulted in a shift of 9 representing the incorporation of nine nitrogen atoms into the molecule (Fig. 30 B)

The amino acid sequence was confirmed as Leu/Ile-Glu-Thr-Val-Leu/Ile-Ser-Leu/Ile-Ser-Leu/Ile for **13** through analysis of the respective MS² fragmentation patterns (Fig. 31).

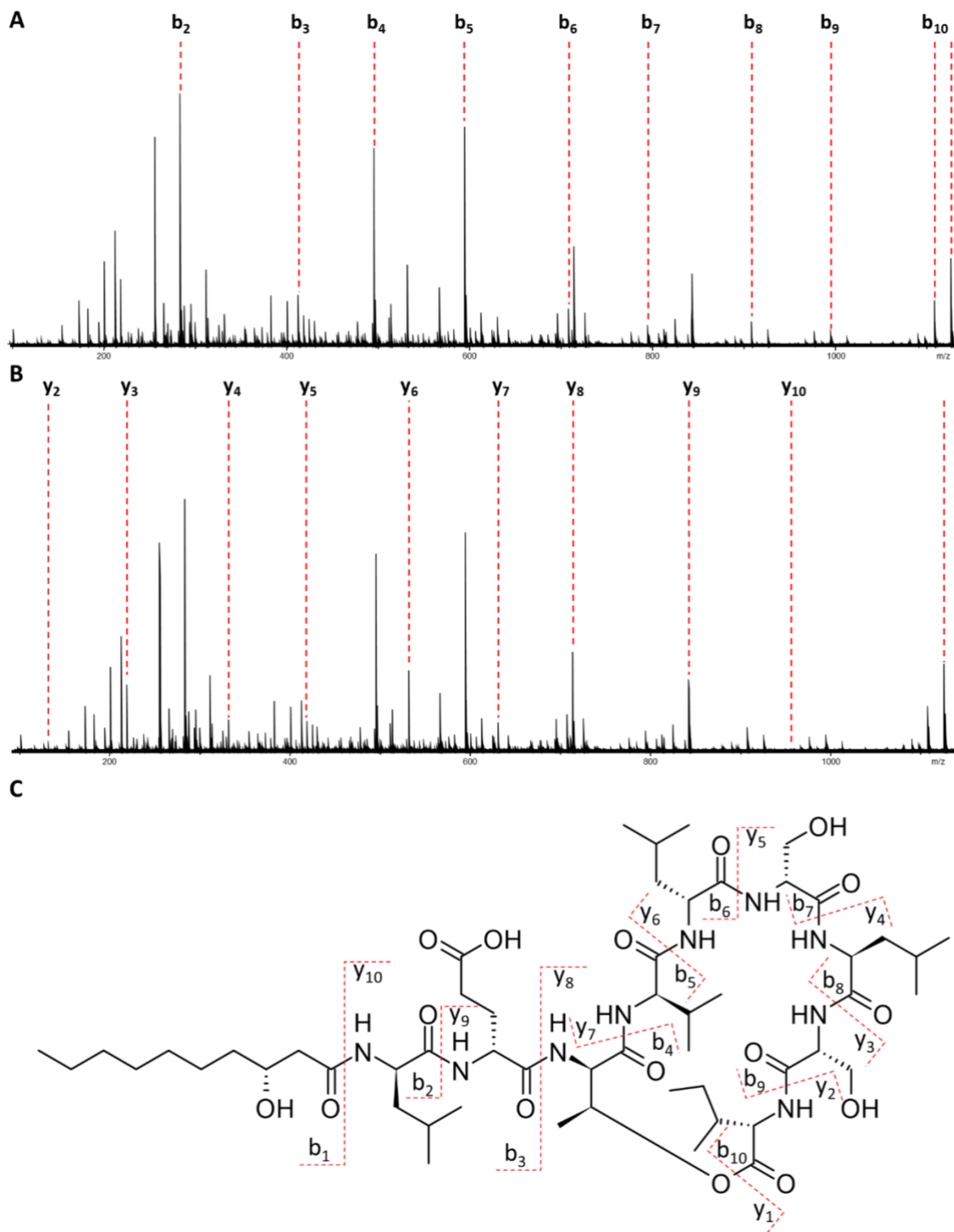


Figure 31. Assignment of MS² fragments to the structure of viscosin (**13**). **A:** b-fragments; **B:** y-fragments; **C:** Structure of viscosin (**13**). All theoretical b- and y-fragments were assigned.

The sequence of **14** was confirmed as Leu/Ile-Glu-Thr-Val-Leu/Ile-Ser-Leu/Ile-Ser-Val through analysis of the respective MS² fragmentation patterns (Fig. 32).

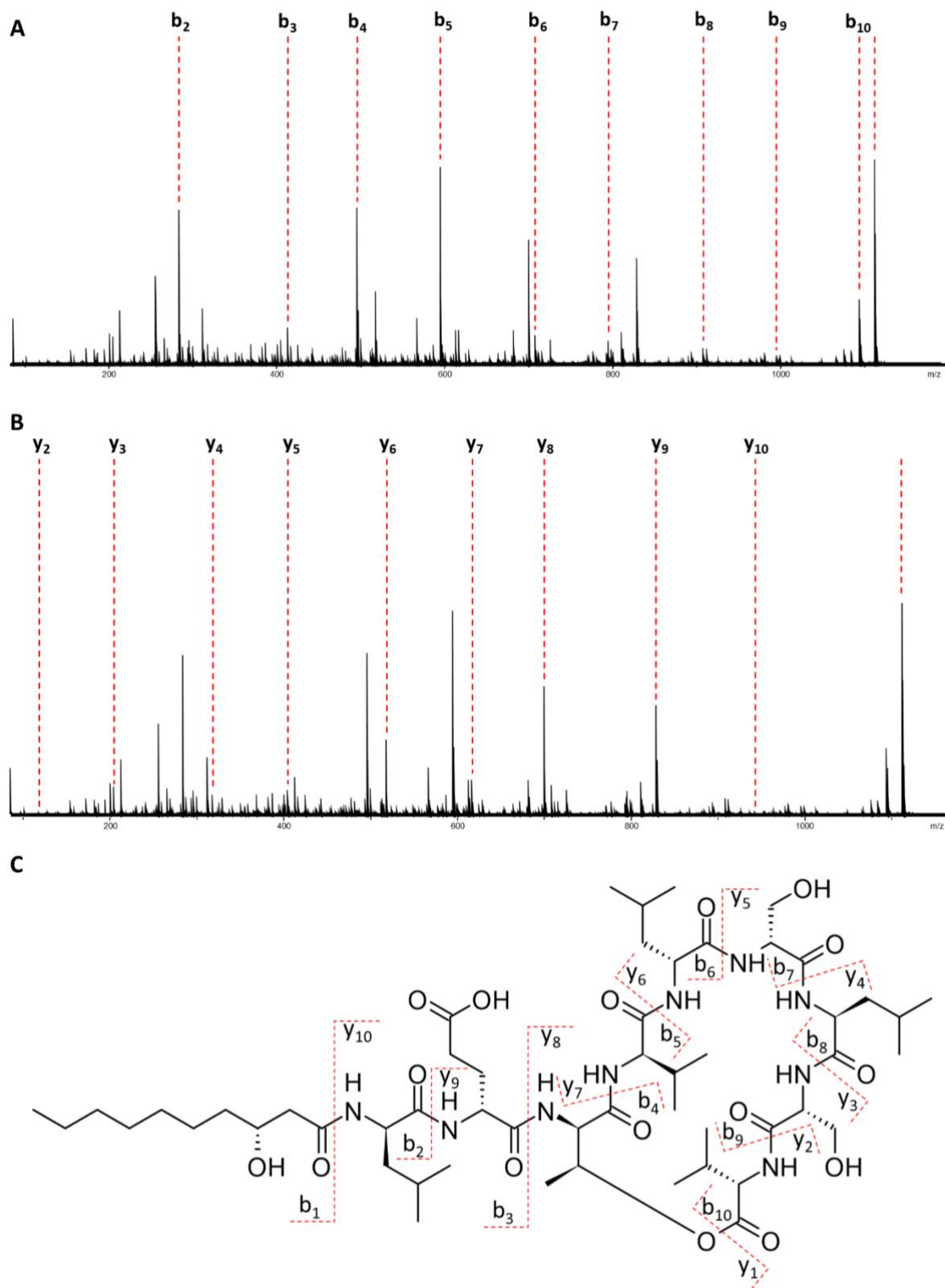


Figure 32. Assignment of MS² fragments to the structure of massetolide E (**14**). **A:** b-fragments; **B:** y-fragments; **C:** Structure of massetolide E (**14**). All theoretical b- and y-fragments were assigned.

To examine whether leucine or isoleucine are incorporated and at which position I performed inverse feeding experiments. ^{12}C amino acids were fed to cultures in ^{13}C ISOGROTM media. The extracts were measured *via* LC-HRMS (Fig. 33). I analysed the isotope pattern in the MS¹ spectra to determine how many molecules leucine, isoleucine and valine are incorporated per molecule of **13** (Fig. 33 A) and **14** (Fig. 33 B), respectively. A shift of 6 corresponds to the incorporation of one leucine or isoleucine molecule and a shift of 5 corresponds to the incorporation of one valine molecule. Furthermore, I fed deuterated leucine (D₁₀) to cultures grown in XPP medium lacking leucine and analysed the fragmentation pattern in comparison to cultures grown in regular XPP medium to determine the positions of leucine and isoleucine in **14** (Fig. 33 C). In this experiment, a neutral loss of 113 corresponds to isoleucine at that position while a neutral loss of 122 corresponds to leucine at that position.

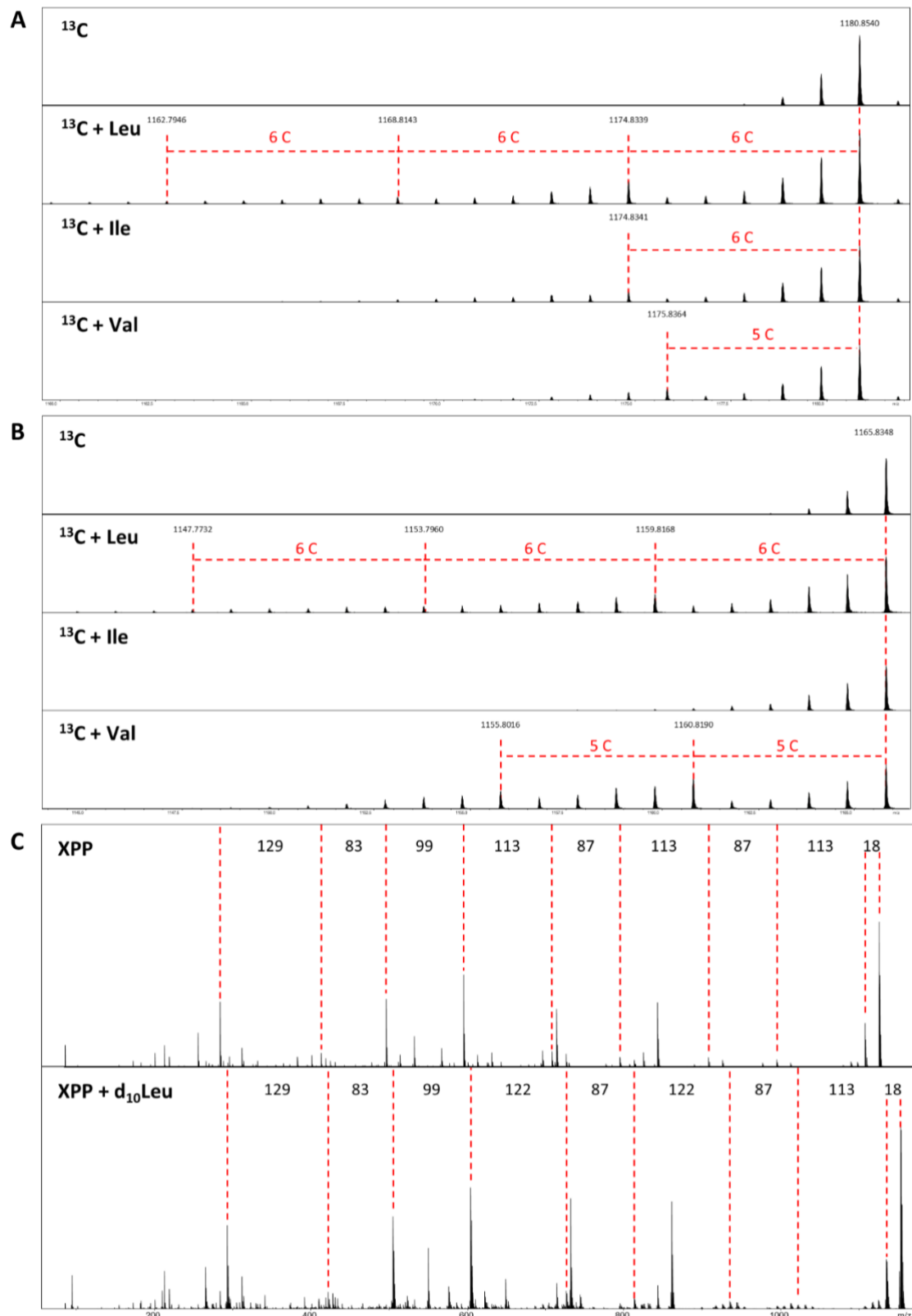


Figure 33. Determination of leucine, isoleucine and valine residues in **13** and **14**. **A:** Mass shifts of **13** resulting from inverse feeding of leucine, isoleucine or valine to cultures grown in ^{13}C -labeled medium. **B:** Mass shifts of **14** resulting from inverse feeding of leucine, isoleucine or valine to cultures grown in ^{13}C -labeled medium.

Feeding experiments with leucine, isoleucine and valine in ^{13}C labelled medium showed that **13** has three leucine, one isoleucine and one valine residue (Figure above A). Furthermore, it showed that **14** has three leucine, two valine and zero isoleucine residues. This indicates that, between **13** and **14**, the single isoleucine residue is “exchanged” for a valine residue. Feeding experiments with deuterated leucine in XPP medium support this assumption. The ion m/z 1154.8554 $[\text{M}+\text{H}]^+$ was picked to analyse the y -fragmentation of **13** labelled with deuterated leucine and m/z 1126.6740 $[\text{M}+\text{H}]^+$ was picked to analyse the y -fragmentation of **13** without isotopic labelling (Fig. 33 C). m/z 1154.8554 $[\text{M}+\text{H}]^+$ corresponds to the incorporation of three deuterated leucine residues. No ions corresponding to the incorporation of fewer or more than three leucine residues were detected. The first neutral loss of m/z 1154.8554 $[\text{M}+\text{H}]^+$ corresponding to either leucine or isoleucine (y_2 ; 113) in the spectrum of the unlabelled sample, was the same in the labelled sample, while all other two neutral losses corresponding to either leucine or isoleucine displayed the mass 122 instead of 113.

All results taken together I could determine the structures of **13** and **14** as shown in Figure 34.

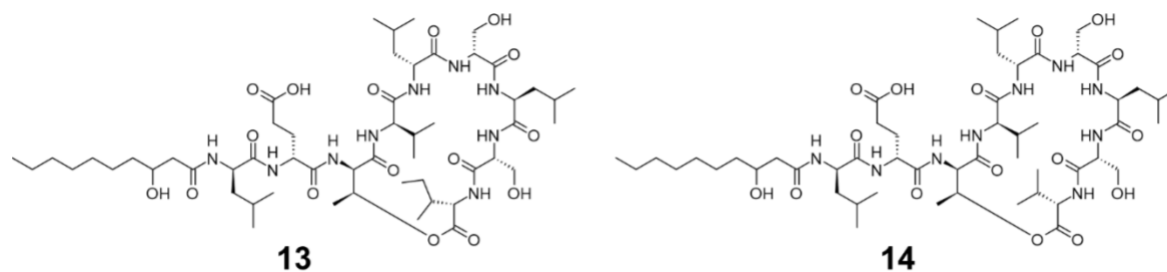


Figure 34. Structures of viscosin (**13**) and massetolide E (**14**).

The amino acid configuration was derived from bioinformatics predictions using antiSMASH. Every C_{dual} domain was assumed to epimerise one amino acid. The amino acid sequence determined by MS^2 analysis in combination with isotopic labelling experiments facilitated the differentiation between leucine and isoleucine moieties and their positions. The 3-hydroxy-decanoic acid moiety that supposedly is incorporated by the C_{starter} domain of the NRPS was derived from the mass (b_2 , Fig. 31 and 32), in the knowledge that it is also

present in other peptides of the viscosin NPF. While a fitting neutral loss was found in MS² spectra of **13** and **14**, the configuration of this moiety remains to be investigated.

3.3.3 Bioactivity guided fractionation

After elucidating the structures of **13** and **14**, the question remained, if one or both compounds are responsible for the bioactivity. To answer this, methanolic XAD extracts of MYb11 and MYb12 were fractionated over time on a HPLC system. The concentration of each fraction was normalized and all of them were tested in a disc diffusion assay against *B. thuringiensis* strains BT247 and BT679 (Fig.35). Here, only the results of the disc diffusion assay with BT679 are shown, since the results for 247 looked identical.

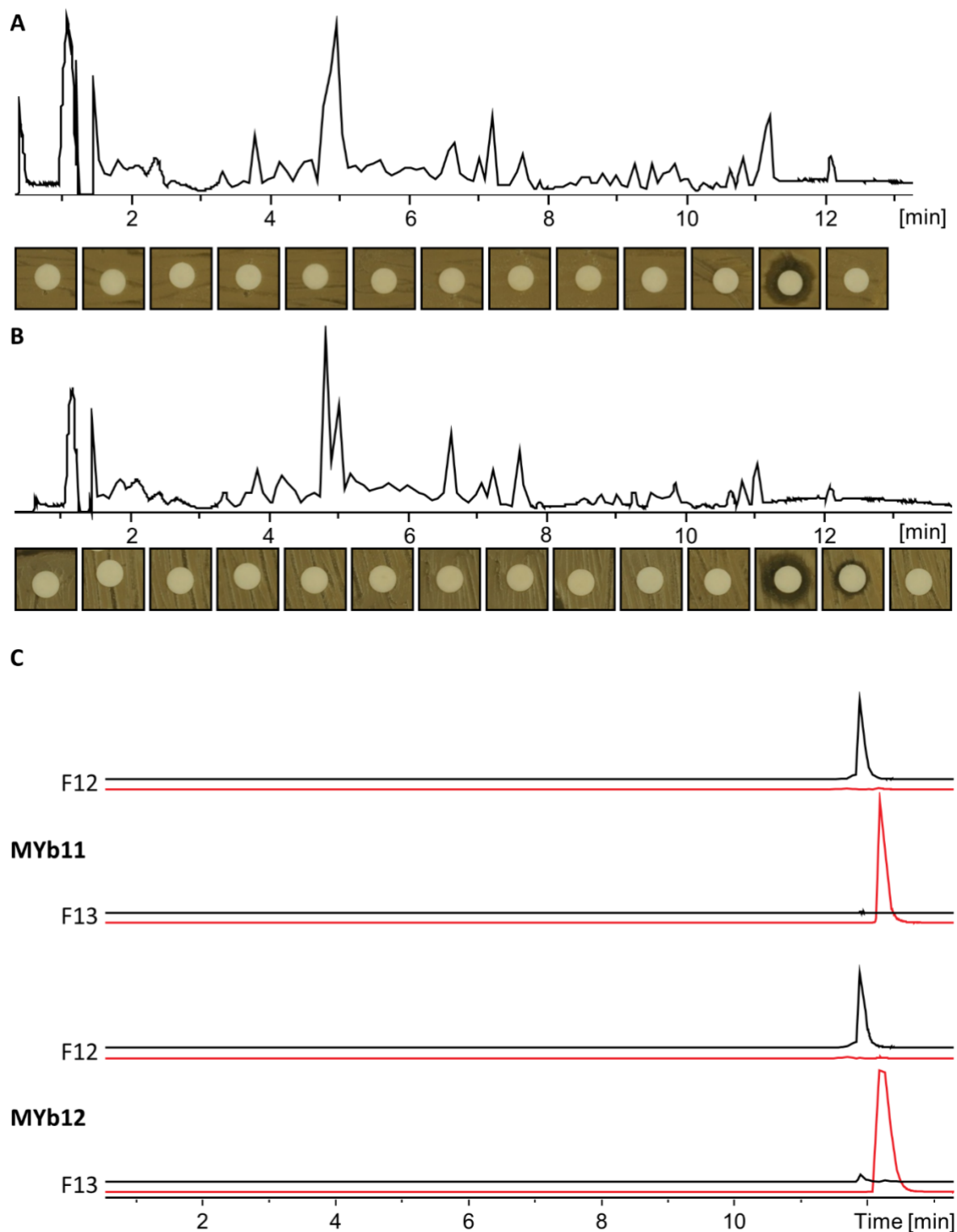


Figure 35. Massetolide E was identified as the bioactive compound against *B. thuringiensis* BT679. **A and B:** Base peak chromatograms of extracts from MYb11 (**A**) and MYb12 (**B**) with images of the disc diffusion assay showing the bioactivity of 0.8-min fractions correlating to the respective portion of the chromatogram against *B. thuringiensis* BT679. **C:** EICs of massetolide E (red; m/z 1112.67 [M+H⁺]) and viscosin (black; m/z 1126.68 [M+H⁺]) in fractions 12 and 13 of MYb11 and MYb12.

While only fraction F12 of MYb11 showed activity against BT679 (Fig. 35 A), fraction 12 and 13 of MYb12 inhibited the growth of the pathogen (Fig. 35 B). LC-MS analysis of fractions F12 and F13 of both strains showed that F12 of MYb11 contained only **14** and F13 contained only **13** (Fig. 35 C). Contrarily, F13 of MYb12 contained both compounds while F12 contained only **14** (Fig. 35 C), showing that the bioactivity against BT247 and BT679 can only be attributed to **14**.

3.4 GameXPeptides from *Xenorhabdus miraniensis*

Using a promoter exchange approach, my colleague Dr. Zhengyi Qian activated a gene cluster in *Xenorhabdus miraniensis* ($gxpS_{Xmira}$) that encodes for a BGC similar, but not identical to the known GameXPeptide BGC ($gxpS_{TT01}$) from *Photorhabdus luminescens* TT01, which is also present in *Xenorhabdus doucetiae*. They both encode for NRPS with five modules and similar specificities predicted by antiSMASH¹⁷⁷ (prediction not shown). While the two BGCs share those features, there are also some differences. $gxpS_{TT01}$ encodes for a NRPS ($GxpS_{TT01}$) with the C domain pattern C/E-C-C/E-C/E (Fig. 36 A). This results in the synthesis of full length NRPs with the configuration D-L-D-D-L¹⁵⁵ (Fig. 36 B). $GxpS_{Xmira}$ instead, displays the C domain pattern C/E-C-C/E-C, thus theoretically resulting in the synthesis of full length NRPs with the configuration D-L-D-L-L (Fig. 36 B).

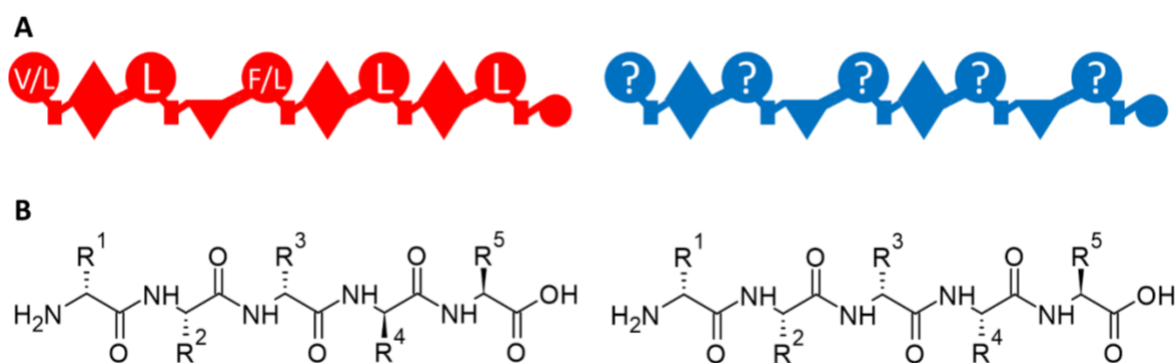


Figure 36. Comparison of GameXPeptide synthetases $GxpS_{TT01}$ and $GxpS_{Xmira}$ (A) and their respective peptide products (B) with focus on stereochemical configuration.

The substrate specificities of the A domains as well as the epimerisation activity of the C/E domains for $GxpS_{TT01}$ are well characterized, yet nothing is known about the same for $GxpS_{Xmira}$. In this work, I characterised $GxpS_{Xmira}$ using LC-MS paired with isotopic labelling and chemical synthesis of candidate peptides, which was performed by Trinetri Goel, with emphasis on the differences between the two NRPS.

3.4.1 Identification of GameXPeptides

3.4.1.1 Promoter exchange in *Xenorhabdus miraniensis*

Promoter exchange based gene cluster activation led to the production of seven peptides (15-21) in *X. miraniensis* (Fig. 37 A). The sum formulae of these peptides was subsequently

determined through stable isotopic labelling experiments with ^{13}C - and ^{15}N -labelled media (Fig. 37 B).

