Technische Universität München

Fakultät für Chemie

Professur für Biosystemchemie

Unlocking the cryptic metabolic potential of bacterial alkaloids

Katharina Lamm

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigten Dissertation.

Vorsitzende(r): Prof. Dr. Shigeyoshi Inoue

Prüfer der Dissertation:

1. Prof. Dr. Tobias A.M. Gulder
2. apl. Prof. Dr. Wolfgang Eisenreich

Die Dissertation wurde am 17. Mai 2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 16. Juli 2021 angenommen.

“Well, I won’t back down.
No, I won’t back down.
You can stand me up at the gates of hell.
But I won’t back down.

No, I’ll stand my ground.
Won’t be turned around.
And I’ll keep this world from draggin’ me down.
Gonna stand my ground.”

- Tom Petty & The Heartbreakers -

Für meine Familie.

Contents

Abstract

Bacterial pyrrolizidine alkaloids (PAs) are a yet underrepresented class of natural products with promising anticancer activity. To date, most of the few identified bacterial PA analogs are derived from Streptomycetes. Bioinformatic analyses of published bacterial genomes, however, suggest a broad variety of novel PAs awaiting discovery. This project sets out to unlock the full metabolic potential of this natural product family and thus to probe its potential as source for novel leads and tools in cancer research.

The few PA hits discovered and studied so far exhibit diverse biological activities, including anticancer, antiinfective and antibacterial activity, making them attractive for the pharmaceutical industry as potential future drugs. Based on the available gene sequences of enzymes involved in PA biosynthesis, bacterial genomes were mined in search of strains possessing novel PA-producing biosynthetic gene clusters (BGCs). Out of the many hits found, a selection of ten bacterial strains was chosen for in-depth characterization, their BGCs varying in cluster size and complexity and thus bearing the potential of finding interesting new molecules.

Target BGCs were accessed by capturing regions of interest from genomic DNA via Direct Pathway Cloning (DiPaC) or homologous recombination strategies. Activation of silent BGCs by heterologous expression of full-length or truncated BGC versions in *E. coli* or *Streptomyces* hosts was tested, cultivation parameters were optimized, and cluster refactoring was performed to facilitate production of novel PA derivatives. Furthermore, a catalytic toolbox consisting of a double promoter expression system was established to study single gene function as well as to promote biosynthesis of further PA derivatives upon flexible combination of functional genes from multiple PA-related pathways. Alongside literature-known pyrrolizixenamides **50** – **52**, longer chained derivatives **53** – **55** could be identified within the course of this thesis, thereby proving the broad distribution of PAs amongst bacteria and their evolutionary conservation. A handful of new molecules with yet unknown identity were found in heterologous expression extracts of BGCs with higher gene complexity.

In addition, innovative methods for reconstitution of enzymatic total synthesis of such compounds *ex vivo* were developed. Nonribosomal peptide synthetase (NRPS) and monooxygenase (MOX) enzymes involved in formation of the pyrrolizidine core structure common to all PA derivatives and independent of their plant, bacterial or fungal origin, were recombinantly expressed and purified. Expression and purification conditions were adapted to obtain protein of as high quantity and quality as possible, in the end yielding up to 5 mg/L of the multidomain NRPS (270 kDa) and 45 mg/L of MOX (45 kDa) protein. Enzyme activity was

confirmed by successful conversion of β-ketoacyl thioesters to the respective pyreudione products (**38**) or transformation of [5,6]-bicyclic pathway intermediates in a Baeyer-Villiger-initiated reaction cascade into pyrrolizixenamide variants (**37**). Moreover, substrate promiscuity as well as the hypothesis of a substrate-mediated reaction cascade was investigated by chemically synthesizing precursor molecules, this way expanding the substrate scope of monomodular NRPS systems and MOXs. Traces of a brabantamide derivative **90** originating from a PA-related lipocyclocarbamate (LCC)-producing pathway hinted at directionality of the MOX-catalyzed reaction depending on the available precursor molecule.

This interdisciplinary project thus contributed to the basic understanding of bacterial PA biosynthesis and displayed the capability of finding novel derivatives of varying complexity with biomedical potential in this domain of life.

Zusammenfassung

Die Naturstoffklasse der bakteriellen Pyrrolizidin Alkaloide (PAs) ist bis dato eine der am wenigsten untersuchten Molekülklassen, die jedoch vielversprechende Wirksamkeit im Kampf gegen bestimmte Krebsformen aufweist. Die Mehrheit der bisher identifizierten PAs stammen aus Streptomyceten. Allerdings haben bioinformatische Analysen von Bakteriengenomen ergeben, dass die Biosynthese dieser Naturstoffe weit verbreitet zu sein scheint und deshalb viele neue, bisher unerkannte PA auf ihre Entdeckung warten. Dieses Dissertationsprojekt verfolgt deshalb das Ziel, einige dieser unerforschten bakteriellen PAs zu finden, diese zu identifizieren und zu charakterisieren und schließlich deren Potenzial und Einsatz als Wirkstoffmoleküle in der Krebsforschung zu prüfen.

Diejenigen Vertreter der PAs, die bereits entdeckt und im Detail untersucht wurden, weisen eine Vielzahl an biologischen Aktivitäten auf, unter Anderem zeigen sie Wirksamkeit als Antiinfektiva, Antibiotika oder Antikrebsmittel und sind daher als zukünftige Arzneimittel von Interesse für die pharmazeutische Industrie. Ausgehend von den Gensequenzen derjenigen Enzyme, die an der PA-Biosynthese publizierter Moleküle beteiligt sind, wurden bakterielle Genome auf der Suche nach ähnlichen PA-produzierenden biosynthetischen Genclustern (BGCs) durchsucht. Aus der Vielfalt an erhaltenen Treffern wurden zehn BGCs für eine detaillierte Untersuchung ausgewählt, wobei neben Clustern mit identischer Organisation wie das Referenzcluster auch BGCs mit erheblich gesteigerter Gesamtgröße und erhöhter Genkomplexität selektiert wurden. Insbesondere anhand der komplexeren Gesamtanordung erhoffte man sich die Entdeckung neuer Moleküle.

Mithilfe der Direct Pathway Cloning Strategie (DiPaC) sowie homologer Rekombination konnte der jeweilige Ziel-DNA-Abschnitt ausgehend von der genomischen DNA des Ursprungsorganismus für die Klonierung zugänglich gemacht werden. Eine Aktivierung sogenannter „stiller“ Sekundärmetabolit-BGCs erfolgte mittels heterologer Expression des jeweiligen Clusters, entweder in seiner vollen Länge oder in gekürzter Form in *E. coli* oder *Streptomyces* Wirtsstämmen. Sämtliche Parameter rund um die Kultivierung, sowie eine Gen-Rearrangierung auf Klonierungsebene wurden im Laufe experimenteller Studien optimiert, um die Naturstoffexpression unter Laborbedingungen bestmöglich zu unterstützen. Des Weiteren wurde eine Art katalytischer Baukasten erstellt, der, basierend auf einem Doppelpromotor-Expressionssystem, nicht nur zur Untersuchung einzelner Genfunktionen verwendet werden sollte, sondern damit auch die Produktion „künstlicher“ neuer PA-Derivate ermöglichte. Neben den bereits Literatur-bekannten Pyrrolizixenamid-Derivaten **50** – **52** konnten weitere Moleküle **53** – **55** mit längeren Acylseitenketten isoliert werden. Anhand dieser Beispiele konnte die

weite Verbreitung der PA-codierenden BGCs in Bakterien und deren evolutionäre Konserviertheit bestätigt werden. Zusätzlich wurden anhand von Masseanalysen einige weitere Moleküle in der Größenordnung der PAs bei der Expression komplexerer BGCs entdeckt. Deren Identität konnte allerdings noch nicht aufgeklärt werden.

Zusätzlich zur Expression ausgewählter PA-codierender BGCs *in vivo*, wurden diverse Methoden angewandt, um die Biosynthese des Pyrrolizidin-Grundgerüsts, welches allen Vertretern der PA-Naturstoffklasse gemein ist, *in vitro* schrittweise zu rekonstruieren. Die beiden an der PA-Biosynthese beteiligten Enzyme, eine nichtribosomale Peptidsynthetase (NRPS) und eine Monooxygenase (MOX), wurden als rekombinante Proteine exprimiert. Das Expressionsvorgehen sowie die Bedingungen für die anschließende Proteinreinigung wurde extensiv optimiert, um ein möglichst sauberes, hoch konzentriertes Protein zu erhalten. Die Megasynthetase NRPS (270 kDa) konnte in Mengen von bis zu 5 mg/L erhalten werden, während die deutlich kleinere MOX (45 kDa) mit bis zu 45 mg/L exprimiert werden konnte. Die Enzymaktivität und volle Funktionalität der Proteine wurde anhand einer erfolgreichen Umsetzung diverser Substrate, wie β-ketoacyl Thioestern zu Pyreudion-Derivaten (**38**) oder die [5,6] -bizyklischen Intermediate über eine Baeyer-Villiger-initiierte Reaktionskaskade zu den entsprechenden Pyrrolizixenamid-Derivaten (**37**), bestätigt. Des Weiteren wurde durch die Variation des Substrates der jeweiligen Reaktion mittels chemisch synthetisierter Vorläufermoleküle die Substratspezifität der beiden Enzyme getestet. Neben den typischen Varianten von **37** konnten in einem MOX-katalysierten Experiment Spuren eines Brabantamid-Derivats **90** nachgewiesen werden, was darauf hindeutet, dass die Art des Substrats den Ausgang der Reaktion steuert und somit auch Moleküle eines verwandten Lipocyclocarbamat (LCC) -Biosyntheseweges entstehen können.

Anhand der beschriebenen Experimente und Untersuchungen konnte dieses interdisziplinäre Projekt somit zum allgemeinen Verständnis der bakteriellen PA-Biosynthese beitragen. Außerdem wurden die Möglichkeiten zur Identifizierung neuer Moleküle mit variabler Komplexität und hohem biomedizinischem Potenzial in dieser Domäne des Lebens aufgezeigt.

List of Abbreviations

Abbreviation	Definition
A	Adenylation domain
AA	Acryl amide
ACN	Acetonitrile
ACP	Acyl carrier protein domain
ADP	Adenosine diphosphate
AdoMet	S-adenoysl methionine
AMP	Adenosine monophosphate
Amp/Amp^R	Ampicillin
antiSMASH	Antibiotica & Secondary Metabolite Analysis SHell
APS	Ammonium persulfate
ATP	Adenosine triphosphate
AT	Acyl transferase domain
BAC	Bacterial artificial chromosome
BGC	Biosynthetic gene cluster
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BVMO	Baeyer-Villiger monooxygenase
C	Condensation domain
CASO	Casein soy peptone bouillon
CoA	Coenzyme A
CTAB	Cetyltrimethyl ammonium bromide
Cy	Cyclization domain
DH	Dehydratase domain
ddH_2O	Bisdistilled water
DHA	Dehydroalanine

Abbreviation	Definition
DiPaC	Direct cloning pathway
DNA	Desoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DMSO	Dimethyl sulfoxide
EDTA	Ethylenediaminetetraacetic acid
E	Epimerization domain
ESI	Electron Spray Ionization
EtOAc	Ethyl acetate
EtOH	Ethanol
F	Formylase domain
FA	Formic acid
FAD	Flavine adenine dinucleotide
FAS	Fatty acid synthase
FMN	Flavine adenine mononucleotide
GA	Gibson assembly
GSH	Glutathione
GYM	Glucose yeast extract malt extract
Hal	Halogenase
HPLC	High performance liquid chromatography
HSS	Homospermidine synthase
H_2O	Water
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISP	Inorganic salt starch agar
Kan/Kan^R	Kanamycin
kbp	Kilobase pair
KR	β-Ketoreductase domain
KS	β-Ketoacyl domain

Abbreviation	Definition
L-AZC	L-azetidine 2-carboxylic acid
LB	Luria Bertrani
LCC	Lipocyclocarbamate
LC-MS	Liquid chromatography-mass spectrometry
LCHR/LLHR	Linear-circular-/linear-linear homologous recombination
MT	Methyltransferase domain
MeOH	Methanol
MOX	Monooxygenase
MS	Mass spectrometry
MS-MS	Tandem mass spectrometry
m/z	Mass to charge ratio
M9	Mineral minimal medium
NAD(P)H	Nicotinamide adenine dinucleotide (phosphate)
NCBI	National Center for Biotechnology Information
NLS	N-lauroyl sarcosine
NMT	N-Methyltransferase domain
NMR	Nuclear magnetic resonance
NRP	Nonribosomal peptide
NRPS	Nonribosomal peptide synthetase
NZA	N-Z-Amine medium
$OD_{\lambda = x}$	Optical density at wavelength λ = x
ori	Origin of replication
Ox	Oxidation domain
PA	Pyrrolizidine alkaloid
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
PKS	Polyketide synthase

Abbreviation	Definition
PPTase	Phosphopantetheinyl transferase
R	Reductase domain
RNase A	Ribonuclease A
rpm	Rounds per minute
SCDH	Short chain dehydrogenase
SD	Shine Dalgarno
SDS	Sodium dodecyl sulphate
SAM	S-adenosylmethionine
T	Thiolation domain
T_a	Oligonucleotide annealing temperature
TAE	Tris/acetate/EDTA buffer
TAR	Transformation-associated recombination
TB	Terrific broth
TE	Thioesterase domain
TE	Tris-EDTA
TFA	Trifluoroacetic acid
TMR-Mal	tetramethylrhodamine-5-(and -6)-C2 maleimide
t_R	Retention time HPLC
vuPA	Vinylogous-urea type PA
YAC	Yeast artificial chromosome

1. Introduction

1.1 Secondary metabolites and their role in pharmacy

In general terms, metabolites are defined as the end products of regulatory processes taking place within the cell of an organism, furthermore distinguishing between molecules originating from primary or secondary metabolism.[1] Primary metabolites are essential for survival of the producing organism as they are needed for cellular growth, maintenance, and development processes. Consequently, these molecules are synthesized permanently by every organism, including fungi, bacteria, and plants. In contrast, secondary metabolites, often also referred to as natural products, do not play a role for essential biosynthetic pathways but instead take over more of an ecological function for the producer.[2] Apart from functioning as protective agents against feeding enemies and weakening food competitors, secretion of natural products allows the respective organism to interact via signaling molecules with its environment.[3] Compared to primary metabolites and due to their lack of necessity concerning cell survival, secondary metabolites are produced in rather small amounts. This, however, does not limit molecular diversity, with natural products showing not only a great variety of chemically but also functionally different molecules. Moreover, a number of secondary metabolites have been found to show biological activities useful for human applications, thus leading to an increasing interest in their potential in future pharmaceutical applications.[4]

The use of natural products for the treatment of human diseases has been ongoing for thousands of years, with the earliest records dating back to ancient Mesopotamia (2600 BC). Furthermore, especially plant-based natural products were applied in ancient Egypt and China to cure or mitigate diseases of all kinds. To date, natural products have found to be the most successful source of potential drug leads. Therefore, original medicinal practices have been extended by elaborate research in the clinical, pharmacological and chemical fields to find novel molecules of interest and learn more about their biological activities and modes of action.[5] The evolution of natural products over many years in combination with long selection processes has led to the formation of structurally complex compounds. The correct absolute configuration of these secondary metabolites is the basis for highly specific interactions with biological target molecules within a biological system, thus allowing the respective molecules to exhibit selective biological activity.[6,7]

Over the past 200 years of pharmaceutical industry, natural products have taken on a significant role concerning the discovery and development of medicines.[8] Until the 1980s, approximately 80 % of commercially available drugs were derived of natural products or based on their structural composition, thus emphasizing their importance for human health care and well-being[9]. With secondary metabolites originating in plants, bacteria and fungi, a large pool of chemically diverse molecules with numerous modes of action, including antimicrobial, antifungal, anti-infective, anti-cancer and anti-diabetic biological activity could be accessed.[9] Alongside the original natural product, several natural product inspired compounds or semisynthetic derivatives have been developed that show increased bioactivity.

Of the many natural products found to show biological activity, most of them are applied in antibacterial treatment, with penicillin (**1**) being one of the most prominent examples (Figure 1). Penicillin (**1**) was first discovered by Alexander Fleming in 1928, who identified the fungus *Penicillium notatum* as the microbial producer and confirmed the antibacterial activity *in vitro.*[10] Large-scale cultivation and subsequent experiments performed by Florey and Chain confirmed the *in vivo* activity of **1** shortly after.[11] In the course of emerging penicillin-resistant *staphylococcus* strains and in search of alternative antibacterial secondary metabolites, the glycopeptide vancomycin (**2**) was isolated from a dirt sample containing *Streptomyces orientalis* in the 1950s. Although it showed activity against most gram-positive strains and even some anaerobic organisms (*clostridia* and *Neisseria gonorrhoeae*) the antibiotic perceives toxicity. Therefore, it was rapidly replaced by less toxic and more efficient antibiotics[12]. The tetracyclines (**3**) form another large group of antibiotics that have been in frequent use for human and animal disease treatment since their discovery in the 1940s. These so-called “broad spectrum agents” not only have a wide range of activity against various types of microorganisms, but they also show hardly any side effects or toxicity issues.[13] **3** and its derivatives intervene with bacterial protein synthesis by inhibiting aminoacyl-tRNA binding to the ribosome so that proper functioning of the transcriptional machinery is prevented. Furthermore, the natural product spectrum originating from *Streptomyces* is constantly augmented by semisynthetic compounds exhibiting increased potency, solubility, bioavailability and aiming to avoid the general issue of antibiotic resistance.[14] A similar mode of action has been reported for the macrolide erythromycin A (**4**), a molecule natively produced in the soil-dwelling bacterium *Saccharopolyspora erythraea* alongside a number of less abundant derivatives.[15] One major drawback of **4** is that it is strongly metabolized by cytochrome P450 complexes, systems that are also involved in the biotransformation of a variety of other therapeutics. Intake of **4** together with other pharmaceuticals that are also

metabolized by cytochrome P450 complexes can therefore lead to the accumulation of chemical agents, which in turn can intensify side effects.[16]

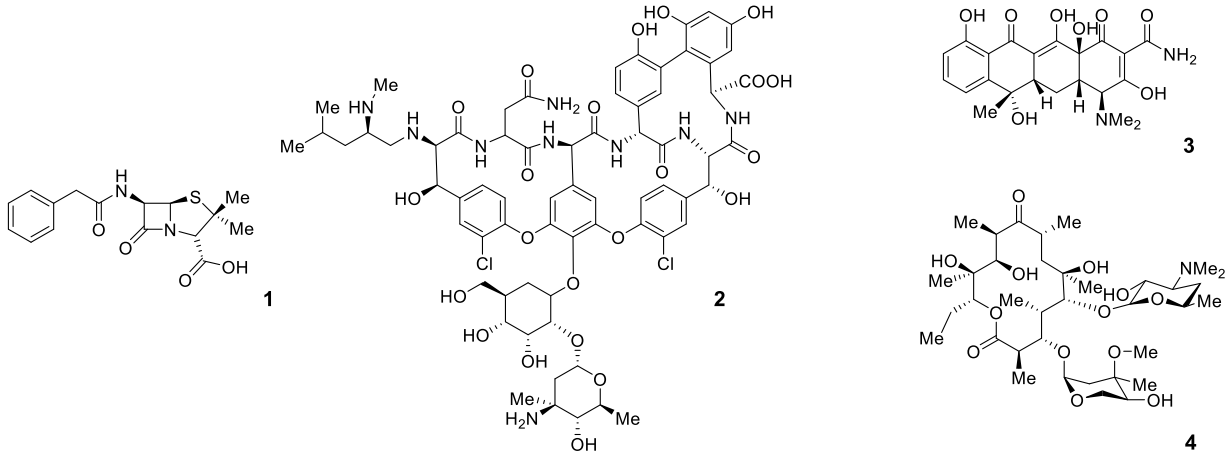


Figure 1. Selection of prominent natural product antibiotics: molecular structures of penicillin (**1**), vancomycin (**2**), tetracycline (**3**) and erythromycin A (**4**).

Ivermectin (**5**), a semi-synthetic derivative of the macrocyclic lactone avermectin, is applied as an antiparasitic agent due to its broad activity against various species of nematodes and arthropods (Figure 2). It functions as a GABA neurotransmitter antagonist, thereby blocking signal transmission from nerves to muscles.[17] Another important natural product applied when treating malaria is the sesquiterpene lactone artemisinin (**6**). This plant-derived molecule isolated from *Artemisia annua* has been an inherent part of fever treatment in traditional Chinese medicine for 2000 years. All derivatives of **6** share the same mode of action against the malaria-causing parasite *plasmodium falciparum*. Compounds are activated upon heme-binding, causing the endoperoxide bridge to break and leading to the presence of reactive oxygen species, which in turn damage intracellular targets. Toxicity is selective to the parasite as it metabolizes blood hemoglobin and is thus comparably rich in heme.[18] Apart from its antimalarial activity, tests have revealed cytotoxic effects on various cancer cells.[19]

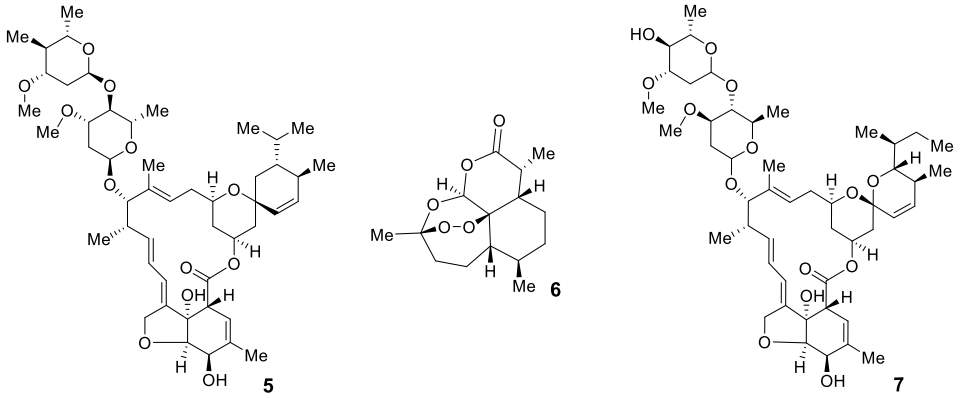


Figure 2. Natural products used as antiparasitics: ivermectin (**5**), artemisinin (**6**) and avermectin (**7**).

With cancer evolving to be one of the leading causes of human deaths worldwide, the urgence of finding drugs to heal or at least prolong patients' lives has become more and more of a priority. Mitomycin C (**8**), discovered in the 1950s in Japan after fermenting *Streptomyces caespitosus*, has been widely utilized in chemo-therapeutic treatment of all kinds of tumors (Figure 3).[20] Mitomycins are termed bioreductive drugs since the active metabolites with cytotoxic potential are produced intracellularly.[21] Another natural product applied especially for ovarian and breast cancer therapy is paclitaxel (**9**), a diterpenoid pseudoalkaloid derived from the pacific yew tree *Taxus brevifolia*.[22] It inhibits cellular mitosis at the level of microtubule spindle dynamics by stabilizing microtubules and preventing tubulin disassembly.[23] With the interest of natural product discovery shifting to finding additional sources for novel compounds in the marine environment, the molecule salinosporamide A (**10**) was identified by Fenical, Jensen and co-workers in 2003. Its anti-cancer activity is based on the inhibition of the 20S proteasome.[24]

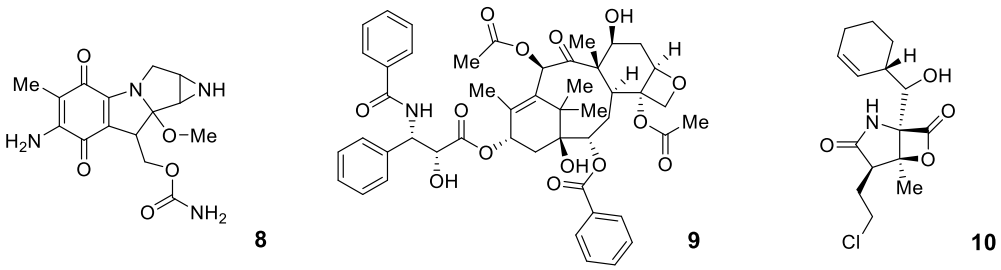
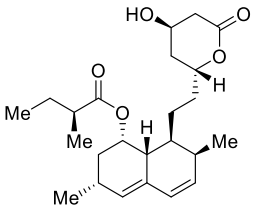
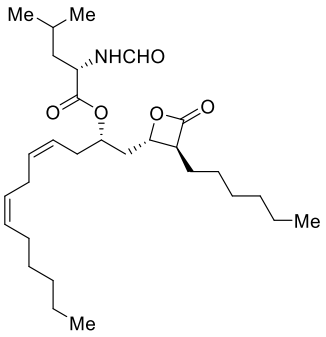


Figure 3. Selected representatives of anticancer natural product leads. Shown are the DNA-crosslinking agent mitomycin C (**8**), mitosis interrupting paclitaxel (**9**), and proteasome inhibitor salinosporamide A (**10**).

Apart from their antibacterial, antiparasitic or antitumoral activities previously mentioned, natural products can also aid in the treatment of diseases such as adiposity and hypercholesterolemia. Lovastatin (**11**) is produced by the fungus *Aspergillus terreus* and intervenes with cholesterol biosynthesis by inhibiting the enzyme HMG-CoA-reductase involved in the rate-limiting step of the pathway (Figure 4).[25] Bioactive tetrahydrolipstatin (**12**) is an approved drug used to combat obesity as it inhibits human pancreatic lipase so that the uptake of fat is blocked.[26]



11



12

Figure 4. Secondary metabolites applied against hypercholesterolemia and obesity. Both lovastatin (**11**) and lipstatin (**12**) belong to the class of statins.

Despite the findings of a number of significant and life-improving drugs, many pharmaceutical companies have reduced their investments in natural product research in the late 20th century.[1] Not only are the organisms of interest often hard to access, but difficulties also arise during cultivation and subsequent obtainment of sufficient amounts of product for further analyses. Additionally, natural product chemistry is complex and time-intensive, making the establishment of a synthetic compound library more compelling in terms of screening efforts. Apart from rediscovering already known molecules, the purity of the isolates plays a crucial role. Whereas synthetic compounds tend to be very pure, natural product isolates often contain more than one compound which interacts with multiple biological targets, thus making mechanism-based drug discovery approaches difficult. In line with the reduction of natural product research by the pharmaceutical industry starting in the 1990s, there has been a significant decrease in new leads as well as new drug approvals (Figure 5).[27]

In contrast to reduced efforts towards natural product discovery by the pharmaceutical industry, academical institutions have continuously maintained their research efforts in this field over the past years. The development and application of new methodologies as well as technological advances in synthetic biology, bioinformatics, analytics, metagenomics, and microbial genomics have led to the formation of an increased number of academic-industrial collaborations and a renewed interest in natural product discovery. One example is the

discovery of the avermectins (**7**) in 1978 in *Streptomyces avermitilis* by the Japanese Kitasato Institute in cooperation with the pharmaceutical company Merck, receiving the Nobel Prize in 2015 for their research efforts.[1] Avermectins were found to exhibit anthelmintic activity and some derivatives thereof can be used to treat river blindness and elephantiasis.[28] Furthermore, it has been seen that combinatorial chemistry in combination with high throughput screening (HTS), which at first seemed to be a promising alternative to natural product-based drug screening, also has disadvantages. Synthetically made compounds often do not show enough structural diversity and therefore a comparatively small number of hits have been identified which match to therapeutic targets. This suggests that combinatorial chemistry is better suited for the expansion and modification of natural product leads rather than creating new ones.[29] In addition, the existing compound libraries only target a staggeringly low percentage of the human genome (1 %), making the necessity of discovering new drug leads for so-called "undruggable targets" desirable.[7]

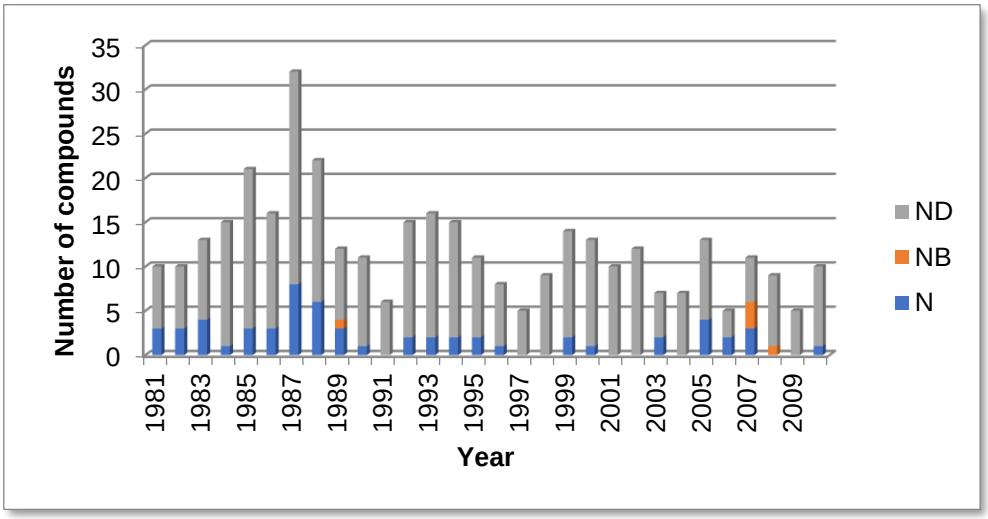


Figure 5. Newly approved natural product drug leads from 1981-2010. N = unmodified natural products, NB = plant-based natural products, ND = derived from a natural product, semi-synthetically modified. Figure taken and adapted from Newman et al. [30]

As mentioned above, several technological advances have been achieved which make research in the field of natural products discovery appealing again. With the introduction of inexpensive next generation sequencing, entire microbial genomes can now be analyzed and their biological potential elucidated. Bioinformatic platforms such as SMURF[31], PRISM[32] and antiSMASH[33] allow the predictions of secondary metabolites and simplify the studies on cryptic pathways. Moreover, the pool of novel microbial sources has expanded decisively, with marine bacteria and cyanobacteria playing an essential role.[34] To save time and reduce the costs in

finding potential pharmaceutical leads, automated screening techniques and improved methodologies such as Top-Down or Bottom-Up approaches have been established.[35] Analytical tools such as LC-MS and NMR for compound isolation and structure elucidation are becoming more and more sensitive to facilitate compound purification and precise information. Also to be mentioned are the advances concerning microorganism cultivation, where researchers have increased the quantity of fermentation cultures in small scale, started to engineer strains to produce higher amounts of compounds or gone over to heterologous expression of biosynthetic pathways of interest.[34] Taken together, the current relevance of natural product research, especially regarding pharmaceutical applications and future potential drug candidates, is immense and bears a great potential with yet a lot to be discovered.

1.2 Biosynthesis of natural products

As a result of genome mining, more than 100.000 bacterial genomes have been sequenced, thus giving access to an enormous pool of unknown secondary metabolite pathways.[36] Sequencing data has shown that genes encoding for enzymes involved in the same pathway which catalyze successive steps are usually organized together and can be co-localized on the chromosome of the producing microorganism. The linkage of biosynthetic, regulatory and resistance genes involved in the biosynthesis of one natural product is defined as a biosynthetic gene cluster (BGC). Genes within one BGC can be subdivided into two categories according to their function in natural product biosynthesis. Signature genes encoding for NRPS, PKS systems or terpene synthases are essential as they build the core structure of the compound, whereas so-called tailoring enzymes, e.g., encoding oxidoreductases, transferases, halogenases and regulatory genes modify the core structure and therefore enhance natural product variety and diversity. The organization of multiple independent BGCs to form a supercluster which produces a completely different secondary metabolite has also been found. In Nature, the production of specific secondary metabolites is dependent on the microorganisms' environment and whether the compound is needed or not, a difficulty arising when investigating BGC metabolite production under laboratory conditions where environmental triggers are absent and the gene cluster may thus be silent.[2]

Despite the enormous diversity of natural products identified and still awaiting to be found, most compounds can be assigned to one of four major secondary metabolite classes, namely the terpenoids, alkaloids, phenylpropanoids and polyketides, depending on their biosynthetic origin and the enzymatic machinery producing them.[37]

1.2.1 Nonribosomal peptide biosynthesis

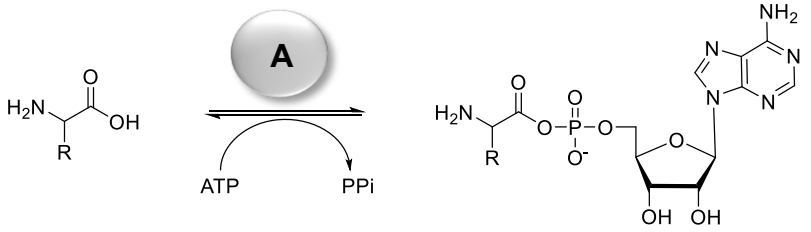
In the 1960s, with the discovery of many novel natural products and the elucidation of their structures progressing rapidly, the fact that some of these molecules containing fatty acid or non-proteinogenic amino acid components could not be produced by the ribosomal machinery became more and more clear.[38] Instead, nonribosomal peptide synthetases (NRPSs), large multienzyme complexes found mainly in bacteria and fungi, were found to be responsible for synthesizing these peptide products with broadranging biological activities. Unlike molecules made by the ribosome, nonribosomal peptides (NRPs) are derived of diverse building blocks, including proteinogenic, non-proteinogenic amino as well as aryl acids which can be posttranslationaly modified, thereby creating structural and functional diversity.[39] Mach et al.[40] and Kleinkauf et al.[41] were the first to find that amino acid incorporation can proceed in alternative ways which differs from protein synthesis in the 1960s. Tests comprised the application of antibiotics known to inhibit the protein synthetic machinery via interacting with the ribosome and showed little effect on natural product synthesis in *Bacillus brevis*, thereby supporting the hypothesis of nonribosomal peptide synthesis. Studies on NRPS enzymology by Lipmann et al.[42-44] in the following years led to the reconstruction of the enzymes' mode of action. Step-by-step elucidation of the respective biosynthetic pathways distinctively contributed to the understanding of nonribosomal peptide synthesis in detail.

As previously mentioned, NRPS systems are multienzyme complexes of varying size and molecular weight. They can further be subdivided into modules, with each module being responsible for the incorporation of a single specific amino acid into the growing peptide chain which is released at the end of the assembly line. Within one module, catalytically active enzymes that function independently from one another are organized as domains. The type of chemical reaction that is catalyzed by each domain, the order they take within a module, as well as the number of modules making up the multienzyme complex defines the primary structure of the resulting peptide molecule.[45] As NRPSs can consist of up to 21 000 residues and reach molecular weights of up to 2.3 MDa, they belong to the megasynthetases.[46] Especially fungal NRPS systems can reach high complexity and large size since natural products are often made by a single synthetase. One example is a synthetase from *Pseudomonas syringae*, a protein comprised of 9454 amino acids and of approximately 1 MDa and the largest known NRPS of bacterial origin to date.[47] NRPSs are multifunctional in terms of being used simultaneously as a template for amino acid selection by each module and harboring all necessary catalytic features for full functionality of the biosynthetic machinery.[38]

1.2.1.1 NRPS domains

There are three core domains essential to peptide formation and thus ubiquitous to all NRPS modules, namely the substrate-activating A-domain, the cofactor-binding and substrate-acylating PCP-domain as well as the peptide bond-promoting condensation (C) domain. Alongside these obligatory domains, a number of peptide-modifying enzymes can accompany the C-A-PCP combination.[48]

With an approximate number of 500 – 550 residues, the adenylation domain (A-domain) is the first entity to come in contact with the substrate needed for peptide formation. It fulfills a recognition and gatekeeper task by selecting amino acids that are to be incorporated into the final natural product and thus determines the peptides' primary sequence. Furthermore, the A domain takes over substrate activation. In a first step, the selected amino acid is converted to an aminoacyl adenylate using ATP as an energy source and in the presence of Mg^{2+}. The aminoacyl adenylate then undergoes a nucleophilic attack by a thiol group of the phosphopantetheinyl (ppan) cofactor attached to an adjacent PCP-domain to form an aminoacyl thioester (Scheme 1).[49] The two-step reaction is made possible by reorganization and conformational change of the catalytic center directly after ATP-cleavage.[38, 50] In some cases, the A domain needs to be activated in advance by small MbtH-like proteins, however the mechanism of action is yet unknown.[51]

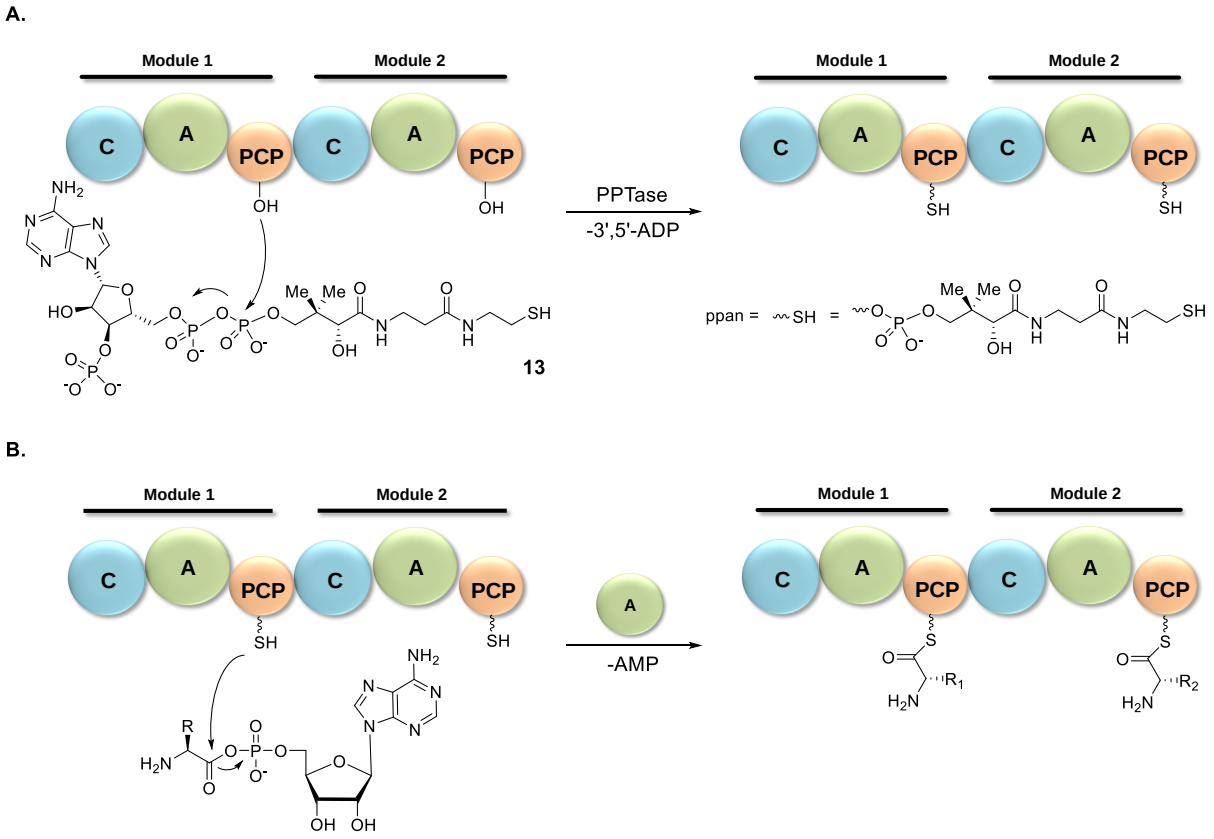


Scheme 1. A-Domain activity for amino acid activation. The amino acid to be incorporated into the peptide molecule is converted to the aminoacyl adenylate via ATP-hydrolysis.

Alongside acyl-CoA synthetases and oxidoreductases such as luciferases, the A-domain belongs to the adenylate-forming enzyme superfamily. Although sequence identities of 30 – 60 % between members of this superfamily are not high, most enzymes share similar core motifs encoding the same functionality.[45] Conserved residues are found to be part of the substrate binding pocket which is located at the interface of the domains' C- and N-terminal subunits. Referring to the superfamily name, the ATP-dependent activation and formation of acyl adenylates is common to all enzymes. One of the first crystal structures solved was the phenylalanine-activating A domain (PheA) belonging to the first module of gramicidin S

synthetase from *Bacillus brevis*.[52] Structure elucidation led to the identification of amino acid residues involved in binding of the phenylalanine substrate as well as coordination of the ATP-cofactor and mutational studies confirmed their catalytic relevance.

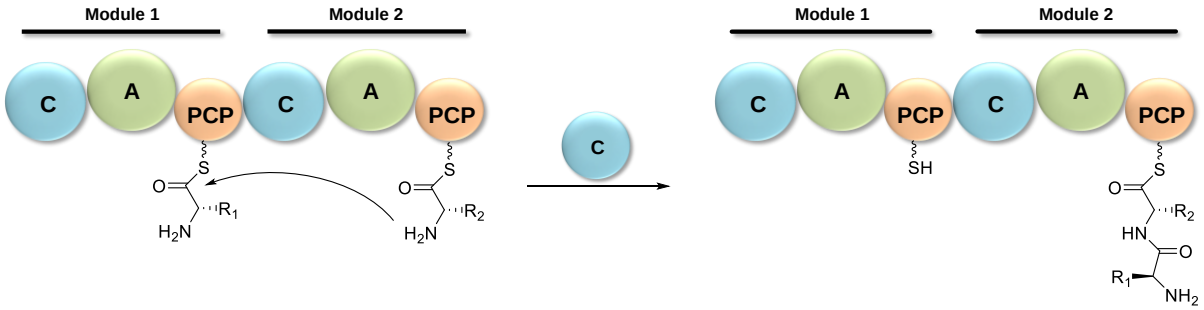
Thiolation (T) or peptidyl carrier protein (PCP) domains function as flexible tethers that can carry building blocks such as substrates or intermediate compounds between different catalytic domains. As part of the carrier protein (CP) superfamily, they share the common four-helix-bundle fold.[50] Interestingly, members of the CP superfamily, including acyl carrier proteins from PKSs and fatty acid synthases, possess similar secondary and tertiary structures but there are only minor sequence similarities on the primary structure level. Merely the amino acid directly surrounding the highly conserved serine residue at the N-terminal end of the second helix is conserved among many PCP-domains.[38] To obtain the active *holo*-state, a 4'-phosphopantetheine (ppan) moiety, which functions as a flexible arm swinging between the catalytic domains, is transferred from coenzyme A (**13**) and covalently bound to the active site serine via a thioester linkage (Scheme 2). Transformation of the PCP-domain from its *apo*- to *holo*-state requires a phosphopantetheinyl transferase (PPTase) which is therefore often co-expressed with the respective gene cluster.[53] With lengths of 70 to 100 amino acids, PCPs are the smallest domains within the NRPS module. Furthermore, they are the only domains incapable of functioning independent from other catalytic domains and the conformational changes occurring are tightly linked to reorganization events within the adjacent A- and C-domains.[54] Protein dynamics were extensively studied for the seventh module PCP-domain of the tyrocidin A NRPS system TycC3-PCP, showing that three conformational states (A, A/H and H) are possible. Helix flexibility and the extent of helix unraveling alter in each of the *apo*- or *holo*-states, these changes going hand in hand with ppan arm movement.[55] Furthermore, interactions of the TycC3-PCP-domain with the PPTase Sfp were investigated in further detail, revealing that the cofactor coenzyme A lies inside a cavity formed upon complexation of both proteins and that coordination of magnesium ions is necessary for the transfer of the coenzyme A moiety from Sfp to the PCP-domain to take place.[53]



Scheme 2. PCP-domain activation by Sfp and subsequent loading of aminoacyl adenylate onto the PCP-domain. The PPTase Sfp transfers the ppan moiety of co-substrate **13** to the central serine residue located in the catalytic pocket of the PCP-domain. An activated amino acid residue can then be covalently bound to the ppan moiety of the PCP-domain which functions as a flexible arm that can transport substrate or intermediates to the adjacent catalytic units.

In contrast to A- and PCP-domains, which participate in peptide initiation, C-domains are compulsory elements of the NRPS module for peptide elongation. Serving as the sites of peptide bond formation, the number of C-domains present within an NRPS usually corresponds to the number of existing peptide bonds of the linear end product. Unlike most enzymes, C-domains exhibit substrate binding sites on the protein surface instead of building a catalytic pocket. Two “crater-like” emarginations function as an acceptor binding site for aminoacyl-S-PCP and a donor binding site for peptidyl-S-PCP. The mode of action includes a nucleophilic attack of an α-amino group of the downstream acceptor unit onto the thioester group of the current module donor group, leading to the formation of an amide bond (Scheme 2).[48] This way, the growing peptide chain is translocated downstream in an assembly line manner to the next module. Whereas the donor-S-PCP is again available for a new peptide, the downstream acceptor site carrying the elongated peptide chain switches functionality to serve as a donor site in the next elongation step (Scheme 3).[56] A conserved HHxxxDG motif

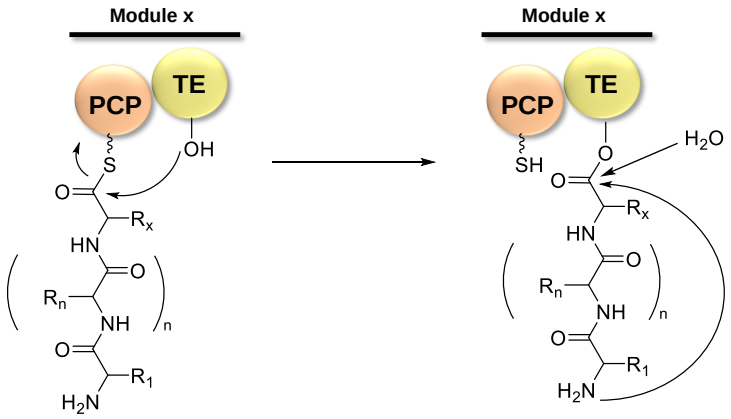
located at the interface of substrate donor and acceptor site has shown to be catalytically essential to C-domain functionality. Identical motifs have been identified in epimerization and heterocyclization domains as well as in acyl transferases, where studies have revealed that the second histidine residue acts as a catalytic base[57]. The structural importance and catalytic activity of this conserved motif have been confirmed by multiple mutational studies, where amino acid exchange led to an impairment of C-domain activity, e.g. for surfactin synthetase[58] and enterobactin synthetase[59]. The siderophore enterobactin is synthesized by biomolecular machinery consisting of three proteins EntE, EntB and EntF. EntF consists of a complete NRPS module, making it an attractive model for studying domain activity including that of C-domains. Structural studies and enzymatic assays revealed C-domain catalyzed amide bond formation between 2,3-dihydroxybenzoate (DHB) and seryl-S-EntF, leading to DHB-seryl-S-EntF intermediate production. With this, the importance of the catalytic triad ser-his-asp for peptidyl-S-enzyme chain growth was confirmed.



Scheme 3. Peptide bond formation on the surface of the C-domain. The amino group of the acceptor substrate performs a nucleophilic attack on the carboxy function of the donor substrate, thereby establishing a novel peptide bond and elongating the existing molecule by another amino acid residue.

Prior to novel peptide initiation, the just assembled molecule must be released from the NRPS machinery at the end of the assembly line, allowing a reactivation of the multienzyme complex. Thioesterase (TE) domains located at the C-terminal end of the final NRPS module catalyze this release reaction in multiple ways.[49] With a size of approximately 250 amino acids, TE domains share homologies with fatty acyl thioesterases. In general, peptides can be released via hydrolysis, thereby generating a linear molecule, or by undergoing an intramolecular cyclization reaction to obtain a circular product. Apart from the final cleavage, both pathways share a common mechanism of action: The peptide is transferred to the active site serine residue of the TE-domain by cleavage from the ppan-arm and subsequently forms an ester-linked intermediate via the terminal carboxyl group of the NRP. An internal nucleophile or water then attacks the acyl-O-intermediate and releases it from the TE-domain (Scheme 4).[56] Internal

nucleophiles can originate from different functional groups, intermediate deacylation and peptide cyclization can occur in various ways. If the peptides' N-terminal amino group is the nucleophile, a head-to-tail cyclization exhibiting maximum ring size, as seen for tyrocidine A, takes place. Branched cyclic molecules, e.g., gramicidin S, result from the attack of a side chain nucleophile and natural products such as surfactin A are made by nucleophilic attack of the β-hydroxy group of a β-hydroxy fatty acid side chain[60]. Surfactin A is a cyclic macrolactam synthesized by surfactin synthetase originating in *Bacillus subtilis*. The SrfTE-domain specifically catalyzes the intramolecular, regio-specific nucleophilic attack of an acyl chain. Its importance was confirmed by experimental studies, which, upon altering the catalytic triad, led to the formation of a linear product by hydrolysis instead of cyclization.[61]



Scheme 4. TE-Domain-mediated peptide release. After transfer of the polypeptide chain from the ppan arm of the PCP-domain onto an active site serine hydroxyl group, peptide release through bond cleavage is catalyzed either by hydrolysis to yield a linear peptide or by intramolecular cyclization yielding a circular product.

Alongside compulsory C-A-T-TE domains present in most NRPS modules, several additional enzymes can be found as part of the natural product biosynthetic machinery. Substrate alteration and peptide modification enhance structural diversity and introduce biological activity into natural product molecules. Tailoring enzymes can thereby act on two different levels: (a) during peptide assembly or (b) after release of the core peptide from the TE-domain of the NRPS system. Moreover, modifications during peptide assembly can be performed by domains directly integrated into NRPS modules or by enzymes interacting with on-line domains (*in trans*-modification).[62]

Methyltransferase (MT) domains are usually integrated between the A- and PCP-domains of a module and methylate the peptide intermediate prior to peptide bond formation. Amino acids are mainly *N*- or *C*-methylated, however, *O*-methylation has also been reported. As a result of

methylation, peptide bond stability increases, whereas susceptibility to proteolytic cleavage decreases. *S*-adenosyl methionine (SAM) serves as the methyl donor co-substrate.[49, 62]

Most A-domains selectively activate and incorporate L-amino acids, thus epimerization (E) domains are necessary for generating D-amino acid residues during peptide assembly. E-domains are located adjacent to the PCP-domain carrying the amino acyl-*S*-ppan bound substrate that is going to be epimerized.[45] In the absence of E-domains, external racemases can provide D-amino acids that are incorporated by a enantioselective A-domain or enantioselective donor sites can assist in the specific integration of the D-amino acid.[63]

The presence of cyclization (Cy) domains in NRPS modules indicates large structural reorganization and a change of residue connectivity along the peptide assembly line. Cy-domains share structural similarities with C-domains and possess a conserved motif DxxxxDxxS in the catalytic pocket.[64] Heterocyclization during elongation follows a two-step mechanism: First, a condensation reaction with resulting peptide bond formation is catalyzed. Afterwards, the cyclization reaction at the thiol or hydroxyl groups of cysteine, serine or threonine residues is initiated by proton abstraction. The resulting (thio)hemiaminal intermediates are then dehydrated to obtain thiazoline or oxazoline ring systems with the typical double bond.[62] Peptides containing ring systems are not only more resistant to hydrolysis, they can also function as metal chelators or intercalating groups and increase bioactivity as well as the affinity of the compound for the target molecule.[64]

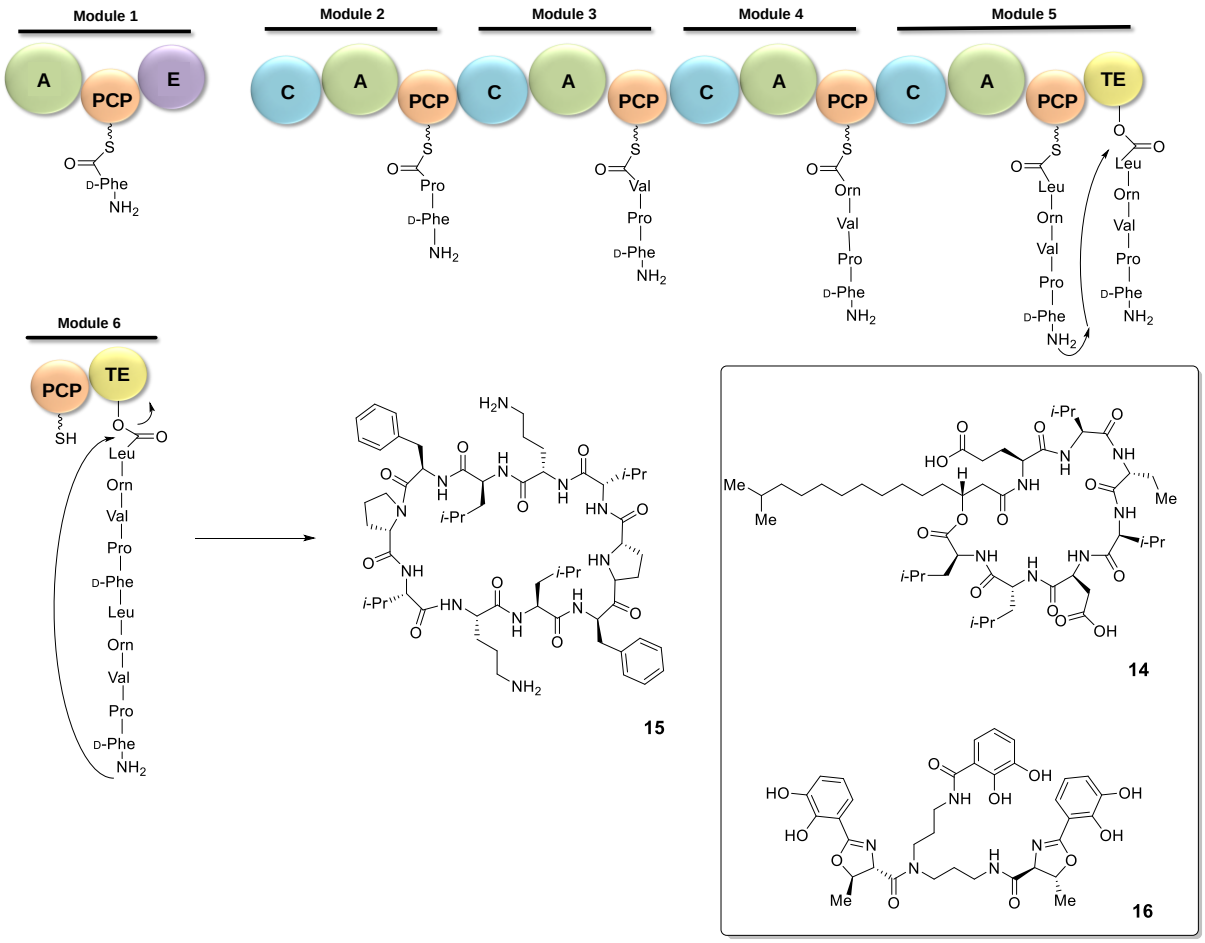
Apart from the previously mentioned tailoring enzymes, several additional auxiliary domains that structurally alter the basic building blocks and enhance structural diversity can be found in the modules of the NRPS biosynthetic machinery, including dehydratase (DH), formylase (F), ketoreductase (KR), monooxygenase (MOX), oxidase (Ox), halogenase (Hal) and reductase (R) domains. Each type of domain is specialized to catalyze a specific reaction and is therefore positioned at strategical parts of the respective module to act in a coordinated action. Furthermore, glycosyl transferases or enzymes catalyzing β-hydroxylation modify the core peptide in a post-assembly line way.[62]

1.2.1.2 Modularity of NRPS systems

The structure and size of the end product released by the TE-domain is usually defined by the number of modules present within the NRPS enzyme and the arrangement of single domains within each module. Assuming that peptide chain growth proceeds from the N-terminal to the C-terminal end of the NRPS machinery, co-linear relations between the order of the modules and the resulting amino acid sequence in the final product exist. Most NRPS systems follow this “co-linearity rule”, however some multienzyme complexes show more complicated domain organization that deviates from the standard C-A-PCP logic. Thus, NRPS systems can be assigned to one of three groups according to the biosynthetic assembly strategy they follow to synthesize a natural product.[65]

Type A NRPS systems assemble peptides following the “co-linearity rule” and incorporate amino acids in a straightforward, linear downstream assembly, with every module incorporating a single residue into the elongating peptide chain.[38] The biosynthesis of surfactin A (**14**) is catalyzed by such an NRPS, where the formation and release of the cyclic lipopeptide occurs after being passed along the complete assembly line consisting of 24 domains organized into seven modules.[66] Iterative NRPSs are classified as type B systems, which in principle also assemble peptides in a linear way. However, they contain a minimal set of modules that are reused multiple times to generate a multimeric product consisting of identical units. Synthesis of an initial peptide does not end with its release from the TE-domain but instead it stays attached while the same peptide is produced again (and again). Only when the correct amount of peptides has been made by the NRPS does the TE-domain release the polypeptide by a cyclization reaction, as can be seen for gramicidin S (**15**) biosynthesis.[67] The gramicidin assembly line consists of five modules that function in a linear way, thereby consecutively producing two identical pentapeptides. The TE-domain of the final module promotes head-to-tail cyclization and homodimerization (Scheme 5).[68] Type C systems are non-linear and their domain arrangement varies from the typical C-A-PCP. Not every module harbors all essential catalytic units. Nevertheless, full functionality is enabled by domains of other modules within the assembly line that take over the missing function. It has been reported that some biosynthetic pathways include nonfunctional domains as well as cases, where one enzyme catalyzes two different reactions. Additionally, small molecules not covalently bound to the NRPS machinery can be incorporated into the NRP.[69] One example is the NRPS system involved in vibriobactin (**16**) biosynthesis. Vibriobactin assembly is rather complicated as it combines variations in domain organization with alternate catalytic reactions. Studies concerning the Cy-Cy-A-C-PCP-C domain arrangement revealed that the first C domain is

catalytically inactive while the second C domain exploits full activity. Furthermore, synthesis is divided into two separate pathways in the middle of the assembly line and norspermidine is incorporated into the molecule although it is not bound to any NRPS unit.[70] Due to their variation in peptide assembly or domain organization, especially type B and C NRPSs contribute to natural product diversity and molecular structural complexity.

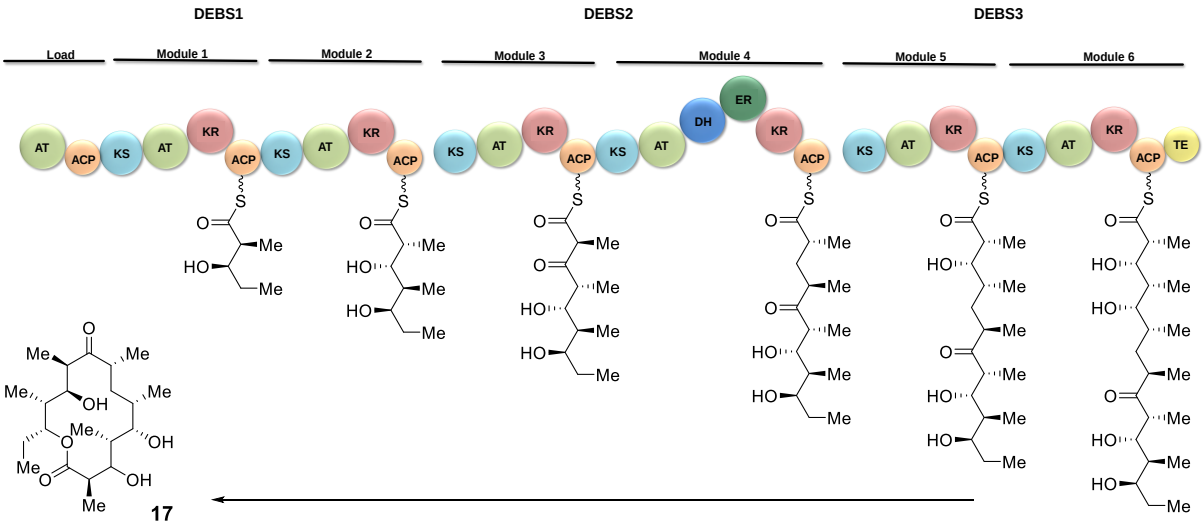


Scheme 5. Type B NRPS-catalyzed gramicidin S biosynthesis. The assembly line consists of two NRPSs, the first consisting of a single module and the second made up of four modules. Each module incorporates one amino acid per round and when working iteratively, two pentapeptides can be coupled followed by intramolecular cyclization to yield **15**. Highlighted in the box are compounds **14**, a molecule built by type A NRPSs and **16** which results from activity of type C systems.

1.2.2 Polyketide biosynthesis

Polyketides make up another class of natural products that exhibit great structural and functional diversity as well as a range of biological activities, thus making them interesting targets for potential future drug applications. They are ubiquitously found in Nature, with many derivatives being isolated from bacteria, fungi and plants.[71] Like nonribosomal peptide biosynthesis, the genes involved in polyketide biosynthesis are clustered as BGCs within the genome. Despite the large number of different polyketides identified, many share a common biosynthetic pathway catalyzed by multienzyme complexes called polyketide synthases (PKSs). Like NRPSs, PKSs are organized as modules which again consist of catalytically active domains and they function in an assembly line mode, sequentially adding units to a growing ketide chain. Variations in the order of unit incorporation or repeated activity of a single domain give rise to three types of PKS systems (I: noniterative, II: iterative, III: ACP-independent & iterative). PKSs show analogies to fatty acid synthases and polyketides are thus synthesized similarly to fatty acids.[72] The principal domains involved include the ketosynthase (KS), acyl transferase (AT), acyl carrier protein (ACP), ketoreductase (KR), dehydratase (DH), enoylreductase (ER) and the thioesterase (TE). In principle, biosynthesis comprises repetitive rounds of decarboxylative Claisen condensation between a malonyl thioester derivative (extender unit) and an acyl thioester (starter unit).[73] AT-domains are one of three obligatory enzymes necessary for polyketide formation and they initiate transacylation of extender units such as malonyl or substituted malonyls bound to coenzyme A to the free thiol of the ACP ppan moiety. The ACP-domain with its flexible ppan arm binds substrates and intermediates and transports these via protein-protein interactions between AT- and KS-domains. The catalytic interface of KSs features a donor and an acceptor site for the receival of the polyketide chain from an upstream ACP-domain and subsequent execution of a Claisen condensation between the received polyketide and a new extender unit, thereby promoting chain elongation.[74] In fatty acid biosynthesis, the resulting β-ketothioester intermediates are then subject to ketoreduction, dehydration and enoylreduction, yielding saturated fatty acids. These steps can be partly or completely abolished in PKSs to generate different types of polyketides.[75] One example of a type I PKS is the 6-deoxyerythonolide B synthase (DEBS) which synthesized the precursor polyketide **17** of **4** (Scheme 6). A linear assembly line consisting of six modules repeatedly incorporates (2*S*)-methyl malonate derived propionate into a growing acyl chain where stereochemistry and β-keto group processing are regulated by each module. Terminal cleavage and cyclization yield the polyketide 6-deoxyerythronolide B.[73] Taken together, polyketide diversity is achieved by using a variety of substrates and extender units, by alternations in the degree of processing after each condensation step and

by stereochemical control after β-ketoacyl thioester reduction. Furthermore, tailoring domains on-line as well as off-line modifying enzymes contribute significantly to natural product versatility.[71]

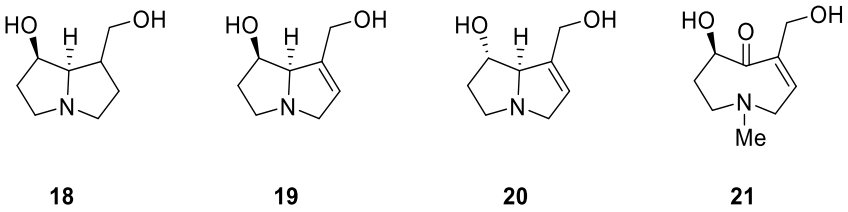


Scheme 6. Type I PKS assembly line for biosynthesis of 6-deoxyerythonolide B (**17**). The compulsory AT-, ACP- and KS- domains catalyze malonyl and acyl unit incorporation, polyketide chain translocation and elongation. Tailoring domains present within a module modify the growing chain and thereby regulate natural product structure.[76]

1.3 Pyrrolizidine alkaloids

1.3.1 Chemistry and function

Pyrrolizidine alkaloids (PAs) are secondary metabolites belonging to the natural product class of alkaloids and are thus amino acid-derived, nitrogen-containing heterocyclic molecules. The core unit consists of a pyrrolizidine ring system made of two five-membered rings which can be saturated or unsaturated and a central nitrogen bridge atom.[77] In Nature, the free pyrrolizidine core hardly ever occurs, instead *N*-oxides or esters resulting from the reaction of an amino alcohol (necine base) with mono- or dicarboxylic aliphatic acids (necic acids) are the predominant species. According to their necine base structure and the functional groups (e.g. carboxyl, hydroxyl groups or double bonds) present in the molecule, PAs are assigned to belong to one of four different types: retronecine (**18**), otonecine (**19**), platynecine (**20**) and heliotridine (**21**; Scheme 7).[78]



18 **19** **20** **21**

Scheme 7. Overview of pyrrolizidine core structure types. **18 – 21** show variations in degree of saturation, stereochemistry and functional groups attached to the pyrrolizidine bicycle.

To date, most PAs identified are plant-derived, especially originating from plants of the *Leguminosae*, *Compositae*, *Orchidaceae* and *Boroginaceae* families. Over 660 PAs and *N*-oxide derivatives have been found in more than 6000 different plants and it is assumed that 3 % of the vegetation worldwide produce these secondary metabolites.[78] Biosynthesis and secretion are undertaken to protect the producing organism against herbivores, which likely is why half of the plant PAs show toxic potential. Although yet underrepresented, PAs of fungal and bacterial origin are predicted to be numerous and chemically diverse and thus await to be discovered (Figure 6).[79]

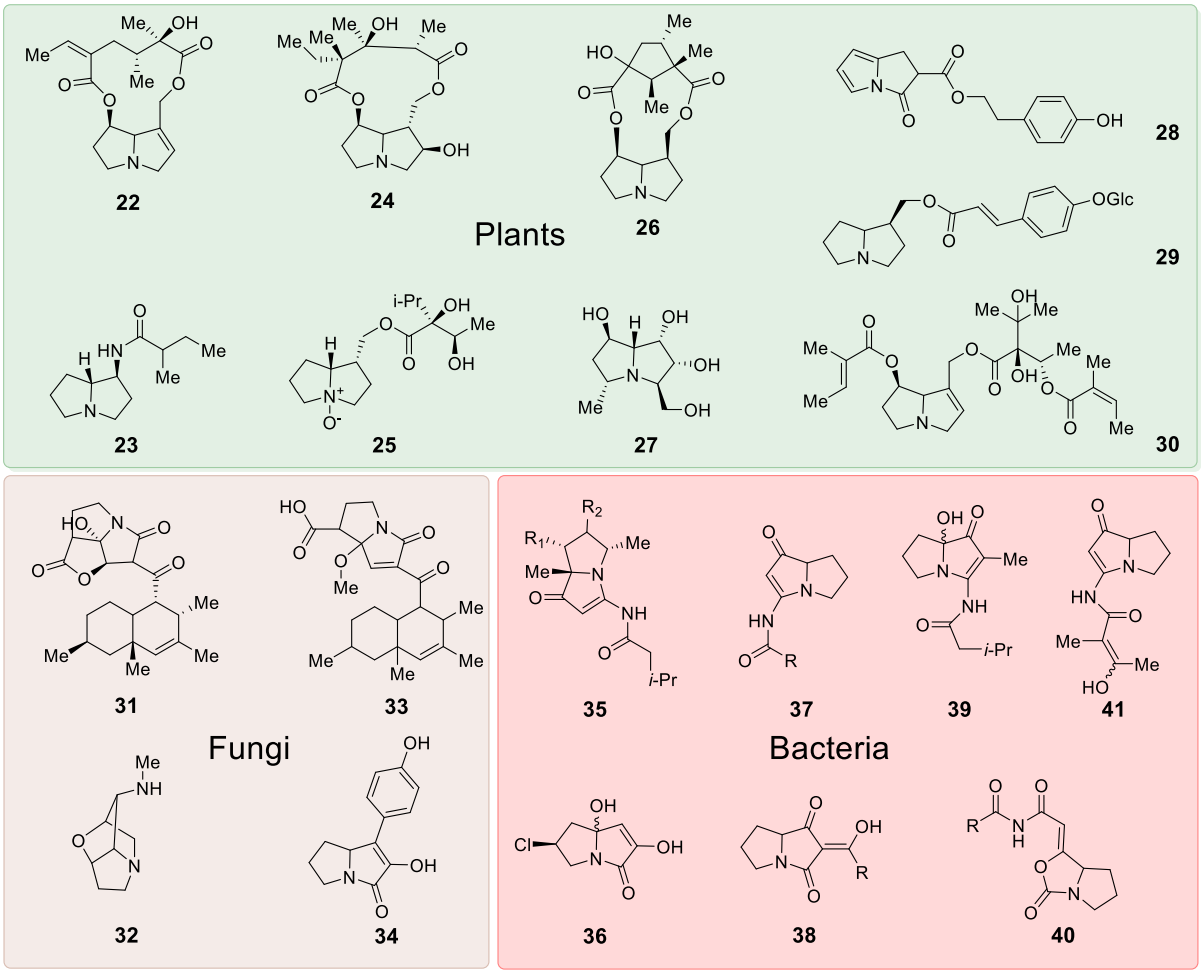
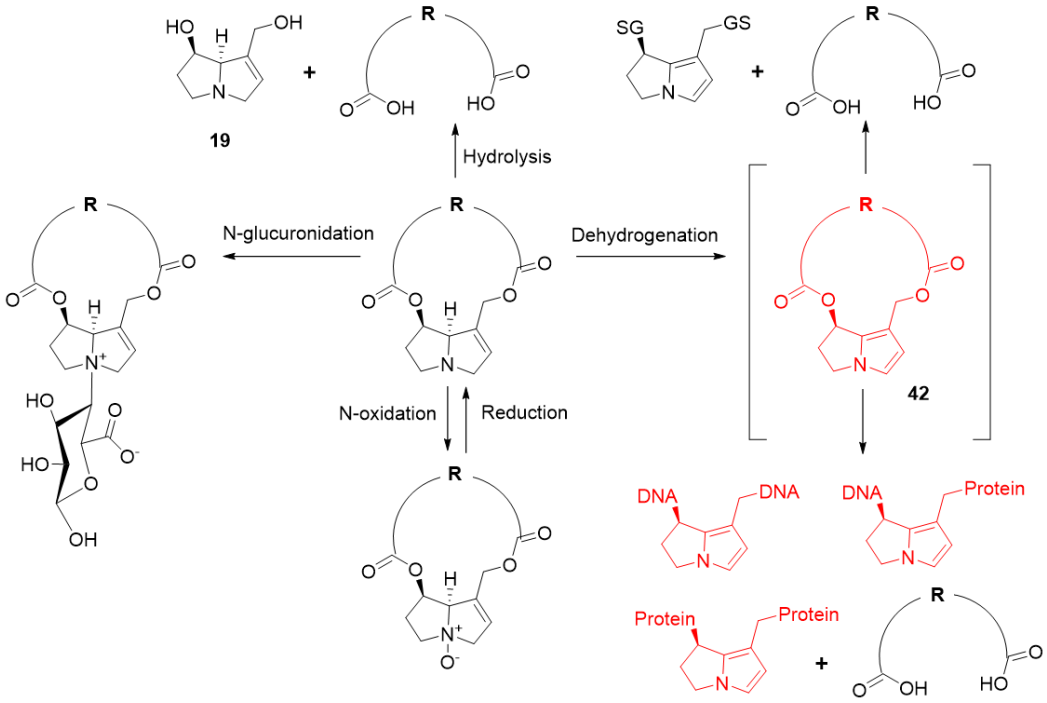


Figure 6. Prominent PA representatives of plant, fungal and bacterial origin. Independent of the species they were discovered in, all examples have the pyrrolizidine core in common.

1.3.2 Toxicity and metabolism

Originally, PAs were discovered while studying symptoms of intoxication occurring in livestock, wildlife and humans after ingesting PA-containing plants or food contaminated with these secondary metabolites, such as honey, herbal teas, salad, grains and milk. Consequently, studies addressing plant PA metabolism and toxicity revealed that metabolic activation upon ingestion is required for toxicity to unfold.[80] In general, unsaturated PAs, including those of the retronecine, heliotridine and otonecine type, exhibit a higher amount of toxicity than saturated platynecine type PAs.[81] Four principal mechanisms are used inside the cell to metabolize PA molecules, with one of them causing the formation of highly reactive and toxic metabolites. The hydrolysis of ester functional groups linked to *C*(7) or *C*(9) carbon atoms splits the molecule back into necine base and necic acid, *N*-oxidation converts necine bases to *N*-oxides

and glucuronidation via uridine diphosphate glucuronosyl transferases yields *N*-oxides and conjugated dienic pyrroles which can be excreted with the urine. Solely the two-step oxidation of PAs via hydroxylation of the necine base followed by a spontaneous dehydration leads to the formation of the reactive alkylating agent dehydropyrrolizidine alkaloid (DHPA, **42**). Processing of otonecine type PAs is initiated by *N*-demethylation of the necine base with subsequent ring closure by elimination of formaldehyde and spontaneous dehydration.[82-83] **42** is highly reactive as it non-specifically binds to cellular protein and intercalates with DNA, thereby promoting cytotoxicity as well as genotoxicity. In addition, pyrrole-protein-protein, pyrrole-DNA-DNA, and pyrrole-DNA-protein crosslinks can lead to loss of cellular function. Excretion and detoxification of reactive PA metabolites is possible when binding to glutathione (GSH) occurs (Scheme 8). Since the bioactivation is predominantly catalyzed in the liver by P450 enzymes, esterases and flavine containing monooxygenases, this organ is especially affected by toxicity issues, including acute and chronic pyrrolizidine alkaloidosis (HSOS, VOD). Apart from hepatotoxicity, PA metabolites can also be carcinogenic, teratogenic, genotoxic, and even pneumotoxic.[78, 83-84]



Scheme 8. Metabolic processing of plant PAs. Hydrolysis, *N*-glucuronidation and *N*-oxidation lead to the formation of non-toxic metabolites that can be excreted via the urine. Dehydrogenation produces reactive intermediate DHPA (**42**) species, which conjugate with DNA or protein, leading to crosslinking and loss of macromolecular functionality. Cellular GSH catches redox-active intermediates and supports excretion, thereby functioning as a detox mechanism.[84]

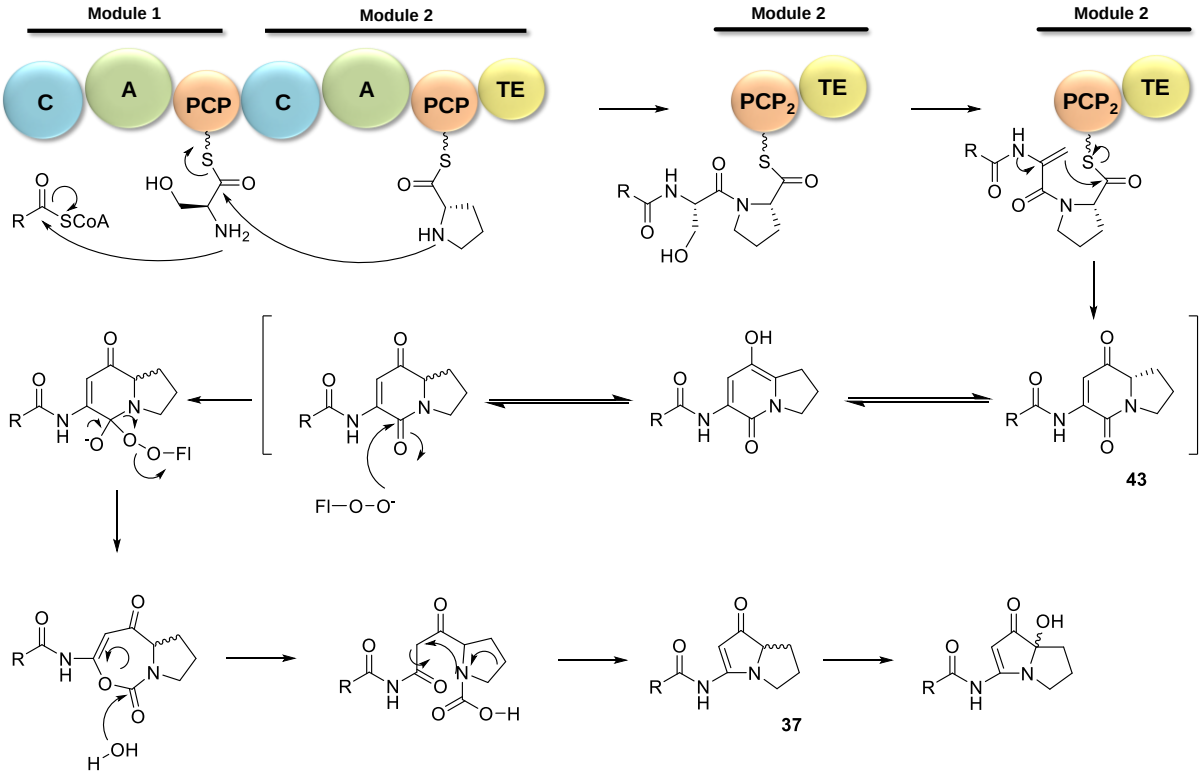
1.3.3 Biosynthesis

Although all PAs share a common core structure and display similar biological activities, the biosynthetic mechanisms involved in secondary metabolite production differ in plants, fungi, and bacteria. Due to their wide distribution, biosynthesis of plant-derived PAs is well studied and includes multiple reaction steps starting with the amino acid arginine. Decarboxylation of arginine gives rise to the intermediate agmatine which is further transformed to putrescine by elimination of amide functionalities from *N*-carbamoyl putrescine. Putrescine can either be directly used as a substrate or further converted to spermidine and spermine. These polyamines are the basic building blocks for NAD^+-dependent homospermidine synthesis catalyzed by homospermidine synthase (HSS). HSS is the initial enzyme specific to PA biosynthesis and actively links primary and secondary metabolism to each other.[85] The aminobutyl group of a spermidine molecule is transferred to putrescine to give a triamine homospermidine, which is further subject to cyclization and reduction steps to yield the PA backbone structure.[86-87]

Up to the present day, only a handful of fungal PA molecules have been identified, therefore little is known about a general biosynthetic pathway. Nevertheless, studies on loline alkaloids (**32**) produced by endophytic fungi living in symbiosis with cool-season grasses revealed a mechanism different to plant PA biosynthesis.[88-89] The amino acids L-proline and L-homoserine are coupled in a condensation reaction to a loline intermediate, which further undergoes oxidative carboxylation and cyclization to establish the core loline ring structure. After ring system formation an oxygen bridge is introduced between *C*(2) and *C*(7), resulting in the installation of a third ring into the final product.[90]

A few more examples of bacterial PA biosynthesis have been identified and analyzed so far and with the rise of whole genome sequencing in combination with bioinformatic tools such as antiSMASH, many more BGCs encoding enzymes involved in PA production were discovered and await in-depth analysis. Biosynthesis is principally executed by NRPSs and monooxygenases (MOXs) and the assembly of the core structure unit of known bacterial vinylogous urea-type PAs (vuPAs) proceeds by a similar mode of action for a variety of derivatives. The amino acids L-serine, L-proline and L-threonine or derivatives thereof in combination with a linear, branched, or hydroxylated ketoacyl thioester are used as the basic building blocks. Additional enzymes can be of need to modify single substrates, e.g., glycosyl transferases or SAM-dependent enzymes. After coupling of the amino and fatty acid moieties in a linear NRPS assembly line, an unusual TE-domain catalyzes the dehydration of the incorporated L-serine residue to dehydroalanine (DHA) with subsequent cyclization and

peptide release from the NRPS to give a [5,6]-bicyclic intermediate **43**. This intermediate is further processed by an FAD-dependent MOX, which catalyzes a reaction cascade initiated by a Baeyer-Villiger oxidation reaction leading to the expansion of the 6-membered ring. Hydrolysis, ring contraction with decarboxylation and hydroxylation reactions finalize conversion of the intermediate to the corresponding [5,5]-bicyclic PA derivative (Scheme 9).[79, 91] Multiple examples of vuPA biosynthesis have shown that these two enzymes are sufficient for pyrrolizidine backbone generation and support the hypothesis that tailoring enzymes contribute in an essential way to natural product diversity.