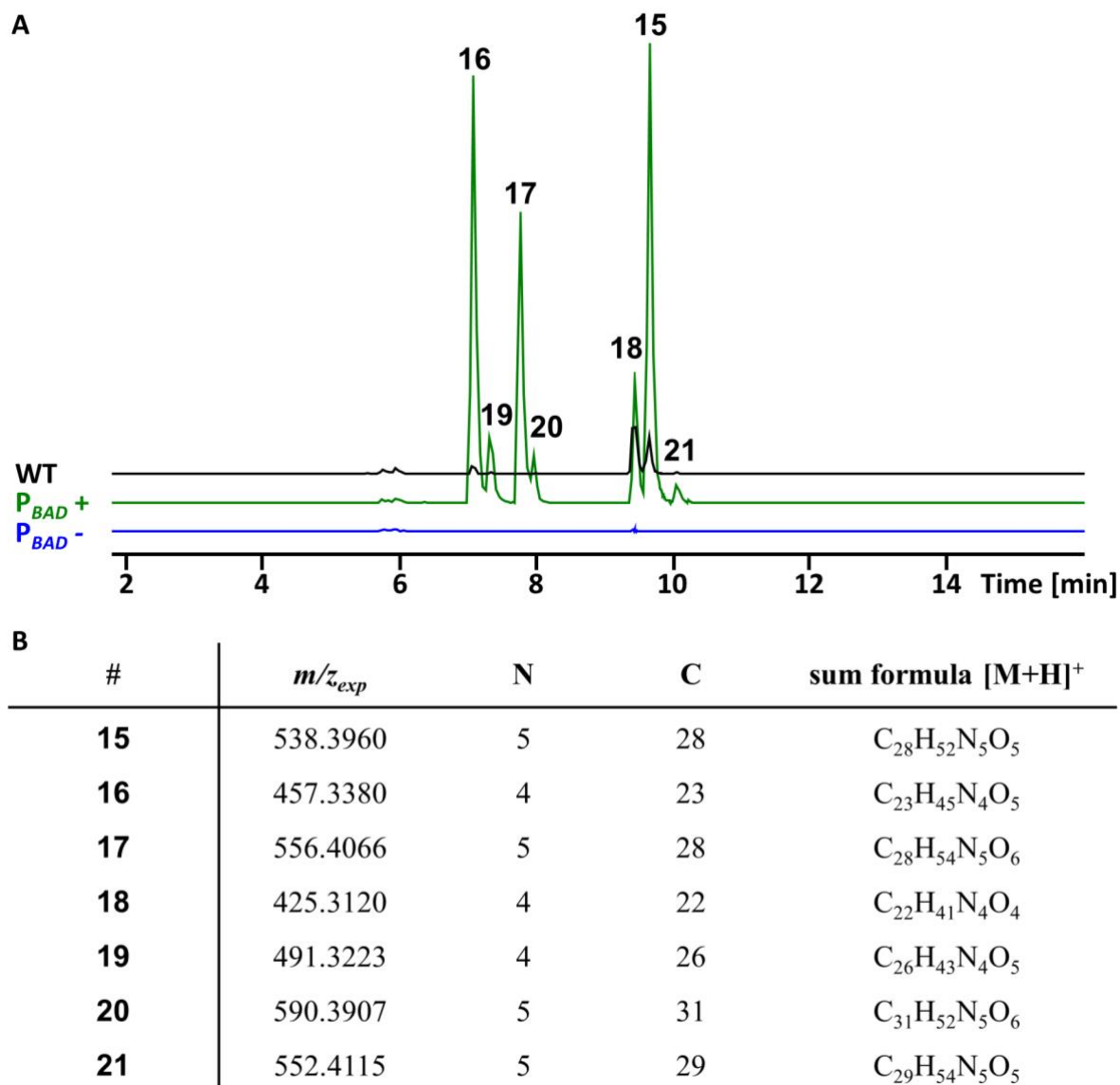


Figure 37. Identification of GameXPeptide compounds **15-21** through promoter exchange and isotopic labelling in *X. miraniensis*. **A:** Combined EIC of compounds **15-21**. Black: wildtype; green: *X. miraniensis* P_{BAD_gxpS} with arabinose; blue: *X. miraniensis* P_{BAD_gxpS} without arabinose. **B:** Properties of compounds **15-21**. Number of carbon and nitrogen atoms was determined through isotopic labelling (data not shown).

Similar to $\text{GxpS}_{\text{TT01}}^{182}$, $\text{GxpS}_{\text{Xmira}}$ seems to produce tetra- and pentapeptides indicated by the number of nitrogen and oxygen atoms in the sum formulae (Fig. 37 B). Due to the mass of the detected peptides and their elemental composition, it can be deduced that $\text{GxpS}_{\text{Xmira}}$ produces cyclic and linear peptides consisting of only hydrophobic amino acids, just like

GxpSTT01. Peptides with an equal number of nitrogen and oxygen atoms are considered cyclic (**15**, **18**, **21**). Peptides with four nitrogen atoms are considered tetrapeptides, while peptides with five nitrogen atoms are considered pentapeptides. The amino acid composition as well as the sequence of **15-21** was determined in further experiments.

3.4.2 Structure elucidation via MS

3.4.2.1 Inverse amino acid feeding

The amino acid composition of compounds **15-21** was determined through inverse feeding of unlabelled amino acids to cultures grown in ^{13}C -labelled medium and subsequent LC-MS analysis. This is exemplarily shown for **17** (Fig. 38).

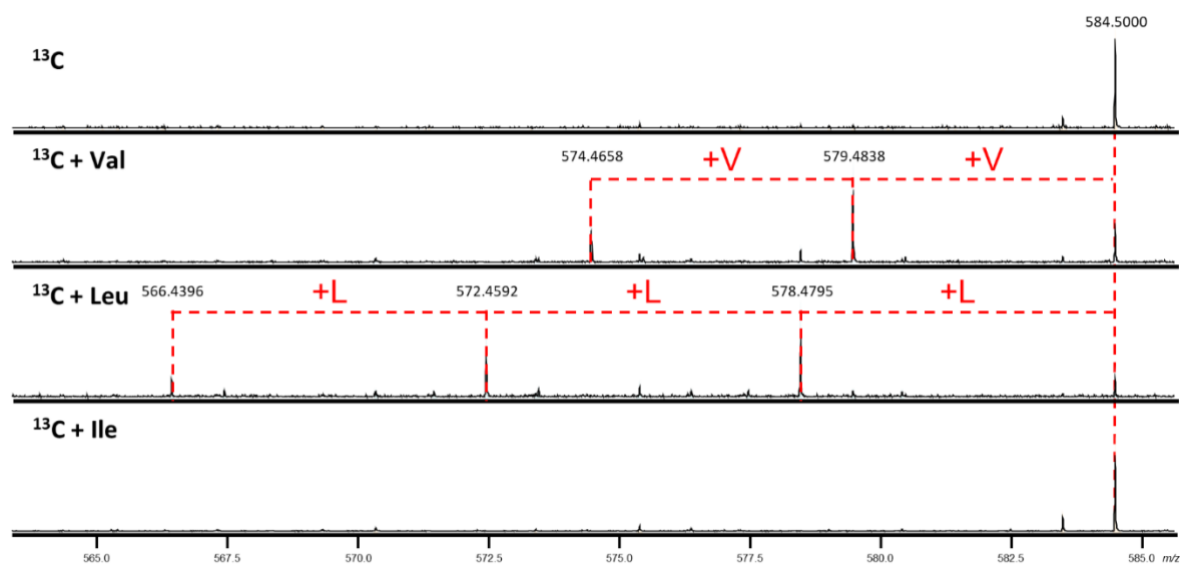


Figure 38. Determination of the amino acid composition of **17** through inverse amino acid feeding experiments.

17 consists of two valine and three leucine residues (Fig. 38). The amino acid composition of all peptides is shown in table 13.

Table 13. Amino acid composition of **15-21**.

#	amino acid composition
15	cyclo(LLL VV)
16	LLLV
17	LLL VV
18	cyclo(LL VV)
19	FLLV
20	FLL VV
21	cyclo(LLLL V)

3.4.2.2 MS₂ sequence analysis

To elucidate which module incorporates which amino acid, I analysed MS² data for peptides **15-21** with particular interest in **17**, **20** and **21** (Fig. 39) as I will explain in the following.

Gxp_{STT01} has three A domains with rigid specificities that only incorporate leucine and two flexible A domains that incorporate valine or leucine (A₁) and phenylalanine or leucine (A₃) (Fig. 36). Something similar was considered for Gxp_{Xmira}, since all peptides contain two to three leucine residues, one to two valine residues and one or no phenylalanine residue (Tab. 13). If every NRPS module is considered to be used once, two modules of Gxp_{Xmira} must have a flexible A domain specificity as can be examined by comparing the pentapeptides **17**, **20** and **21**. By comparing **17**, **20** and **21**, one can infer that two leucine incorporating A domains have flexible specificities, one being L/V and the other one being L/F, similar to A₁ and A₃ from Gxp_{STT01}. I used MS² sequence analysis of those three compounds to determine the flexible positions in Gxp_{Xmira}. Because of the collinearity rule of NRPS, linear peptides can be investigated to gain insights on the A domain specificity of the respective NRPS module. This is not possible with cyclic peptides, since multiple b and y fragmentation patterns overlay, which makes it impossible to determine, the C- and N-termini of the peptide. **17** was selected because it is a linear pentapeptide with two valine

residues. The goal was to determine the positions of the two valine residues, of which one is flexible. **20** was chosen to determine the position of the flexible leucine/phenylalanine residue, because it is a linear pentapeptide with a phenylalanine residue. Hence, the flexible L/F residue can be determined by comparison of **20** (FLLVV) with **17** (LLLVV)

Lastly, **21** was selected because it is the only GameXPeptide identified from *X. miraniensis* whose mass matches with the known GameXPeptide C from TT01. This makes it an interesting candidate to study the differences between the two NRPS and their respective products.

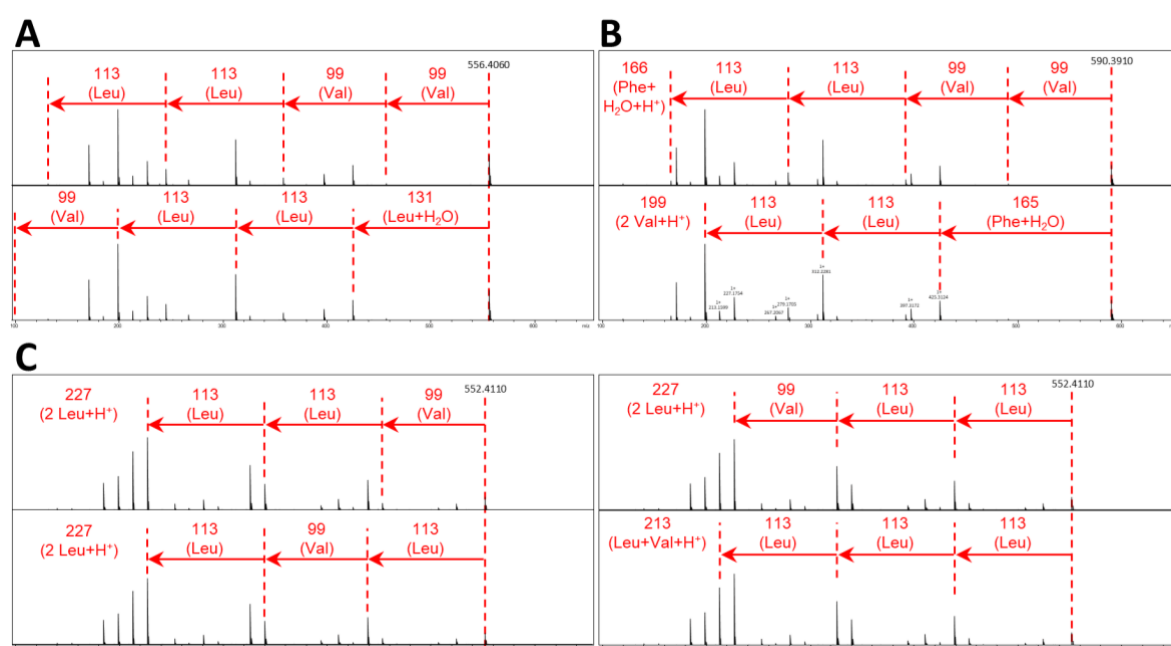


Figure 39. MS² peptide sequence analysis of **17** (A), **20** (B) and **21** (C) produced by *X. miraniensis*.

17 (m/z 556.4060 $[M+H]^+$) displayed neutral losses of 99 (Val), 99 (Val), 113 (Leu) and 113 (Leu) in b fragmentation and 131 (Leu+H₂O), 113 (Leu), 113 (Leu) and 99 (Val) in y fragmentation (Fig. 39 A). Hence, the sequence of **17** was determined as VVLLL.

20 (m/z 590.3910 $[M+H]^+$) displayed neutral losses of 99 (Val), 99 (Val), 113 (Leu) and 113 (Leu) in b fragmentation and 165 (Phe+H₂O), 113 (Leu) and 113 (Leu) in y fragmentation (Fig.39 B). Therefore, the sequence of **20** was determined as VVLLF.

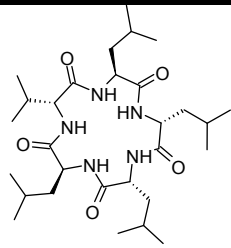
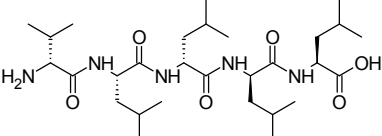
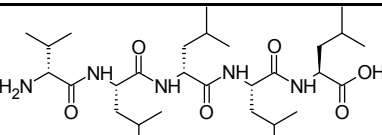
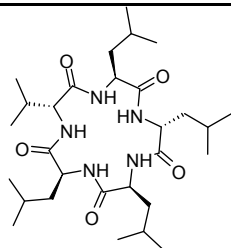
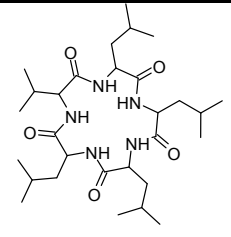
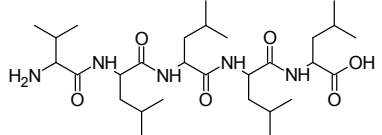
The sequence of **21** (m/z 552.4110 [M+H]⁺) could not be determined through this experiment. Since **21** is a cyclic peptide the neutral loss of 99 corresponding to valine can be found at multiple positions throughout different fragmentation patterns (Fig. 39 C).

By combining the sequences of **17** (VVLLL) and **20** (VVLLF), we can determine that the A₅ domain of GxpS_{Xmira} is specific for leucine and phenylalanine. Yet, the flexible V/L A domain could not be determined as either the first or the second A domain, both incorporating valine, is also specific for leucine.

3.4.2.3 Comparison of synthetic and natural peptides

With the acquired structural information, I was able to formulate the structure of candidate peptides. I chose compound **21** to elucidate the structure, especially the configuration of the amino acids, since its stereoisomer is GameXPeptide C produced by *P. luminescens* TT01. My colleague Trinetri Goel synthesised four candidate peptides based on the structural information I had gathered with previous experiments (Tab. 14). I compared the synthesised peptides to extracts from TT01 and *X. miraniensis* containing cyclic and linear peptides with the amino acid sequence VLLLL. All peptides used in this experiment are displayed in table 14. Stereochemistry of **21** and **22** was to be elucidated in this experiment and therefore is not indicated here.

Table 14. Synthetic candidate peptides with defined stereochemistry and selected naturally occurring GameXPeptides from TT01 and *X. miraniensis*.

name	m/z [M+H] ⁺	origin	structure
MTG55 cyclo(GXP C)	552	synthesis/TT01	
TG3 GXP C	570	synthesis/TT01	
TG4	570	synthesis	
TG5	552	synthesis	
21	552	<i>X. miraniensis</i>	
22	570	<i>X. miraniensis</i>	

First, the synthetic peptides and the extracts from TT01 and *X. miraniensis* were measured in separate LC-MS runs to determine the retention time of each peptide. Thereafter, the synthetic peptides **MTG55**, **TG3**, **TG4** and **TG5** were coinjected one by one with the extracts from TT01 and *X. miraniensis* (Fig. 40). Wherever only one single peak occurred in

the chromatogram during the coinjection experiment, the synthetic and the naturally occurring peptide were the same. If two peaks occurred, they differed from each other.

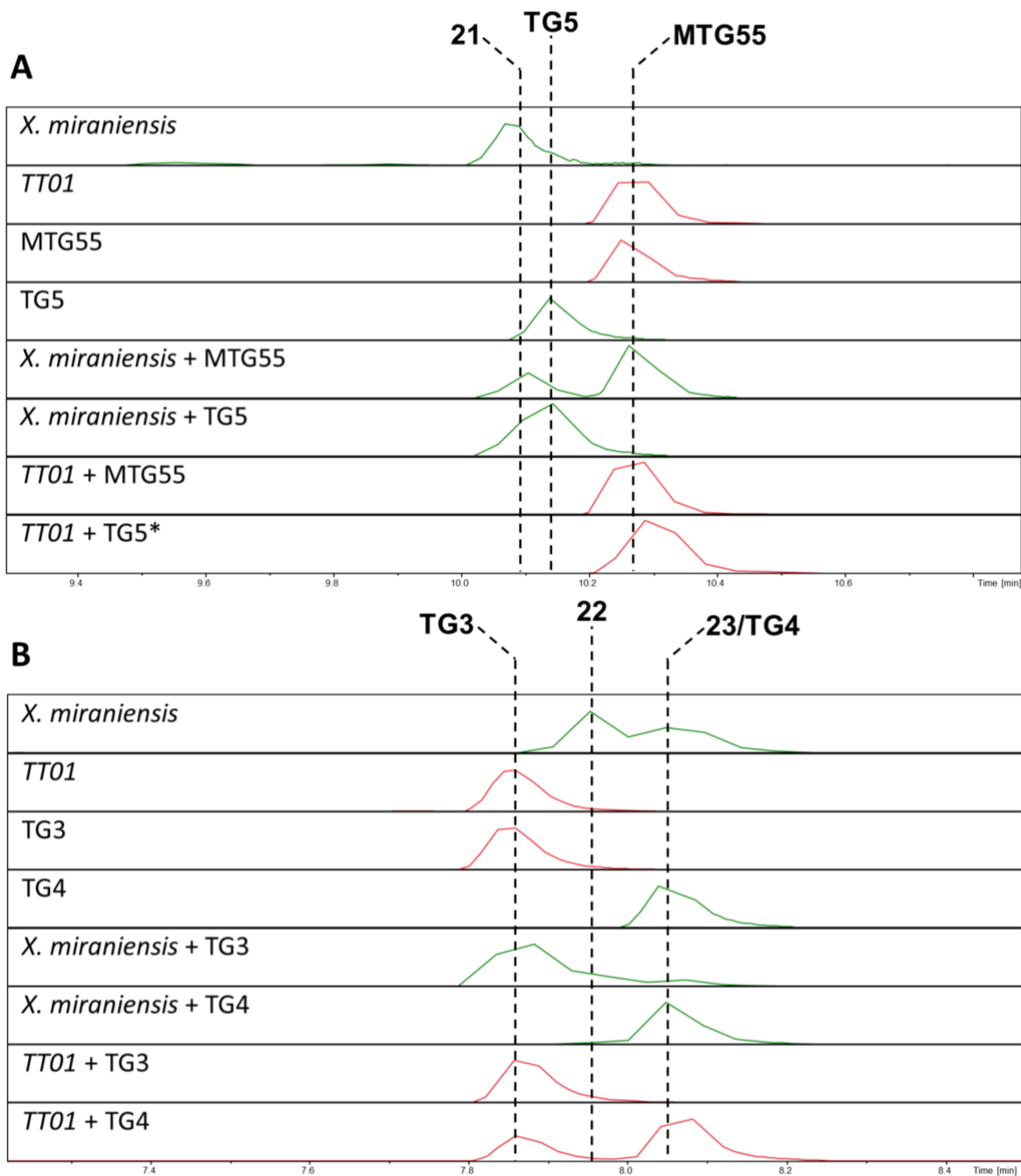


Figure 40. Coinjection experiment to determine the stereochemical configuration and peptide sequence of **21** and **22**. EICs for each mass are shown. **A:** Extracts from *X. miraniensis* and TT01 were coinjected with MTG55 and TG5. **B:** Extracts from *X. miraniensis* and TT01 were coinjected with TG3 and TG4. *: Intensity of TG5 was too low compared to the signal from TT01, however presence of TG5 was confirmed.

The RT of neither TG5 nor MTG55 matched with the RT of **21** (Fig. 40 A). Thus, the structure of **21** also differs from those of TG5 and MTG55 and has to be investigated in further experiments. TT01 produces MTG55 (cyclo(Gxp C)) as expected.

Coinjection experiments with linear peptides TG3 (GxpC) and TG4 revealed two compounds with the same mass (m/z 570 [M+H]⁺) that are produced by *X. miraniensis* (Fig. 40 B). The compounds were subsequently named **22** and **23**. As expected, TT01 only produced one compound with this mass, TG3. **23** displayed the same RT as TG4. Hence, **23** and TG4 are identical. The structural properties of **22** could not be solved as neither **TG3** nor **TG4** displayed the same retention time. To elucidate the structure of **22** I compared MS² data from **22** and **23** (Fig. 41), whose structure is now known.

The LC chromatogram showed two peaks for **22** (RT 7.96 min) and **23** (RT 8.07 min) (Fig. 41 A). MS² analysis of **22** revealed neutral losses of 113 (Leu), 99 (Val), 113 (Leu) and 113 (Leu) in b fragmentation, whereas the same analysis of **23** revealed neutral losses of 99 (Val), 113 (Leu), 113 (Leu) and 113 (Leu) (Fig. 41 B).

Based on these results, I was able to deduce the structures of compounds **21**, **22** and **23** (Fig. 41 C). The structure of **23** (vLILL) was already confirmed through the coinjection experiment (Fig. 40). **22** differs from **23** only in the position of the valine residue (Fig. 41 B). The stereochemical configuration at each position can be considered equal to the configuration in **23**, because of the collinearity rule for NRP biosynthesis. Therefore, **22** has the sequence IVILL (Fig. 41 C). **22** (IVILL) and **23** (vLILL) are both produced by *X. miraniensis*, as is **21** (Fig. 40). Since the fragmentation pattern of **21** (Fig. 39 C) matches the cyclic peptides with the sequences of both **22** and **23**, cyclo(IVILL) and cyclo(vLILL) were considered. Yet, the RT of **21** did not match with the RT of cyclo(vLILL) (**TG5/23**) (Fig. 40 A). Hence, the structure of **21** was determined as cyclo(IVILL) (Fig. 41 C).

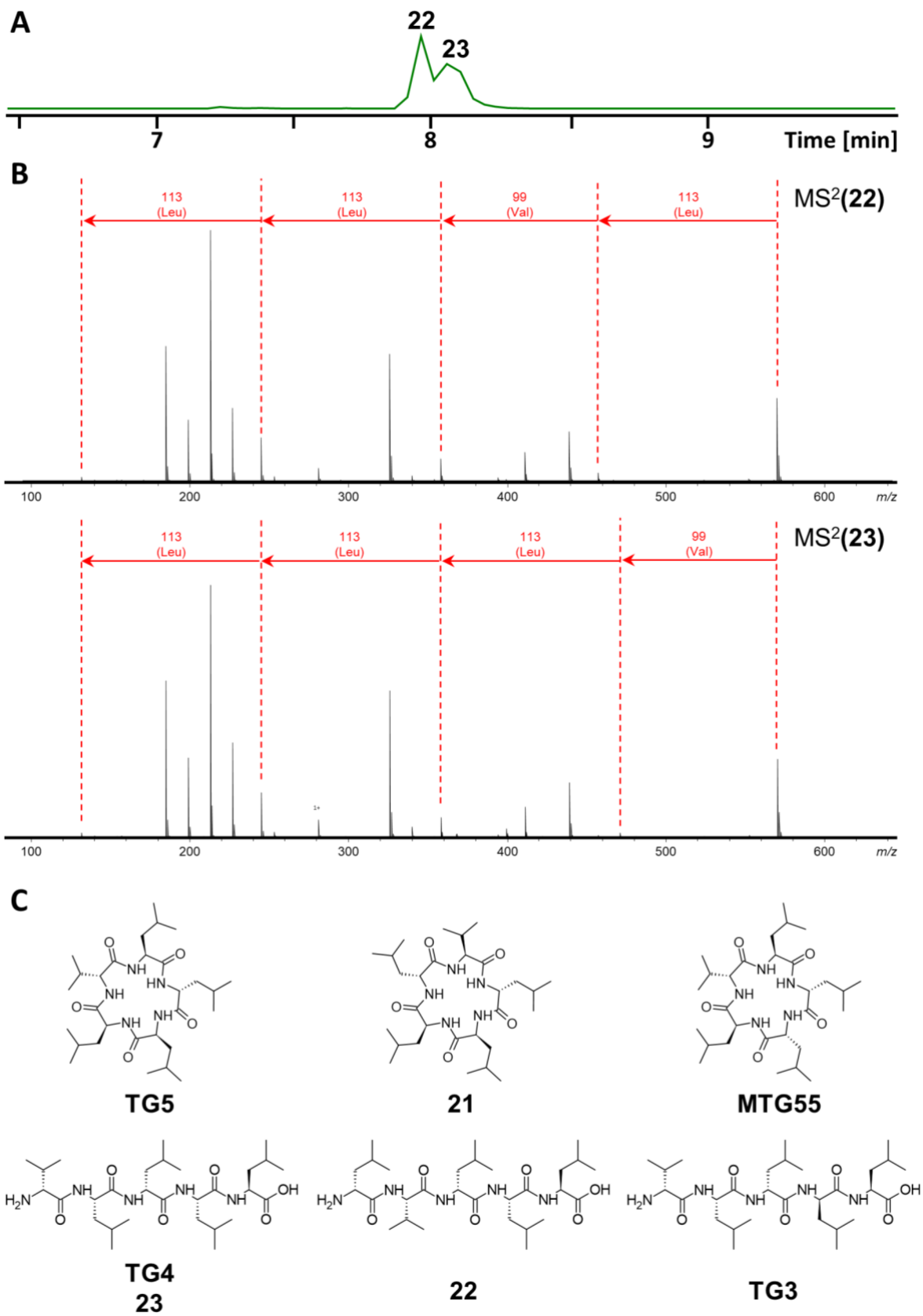


Figure 41. Determination of structural differences between **22** and **23**. **A:** EIC of **22** and **23**. **B:** MS² fragmentation of **22** and **23**. **C:** Deduced structures of **21**, **22** and **23**.

4 DISCUSSION

4.1 Inthraszentin

In this work, I identified the full-length product of the *isz* BGC in *X. szentirmaii*, *X. indica* and *P. thracensis*. Previous attempts to activate said BGC in *X. szentirmaii* failed (personal communication Prof. Dr. Helge B. Bode, Dr. Jürgen Breitenbach and Svenja Simonyi). To my knowledge, I was the first person to investigate the *isz* BGC in *P. thracensis*. However, former group member Dr. Olivia Schimming successfully cloned the *isz* BGC from *X. indica*, but accidentally truncated the BGC during cloning (Fig. 42). This resulted in the production of xenolindicin¹⁸³ (Fig. 43).

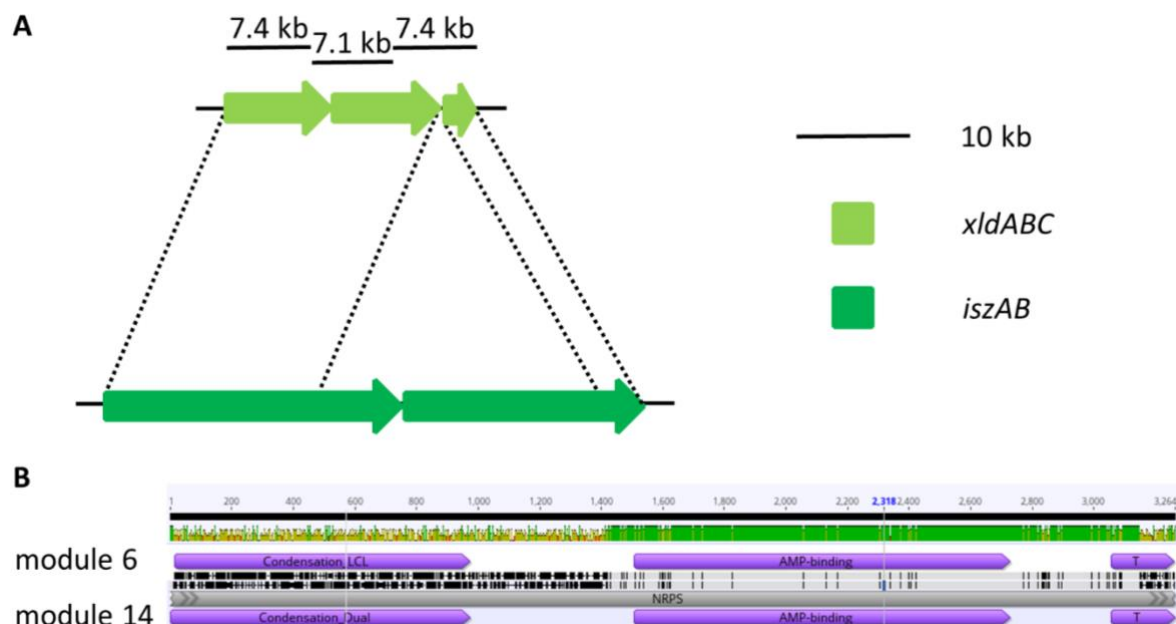


Figure 42. Origin of the xenolindicin BGC *xldABC*. **A:** Truncation of the *isz* BGC led to formation of the *xldABC* BGC. **B:** Homology of threonine incorporating modules 6 and 14 of *iszAB*.

This accident had its roots in the homology of A₆ and A₁₄. Both A domains are specific for threonine. The genome sequence at the time was incomplete and the homology of the sequences resulted in an imperfect assembly, where the sequence read skipped from A₆ to A₁₄ at some point. Thus, Dr. Schimming thought she needed to clone a BGC of 22 kb instead of 47 kb.

Curiously enough, she was able to amplify a PCR fragment that resulted from an error, which cannot be traced back to its roots. It is conceivable that the polymerase skipped from the sequence of module 6 to that of module 14, mimicking the error in the genome assembly. She aimed to amplify three split parts of the BGC with the size of 7.4 kb, 7.1 kb and 7.4 kb¹⁸³, of which she amplified the first two correctly. However, she was not able to amplify the last one correctly because it would have ranged 33 kb instead of 7.4 (Fig. 42). She amplified a fragment of 3.8 kb and neglected the unexpected size of this fragment. She then assembled a plasmid containing the truncated version of the *isz* BGC via yeast homologous recombination cloning and subsequently transformed into *E. coli*¹⁸³. She was able to achieve the heterologous production of xenolindicins A-C (Fig. 43).

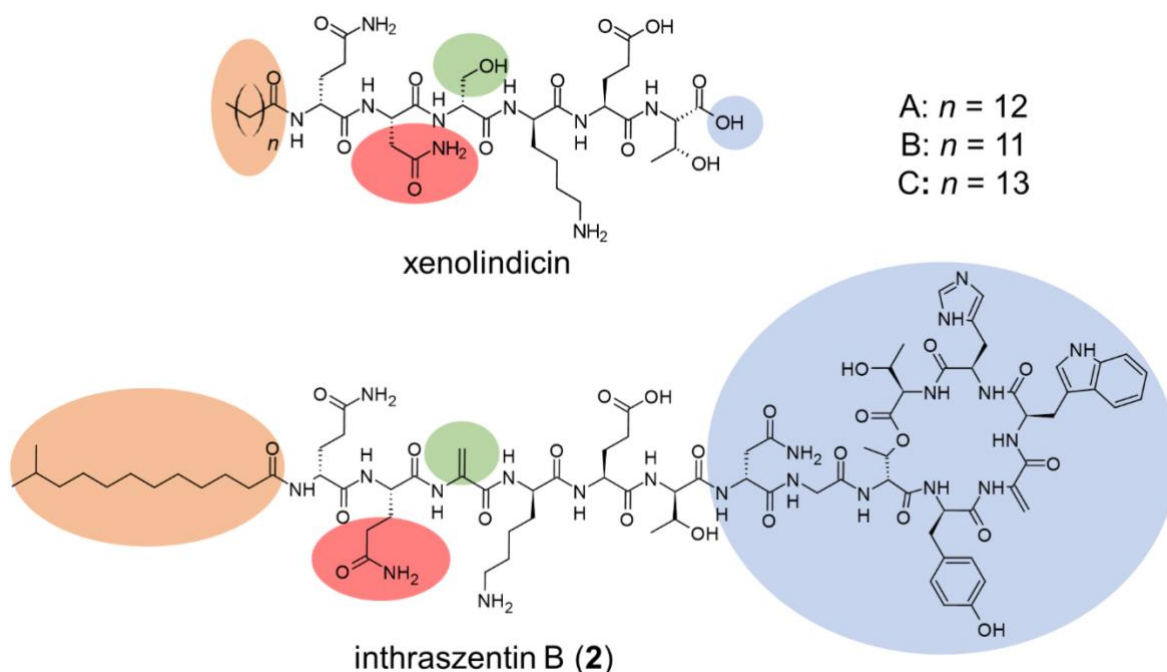


Figure 43. Structural differences between xenolindicins A-C and intraszentin B (2).

By comparing the proposed intraszentin B structure with the published xenolindicin structure, I identified four major differences (Fig. 43): (i) intraszentin B contains *iso*-tridecanoic acid, while all three xenolindicin derivatives contain straight chain fatty acids (Fig. 43, orange); (ii) the second amino acid incorporated in xenolindicins is asparagine, in intraszentin B it is glutamine (Fig. 43, red); (iii) xenolindicin contains a serine residue at position three, while I predict intraszentin B to contain dehydroalanine at said position (Fig.

43, green), as well as at position eleven; (iv) xenolindicin is a linear lipohexapeptide and all intraszentins (**1-3**) are cyclic lipodepsitradecapeptides (Fig. 43, blue).

All intraszentins were produced in their natural host, while xenolindicins were discovered through heterologous production in *E. coli*¹⁸³. However, *E. coli* is not able to produce branched chain fatty acids. In fact, *E. coli* incorporates straight chain fatty acids instead of branched chain fatty acids, when it is deployed as a production host for natural products from *Photorhabdus* spp.^{29,184,185}. Thus, the difference in the incorporated fatty acids can be attributed to the production hosts.

The different lengths of the peptides stems from the accidental engineering of the *isz*-NRPS into a truncated version by Dr. Olivia Schimming. Though analysis of the fragmentation pattern of intraszentin B indicates the incorporation of glutamine at position two (Fig. 21), xenolindicin contains an asparagine residue at said position (Fig. 43). While glutamine and asparagine are structurally very similar, the A domain should be very specific, indicated through the fact that no derivatives with different amino acids were detected of neither xenolindicin nor intraszentin. To my knowledge, A domain specificity is mostly consistent throughout different production hosts, yet some exceptions have been observed in our group (Kenan A. J. Bozhüyük, personal communication). Hence, there is an urgent need to confirm the proposed xenolindicin structures with complementary methods. Although *E. coli* and *Xenorhabdus* spp. are closely related, the natural version of this natural product with its natural composition and natural function is produced in its natural host.

I predicted the presence of dehydroalanine in intraszentin A-C by balancing out experimental and bioinformatic data (Fig. 20). The only possibility to fit everything into a plausible structure proposal was by assuming that every bioinformatically predicted serine residue in all three intraszentins is in fact a dehydroalanine residue. α,β -dehydro amino acids like dehydrobutyrine (Dhb) and dehydroalanine (Dha) are common in some NRPs and introduce proteolytic stability and reactive functionality to the peptides¹⁸⁶. These reactive properties can even facilitate binding to target structures¹⁸⁷. However, there is no evidence yet, that Dha is incorporated through the intraszentin synthetase. Unfortunately, bioinformatic analysis cannot lead to the desired clarity. Although dehydroamino acids are typically incorporated by C domains that incorporate modified amino acids (C_{modAA})¹⁸⁸, dual condensation/epimerisation domains (C_{dual}) have also been linked to the incorporation of

dehydroamino acids in NRPs¹¹⁴. This is especially true for cyclic lipopeptides¹⁸⁸. Patteson *et al.* analysed all C_{dual} domains involved in nunapeptin, syringomycin E, syringopeptin and jessenipeptin biosynthesis and found that C_{dual} domains that catalyse the dehydration cannot be phylogenetically distinguished from those who do not¹⁸⁸. Since all A domains specific for the incorporation of serine in IszAB are followed by a C_{dual} domain, it remains plausible that those C_{dual} domains catalyse the dehydration of serine to Dha.

Many natural products containing dehydroamino acids display potent bioactivities including the lantibiotic nisin¹⁸⁹, the histone deacetylase inhibitor romidepsin¹⁹⁰ and the cyanobacterial toxin microcystin-LR¹⁹¹, with the latter two being non-ribosomal peptide/polyketide hybrids. This hints at a potential bioactivity of intraszentins A-C.

Furthermore, intraszentins A-C share some structural properties with the NRPS-derived potent antibiotics daptomycin¹⁷⁴ and teixobactin¹⁷⁵ (Fig. 44 B). Both teixobactin and daptomycin kill Gram-positive bacteria by interrupting the biosynthesis of the cell envelope^{192,193}. Among the similarities between the three peptides is the capping of the N-terminal amine (Fig. 44 B). N-terminal capping hinders peptide degradation by proteases and peptidases¹⁹⁴. Furthermore, a hydrophobic N-terminus may be critical for interaction and disruption of bacterial membranes¹⁹⁵. Hydrophobicity can be either achieved through capping or through incorporation of aromatic amino acids, e.g. teixobactin, daptomycin. Teixobactin contains many residues with cationic properties that facilitate coordination of its target and are essential for killing activity¹⁹². Intraszentins also contain some positively charged residues under physiological conditions, e.g. lysine, histidine (Fig. 44 B). Although not identical, the unusual L-*allo*-enduracididine residue of teixobactin and the D-histidine residue of intraszentin are both cationic side chains of the cyclic C-terminal part of the peptides and may act in similar fashion. L-*allo*-enduracididine is directly involved in the coordination of the target of teixobactin¹⁹². Incorporation of D-amino acids, which occur in all three peptides, can make peptides more stable and may cause structural properties that foster bioactivity¹⁹⁶. The ester bond of both teixobactin and intraszentin is formed between the C-terminal carboxyl group and the side chain of D-threonine residues, which may further stabilise the molecule. It can only be speculated, whether the ring forming threonine residue is configured as 2*R*, 3*S* (D-threonine) as in teixobactin or 2*R*, 3*R* (D-*allo*-threonine).

In addition, the architecture of the daptomycin, teixobactin and inthraszentin BGCs display striking similarities (Fig. 44 A). All three BGCs, *iszAB*, *txbAB* and *dptABCD*, encode for NRPS with ABC transporter-related genes in close proximity thereof (Fig. 44 A). The resemblance is especially striking between *iszAB* and *txbAB*, where the ABC transporter is encoded almost directly downstream of the two NRPS genes. These transporters are typical for antibiotic producing organisms, presumably facilitating the secretion of the antibiotic and thereby introducing resistance to the produced antibiotic¹⁹⁷.

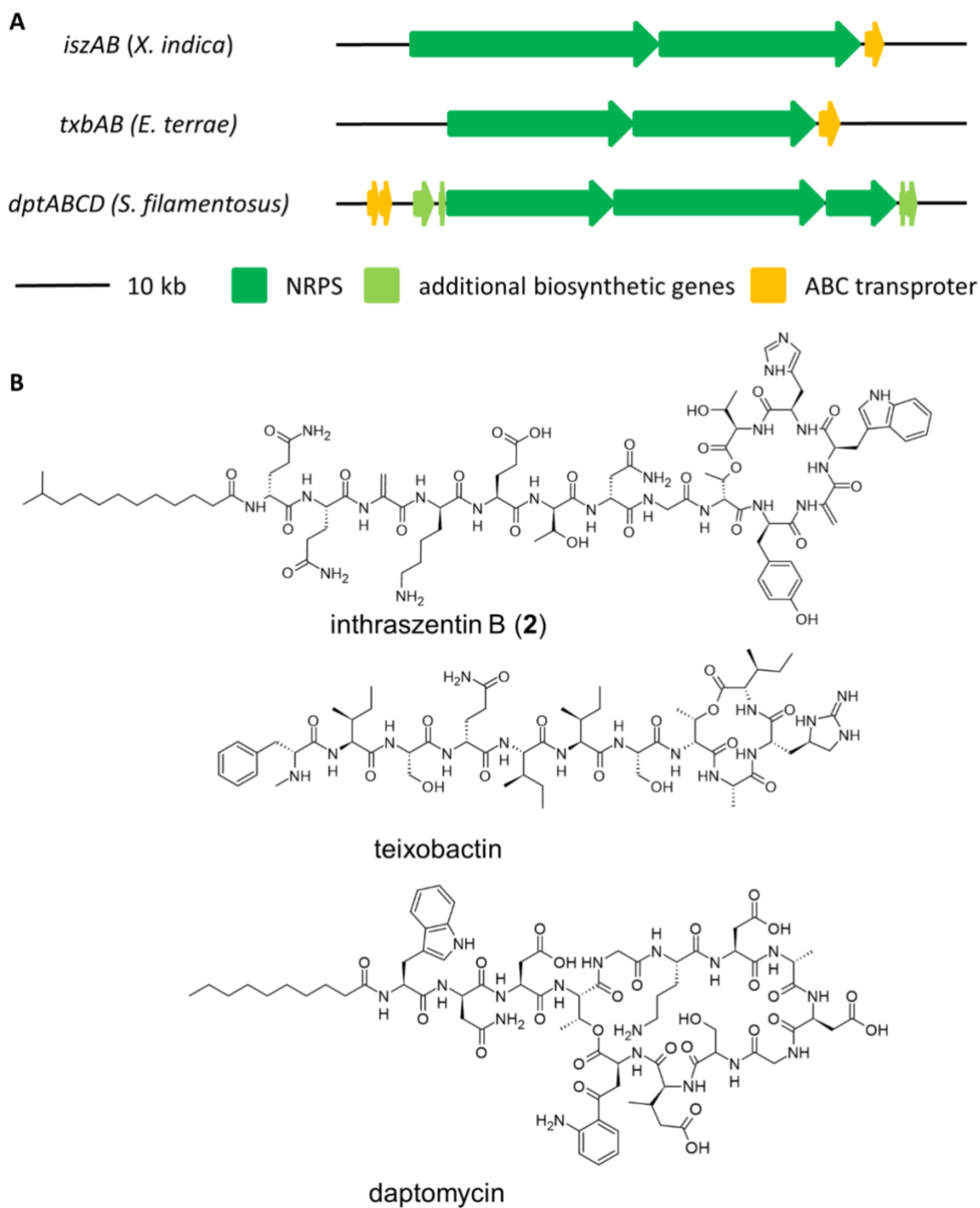


Figure 44. Comparison of intrazasentín B (2), teixobactín and daptomycín and the corresponding BGCs. **A:** Overview of the *isz*, *txb* and *dpt* BGCs. **B:** Structural similarities between intrazasentín B (2), teixobactín and daptomycín.

I was able to achieve the identification of the intraszentins and the corresponding gene cluster by using GNPS molecular networking to compare the metabolome of wild type and mutant strains. I combined the available bioinformatic data on BGCs, with the established promoter exchange approach and GNPS network analysis. I deployed promoter exchange strains, in which no activation of the production of new natural products were discovered, as knockout mutants and compared them to the corresponding wild type strains. Thereby, I was able to link the natural products produced only in the wild type to the corresponding knocked out gene cluster. Using GNPS network analysis aids in the discovery of MS signals that may be overlooked in manual or less sensitive automated analysis. In addition, this workflow offers an expedient if activation of NPBGCs is not successful. By exchanging the promoter, inactivation of the targeted BGC is probably still achieved. The generated knock out mutant strain can then be compared to the corresponding wild type strain to identify the natural product. Moreover, this method can be used to identify natural products from mutants generated by transposon mutagenesis, thereby increasing the throughput.

Although I gained as much structural information as possible, structure elucidation *via* complementary techniques remains necessary to confirm the herein proposed structures. Purification specifically remains a bottleneck in that regard, as it enables further experiments such as NMR, crystallisation as well as partial hydrolysis, which can be deployed for structure elucidation of lipopeptides¹⁹⁸. Since I generated structure proposals based on experimental data, I view the chemical synthesis and subsequent analysis of candidate peptides as the most promising approach for structure elucidation. The Q, E and N residues warrant special attention, because the structures of intraszentins and xenolindicin vary in regard to those amino acids and their position in the peptide. When the structures of intraszentins A-C are elucidated, the pure compounds have to be tested for potential bioactivities, because of their promising structural properties and similarities to known antibiotics.

4.2 Protective Sphinganine from *Pseudomonas* MYb115

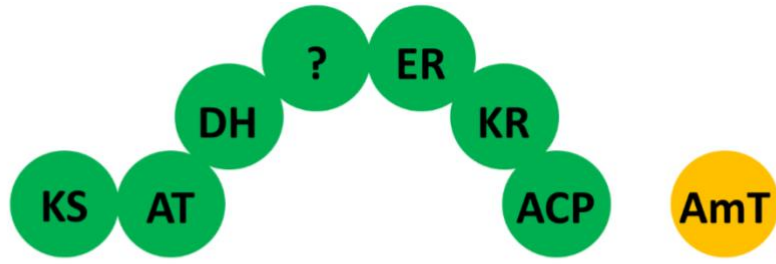
Pseudomonas fluorescens MYb115 protects *C. elegans*, its natural host organism, from negative effects, which are otherwise caused by infection with *Bacillus thuringiensis* BT247¹⁶⁴. Promoter-exchange and deletion strains of the *sgaAB* BGC in MYb115 led to the identification of three long chain sphinganine compounds that promote this effect. Together with our collaborators, Lena Peters and Dr. Katja Dierking, we showed that inactivation of said BGC abolishes the protecting effect of MYb115 on *C. elegans*, while targeted activation restored the effect (Fig. 24). The gene cluster encodes for a fully reducing type I polyketide synthase, which also contains a cryptic domain with unknown function, followed by a gene encoding for an aminotransferase (Fig. 23). Although I could not find a similar cluster in literature, the domain architecture gives some indications about a potential biosynthesis (Fig. 45).

Proposedly, the biosynthesis starts with long chain acyl-CoA that is subsequently elongated with shorter acyl-CoA units, most likely malonyl-CoA. Palmitoyl-CoA is a likely starter unit candidate, though palmitic acid has to be activated first. In *E. coli*, this reaction is catalysed by FadD, an acyl coenzyme A synthase¹⁹⁹. NCBI blastp on this protein against the translated genome sequence of MYb115 revealed that MYb115 contains at least three different versions of FadD, FadD1-3 (data not shown). The pathogen *P. aeruginosa* uses at least two different FadD enzymes, FadD1 and FadD2, to metabolise short and long chain fatty acids²⁰⁰. Hence, MYb115 probably has the ability to activate palmitic acid and other long chain fatty acids. Palmitoyl-CoA is subsequently loaded onto the T domain of the PKS and subsequently elongated in two to four cycles with malonyl-CoA (Fig. 45). Every elongation unit is fully reduced after each step (Fig. 45). The role of the cryptic domain has yet to be determined.

NCBI blastp on the region between revealed between the dehydratase (DH) and the enoyl reductase (ER) domains revealed similarities to the short-chain dehydrogenase/reductase (SDR) family. SDR family proteins are very diverse in function²⁰¹. Among them, epimerase- and dehydratase-activity on the serine residue have to be considered as possible functions of this cryptic domain. Alternatively, this domain could be an aminohydrolase domain, another member of the SDR family. Enzymes of this class typically convert hydroxyl moieties to amino moieties under the consumption of water and the formation of ammonia and vice

versa. An aminohydrolase is involved in the biosynthesis of the polyketide azalomycin F3a, where it catalyses the reaction from 4-guanidinobutyric acid amide to 4-guanidinobutyric acid. This reaction is part of a mechanism to convert arginine to 4-guanidinobutyryl-CoA, which is subsequently used as a starter unit²⁰². Hence, it remains in question if a fatty acid or a natural occurring amide is deployed as starter unit for sphinganine biosynthesis in MYb115.

A



B

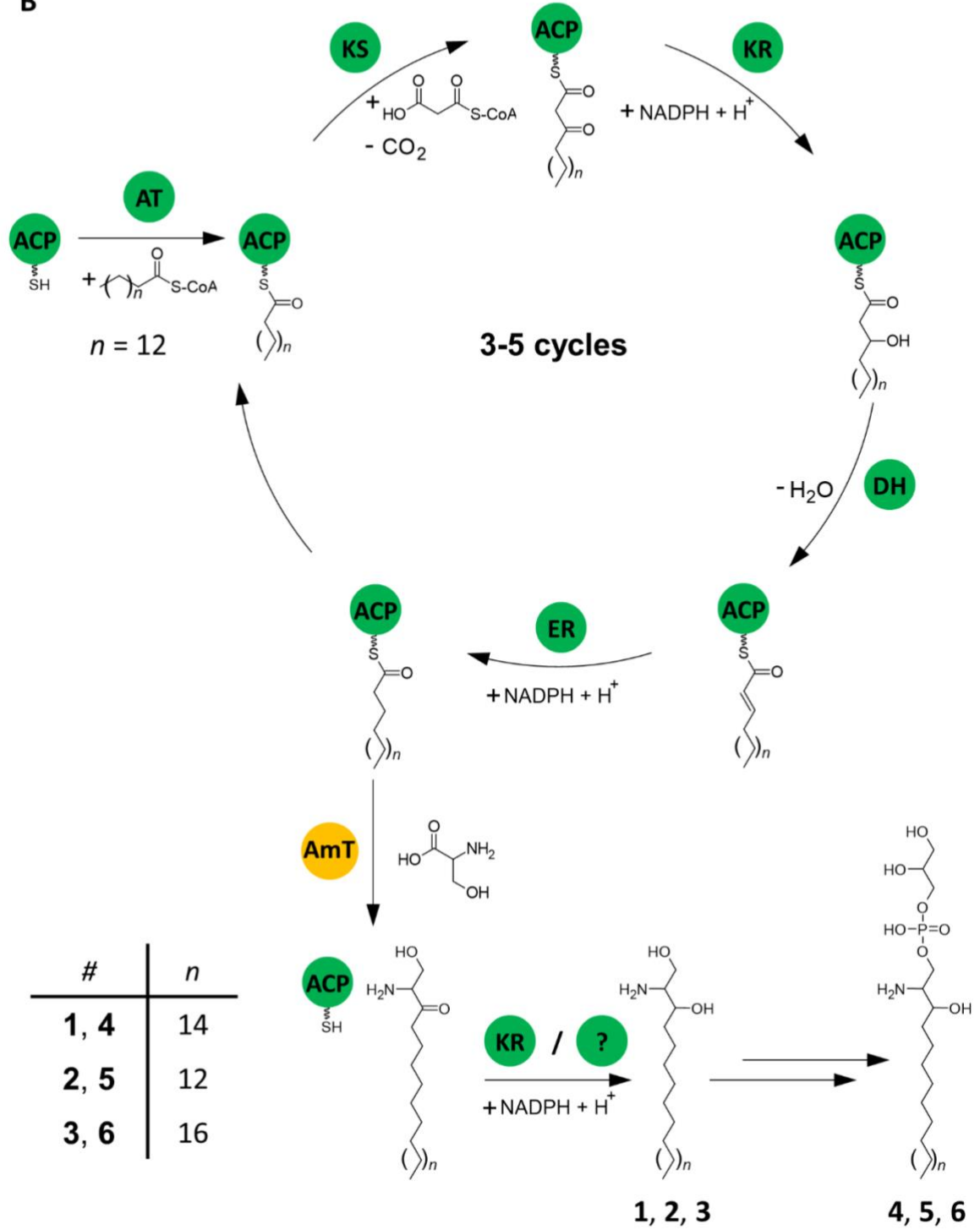
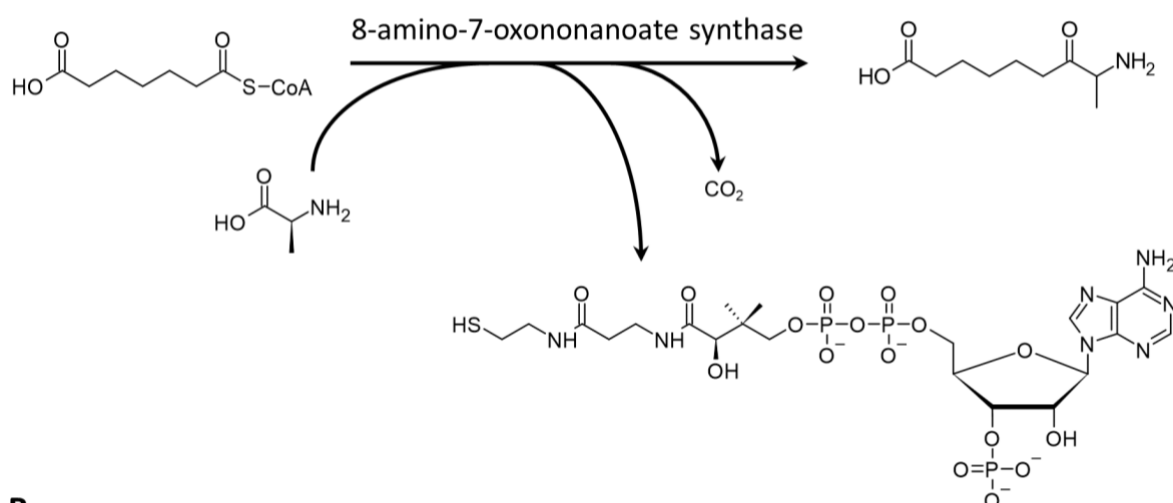


Figure 45. **A:** Ball scheme of the PKS SgaA and the aminotransferase SgaB. **B:** Possible biosynthesis scheme of long chain sphinganine **5**, **6** and **7**. ACP = acyl-carrier protein; AmT = aminotransferase; AT = acyl transferase; DH = dehydratase; ER = enoyl reductase; KR = ketoreductase; KS = ketosynthase; ? = cryptic domain. Only the domains responsible for the respective reactions are shown.