Scheme 9. Principle mechanism of vuPA biosynthesis. NRPS enzymes initiate the formation of a heterocyclic intermediate before an FAD-dependent MOX catalyzes the conversion to the PA core structure by a Baeyer-Villiger oxidation-initiated reaction cascade.

1.3.4 Monooxygenase-catalyzed reactions

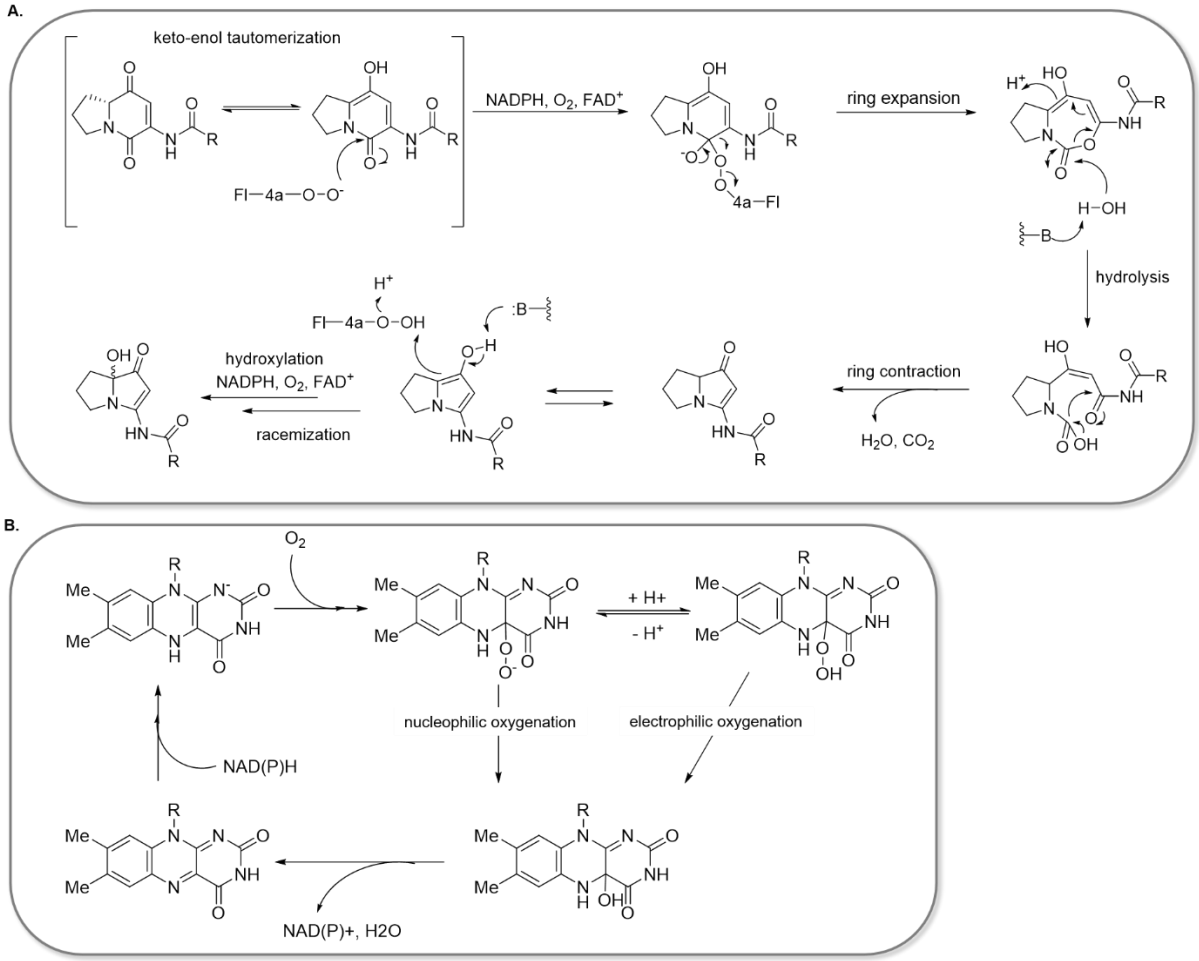
Monooxygenases (MOX) are present in all kingdoms of life and of special interest for PA biosynthesis. Its functionality as an oxidoreductase, which catalyzes electron transfer for molecular oxygen activation, is of great importance for many metabolic pathways. In particular, MOXs catalyze insertion of a single *O*-atom, leading to oxygenation or hydroxylation of the substrate. For generation of the reactive oxygen intermediate species, most MOXs depend on the presence of a cofactor, such as flavine, heme, pyrroquinoline quinone or metal ions.[92]

For instance, heme-dependent cytochrome P450 MOXs of microbial origin play a pivotal role in fatty acid degradation by hydroxylating *n*-alkanes to fatty acids, as described by Maier et al. in 2001 for *Acinetobacter sp.* EB104 CYP153.[93] For methanotrophic bacteria, such as *Methylococcus capsulatus*, copper-dependent MOXs were found to oxidize methane, thus allowing the resulting products to be used as carbon and energy sources by the bacterium.[94]

Flavine-dependent MOXs are ubiquitous amongst prokaryotes and they catalyze multiple reactions, including epoxidations, halogenations, sulfoxidations, and Baeyer-Villiger oxidations, the latter taking place during PA biosynthesis.[95] The type of oxygenations being catalyzed by a specific MOX depend on the shape and chemical nature of the active site of the enzyme.[96]

In PA biosynthesis, formation of the pyrrolizidine core depends on a Baeyer-Villiger MOX (BVMO), which, in many cases, initiates an entire reaction cascade. This sequence of reactions includes ring expansion, hydrolysis, decarboxylation and ring contraction as well as hydroxylation to convert a 5,6-bicyclic NRPS-derived intermediate into the final 5,5-bicyclic backbone (Scheme 10A).[79]

However, independent of the reaction being catalyzed, the mechanism of substrate activation through generation of a reactive intermediate remains the same for all oxygenations. Since the reduced flavine species is needed to activate molecular oxygen, NADPH functions as an electron donor. Reduced flavine can then build a covalent bond to molecular oxygen at its C(4a)-position to form a reactive (hydro)peroxyflavine intermediate. In the following, the substrate can be oxygenated by nucleophilic or electrophilic attack of the reactive species, resulting in the incorporation of a single *O*-atom, whereas the second oxygen atom is reduced to water (Scheme 10B).

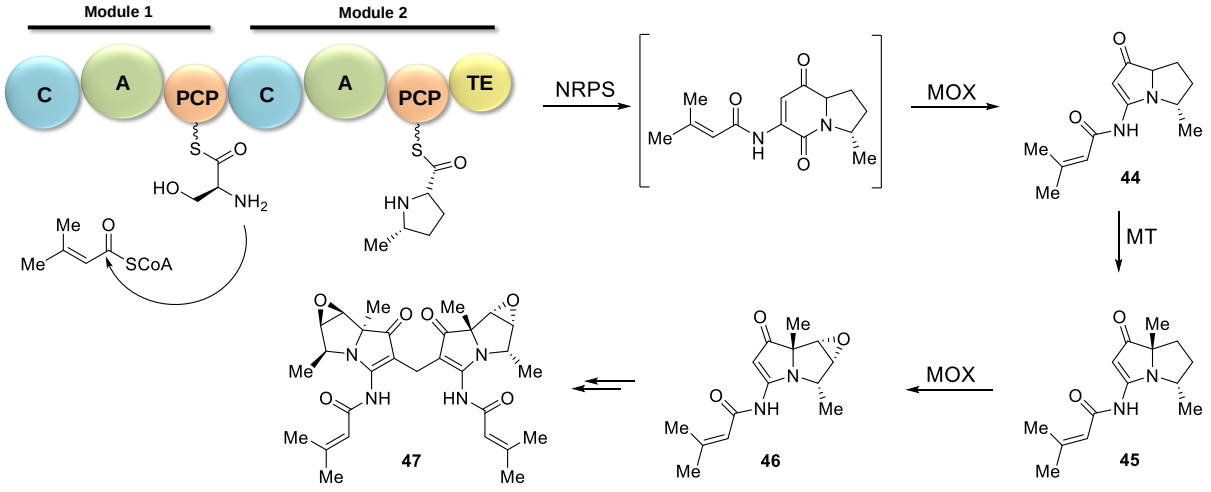


Scheme 10. FAD-dependent BVMO-catalyzed reaction cascade in PA biosynthesis. **A.** Multistep conversion of a 5,6-bicyclic intermediate to the 5,5-bicyclic PA product.[79] **B.** Mechanism of BVMO-catalyzed oxygenation reactions in dependence of FAD and NADPH.[96]

1.3.5 Potential of bacterial PAs in disease treatment

Of the bacterial vuPA molecules identified to date, a few exhibit interesting biological activities, making them appealing for putative future drug applications. Many of them show antifungal or antibacterial activity against a broad range of human pathogens, e.g., brabantamides (**40,** see Scheme 6).[97] PA derivatives also find application in atherosclerosis treatment, where lipocyclocarbamates (LCCs) selectively inhibit lipoprotein associated phospholipases (Lp-PLA_2), proteins involved in the inflammation of atherosclerotic plaques. The chemically synthesized LCC derivative darapladib is currently undergoing phase III clinical trials.[98] One bacterial vuPA, mitomycin C (**8,** see Figure 3), already finds broad application as an antitumor drug. After metabolic activation by quinone reduction, the bacterial cell cycle can be arrested

and subsequently apoptosis is induced as a result of DNA-crosslinking activity.[20] Furthermore, **8** was found to suppress T-cell mediated autoimmune response reactions, as found for transplanted organs or in multiple sclerosis.[99] Isolated from *Streptomyces* species, the clazamycins (**36**) drew attention due to their unique structure including a chlorine substituent which has been seen in several established antitumor agents. Compared to many other PAs, **36** shows no signs of hepato-, nephron- or spleen toxicity, making them even more favorable for drug development.[100] The mechanism of action against leukemia cells L-1210 includes the inhibition of cell surface components participating in transmembrane transport of nutrients such as amino acids and sugars essential for DNA replication, leading to cell disintegration and death.[101] Antiproliferative activity and cell cytotoxicity against the lung cancer cell line A549 and liver cell cancer cell line have also been demonstrated for dimeric analogues of the bohemamines (**46**), the dibohemamines (**47;** Scheme 11).[102-103] The mentioned examples of vuPA derivatives originating in bacteria give an insight into their variety of biological activities, broad applicability, and the great potential for future drug leads, thus making detailed studies on mechanistic procedures including biosynthesis as well as the mode of action on target molecules desirable.



Scheme 11. Bohemamine A biosynthetic pathway. Modified basic building blocks are coupled to establish the pyrrolizidine core following the general NRPS-MOX-catalyzed mechanism. A MT installs an additional methyl group to obtain NP25302 (**45**) before MOX catalyzes epoxide formation to yield vuPA **46**. Dimerization of **46** leads to generation of cytotoxic dibohemamine A (**47**). Figure adapted from Nah et al. [103]

2. Aims of this thesis

The existing examples of vuPAs originating in bacteria have given an insight into this yet underrepresented group of natural products and the potential they bring for future drug development. The variety of bacteria producing PAs gives rise to a huge pool of novel molecules and unlocking their potential will be of great interest and importance. It is thus the aim of this thesis to identify, isolate and characterize novel members of the vuPA family, especially derivatives of **37**, from a selection of bacteria encoding genes predicted to participate in molecular assembly.

2. 1 Heterologous expression of novel PA-encoding gene clusters

Initially, a selection of bioinformatic tools is applied to identify BGCs of different bacterial origin harboring the core genes encoding the NRPS and FAD-dependent (MOX) enzymes proven to be involved in the biosynthesis of **37**. Based on the available DNA sequences of these core genes, genome mining led to the identification of many novel putative vuPA biosynthetic pathways with highly diverse BGC organization. Ten phylogenetically different bacterial strains of both proteobacterial and actinobacterial origin possessing homologues of the core vuPA genes were chosen for further investigation. Whereas most selected PA-related BGCs show an increased gene complexity compared to the **37**-producing BGC of *Xenorhabdus stockiae* DSM17904, two share an identical setup and will be tested as a proof-of-principle approach.

Heterologous expression of vuPA BGCs is tested in various host systems, including several *E. coli* and *Streptomyces* strains. BGCs of interest must be captured from genomic DNA either using PCR-amplification or restriction digest and subsequently inserted into suitable vector systems. For expression in *E. coli*, vectors pET28bSUMO, pET28bpTetGFPv2, pCX2 and pBAD_RSF1031K are used, whereas pSET152ermE*rev is applied for *Streptomyces* expression. Cloning is conducted either using the “Direct Pathway Cloning” (DiPaC) strategy or by proceeding with linear-linear / linear-circular homologous recombination (LLHR / LCHR). As soon as target BGCs have been captured successfully, the constructed plasmids will be transformed into adequate host cells (*E. coli* BL21DE3 and BAP1 and *Streptomyces albus* B2P1 and Del14, *S. coelicolor* M1154, *S. lividans* TK24) for BGC expression. Due to their high GC-content and overall large cluster size (8000 to 30000 kilo base pairs (kbp)), refactoring of BGCs may be inevitable to remove putative regulatory elements, such as transcription terminators, and to thus enable vuPA biosynthesis. Expression parameters, like cultivation

temperature and time, type of culture medium and pH, batch size and even supplementation with additives to support biosynthesis, will be altered and optimized along the way. Isolation of produced secondary metabolites will be performed using Amberlites XAD-16 and ethyl acetate extraction. The resulting raw extracts are then analyzed by LC-MS and promising compounds can be further purified with preparative HPLC to undergo characterization studies (NMR, MS-MS). Once a novel vuPA derivative has been successfully produced and characterized, an optimization of the heterologous expression conditions followed by upscaling is intended to increase the natural product yield.

2.2 Stepwise reconstitution of the vuPA biosynthetic pathway

The second part of this thesis addresses the mechanisms of vuPA biosynthesis in more detail. *In vitro* approaches include the expression of single enzymes participating in vuPA assembly, with a focus set to NRPS systems and MOXs. After expression in *E. coli* host systems, proteins will be purified by affinity chromatography and subsequently tested for activity in isolated enzyme assays. Once a fully functional testing system has been established for single enzymes, it is the aim to obtain vuPA derivatives in traceable amounts inside a reaction tube by combining multiple purified enzymes together with the necessary co-factors and co-substrates. Furthermore, *in vitro* total biosynthesis can be manipulated by variation of the basic building blocks, e.g., fatty acids of different length and degree of saturation, and will hopefully give detailed information on substrate promiscuity and the enzymes' modes of action. Apart from the bimodular NRPS system, the substrate scope of the monomodular NRPS system Pys involved in pyreudione (**38**) biosynthesis will be investigated. The generation of novel putative vuPA derivatives *in vitro* is then going to be tested by the combination of enzymes from different predicted vuPA pathways. To support these mechanistic studies, *in vivo* experiments expressing truncated versions of the respective BGC will be done to compare product outcome. Isolation of new products and intermediates by varying substrates in feeding experiments, heterologously expressing one BGC in different gene constellations or combining functional genes from different pathways is desirable to shed light onto the biosynthetic pathway of this class of molecules. If possible, the MOX enzyme is going to be crystallized alone as well as co-crystallized with co-factors (FAD, NADPH) and substrates to gain insight on the binding states during Baeyer-Villiger oxidation and find out what causes MOX homologues to synthesize different molecules, as was observed for the pyrrolizixenamides (**37**) versus the related brabantamides (**40**).

3. Results and Discussion

3.1 Bioinformatics

3.1.1 Genome mining

Within the course of this work, mining for genes encoding core biosynthetic enzymes involved in PA biosynthesis, namely the NRPS and MOX systems, was performed using the NCBI database BLAST function to search for sequence homologies. The obtained data was combined with AntiSMASH[33], a tool used to identify and analyze BGCs within whole genome sequences. Bacterial genomes were mined using the available DNA and amino acid sequences of the NRPS and MOX genes from *Xenorhabdus stockiae* DSM17904 which produce **37**[79] and those encoding both genes of interest were preselected. Finally, approximately ten BGCs of varying size and complexity were chosen to further work with throughout the course of this thesis (Figure 7).

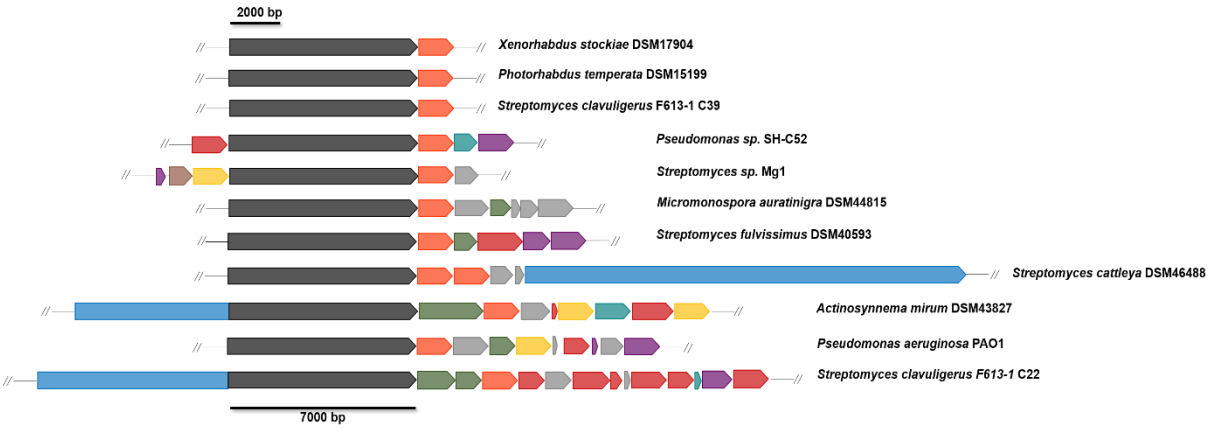


Figure 7. Overview of BGC organization in selected proteo- and actinobacteria. Gene functions are visualized by color coding: NRPS (black), FAD-dependent MOX (orange), short chain dehydrogenase (green), cytochrome P450 (yellow), S-adenosylmethionine transferase AdoMeT (brown), type-I PKS (blue), transport systems/cell signaling components (purple), other modifying enzymes (red) and transcriptional regulators (turquoise).

As can be seen in Figure 7, representatives of both proteobacteria and actinobacteria phyla were selected to study. Despite their variation in BGC complexity as well as cluster sizes, ranging from eight to over thirty kilo base pairs (kbp), all comprise the core NRPS and MOX genes. Due to the variability of tailoring genes, the potential for finding novel PA derivatives by expression of the selected BGCs was promising.

With the help of AntiSMASH the localization of each BGC of interest within the bacterial genome, its gene composition as well as the putative function of tailoring genes was analyzed.

Furthermore, predictions concerning the core structure synthesized by the NRPS system were made, which indicated whether the respective BGC is of interest to our search for PA-producing gene clusters or not. For target BGCs, NRPS systems were predicted to incorporate L-serin and L-proline to generate the heterocyclic PA core. Initially, genomes of interest encoding putative PA BGCs were analyzed using AntiSMASH version 2. Throughout the project, updated versions of AntiSMASH up to recent version 5.2.0 reassigned BGC numbers and elucidated similarities to known clusters. For *Pseudomonas aeruginosa* PAO1, functions to former hypothetical genes (when analyzed in 2017) could be assigned throughout the course of this thesis, with the first hypothetical protein being annotated as protein PvdD and the second hypothetical protein being annotated as an isomerase (in 2020). Furthermore, additional genes could be identified, and revision of gene functionality took place for one of the two BGCs of interest encoded in the *Streptomyces clavuligerus* F613-1 genome. An overview of known data for the chosen bacterial BGCs of interest is given in Table 1.

Table 1. Overview of collected information on BGC organization, size, GC-content, and localization within the native bacterial genome.

Strain name	Notation	BGC # (2017) or region (2020)	BGC size [bp]	GC content [%]	Number of genes (genes with known function)
Micromonospora auratinigra	DSM44815	# 21	13613	70	7 (3)
Streptomyces fulvissimus	DSM40593	# 2 Region 1.2	12426	71	6 (6)
Xenorhabdus stockiae	DSM17904	# 19	8388	43	2 (2)
Photorhabdus temperata	DSM15199	# 12 Region 1.11	8382	44	2 (2)
Streptomyces cattleya	DSM46488	# 37 Region 1.14	29109	75	6 (4)
Actinosynnema mirum	DSM43827	# 53 Region 1.19	24755	77	9 (8)
Pseudomonas aeruginosa	PAO1	# 29 Region 1.9	15511	70	10 (7)
Streptomyces sp.	Mg1	# 5 Region 1.3	13491	71	6 (5)
Streptomyces clavuligerus	F613-1	# 22 Region 1.13	31002	74	14 (9)
Streptomyces clavuligerus	F613-1	# 39 Region 1.21	8356	75	2 (2)
Pseudomonas sp.	SH-C52	Region 1.1	12786	63	5 (5)

To support the hypothesis that NRPS and MOX core enzymes of all BGCs of interest are related and thus catalyze a similar PA backbone formation reaction, studies of homology based on the nucleotide sequences of the mentioned genes were performed. Sequence homologies and nucleotide conservation of the core biosynthetic PA-producing enzymes within the whole genomes of selected bacteria were elucidated using the Multiple Sequence Alignment Viewer 1.18.0 (NCBI) and MAFFT (Multiple Alignment using Fast Fourier Transformation) tools (EMBL-EBI). Phylogenetic relationships of bacterial strains encoding PA-producing BGCs were visualized with Evolview v3.[104]

3.1.2 Homology analysis of NRPS systems

The bar diagram in Figure 8 expresses a high level of NRPS residue conservation amongst all chosen bacterial strains. Interestingly, even hybrid PKS-NRPS systems such as the ones found in *Actinosynnema mirum* DSM43827 and *Streptomyces clavuligerus* F613-1 cluster 22 feature high sequence homology in the NRPS part of the gene.

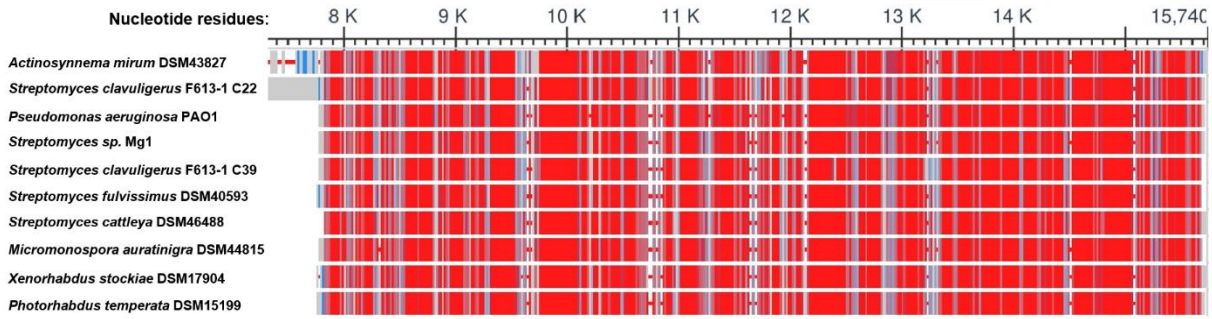


Figure 8. Representation of NRPS system sequence homology throughout the selection of bacterial strains. Highly conserved regions are shown in red, less conserved regions in blue and nucleotide variations are represented in grey.

To further emphasize the genetic similarities of the selected BGCs, the multiple sequence alignment coverage was illustrated as a heat map (Figure 9). The heat map shows that most of the bacterial strains have relatively high sequence similarities ranging between 60 – 75 % coverage. Interestingly, some strains belonging to the proteobacteria phylum such as *Xenorhabdus stockiae* DSM17904 and *Photorhabdus temperata* DSM15199 possess high sequence homology to each other, whereas the sequence coverage with actinobacterial strains is less than 50 %.

% similarity	DSM44815	DSM40593	DSM17904	DSM15199	DSM46488	DSM43827	PAO1	SMg1	F613-1 C39	F613-1 C22
DSM44815	100.0	64.40	47.12	48.56	66.47	59.15	63.23	64.50	65.50	58.51
DSM40593	64.40	100.0	47.39	48.43	65.86	60.70	62.30	66.28	71.26	59.98
DSM17904	47.12	47.39	100.0	61.27	46.19	42.87	46.05	46.31	45.90	43.50
DSM15199	48.56	48.43	61.27	100.0	46.92	43.74	47.40	47.54	47.55	44.30
DSM46488	66.47	65.86	46.19	46.92	100.0	63.29	65.34	65.53	68.44	61.21
DSM43827	59.15	60.70	42.87	43.74	63.29	100.0	59.56	59.95	61.36	73.07
PAO1	63.23	62.30	46.05	47.40	65.34	59.56	100.0	62.41	63.38	57.71
SMg1	64.50	66.28	46.31	47.54	65.53	59.95	62.41	100.0	67.36	58.67
F613-1 C39	65.50	71.26	45.90	47.55	68.44	61.36	63.38	67.36	100.0	61.21
F613-1 C22	58.51	59.98	43.50	44.30	61.21	73.07	57.71	58.67	61.21	100.0

Figure 9. Comparison of NRPS nucleotide sequences. The amount of similarity between two nucleotide sequences is given in %. Red coloring represents higher sequence homology, yellow represents lower sequence homology.

Figure 10 visualizes an overall close phylogenetic relationship within the phyla and decreased homology between proteo- and actinobacterial representatives. Nevertheless, the upper branching of the phylogenetic tree highlights a closer relationship between the NRPS systems of *Xenorhabdus stockiae* DSM17904 and *Photorhabdus temperata* DSM15199 with the NRPS sequences of actinobacterial hybrid PKS-NRPS systems found in *Actinosynnema mirum* DSM43827 and *Streptomyces clavuligerus* F613-1 cluster 22 than with that of the proteobacterium *Pseudomonas aeruginosa* PAO1. Furthermore, members of the actinobacterial phylum can be relatively distant from each other despite their high amount of nucleotide conservation of up to 75 %, as it is the case for *Streptomyces cattleya* DSM46488 in relation to *Streptomyces fulvissimus* DSM40593, *Streptomyces clavuligerus* F613-1 cluster 39 and *Streptomyces sp.* Mg1 NRPS genes. These differentiations can most likely be explained as the result of evolutionary processes and adaptation mechanisms of the organisms to their respective environment.

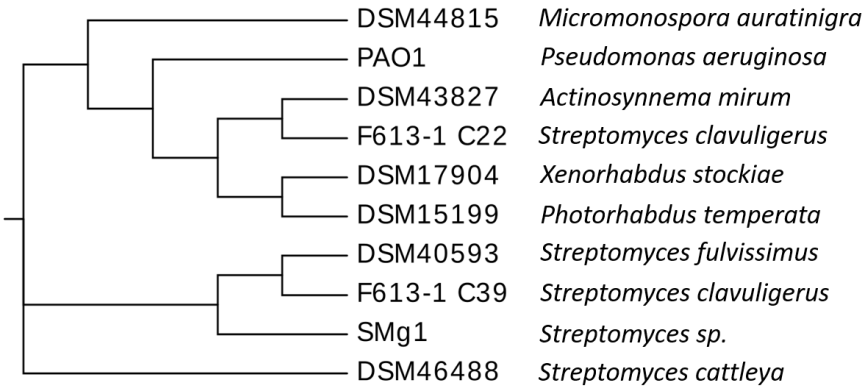


Figure 10. Phylogenetic tree of bacterial strains encoding PA-producing BGCs established by multiple alignment of NRPS nucleotide sequences.

3.1.3 Homology analysis of FAD-dependent MOXs

A similar sequence homology trend as observed for NRPS systems can be seen for the nucleotide sequences of MOX genes encoded within PA-producing BGCs. Figure 11 makes it clear that nucleotide sequence differences increase, the larger and more complex MOX systems get, for example regarding *Streptomyces sp.* Mg1 C5 compared to *S. clavuligerus* F613-1 C39.

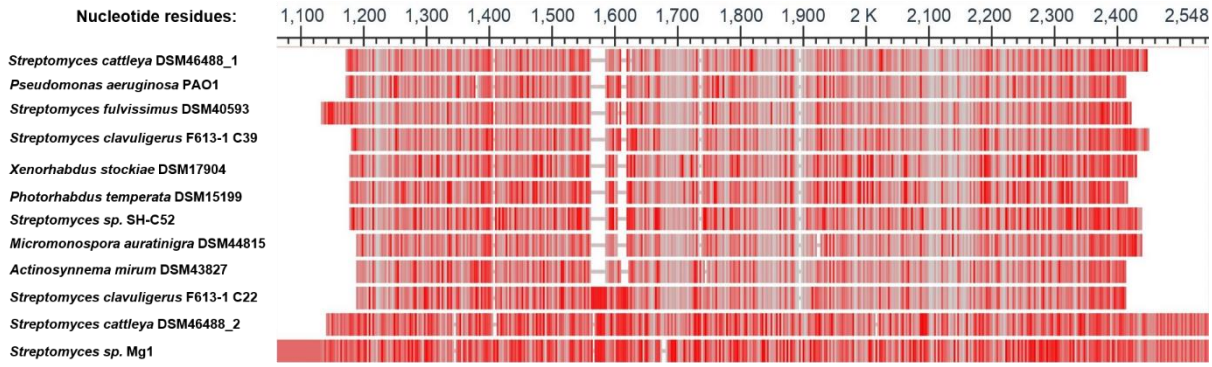


Figure 11. Representation of MOX system sequence homology throughout the selection of bacterial strains. Highly conserved regions are shown in red, nucleotide variations are represented in grey.

Again, homology within the actinobacterial and proteobacterial phyla is high (60 – 75 %) whereas sequence similarities less than 50 % are comparatively low between members of the different phyla (Figure 12).

% similarity	DSM44815	DSM40593	DSM17904	DSM15199	DSM46488_1	DSM46488_2	DSM43827	PAO1	SMg1	F613-1 C39	F613-1 C22	SH-C52
DSM44815	100	66.31	51.10	50.22	68.28	43.61	63.42	66.31	62.80	63.19	61.25	52.41
DSM40593	66.31	100	53.40	50.30	66.92	43.83	60.79	65.19	67.45	76.09	61.61	49.48
DSM17904	51.10	53.40	100	61.07	53.95	37.03	52.81	53.43	52.89	53.16	50.35	52.05
DSM15199	50.22	50.30	61.07	100	50.95	33.92	46.45	51.08	49.21	49.78	46.98	48.83
DSM46488_1	68.28	66.92	53.95	50.95	100	44.56	63.31	68.85	64.72	65.91	63.25	50.73
DSM46488_2	43.61	43.83	37.03	33.92	44.56	100	46.16	45.23	42.88	43.65	45.10	35.94
DSM43827	63.42	60.79	52.81	46.45	63.31	46.16	100	61.26	60.91	62.23	72.89	50.09
PAO1	66.31	65.19	53.43	51.08	68.85	45.23	61.26	100	65.27	63.64	58.83	52.56
SMg1	62.80	67.45	52.89	49.21	64.72	42.88	60.91	65.27	100	65.23	59.97	48.20
F613-1 C39	63.19	76.09	53.16	49.78	65.91	43.65	62.23	63.64	65.23	100	60.28	51.15
F613-1 C22	61.25	61.61	50.35	46.98	63.25	45.10	72.89	58.83	59.97	60.28	100	47.38
SH-C52	52.41	49.48	52.05	48.83	50.73	35.94	50.09	52.56	48.20	51.15	47.38	100

Figure 12. Comparison of MOX nucleotide sequences. The amount of similarity between two nucleotide sequences is given in %. Red coloring represents higher sequence homology, yellow represents less high and green low sequence homology.

The heat map underlines a low percentage of conserved residues in the second MOX gene of the *Streptomyces cattleya* DSM46488 BGC which leads to the assumption that it takes part in PA biosynthesis. However, it is probably not directly involved in the assembly of the PA core structure. Furthermore, the MOX nucleotide sequence from *Pseudomonas sp.* SH-C52 shows relatively low homology with bacterial strains of both proteo- and actinobacterial origin. A reason could be that this specific MOX termed BraC is known to be active in a related biosynthetic pathway, thereby catalyzing the synthesis of **40**[97] from a common precursor. Variations in sequence thus most likely coincide with the enzymes' altered mode of action and have an influence on natural product outcome.

The phylogenetic tree in Figure 13 emphasizes sequence similarities of both actino- and proteobacteria and shows increasing deviations and less relatedness the higher gene cluster complexity gets.

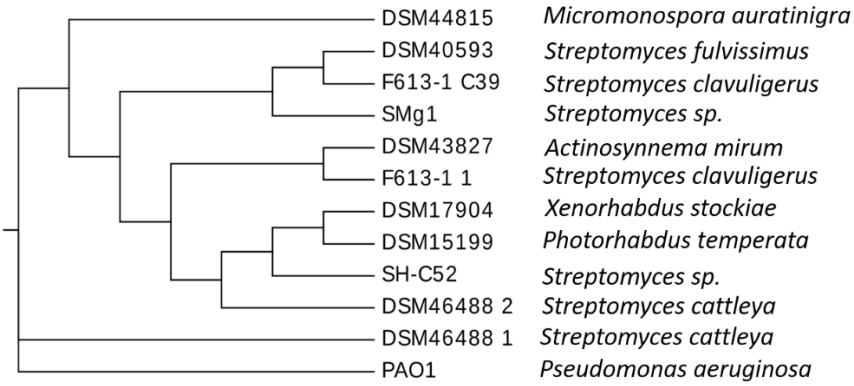


Figure 13. Phylogenetic tree of bacterial strains encoding PA-producing BGCs constructed based on MOX nucleotide sequences obtained by genome mining.

NRPS and MOX nucleotide sequence homologies of PA-producing bacteria are consistent upon comparison of NRPS and MOX phylogenetic trees, heat maps and alignments, thereby underlining the potential to find novel PA derivatives within these BGCs. Phylogenetic relationships and conserved gene homology support the hypothesis that similar molecules are produced and gene cluster complexity hints at finding novel PA derivatives.

3.2 Heterologous expression of vuPA-encoding BGCs

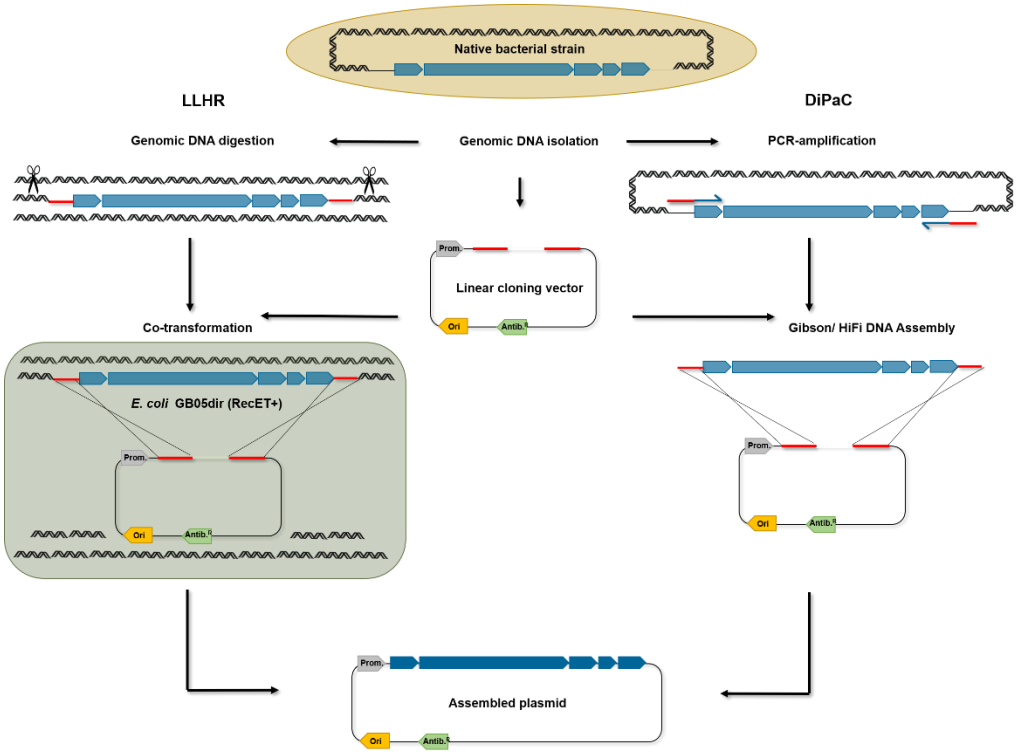
3.2.1 Cloning strategies to capture BGCs

The genome mining approach allowed screening of many bacterial genomes in search of putative PA-producing BGCs and the selection of a smaller number of candidate BGCs to work within the course of this thesis. Nevertheless, the challenge of accessing these genes of interest and furthermore finding methods to activate them under laboratory conditions remains, since many BGCs are silent under native conditions and do not produce the natural product. One strategy to tackle these difficulties is to clone the entire BGC with subsequent heterologous expression in a suitable host system.[105]

The conventional way of accessing secondary metabolite BGCs has been based on the construction of large insert genomic libraries where partially digested DNA is captured on fosmids/cosmids or in BAC/YAC vector systems. This method has allowed transfer of target genome sequences of more than 100 kbp size into a single vector without needing sequence data. However, library establishment and screening of colonies remains time consuming and work intensive. In addition, especially large BGCs of interest can be separated and distributed over several library clones, making identification of the respective clones and subsequent reassembly laborious. *In vivo* direct cloning strategies such as transformation-associated recombination (TAR cloning)[106] in yeast or linear-linear / linear-circular homologous recombination (LLHR and LCHR)[107] in *E. coli* as well as *in vitro* recombination systems like Gibson / HiFi DNA Assembly used in Direct Pathway Cloning (DiPaC)[108] have become efficient alternatives for capturing BGCs of varying size and complexity directly from genomic DNA (Scheme 12).

DiPaC enables the assembly of vector systems suited for subsequent BGC heterologous expression by combining long-amplicon PCR for DNA capturing with homology-based assembly to build functional expression constructs.[108] This *in vitro* recombination method is suited for small to midsize BGCs and is less time consuming than in *vivo* techniques due to a reduction of cloning steps. Furthermore, DiPaC allows a free choice of vector backbone and cluster refactoring is possible directly during initial cloning, for example, to remove terminator regions.[109] Homologous recombination was applied as an efficient alternative to PCR-based DiPaC, for target BGCs larger than 20 kbp or BGCs extremely difficult to amplify, which could otherwise only be cloned in many small fragments in laborious working steps or not at all. Using this cloning strategy, nucleotide sequences of interest can be captured in a single piece from genomic DNA and then directly transferred into a suitable vector system for expression,

making the construction of large fosmid or BAC libraries coupled with extensive screening unnecessary. For LCHR, prophage derived Redαβ proteins mediate homologous recombination at the replication fork when replication is active between a linear and a circular DNA fragment. In contrast, homologous recombination of two linear DNA fragments (LLHR) is promoted by RecET proteins and follows a replication-independent mechanism.[107, 110]



Scheme 12. Overview of *in vivo* and *in vitro* recombination strategies. For LLHR, isolated genomic DNA needs to be digested before co-transforming with a linearized capturing vector into cells capable of homologous recombination. The capturing vector itself was assembled prior to co-transformation to harbor regions of homology. In contrast, target BGC DNA can be amplified directly from a genomic template and the cloning vector must only be linearized by restriction digest when proceeding with the DiPaC strategy.