After reaching the final chain length, the aminotransferase domain catalyses the release of the long chain sphinganine product (Fig. 45). This domain is similar to 8-amino-7-oxononanoate synthases that catalyse the reaction from pimeloyl-CoA and L-alanine to 8-amino-7-oxononanoate and CO₂, thereby releasing coenzyme A. This reaction is very similar to the release of the sphinganine from the Ppant arm of the T domain by attaching the serine residue (Fig. 46).

A



B

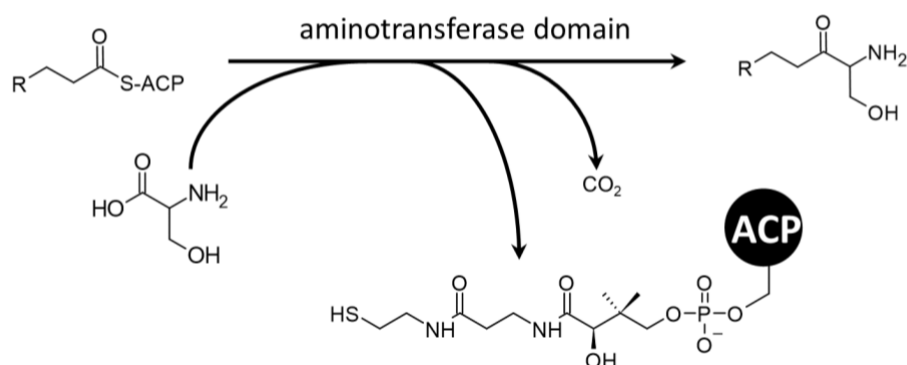


Figure 46. Similar reaction mechanisms of the 8-amino-7-oxononanoate synthase (**A**) and the proposed reaction of the aminotransferase domain of SgaAB (**B**). **A:** Pimeloyl-CoA and L-alanine react to 8-amino-7-oxononanoate, thereby releasing coenzyme A and CO₂. **B:** Acyl-ACP and serine react to a sphinganine, thereby releasing *holo*-ACP and CO₂.

Based on the assumption that palmitoyl-CoA is used as starter unit for this sphinganine biosynthesis, other long chain fatty acids can be used as well, if they can be activated by one of the FadD homologues in MYb115. Those fatty acids could be synthesized by the nematode host of MYb115. This could include *iso*-branched fatty acids²⁰³, as well as very long-chain fatty acids like behenic acid²⁰⁴, thus leading to a different product spectrum or less elongation cycles. Biochemical characterisation of this unusual iT1PKS is of great interest to understand the underlying mechanism of this hitherto undescribed route for sphinganine biosynthesis.

Furthermore, promoter exchange in front of *sgaAB* led to overproduction of compounds **5-7** (3.2.2.1), while production in MYb115 wild type samples was barely detectable under laboratory conditions (3.2.5.1). However, both strains, induced MYb115 $P_{BADsgaAB}$ and MYb115 wild type, displayed the same ability to protect *C. elegans* (3.2.2.2). This inspires two thoughts. First, the activation of the *sgaAB* BGC in nature must be triggered by the host or, if it is a direct response to infection with *B. thuringiensis*, by the pathogen. Either way, the underlying regulation needs to be investigated to understand how MYb115 protects its host. Second, we do not know how the sphinganines act in the protection mechanism. Although we found related PG-sphingolipids (3.2.4), their concentrations were very similar in the wild type and the induced promoter exchange strain, whereas sphinganine levels differed. Hence, we can conclude that sphinganine production is essential for the MYb115-mediated protection, but we can only speculate whether the related PG-sphingolipids play a role in the protection.

One conceivable way of protection involves a capsule formed by MYb115. Several genes encoding for capsule biosynthesis are encoded in close proximity to *sgaAB* (Fig. 47 and Tab. S7). In addition, we found many strains, mainly of the α -proteobacteria clade, that carry the *sga* BGC. We analysed three strains that live in close association with eukaryotic hosts. *Sinorhizobium meliloti* and *Methylorubrum extorquens* live in symbiotic relationships with plants and fixate nitrogen^{205,206}, while *Legionella longbeachae* is a human pathogen that has also been isolated from soil²⁰⁷. Strikingly, in all three cases, genes that encode for capsule biosynthesis related proteins can be found in close proximity to the *sga* BGC (Fig. 47 and Tab. S7). Along with capsule biosynthesis proteins and capsular polysaccharide biosynthesis

proteins, glycosyltransferases and SDR family oxidoreductases are among the most prominent gene annotations that we found in proximity to *sgaAB*. Both, glycosyltransferases and SDR family oxidoreductases play a key role in biosynthesis of capsular polysaccharides^{208,209}.

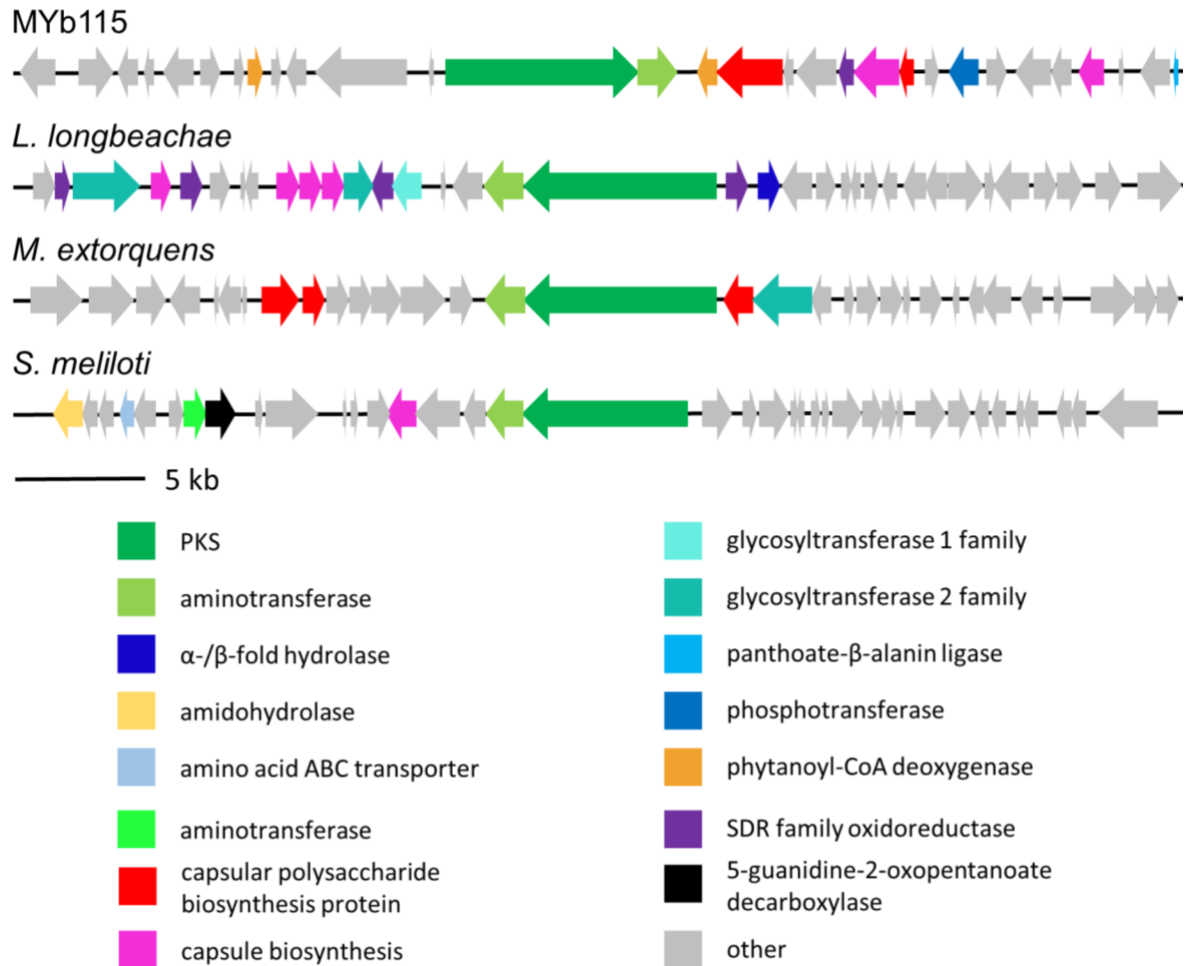


Figure 47. Biosynthesis-related genes in genomic proximity to *sgaAB*. 45 kb region centered around *sgaAB* was analysed in *Pseudomonas* MYb115, *Legionella longbeachae* NSW150, *Methylobacterium extorquens* AM1 and *Sinorhizobium meliloti* 2011 using antiSMASH¹⁷⁷. All biosynthesis-related genes were analysed using NCBI blastp. The resulting annotations for each gene are shown here.

The capsules of Gram-negative bacteria consist of lipopolysaccharides (LPSs), which are attached to the outer membrane with a lipid membrane anchor²¹⁰. The composition of the LPSs and the membrane anchors vary between species^{210–212}. Although microbial capsules are important virulence factors for pathogenic bacteria²⁰⁸, some are reported to facilitate the

maturation of the host immune system¹³ and to grant protection from inflammatory diseases²¹³.

All in all, we observed multiple reasons why the long chain sphinganes described in this work and their protective effect in *C. elegans* are probably due to an involvement in capsule biosynthesis. Firstly, the genomic context. *sgaAB* is encoded in close proximity to capsule biosynthesis related genes in all four strains, that we analysed (Fig. 47 and Tab. S7). It is well known that genes that are related to the same biological processes are encoded in proximity of one another²¹⁴. Secondly, microbial capsules are involved in other protection and pathogenicity mechanisms as stated above. Thirdly, we detected PG-sphingolipids in which the glycerol moiety might act as a linker between a polysaccharide and the sphinganes. Last but not least, all three deletion mutants as well as MYb115 *P_{BAD}sgaAB* without induction with arabinose exhibit an altered phenotype (data not shown; Lena Peters, personal communication). The colonies appear smaller and less mucoid after incubation on agar plates compared to the wild type strain.

All of the reasons above lead me to the conclusion that the long chain sphinganes (**5-7**) may act as lipid membrane anchors for the LPS capsule of MYb115 and other strains that carry the *sgaAB* BGC. This hypothesis should be tested by inactivating the nearby capsule biosynthesis genes in MYb115. Furthermore, the influence of inactivation of *sgaAB* in *Sinorhizobium meliloti*, *Methylobacterium extorquens* and *Legionella longbeachae* on the interaction with their respective host organisms has to be investigated.

Although a mechanism including the capsule is likely, a different one that directly involves sphinganes or phosphorylated sphingane derivatives cannot be eliminated from consideration, as sphingane derivatives show a variety of effects on eukaryotic cells^{215,216}. In addition, the sphingane derivative safringol alias *L-threo*-sphingane is a reversible inhibitor of *lyso*-sphingolipid protein kinase C and was subject of a clinical trial for cancer treatment^{217,218}.

The stereochemical properties remain the last missing piece to fully elucidate the structure of compounds **5-7**. In order to assess possible protective effects of **5-7** or their phosphorylated derivatives, their stereochemistry needs to be solved. This proves to be especially challenging as only the relative configuration of the four possible stereoisomers

(Fig. 48) can be determined by NMR. Total synthesis and subsequent comparative analysis of all four stereoisomers is necessary to determine the absolute configuration. Yet, this step is possible²¹⁹ but difficult to perform and commercial synthesis of the four possible compounds is very expensive.

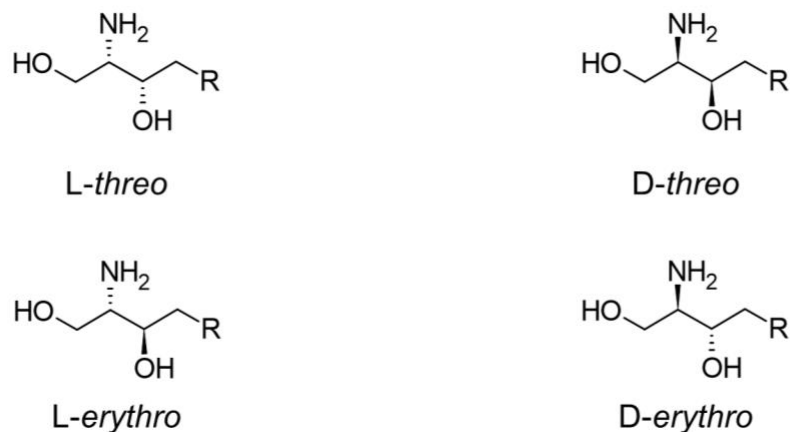


Figure 48. Possible sphinganine stereoisomers.

When the pure compound with defined stereochemical properties is available, either through synthesis or purification, we can test if it mediates protection of *C. elegans* against infection. A possible reason why MYb115 uses SgaAB instead of an established to produce these sphinganine derivatives might be the unusual stereochemistry of the product compared to the natural short chain *D-erythro*-sphingosine and -sphinganine found in eukaryotes. This can also explain why *C. elegans* is dependent on its microbiota to produce these compounds.

Neither of the two proposed mechanisms can be proven right nor wrong with the available experimental data. Hence, they should be extensively investigated as insights into microbiota-mediated protection of *C. elegans* possibly leads to microbiota based therapies for humans.

4.3 Antimicrobial peptides from *Pseudomonas* MYb11 and MYb12

I identified two natural products of the viscosin group, viscosin and massetolide E, from the natural *C. elegans* microbiota *Pseudomonas* MYb11 and MYb12. The structure of both natural products was elucidated using isotopic labelling experiments and LC-MS analyses.

4.4 GameXPeptides from *Xenorhabdus miraniensis*

GameXPeptides have been extensively studied in our group. They have been first described in 2012¹⁵⁵ and have been used as model natural products to study non-ribosomal peptide synthesis and most importantly to engineer NRPS that produce non-natural NPs^{222–224}. The respective NRPS, GxpS, produces a library of linear and cyclic pentapeptides in its original host, *Photorhabdus luminescens* TT01, but also when heterologously expressed in *E. coli*^{155,225}. GameXPeptides are widespread among *Photorhabdus* and *Xenorhabdus* bacteria⁴³. Tobias *et al.* analysed a large collection of *Photorhabdus* and *Xenorhabdus* strains on genomic and metabolic levels. Interestingly, they found that all five strains, in which they could find neither a *gxpS* gene nor a GameXPeptide-related natural product, produce xefoampeptides. The natural function of GameXPeptide and xefoampeptides (Fig. 50) remains hitherto unknown. Nonetheless, the fact that every single *Photorhabdus* and *Xenorhabdus* strain produces either one or the other indicates that GameXPeptides and xefoampeptides inherit similar biological functions.

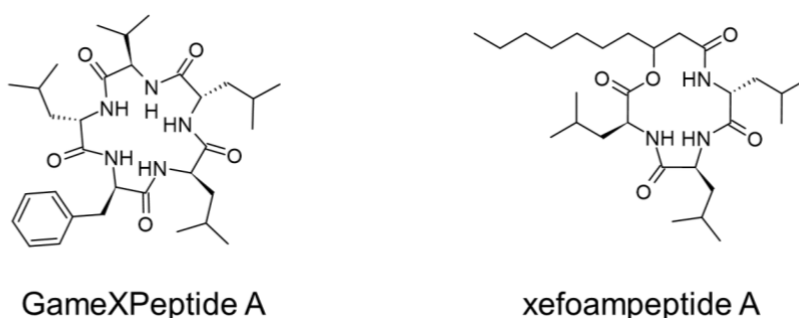


Figure 50. Structures of GameXPeptide A and xefoampeptide A.

In this work, I studied the natural products produced by the unusual GxpS of *X. miraniensis* DSM 17902. In comparison with the well-studied GxpSs of *P. luminescens* TT01 and *X. doucetiae* DSM 17909, this NRPS contains one fewer C_{Dual} domain (Fig. 36). GxpS_{Xmira} produces different GameXPeptide derivatives than GxpS_{TT01} and GxpS_{Xdou}, which is partly because of the changed domain architecture. As the generation of new NRPs is driven by evolution through recombination²²⁶, the altered *gxpS* BGC from *X. miraniensis* and the

previously known *gxps* BGCs from TT01 and *X. doucetiae* are closely related. Yet, their ancestry remains to be investigated.

The linkage of the *gxps* BGC of *X. miraniensis* to the production of the xenoinformycin NPF²²⁷ was the initial reason to study this cluster in detail. However, we observed no relation between this cluster and the production of xenoinformycins. While we observed the production of some xenoinformycins in *X. miraniensis*, it was not affected by promoter exchange in front of the *gxps* BGC (Fig. 51).

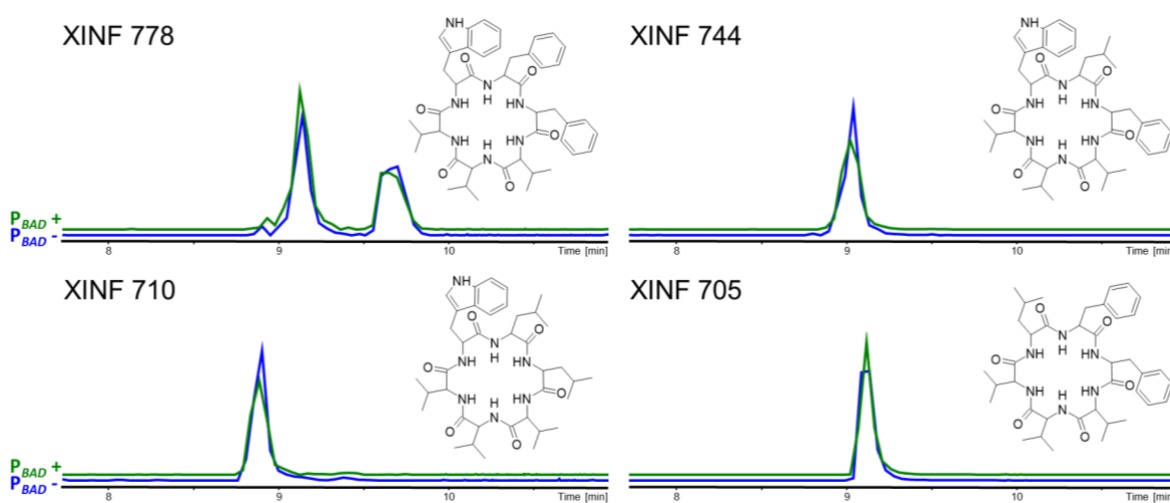


Figure 51. Comparison of production of xenoinformycins (XINF) in *X. miraniensis* $P_{BAD}gxps$ with arabinose added (green) and without (blue). Extracted ion chromatograms (EICs) of xenoinformycin (XINF) derivatives 778, 744, 710 and 705 that were identified by Behsaz *et al.* (2021)²²⁷.

Although, we could not find the origin of the xenoinformycins, we shed some light on the alternative GameXPeptides that are produced by *X. miraniensis* and elucidated their stereochemistry by comparative analysis of NPs and chemically synthesised candidate peptides by LC-MS.

We were able to solve the structure of **23** through comparison with synthetic peptides. Furthermore, I could deduce the structures of **21** and **22** from the structure of **23** as well as the known structures of the synthetic peptides and the domain architecture of GxpS_{Xmira}. The structures and peptide sequences that were confirmed experimentally, affirmed the flexibility of A domain specificities of A₁, A₂ and A₅ of GxpS_{Xmira} (Fig. 52).



Figure 52. Schematic image of the GameXPeptide synthetase GxpS_{Xmira}.

5 REFERENCES

1. Raaijmakers, J. M., Paulitz, T. C., Steinberg, C., Alabouvette, C. & Moënne-Loccoz, Y. The rhizosphere: a playground and battlefield for soilborne pathogens and beneficial microorganisms. *Plant Soil* 321, 341–361 (2009).
2. Deveau, A. *et al.* Bacterial–fungal interactions: ecology, mechanisms and challenges. *Fems Microbiol Rev* 42, 335–352 (2018).
3. Turnbaugh, P. J. *et al.* The Human Microbiome Project. *Nature* 449, 804–810 (2007).
4. Costello, E. K. *et al.* Bacterial Community Variation in Human Body Habitats Across Space and Time. *Science* 326, 1694–1697 (2009).
5. Fay, K. T. *et al.* The gut microbiome alters immunophenotype and survival from sepsis. *Faseb J* 33, 11258–11269 (2019).
6. Preiswerk, D., Walser, J.-C. & Ebert, D. Temporal dynamics of microbiota before and after host death. *Isme J* 12, 2076–2085 (2018).
7. Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nat Rev Microbiol* 19, 55–71 (2021).
8. Caza, M., Lépine, F. & Dozois, C. M. Secretion, but not overall synthesis, of catecholate siderophores contributes to virulence of extraintestinal pathogenic *Escherichia coli*. *Mol Microbiol* 80, 266–282 (2011).
9. Bearden, S. W., Fetherston, J. D. & Perry, R. D. Genetic organization of the yersiniabactin biosynthetic region and construction of avirulent mutants in *Yersinia pestis*. *Infect Immun* 65, 1659–1668 (1997).
10. Fischbach, M. A. *et al.* The pathogen-associated *iroA* gene cluster mediates bacterial evasion of lipocalin 2. *Proc National Acad Sci* 103, 16502–16507 (2006).
11. Correnti, C. & Strong, R. K. Mammalian Siderophores, Siderophore-binding Lipocalins, and the Labile Iron Pool*. *J Biol Chem* 287, 13524–13531 (2012).
12. Buffie, C. G. & Pamer, E. G. Microbiota-mediated colonization resistance against intestinal pathogens. *Nat Rev Immunol* 13, 790–801 (2013).
13. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. An Immunomodulatory Molecule of Symbiotic Bacteria Directs Maturation of the Host Immune System. *Cell* 122, 107–118 (2005).

14. Zipperer, A. *et al.* Human commensals producing a novel antibiotic impair pathogen colonization. *Nature* 535, 511–516 (2016).
15. Gorwitz, R. J. *et al.* Changes in the Prevalence of Nasal Colonization with *Staphylococcus aureus* in the United States, 2001–2004. *J Infect Dis* 197, 1226–1234 (2008).
16. Diekema, D. J. *et al.* Survey of Infections Due to *Staphylococcus* Species: Frequency of Occurrence and Antimicrobial Susceptibility of Isolates Collected in the United States, Canada, Latin America, Europe, and the Western Pacific Region for the SENTRY Antimicrobial Surveillance Program, 1997–1999. *Clin Infect Dis* 32, S114–S132 (2001).
17. DeLeo, F. R., Otto, M., Kreiswirth, B. N. & Chambers, H. F. Community-associated methicillin-resistant *Staphylococcus aureus*. *Lancet* 375, 1557–1568 (2010).
18. Geers, A. U., Buijs, Y., Strube, M. L., Gram, L. & Bentzon-Tilia, M. The natural product biosynthesis potential of the microbiomes of Earth – Bioprospecting for novel antimicrobial agents in the meta-omics era. *Comput Struct Biotechnology J* 20, 343–352 (2022).
19. Fonseca, V. G. *et al.* Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nat Commun* 1, 98 (2010).
20. Coghlan, A. Nematode genome evolution. *Wormbook* 1–15 (2005)
doi:10.1895/wormbook.1.15.1.
21. McFall-Ngai, M. *et al.* Animals in a bacterial world, a new imperative for the life sciences. *Proc National Acad Sci* 110, 3229–3236 (2013).
22. Thomas, G. M. & Poinar, G. O. *Xenorhabdus* gen. nov., a Genus of Entomopathogenic, Nematophilic Bacteria of the Family Enterobacteriaceae. *Int. J. Syst. Bacteriol.* 29, 352–360 (1979).
23. Tailliez, P., Pagès, S., Ginibre, N. & Boemare, N. New insight into diversity in the genus *Xenorhabdus*, including the description of ten novel species. *Int. J. Syst. Evol. Microbiol.* 56, 2805–2818 (2006).
24. Tailliez, P. *et al.* Phylogeny of *Photorhabdus* and *Xenorhabdus* based on universally conserved protein-coding sequences and implications for the taxonomy of these two genera. Proposal of new taxa: *X. vietnamensis* sp. nov., *P. luminescens* subsp. *caribbeanensis* subsp. nov., *P. luminescens* subsp. *hainanensis* subsp. nov., *P. temperata* subsp. *khanii* subsp. nov., *P. temperata* subsp. *tasmaniensis* subsp. nov., and the reclassification of *P. luminescens* subsp. *thracensis* as *P. temperata* subsp. *thracensis* comb. nov. *Int. J. Syst. Evol. Microbiol.* 60, 1921–1937 (2010).
25. Maneesakorn, P. *et al.* Phylogenetic and cophylogenetic relationships of entomopathogenic nematodes (Heterorhabditis: Rhabditida) and their symbiotic bacteria (*Photorhabdus*: Enterobacteriaceae). *Mol. Phylogenet. Evol.* 59, 271–280 (2011).

26. Sajnaga, E. & Kazimierczak, W. Evolution and taxonomy of nematode-associated entomopathogenic bacteria of the genera *Xenorhabdus* and *Photorhabdus*: an overview. *Symbiosis* 80, 1–13 (2020).
27. French-Constant, R. *et al.* *Photorhabdus*: towards a functional genomic analysis of a symbiont and pathogen. *FEMS Microbiology Reviews* 26, 433–456 (2003).
28. Ciche, T. A., Darby, C., Ehlers, R.-U., Forst, S. & Goodrich-Blair, H. Dangerous liaisons: The symbiosis of entomopathogenic nematodes and bacteria. *Biol Control* 38, 22–46 (2006).
29. Shi, Y.-M. & Bode, H. B. Chemical language and warfare of bacterial natural products in bacteria-nematode-insect interactions. *Nat Prod Rep* 92, fiw007 (2018).
30. Goodrich-Blair, H. & Clarke, D. J. Mutualism and pathogenesis in *Xenorhabdus* and *Photorhabdus*: two roads to the same destination. *Molecular Microbiology* 64, 260–268 (2007).
31. Kaya, H. K. & Gaugler, R. Entomopathogenic Nematodes. *Annu Rev Entomol* 38, 181–206 (1993).
32. Rasmann, S. *et al.* Recruitment of entomopathogenic nematodes by insect-damaged maize roots. *Nature* 434, 732–737 (2005).
33. Ciche, T. A. & Ensign, J. C. For the Insect Pathogen *Photorhabdus luminescens*, Which End of a Nematode Is Out? *Appl. Environ. Microbiol.* 69, 1890–1897 (2003).
34. Brown, S. E., Cao, A. T., Hines, E. R., Akhurst, R. J. & East, P. D. A Novel Secreted Protein Toxin from the Insect Pathogenic Bacterium *Xenorhabdus nematophila* *. *J Biol Chem* 279, 14595–14601 (2004).
35. Tobias, N. J. *et al.* Genome comparisons provide insights into the role of secondary metabolites in the pathogenic phase of the *Photorhabdus* life cycle. *BMC Genomics* 17, 537 (2016).
36. Richards, G. R. & Goodrich-Blair, H. Masters of conquest and pillage: *Xenorhabdus nematophila* global regulators control transitions from virulence to nutrient acquisition. *Cell Microbiol* 11, 1025–1033 (2009).
37. Forst, S. & Neilson, K. Molecular biology of the symbiotic-pathogenic bacteria *Xenorhabdus* spp. and *Photorhabdus* spp. *Microbiol Rev* 60, 21–43 (1996).
38. Akhurst, R. J. Morphological and Functional Dimorphism in *Xenorhabdus* spp., Bacteria Symbiotically Associated with the Insect Pathogenic Nematodes *Neoaplectana* and *Heterorhabditis*. *Microbiology* 121, 303–309 (1980).

39. Akhurst, R. J. & Boemare, N. E. A Numerical Taxonomic Study of the Genus *Xenorhabdus* (Enterobacteriaceae) and Proposed Elevation of the Subspecies of *X. nematophilus* to Species. *Microbiology* 134, 1835–1845 (1988).
40. Sugar, D. R. *et al.* Phenotypic variation and host interactions of *Xenorhabdus bovienii* SS-2004, the entomopathogenic symbiont of *Steinernema jolietii* nematodes. *Environ Microbiol* 14, 924–939 (2012).
41. Eckstein, S., Dominelli, N., Brachmann, A. & Heermann, R. Phenotypic Heterogeneity of the Insect Pathogen *Photorhabdus luminescens*: Insights into the Fate of Secondary Cells. *Appl Environ Microb* 85, (2019).
42. Cao, M. & Goodrich-Blair, H. *Xenorhabdus nematophila* bacteria shift from mutualistic to virulent Lrp-dependent phenotypes within the receptacles of *Steinernema carpocapsae* insect-infective stage nematodes. *Environ Microbiol* 22, 5433–5449 (2020).
43. Tobias, N. J. *et al.* Natural product diversity associated with the nematode symbionts *Photorhabdus* and *Xenorhabdus*. *Nat Microbiol* 1354, 82–35 (2017).
44. Dreyer, J., Malan, A. P. & Dicks, L. M. T. Bacteria of the Genus *Xenorhabdus*, a Novel Source of Bioactive Compounds. *Front Microbiol* 9, 3177 (2018).
45. Diggle, S. P. & Whiteley, M. Microbe Profile: *Pseudomonas aeruginosa*: opportunistic pathogen and lab rat. *Microbiology* 166, 30–33 (2020).
46. Elborn, J. S. Cystic fibrosis. *Lancet* 388, 2519–2531 (2016).
47. Dirksen, P. *et al.* The native microbiome of the nematode *Caenorhabditis elegans*: gateway to a new host-microbiome model. *BMC Biol.* 14, 38 (2016).
48. Novik, G., Savich, V. & Kiseleva, E. Microbiology in Agriculture and Human Health. (2015) doi:10.5772/60502.
49. Sampedro, I., Parales, R. E., Krell, T. & Hill, J. E. *Pseudomonas chemotaxis*. *Fems Microbiol Rev* 39, n/a-n/a (2014).
50. Sørensen, J. & Nybroe, O. *Pseudomonas*, Volume 1 Genomics, Life Style and Molecular Architecture. 369–401 (2011) doi:10.1007/978-1-4419-9086-0_12.
51. Liu, Y. *et al.* Diversity of Aquatic *Pseudomonas* Species and Their Activity against the Fish Pathogenic Oomycete *Saprolegnia*. *Plos One* 10, e0136241 (2015).
52. Centre, J. R. *et al.* *European atlas of soil biodiversity*. (Publications Office, 2010). doi:doi/10.2788/94222.
53. Frézal, L. & Félix, M.-A. *C. elegans* outside the Petri dish. *Elife* 4, e05849 (2015).

54. Gross, H. & Loper, J. E. Genomics of secondary metabolite production by *Pseudomonas* spp. *Nat Prod Rep* 26, 1408–1446 (2009).
55. Pan, S.-Y. *et al.* Historical Perspective of Traditional Indigenous Medical Practices: The Current Renaissance and Conservation of Herbal Resources. *Evid-based Compl Alt* 2014, 525340 (2014).
56. Waldvogel, S. R. Caffeine—A Drug with a Surprise. *Angew. Chem. Int. Ed.* 42, 604–605 (2003).
57. Huxtable, R. J. & Schwarz, S. K. The isolation of morphine--first principles in science and ethics. *Mol Interv* 1, 189–91 (2001).
58. O'Brien, R. D. Poisons of Plant Origin. 157–178 (1974) doi:10.1007/978-1-4684-2943-5_4.
59. 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* 3, 226–236 (1929).
60. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J Nat Prod* 83, 770–803 (2020).
61. Capecchi, A. & Reymond, J.-L. Classifying natural products from plants, fungi or bacteria using the COCONUT database and machine learning. *J Cheminformatics* 13, 82 (2021).
62. Bode, H. B. Entomopathogenic bacteria as a source of secondary metabolites. *Current Opinion in Chemical Biology* 13, 224–230 (2009).
63. Challis, G. L. & Hopwood, D. A. Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. *Proc National Acad Sci* 100, 14555–14561 (2003).
64. Donald, L. *et al.* *Streptomyces*: Still the Biggest Producer of New Natural Secondary Metabolites, a Current Perspective. *Microbiol Res* 13, 418–465 (2022).
65. Bérdy, J. Thoughts and facts about antibiotics: Where we are now and where we are heading. *J Antibiotics* 65, 385–395 (2012).
66. Joyce, S. A. *et al.* Bacterial biosynthesis of a multipotent stilbene. *Angew. Chem. Int. Ed. Engl.* 47, 1942–1945 (2008).
67. Akhurst, R. J. Antibiotic Activity of *Xenorhabdus* spp., Bacteria Symbiotically Associated with Insect Pathogenic Nematodes of the Families Heterorhabditidae and Steinernematidae. *Microbiology* 128, 3061–3065 (1982).

68. Webster, J., Li, J. & Hu, K. Nematicidal metabolites produced by *Photorhabdus luminescens* (Enterobacteriaceae), bacterial symbiont of entomopathogenic nematodes. *Nematology* 1, 457–469 (1999).
69. Abebew, D., Sayedain, F. S., Bode, E. & Bode, H. B. Uncovering Nematicidal Natural Products from *Xenorhabdus* Bacteria. *J Agr Food Chem* 70, 498–506 (2022).
70. Tobias, N. J. *et al.* *Photorhabdus*-nematode symbiosis is dependent on hfq-mediated regulation of secondary metabolites. *Environ Microbiol* 19, 119–129 (2016).
71. Neubacher, N. *et al.* Symbiosis, virulence and natural-product biosynthesis in entomopathogenic bacteria are regulated by a small RNA. *Nat Microbiol* 5, 1481–1489 (2020).
72. Bode, E. *et al.* Promoter Activation in Δ hfq Mutants as an Efficient Tool for Specialized Metabolite Production Enabling Direct Bioactivity Testing. *Angew. Chem. Int. Ed.* 58, 18957–18963 (2019).
73. Gulsen, S. H. *et al.* Antiprotozoal activity of different *Xenorhabdus* and *Photorhabdus* bacterial secondary metabolites and identification of bioactive compounds using the easyPACId approach. *Sci Rep-uk* 12, 10779 (2022).
74. Cai, X. *et al.* Entomopathogenic bacteria use multiple mechanisms for bioactive peptide library design. *Nature Chemistry* 9, 379–386 (2016).
75. Reimer, D., Luxenburger, E., Brachmann, A. O. & Bode, H. B. A New Type of Pyrrolidine Biosynthesis Is Involved in the Late Steps of Xenocoumacin Production in *Xenorhabdus nematophila*. *Chembiochem* 10, 1997–2001 (2009).
76. Reimer, D., Pos, K. M., Thines, M., Grün, P. & Bode, H. B. A natural prodrug activation mechanism in nonribosomal peptide synthesis. *Nat. Chem. Biol.* 7, 888–890 (2011).
77. Wenski, S. L., Kolbert, D., Grammbitter, G. L. C. & Bode, H. B. Fabclavine biosynthesis in *X. szentirmaii*: shortened derivatives and characterization of the thioester reductase FclG and the condensation domain-like protein FclL. *J Ind Microbiol Biot* 46, 565–572 (2019).
78. Cimen, H. *et al.* Antifungal activity of different *Xenorhabdus* and *Photorhabdus* species against various fungal phytopathogens and identification of the antifungal compounds from *X. szentirmaii*. *Appl Microbiol Biotechnol* 105, 5517–5528 (2021).
79. Fuchs, S. W., Grundmann, F., Kurz, M., Kaiser, M. & Bode, H. B. Fabclavines: Bioactive Peptide–Polyketide–Polyamino Hybrids from *Xenorhabdus*. *Chembiochem* 15, 512–516 (2014).

80. Fuchs, S. W., Proschak, A., Jaskolla, T. W., Karas, M. & Bode, H. B. Structure elucidation and biosynthesis of lysine-rich cyclic peptides in *Xenorhabdus nematophila*. *Organic & Biomolecular Chemistry* 9, 3130–3132 (2011).
81. Gualtieri, M., Aumelas, A. & Thaler, J.-O. Identification of a new antimicrobial lysine-rich cyclolipopeptide family from *Xenorhabdus nematophila*. *J. Antibiot.* 62, 295–302 (2009).
82. McInerney, B. V., Taylor, W. C., Lacey, M. J., Akhurst, R. J. & Gregson, R. P. Biologically Active Metabolites from *Xenorhabdus* Spp., Part 2. Benzopyran-1-one Derivatives with Gastroprotective Activity. *J Nat Prod* 54, 785–795 (1991).
83. McInerney, B. V. *et al.* Biologically Active Metabolites from *Xenorhabdus* Spp., Part 1. Dithiolopyrrolone Derivatives with Antibiotic Activity. *J Nat Prod* 54, 774–784 (1991).
84. Furgani, G. *et al.* *Xenorhabdus* antibiotics: a comparative analysis and potential utility for controlling mastitis caused by bacteria. *J Appl Microbiol* 104, 745–758 (2008).
85. Lang, G., Kalvelage, T., Peters, A., Wiese, J. & Imhoff, J. F. Linear and Cyclic Peptides from the Entomopathogenic Bacterium *Xenorhabdus nematophilus*. *J Nat Prod* 71, 1074–1077 (2008).
86. Li, J., Chen, G. & Webster, J. M. Nematophin, a novel antimicrobial substance produced by *Xenorhabdus nematophilus* (Enterobacteriaceae). *Can J Microbiol* 43, 770–773 (1997).
87. Kennedy, G. *et al.* Studies on the novel anti-staphylococcal compound nematophin. *Bioorg Med Chem Lett* 10, 1751–1754 (2000).
88. Cai, X. *et al.* Biosynthesis of the Antibiotic Nematophin and Its Elongated Derivatives in Entomopathogenic Bacteria. *Org Lett* 19, 806–809 (2017).
89. Winsor, G. L. *et al.* Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic Acids Res* 44, D646–D653 (2016).
90. Paulsen, I. T. *et al.* Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat Biotechnol* 23, 873–878 (2005).
91. Bender, C. L., Alarcón-Chaidez, F. & Gross, D. C. *Pseudomonas syringae* Phytotoxins: Mode of Action, Regulation, and Biosynthesis by Peptide and Polyketide Synthetases. *Microbiol Mol Biol R* 63, 266–292 (1999).
92. Raaijmakers, J. M., Bruijn, I. de & Kock, M. J. D. de. Cyclic Lipopeptide Production by Plant-Associated *Pseudomonas* spp.: Diversity, Activity, Biosynthesis, and Regulation. *Mol Plant-microbe Interactions* 19, 699–710 (2006).

93. Götze, S. & Stallforth, P. Structure, properties, and biological functions of nonribosomal lipopeptides from pseudomonads. *Nat Prod Rep* 37, 29–54 (2019).
94. Feigin, A. M., Takemoto, J. Y., Wangspa, R., Teeter, J. H. & Brand, J. G. Properties of Voltage-gated Ion Channels Formed by Syringomycin E in Planar Lipid Bilayers. *J Membr Biology* 149, 41–47 (1996).
95. Becucci, L. *et al.* Channel-forming activity of syringomycin E in two mercury-supported biomimetic membranes. *Biochimica Et Biophysica Acta Bba - Biomembr* 1848, 932–941 (2015).
96. Arnold, D. L. & Preston, G. M. *Pseudomonas syringae*: enterprising epiphyte and stealthy parasite. *Microbiology+* 165, 251–253 (2018).
97. Nielsen, T. H., Christophersen, C., Anthoni, U. & Sørensen, J. Viscosinamide, a new cyclic depsipeptide with surfactant and antifungal properties produced by *Pseudomonas fluorescens* DR54. *J Appl Microbiol* 87, 80–90 (1999).
98. Geudens, N. & Martins, J. C. Cyclic Lipodepsipeptides From *Pseudomonas* spp. - Biological Swiss-Army Knives. *Front Microbiol* 9, 1867 (2018).
99. Reder-Christ, K. *et al.* Model membrane studies for characterization of different antibiotic activities of lipopeptides from *Pseudomonas*. *Biochimica Et Biophysica Acta Bba - Biomembr* 1818, 566–573 (2012).
100. Alam, K. *et al.* Genome Mining of *Pseudomonas* Species: Diversity and Evolution of Metabolic and Biosynthetic Potential. *Molecules* 26, 7524 (2021).
101. Rodríguez-Beltrán, J., Elabed, H., Gaddour, K., Blázquez, J. & Rodríguez-Rojas, A. Simple DNA transformation in *Pseudomonas* based on the Yoshida effect. *J Microbiol Meth* 89, 95–98 (2012).
102. CRICK, F. H. On protein synthesis. *Sym Soc Exp Biol* 12, 138–63 (1958).
103. Mach, B., Reich, E. & Tatum, E. L. SEPARATION OF THE BIOSYNTHESIS OF THE ANTIBIOTIC POLYPEPTIDE TYROCIDINE FROM PROTEIN BIOSYNTHESIS*. *Proc National Acad Sci* 50, 175–181 (1963).
104. BERG, T., FROHOLM, L. & LALAND, S. THE BIOSYNTHESIS OF GRAMICIDIN S IN A CELL-FREE SYSTEM. *Biochem J* 96, 43–52 (1965).
105. Sieber, S. A. & Marahiel, M. A. Molecular Mechanisms Underlying Nonribosomal Peptide Synthesis: Approaches to New Antibiotics. *Chem Rev* 105, 715–738 (2005).
106. Challis, G. L. & Naismith, J. H. Structural aspects of non-ribosomal peptide biosynthesis. *Curr Opin Struc Biol* 14, 748–756 (2004).

107. Fischbach, M. A. & Walsh, C. T. Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. *Chem Rev* 106, 3468–3496 (2006).
108. Wenzel, S. C., Meiser, P., Binz, T. M., Mahmud, T. & Müller, R. Nonribosomal Peptide Biosynthesis: Point Mutations and Module Skipping Lead to Chemical Diversity. *Angew. Chem. Int. Ed.* 45, 2296–2301 (2006).
109. Cai, X. *et al.* Entomopathogenic bacteria use multiple mechanisms for bioactive peptide library design. *Nature Chemistry* 9, 379–386 (2017).
110. Süßmuth, R. D. & Mainz, A. Nonribosomal Peptide Synthesis—Principles and Prospects. *Angew. Chem. Int. Ed.* 56, 3770–3821 (2017).
111. Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* 6, 493–505 (1999).
112. Röttig, M. *et al.* NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res* 39, W362–W367 (2011).
113. Lambalot, R. H. *et al.* A new enzyme superfamily - the phosphopantetheinyl transferases. *Chemistry & Biology* 3, 923–936 (1996).
114. 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* 12, 1189–1200 (2005).
115. Mullooney, M. W., McClure, R. A., Robey, M. T., Kelleher, N. L. & Thomson, R. J. Natural products from thioester reductase containing biosynthetic pathways. *Nat Prod Rep* 35, 847–878 (2018).
116. Mootz, H. D., Schwarzer, D. & Marahiel, M. A. Ways of Assembling Complex Natural Products on Modular Nonribosomal Peptide Synthetases. *Chembiochem* 3, 490–504 (2002).
117. Rawlings, B. J. Biosynthesis of fatty acids and related metabolites. *Nat Prod Rep* 15, 275–308 (1998).
118. Hertweck, C. The Biosynthetic Logic of Polyketide Diversity. *Angew. Chem. Int. Ed.* 48, 4688–4716 (2009).
119. O’Hagan, D. Biosynthesis of fatty acid and polyketide metabolites. *Nat Prod Rep* 10, 593–624 (1993).
120. Chen, A., Re, R. N. & Burkart, M. D. Type II fatty acid and polyketide synthases: deciphering protein–protein and protein–substrate interactions. *Nat Prod Rep* 35, 1029–1045 (2018).

121. Sandmann, A. *et al.* A Type II Polyketide Synthase from the Gram-Negative Bacterium *Stigmatella aurantiaca* Is Involved in Aurachin Alkaloid Biosynthesis. *Angew. Chem. Int. Ed.* 46, 2712–2716 (2007).
122. Brachmann, A. O. *et al.* A type II polyketide synthase is responsible for anthraquinone biosynthesis in *Photobacterium luminescens*. *Chembiochem* 8, 1721–1728 (2007).
123. Yu, D., Xu, F., Zeng, J. & Zhan, J. Type III polyketide synthases in natural product biosynthesis. *Iubmb Life* 64, 285–295 (2012).
124. Austin, M. B. & Noel, J. P. The chalcone synthase superfamily of type III polyketide synthases. *Nat Prod Rep* 20, 79–110 (2002).
125. Moore, B. S. & Hopke, J. N. Discovery of a New Bacterial Polyketide Biosynthetic Pathway. *Chembiochem* 2, 35–38 (2001).
126. Pfeifer, V. *et al.* A Polyketide Synthase in Glycopeptide Biosynthesis THE BIOSYNTHESIS OF THE NON-PROTEINOGENIC AMINO ACID (S)-3,5-DIHYDROXYPHENYLGLYCINE*. *J Biol Chem* 276, 38370–38377 (2001).
127. Seshime, Y., Juvvadi, P. R., Fujii, I. & Kitamoto, K. Discovery of a novel superfamily of type III polyketide synthases in *Aspergillus oryzae*. *Biochem Bioph Res Co* 331, 253–260 (2005).
128. Chan, Y. A., Podevels, A. M., Kevany, B. M. & Thomas, M. G. Biosynthesis of polyketide synthase extender units. *Nat Prod Rep* 26, 90–114 (2008).
129. Coderre, P. E. & Earhart, C. F. The entD Gene of the Escherichia coli K12 Enterobactin Gene Cluster. *Microbiology+* 135, 3043–3055 (1989).
130. Nakano, M. M., Corbell, N., Besson, J. & Zuber, P. Isolation and characterization of *sfp*: a gene that functions in the production of the lipopeptide biosurfactant, surfactin, in *Bacillus subtilis*. *Molec. Gen. Genet.* 232, 313–321 (1992).
131. Sarowska, J. *et al.* Virulence factors, prevalence and potential transmission of extraintestinal pathogenic *Escherichia coli* isolated from different sources: recent reports. *Gut Pathog* 11, 10 (2019).
132. Frueh, D. P. *et al.* Dynamic thiolation–thioesterase structure of a non-ribosomal peptide synthetase. *Nature* 454, 903–906 (2008).
133. Arima, K., Kakinuma, A. & Tamura, G. Surfactin, a crystalline peptidelipid surfactant produced by *Bacillus subtilis*: Isolation, characterization and its inhibition of fibrin clot formation. *Biochem Bioph Res Co* 31, 488–494 (1968).

134. Goussous, S. A. *et al.* Structure of the Fundamental Lipopeptide Surfactin at the Air/Water Interface Investigated by Sum Frequency Generation Spectroscopy. *J Phys Chem B* 121, 5072–5077 (2017).
135. Kluge, B., Vater, J., Salnikow, J. & Eckart, K. Studies on the biosynthesis of surfactin, a lipopeptide antibiotic from *Bacillus subtilis* ATCC 21332. *Febs Lett* 231, 107–110 (1988).
136. Beld, J., Sonnenschein, E. C., Vickery, C. R., Noel, J. P. & Burkart, M. D. The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. *Nat Prod Rep* 31, 61–108 (2014).
137. Ciche, T. A., Bintrim, S. B., Horswill, A. R. & Ensign, J. C. A Phosphopantetheinyl Transferase Homolog Is Essential for *Photorhabdus luminescens* To Support Growth and Reproduction of the Entomopathogenic Nematode *Heterorhabditis bacteriophora*. *Journal of Bacteriology* 183, 3117–3126 (2001).
138. Duchaud, E. *et al.* The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. *Nat. Biotechnol.* 21, 1307–1313 (2003).
139. Drechsler, M. Analyse und Regulation der Naturstoff-Biosynthese in *Photorhabdus luminescens*. (Goethe-Universität, Frankfurt, 2018).
140. Finking, R. *et al.* Characterization of a New Type of Phosphopantetheinyl Transferase for Fatty Acid and Siderophore Synthesis in *Pseudomonas aeruginosa* *. *J Biol Chem* 277, 50293–50302 (2002).
141. Zhou, B., Xiao, J. F., Tuli, L. & Ransom, H. W. LC-MS-based metabolomics. *Mol Biosyst* 8, 470–481 (2011).
142. Louie, K. B. *et al.* Comprehensive Natural Products III. 263–306 (2020) doi:10.1016/b978-0-12-409547-2.14834-6.
143. Sargent, M. *Guide to achieving reliable quantitative LC-MS measurements*. (RSC Analytical Methods Committee, 2013).
144. Johnson, J. V., Yost, R. A., Kelley, P. E. & Bradford, D. C. Tandem-in-space and tandem-in-time mass spectrometry: triple quadrupoles and quadrupole ion traps. *Anal Chem* 62, 2162–2172 (1990).
145. Morris, H. R. *et al.* High Sensitivity Collisionally-activated Decomposition Tandem Mass Spectrometry on a Novel Quadrupole/Orthogonal-acceleration Time-of-flight Mass Spectrometer. *Rapid Commun. Mass Spectrom.* 10, 889–896 (1996).
146. Fiehn, O. Metabolomics – the link between genotypes and phenotypes. *Plant Mol Biol* 48, 155–171 (2002).

147. Chen, Z.-Z. *et al.* Nontargeted and Targeted Metabolomic Profiling Reveals Novel Metabolite Biomarkers of Incident Diabetes in African Americans. *Diabetes* 71, 2426–2437 (2022).
148. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* 34, 828–837 (2016).
149. Little, J. L., Cleven, C. D. & Brown, S. D. Identification of “Known Unknowns” Utilizing Accurate Mass Data and Chemical Abstracts Service Databases. *J Am Soc Mass Spectr* 22, 348–359 (2011).
150. *Analyzing Biomolecular Interactions by Mass Spectrometry*. vol. 1 (Wiley-VCH, 2015).
151. Kim, S., Rodgers, R. P. & Marshall, A. G. Truly “exact” mass: Elemental composition can be determined uniquely from molecular mass measurement at ~0.1mDa accuracy for molecules up to ~500Da. *Int J Mass Spectrom* 251, 260–265 (2006).
152. Badertscher, M., Bischofberger, K., Munk, M. E. & Pretsch, E. A Novel Formalism To Characterize the Degree of Unsaturation of Organic Molecules. *J Chem Inf Comp Sci* 41, 889–893 (2001).
153. Pretsch, E., Bühlmann, P. & Affolter, C. *Structure Determination of Organic Compounds*. (Springer, 2013). doi:<https://doi.org/10.1007/978-3-662-04201-4>.
154. Comisarow, M. B. & Marshall, A. G. The Early Development of Fourier Transform Ion Cyclotron Resonance (FT-ICR) Spectroscopy. *J. Mass Spectrom.* 31, 581–585 (1996).
155. Bode, H. B. *et al.* Determination of the Absolute Configuration of Peptide Natural Products by Using Stable Isotope Labeling and Mass Spectrometry. *Chemistry - A European Journal* 18, 2342–2348 (2012).
156. Roepstorff, P. & Fohlman, J. Letter to the editors. *Biomed Mass Spectrom* 11, 601–601 (1984).
157. Papayannopoulos, I. A. The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass Spectrom. Rev.* 14, 49–73 (1995).
158. Marfey, P. Determination of D-amino acids. II. Use of a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzene. *Carlsberg Res Commun* 49, 591 (1984).
159. Bhushan, R. & Brückner, H. Marfey’s reagent for chiral amino acid analysis: A review. *Amino Acids* 27, 231–247 (2004).
160. Thoma, S. & Schobert, M. An improved Escherichia coli donor strain for diparental mating. *FEMS Microbiol. Lett.* 294, 127–132 (2009).