In principle, *in vitro* DiPaC and *in vivo* LLHR and LCHR techniques were used to access the target BGCs addressed in this thesis. Cluster size, GC-content, gene complexity and origin were some of the criteria which influenced the choice of cloning strategy. Furthermore, the presence of transcription terminators within BGCs of interest, especially in noncoding regions between genes, was examined to ensure transcription of the entire nucleotide sequence. The ARNold transcription terminator prediction web tool was applied for screening[111]. Refactoring of genes would have been necessary in the case of intermitting terminator sequences. However, none were found in any of the selected bacterial BGCs. Biomolecular techniques,

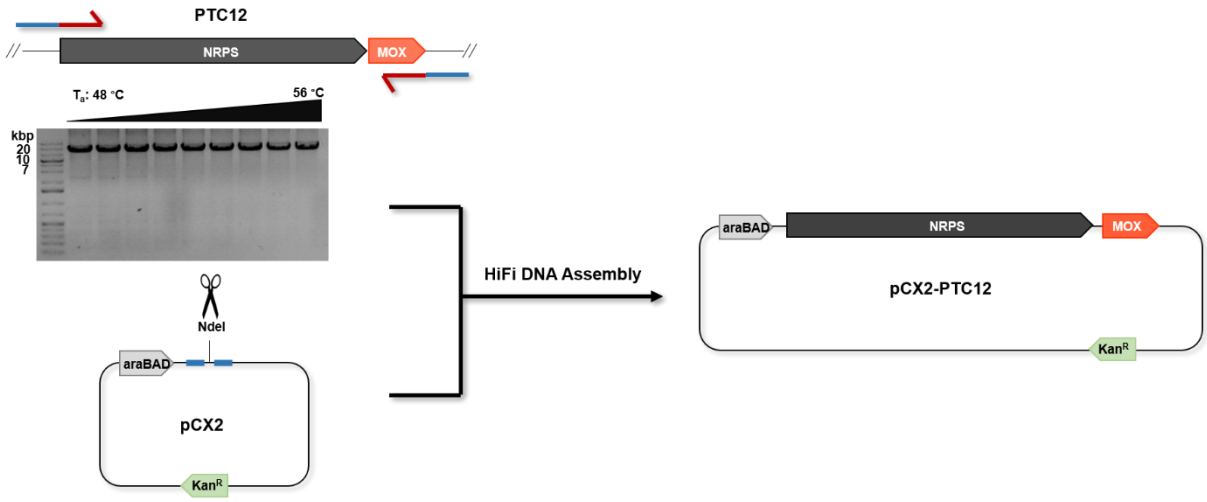
including screening PCR, analytical restriction digest and DNA sequencing, were applied to confirm the identity of the captured genomic DNA to validate successful plasmid assembly.

3.2.1.1 DiPaC - Single fragment assembly

BGCs of smaller size (less than 9 kbp) such as cluster # 12 from *Photorhabdus temperata* DMS15199 or cluster # 39 of *Streptomyces clavuligerus* F613-1 carrying only the two core genes encoding the PA-producing NRPS and MOX systems, were cloned using DiPaC. PCR amplification of the region of interest from isolated genomic DNA was performed in one piece using specifically designed primers flanking the ends of the cluster. In addition to the gene-specific nucleotide sequence, primers were designed to carry 20 – 25 bp overhangs complementary to the BGC insertion site of the linearized cloning vector. The vector of choice was linearized prior to plasmid assembly either by enzymatic digestion or PCR-amplification.

***Photorhabdus temperata* DSM15199 cluster # 12 (PTC12)**

Cluster # 12 of the proteobacterium *Photorhabdus temperata* DSM15199 was successfully amplified and inserted into the NdeI-linearized arabinose-inducible pCX2 vector system for heterologous expression in *E. coli* (Scheme 13*)*. *E. coli* was chosen as a heterologous host for this specific gene cluster due to the relatively low GC-content (44 %) and the similar codon usage. Furthermore, expression using the pCX2 vector served as a proof of principle for PA-production in heterologous host systems, since cluster composition is identical to the previously successfully cloned and expressed *Xenorhabdus stockiae* DSM17904 cluster # 19 in pCX2.[79]



Scheme 13. Plasmid assembly of pCX2-PTC12 using DiPaC. BGC cluster # 12 insert DNA carries homology arms at 5' and 3' ends compatible to the region flanking the insertion site of the pCX2 vector backbone (marked in blue). The agarose gel displays high specificity of designed primers as no significant side products were amplified during PCR.

Alongside pCX2, a second vector system harboring PTC12 was assembled for heterologous expression in *E. coli*. In this case, gene transcription was regulated by a tetracycline-inducible system, allowing not only the tight control of expression but also enabling induction to be reversed. The repression-based configuration was applied for expression purposes where the Tet operator (TetO) lies between the constitutive P_{Tet} promoter and the gene of interest. In the absence of tetracycline, TetO is associated with the Tet repressor (TetR) which prohibits downstream gene expression. Upon addition of tetracycline, the antibiotic binds to TetR, causing disruption of the TetOR complex which leads to the induction of TetO-dependent gene expression (Figure 14). [112] A modified version of the pET28bpTetO vector was used for cloning, where the GFP gene had previously been inserted to monitor translation efficiency during heterologous expression. Both NRPS and MOX genes were inserted at the multiple cloning site as a single DNA fragment to yield the pET28bpTetOGFPv2-PTC12 expression plasmid.

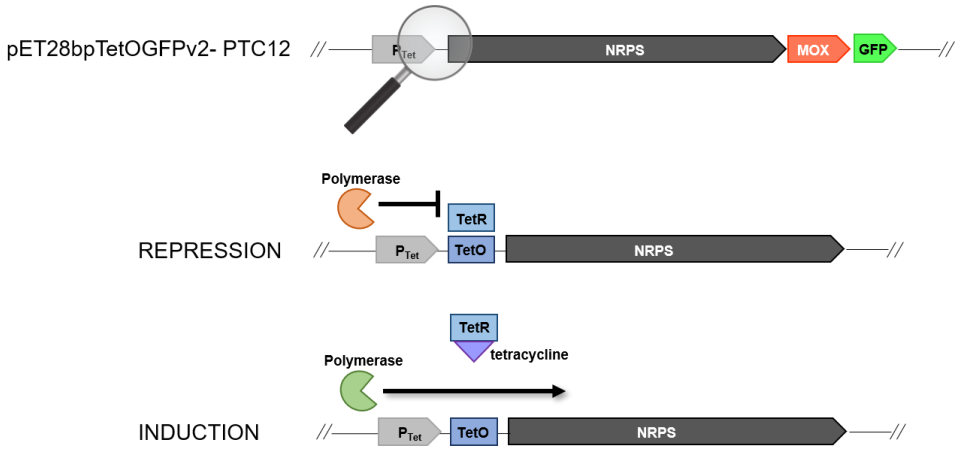


Figure 14. Regulation of cluster # 12 translation using a tetracycline-inducible expression system. In the absence of tetracycline, gene expression is repressed due to complex formation of TetR with TetO. Downstream gene expression is initiated when tetracycline binds to TetR, causing dissociation from TetO and allowing polymerase to start transcription.

Streptomyces clavuligerus F613-1 cluster # 39 (SclavuC39)

Due to a high GC-content of 75 % as well as its actinobacterial origin, *Streptomyces clavuligerus* F613-1 cluster # 39 was inserted into a vector system suitable for heterologous expression in *Streptomyces* host strains. The pSET152ermE*rev vector system possesses a constitutive promoter, allowing continuous transcription of the associated genes. Captured DNA was assembled with the StuI-linearized pSET152ermE*rev vector via SLIC assembly[113], a more cost-effective alternative to HiFi DNA Assembly following a similar reaction mechanism. In contrast to PTC12, which was captured and multiplied using Q5® High-Fidelity DNA Polymerase (NEB), PCR amplification from *Streptomyces clavuligerus* F613-1 genomic DNA was only possible with KOD Xtreme™ Hot Start DNA Polymerase (MerckMillipore), a polymerase specialized to amplify from highly GC-rich DNA templates. This specific heterologous expression approach served to prove successful BGC activation in a host system other than *E. coli*.

Concerning plasmid design, several regulatory elements seem to be of importance for efficient BGC expression in *Streptomyces*. As previously reported, the distance between the Shine-Dalgarno (SD) nucleotide sequence and the start codon of the first gene is crucial for translation initiation and efficiency is predicted to be highest with a spacer region of six to nine nucleotides.[114] Thus, alongside the commonly used eight base pair (bp) spacer a plasmid exhibiting only one nucleotide between SD sequence and start codon was assembled for comparative evaluation (Figure 15).

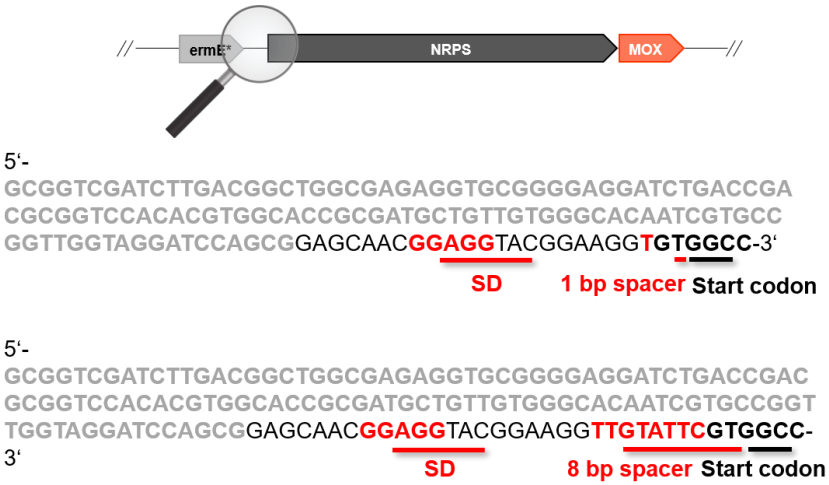


Figure 15. Variation of nucleotide distance between Shine-Dalgarno (SD) sequence and start codon of the first gene being translated to increase chances of silent BGC activation under laboratory conditions.

To further increase BGC translation efficiency, NRPS and MOX genes of cluster # 39 were separated and put under the control of two different constitutive promoters. The establishment of a double promoter system was performed in multiple steps (Figure 16). Identical to the constructs where translation of the entire BGC was regulated by a single promoter, the NRPS gene was inserted at the StuI restriction site and kept under the control of the ermE* promoter. After gene insertion, the StuI restriction site was restored by nucleotide insertion. Incorporation of the second promoter downstream of the NRPS gene was achieved by ligation. For ligation, the assembled plasmid was linearized by enzymatic digest with StuI and XbaI. Simultaneously, a set of three different promoters, namely kasOP, gapdh(EL) and SF-14, were PCR-amplified using Q5® High-Fidelity DNA Polymerase. Primers were designed to carry a StuI restriction site at the 5’ and an XbaI restriction site at the 3’ end so that consecutive digest with both enzymes generated sticky ends complementary to those generated on the vector backbone. Double promoter plasmid assembly was performed using T4 DNA Ligase (Jena Bioscience). In a final step, the cluster # 39 MOX gene was introduced into the three XbaI-linearized double promoter system variants (ermE*-kasOP, ermE*-gapdh(EL) or ermE*-SF-14) via HiFi DNA Assembly.

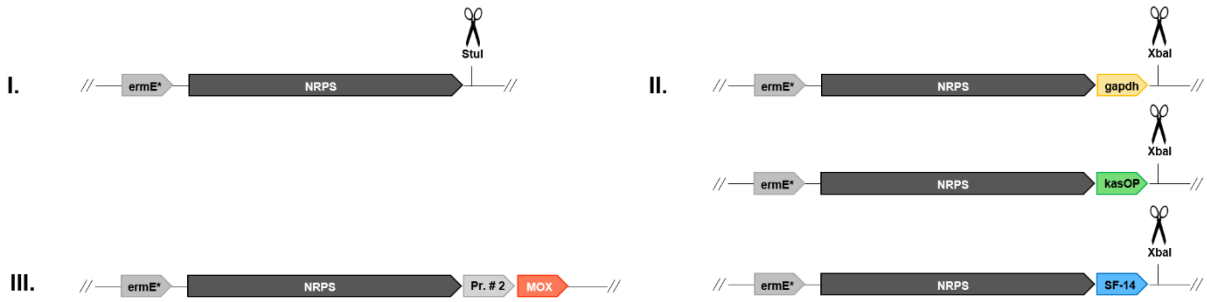


Figure 16. Multistep assembly of double promoter vector systems for BGC expression in *Streptomyces*. Cluster # 39 NRPS gene is inserted behind ermE* promoter (I.). StuI restriction site is incorporated for subsequent linearization and insertion of second promoter via ligation (II.) Finally, cluster # 39 MOX gene is put under regulation of a second promoter by insertion at the XbaI restriction site (III.).

Whereas steps I. and II. could be executed without further difficulties, incorporation of the MOX gene remained challenging. It was originally attempted to perform a second round of ligation for the final insertion step by simultaneous digestion of vector and insert DNA with XbaI. However, recircularization and inefficient dephosphorylation resulted in massive growth of transformed *E. coli* colonies harboring only the plasmid construct lacking the MOX gene. As an alternative, 20 bp overhangs complementary to the respective pSET152ermE*rev-SclavuC39 NRPS-gapdh/kasOP/SF-14 backbones were attached to 5' and 3' ends of SclavuC39 MOX DNA and plasmid construction was performed using HiFi DNA Assembly. To finally achieve insertion of the MOX gene, multiple assembly attempts were necessary, where different insert-to-vector ratios (2:1 to 5:1), growth periods after transformation, enzymatic digestion and dephosphorylation times and *E. coli* host cell types were tested. Despite all these optimization attempts, successful incorporation of SclavuC39 MOX at the designated insertion site remained to be achieved.

Micromonospora auratinigra DSM44815 cluster # 21 (MiC21)

When work on this thesis commenced, the genome sequence of *Micromonospora sp.* M71-S20 (later designated as *Micromonospora auratinigra* DSM44815), an actinobacterial strain received from the ZIEL Institute at TUM was not yet deciphered. Genome mining and bioinformatic analyses performed by Dr. Paul D'Agostino revealed the presence of a BGC encoding putative genes involved in PA biosynthesis. Besides the identified core NRPS and MOX genes BGC # 21 was of particular interest, since four of the seven encoded genes could not be associated with known protein functions at the time.

The 13613 bp and 70 % GC-content DNA fragment of interest was successfully amplified in one piece using KOD Xtreme™ Hot Start DNA Polymerase. Analogous to cloning of SclavuC39, MiC21 was inserted into the *Streptomyces* expression vector system pSET152ermE*rev with one and eight bp spacer regions in between SD-sequence and start codon. Incorporation of the DNA insert via SLIC assembly proved to be challenging due to recircularization of linearized and dephosphorylated vector DNA. Optimization of insert-to-vector ratios and extensive digestion and dephosphorylation times led to construction of the desired heterologous expression plasmid constructs (Figure 17).

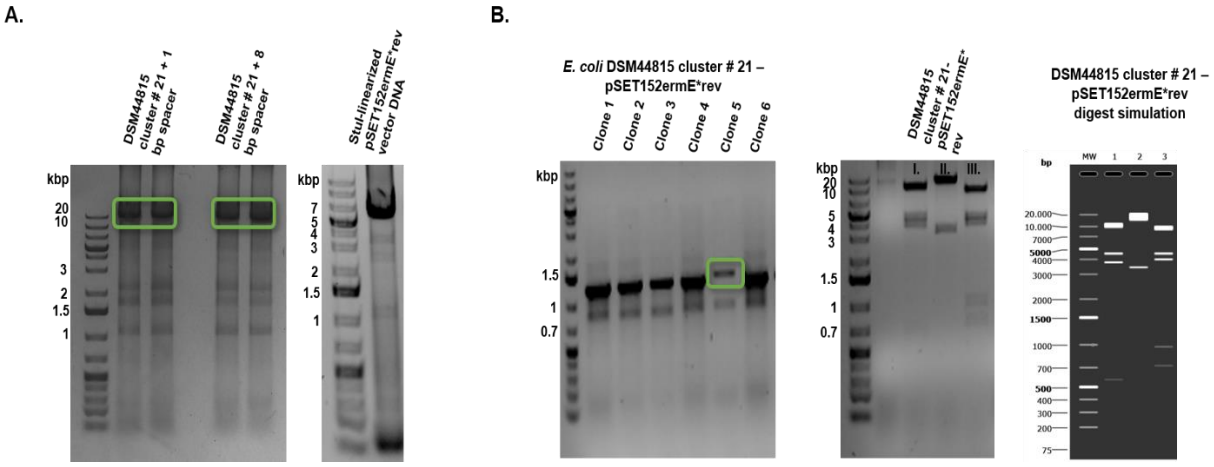


Figure 17. PCR-amplification of insert DNA carrying homology arms varying in spacer size. **A.** BGC # 21 was amplified with either a 1 bp or 8 bp spacer region in the 5'-primer nucleotide overhang. Vector DNA was enzymatically linearized and purified by gel extraction. **B.** Screening of transformed colonies using cluster-specific annealing primers yielded clones harboring the plasmid of interest. Analytical restriction digest confirmed correct assembly. Lane 1: undigested plasmid DNA; I. digest with XmnI (573, 3799, 4509, 10570 bp), II: digest with BmtI (3411, 16040 bp) and III. digest with HindIII (723, 969, 3912, 4491, 9356 bp).

3.2.1.2 DiPaC - Multiple fragment assembly

With increasing size and cluster complexity, capturing of target DNA in a single fragment became inefficient since neither Q5® High-Fidelity DNA Polymerase nor KOD Xtreme™ Hot Start DNA Polymerase could amplify the desired sequence from genomic DNA. High GC-content (> 70 %) and repetitive sequences especially at BGC boundaries support hairpin formation, making primer annealing difficult. To overcome these problems, BGCs of interest affected by these restrictions were divided into multiple fragments and PCR-amplified independently before sequentially inserting them into the vector expression system. After each

round of fragment insertion, the unique restriction site was either restored or a novel single cutting enzyme restriction site was introduced downstream of the respective insert. All assembly reactions were prepared using SLIC or HiFi DNA methods.

Single fragment capturing of 15.5 kbp gene cluster # 5 from *Streptomyces sp*. Mg1 and 13.5 kbp cluster # 29 originating in *Pseudomonas aeruginosa* PAO1 had previously been unsuccessful, therefore selected BGCs were subdivided into two or three fragments and inserted stepwise into the growing plasmid. Heterologous expression was to be tested in both *E. coli* and *Streptomyces* hosts, therefore clusters # 5 and # 29 were introduced into vector systems suitable for expression in either of the hosts.

Streptomyces sp. **Mg1 cluster # 5 (SMg1C5)**

Streptomyces sp. Mg1 BGC # 5 was split into two sections, the first fragment consisting of three genes encoding tailoring enzymes located upstream of the core NRPS gene, which itself was amplified together with the MOX gene and a third gene of yet unknown function as a second DNA fragment. Separation of tailoring genes from the core genes involved in PA-production enabled cloning and expression of truncated BGC versions which opened opportunities for studying the function of tailoring genes considering natural product outcome. For heterologous expression in *Streptomyces*, three pSET152ermE*rev-based plasmids were constructed, either harboring only one of the two fragments or the entire cluster. The pET28bpTetOGFPv2 vector system was chosen for *E. coli* expression due to tight induction regulation and transcription monitoring via GFP (Figure 18).

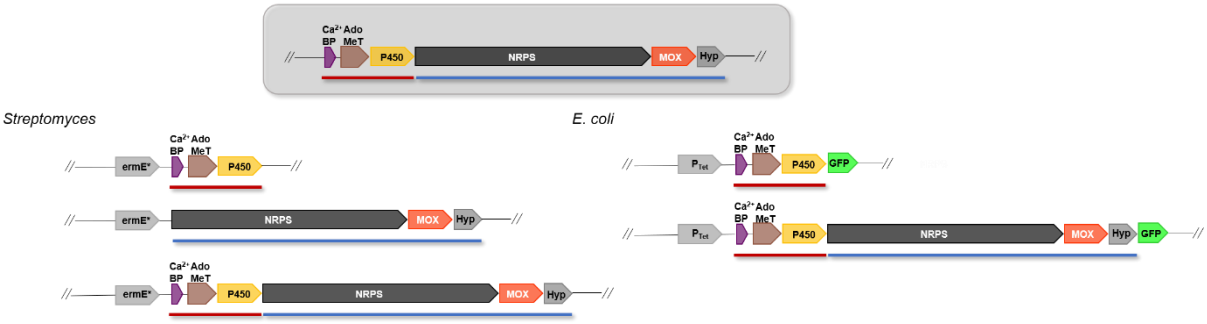


Figure 18. Illustration of SMg1C5 multi-fragment assembly. Gene functionality was tested in *Streptomyces* by expression of truncated BGC versions. Successful transcription of all genes was monitored by GFP expression at the 3' end of the cluster.

Pure DNA samples of both cluster pieces could be obtained after PCR using purified genomic DNA as a template. Furthermore, linearized and dephosphorylated pET28bpTetOGFPv2 vector DNA was purified by gel extraction (Figure 19A.) Analytical restriction digests performed

after insertion of cluster # 5 part 1 and in a later step of part 2 were unique to the desired plasmids and confirmed successful assembly (Figure 19B). Multiple digests were set up for each of the constructs to eliminate putative false-positive clones which show similar restriction patterns but with slightly varying sizes. Due to the logarithmic running behavior of the marker substance, it can be hard to discriminate between two sizes, especially in the high kbp regions above 5 kbp. To further simplify evaluation and identification of clones harboring the desired plasmid of interest, enzymes were chosen which cut specifically within the inserted piece of DNA but not on the vector backbone.

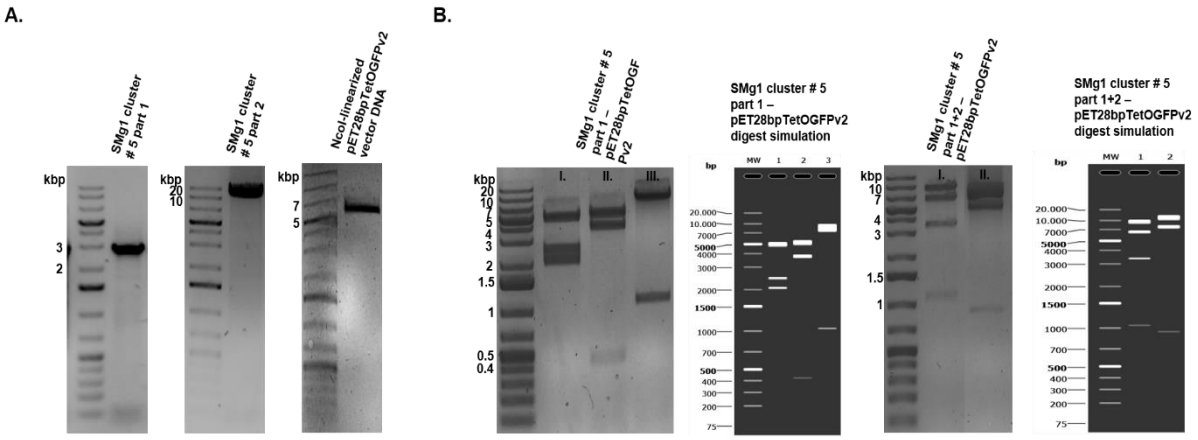


Figure 19. Separation of total BGC into pieces for DNA accession. **A**. BGC # 5 was captured as two separate pieces from genomic DNA. Linearized vector DNA was purified and prepared for HiFi DNA Assembly by gel extraction. **B.** Multiple analytical restriction digest of either construct harboring only cluster part 1 or construct with both cluster parts 1 and 2 was executed for validation purposes. Part 1 was digested with I. XmnI (2033, 2402, 4985 bp), II. NdeI (420, 3769, 5231 bp) and III. NcoI (1055, 8365 bp), whole BGC introduction was confirmed by digestion with I. NcoI (1055, 3370, 6501, 9172 bp) and II. BglI (947, 7832, 11319 bp).

Pseudomonas aeruginosa PAO1 cluster # 29 (PAO1C29)

PAO1 cluster # 29 was cloned into the pSET152ermE*rev vector system the same way as described above for *Streptomyces sp.* Mg1 BGC # 5, with the only difference being that the DNA of interest was captured in three pieces (Figure 20). In a short version, solely the genes encoding proteins for PA core assembly were inserted (8222 bp), whereas the complete cluster sequence was introduced consecutively, with the final cloning construct consisting of four pieces of DNA (21349 bp total size). The second fragment comprised five genes (4912 bp) whereas the last fragment to be inserted consisted of the last three genes of the cluster and had a size of 2368 bp. The shortest fragment was amplified using Q5® High-Fidelity DNA

Polymerase from genomic DNA. In contrast, insert DNA pieces one and two could only be captured by PCR with KOD Xtreme™ Hot Start DNA Polymerase.

In 2019, during cloning and first heterologous expression experiments of PAO1C29, two publications addressing this BGC of interest were released. Hong et al.[115] proposed the establishment of the PA core structure to proceed as we expected it, with the difference that L-azetidine 2-carboxylic acid (L-AZC) is incorporated instead of L-proline and subsequent ring modifications take place to yield azetidomonamide A (**48**). The ninth gene of the cluster was identified as a SAM-dependent enzyme predicted to catalyze biosynthesis of L-AZC from SAM. These data suggest that expression of a short version of the cluster as we designed most likely does not result in natural product production since the substrate L-AZC is unavailable for incorporation. Nevertheless, expression experiments were conducted to examine if the NRPS system can utilize other precursors in the absence of L-AZC to produce alternative molecules. Furthermore, Patteson et al.[116] could prove that core NRPS and MOX enzymes of cluster # 29 produce azabicyclenes (**49**) as pathway intermediates based on L-AZC. This information was used to design co-transformation constructs, with one of them carrying the minimal BGC (AzeBC = NRPS and MOX), the other expressing the SAM-dependent enzyme AzeJ (Figure 19).

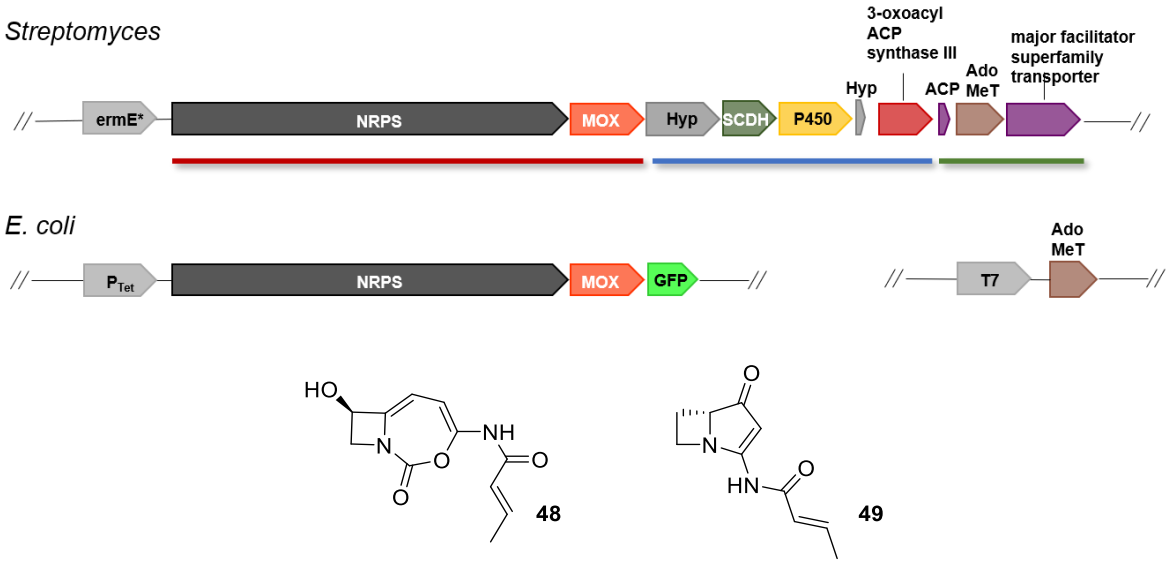


Figure 20. PAO1C29 DNA capturing for heterologous expression in *Streptomyces* and *E. coli*. The BGC was divided into three pieces and sequentially assembled for expression in *Streptomyces*. Co-transformation of minimal gene set in combination with SAM-dependent enzyme in *E. coli*. Full BGC expression yields **48**, whereas **49** is synthesized upon core enzyme activity in combination with AzeJ only.

Whereas amplification of fragments one and three was specific, major side product bands were observed for fragment two, making purification by gel extraction of the correct band necessary (Figure 21A). The designed fragment two forward primer has multiple annealing sites along the genomic DNA, including binding to regions immediately upstream or within the target sequence itself, thus side product formation can be explained. In a similar case, multiple DNA bands were detected in PCR samples containing the AzeBC insert (Figure 21B). However, the designed primer pair was predicted to be specific, letting us assume that the polymerase might have dissociated from template DNA during processing due to extensive secondary structure formation, e.g., hairpins, resulting in unwanted shortened nucleotide sequences. In addition, secondary structures occurring directly in or adjacent to the primer binding sites could have reduced overall product outcome.[117] Assembled plasmids were successfully co-transformed into *E. coli* BAP1 host cells (Figure 21C).

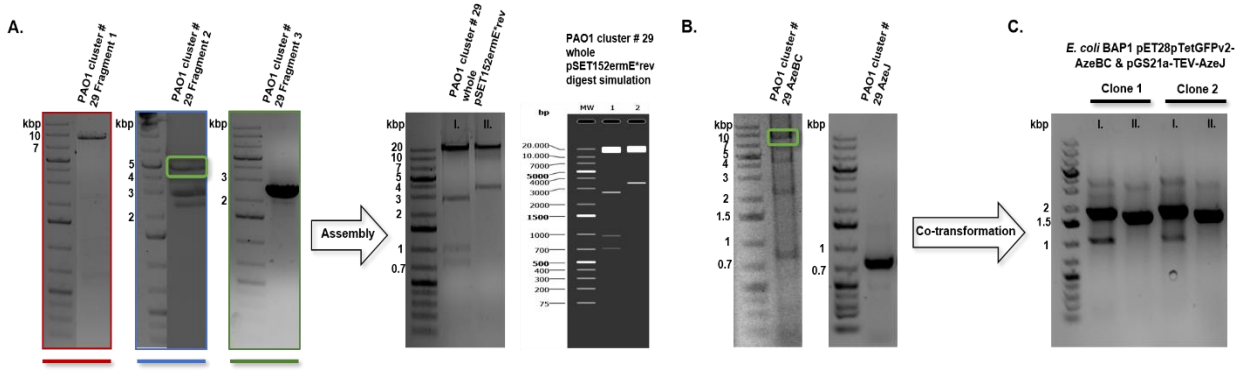


Figure 21. PAO1 BGC capturing in multiple fragments from genomic DNA. **A.** Cluster # 29 was split into three pieces, which were PCR-amplified and inserted one fragment at a time into the pSET152ermE*rev expression system. Validation of the fully assembled BGC was performed by analytical restriction digest with I. HindIII (723, 969, 2678, 16975 bp) and II. BmtI/ScaI (3488, 17857 bp). **B.** Insert DNA production of minimal gene set AzeBC and SAM-dependent enzyme AzeJ. **C.** Genes were inserted into vectors pET28pTetOGFPv2 and pGS21a-TEV and co-transformed into *E. coli* BAP1 cells. Presence of both plasmids was confirmed by screening PCR with I. AzeBC-specific and II. AzeJ-specific screening primer pairs.

3.2.1.3 DiPaC - "Mix & Match"- cloning to study MOX mechanism of action

Being one of the essential enzymes involved in PA core structure assembly, the MOX gene is present in all BGCs related to PA-biosynthesis. Baeyer-Villiger Monooxygenases (BVMOs) catalyze the oxidation of linear or cyclic ketones to form esters or lactones by incorporation of an oxygen atom and thus can take part in the construction of complex secondary metabolites.

For full catalytic function, the majority of BVMOs acting during PA-biosynthesis depend on flavin- (FAD or FMN) and nicotinamide-(NAD(P)H) containing co-factors which are associated with the enzymes' active site.[118] As shown in chapter 3.1.2, Figure 12, flavin-dependent BVMOs from *Xenorhabdus stockiae* DSM17904 and *Pseudomonas sp*. SH-C52 share 52 % nucleotide sequence identity although they originally catalyze Baeyer-Villiger oxidation reactions leading to different natural products (**37** or **40**). To gain a more detailed insight into the mechanism of action of this type of enzyme, several MOX genes taken from some of the selected BGCs were introduced into an *E. coli* expression system already harboring the functional NRPS gene *pxaA* from *Xenorhabdus stockiae* DSM17904 and the resulting natural product outcome was studied.

The plasmid pXst4, consisting of the pCX2 vector backbone together with the NRPS insert DNA *pxaA,* was received from Helge Bode (University of Frankfurt). A selection of MOX genes, including the ones from *Xenorhabdus stockiae* DSM17904, *Pseudomonas sp*. SH-C52, *Micromonospora auratinigra* DSM44815 and *Pseudomonas aeruginosa* PAO1, were PCR-amplified as single DNA fragments and inserted downstream of the NRPS gene via a unique BglII restriction site using DNA HiFi Assembly (Figure 22). As can be seen in the agarose gel, all MOX inserts have sizes of about 1200 bp and could be obtained in rather pure and high amounts. For DSM17904 MOX, DNA was extracted from the gel to eliminate side products (bands at 400 – 500 bp) and to thus ensure insertion of the correct piece of DNA during plasmid assembly.

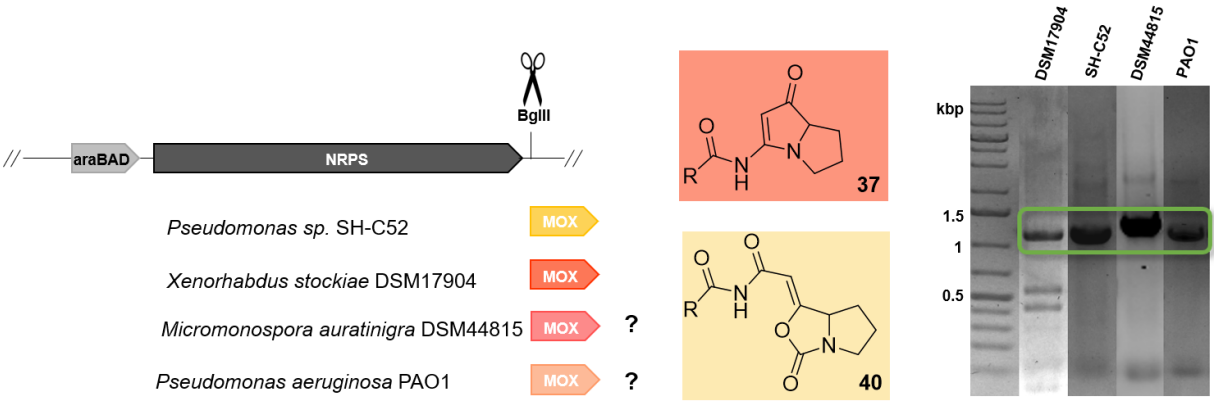


Figure 22. BVMO cloning for *in vivo* mechanism of action studies. MOX genes of selected bacterial PA-producing BGCs were PCR-amplified (agarose gel) and inserted into a functional PA-intermediate-producing vector system to examine MOX catalytic activity upon expression.

3.2.1.4 DiPaC - Establishment of double promoter systems

Investigations on the production of PA-derivatives were further elaborated by extending the described “Mix & Match”-cloning strategy (3.2.1.3) in terms of inserting either single or multiple genes encoding tailoring enzymes downstream of the MOX gene. For testing matters, the plasmid containing both NRPS and MOX core genes was StuI-linearized and truncated versions of the selected BGC were introduced.

As the MOX gene originating from *Micromonospora auratinigra* DSM44815 had previously already been incorporated into the pXst4 vector system, the remaining genes of cluster # 21 were used as a test system for PA-derivative production. In total, three different versions of the target BGC were inserted at the StuI restriction site, the first fragment consisting of only the short chain dehydrogenase (SCDH) encoding gene. The second DNA insert represents a short cluster version including two hypothetical genes and the SCDH, whereas all five genes located downstream of the MOX were present in the third long insert version (Figure 23). KOD Xtreme™ Hot Start DNA Polymerase was used to amplify the short and long version of cluster # 21, for SCDH Q5® High-Fidelity DNA Polymerase was sufficient. All variations of the chosen cluster were amplified and successfully inserted into the heterologous expression construct.

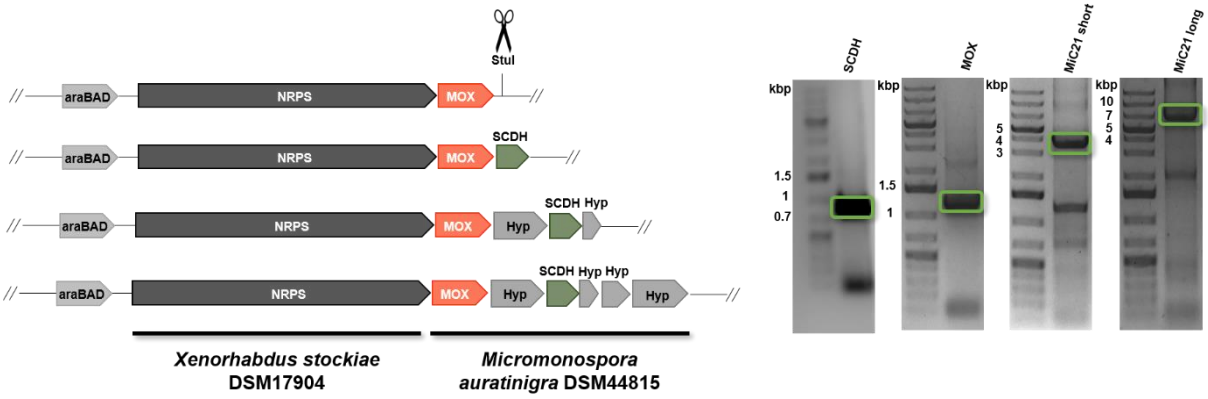


Figure 23. “Mix & Match” -cloning of MiC21 in combination with the XSC19 NRPS gene. Variations of the BGC (SCDH: 879 bp; MOX: 1200 bp; short: 3808 bp, long: 6516 bp) were inserted to examine PA-derivative production. Gel extraction of amplified insert DNA was necessary to remove side products resulting from unspecific primer annealing to similar or repetitive sequences occurring in template DNA.

Since heterologous expression and activation of silent BGCs has proven to be a challenge, it remains unclear whether a single promoter is strong enough to control transcription of a medium to large sized (> 10 kbp) nucleotide sequence and thus promote recombinant production of sufficient protein amounts necessary for natural product biosynthesis. With this uncertainty in mind, the idea of building a double promoter system to initiate and ensure

transcription of an entire BGC became reasonable. Transcription of genes encoded on the previously assembled pXst4-MiC21 MOX plasmid construct is regulated via an arabinose-inducible araBAD promoter. Introduction of a second araBAD promoter downstream of the MOX gene resulted in constructing a simultaneously operating double promoter system, whereas insertion of an IPTG-inducible T7 promoter yielded a consecutively operating promoter system, allowing gene transcription to be induced sequentially[119]. In principle, induction of genes involved in different parts of the natural product biosynthetic pathway taking place at different time points can be a powerful tool. However, it is important to consider potential inhibitory effects one inducer can have on the activity of the other promoter (Figure 23).[120]

The order in which the second promoter sequence and the genes it controls were cloned, differed depending on the type of promoter inserted. For the consecutively operating double promoter system, the T7 sequence was introduced first and the 3' primer overhang of the PCR product was designed to install a BglII restriction site for subsequent BGC insertion. In contrast, the simultaneously operating double promoter system was built by inserting the BGC tailoring genes in a first step, followed by assembly of the araBAD promoter in a last step. Here, the BglII restriction site was encoded in the 5' primer overhang of the PCR-amplified BGC fragment. Reversed assembly was performed for construction of the double araBAD system to omit difficulties concerning validation of cloning steps, especially by screening PCR. Additionally, insertion of the BGC insert DNA could be ineffective or even fail because 5' primer overhangs exhibit complementary nucleotide sequences to a region with multiple annealing sites along the plasmid.