161. Lengyel, K. *et al.* Description of four novel species of *Xenorhabdus*, family Enterobacteriaceae: *Xenorhabdus budapestensis* sp. nov., *Xenorhabdus ehlersii* sp. nov., *Xenorhabdus innexi* sp. nov., and *Xenorhabdus szentirmaii* sp. nov. *Syst. Appl. Microbiol.* 28, 115–122 (2005).
162. Somvanshi, V. S. *et al.* A novel species of *Xenorhabdus*, family Enterobacteriaceae: *Xenorhabdus indica* sp. nov., symbiotically associated with entomopathogenic nematode *Steinernema thermophilum* Ganguly and Singh, 2000. *Syst Appl Microbiol* 29, 519–525 (2006).
163. Hazir, S. *et al.* Two new Subspecies of *Photorhabdus luminescens*, Isolated from *Heterorhabditis bacteriophora* (Nematoda: Heterorhabditidae): *Photorhabdus luminescens* subsp. *kayaii* subsp. nov. and *Photorhabdus luminescens* subsp. *thracensis* subsp. nov. *Syst Appl Microbiol* 27, 36–42 (2004).
164. Kissoyan, K. A. B. *et al.* Natural *C. elegans* Microbiota Protects against Infection via Production of a Cyclic Lipopeptide of the Viscosin Group. *Current Biology* 29, 1–20 (2019).
165. Saux, M. F.-L., Viillard, V., Brunel, B., Normand, P. & Boemare, N. E. Polyphasic classification of the genus *Photorhabdus* and proposal of new taxa: *P. luminescens* subsp. *luminescens* subsp. nov., *P. luminescens* subsp. *akhurstii* subsp. nov., *P. luminescens* subsp. *laumondii* subsp. nov., *P. temperata* sp. nov., *P. temperata* subsp. *temperata* subsp. nov. and *P. asymbiotica* sp. nov. *Int. J. Syst. Bacteriol.* 49 Pt 4, 1645–1656 (1999).
166. Bode, E. *et al.* Simple “on-demand” production of bioactive natural products. *Chembiochem* 16, 1115–1119 (2015).
167. Green, M. R. & Sambrook, J. *Molecular Cloning: A laboratory Manual.* Cold Spring Harbor Laboratory Press (2012).
168. Fu, C., Donovan, W. P., Shikapwashya-Hasser, O., Ye, X. & Cole, R. H. Hot Fusion: an efficient method to clone multiple DNA fragments as well as inverted repeats without ligase. *PLoS ONE* 9, e115318 (2014).
169. Dower, W. J., Miller, J. F. & Ragsdale, C. W. High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res* 16, 6127–6145 (1988).
170. Brown, E. M. *et al.* Bacteroides-Derived Sphingolipids Are Critical for Maintaining Intestinal Homeostasis and Symbiosis. *Cell Host Microbe* 25, 668-680.e7 (2019).
171. Weber, T. *et al.* antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research* 43, W237–W243 (2015).
172. Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research* 47, W81–W87 (2019).

173. Zhou, Q. *et al.* Structure and biosynthesis of xenoamicins from entomopathogenic *Xenorhabdus*. *Chemistry* 19, 16772–16779 (2013).
174. Beiras-Fernandez, A., Vogt, F., Sodian, R. & Weis, F. Daptomycin: a novel lipopeptide antibiotic against Gram-positive pathogens. *Infect Drug Resist* 3, 95–101 (2010).
175. Ling, L. L. *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* 517, 455–459 (2015).
176. Kaneda, T. Iso- and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. *Microbiol Rev* 55, 288–302 (1991).
177. Blin, K. *et al.* antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res* 49, gkab335- (2021).
178. Li, Y. *et al.* Sphingolipids in marine microalgae: Development and application of a mass spectrometric method for global structural characterization of ceramides and glycosphingolipids in three major phyla. *Anal Chim Acta* 986, 82–94 (2017).
179. Harrison, P. J., Dunn, T. M. & Campopiano, D. J. Sphingolipid biosynthesis in man and microbes. *Nat Prod Rep* 35, 921–954 (2018).
180. Gerard, J. *et al.* Massetolides A–H, Antimycobacterial Cyclic Depsipeptides Produced by Two *Pseudomonads* Isolated from Marine Habitats. *J. Nat. Prod.* 60, 223–229 (1997).
181. Bruijn, I. de, Kock, M. J. D. de, Waard, P. de, Beek, T. A. van & Raaijmakers, J. M. Massetolide A biosynthesis in *Pseudomonas fluorescens*. *Journal of Bacteriology* 190, 2777–2789 (2008).
182. Nollmann, F. I. *et al.* Insect-specific production of new GameXPeptides in *Photobacterium luminescens* TTO1, widespread natural products in entomopathogenic bacteria. *Chembiochem* 16, 205–208 (2015).
183. Schimming, O., Fleischhacker, F., Nollmann, F. I. & Bode, H. B. Yeast Homologous Recombination Cloning Leading to the Novel Peptides Ambactin and Xenolindicin. *Chembiochem* 15, 1290–1294 (2014).
184. Bian, X., Plaza, A., Zhang, Y. & Müller, R. Luminmycins A–C, Cryptic Natural Products from *Photobacterium luminescens* Identified by Heterologous Expression in *Escherichia coli*. *J Nat Prod* 75, 1652–1655 (2012).
185. Fu, J. *et al.* Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat Biotechnol* 30, 440–446 (2012).
186. Bogart, J. W. & Bowers, A. A. Dehydroamino acids: chemical multi-tools for late-stage diversification. *Org Biomol Chem* 17, 3653–3669 (2019).

187. MacKintosh, R. W. *et al.* The cyanobacterial toxin microcystin binds covalently to cysteine-273 on protein phosphatase 1. *Febs Lett* 371, 236–240 (1995).
188. Patteson, J. B. *et al.* Structure and Function of a Dehydrating Condensation Domain in Nonribosomal Peptide Biosynthesis. *J Am Chem Soc* 144, 14057–14070 (2022).
189. Li, B. *et al.* Structure and Mechanism of the Lantibiotic Cyclase Involved in Nisin Biosynthesis. *Science* 311, 1464–1467 (2006).
190. Cheng, Y.-Q., Yang, M. & Matter, A. M. Characterization of a Gene Cluster Responsible for the Biosynthesis of Anticancer Agent FK228 in *Chromobacterium violaceum* No. 968. *Appl Environ Microb* 73, 3460–3469 (2007).
191. Tillett, D. *et al.* Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide–polyketide synthetase system. *Chem Biol* 7, 753–764 (2000).
192. Shukla, R. *et al.* Teixobactin kills bacteria by a two-pronged attack on the cell envelope. *Nature* 608, 390–396 (2022).
193. Müller, A. *et al.* Daptomycin inhibits cell envelope synthesis by interfering with fluid membrane microdomains. *Proc National Acad Sci* 113, E7077–E7086 (2016).
194. Huan, Y., Kong, Q., Mou, H. & Yi, H. Antimicrobial Peptides: Classification, Design, Application and Research Progress in Multiple Fields. *Front Microbiol* 11, 582779 (2020).
195. Schmidtchen, A., Pasupuleti, M. & Malmsten, M. Effect of hydrophobic modifications in antimicrobial peptides. *Adv Colloid Interfac* 205, 265–274 (2014).
196. Kapil, S. & Sharma, V. d-Amino acids in antimicrobial peptides: a potential approach to treat and combat antimicrobial resistance. *Can J Microbiol* 67, 119–137 (2021).
197. Méndez, C. & Salas, J. A. The role of ABC transporters in antibiotic-producing organisms: drug secretion and resistance mechanisms. *Res Microbiol* 152, 341–350 (2001).
198. Götze, S. & Stallforth, P. Structure elucidation of bacterial nonribosomal lipopeptides. *Org Biomol Chem* 18, 1710–1727 (2020).
199. Black, P. N., DiRusso, C. C., Metzger, A. K. & Heimert, T. L. Cloning, sequencing, and expression of the *fadD* gene of *Escherichia coli* encoding acyl coenzyme A synthetase. *J Biol Chem* 267, 25513–25520 (1992).
200. Kang, Y., Zarzycki-Siek, J., Walton, C. B., Norris, M. H. & Hoang, T. T. Multiple FadD Acyl-CoA Synthetases Contribute to Differential Fatty Acid Degradation and Virulence in *Pseudomonas aeruginosa*. *Plos One* 5, e13557 (2010).

201. Kavanagh, K. L., Jörnvall, H., Persson, B. & Oppermann, U. Medium- and short-chain dehydrogenase/reductase gene and protein families. *Cell Mol Life Sci* 65, 3895 (2008).
202. Ray, L. & Moore, B. S. Recent advances in the biosynthesis of unusual polyketide synthase substrates. *Nat Prod Rep* 33, 150–161 (2015).
203. Kniazeva, M., Crawford, Q. T., Seiber, M., Wang, C.-Y. & Han, M. Monomethyl Branched-Chain Fatty Acids Play an Essential Role in *Caenorhabditis elegans* Development. *Plos Biol* 2, e257 (2004).
204. Wang, F. *et al.* Saturated very long chain fatty acid configures glycosphingolipid for lysosome homeostasis in long-lived *C. elegans*. *Nat Commun* 12, 5073 (2021).
205. Mergaert, P. *et al.* Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*–legume symbiosis. *Proc National Acad Sci* 103, 5230–5235 (2006).
206. Sy, A., Timmers, A. C. J., Knief, C. & Vorholt, J. A. Methylotrophic Metabolism Is Advantageous for *Methylobacterium extorquens* during Colonization of *Medicago truncatula* under Competitive Conditions. *Appl Environ Microb* 71, 7245–7252 (2005).
207. Steele, T. W., Lanser, J. & Sangster, N. Isolation of *Legionella longbeachae* serogroup 1 from potting mixes. *Appl Environ Microb* 56, 49–53 (1990).
208. Willis, L. M. & Whitfield, C. Structure, biosynthesis, and function of bacterial capsular polysaccharides synthesized by ABC transporter-dependent pathways. *Carbohydr Res* 378, 35–44 (2013).
209. Kallberg, Y., Oppermann, U. & Persson, B. Classification of the short-chain dehydrogenase/reductase superfamily using hidden Markov models. *Febs J* 277, 2375–2386 (2010).
210. Whitfield, C., Wear, S. S. & Sande, C. Assembly of Bacterial Capsular Polysaccharides and Exopolysaccharides. *Annu Rev Microbiol* 74, 1–23 (2020).
211. Willis, L. M. *et al.* Conserved glycolipid termini in capsular polysaccharides synthesized by ATP-binding cassette transporter-dependent pathways in Gram-negative pathogens. *Proc National Acad Sci* 110, 7868–7873 (2013).
212. Liston, S. D., Ovchinnikova, O. G. & Whitfield, C. Unique lipid anchor attaches Vi antigen capsule to the surface of *Salmonella enterica* serovar Typhi. *Proc National Acad Sci* 113, 6719–6724 (2016).
213. Mazmanian, S. K., Round, J. L. & Kasper, D. L. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* 453, 620–625 (2008).
214. Tamames, J., Casari, G., Ouzounis, C. & Valencia, A. Conserved Clusters of Functionally Related Genes in Two Bacterial Genomes. *J Mol Evol* 44, 66–73 (1997).

215. Spiegel, S. & Milstien, S. Sphingosine 1-Phosphate, a Key Cell Signaling Molecule*. *J Biol Chem* 277, 25851–25854 (2002).
216. Gutner, U. A. & Shupik, M. A. The Role of Sphingosine-1-Phosphate in Neurodegenerative Diseases. *Russ J Bioorg Chem* 47, 1155–1171 (2021).
217. Coward, J. *et al.* Safingol (l-threo-sphinganine) induces autophagy in solid tumor cells through inhibition of PKC and the PI3-kinase pathway. *Autophagy* 5, 184–193 (2009).
218. Dickson, M. A. *et al.* A Phase I Clinical Trial of Safingol in Combination with Cisplatin in Advanced Solid Tumors. *Clin Cancer Res* 17, 2484–2492 (2011).
219. Hoffman, R. V. & Tao, J. A Synthesis of d-erythro- and l-threo-Sphingosine and Sphinganine Diastereomers via the Biomimetic Precursor 3-Ketosphinganine. *J Org Chem* 63, 3979–3985 (1998).
220. Reybroeck, W. *et al.* Cyclic Lipodepsipeptides Produced by *Pseudomonas* spp. Naturally Present in Raw Milk Induce Inhibitory Effects on Microbiological Inhibitor Assays for Antibiotic Residue Screening. *Plos One* 9, e98266 (2014).
221. Tran, H., Ficke, A., Asiimwe, T., Höfte, M. & Raaijmakers, J. M. Role of the cyclic lipopeptide massetolide A in biological control of *Phytophthora infestans* and in colonization of tomato plants by *Pseudomonas fluorescens*. *New Phytol* 175, 731–742 (2007).
222. Bozhüyük, K. A. J. *et al.* De novo design and engineering of non-ribosomal peptide synthetases. *Nat Chem* 10, 275–281 (2018).
223. Bozhüyük, K. A. J. *et al.* Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nat Chem* 11, 653–661 (2019).
224. Bozhueyuek, K. A. J., Watzel, J., Abbood, N. & Bode, H. B. Synthetic Zippers as an Enabling Tool for Engineering of Non-Ribosomal Peptide Synthetases**. *Angew. Chem. Int. Ed.* 60, 17531–17538 (2021).
225. Kegler, C. *et al.* Rapid Determination of the Amino Acid Configuration of Xenotetrapeptide. *Chembiochem* 15, 826–828 (2014).
226. Baunach, M., Chowdhury, S., Stallforth, P. & Dittmann, E. The Landscape of Recombination Events That Create Nonribosomal Peptide Diversity. *Mol Biol Evol* 38, msab015- (2021).
227. Behsaz, B. *et al.* Integrating genomics and metabolomics for scalable non-ribosomal peptide discovery. *Nat Commun* 12, 3225 (2021).

6 ATTACHMENTS

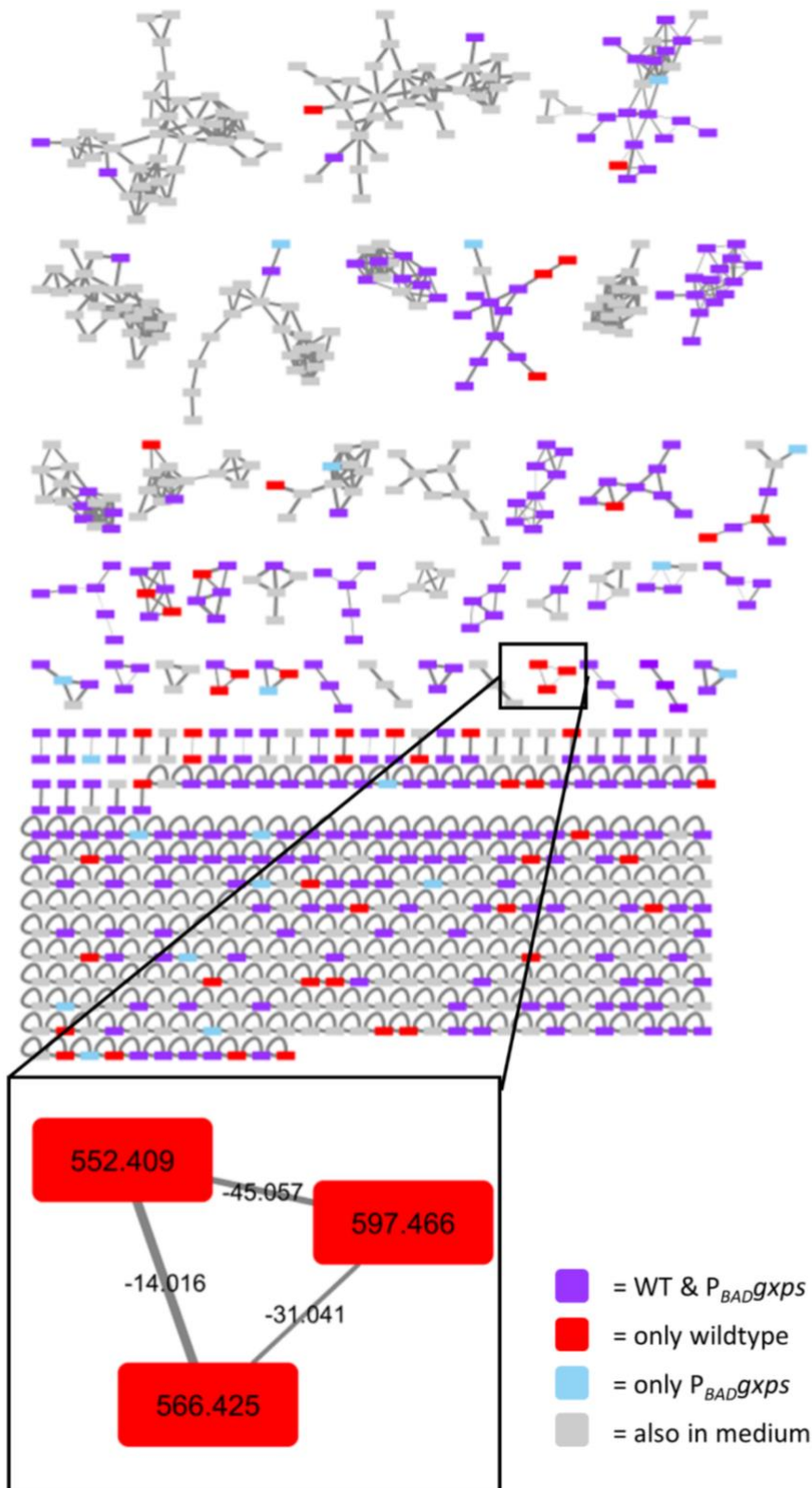


Figure S 1. GNPS network analysis comparing *X. szentirmaii* DSM wild type and the silenced promoter exchange strain of BGC *xsze00346* (MW94) with the enlarged GameXPeptide subnetwork.

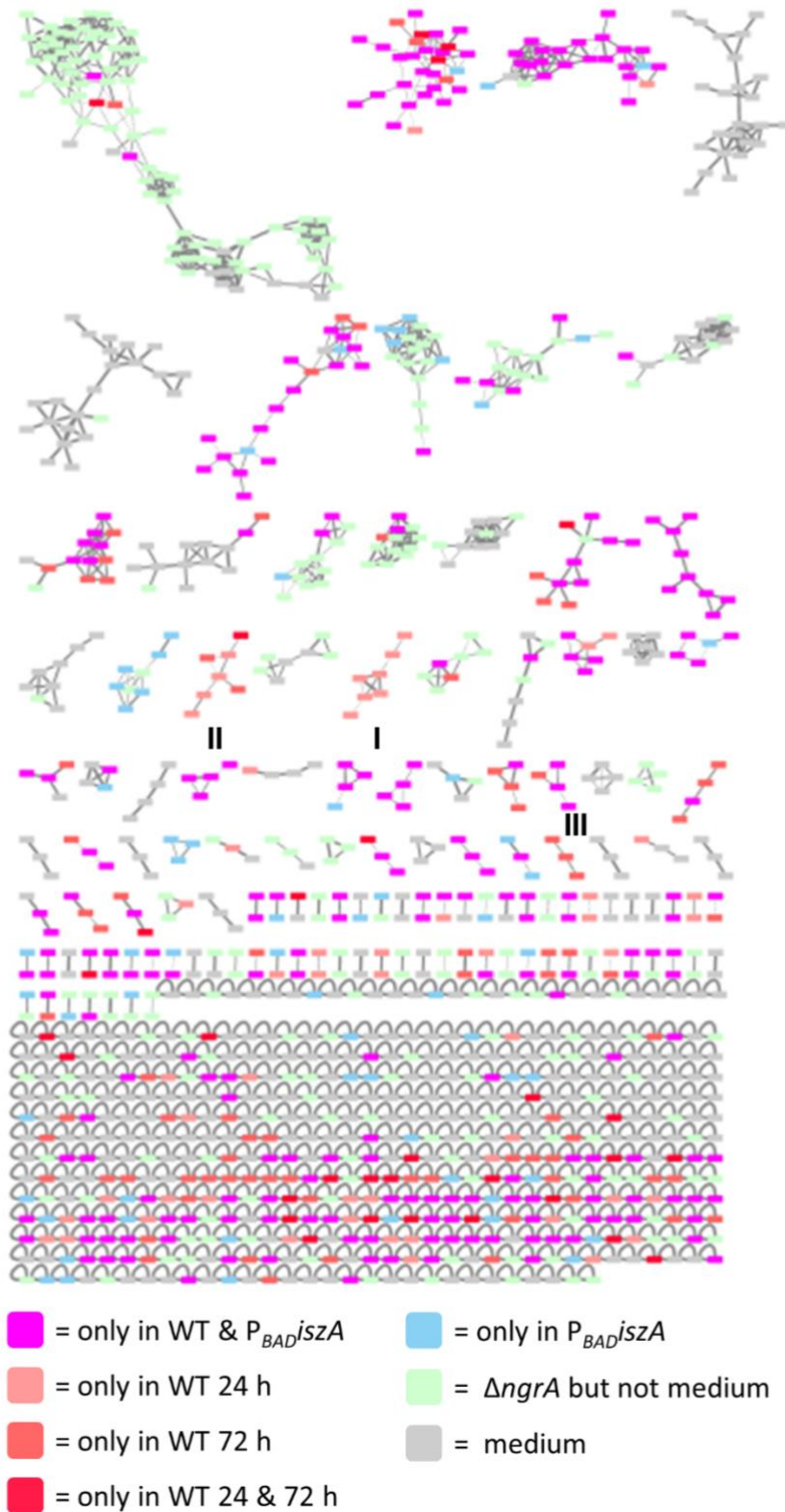


Figure S 2. Detailed view of the GNPS network analysis comparing *X. szentirmaii* DSM wild type, $\Delta ngrA$ and the silenced promoter exchange strain of BGC *xsze02985-xsze02984* with samples taken after 24 h and 72 h. Subnetworks I, II and III are shown in detail in figure S 3.

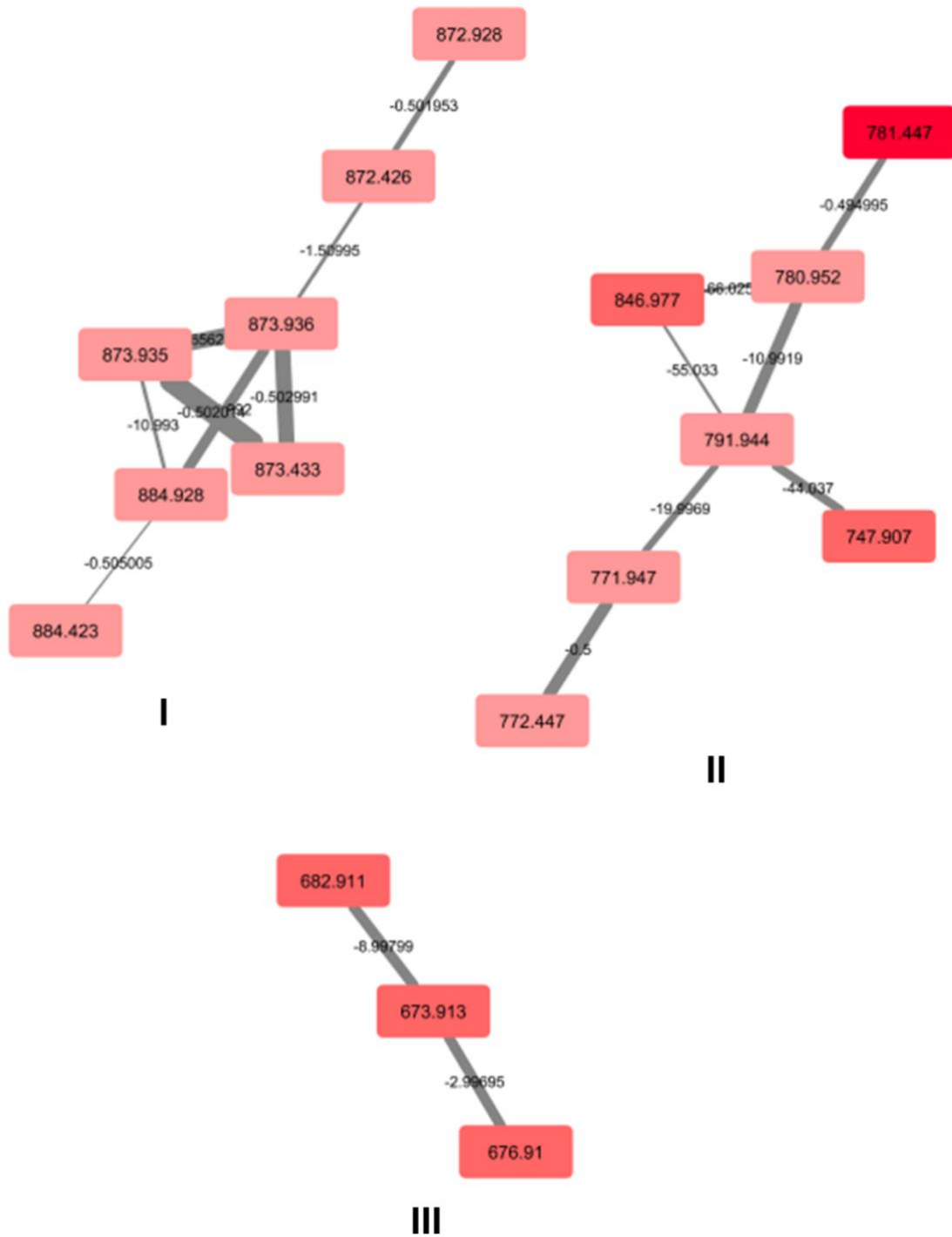


Figure S 3. Detailed view of subnetworks I, II and III from figure S 2.

Meas. m/z	Ion Formula	Score	err [ppm]
873.435944	C85H130N6O32	100.00	-0.0
	C83H118N20O22	89.59	-0.0
	C113H235N11	1.91	-0.2

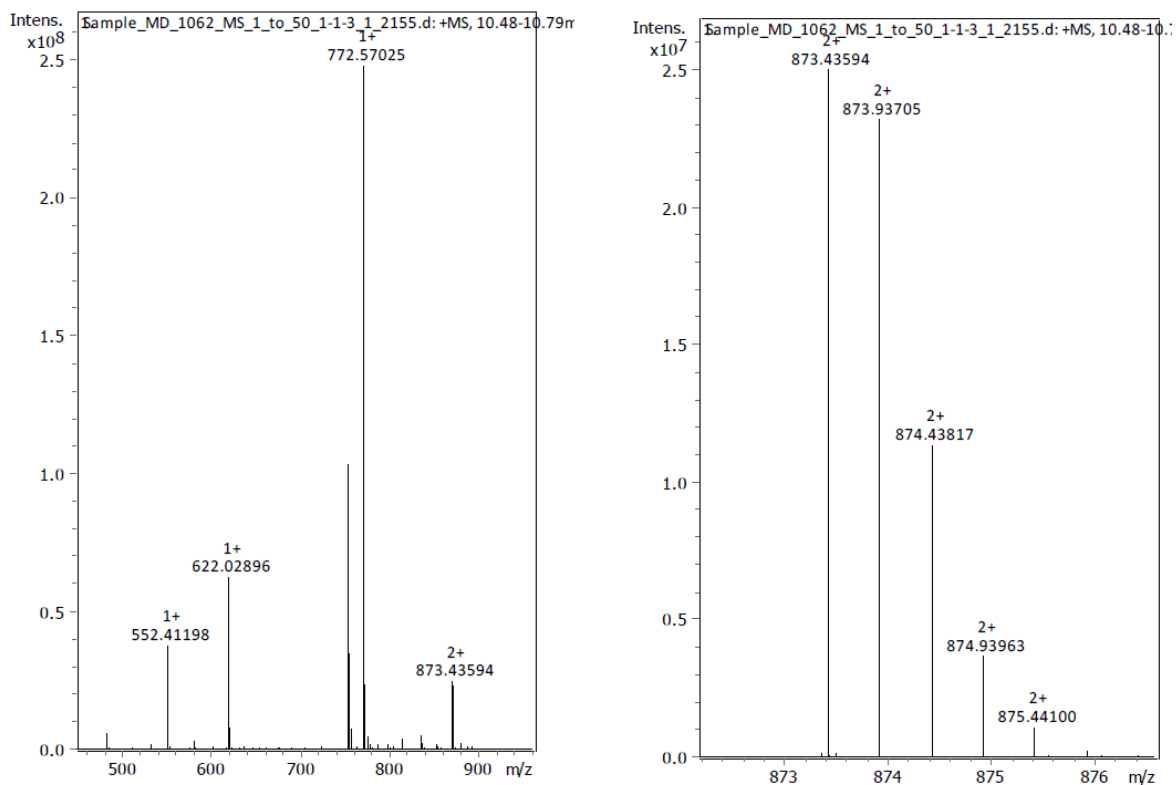


Figure S 4. MRMS data from Bruker on intraszentin A (1). Data was recorded and evaluated by Matthias Witt.