Simultaneous induction of gene transcription was tested for both MiC21 and PAO1C29 clusters, with the pXst4-MiC21 MOX plasmid harboring all modifying genes of the respective BGC located downstream of the MOX core PA-producing gene (Figure 24).

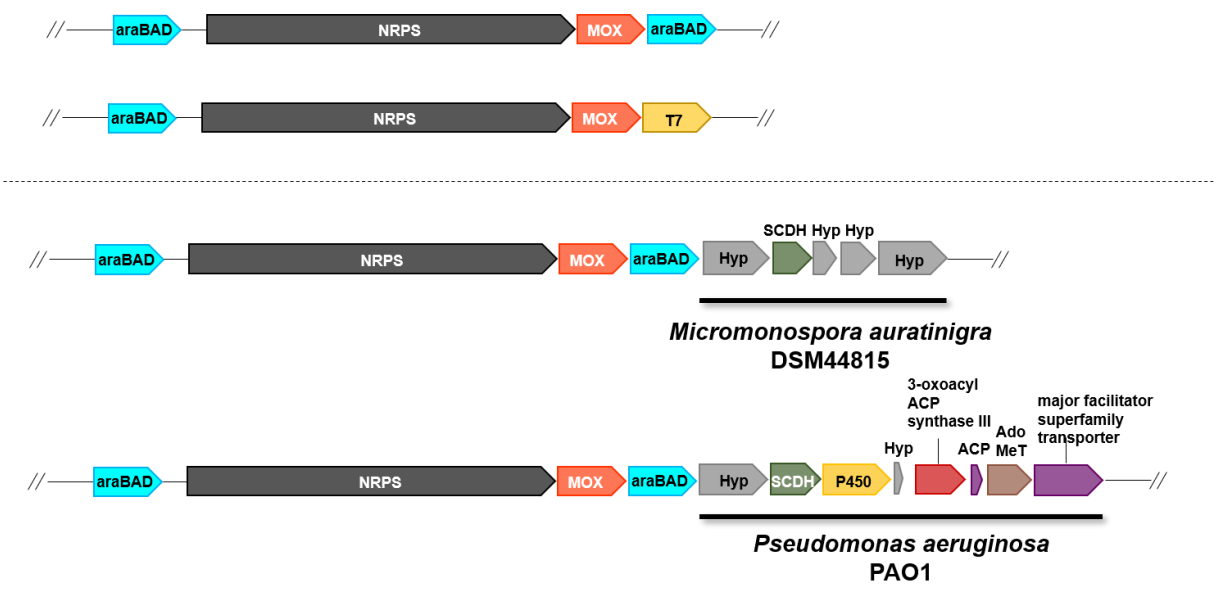


Figure 24. Construction of simultaneously and consecutively operating double promoter expression systems. Core genes involved in PA backbone assembly are controlled by a first L-arabinose-inducible promoter whereas a second promoter introduced downstream of the core MOX regulates transcription of additional modifying genes. Tailoring genes of both MiC21 and PAO1C29 were inserted for testing of the established system.

Construction of double promoter expression systems was a multi-step process, starting with PCR-amplification of both promoter sequences T7 and araBAD as well as the back parts of PA-encoding BGCs from DSM44815 and PAO1 (Figure 25A) followed by their gradual insertion of promoter and tailoring gene sequences. Analytical restriction digest (Figure 25B and C) and plasmid sequencing was performed after each step for validation of correct assembly. Full assembly of the double L-arabinose-inducible promoter system could be achieved, whereas insertion of the tailoring genes via DNA HiFi Assembly downstream of the T7 promoter failed. Neither repetition of the entire cloning process nor alteration of restriction site, primer nucleotide overhang design and choice of transformation host cells could help overcome these difficulties. This led to the assumption that the insertion site for tailoring genes exhibits a different nucleotide sequence than predicted during *in silico* cloning and therefore plasmid linearization and insertion of additional genes was prohibited. However, plasmid sequencing of the region of interest excluded this hypothesis.

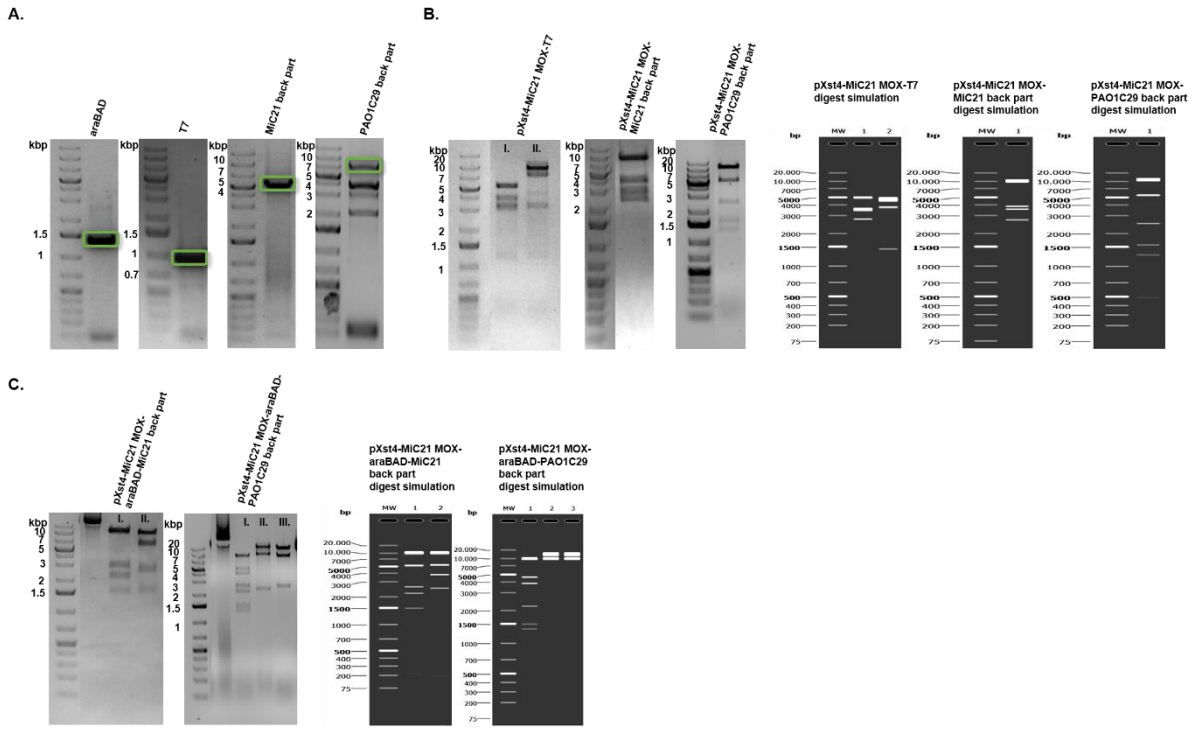


Figure 25. Multi-step assembly of *E. coli* double promoter expression systems. **A.** Amplified insert DNA encoding either promoter sequences (araBAD: 1334 bp; T7: 851 bp) or BGC parts downstream of MOX core gene (MiC21 back part: 5372 bp; PAO1C29 back part: 7292 bp). **B.** In a first round of plasmid assembly, the T7 promoter sequence (for consecutive double promoter system) or the genes involved in PA-modification (for simultaneous double promoter system) were successfully inserted. pXst4-MiC21MOX-T7 digested with I. NdeI/HindIII/ApaLI (2804, 3544, 3561, 4964 bp) and II. XbaI/ApaLI (1080, 2965, 3709, 7119 bp). pXst4-MiC21 MOX-MiC21 back part digested with BmtI/HindIII (2275, 2737, 3886, 11172 bp), pXst4-MiC21 MOX-PAO1C29 back part with KpnI/HindIII (489, 1271, 2487, 2895, 5395, 10815 bp). **C.** Introduction of the araBAD promoter sequence was achieved for both test constructs. Undigested plasmid DNA was applied to the first lane of the agarose gel, plasmids were test digested with multiple enzymes. pXst4-MiC21 MOX-araBAD-MiC21 back part digested with I. AflII/HindIII/XbaI (298, 1503, 2206, 2608, 5120, 9771 bp), II. NdeI/SnaBI/NotI (194, 2508, 3759, 5384, 9571 bp), pXst4-MiC21 MOX-araBAD-PAO1C29 back part with I. HindIII/NdeI/XbaI (1338, 1503, 2206, 3892, 4642, 9771 bp), II. BmtI (9771, 13581 bp), III. NdeI (9771, 13581 bp).

3.2.1.5 LLHR/LCHR for full-length BGC accession

Both LLHR and LCHR were used for BGC capturing of *Streptomyces fulvissimus* DSM40593 cluster # 2 (SFC2), *Streptomyces cattleya* DSM46488 cluster #37 (SCC27) and *Actinosynnema mirum* DSM43827 cluster # 53 (AMC53). Target DNA accession and insertion into an initial vector system was mediated by LLHR, vector backbone exchange was conducted by LCHR.

Recombination was performed following the general protocol for all three mentioned BGCs. Initially, the so-called capturing vector was established by three fragment HIFi DNA Assembly, consisting of the vector backbone of choice as well as two 500 bp long capturing site inserts. These capturing sites share homologous nucleotide sequences to the 5' and 3' end regions of the target BGC and are the place of recombination with purified genomic DNA. It is therefore of importance that the sites directly flank the cluster of interest, ideally starting with the start codon of the first gene and ending as close to the end of the target region as possible. The left and right 500 bp capturing sites were PCR-amplified using Q5® High-Fidelity DNA Polymerase, thereby carrying 20 – 25 bp overhangs complementary to the pSET152ermE*rev vector backbone (Figure 25). The vector itself was linearized by enzymatic digest with StuI. In addition, a unique restriction site (NotI) was inserted in between the two capturing sites to enable vector linearization for LLHR (Figure 26).

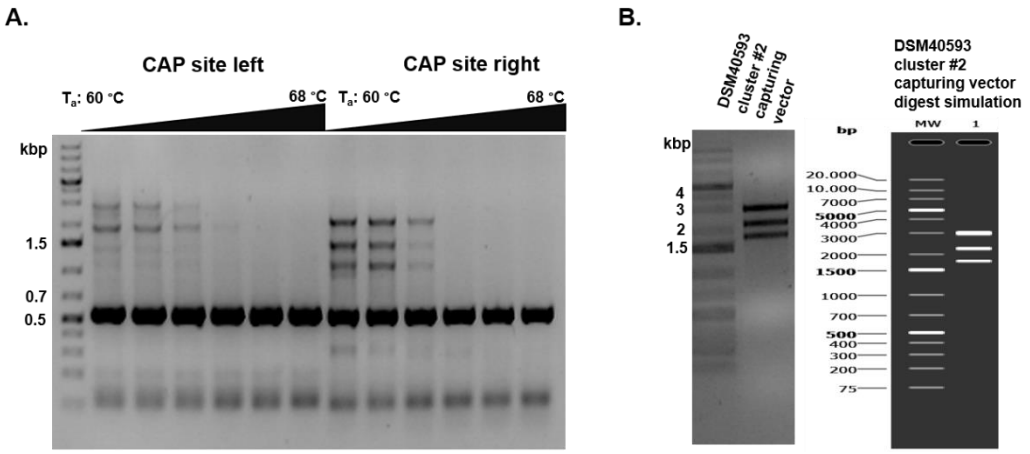
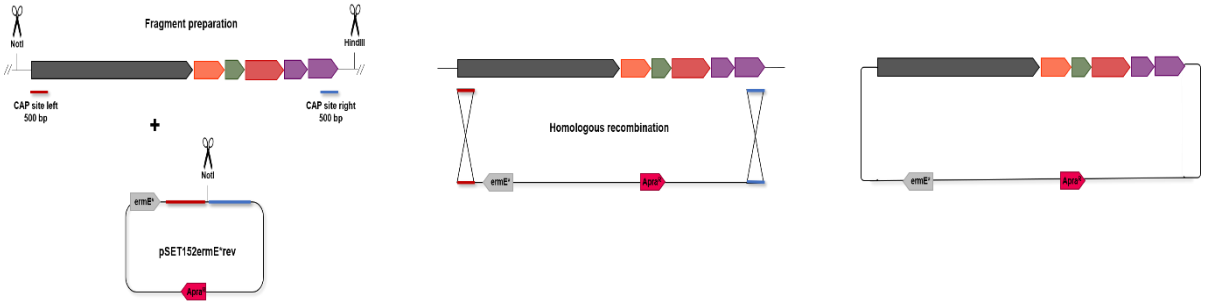


Figure 26. Assembly of capturing vector by three-fragment HiFi DNA Assembly. **A.** Homology arms of 500 bp flanking the BGC of interest were PCR-amplified using a gradient to determine the optimal annealing temperature. **B.** Analytical restriction digest of assembled DSM40593 cluster # 2 capturing vector using NcoI/NotI/BmtI (1734, 2183, 2930 bp). Theoretical restriction digest pattern simulated using SnapGene.[121]

Construction of the respective capturing vector for both BGC SFC2 and SCC37 was achieved by three-fragment HiFi DNA Assembly. Successful insertion of both 500 bp-homology arms into pSET152ermE*rev was validated by analytical restriction digest (Figure 25) as well as plasmid sequencing. Following preparative restriction digest of the capturing vector with NotI, clusters # 2 or # 37 were introduced into the expression system by co-transforming vector and insert DNA into *E. coli* GB05dir host cells capable of performing LLHR since they express all necessary proteins (Scheme 14).



Scheme 14. General LLHR procedure. DNA fragments are prepared for homologous recombination by enzymatic restriction digest. Recombination takes place between two linear sequences which share homologous nucleotide regions of 500 bp with the result of a fully assembled plasmid ready for heterologous expression.

Success of homologous recombination is dependent on genomic DNA quality and accessibility of the BGC of interest. For target DNA-capturing, genomic DNA was digested with unique restriction enzymes flanking the region of interest without cutting inside the BGC nucleotide sequence (Figure 27).

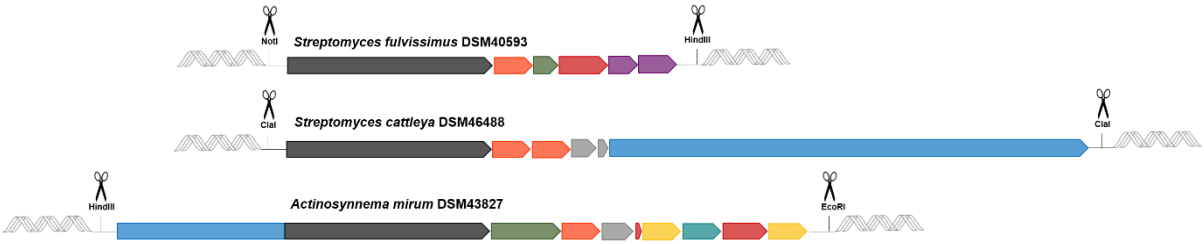


Figure 27. DNA-capturing by restriction digest of bacterial genomic DNA. Clusters # 2 (*Streptomyces fulvissimus* DSM40593; 12.5 kbp), # 37 (*Streptomyces cattleya* DSM46488; 29.1 kbp) and # 53 (*Actinosynnema mirum* DSM43827; 24.8 kbp) were excised from genomic DNA using enzymes that cut in the flanking regions of the BGC.

Streptomyces cattleya DSM46488 cluster # 37 (SCC37)

The cluster of interest was excised from genomic DNA by restriction digest with the enzyme ClaI which cuts 580 bp upstream and 615 bp downstream of the BGC boundaries. The left capturing site was designed to be homologous to the start codon of the NRPS gene whereas the right capturing site showed homology to the final type I PKS gene. Unfortunately, numerous attempts to assemble the pST152ermE*rev-SCC37 plasmid remained unsuccessful since the nucleotide sequence of interest failed to be captured by LLHR. Most likely, the high GC-content of more than 70 %, especially within the 500 bp capturing sites limits homologous recombination. Elaborated secondary structure formation and increased inverted repeat stability due to additional hydrogen bond formation compared to GC-low sequences make unwinding of the DNA more difficult and time consuming. Furthermore, formation of a complex with Rec proteins catalyzing LLHR differs between GC-rich and GC-low DNA and it is proposed that a GC-rich DNA-Rec protein complex exhibits enhanced sterical hinderance leading to blockage of recombination events.[122] In addition, due to repetitive sequences throughout the target BGC as well as its large size of almost 30 kbp, it could also be possible that while LLHR occurred, the assembled plasmid was instable and too much of a metabolic burden for the host cell. As an alternative, a low copy plasmid could be tested for BGC integration and expression in *Streptomyces*.[123]

Actinosynnema mirum DSM43827 cluster # 53 (AMC53)

LLHR of BGC # 53 into pSET152ermE*rev also remained without success, but in this case the target BGC could not be excised from genomic DNA using HindIII and EcoRI. Identity testing of isolated DSM43827 genomic DNA using general 16S primers confirmed the presence of bacterial DNA. However, annealing of strain specific primers failed multiple times. Once strain specific primer binding could finally be achieved, multiple bands were observed in the agarose gel, which could result in unspecific binding or the presence of impurities, thus questioning nucleotide identity and gDNA purity. Recultivation, isolation, and purification of gDNA from cells grown on solid and in liquid media did not solve this issue (Figure 28).

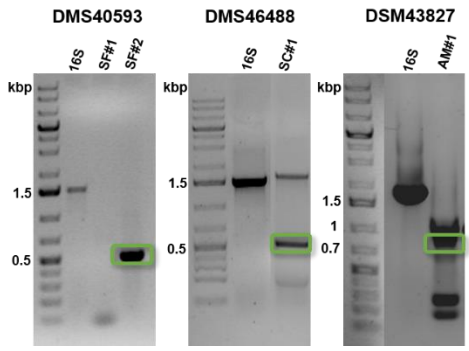


Figure 28. Confirmation of genomic DNA identity. General primers amplifying 16S-rDNA (27F/1492R) as well as strain specific primer pairs were applied to verify strain identity. Isolated genomic DNA from cultivated DSMZ strains proved to be both of bacterial origin and consistent with the expected specific bacterial strain.

Streptomyces fulvissimus DSM40593 cluster # 2 (SFC2)

The 12.5 kbp large cluster was captured from genomic DNA by NotI/HindIII combined restriction digest and inserted into the pSET152ermE*rev vector system by LLHR. Unfortunately, design of the left region of homology at the 5' end of the cluster happened to be unfit for heterologous expression as the start codon of the NRPS gene was too distant from the SD-sequence to be transcribed. Therefore, the original pSET152ermE*rev capturing vector needed to be replaced by a system containing a left capturing site directly homologous to the BGC start. Vector backbone exchange was achieved by consecutive rounds of LCHR, first transferring the BGC of interest to the pCC1FOS vector (Figure 29) before retransferring it to the adjusted pSET152ermE*rev capturing vector backbone (Scheme 15).

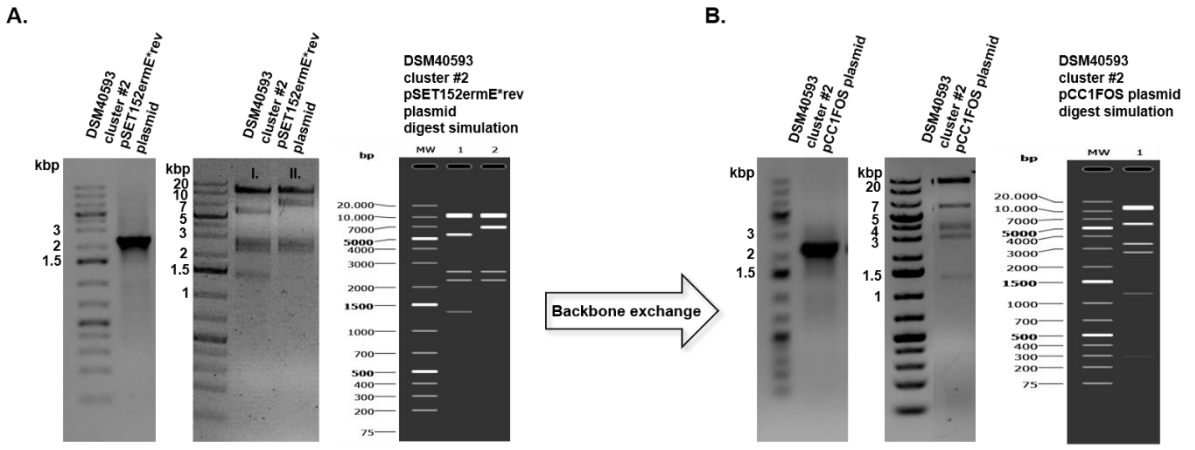
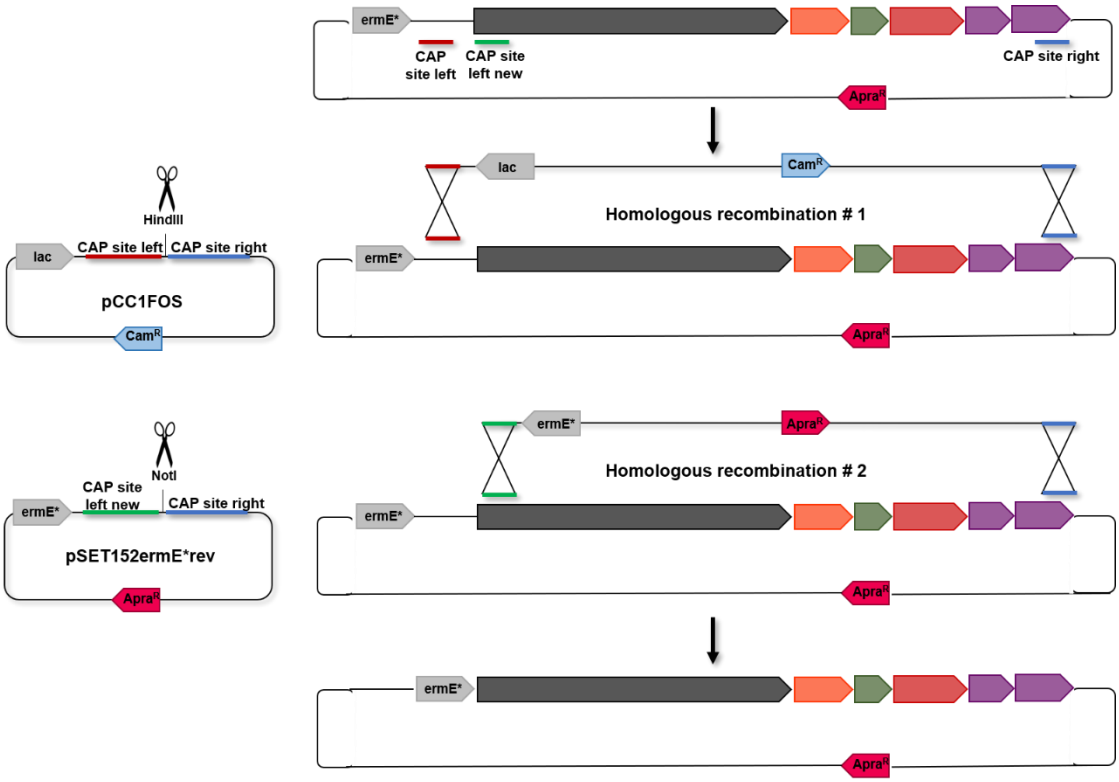


Figure 29. Initial round of vector backbone exchange by LCHR. Exchange of pSET152ermE*rev to pCCFOS backbone for SFC2 construct was performed in *E. coli* GBO8red cells. For screening of cells harboring either of the assembled plasmids, a primer pair was chosen which anneals to the vector backbone on one side and to the inserted BGC DNA on the other (**A.** pSET152ermE*rev construct: 2460 bp; **B.** pCC1FOS plasmid: 2354 bp). Validation of correct plasmid assembly was monitored using analytical restriction digest (**A.** pSET152ermE*rev plasmid: I. BmtI/SpeI/XmnI (1343, 2215, 2571, 5609, 10432 bp); II. BmtI/XmnI (2215, 2571, 6952, 10432 bp); **B.** pCC1FOS plasmid: SpeI/XmnI (70, 287, 1197, 2711, 3328, 5609, 11772 bp).

Agarose gels in Figure 29 show transfer of the target BGC DNA in between vector backbones after co-transformation of linearized target backbone with circular, insert-containing plasmid into LCHR-competent *E. coli* GB08red host cells. During the process of vector backbone exchange, KOD Xtreme™ Hot Start DNA Polymerase was purchased, allowing SFC2 to be PCR-amplified in a single piece, as described above in 3.2.1.1 for PTC12 and SclavuC39. The BGC of interest was subsequently introduced into pSET152ermE*rev using DiPaC, resulting in the same final plasmid as obtained by LLHR in combination with LCHR. The consecutive exchange of the vector backbone, first from pSET152ermE*rev to pCC1FOS and second back to pSET152ermE*rev possessing the novel capturing sites directly flanking SFC2, is illustrated in Scheme 15.



Scheme 15. Vector backbone exchange using LCHR. The BGC of interest can be transferred to alternative vector systems by recombination events of the circular cluster-containing plasmid with the novel linearized capturing vector DNA.

With the help of DiPaC and homologous recombination techniques most of the selected BGCs of interest were made accessible as it was possible to capture DNA segments of up to 16 kbp in size and over 70 % GC-content from genomic DNA. Target DNA was successfully introduced into vectors suited for heterologous expression of natural products, thereby controlling gene expression either using inducible (*E. coli*) or constitutive (*Streptomyces*) promoters. Furthermore, molecular genetic tools were applied to establish an expression system that allows more efficient natural product biosynthesis (e.g., by varying spacer region between RBS and start codon) as well as fast and easy manipulation (generation of double promoter systems and mix & match cloning).

3.2.2 Expression of PA derivative-producing BGCs

Over the past two decades heterologous expression of secondary metabolite-producing BGCs has become an efficient method to discover novel natural products and potential future drug candidates. Capturing of BGCs with subsequent expression in suitable heterologous host systems has given access to molecules available in only small titers or not at all in Nature. Often, cultivation of the native producers in the laboratory is complex and difficult or impossible and even if culturable, the BGCs of interest might remain silent.[124] Thus, heterologous expression *in vivo* gives the opportunity to either express entire BGCs proposed to be involved in biosynthesis of diverse PA derivatives or to generate pathway intermediates by expression of single enzymes.

Model systems such as *E. coli* and yeast have been intensively studied and to date multiple tools for host engineering have been established and applied to improve metabolite production while simultaneously reducing background signal. In general, *E. coli* is a frequently used model organism which features rapid growth, comparatively little demand concerning cultivation conditions and genome simplicity, making it ideal for studying and understanding diverse metabolic pathways in molecular biology and biochemistry.[125] Nevertheless, *E. coli* has limitations concerning heterologous expression of BGCs and their secondary metabolites derived from gram positive bacteria such as actinomycetes. Apart from variations concerning codon usage and GC-content of genetic information, posttranslational modifications of enzymes involved in natural product biosynthesis is performed by endogenous producers whereas native *E. coli* PPTases fail and thus it is necessary to co-express adequate proteins. Furthermore, specific precursors needed for molecule buildup may not be produced by *E. coli* hosts. It is therefore inevitable to select suitable (engineered) host systems to overcome model organism-related limitations.[126] Like *E. coli*, several *Streptomyces* strains have been genetically optimized and cultivation conditions were improved to facilitate secondary metabolite production of gram-positive strains of actinobacterial origin.[127]

Within the course of this thesis, both complete BGCs as well as truncated BGC versions were heterologously expressed in engineered *E. coli* or *Streptomyces* strains to activate PA-producing enzymes under laboratory conditions and to identify novel PA derivatives. Multiple parameters associated with cultivation conditions were varied to enable and improve secondary metabolite production and titers.

In general, all *E. coli* host strains used for heterologous expression experiments possessed a copy of a gene encoding for a PPTase, an enzyme necessary for *in vivo* activation of NRPS systems present in the BGCs to be expressed. Alongside BAP1 cells, which carry the Sfp

PPTase derived from *Bacillus subtilis*, the bacterial host DH10β::mtaA, which co-expresses the PPTase MtaA found in myxobacterium *Stigmatella aurantiaca*, was applied. Both PPTases show a relaxed substrate specificity towards several NRPS and PKS as well as hybrid NRPS/PKS systems and are thus broadly applicable to a wide range of heterologous expression experiments.[128]

Heterologous expression in *Streptomyces* was performed for BGCs of actinobacterial origin due to a similar codon usage and GC-content. The closely related chassis strains *Streptomyces coelicolor* M1154 and *Streptomyces lividans* TK24 are well-studied, easy to genetically manipulate model systems where endogenous competing BGCs have been deleted to enhance precursor availability and increase production of secondary metabolites of interest.[129] In addition, mutants Del14 and B2P1 of *Streptomyces albus* J1074 were used, host systems distinguished by their small genome and fast growth as well as a clean metabolic background.[130] Deletion of several if not all endogenous BGCs associated with secondary metabolism in combination with an increase of bacterial integration sites supports target BGC integration and simplifies detection of novel secondary metabolites.[131]

3.2.2.1 Proof of Principle – Heterologous expression of BGCs synthesizing 37

Regarding the close organizational resemblance to the pyrrolizixenamide-producing BGC identified in *Xenorhabdus stockiae* DSM17904, PTC12 and SclavuC39 were heterologously expressed as proof of principle (see 3.1.1, Figure 7). Since both BGCs consist of the minimal genes necessary for PA core structure assembly, the expectation was to find variations of **37** or slightly modified derivatives thereof.

PTC12

Being derived of a gram-negative bacterium and having similar codon usage as in *Xhenorhabdus stockiae*, DSM15199 cluster 12 seemed to be suited for heterologous expression in *E. coli*. Since the pyrrolizixenamide-encoding BGC of *X. stockiae* had successfully been expressed when under the control of an arabinose-inducible promoter, the same pCX2 expression vector received from the Bode group (University of Frankfurt) was used for cluster 12 experiments. Successfully established plasmids were transformed into and corresponding enzymes expressed by BAP1 cells. Multiple clones were grown at the same time and expression profiles were compared to a pCX2 empty vector negative control.

Multiple signals exhibiting low peak intensities were detected in the HPLC chromatograms of at least two of the cultivated *E. coli* BAP1 PTC12-pCX2 mutants (Figure 30, zoom in). Compounds with molecular ion sizes correlating to variations of **37** solely differing in the length of the hydrocarbon side chain moiety (C_8 – C_{11}) could be identified eluting at an ACN gradient of 43 – 52 %. Furthermore, when searching mass spectra for derivatives of shorter chain length such as pyrrolizixenamides A (**50**) and B (**51**), two additional molecule signals of interest were identified at t_R = 18.44 min (33 % ACN) and t_R = 21.22 (38 % ACN), where no adequate mass adduct was found in the negative control. With an *m/z* ratio of 237.1, the unique analogue ion of **50** could only be detected in the mass spectrum since the characteristic PA peak was overlaid by a more intense signal also found in the empty vector HPLC chromatogram. A total of six ion adducts with m/z ratios matching the $[M+H]^+$ masses of PA derivatives was identified by LC-MS, with relative abundances of compounds bearing an even number of carbon atoms in the side chain being significantly higher than those of molecules with an odd numbered carbon chain. Unequal distribution may be a result of interactions with endogenous *E. coli* fatty acid metabolism where general availability of even-chain fatty acids is much higher than that of odd-chain fatty acids.[132] Whereas glucose is used by microorganisms as a carbon source for generating even chains from acetyl- and malonyl-CoA building blocks, propionyl-CoA is the basic building block for odd chains and requires supplementation of the cultivation broth with propionate.[133] As a result, fatty acids and acyl-CoA with even carbon side chains are much more abundant and thus more likely to be incorporated into the respective pyrrolizixenamide structure compared to the less abundant odd-chained substrates.

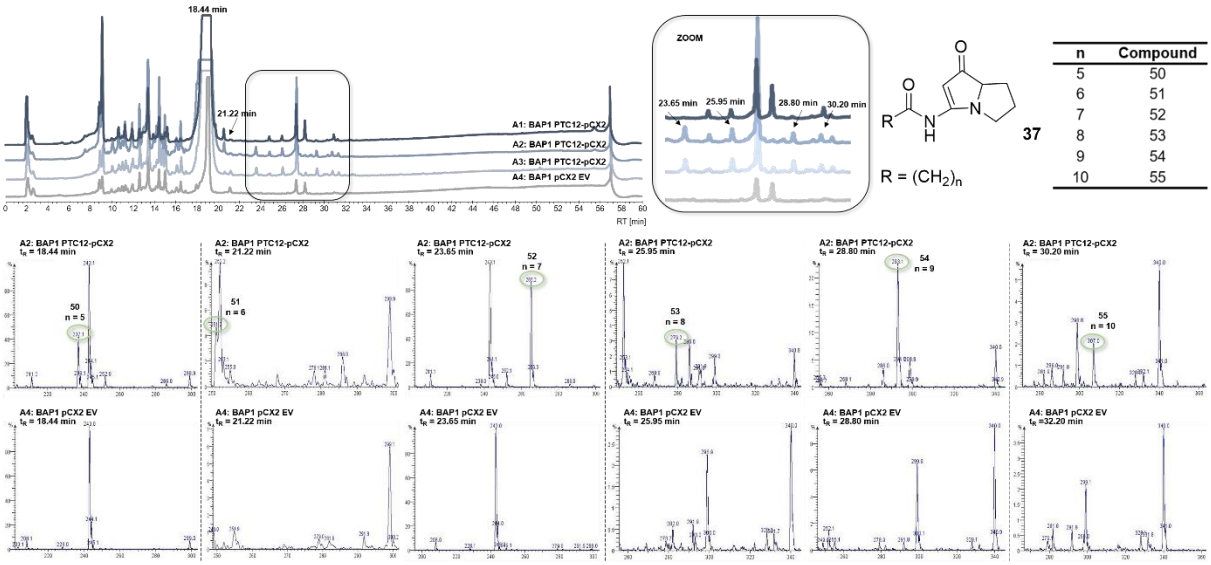


Figure 30. LC-MS data of PTC12 heterologously expressed in *E. coli* BAP1 cells. Boxed region in the HPLC expression profile was enlarged to allow detailed identification of secondary metabolite signals characteristic for BGC-containing clones. Mass spectra highlight relevant *m/z* ratios of analogues of **37** found in bacterial extracts. Variants of **37** differ in length of hydrocarbon side chain (see structure and table on upper right).

Raw extracts of grown cultures were pooled and purified by preparative HPLC (Figure S1). Four out of six initially identified PA derivatives could be rediscovered in high quantities and separated from each other. Unluckily, the molecule with a detected *m/z* ratio of 237 (peak 1, t_R = 19.8 min prep-HPLC) could not be obtained in amounts high and pure enough for further characterization as separation from the overlay peak was not possible. Compounds represented by peaks 2, 3 and 4a in the HPLC chromatogram on the other hand were of sufficient purity for NMR measurements. ^{1}H-NMR spectra of all three compounds looked like published data of **50** - **52** with the major difference being the number of CH_2 groups (Figures S2 - 4). According to the designated ppm shifts and peak integrals, the compound of peak 2 is the analogue of **54** and corresponds to the pyrrolizidine core carrying a C_9 moiety whereas peak 4a falls in line with a C_{10}-bearing pyrrolizixenamide analogue **55**.

In summary, heterologous expression of the minimal BGC of *Photorhabdus temperata* DSM15199 in *E. coli* yielded variants of secondary metabolite **37** with varying hydrocarbon side chain lengths and successfully served as a proof of principle for *in vivo* PA production under laboratory conditions. Not only could the previously identified pyrrolizixenamides A, B and C[79] be reisolated from this related bacterial species as compounds **50** – **52**, but also three additional, longer chained derivatives **53** – **55** were found. To our knowledge, the latter have not been isolated from any other BGC so far.

SclavuC39

The 8.5 kbp-sized SclavuC39 *cluster* was heterologously expressed in a variety of *Streptomyces* strains, including *S. coelicolor* M1154, *S. lividans* TK24 as well as multiple *S. albus* deletion strains. Production of secondary metabolites was detected for *S. albus* B2P1 and B3P1 clones (Figure 31).

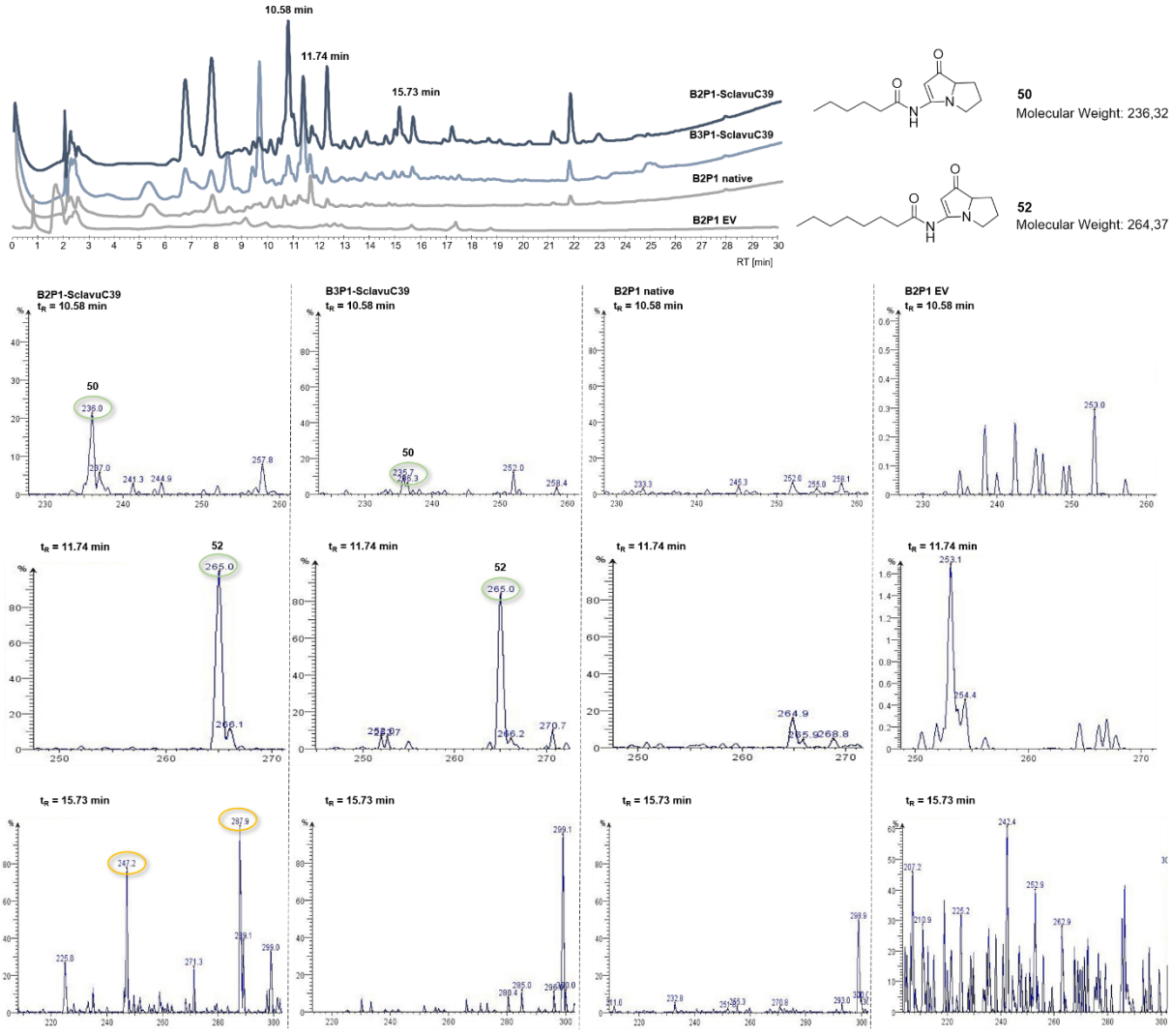


Figure 31. Heterologous expression of SclavuC39 in *Streptomyces*. Extracts of cultivated *S. albus* B2P1 and B3P1 host strains carrying cluster 39 were analyzed by LC-MS in search of novel PA-associated secondary metabolites.

As previously mentioned, predictions for product outcome predominantly included pyrrolizixenamide derivatives, therefore mass analysis concentrated on *m/z* ratios between 200 and 400. Two prominent molecule peaks were identified via HPLC at retention times of 10.58 and 11.74 min where mass ion peaks of *m/z* = 237 and 265 were abundant, which

correspond to masses of **50** and **52**. Comparison with both, an empty vector control and a natively grown *S. albus* B2P1 culture, confirmed the absence of these molecule types when cluster 39 was not available, leading to the assumption that secondary metabolites were produced by expression of the BGC. These findings go in line with the results obtained for PTC12 expression (Figure 29) where derivatives of **37** were also eluted at a gradient of 30 – 40 % ACN. Thus, it can be assumed that the detected molecules were indeed the expected products. Regarding the chromatogram of the B2P1-SclavuC39 extract, additional, less polar compounds were eluted from the column after 15 min. At t_R = 15.73 min (53 % ACN), a ion mass peak with *m/z* = 287.9 appeared with high relative abundance, which could correspond to the $[M+Na]^+$ mass of **52** (265 g/mol). A second ion (*m/z* = 247.2) could not be linked to production of **50**/**52**. For detailed analyses of the detected compounds, cultivation of the producing *Streptomyces* clones was repeated under identical conditions on a larger scale. However, the displayed results could not be reproduced and final molecule characterization remained incomplete.

As for minimal BGC heterologous expression events of PTC12 observed in *E. coli*, biosynthesis of predicted pyrrolizixenamides as a result of cluster 39 translation was verified in *Streptomyces* hosts. Alongside the expected secondary metabolites, additional signals exhibiting mass to charge ratios in the range of PA derivatives indicated the presence of further molecules of interest. Their identity, however, is yet to be disclosed.

3.2.2.2 Identification of novel PA derivatives via expression of complex BGCs

Heterologous expression of PA-related BGCs bearing additional genes up- and/or downstream of the core genes necessary for establishment of the pyrrolizidine core structure was executed by transfer of the respective BGCs into *E. coli* and *Streptomyces* hosts. Putative novel secondary metabolites were identified and analyzed by LC-MS.

SMg1C5

With a total of six genes and a length of 13.5 kbp, cluster 5 is almost double in size compared to the minimal gene cluster set consisting of NRPS and MOX genes. This cluster was of particular interest as it encodes a SAM-dependent MT, an enzyme known to either methylate small molecules at *C*, *O*, *N* and *S*-atoms[134] or to use SAM as a precursor for L-AZC biosynthesis.[135] As mentioned in section 3.2.1.2, L-AZC can be incorporated instead of L-proline into the PA backbone and thus functions as a basic PA building block. Furthermore, the presence of a cytochrome P450 encoding gene may hint at hydroxylation of the aromatic

ring system or heteroatom oxidation.[136] Thus, if expression works, the chance of finding a novel secondary metabolite is high.

Being of actinobacterial origin, BGC 5 was initially captured, conjugated, and expressed in host strains *Streptomyces coelicolor* M1154 and *Streptomyces lividans* TK24. Liquid cultures of multiple clones were cultivated to examine differences in the respective expression profiles and to support reproducibility of the experiment. Expression profiles of two cultivated M1154-SMg1C5 clones as well as an M1154 empty vector negative control are shown in Figure 32A.

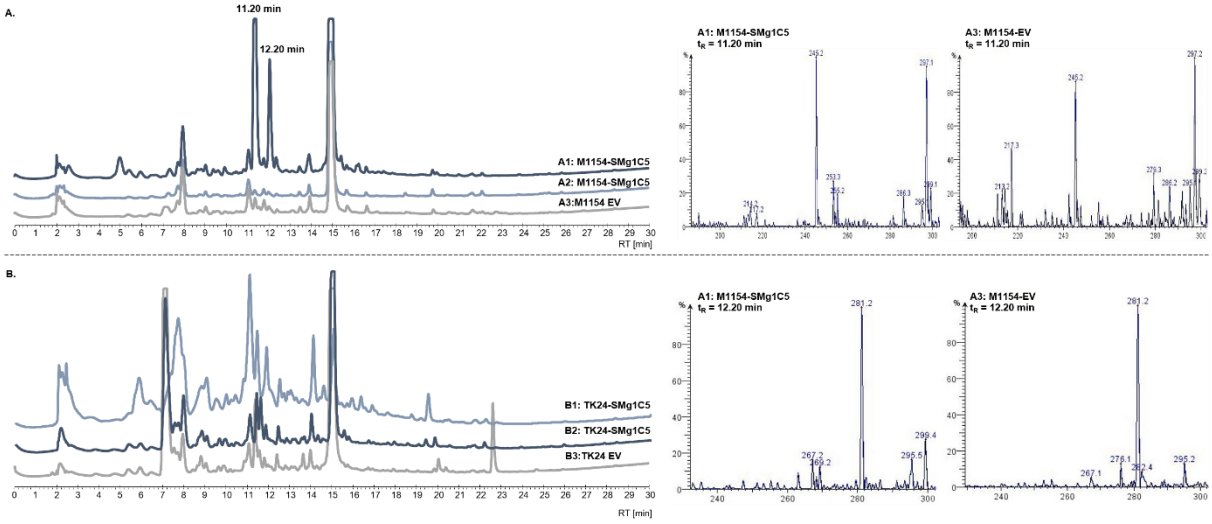


Figure 32. LC-MS heterologous expression profiles of SMg1 cluster 5 in *Streptomyces* host strains *S. coelicolor* M1154 (**A**) and *S. lividans* TK24 (**B**). EV = Empty vector. Negative control comprises conjugated pSET152ermErev* empty vector expressed in the respective host strain. Peak at t_R = 15 min corresponds to nalidixic acid (*m/z* = 233).

Clone A1 clearly shows metabolite production associated with the presence of the SMg1C5 genomic information at t_R = 11.2 min and t_R = 12.2 min. Unfortunately, evidence for the presence of a PA-related compound could not be confirmed by mass analysis since all mass ions occurring at the respective retention times were also present in the negative control. Furthermore, the obtained result was not reproducible as the chromatogram of clone A2 does not show these peaks and repetition of the experiment using the same clone also yielded an empty chromatogram. One possibility for the absence of characteristic ions for these peaks could be that compounds were not ionized properly and thus fragments did not “fly” to the detector. Optionally, the produced secondary metabolites do not show PA character, thus varied distinctly in size and were not detected within the molecular weight range of PA derivatives. Unlike detection of two distinct peaks in the extracts of M1154 clones, SMg1C5 expressed in TK24 did not result in the production of any novel molecules (Figure 32B). These

experiments suggest that cluster 5 remained silent upon heterologous expression and chosen laboratory cultivation conditions failed to activate secondary metabolite production.

Simultaneously, SMg1C5 integrated into the pETpTetGFPv2 vector was expressed in *E. coli* BAP1 cells for five days at 20 °C. Green fluorescence of the obtained cell pellet demonstrated successful gene transcription and amino acid translation of the entire BGC. Extraction of putative novel natural compounds was performed either by extracting supernatant and pellet of the cultured cell broth separately or alternatively by addition of Amberlites XAD-16 24 h prior to extraction. Amberlite polymeric adsorbents assist in removal of nonpolar compounds from a polar solvent such as the water-based cultivation medium. By extracting molecules with the resin, the overall background signal could be reduced significantly (Figure 33).

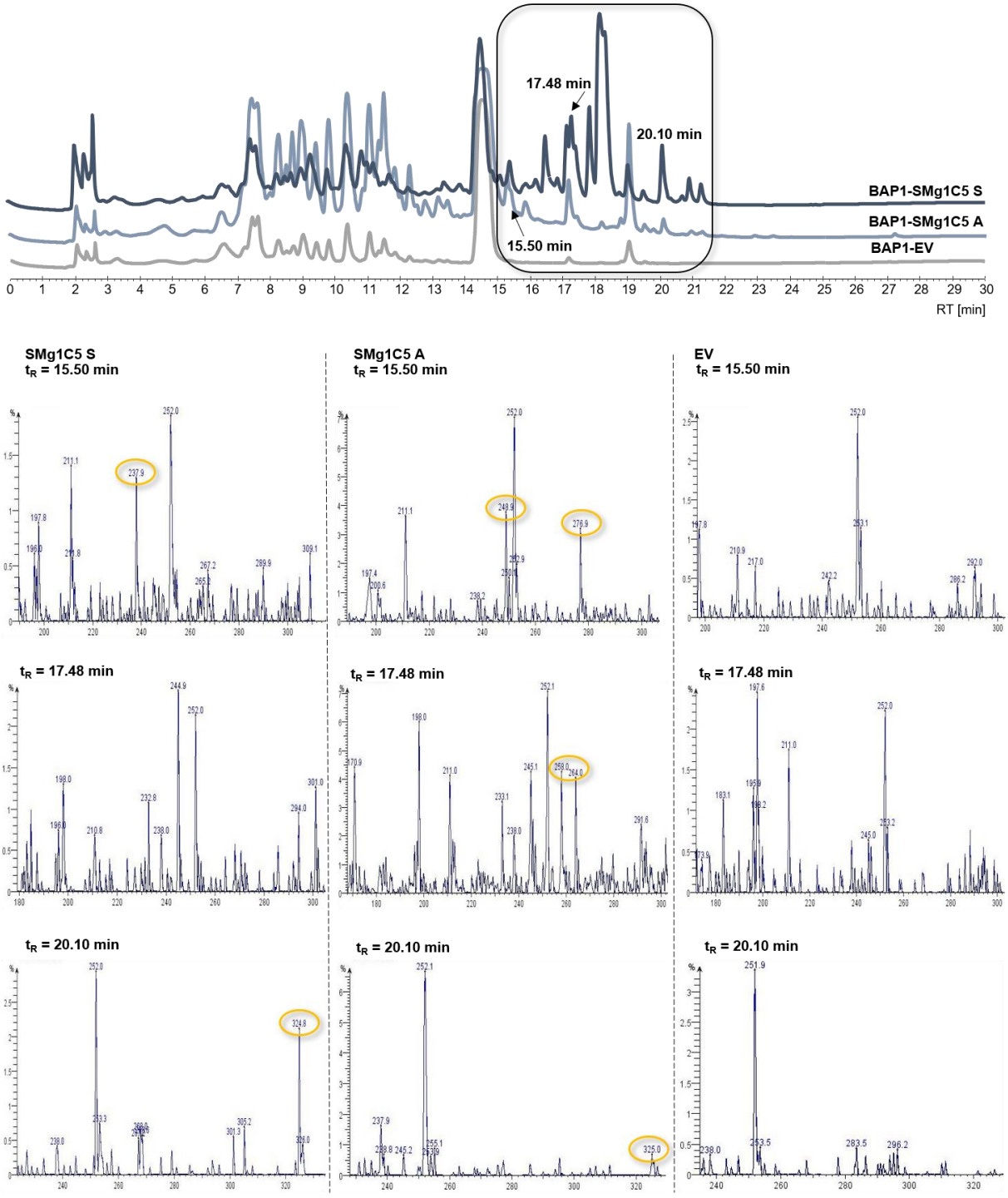


Figure 33. Heterologous expression of SMg1 cluster 5 in *E. coli*. LC-MS data of verified clones in comparison to pETpTetGFPv2 empty vector negative control. Extracts were processed either by supernatant extraction or extraction of Amberlite XAD-16 loaded resin. S = supernatant fraction; A = Amberlite XAD-16 extracted fraction; EV = Empty vector.

HPLC expression profiles of both supernatant and Amberlite-extracted fractions showed multiple peaks after a retention time of 15 min and when using the resin, a distinct reduction

of signals could be observed. Mass spectra of signals occurring in both extracts showed traces of ions within the size range of putative PA derivatives (*m/z* = 200 – 350). The low relative abundance of these mass ions made a conclusive identification of the corresponding molecules difficult and measurement repetition using high resolution MS could not overcome these hurdles. Regarding the detected mass ion peak of *m/z* = 325 at t_R = 20.1 min and referring to gene cluster organization, a hypothesized PA natural product could share similarities with legonmycin A (**39**), a molecule which bears both a hydroxyl as well as a methyl group at the core structure.[91] Alternatively, either a hydroxyl or a methyl function may be attached to the C10-hydrocarbon side chain. Preparative HPLC was performed for a 4 L culture extract processed without Amberlites XAD-16 to purify peaks of interest and gain further information on secondary metabolite production. LC-MS as well as HR-MS measurements, however, did not bring clarity, since multiple masses were identified. Thus, isolated peak fractions needed to be further purified, e.g., by using isocratic elution to remove remaining impurities.

Alongside signature peaks detected in both processed extracts, the molecule peak found in the supernatant fraction at t_R = 18 min was of interest due to its high abundance within the sample (Figure 33).

In total, four rounds of preparative HPLC were conducted to obtain 8 mg of a fairly pure substance. Whereas most impurities of the raw extract could be removed, the peak of interest was actually composed of two fractions and separation proved to be challenging. The initial ACN gradient of 5 – 100 % was first adjusted to 25 – 100 % followed by 40 – 100 % to obtain better separation results, but separation was less than ideal (Figure 34). Therefore, an isocratic elution was applied for final fractionation. Based on the expression profile of the substance of interest with elution between 55 and 65 % ACN + 0.05 % TFA solvent when using a gradient run, a proportion of 60 % ACN was chosen for isocratic separation. Despite these extensive efforts to isolate the compound, the obtained material was not of sufficient purity and quantity. Therefore, renewed expression at larger scale followed by isolation and structure elucidation needs to be conducted in the future.

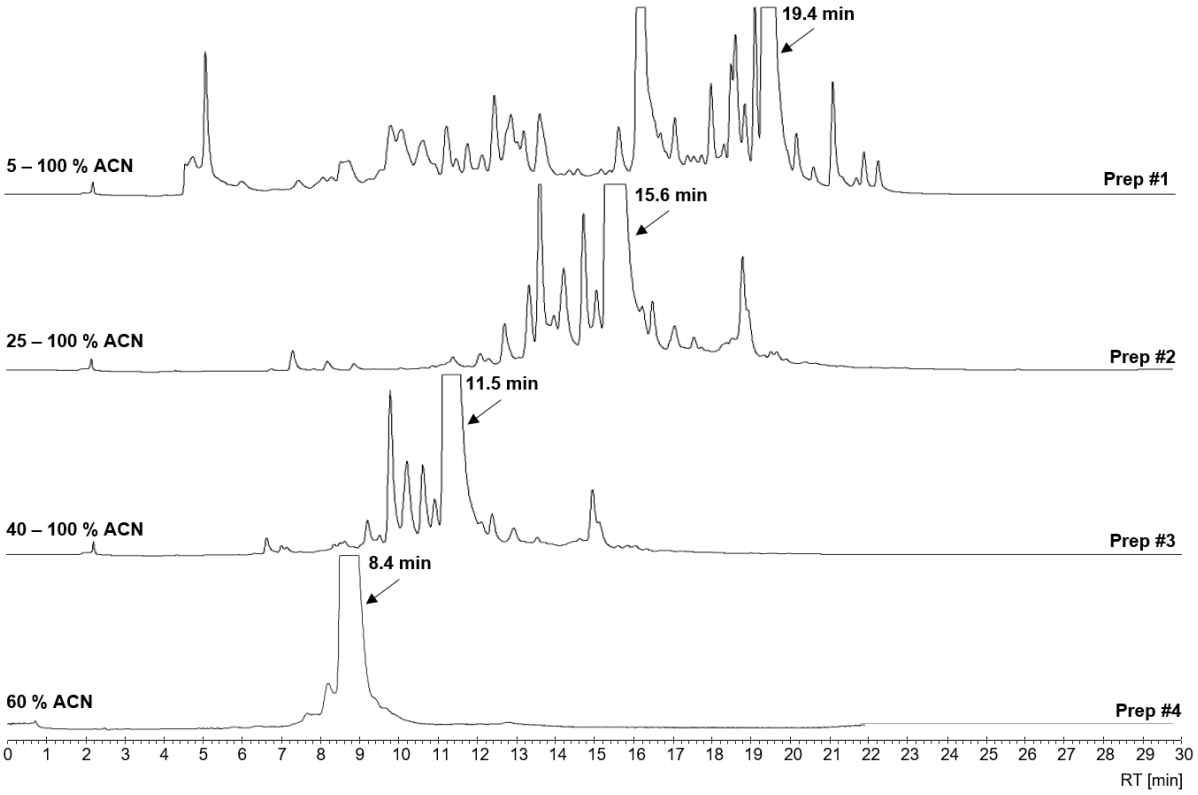


Figure 34. Preparative HPLC runs for stepwise SMg1C5 supernatant main peak purification. Three consecutive ACN gradients were run before using an isocratic method as a final separation step.

Whereas BGC expression in the *E. coli* host system seemed to be of success, no explicit production of PAs was found in *Streptomyces* clones which gave clear evidence of cluster activation. Although two distinct peaks were detected by HPLC, no typical masses underlining presence of the molecules of interest could be assigned. In contrast, a unique expression profile was seen for *E. coli* cultures harboring SMg1C5. Multiple signals indicating secondary metabolite production were isolated and prepared for structure identification and elucidation. Since compounds of interest exhibited molecular ions within the size range of PA derivatives but did not coincide with molecular weights of **50** - **52**, the chance of identifying a novel derivative was high.

SFC2 (DSM40593)

A similar-sized BGC of *Streptomyces fulvissimus* DSM40593 origin was also studied by heterologous expression in *S. coelicolor* M1154 and *S. lividans* TK24. The full-length cluster consisting of six genes, including the two core NRPS and MOX genes as well as four genes encoding for modifying enzymes located downstream, is of 14.5 kbp size. Alongside two oxidoreductases, which are most likely involved in substrate oxidation via hydrogen transfer to a nicotinamide or flavine co-factor[137], an ATPase-like protein involved in energy supply as well as a phenylacetic acid degrading protein PaaN putatively providing the acyl-CoA precursor via a ring-opening reaction could be identified.[138] Based on the encoded enzymes, it may be assumed that the secondary metabolite produced putatively resembles a pyrrolizixenamide-like product.

Similar to the previously described expression of SMg1C5, no production of secondary metabolites could be detected in TK24-SFC2 extracts (Figure 35A). These findings indicate that the tested *S. lividans* strain is presumably unsuited for the experimental challenges of this study. In contrast, several novel peaks were detected in two of the four M1154-SFC2 extracts. Clone B1 showed three distinct peaks absent in all other extracts including the negative control. Therefore, produced metabolites were assumed to be linked to SFC2 expression.

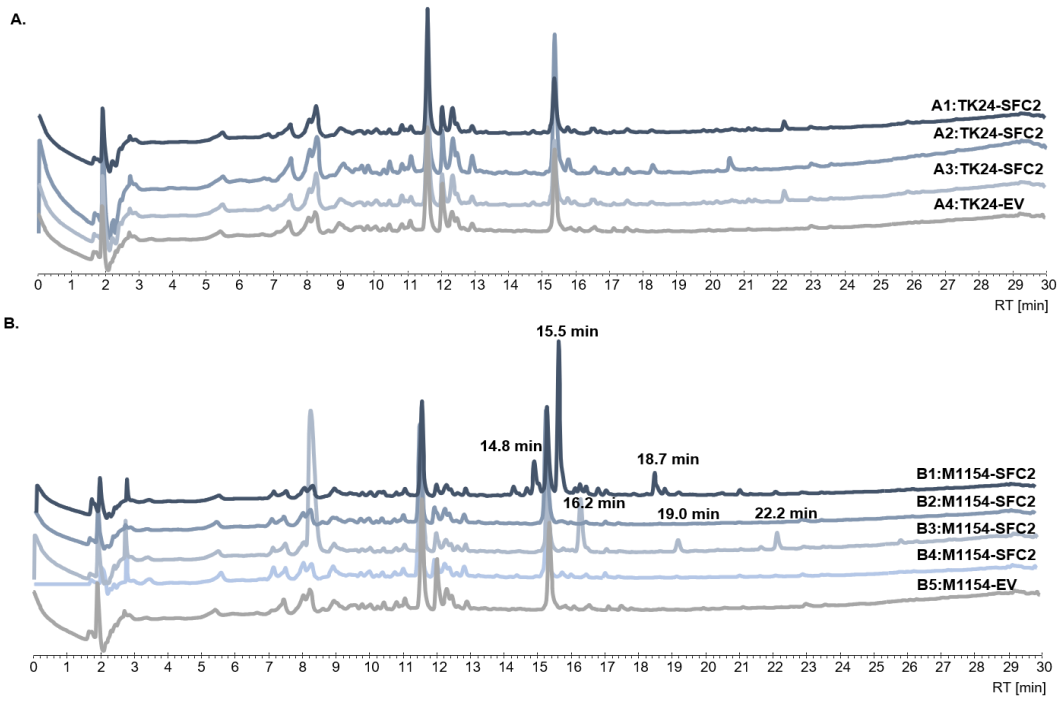


Figure 35. LC-MS expression profiles of SFC2 heterologously expressed in **A**. *S. lividans* TK24 and **B.** *S. coelicolor* M1154. Multiple clones were picked and expressed to support reproducibility of the results. EV = Empty vector pSET152ermErev* negative control.

For mass analysis of the detected peaks, an *m/z* segment ranging from 200 – 300 is displayed in Figure 36, because most published PA derivatives possess corresponding molecular weights. Nevertheless, the full mass spectrum from *m/z* 10 to 2000 was searched for adducts which might be associated with novel PA molecules.

For clone B1 (Figure 35B), the first metabolite eluting at t_R = 14.8 min did not show highly abundant adducts within the size range of putative PA derivatives, since all ions produced by ESI were also seen in the negative control. Simply the presence of an ion of *m/z* = 249 might be associated with secondary metabolite production due to SFC2 expression. The molecule found in all extracts of both TK24 and M1154 expressions at ca. 15 min retention time corresponds to the antibiotic nalidixic acid (232.235 g/mol), which was supplemented to the culture broth at the beginning of cultivation. The main peak found in the B1 extract at t_R = 15.5 min showed an abundant adduct of an *m/z* ratio of 296 (30 % relative abundance). Unlike ions detected in all fractions independent of elution time (e.g., *m/z* = 227, 241, 255 and 264), this adduct was unique to the retention time. Unfortunately, the same adduct but with less relative abundance was also seen both in clone B3 (20 %) and the negative control (15 %). The differences in relative intensity of the signal can be explained by variations in raw extract concentration of the respective sample injected. The third molecule of interest at t_R = 18.7 min looked to be more promising. An ion adduct with a *m/z* ratio of 238 was detected with decisively higher abundance than found in extract B3 and the negative control (Figure 36). Compared to other fractions at different retention times where clones B1, B3 and the negative control displayed this molecular weight with similar intensities, the clear difference in ion abundance may give evidence that at t_R = 18.7 min the ion peak at 238 could be a specific product of SF cluster 2 expression. In addition, the low abundance in B3 might explain why no molecule peak was detected in the HPLC chromatogram (Figure 36).

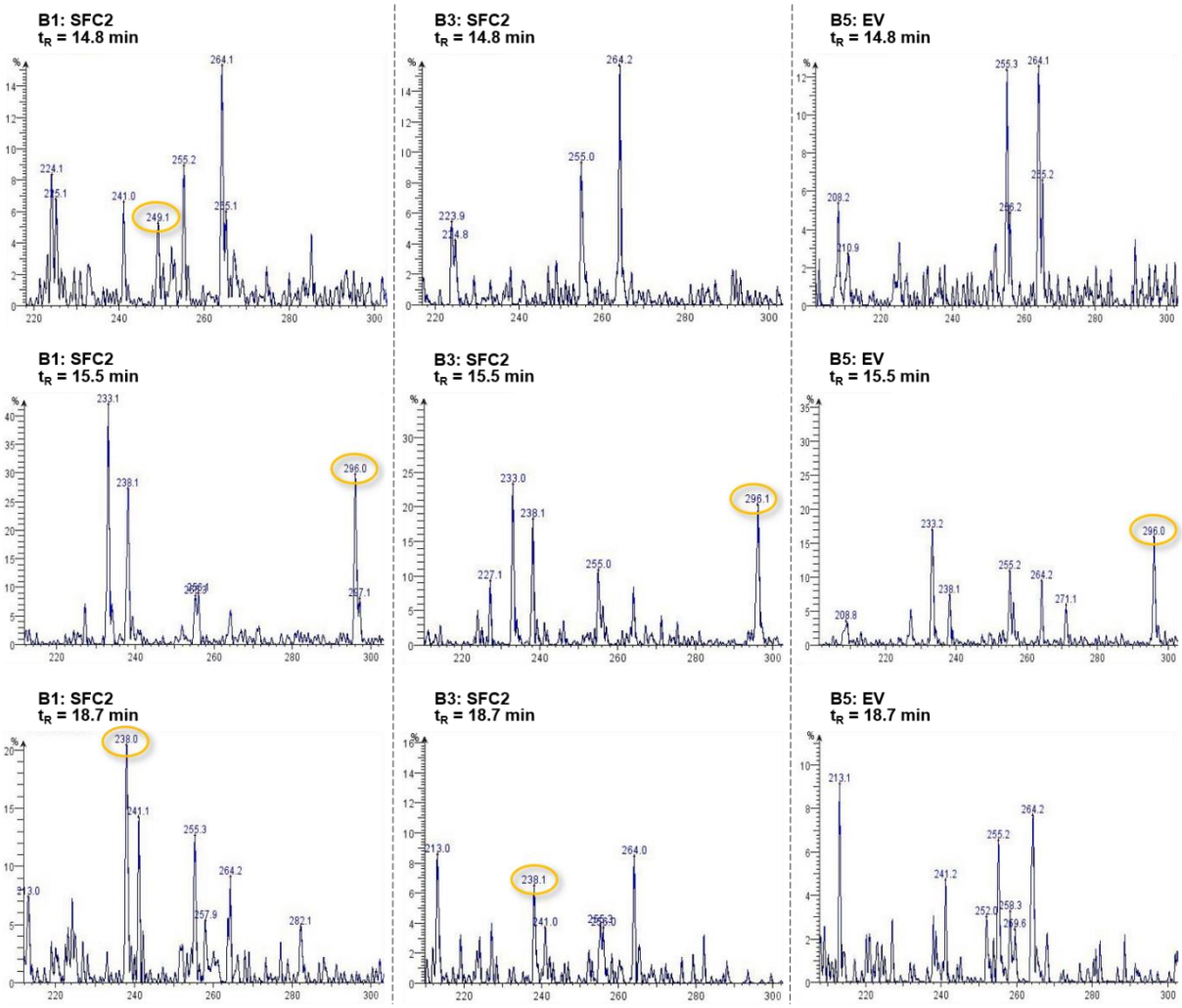


Figure 36. ESI-MS data obtained for significant molecule peaks detected by coupled liquid chromatography for M1154-SFC2 clone B1. Putative ion adducts of interest marked in yellow.

Secondary metabolites seen in the HPLC chromatogram of clone B3 (Figure 35B) turned out to be of even higher interest in terms of finding pyrrolizixenamide-related derivatives as two of the three products seemed to be unique to the extract. Whereas the peak with the highest UV signal (t_R = 16.2 min) exhibited an ion at *m/z* = 258 in both extracts B1 and B3 but not in the negative control, a unique mass signal was observed at t_R = 19 min. With a mass of 274 g/mol, the molecule has a similar size compared to pyrrolizixenamide derivatives. Furthermore, at higher retention times such as the metabolite eluted after 22.2 min, adducts with even higher *m/z* ratios (> 270) could be found that only appeared in the B3 extract (Figure 37).

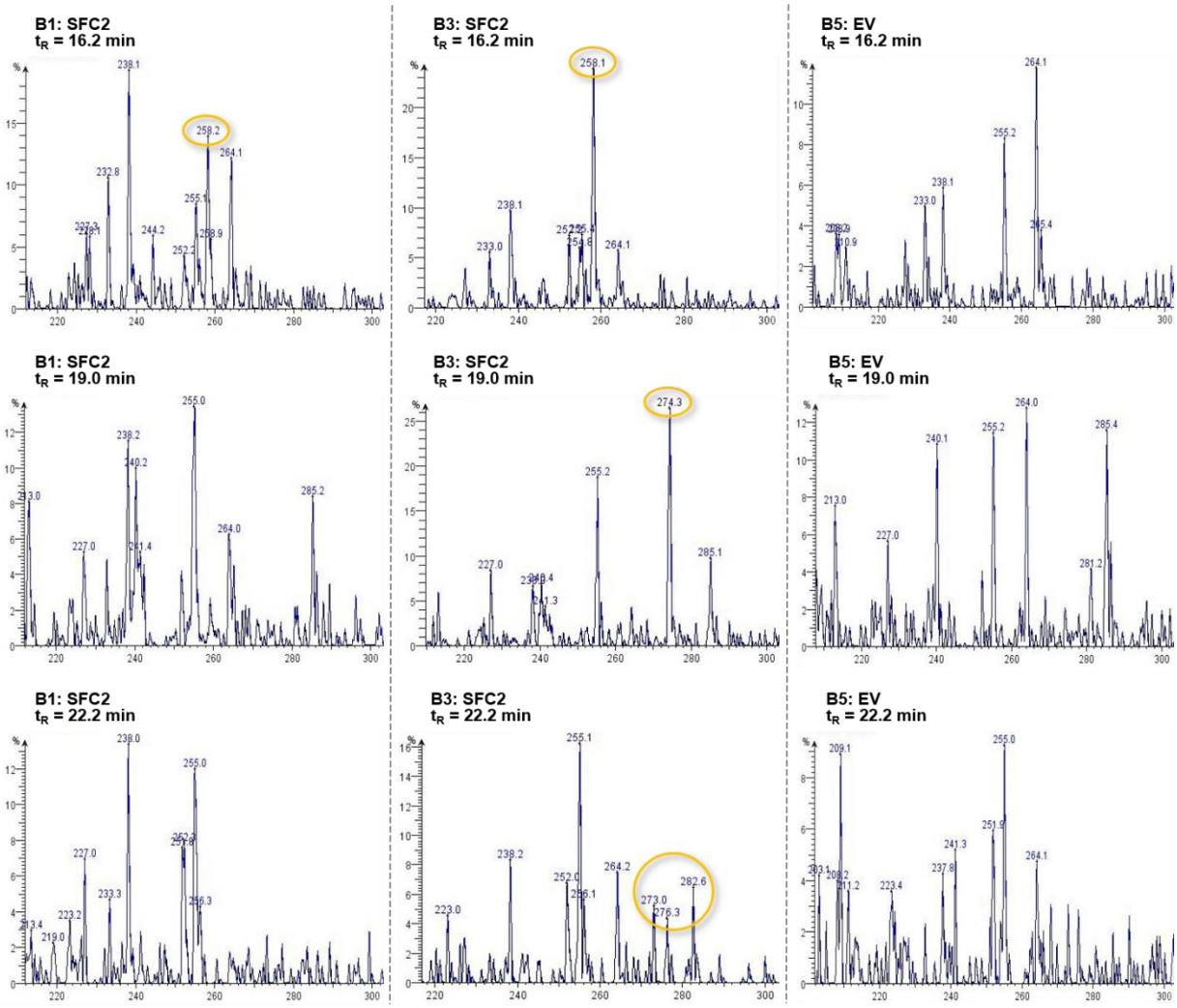


Figure 37. ESI-MS data obtained for significant molecule peaks detected in HPLC expression profile of M1154-SFC2 clone B3. Putative ion adducts of interest marked in yellow.

Taken together, heterologous expression of SFC2 seemed to be successful for individual cases, with the M1154 host strain being more suitable than TK24. Multiple peaks representing putative PA derivatives were detected in the HPLC chromatograms of two clones and masses could be assigned which are in line with molecular weights of published PAs. It is of interest that two clones bearing an identical BGC exhibited completely different expression profiles, as one would assume production of the same molecules upon BGC activation. This could be the case, if one of the clones did not express the full-length BGC of interest, whereas full translation proceeded in the other clone, thereby leading to the presence of different expression metabolites in the extract. Alternatively, BGC activation did not take place in one of the clones and signals were the result of endogenous secondary metabolism. To gain further insight into the ongoing expression mechanisms, both clones B1 and B3 were recultivated at a 200 mL scale. Unfortunately, the originally detected products could not be identified in the scale-up

(Figure S5). Because reproducibility of the experiment was not possible and the existing amount of raw extract of the 50 mL culture was insufficient for preparative purification and subsequent characterization, outcome of the heterologous expression experiment and identification of the produced molecules remained open. Thus, speculations on putative molecular compositions would be very vague at this stage of investigation.

PAO1C29

Pseudomonas aeruginosa PAO1 is known to be an opportunistic human pathogen due to its resistance towards multiple antibiotics and disinfectants, making it a major risk for hospital patients[139]. Interestingly, this gram-negative bacterial strain categorized as a class 2 biological safety hazard possesses a BGC encoding genes potentially involved in PA biosynthesis. Alongside the core NRPS and MOX enzymes, genes encoding proteins involved in ring modification and acyl side chain synthesis are included in the BGC. Simultaneously to expression efforts made in both *Streptomyces* and *E. coli* during this thesis, Hong et al.[115] identified azetidomonamides (**48**, **49**) as a novel set of PA molecules when cultivating the native as well as mutant strains of PAO1. They could link biosynthesis to expression of the full-length cluster 29 and set up a proposal for molecule biosynthesis where the last gene of the cluster, a SAM-dependent enzyme, plays a decisive role in amino acid precursor supplementation.

Based on the published information, the objective was to study promiscuity of the minimal gene cluster in the absence of downstream genes. To investigate whether alternative PA derivatives, like **37**, are produced when substrates specific to formation of the azetidine ring are not available, either only NRPS and MOX enzymes were expressed or co-expression with the SAM-dependent enzyme AzeJ was initiated. All culture broths were supplemented with L-Ser, L-Pro and in some cases also L-Met, a pre-precursor of L-AZC, the proline derivative incorporated into the ring system of **48** and **49**.

Heterologous expression of the minimal gene cluster designated PAO1C29**s** in *S. lividans* TK24 and *S. coelicolor* M1154 led to the identification of three distinct secondary metabolite signals associated with the presence of the BGC.

For M1154 clones, two to three main signals as well as a few smaller, less intense peaks were observable in extracted supernatants which differed from the empty vector expression (Figure 38A). As some of the detected signals were also present with comparatively low intensity in the negative control, mass analysis was necessary to distinguish between unique molecule

peaks and background signal. Unlike previously expressed BGCs, all clones A1 – A3 showed identical peak distribution, supporting reproducibility of the expression result. Furthermore, the major signals of interest visible at retention times between 10 and 13 min could also be found in the corresponding pellet fractions.

Looking at supernatant extractions of TK24 clones, expression profiles of empty vector sample B4 and PAO1C29s samples B1 – B3 were identical (Figure 38B). However, pellet fraction B1 indicates production of molecules of interest since three peaks eluting at the same retention times (t_R = 10.6, 11.4 and 12.2 min) as the M1154 supernatant fraction metabolites were observed.

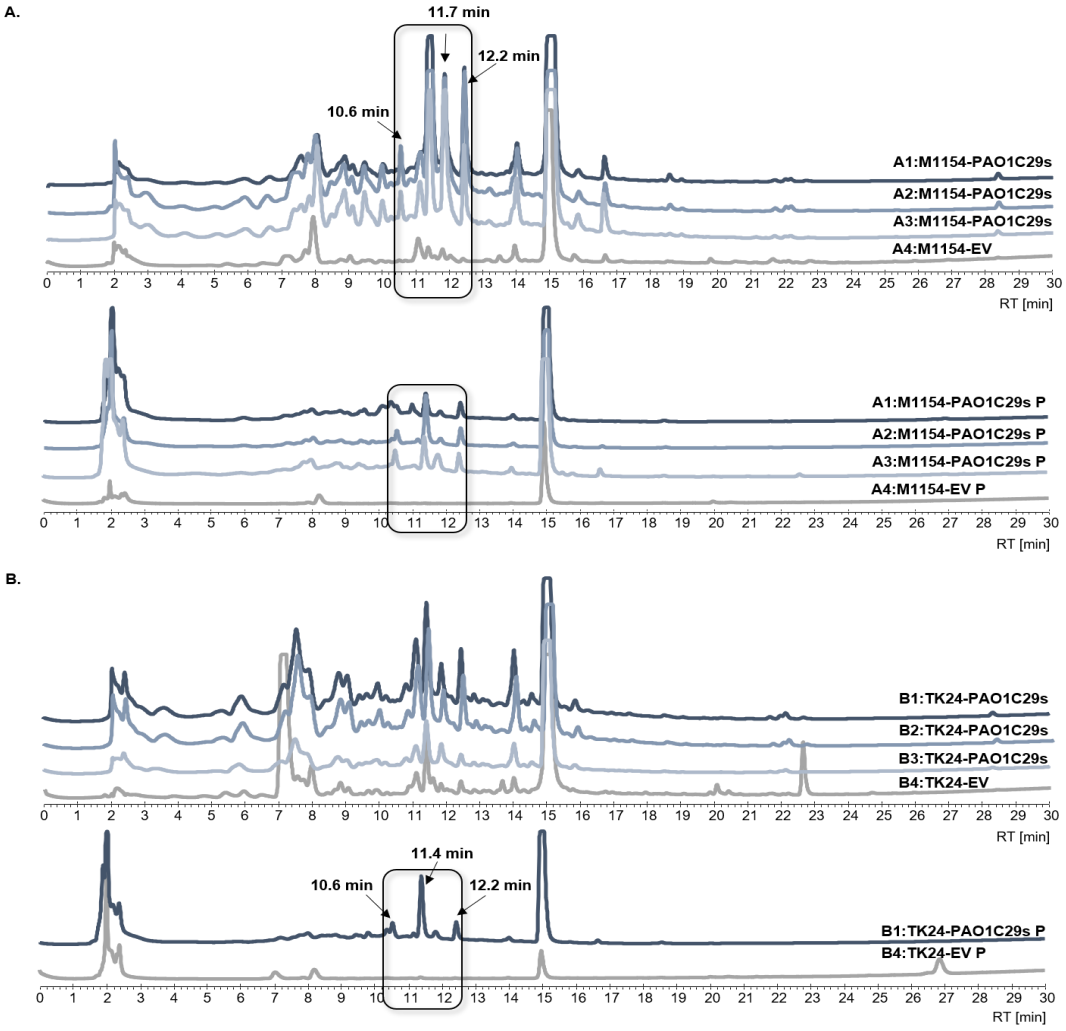


Figure 38. HPLC expression profiles of PAO1C29s heterologously expressed in *Streptomyces*. **A.** *S. coelicolor* M1154 PAO2C29s clones compared to empty vector negative control. Supernatant (top chromatogram) and pellet (lower chromatogram) fractions were extracted and measured separately. **B.** Expression of multiple *S. lividans* TK24 PAO1C29s cultures alongside negative control. Expression profiles of both supernatant (top) and pellet (bottom) fractions given.