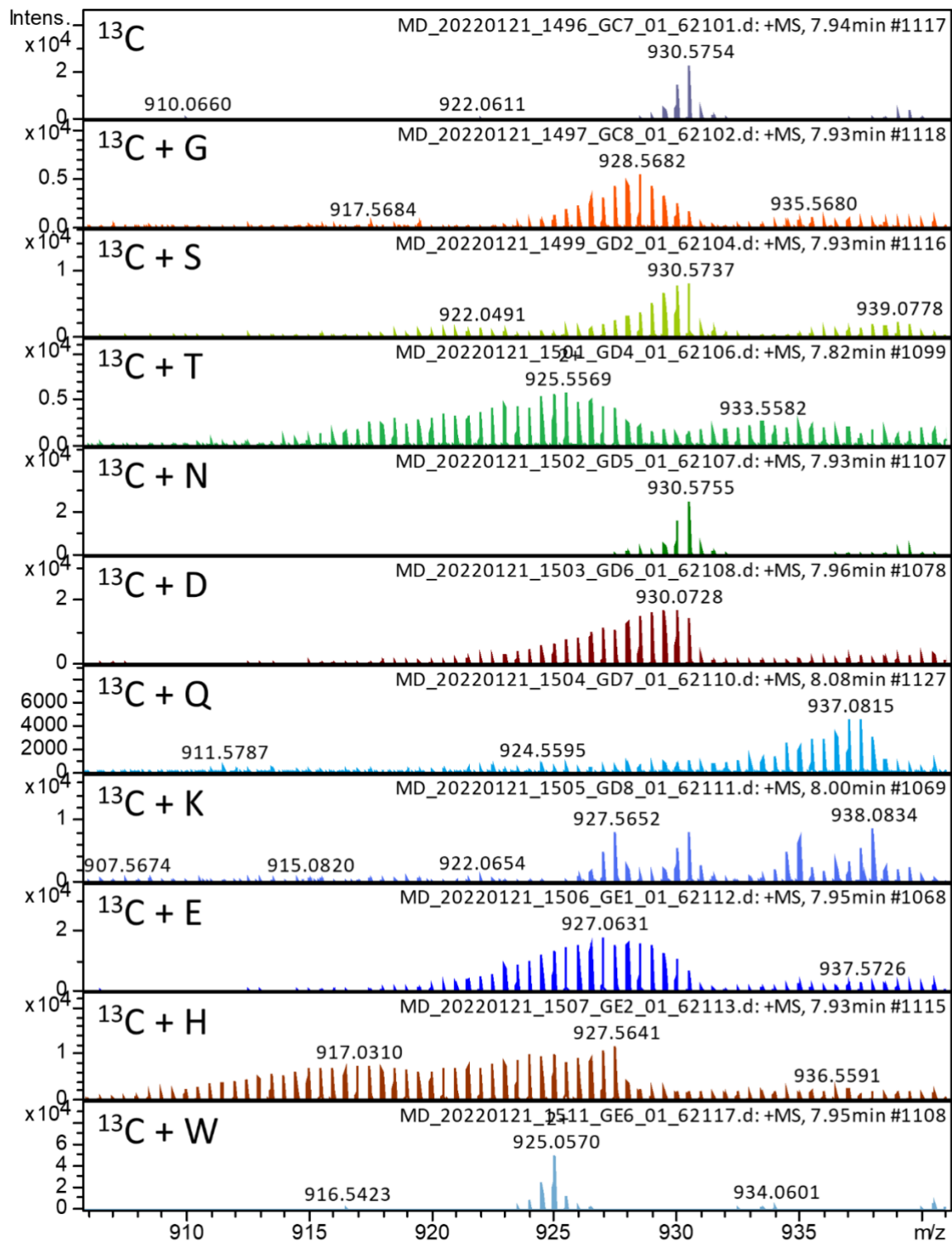


Figure S 5. Selected spectra resulting from inverse feeding of amino acids to *P. thracensis* cultures grown in ^{13}C -labeled medium.

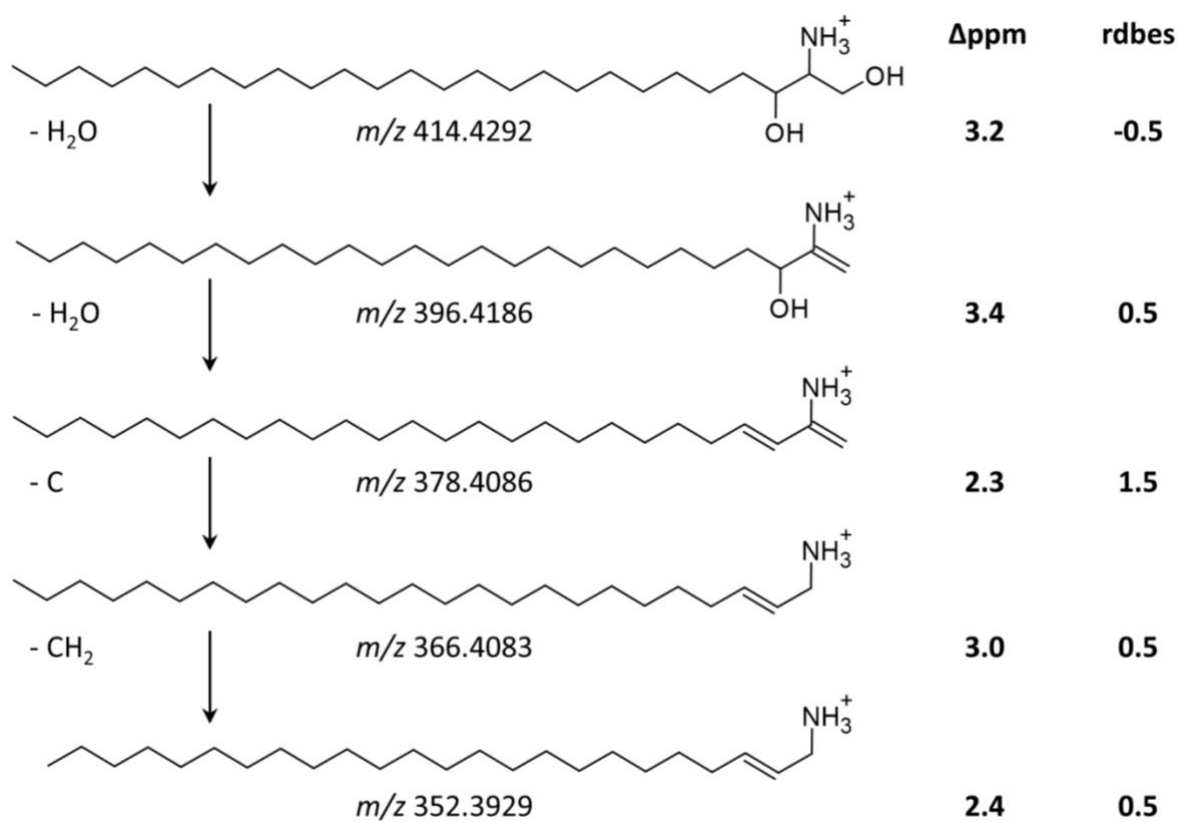
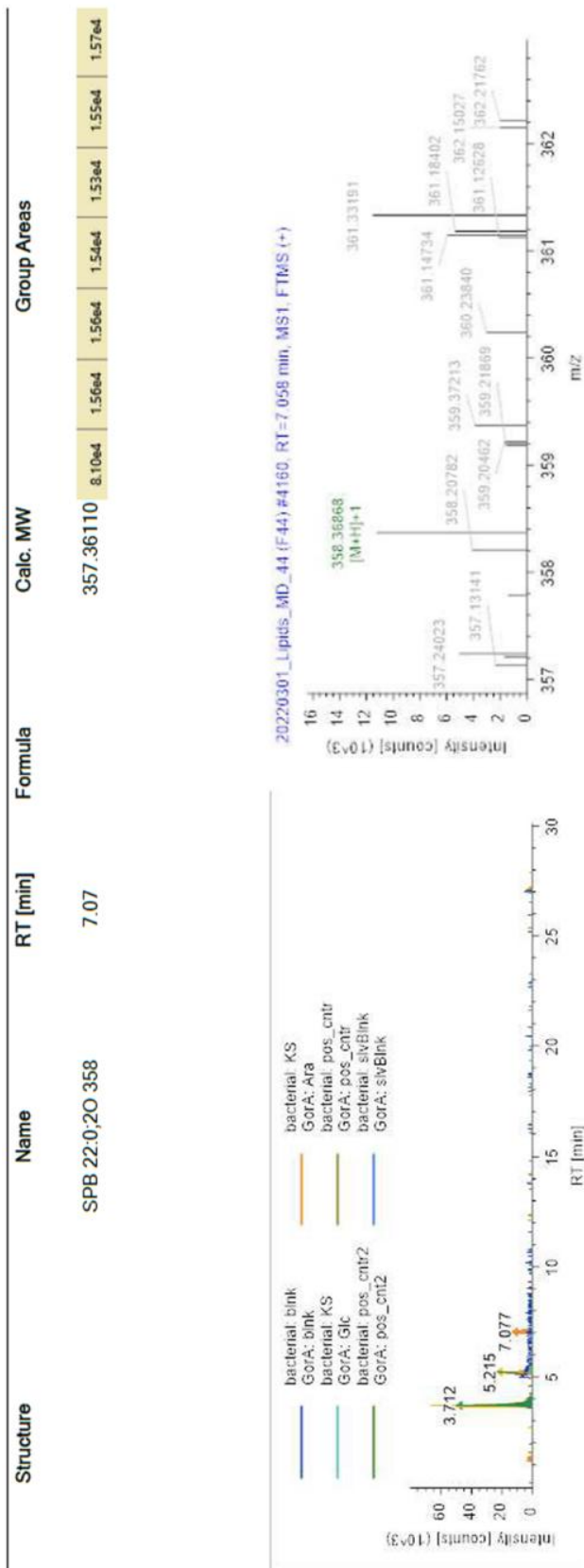


Figure S 6. Fragment ions of **5**.



© Reported with Compound Discoverer 3.3

Figure 7. LC chromatogram and mass spectrum of **8**. This graphic was created by Dr. Georgia Angelidou with Compound Discoverer 3.3.

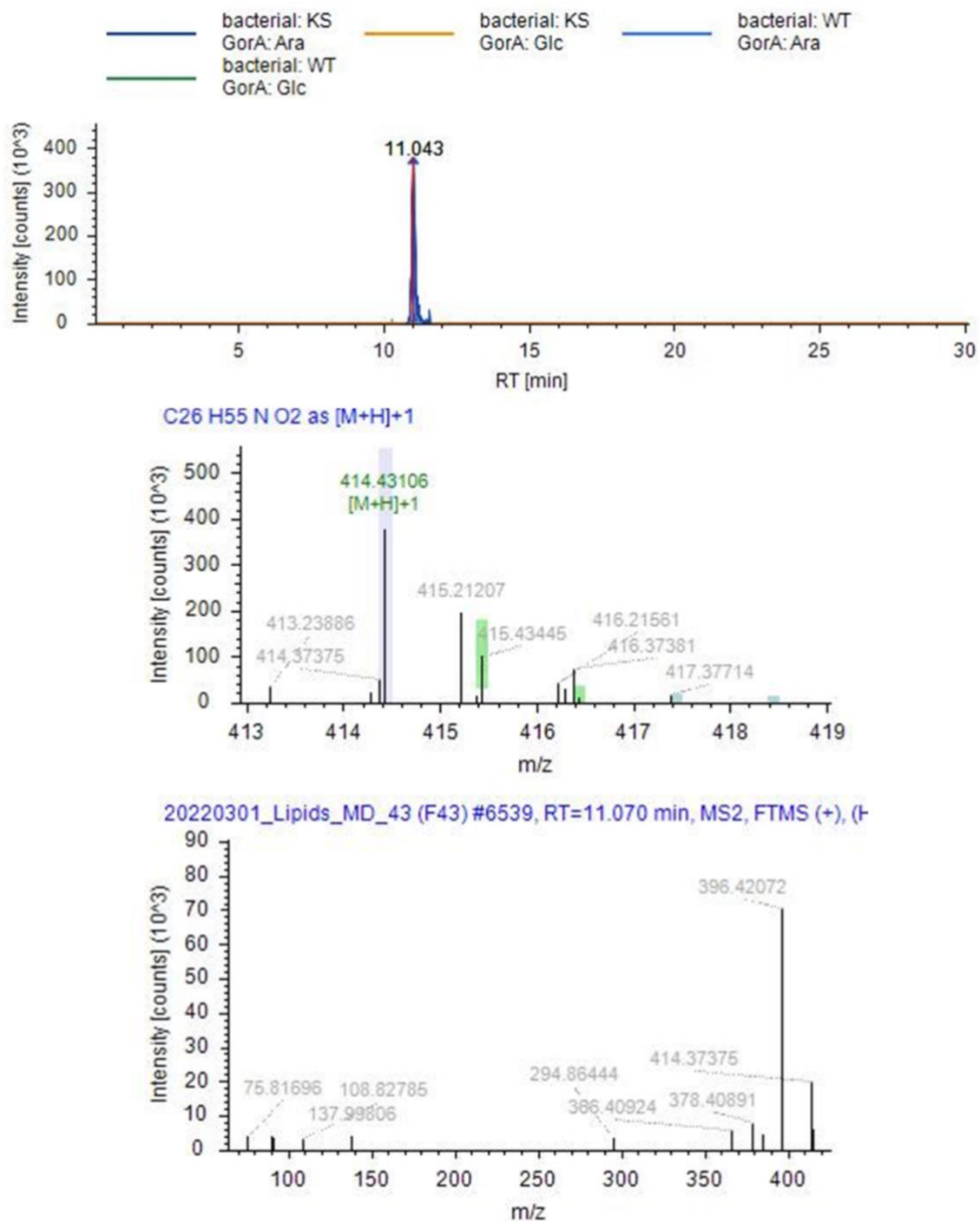
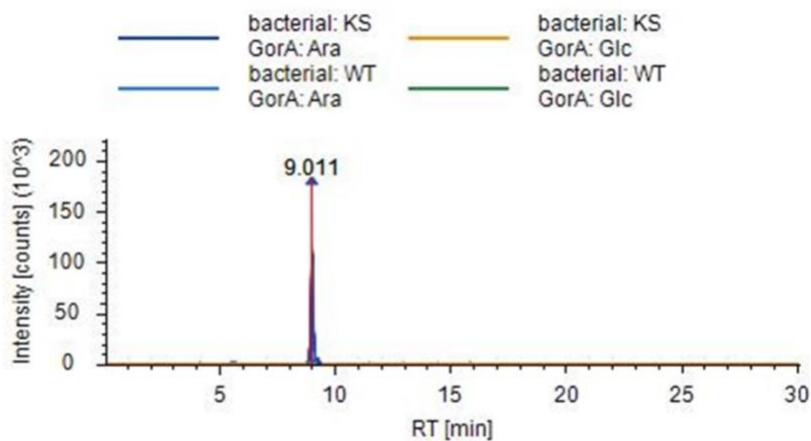
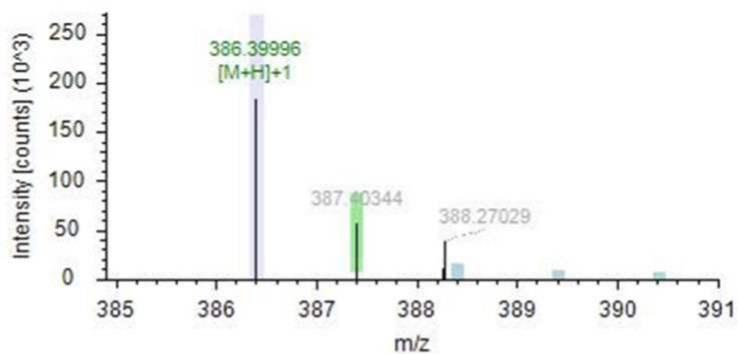


Figure S 7. LC chromatogram, MS¹ and MS² spectra of **5**. This graphic was created by Dr. Georgia Angelidou with Compound Discoverer 3.3.



20220301_Lipids_MD_45 (F45) #5337, RT=9.011 min, MS1, FTMS (+)
C24 H51 N O2 as [M+H]⁺1



20220301_Lipids_MD_45 (F45) #5321, RT=8.985 min, MS2, FTMS (+), (H)

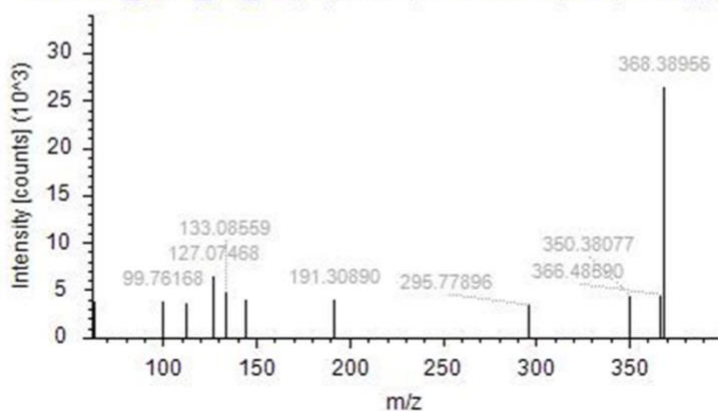


Figure S 8. LC chromatogram, MS¹ and MS² spectra of **6**. This graphic was created by Dr. Georgia Angelidou with Compound Discoverer 3.3.

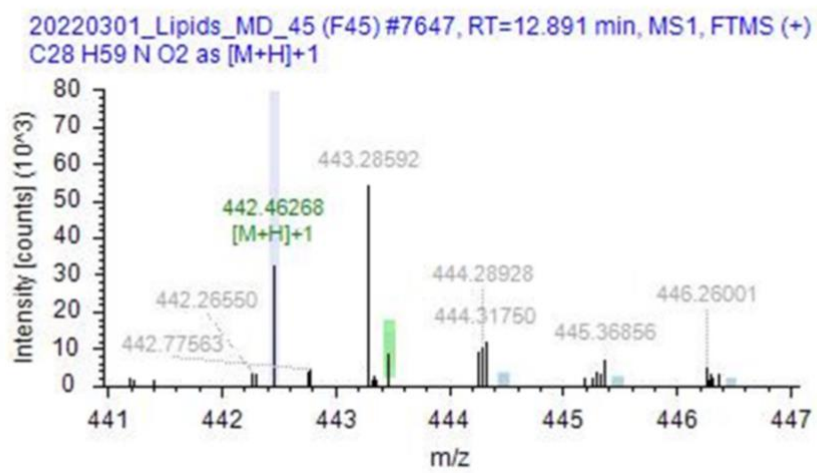
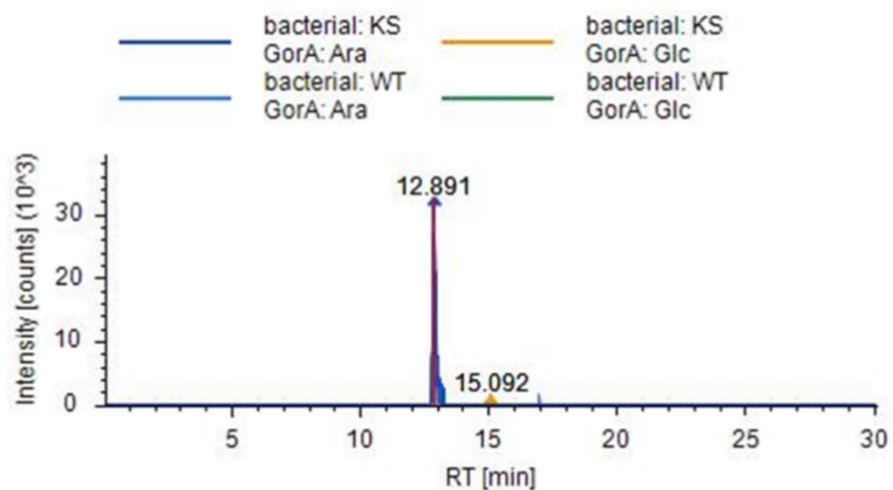


Figure S 9. LC chromatogram and MS¹ spectra of 7. No MS² spectra was obtained. This graphic was created by Dr. Georgia Angelidou with Compound Discoverer 3.3.

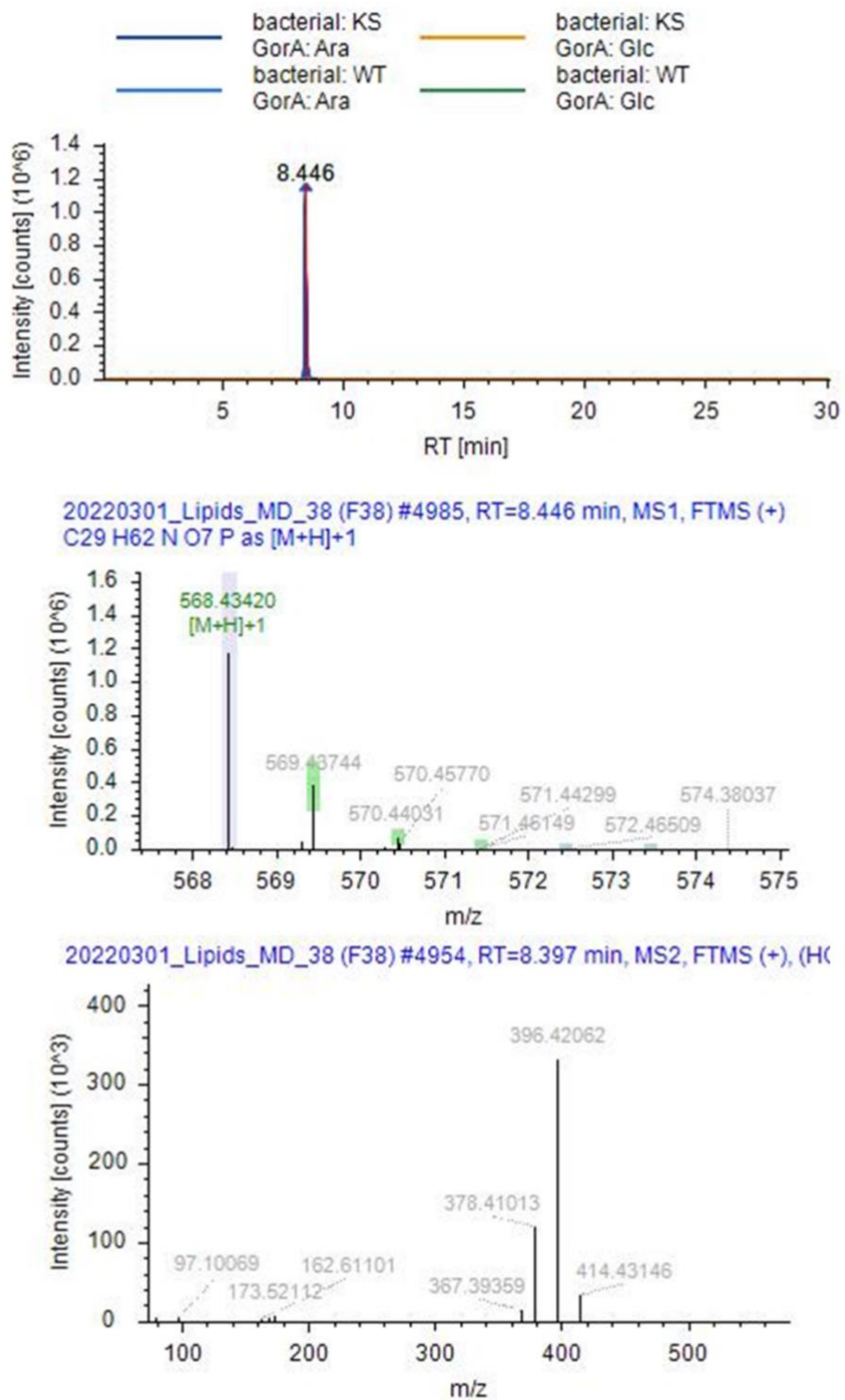
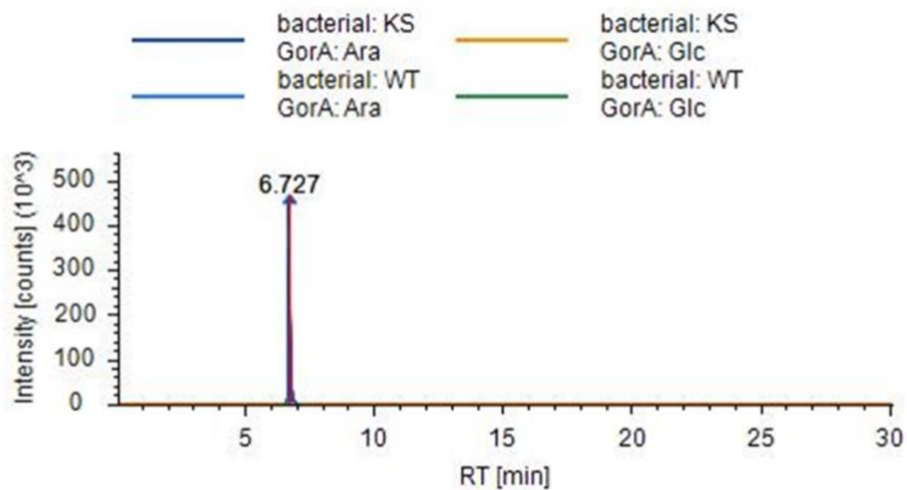
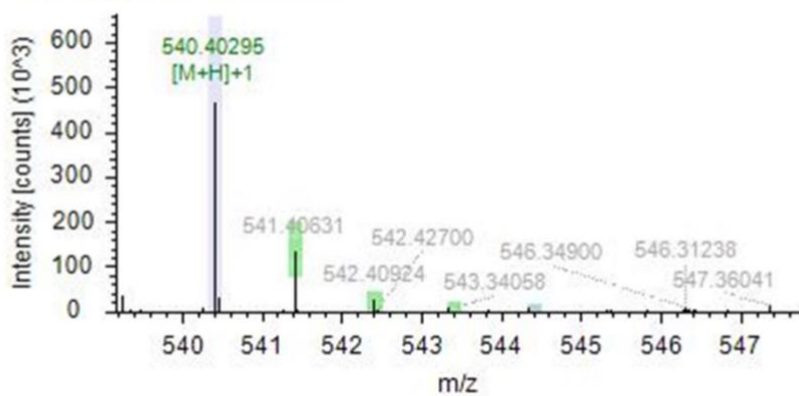


Figure S 10. LC chromatogram, MS¹ and MS² spectra of **9**. This graphic was created by Dr. Georgia Angelidou with Compound Discoverer 3.3.