Mass analyses of the molecules with retention times of t_R = 10.6, 11.7 and 12.2 min revealed the presence of unique ion adducts for each of the three peaks of interest (Figure 39A and B). A fourth signal at 11.3 min did not exhibit any characteristic molecular ions differing from the negative control. Compared to **48** and **49** which exhibit masses of 236.1 and 192.1 g/mol, the identified compounds displayed decisively larger molecular weights, nevertheless staying in the size range of PA derivatives. The highly abundant molecule eluted at t_R = 12.2 min with an *m/z* ratio of 249.1 was of particular interest, since the size could correspond to the $[M+ACN+H]^+$ adduct of a predicted molecule **60**, a derivative of **49** bearing an L-proline residue instead of the L-AZC amino acid. Theoretically, the present metabolite could also be an L-proline containing derivative (**61**) of **48**, however the proposed biosynthetic pathway suggests that **49** is synthesized by the minimal gene cluster whereas **48** includes additional ring modification steps executed by tailoring enzymes which were absent in the expressed PAO1C29s construct (Figure 40). On the other hand, incorporation of a heptanoyl moiety in combination with the L-proline building block by the minimal BGC would lead to formation of **62** instead of **48**. With a molecular weight of 249.2 g/mol, this molecule would share the same mass as a derivative of **51** bearing an extra double bond. Preparative HPLC was performed to isolate the molecule of interest and prepare it for characterization via NMR analysis. Although multiple rounds of preparative HPLC were performed to purify the product of interest, initial 1D- and 2D-NMR analysis revealed remaining impurities as well as low resolution due to only small amounts of final compound available, making distinct molecule identification difficult. Thus, the expression experiment needs to be repeated at bigger scale in the future.

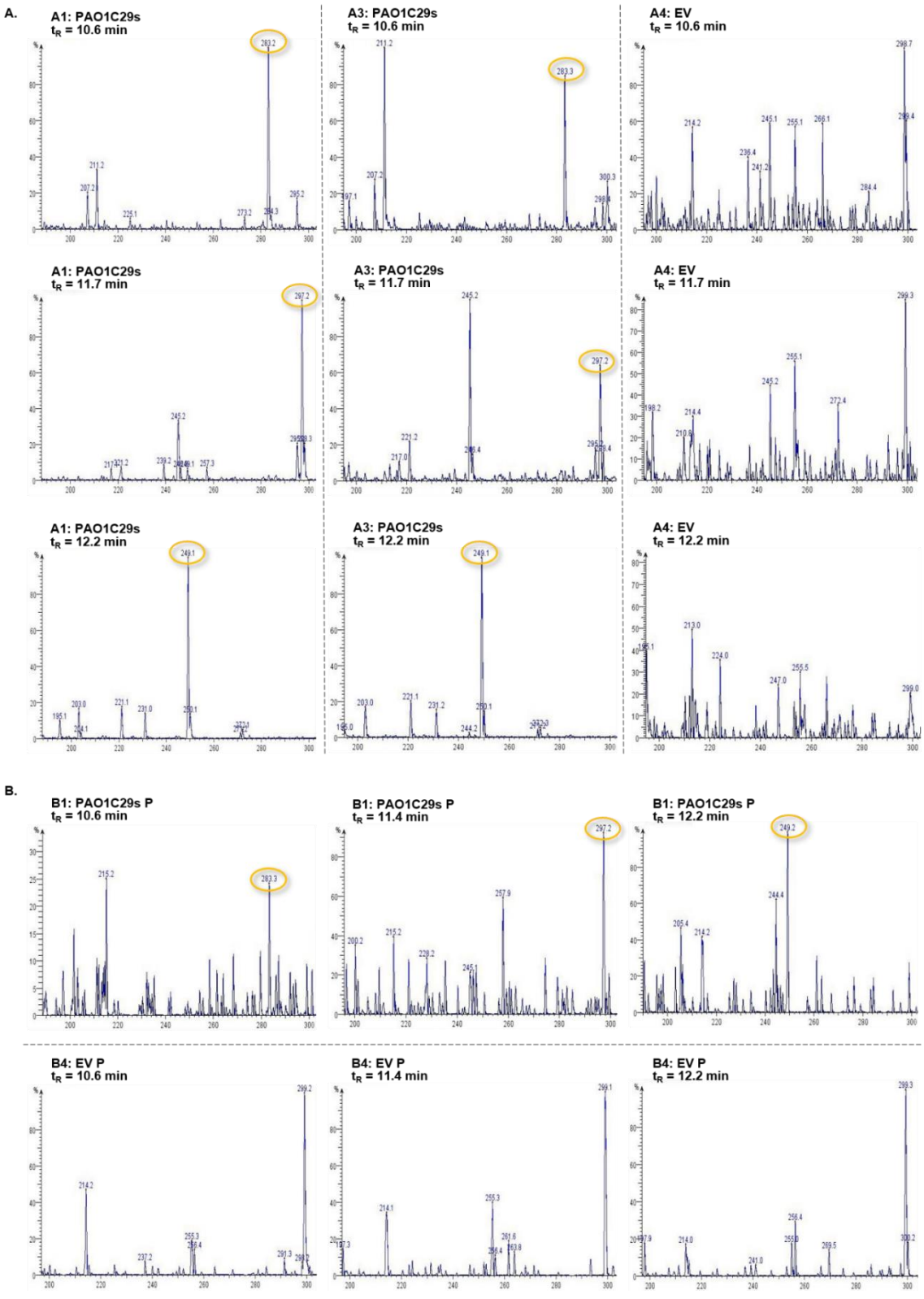
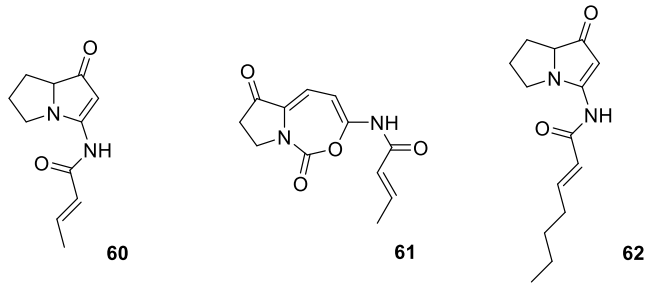


Figure 39. Mass spectra of characteristic PAO1C29s secondary metabolites. **A.** Mass data given for samples A1 and A3 representative for all M1154 clones. **B.** Confirmation of compounds in pellet fractions shown for TK24 B1 sample. EV = Empty vector negative control in respective host strains.



Compound	Molecular weight (g/mol)
60	206.11
61	248.08
62	248.15

Figure 40. Putative azetidomonamide-like natural products **60** - **62** as a result of PAO1C29s heterologous expression in *Streptomyces*.

The other prominent molecules present in the raw extract displayed abundant molecular ion peaks of *m/z* = 283.3 (t_R = 10.6 min) and *m/z* = 297.2 (t_R = 11.4 min) and are considerably larger in size than **48** or **49** variants. These compounds were likely to be associated with biosynthesis of alternative secondary metabolites based on the substrates available in the cultivation medium, making them interesting for further investigation. Unfortunately, despite being able to reproduce the experimental results on a small scale, not enough purified compound material was left after preparative HPLC of the peak at 10.6 min to proceed with NMR analysis and the molecule detected at 11.4 min retention time seemingly decomposed as the signal was already absent in preparative HPLC chromatograms.

Simultaneously to expression in *Streptomyces*, PAO1C29s was introduced into the pET28bpTetGFPv2 vector and cultivated in *E. coli* BAP1. By using this specific vector system, monitoring of full gene translation was possible due to a C-terminally encoded GFP protein. Positively screened clones were cultured in both nutritious media LB and TB to ensure availability of carbon sources and substrate precursors as well as the minimal medium M9 to reduce background signal. In addition, heterologous expression was performed as a time-course experiment over five days to monitor secondary metabolite production.

Whereas clones cultivated in M9 did not produce any compounds of interest (Figure 41A), many unique molecule signals were detected in the HPLC expression profiles of cultures grown in LB and TB. Because peaks were spread over an ACN gradient from 33 % to 68 %, it was difficult to identify potential PA derivatives. A selection of six promising signals were selected for further investigation (Figure 41B and C).

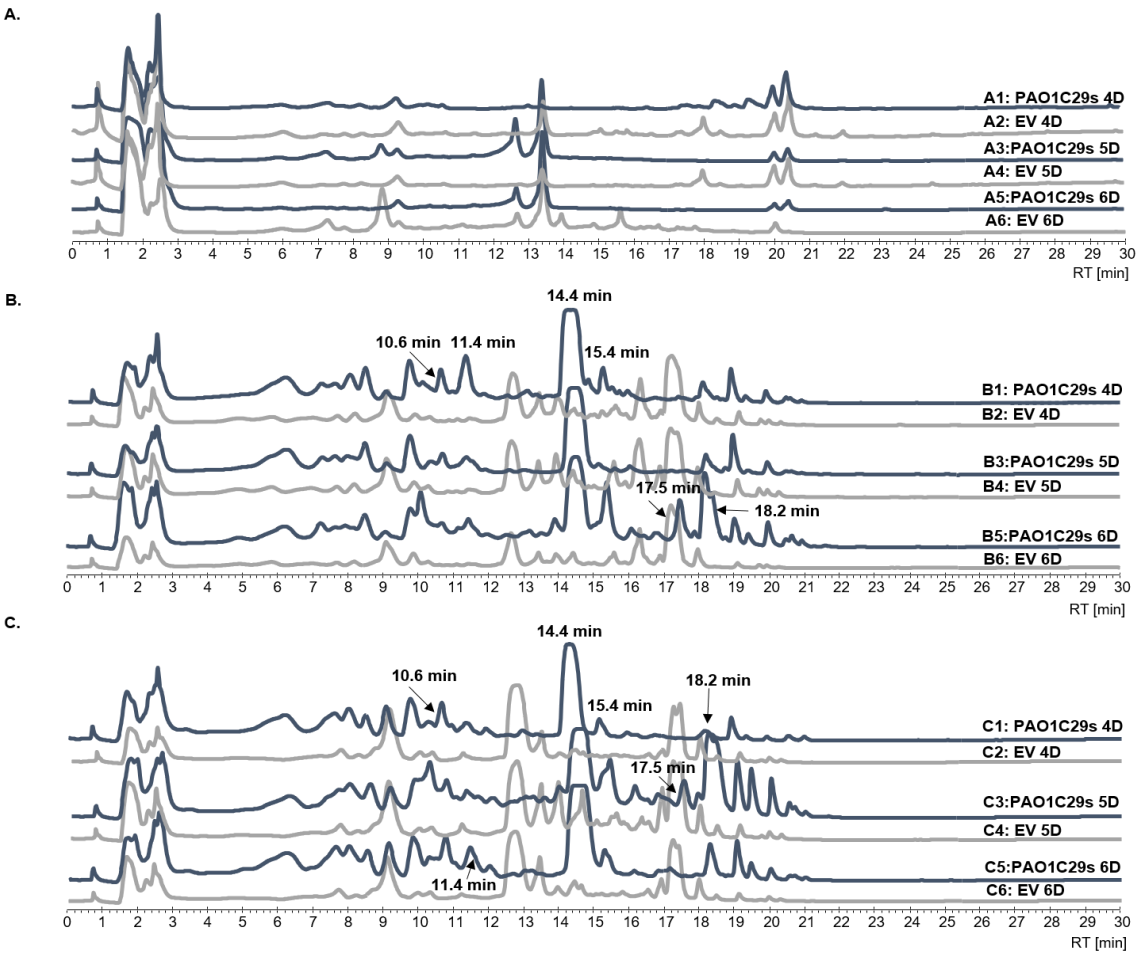


Figure 41. HPLC chromatograms of PAO1C29s time-course heterologous expression experiment in *E. coli* BAP1 cells. Grown bacteria were extracted after four, five and six days of incubation time at 20 °C. **A.** Supernatant fractions of clones cultivated in M9 minimal medium. **B.** and **C.** Expression of PAO1C29s-containing clones in nutrient-rich media LB and TB. Pellets of all cultures extracted separately. Only supernatant fractions are displayed.

After four days incubation time, first metabolites were detected by HPLC and hereafter both number and abundance of characteristic molecule peaks within the supernatant raw extract increased in accordance with the growth period. Interestingly, none of the detected peaks seen after four days disappeared with ongoing experimental duration as may have been expected when initial intermediate molecule formation is followed by transformation into the final product over time. It can thus be assumed that secondary metabolite formation was catalyzed within four days and products accumulate throughout the course of time. Distribution of novel compound signals is remarkably similar for cultures grown in LB and TB medium, therefore detected peaks were likely to be associated with the presence and expression of the PAO1C29s minimal gene cluster and less with media-specific background signals.

Two significant metabolite peaks were observed at retention times corresponding to molecules that had also been produced during PAO1C29s expression in *Streptomyces* (t_R = 10.6 and 11.4 min). Despite identical times of elution from the column, *m/z* ratios of the most prominent molecular adducts found in mass spectra of LB- or TB-grown cultures did not match with *Streptomyces* product ions (Figure 42). However, with an *m/z* ratio of 284.3, the ion detected at t_R = 10.6 min differed only by the mass of a proton, leading to the assumption that it might indeed be the same compound. At t_R = 11.4 min, two unique ion signals of *m/z* = 279.1 and 307.3 were detected, which could be linked to PA derivatives based on their molecular weight. Although a mass difference of 28 would indicate metabolites differing by two -CH_2 groups, it is unlikely that they were retained on the column equally long. Pyrrolizixenamides with identical ion masses bearing long fatty acid side chains (C_9 and C_{11}) elute at an ACN gradient of over 55 % compared to 38 % observed. Therefore, the produced compounds were probably of different PA origin. The most abundant metabolite produced by *E. coli* BAP1 PAO1C29s cultures was measured at 14.4 min retention time. With *m/z* = 159.1, the characteristic adduct of this fraction was decisively smaller than any published PA, including **49** (192.1 g/mol), making it rather uninteresting for identification of novel PA derivatives despite its high abundance. Presumably, the detected molecule is a shunt or degradation product of PAO1C29s metabolism. Analysis of ESI-ionized secondary metabolites eluted after 15.5 and 17.5 min showed further promising mass distributions within the range of PA derivatives. Like mass ions detected at t_R = 10.6 min, two equally abundant spectral peaks at 336.3 and 364.2 were seen after 18.2 min retention time, molecular ions differing again by two -CH_2 groups. Broadness of the elution peaks recorded by HPLC could possibly explain occurrence of two masses at once, indicating the presence of more than one signal hidden underneath a main peak. Alternatively, an altered ring configuration, for example a heptane instead of a pentane ring, with otherwise identical chemical structure may lead to simultaneous occurrence of two mass adducts at the same retention time (Figure 42).

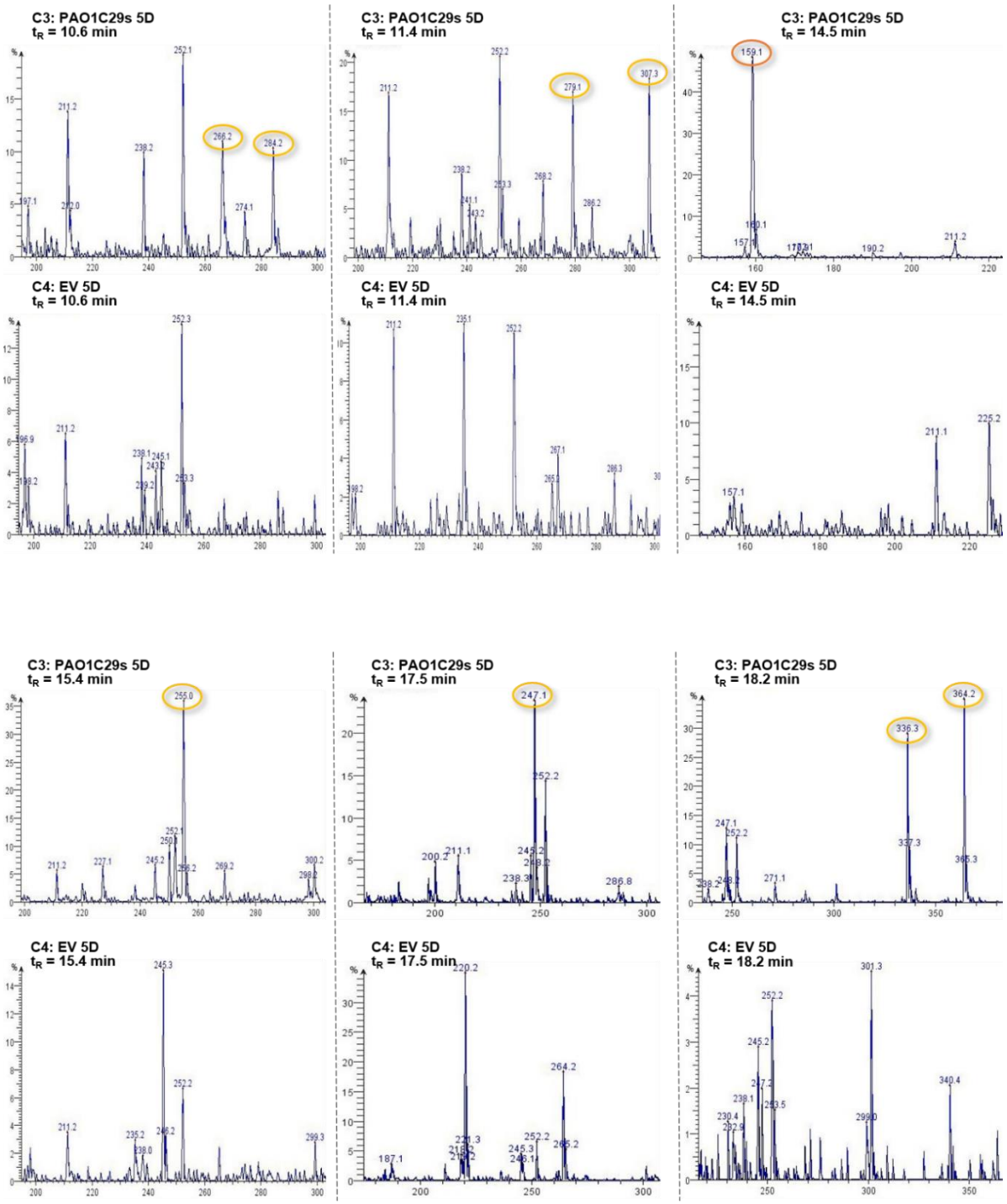


Figure 42. Mass analysis of selected secondary metabolite peaks of interest. Excerpts of most abundant ion adducts in comparison to the empty vector negative control shown within the *m/z* ratio range of known PAs (150 – 350 g/mol). All selected peaks were present after five and six days of incubation time in both LB and TB medium.

To gain further insight on the heterologous expression of BGCs involved in secondary metabolite biosynthesis, the described experiments must be repeated at bigger scale to isolate, purify, and further characterize any molecules of interest with additional analytical methods such as NMR.

Establishment of the azetidine ring system of **48** and **49** was studied by combining the minimal gene cluster with the SAM-dependent AzeJ enzyme in co-expression experiments and additional supplementation of L-methionine to the culture broth. Since all modifying BGC-specific enzymes were absent, biosynthesis of SAM from L-methionine and thus the outcome of the produced secondary metabolites was dependent of *E. coli* native proteins catalyzing the assembly of the precursor molecule needed for azetidine formation. As an alternative to standard extraction of supernatant and pellet fractions, designated bacterial cultures were treated with Amberlites XAD-16 after a 24 h growth period (Figure 43A). As can be seen in Figure 43, HPLC expression profiles of supernatants A1 and A2 exhibit additional peaks compared to the Amberlite resin-treated culture A3. Two distinct signals at t_R = 9.6 min and t_R = 10.9 min could be identified, which were present in all three samples, but absent in the negative control. Mass analysis showed a unique ion with *m/z* = 262.1 only for the supernatant fraction of A1, whereas spectra of A2 and A3 did not differ much from the empty vector although HPLC chromatograms displayed a less intense, but analogous peak to A1. The size of the mass adduct is within proximity of **52**, speculating that the molecule may be unsaturated**.** Interestingly, the most abundant molecule detected by HPLC and adduct observed in mass spectra at a retention time of 10.9 min showed an *m/z* ratio of 159.1, a mass that had previously been found when expressing the minimal gene cluster, however at a later retention time (t_R = 14.4 min; Figure 42). These findings support the hypothesis that the expressed molecule is rather a side product or intermediate of BGC 29 short expression and less the desired final natural product as it is produced both in the absence and presence of AzeJ. Peaks characteristic to the supernatant fractions (A1 and A2) did not feature any further metabolites of interest (Figure 43B). Furthermore, LC-MS data of minimal gene cluster expression in comparison to co-expression with the SAM-dependent enzyme revealed only minor differences that would support functionality of AzeJ and alter PA derivative outcome. However, if co-expression had failed and only the minimal gene cluster was functional, an identical HPLC expression profile, molecular ion distribution and relative abundance would have been expected. As the obtained cell pellets were green fluorescent, translation had been successful, ruling out the possibility of complete malfunction of gene expression. Nevertheless, an azetidine-like ring system could not be identified in the presence of AzeJ. In this case, *in vivo* biosynthesis of L-AZC most likely did not take place since the precursor SAM was not

supplemented but instead only the pre-precursor L-methionine was available. This extra catalytic step probably restricted NRPS-initiated PA derivative formation.

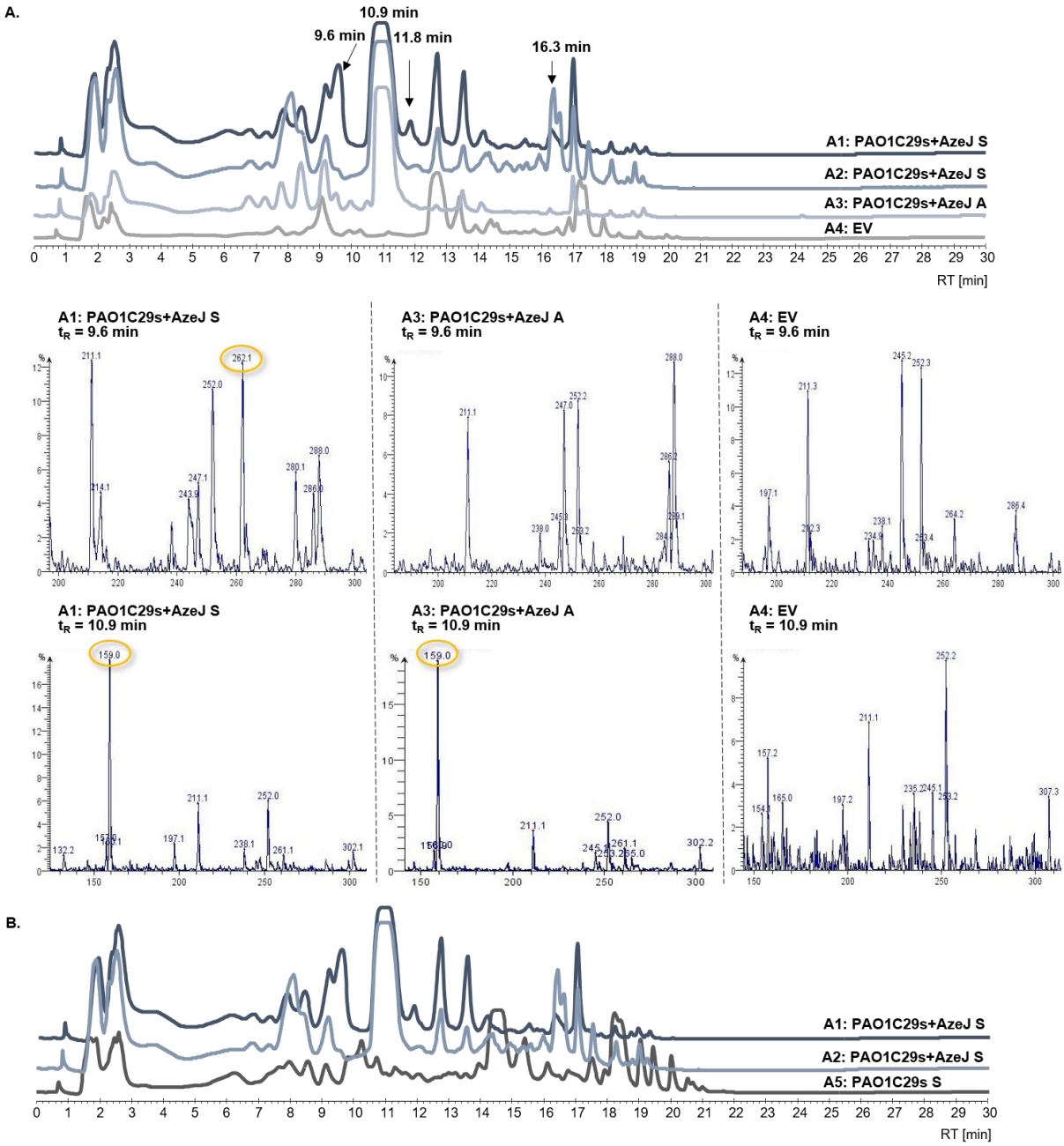


Figure 43. LC-MS data of PAO1C29 co-expression experiments. **A.** Extract preparation and compound isolation with or without Amberlite XAD-16 resin. **B.** Comparison of minimal BGC expression with AzeJ co-expression. S = extracted supernatant fraction; A = extracted Amberlite XAD-16 fraction.

All in all, characteristic peaks and masses detected by LC-MS hint at successful biosynthesis of natural products for both heterologous expression of the PAO1C29 minimal BGC in *E. coli* and *Streptomyces*. By identification of identical molecular adducts in the two expression systems, it can be assumed that BGC activation under the given circumstances was possible and that the produced metabolites are promising candidates for novel PA derivatives. The absence of **48** and **49** upon expression of the short cluster version alone or in co-expression with AzeJ further supports the necessity of specific precursors and synthesizing enzymes for a specific natural product outcome and supplementation with SAM seems to be obligatory.

***Micromonospora auratinigra* DSM44815**

As previously mentioned, cell pellets of *Micromonospora sp.* M71-S20 (DSM44815) were received from the ZIEL Institute for study purposes. In initial tests, the native gram-positive strain was cultivated in the laboratory using a variety of media types and under application of both solid and liquid media. This way, not only cultivability of the soil-dwelling bacterium was examined but also the intension of identifying functionalities of unknown genes encoded in the BGC of interest as well as the possibility of obtaining a reference for heterologous expression attempts was followed. For native strain cultivation, cell pellet was resuspended in TB buffer, plated on GYM agar plates, and left to grow for up to three weeks until bacteria spread to give a black bacterial lawn. Bacteria were then transferred to CASO medium for a seven-day liquid pre-cultivation phase. Grown cells were finally used to inoculate 200 mL main cultures containing either ISP-4, GYM or NZA broth and left to grow for further 8 days at 28 °C. Supernatant and pellet fractions were extracted separately and studied using LC-MS for putative secondary metabolite biosynthesis.

Native expression experiments revealed a tight linkage between nutrient availability and secondary metabolite production (Figure 44). Whereas no significant compound production could be detected in extracts of DSM44815 grown in the salt-rich ISP-4 medium, multiple unique molecule peaks were seen in expression profiles of cultures grown in nutrient-rich GYM and NZA media between 15 and 18 min of retention time (50 – 58 % ACN). Moreover, it seems as if different molecules were synthesized in the nutrient-rich media, since elution times and the number of molecules eluted vary. This shows that regulation of BGCs involved in secondary metabolite production is sensitive in DSM44815 and that slightest changes in nutrient availability and type alter the expression pattern.

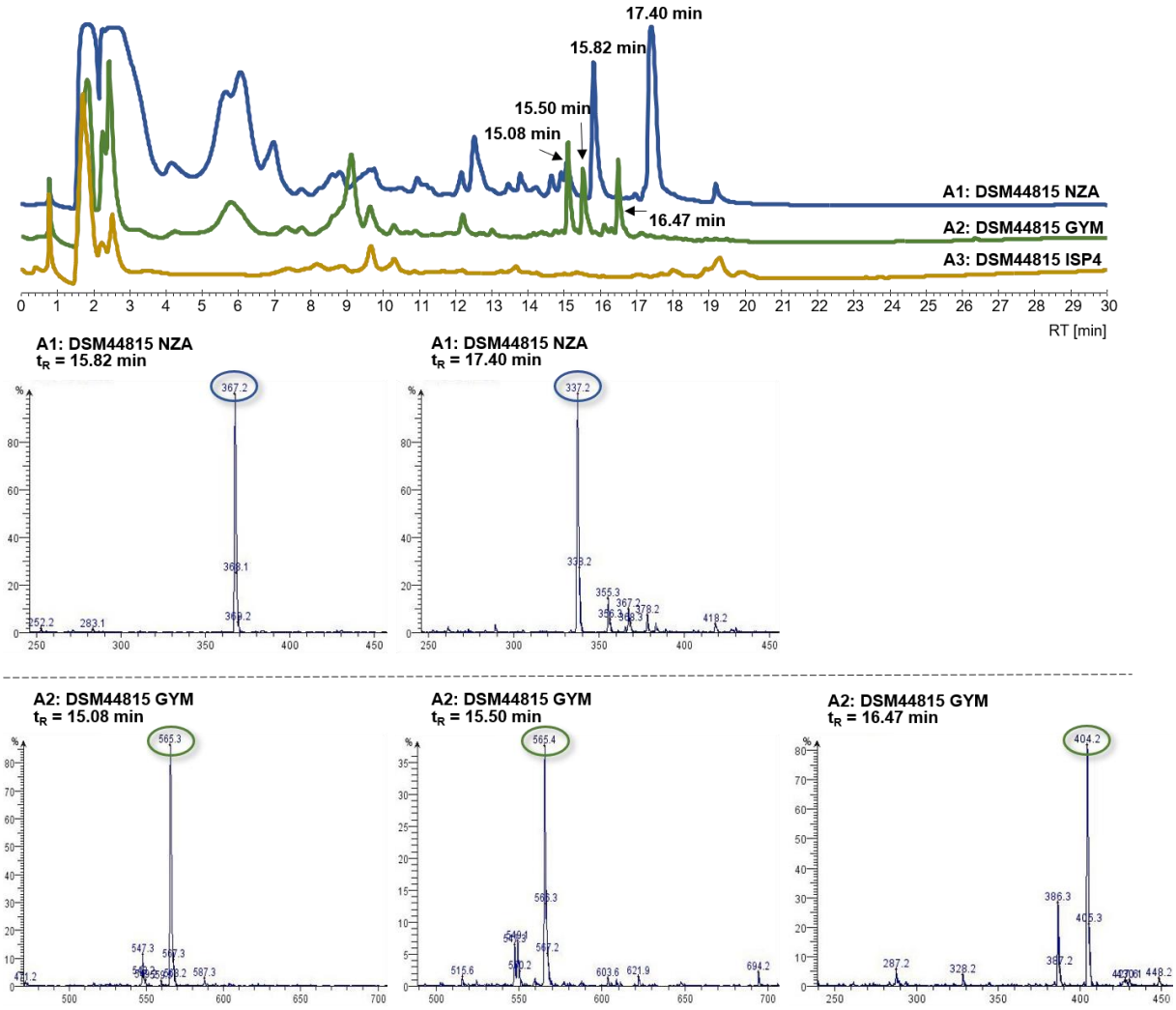


Figure 44. DSM44815 native strain expression in a selection of cultivation media. Mass spectra highlight most abundant molecular ions occurring in fractions of interest labeled in the corresponding HPLC chromatograms of GYM and NZA extracts.

Cultivation of *Micromonospora auratinigra* in NZA medium led to production of two prominent compounds with characteristic and highly abundant ion adducts at t_R = 15.82 min and t_R = 17.40 min. With *m/z* ratios of 367.2 and 337.2, both molecules carry masses comparatively high to published PA derivatives, thereby hinting at modification of the core structure catalyzed by additional BGC-encoded tailoring enzymes. Alternatively, the ion adducts can be associated with another class of metabolites other than PAs as BGC 21 is just one of many clusters encoded in the genome of DSM44815 and all have the potential of being expressed under the given laboratory conditions. This may also be an explanation for variations in the expression profiles of GYM- and NZA-grown cultures mentioned earlier. Three unique signals were detected for bacterial cells grown in GYM medium. Although eluted from the C_{18}-column at similar ACN-percentages as NZA-related compounds, ion masses were almost double as for

example variants of **37**. Again, extracted metabolites may be of different origin or it could be that isolated peaks represent dimeric analogues of common PA derivatives as has been reported for the dibohemamines[102] (**47**). All other peaks detected in HPLC chromatograms did not exhibit masses of particular interest. Selected compounds were isolated in amounts of 2 – 5 mg from the raw extracts using preparative HPLC (Figure S6) and subsequently characterized by NMR. ^{1}H-NMR analysis of the DSM44815 NZA compound eluted after 20.3 min during preparative HPLC indicated a structure similar to pyrrolizixenamides and was thus of particular interest, however the final product material was of insufficient quality for further 2D-analysis. Furthermore, reproducibility was not given upon repetition of the experiment.

MiC21

For heterologous expression of the full MiC21 cluster, *Streptomyces albus* deletion strains B2P1 and Del14 were applied as host strains in place of the commonly used *S. lividans* TK24 and *S. coelicolor* M1154 strains. Both deletion strains are mutants of *S. albus* J1074, with Del14 having all 15 intrinsic BGCs involved in secondary metabolism deleted to reduce production of any native secondary metabolites and thus enable detection of products of interest occurring only in minimal amounts. In addition, Del14 (= B1P1) and B2P1 exhibit one or more bacterial integration (B) and placebo integration (P) sites to support insertion of multiple copies of the pSET152ermE*rev-MiC21 plasmid into the *Streptomyces* host genome and therefore increase compound production levels.[131] Furthermore, the nutrient-rich cultivation broth YEME was tested as an alternative to ISP-4 to boost production levels. As previously shown for the native DSM44815 strain, expression can be sensitive to changes in media composition and lead to activation of different BGCs involved in various secondary metabolite production pathways. Heterologous expression was monitored for five and seven days to detect putative unstable products or intermediates.

Variation of the main culture growth period affected secondary metabolite production of Del14-MiC21 clones (Figure 45). Whereas the HPLC expression profile of the five-day extract showed no significant deviation from the negative control, four distinct signals were observed in the seven-day extract. The most abundant molecular ions present in the corresponding fractions possessed molecular weights from 250 – 300 g/mol, thereby being within the size range of PA derivatives. Notably, *m/z* ratios of molecules eluted after 13.2 and 16.4 min coincide with the $[M+H]^+$ masses of **52** (C_8) and a further derivative thereof bearing a C_{10} side chain moiety (**54**). Although the observed retention times deviate from the ones seen for **52** and **54** due to a variation in the duration of the HPLC method (30 min run time and 60 min run time, see Figure 29), the respective molecules elute at approximately the same % of ACN (42 – 44 % for **52**

and analogue, 51 – 53 % for **54** and analogue). This is a further indication that identical molecules were produced by different BGCs. Metabolites with retention times of 15.11 and 18.00 min were of identical molecular weight (263.3 g/mol). However, it is likely that two different compounds with altered polarity were biosynthesized and by coincidence share the same mass. Comparing to the native expression profiles, retention times of heterologously expressed compounds are very similar to natively expressed molecules in GYM and NZA media (15.11↔15.08 min, 16.35↔16.47 min and 18.00↔17.40 min) but molecular ion masses do not match at all. Thus, identification of putative PAs by native expression as a reference was not possible. Unfortunately, like heterologous expression attempts of other BGCs potentially involved in PA biosynthesis, observed results could not be reproduced by one-to-one repetition of the experiment and additional scale-up attempts, restricting further investigations on compound characterization.

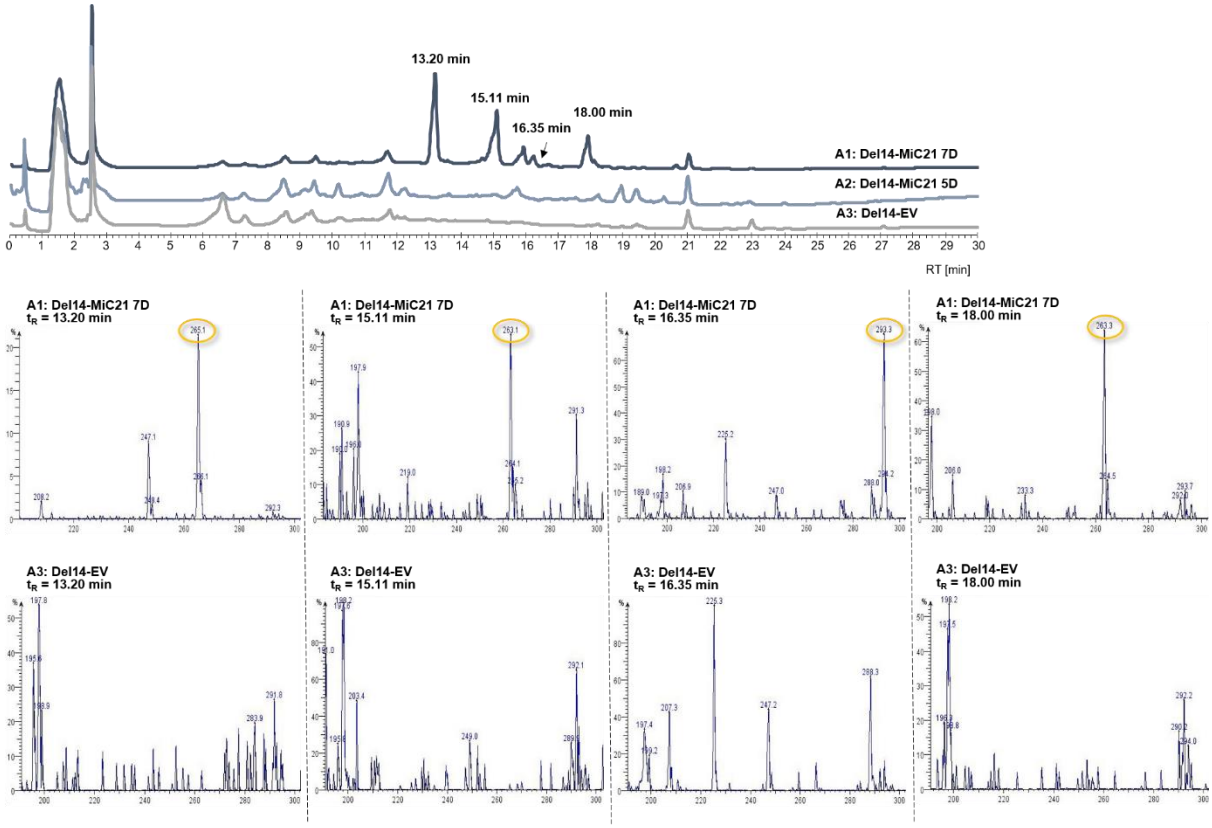


Figure 45. Studies on heterologous expression of MiC21 in *S. albus* deletion strain Del14. Bacterial cultures were incubated for five or seven days and left to grow in ISP-4 medium before extracting with ethyl acetate. Mass spectra show most abundant molecular ion adducts of designated HPLC signals.

Unlike heterologous expression in *S. albus* Del14 where background noise was minimal, as can be observed in the HPLC chromatograms of Figure 45, endogenous BGC secondary metabolite expression was still active in *S. albus* B2P1, which led to increased signal

occurrence. Simultaneously, identification of molecule peaks significant to MiC21 expression was made more difficult.

Interestingly, the number of unique peaks was maximal after a five-day growth period and signal intensity reduced with ongoing time, once again highlighting sensitivity of secondary metabolite production based on minimal changes of the expression experiment. Again, characteristic MiC21 BGC peaks were observed in the HPLC chromatograms, which are eluted from the column at retention times of 15 to 17 mins similar to the natively expressed metabolites (Figure 30). As previously observed for **50** – **54**, multiple compounds eluted consecutively at 50 – 58 % ACN. This could indicate the presence of derivatives which alter only slightly in their structure from one another, e.g., by a single atom. However, mass analysis of designated fractions did not reveal presence of molecular ions of interest worth following up on (Figure 46).