20220301_Lipids_MD_38 (F38) #3962, RT=6.727 min, MS1, FTMS (+)
 C₂₇H₅₈N₀O₇P as [M+H]⁺1



20220301_Lipids_MD_38 (F38) #4007, RT=6.802 min, MS2, FTMS (+), (H

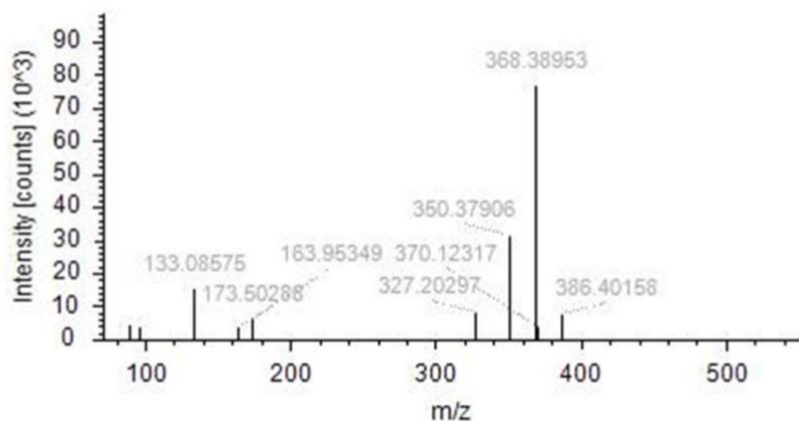
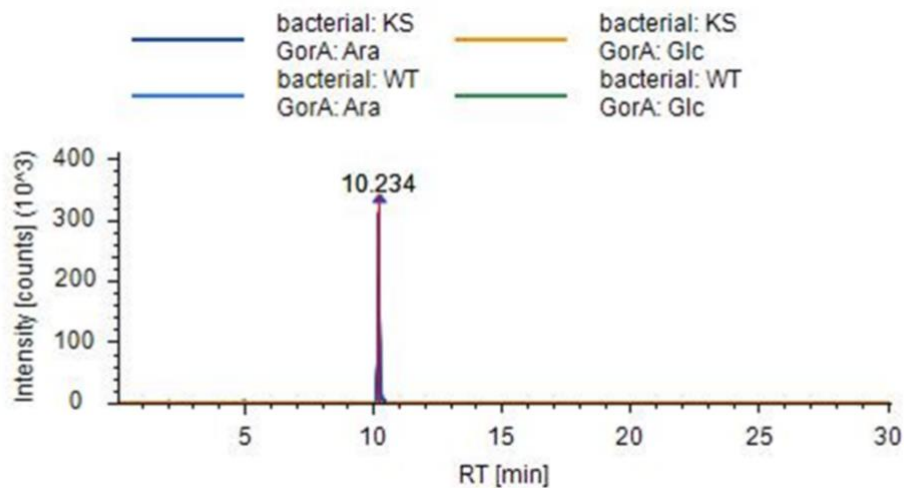
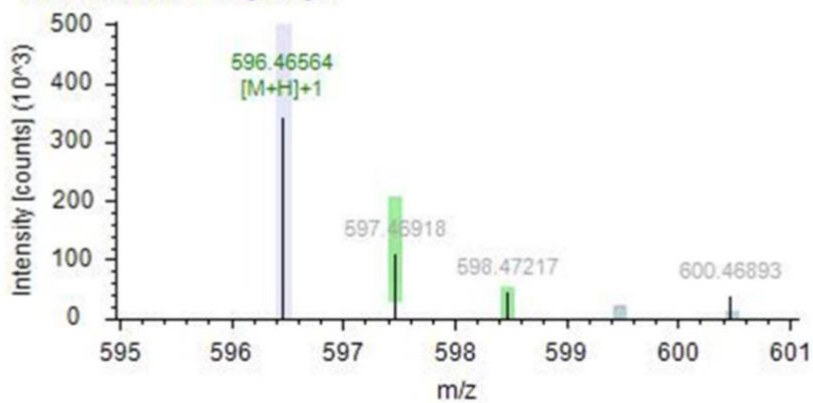


Figure S 11. LC chromatogram, MS¹ and MS² spectra of **10**. This graphic was created by Dr. Georgia Angelidou with Compound Discoverer 3.3.



20220301_Lipids_MD_38 (F38) #6041, RT=10.234 min, MS1, FTMS (+)
 C31 H66 N O7 P as [M+H]⁺1



20220301_Lipids_MD_38 (F38) #6028, RT=10.212 min, MS2, FTMS (+), (-)

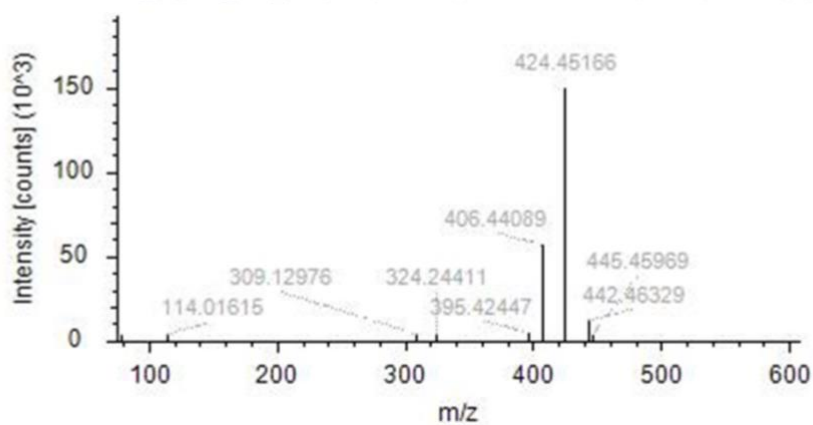


Figure S 12. LC chromatogram, MS¹ and MS² spectra of **11**. This graphic was created by Dr. Georgia Angelidou with Compound Discoverer 3.3.

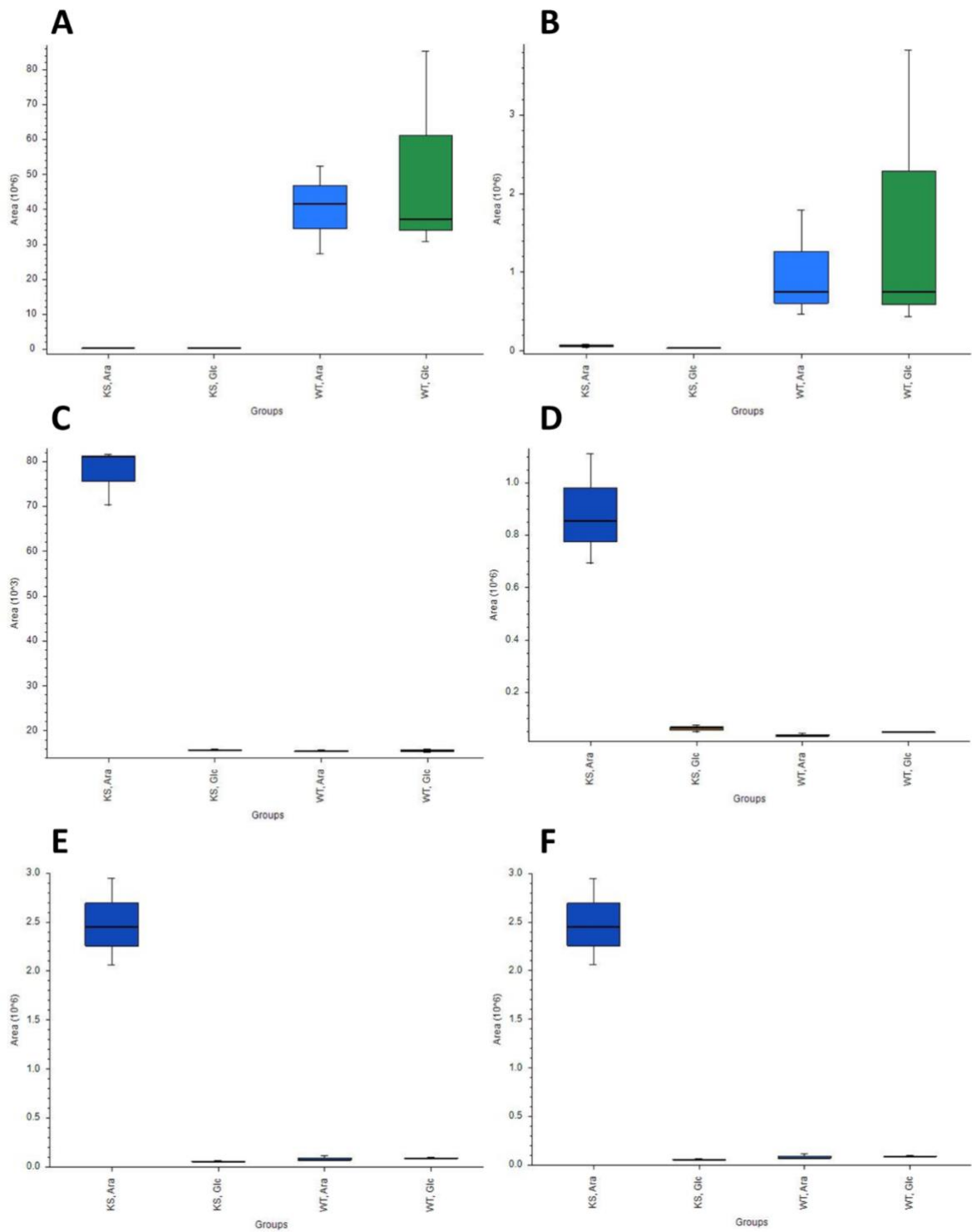


Figure S 13. Relative abundance of sphinganine in MYb115 P_{BADpks} (KS) and wild type (WT). Blue = with arabinose; Green = without arabinose. **A:** sphinganine 302; **B:** sphinganine 330; **C:** 8; **D:** 6; **E:** 5; **F:** 7.

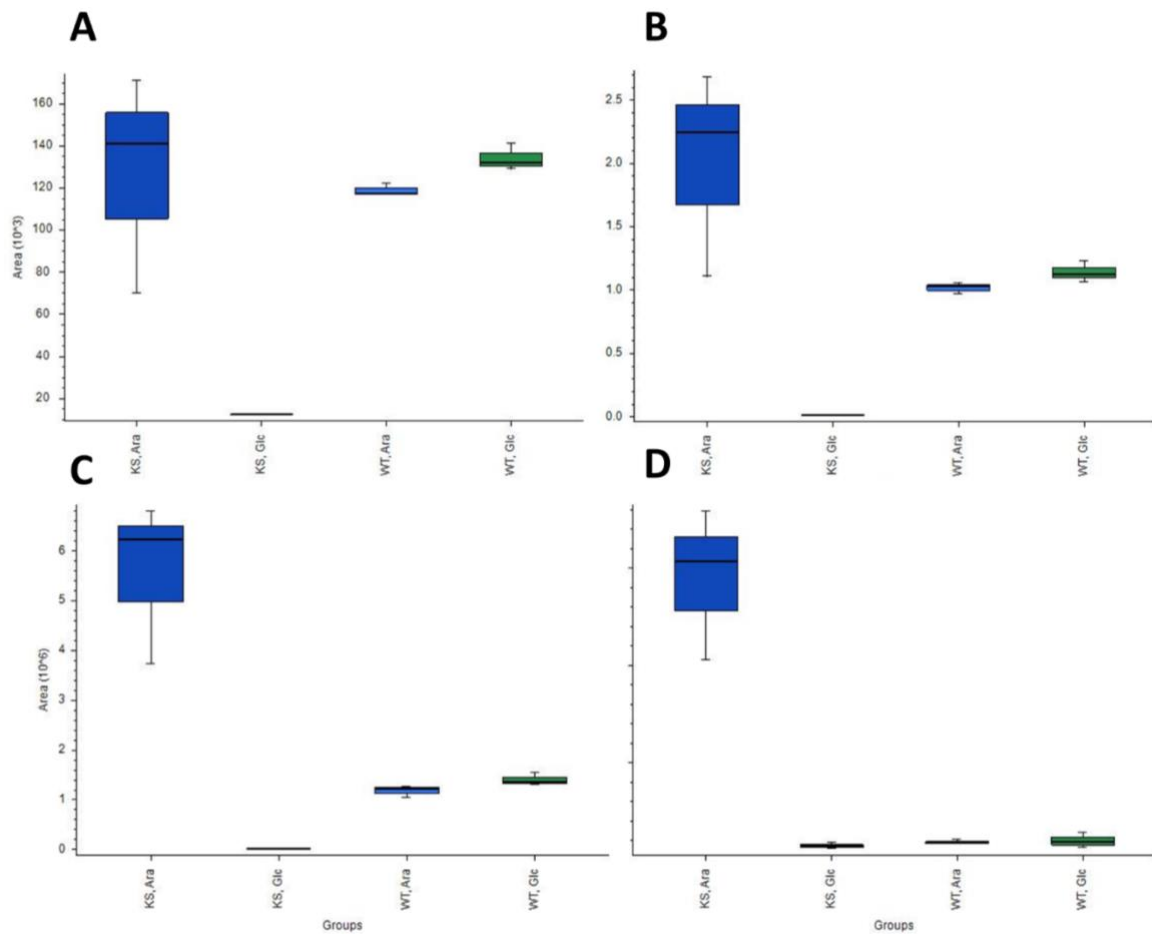


Figure S 14. Relative abundance of phosphoglycerol sphingolipids in MYb115 P_{BADpks} (KS) and wild type (WT). Blue = with arabinose; Green = without arabinose. **A: 12; B: 10; C: 9; D: 11.**

Table S 1. Pairwise identity of IszA proteins of *P. thracensis*, *X. indica*, *X. szentirmaii* DSM and *X. szentirmaii* US.

	<i>P. thracensis</i>	<i>X. indica</i>	<i>X. szentirmaii</i> DSM	<i>X. szentirmaii</i> US
<i>P. thracensis</i>		80.120%	85.584%	85.572%
<i>X. indica</i>	80.120%		78.053%	78.042%
<i>X. szentirmaii</i> DSM	85.584%	78.053%		99.988%
<i>X. szentirmaii</i> US	85.572%	78.042%	99.988%	

Table S 2. Pairwise identity of *iszA* genes of *P. thracensis*, *X. indica*, *X. szentirmaii* DSM and *X. szentirmaii* US.

	<i>P. thracensis</i>	<i>X. indica</i>	<i>X. szentirmaii</i> DSM	<i>X. szentirmaii</i> US
<i>P. thracensis</i>		78.917%	84.814%	84.811%
<i>X. indica</i>	78.917%		77.035%	77.047%
<i>X. szentirmaii</i> DSM	84.814%	77.035%		99.981%
<i>X. szentirmaii</i> US	84.811%	77.047%	99.981%	

Table S 3. Pairwise identity of IszB proteins of *P. thracensis*, *X. indica*, *X. szentirmaii* DSM and *X. szentirmaii* US.

	<i>P. thracensis</i>	<i>X. indica</i>	<i>X. szentirmaii</i> DSM	<i>X. szentirmaii</i> US
<i>P. thracensis</i>		82.419%	88.029%	88.153%
<i>X. indica</i>	82.419%		80.359%	80.722%
<i>X. szentirmaii</i> DSM	88.029%	80.359%		98.597%
<i>X. szentirmaii</i> US	88.153%	80.722%	98.597%	

Table S 4. Pairwise identity of *iszB* genes of *P. thracensis*, *X. indica*, *X. szentirmaii* DSM and *X. szentirmaii* US.

	<i>P. thracensis</i>	<i>X. indica</i>	<i>X. szentirmaii</i> DSM	<i>X. szentirmaii</i> US
<i>P. thracensis</i>		79.039%	86.268%	86.469%
<i>X. indica</i>	79.039%		78.008%	78.036%
<i>X. szentirmaii</i> DSM	86.268%	78.008%		98.744%
<i>X. szentirmaii</i> US	86.469%	78.036%	98.744%	

Table S 5. Pairwise identity of IszC proteins of *P. thracensis*, *X. stockiae*, *X. indica*, *X. szentirmaii* DSM and *X. szentirmaii* US.

	<i>P. thracensis</i>	<i>X. stockiae</i>	<i>X. indica</i>	<i>X. szentirmaii</i> DSM	<i>X. szentirmaii</i> US
<i>P. thracensis</i>		79.964%	88.525%	94.718%	94.718%
<i>X. stockiae</i>	79.964%		78.097%	79.533%	79.533%
<i>X. indica</i>	88.525%	78.097%		88.151%	88.151%
<i>X. szentirmaii</i> DSM	94.718%	79.533%	88.151%		100%
<i>X. szentirmaii</i> US	94.718%	79.533%	88.151%	100%	

Table S 6. Pairwise identity of *iszC* genes of *P. thracensis*, *X. stockiae*, *X. indica*, *X. szentirmaii* DSM and *X. szentirmaii* US.

	<i>P. thracensis</i>	<i>X. stockiae</i>	<i>X. indica</i>	<i>X. szentirmaii</i> DSM	<i>X. szentirmaii</i> US
<i>P. thracensis</i>		76.265%	88.242%	93.030%	93.030%
<i>X. stockiae</i>	76.265%		74.362%	76.380%	76.380%
<i>X. indica</i>	82.242%	74.362%		81.422%	81.422%
<i>X. szentirmaii</i> DSM	93.030%	76.380%	81.422%		100%
<i>X. szentirmaii</i> US	93.030%	76.380%	81.422%	100%	

Table S 7. Biosynthesis-related genes in genomic proximity to *sgaAB*. 45 kB region centered around *sgaAB* was analysed in *Pseudomonas* MYb115, *Legionella longbeachae* NSW150, *Methylobacterium extorquens* AM1 and *Sinorhizobium meliloti* 2011 using antiSMASH¹⁷⁷. All biosynthesis-related genes were analysed using NCBI blastp. The resulting annotations for each gene are shown here. The abundance of a gene with the corresponding annotation in each organism is indicated by x's (x=one gene with the corresponding annotation; xx = two; xxx = three or more).

	MYb115	<i>L.</i> <i>longbeachae</i>	<i>M.</i> <i>extorquens</i>	<i>S. meliloti</i>
capsular polysaccharide biosynthesis protein	xx	xxx	xxx	
capsule biosynthesis	xx			x
glycosyltransferase family 1		x		
glycosyltransferase family 2		xx	x	
SDR family oxidoreductase	x	xxx		
α -/ β -fold hydrolase		x		
amidohydrolase				x
aminotransferase				x
amino acid ABC transporter				x
pantoate- β -alanine ligase	x			
phosphotransferase	x			
phytanoyl-CoA dioxygenase family	xx			
5-guanidine-2-oxopentanoate decarboxylase				x

7 CURRICULUM VITAE

Moritz Drechsler

Geburtsdatum und -ort 19.08.1993 in Darmstadt

Nationalität Deutsch

Akademischer Werdegang

02/2018-heute Promotion am Institut für Molekulare Biowissenschaften der Johann Wolfgang Goethe Universität in Frankfurt am Main in der Arbeitsgruppe für Molekulare Biotechnologie von Prof. Dr. Helge B. Bode
Thema der Dissertation:
Identifizierung neuer Naturstoffe aus nematoden-assoziierten Bakterien mithilfe der Massenspektrometrie

10/2015-02/2018 M.Sc. Molekulare Biowissenschaften an der Johann Wolfgang Goethe Universität in Frankfurt am Main
Thema der Abschlussarbeit:
Analyse und Regulation der Naturstoff-Biosynthese in *Photorhabdus luminescens*
Angefertigt in der Arbeitsgruppe von Prof. Dr. Helge B. Bode
Gesamtnote: 1,3

10/2012-09/2015 B.Sc. Biowissenschaften an der Johann Wolfgang Goethe Universität in Frankfurt am Main
Thema der Abschlussarbeit:
Molekulare Analysen des DNA-Translokators und Etablierung eines markerlosen Mutagenesesystems in *Thermus thermophilus* HB27
Angefertigt in der Arbeitsgruppe von Prof. Dr. Beate Averhoff
Gesamtnote: 2,1

08/2003-06/2012 Abitur an der Lichtenbergschule in Darmstadt
Gesamtnote: 2,1

8 PUBLICATIONS

Singh, G., Calchera, A., Schulz, M., Drechsler, M., Bode, H. B., Schmitt, I. & Grande, F. D. Climate-specific biosynthetic gene clusters in populations of a lichen-forming fungus. *Environ Microbiol* 23, 4260-4275 (2021).

Kissoyan, K. A. B., Drechsler, M., Stange, E.-L., Zimmermann, J., Kaleta, C., Bode, H. B. & Dierking, K. Natural *C. elegans* Microbiota Protects against Infection via Production of a Cyclic Lipopeptide of the Viscosin Group. *Current Biology* 29, 1–20 (2019).

Salzer, R., D’Imprima, E., Gold, V. A. M., Rose, I., Drechsler, M., Vonck, J. & Averhoff, B. Topology and Structure/Function Correlation of Ring- and Gate-forming Domains in the Dynamic Secretin Complex of *Thermus thermophilus*. *J Biol Chem* 291, 14448–14456 (2016).

9 RECORD OF CONFERENCES

Advances in Chemical Biology 2021, poster presentation. “A natural product from a native *C. elegans* microbiota isolate increases host resistance to infection against pathogenic bacteria”, 26. – 28.01.2021, online

10 ERKLÄRUNG

Ich erkläre hiermit, dass ich mich bisher keiner Doktorprüfung im Mathematisch-Naturwissenschaftlichen Bereich unterzogen habe.

Frankfurt am Main, den 21.04.2023

Moritz Drechsler

11 VERSICHERUNG

Ich erkläre hiermit, dass ich die vorgelegte Dissertation selbstständig angefertigt und mich anderer Hilfsmittel als der in ihr angegebenen nicht bedient habe, insbesondere, dass alle Entlehnungen aus anderen Schriften mit Angabe der betreffenden Schrift gekennzeichnet sind.

Ich versichere, die Grundsätze der guten wissenschaftlichen Praxis beachtet, und nicht die Hilfe einer kommerziellen Promotionsvermittlung in Anspruch genommen zu haben.

Frankfurt am Main, den 21.04.2023

Moritz Drechsler

12 DANKSAGUNG

Ich danke meinem Doktorvater **Prof. Dr. Helge B. Bode** dafür, dass er mir die Chance gegeben hat neue Naturstoffe zu entdecken, auch wenn manche Naturstoffe „nicht ohne Grund immer noch unbekannt sind“. Außerdem danke ich ihm für die anhaltende Unterstützung und hilfreichen Ratschläge auch wenn die Ausfertigung dieser Arbeit etwas länger gedauert hat als geplant.

Ich bedanke mich herzlichst bei **Prof. Dr. Martin Grininger** für die Übernahme des Zweitgutachtens dieser Arbeit.

Für das Korrekturlesen dieser Arbeit danke ich **Dr. Carsten Kegler**.

Außerdem danke ich meinen Kooperationspartnern mit denen ich viel Freude an der Beantwortung interdisziplinärer wissenschaftlicher Fragestellungen hatte. Besonders hervorheben möchte ich in diesem Zusammenhang **Lena Peters**, **Dr. Kohar A. Kissoyan** und **Dr. Katja Dierking** von der CAU zu Kiel, sowie **Dr. Georgia Angelidou** und **Dr. Nicole Paczia** vom MPI für terrestrische Mikrobiologie in Marburg.

Des Weiteren bedanke ich mich bei allen aktiven und ehemaligen Mitgliedern der Arbeitsgruppe Bode für das tolle Arbeitsklima und die vielen aufschlussreichen Diskussionen. In diesem Zuge möchte ich mich auch bei **Dr. Yi-Ming Shi**, **Dr. Yan-Ni Shi** und **Dr. Zhengyi Qian** für die tolle Zusammenarbeit bedanken.

Ein besonderer Dank gilt meinen Büropartnern für die extensiven Gummibärchentests. Darüberhinaus danke ich **Dr. Lukas Kreling**, **Dr. Janik „Kranzelmaier“ Kranz** und **Dr. Nick Neubacher** für die unfassbar lustige Zeit und die unvergesslichen Dienstage im hohen Gras. Ich danke **Dr. Margaretha Westphalen**, **Dr. Andreas „DoktorAndi“ Tietze** und **Peter Grün** für die zahlreichen Stunden im MS-Raum und die noch zahlreicheren gemeinsamen Tassen Kaffee.

Zu guter Letzt möchte ich meiner Familie danken. Ich danke meinen Eltern dafür, dass sie mich auf meinem Weg immer unterstützt haben. Ich danke meinen Söhnen **Fritz** und **Kalle** dafür, dass sie die Zeit in der diese Arbeit entstanden ist zwar nicht kürzer aber schöner gemacht haben. Mein größter Dank gilt meiner Frau **Ann-Kathrin** dafür, dass sie mich immer bedingungslos unterstützt, motiviert und auch ertragen hat, auch wenn der ein oder andere Zeitplan nicht aufging.