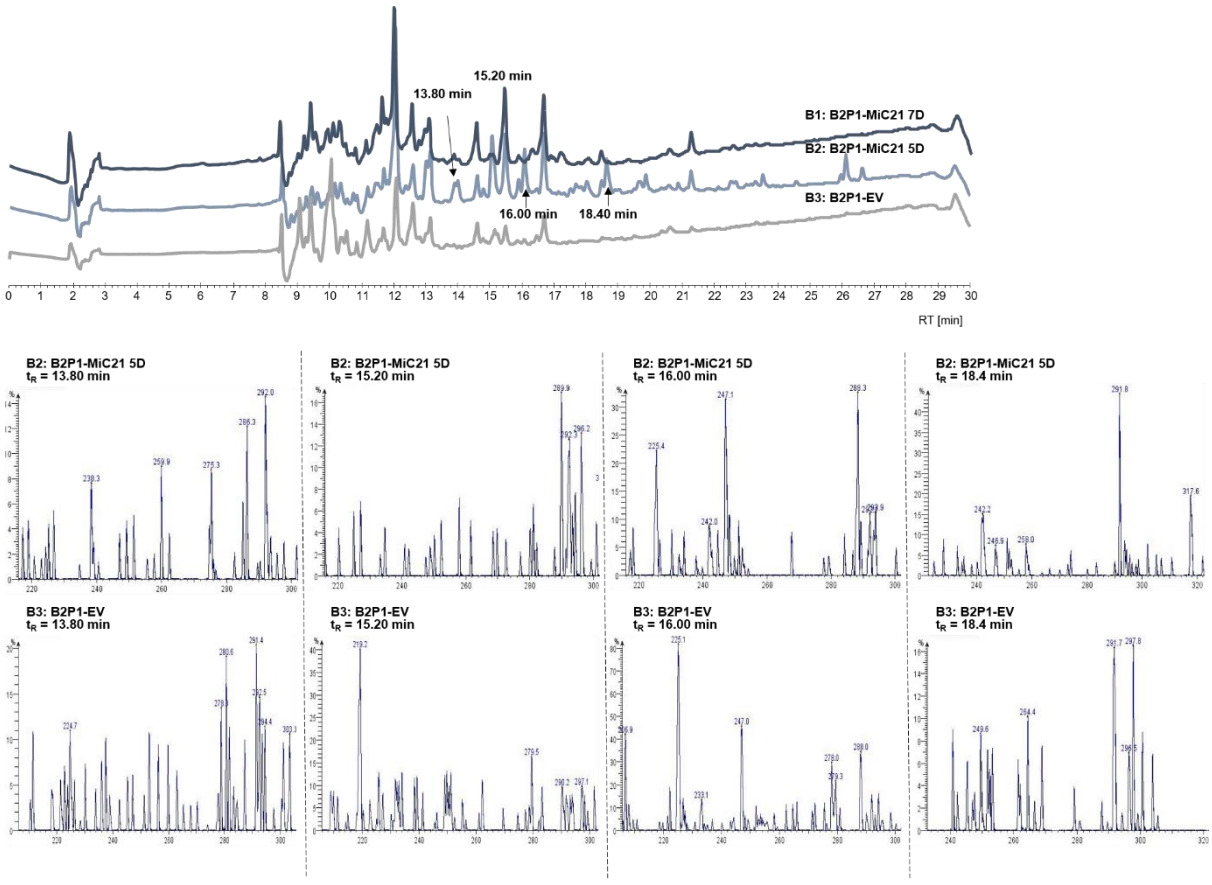


Figure 46. LC-MS analysis of MiC21 encoding *S. albus* B2P1 heterologous expression cultures. Comparison of characteristic peak fractions found in raw product extracted after five days with empty vector control.

Substitution of the salt-rich minimal medium ISP-4 with the nutrient-rich medium YEME did not enhance compound production of neither Del14-MiC21 nor B2P1-MiC21 bacteria as anticipated (Figure S7).

All in all, expression of native *Micromonospora auratinigra* DSM44815 in different types of culture broth promoted activation of secondary metabolism. The choice of medium thereby caused unintended control over the compound expression distribution. Experimental results underline sensitivity concerning regulation of secondary metabolite expression in bacteria and emphasize the difficulties occurring during heterologous expression. Although identified molecules may be associated with PA biosynthesis due to their measured molecular ion masses, it remains unclear whether they indeed result from expression of BGC 21 or are the result of biosynthetic activities catalyzed by enzymes encoded in other endogenous BGCs. A single molecule found in the extracts of bacteria cultivated in NZA-medium was characterized using NMR and indicated similarities to **37,** making in depth studying desirable in the future. Heterologous expression of cluster 21 in *S. albus* deletion strains aimed at optimizing conditions in a way that potential molecules of interest could be detected even more easily by enhancing chances of plasmid integration while reducing background noise at the same time. Presence of promising candidates was documented one time for *S. albus* Del14-MiC21, but further metabolite characterization was restricted due to lack of experimental reproducibility.

3.2.2.3 Mix & Match combination of BGC parts to generate novel PAs

The *E. coli* heterologous expression system was used as a basis for the establishment of a biomolecular toolbox to study PA derivative biosynthesis in more detail. Hereby, functional genes of different PA-producing BGCs encoding either the core NRPS system and MOX or enzymes involved in tailoring reactions were combined from two or multiple pathways to obtain a fully functional PA-producing BGC *in vivo*. Furthermore, the concept of mixing and matching single genes with homologues of related pathways permits studying the function of yet uncharacterized enzymes, their role in PA biosynthesis and contributes to the elucidation of single catalytic steps within the molecular mechanism of PA synthesis. Lastly, flexible insertion of genes into an existing, successfully expressing plasmid allows manipulation of the biosynthetic pathway in terms of generating novel PA derivatives.

3.2.2.3.1 MOX substrate specificity

One central question addressed in this thesis is the substrate specificity of the MOX-catalyzed Baeyer-Villiger oxidation reaction. In 2015, Schimming et al.[79] had proposed an initial NRPS-catalyzed reaction step to generate the [5,6]-bicyclic intermediate ring structure in PA biosynthesis. Based on the performed heterologous expression experiments and upon comparison of their results to those published by Schmidt et al.[97] on the biosynthetic pathway of lipocyclocarbamates, Shimming and coworkers hypothesized that establishment of this specific intermediate was independent of the acyl-CoA incorporated. Thus, an identical buildup of both **37** and **40** variants was assumed and directionality was predicted to be introduced at a later step in biosynthesis.

In consideration of these previous findings and hypotheses, we set out to investigate the mechanism driving the pathway-specific directionality of natural product outcome in more detail. Our working hypothesis addressed the MOX reaction, where the arrangement and complexity of the acyl side chain moiety is predicted by us to influence the directionality of the reaction cascade performed by the BVMO of the respective gene cluster. This would then in theory lead to the formation of either the lipocyclocarbamate (LCC) or the pyrrolizixenamide compound.

The NRPS enzyme PxaA originating in *Xhenorhabdus stockiae* DSM17904 was heterologously expressed in combination with MOX homologues of PxaB, amongst others from *Pseudomonas aeruginosa* PAO1, *Micromonospora auratinigra* DSM44815 and *Pseudomonas sp.* SH-C52, to study secondary metabolite production. Amino acids as well as a variety of fatty

acid precursors and sugars were supplemented to the cultivation broth to support biosynthetic processes and presumably direct product outcome.

Pyrrolizixenamide intermediate production (**63**-**67**) was achieved by expression of the standalone PxaA enzyme in both *E. coli* BAP1 and DH10β cells, thereby confirming successful priming of the NRPS PCP-domain upon co-expression with Sfp or MtaA PPTases.

Four [5,6]-bicyclic pathway intermediates differing only in chain length of the fatty acid moiety were identified in the raw extract of expression cultures treated with Amberlites XAD-16 (Figure 47), three of them exhibiting identical molecular ion *m/z* ratios to the $[M+H]^+$ molecular weights of published intermediate structures **63** - **65** (C_6 – C_8). With increasing hydrocarbon chain length, the molecules become less polar, causing a shift in retention time of around two minutes per additional CH_2 group. Apart from these shorter chained molecules, a less intense signal was detected at t_R = 21.75 min, displaying the adduct of a metabolite **67** bearing a C_{12} side chain residue (*m/z* = 349.14) as expected when supplementing with lauric acid (L). Since HPLC-recorded peak intensity stands in direct relation to the abundance of the specific compound within the raw extract, it was surprising to find the shortest chain intermediate in the highest amounts whereas the predicted long chain intermediate was present in comparatively low amounts. Most likely, lauric acid is processed by endogenous fatty acid metabolism, thereby creating truncated acyl-CoA units which are subsequently used by the NRPS system as precursors and coupled to amino acids L-serine and L-proline to generate the shorter-chained PA intermediates.

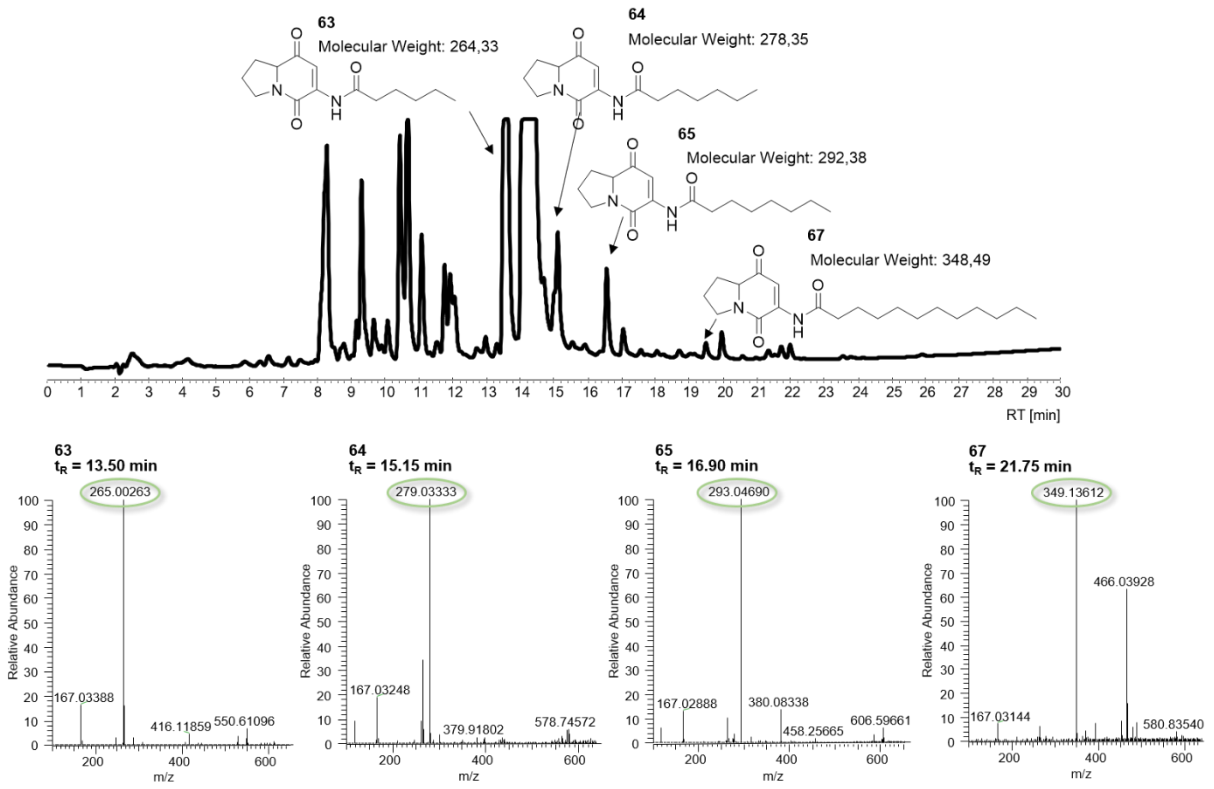


Figure 47. LC-MS monitoring of common PA intermediate production by heterologous expression of an NRPS system under the control of an L-arabinose inducible promoter in *E. coli* DH10β host strain. Detected mass adducts are in accordance with molecular weights of published intermediate structures.

As mentioned earlier, the biosynthetic pathway of **40** deviates from biosynthesis of **37** most likely due to the character of the acyl side chain. The NRPS system involved in brabantamide biosynthesis incorporates an activated 3-hydroxy fatty acid into the molecular skeleton where a rhamnose sugar moiety is attached to the hydroxyl group within the course of compound establishment. Although it is yet to be validated whether coupling of the sugar to the fatty acid moiety takes place prior to peptide synthesis or as a final tailoring modification, an incorporation prior to peptide bond establishment was assumed by Schmidt et al., thereby supporting the hypothesis of a substrate-controlled mechanism.[97] One experimental setup therefore set out to produce intermediates carrying a β-hydroxy fatty acid residue to mimic the fatty acid type used for biosynthesis of **40**. For this, (*R*)-3-hydroxymyristic (M) and (*R*)-3-hydroxypalmitic acids (P) were supplemented to the medium at concentrations of 100 mg/L. HPLC expression profiles of *E. coli* BAP1 PxaA-pCX2 cultures incubated with either M or P showed production of two unique secondary metabolites eluting from the C_{18}-column after 13.50 and 15.45 min that were absent in the empty vector (Figure 48).

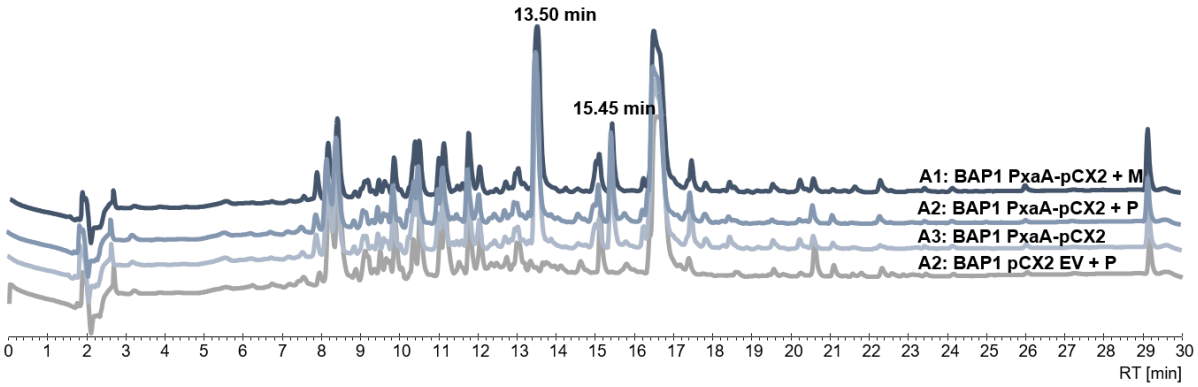


Figure 48. Introduction of alternative acyl side chains into the backbone structure of pathway intermediates by supplementing culture broth with 3-hydroxy fatty acid precursors. M or P was added at 100 mg/L alongside standard amino acids.

Upon comparison with a culture supplemented with only the amino acids L-serine and L-proline but lacking any kind of fatty acid precursor component, chromatograms showed identical peak distribution. Furthermore, secondary metabolite retention times were in line with those of intermediates **63** and **64** identified earlier (see Figure 47), leading to the conclusion that feeding of β-hydroxy fatty acids did not influence fermentation outcome. Starting from M or P, pathway intermediates were predicted to be of sizes larger than 300 g/mol, molecules which would have been expected to elute at a higher ACN percentages despite the additional hydroxyl group making them more polar than without. These findings also support speculations that fatty acids supplemented to the medium are not taken up by the *E. coli* cell to be used for PA generation, but instead only endogenous acyl-CoA precursors are incorporated. This in turn may be the reason why short chain intermediates were produced at much higher titers than long chain ones. Solutions for enabling fatty acid uptake by increasing cell membrane permeability were tested by supplementing the cultivation broth with selected additives.

Numerous agents have been found that disturb the lipopolysaccharide (LPS) layer of the outer membrane and thus make it more permeable and susceptible, for example towards antibiotics. Amongst many others, divalent cations such as Ca^{2+} or Mg^{2+}, sugars as well as chelators of all kinds are capable of increasing membrane permeability and therefore can be used as effective additives in heterologous expression experiments.[140-141] To facilitate uptake of supplemented fatty acid precursors needed for PA biosynthesis, *E. coli* BAP1 PxaA-pCX2 clones were cultivated in TB and LB medium containing either 0.5 mM $CaCl_2$ or 0.5 mM sucrose or a combination of 0.5 mM each $CaCl_2$, sucrose and EDTA. Evidence of long chain **37** derivative production was used as a measure for fatty acid uptake.

In the presence of $CaCl_2$, biosynthesis of two additional secondary metabolites at t_R = 28.15 min and 28.60 min on top of the previously identified pathway intermediates **50 - 52** were observed, with one of them exhibiting an abundant molecular ion of *m/z* = 349.4 corresponding to intermediate **66** (Figure 49). Although not as obvious as seen in the HPLC chromatograms of samples A1 and B1, test cultures grown with sucrose or with a combination of salt, sugar and EDTA chelator also produced an identical compound. Comparison of relative peak intensities with the expression profile of cultures grown without additives (see Figure 48) revealed an increased titer of the long chain derivative, hinting at a higher availability of lauric acid precursor inside the cell. Enhanced amounts of lauric acid in turn can be associated with successful permeation of the outer membrane and support efficacy of the selected agents.

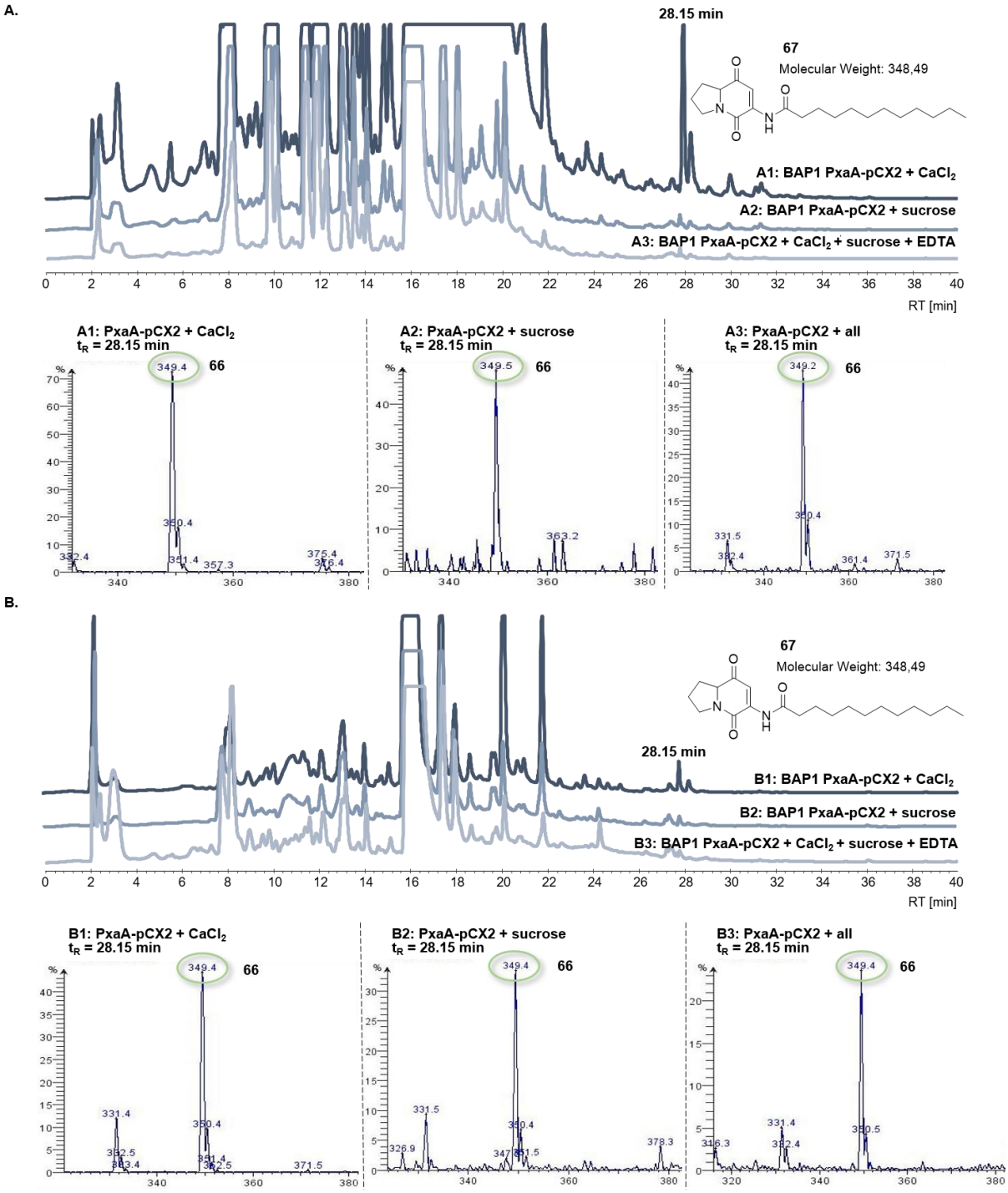
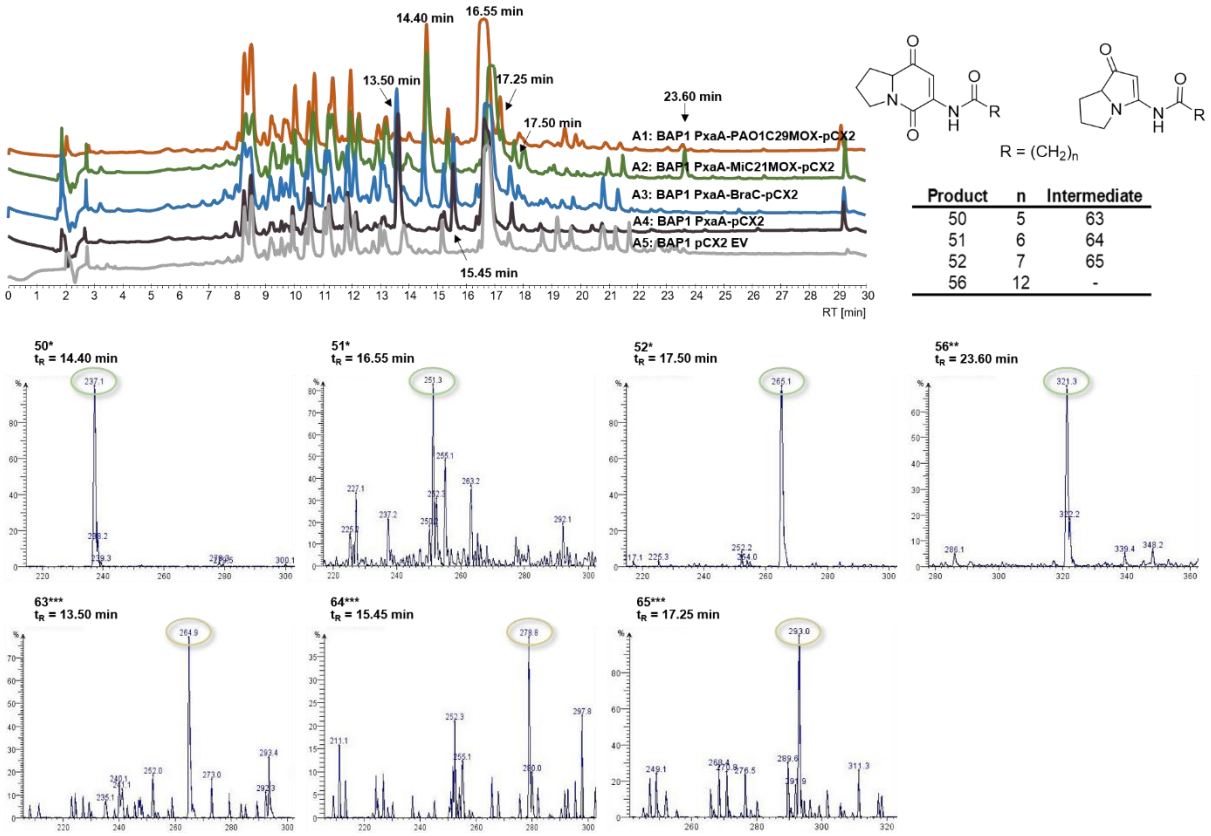


Figure 49. Increasing cell membrane permeability for precursor uptake by addition of permeability-enhancing agents. Salts such as $CaCl_2$ provide polycations, sugars such as sucrose alter transport across the membrane and EDTA functions as a chelator of divalent cations, leading to LPS release from the cells' outer membrane and barrier disruption. Expression experiments were performed in nutrient-rich media TB (**A**) and LB (**B**).

Heterologous expression of the minimal BGC thought to be necessary for PA biosynthesis was achieved by insertion of the respective MOX homologue downstream of PxaA (Figure 50). PAO1C29 MOX and MiC21 MOX were tested for production of **37** whereas BraC from *Pseudomonas sp.* SH-C52 was inserted to validate biosynthesis of **40** and to examine divergence of the reaction into the respective biosynthetic pathways.

To identify products of secondary metabolism more easily, all cultivated *E. coli* BAP1 clones were treated with Amberlite XAD-16 resin 18 h prior to compound extraction. HPLC expression profiles of all tested PxaB MOX homologues are displayed in Figure 50, revealing molecular compositions of raw extracts which differ from the empty vector negative control. Specific expression of the PA-producing BGC led to biosynthesis of multiple secondary metabolites that were eluted from the column starting at an ACN-gradient of 44 %. As a result of inserting MOX genes from *P. aeruginosa* and *M. auratinigra* downstream of the NRPS system, $[M+H]^+$ mass adducts of 237.1, 251.3 and 265.1 identical to **50, 51** and **52** were found. Preparative HPLC for compound isolation followed by NMR analysis confirmed molecule identity. Furthermore, an additional variant of **37** bearing an acyl side chain with 12 carbon atoms was detected at t_R = 23.60 min and most likely corresponds to **56**. Retention times of final product molecules shifted around 0.5 – 1 min towards higher values compared to PA intermediates which can be explained by core structures becoming less polar due to the loss of a keto group and thus elution from the column at higher ACN-percentages. **50** - **52** were also identified as expression products of BraC (involved in brabantamide biosynthesis) together with PxaA whereas no mass evidence of either **40** or derivatives thereof lacking the sugar moiety were seen. These findings support the idea of a substrate-controlled MOX enzyme reaction cascade. Interestingly, in addition to pyrrolizixenamide molecules, metabolites **63** and **64** were present in the raw extract of the BraC-expressing clone as well. It remains unclear why a certain proportion of intermediate was not converted to the final PA derivative when simultaneous natural product formation clearly highlights BraC catalytic activity. One possibility could be that expressed BraC exhibited a significantly lower substrate turnover number, leading to accumulation of intermediate before it can be further transformed into the final compound.



Product	n	Intermediate
50	5	63
51	6	64
52	7	65
56	12	-

Figure 50. Analysis of secondary metabolite production of *E. coli* cultures expressing MOX homologues of different native bacterial origin. Comparison of HPLC minimal BGC expression profiles A1, A2 and A3 with NRPS only (A4) and empty vector expression (A5). *Molecular ions detected in mass spectra of all three different MOX homologues. **Adducts detected only in mass spectrum of A2 clone expressing MiC21 MOX. ***Pathway intermediates found in LC-MS data of A3 and A4.

To further study directionality of the BraC-catalyzed reaction cascade depending on the type of substrate available, expression tests were performed which included media supplementation with diverse fatty acids (lauric acid, octanoic acid, (*R*)-3-hydroxypalmitic acid) and sugars (D-glucose, L-rhamnose) alongside standard amino acids and $CaCl_2$ for increased cell permeability.

Structural features of [5,6]-bicyclic PA intermediates and product molecules bearing a pyrrolizidine core result in moderate polarity and thus cause secondary metabolites to be eluted from a C_{18} reversed phase column starting at ACN-percentages of 44 % (t_R = 13.0 min). Independent of the additive combination available during heterologous expression, all grown bacterial cultures showed identical distribution of produced secondary metabolites extracted from Amberlite XAD-16 resin (Figure 51). Furthermore, comparison of obtained chromatograms to profiles recorded for samples A6 and A7 displayed no unique metabolite signals that would have indicated an additive-induced alteration of compound biosynthesis. As

previously observed, combination of the NRPS system PxaA with the BraC MOX leads to simultaneous expression and occurrence of pathway intermediates and products. Mass analysis of **63** (t_R= 13.50 min) not only showed the predicted $[M+H]^+$ mass of 265.1 but also an abundant ion adduct with an *m/z* ratio of 247.1. Since both molecular ions differ by a molecular weight of 18 g/mol corresponding to a molecule of water, dehydration of the product molecule may have occurred. Preparative HPLC and NMR analysis confirmed the identity of the target compound to be identical to **63** (Figure S8). Despite enhancing outer membrane permeability and adjusted supplementation of appropriate building blocks, the reaction mechanism of BraC could thus not be steered towards biosynthesis of variants of **40**.

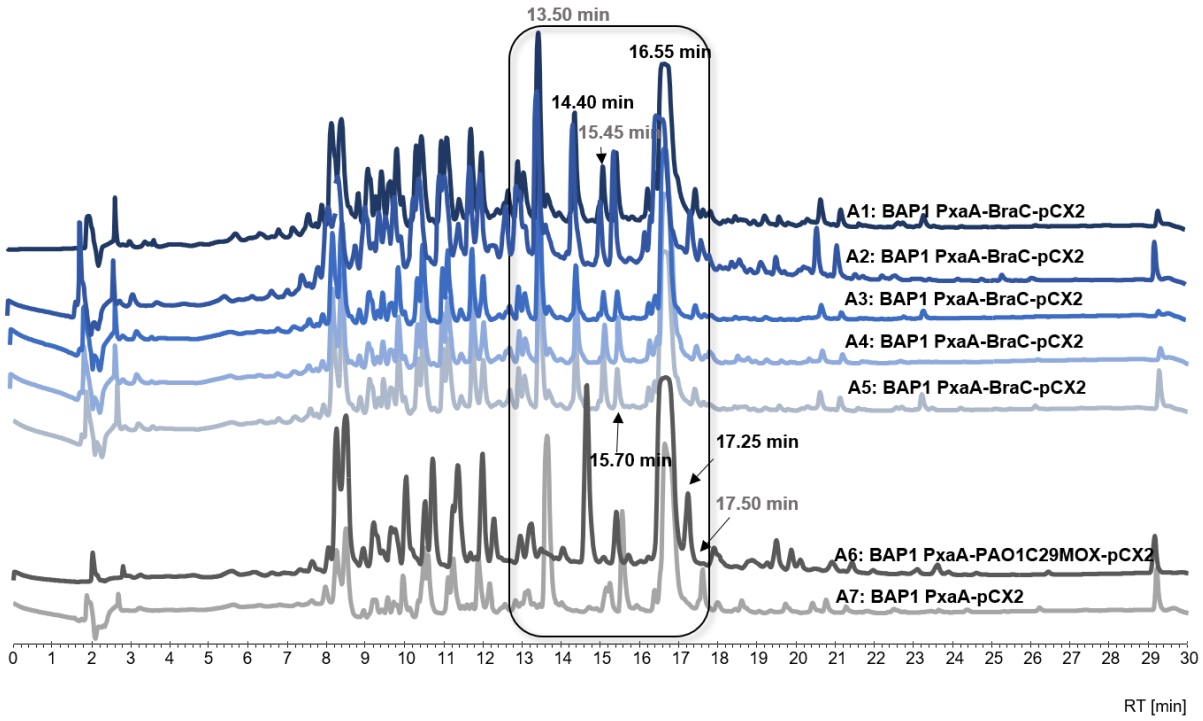


Figure 51. HPLC data of extracts from *E. coli* PxaA-BraC-pCX2 test expression cultures using various precursors (A1-A5). Comparison with pathway intermediate- (A6) and pyrrolizixenamide (A7)-producing clones. Apart from amino acids L-serine and L-proline, supplements added include A1: lauric acid; A2: (*R*)-3-hydroxypalmitic acid; A3: lauric acid and D-glucose; A4: lauric acid and L-rhamnose; A5: octanoic acid and L-rhamnose. Retention times of PA intermediate molecule signals highlighted in grey, pyrrolizixenamide natural products labeled in black writing.

By combination of functional genes involved in PA assembly, derivatives of **37** were obtained in amounts high enough for identification and characterization, independent of the MOX homologue that had been inserted and expressed. The expression tests showed that all selected BVMOs, including BraC of the related brabantamide pathway, share a common mechanism of action and therefore underline the potential role of the substrate type concerning

product outcome. Accessibility to modified basic building blocks, which may direct the reaction towards one specific natural product, remained restricted, thereby limiting the number of novel putative compounds being found.

3.2.2.3.2 Generation of novel PA derivatives

By means of heterologous expression of a standalone functional PxaA NRPS system or by combination with a functional BVMO, biosynthesis of literature-known PAs **50** – **52** was achieved. In addition, longer chained derivatives thereof were biosynthesized (**53** – **56**) which have, to our knowledge, not been found in any other native or heterologous expression attempt so far. In line with these findings, the corresponding pathway intermediates (**63** – **67**) were identified and isolated from *E. coli*. Moreover, an exchange of the MOX gene with homologues originating in a variety of bacterial strains demonstrated a common catalytic reaction mechanism. The established mix & match system was further used with the intension of generating novel PA derivatives. Genes encoding for modifying enzymes were incorporated and expressed to study changes made to the PA backbone. Since intermediates had successfully been transformed into pyrrolizixenamides by catalytically active MiC21 MOX, attempts were made to express additional enzymes of the same BGC. This specific BGC was of particular interest as the identities of some genes had not been solved. Therefore, finding of novel PA derivatives could yield information concerning enzyme function. Three variations of BGC composition were tested and target genes were inserted downstream of the BVMO.

Apart from the known identity of the core genes involved in PA biosynthesis, only one further gene function was annotated for *Micromonospora auratinigra* DSM44815 cluster 21, namely a short chain dehydrogenase (SCDH). Although SCDHs are a large and versatile group of enzymes, most catalyze proton transfer reactions in the presence of a co-factor such as NADPH.[142] In the context of PA compound modification, they are predicted to be involved in a reduction reaction of the pyrrolizidine core structure ketone to an alcohol. This minor modification most likely leads to a minimal shift in peak retention time as the molecule becomes more polar. However, detection of this shift could be difficult to identify in the HPLC chromatogram. The SCDH-encoding gene was inserted downstream of the MOX homologue of the same biosynthetic pathway, heterologously expressed and indicators of successful modification and SCDH activity were studied based on mass spectra analysis.

Next to a single gene insertion, two versions of the MiC21 BGC were to be expressed, one including the SCDH as well as the two directly adjacent hypothetical proteins flanking the

SCDH (short), the other variant bearing the entire BGC as a whole except for the initial gene encoding for the NRPS (long).

Despite heterologous test expressions of multiple clones and supplementation with membrane-permeabilizing agents, production of novel PA derivatives was not detected for either of the three BGC variations (Figure 52). Furthermore, for cases A1 – A3 and B1 – B3 displaying HPLC profiles of SCDH insertion as well as the short BGC variant, even characteristic signals corresponding to **50** - **52** as can be seen in sample A4 were not present. These findings indicate that none of the encoded enzymes were expressed and the overall experiment failed, as the production of the PA backbone would have been expected even in cases of malfunction of any tailoring enzymes. The overly broad signal (t_R = 21 – 23 min) detected for all experiments supplemented with Triton X-100 can be associated with the hydrophobic detergent. Extracts of the full-length BGC (C1 and C2) showed traces of PA intermediates **64** and **65** as well as PA products **50** and **52** and thus support failure of the other expression variations, where traces of any PA-related molecule were absent. According to secondary metabolite signal distribution, downstream insertion of all encoded BGC genes did not alter molecule composition of the extract or lead to the identification of novel peaks which could be linked to PA core structure modification. In summary, neither single gene nor multiple gene incorporation into the established pyrrolizixenamide-producing plasmid led to generation of novel PA derivatives as expected. Unsuccessful expression of tailoring enzyme may be a result of insufficient translation caused by limited promoter strength, resulting in the absence of new and interesting compounds.

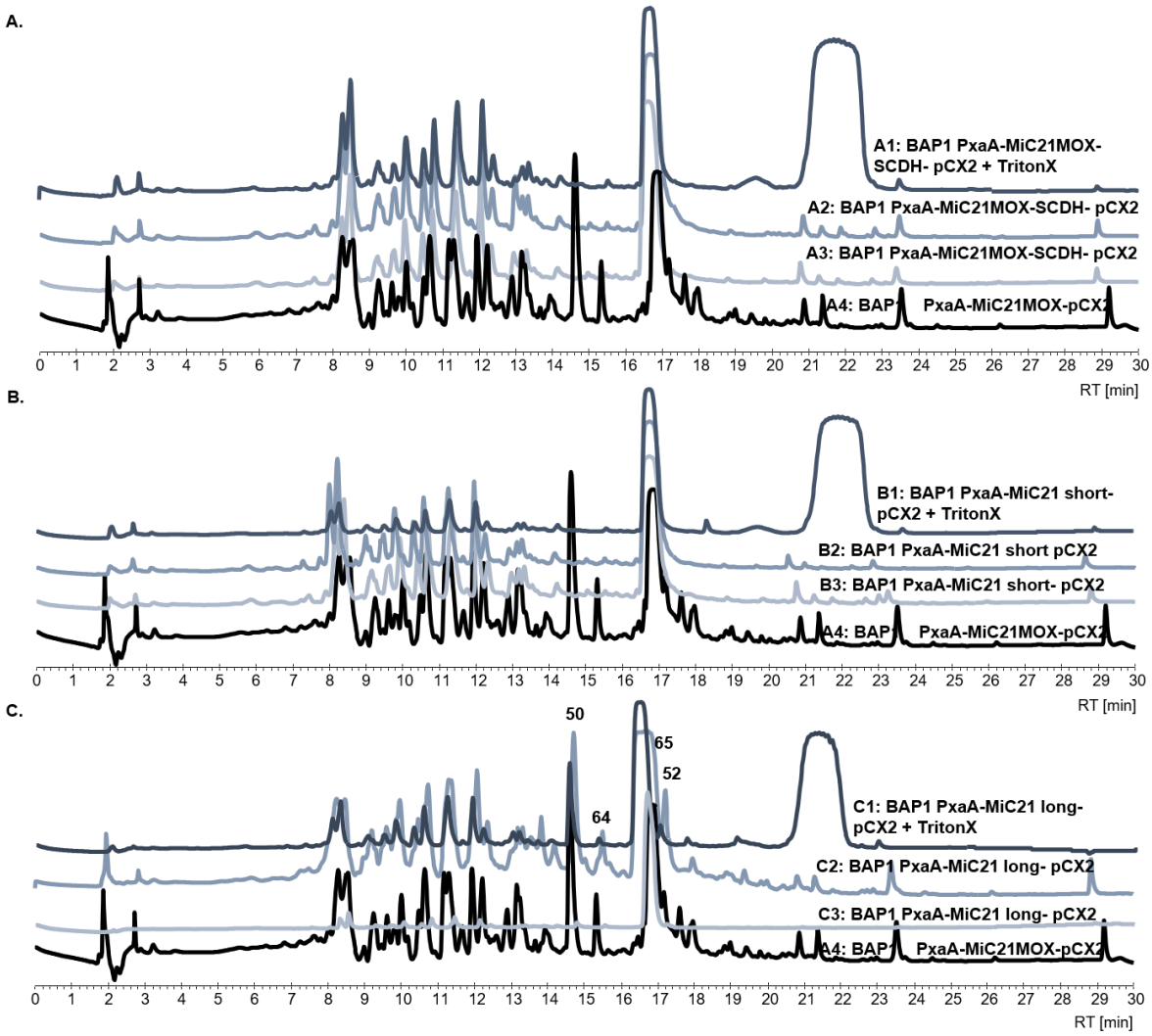


Figure 52. Experiments addressing the biosynthesis of novel PA derivatives by heterologous expression of additional tailoring enzymes. All tests were performed in the presence of 0.1 M $CaCl_2$, one test for each cluster variant additionally contained 0.1 % (v/v) Triton X-100 to further enhance cell membrane permeability. **A.** Test expressions performed to study modifying reaction catalyzed by SCDH. **B.** Heterologous expression of enzymes with unknown function alongside SCDH and PA core enzymes. **C.** Expression of full-length MiC21 BGC lacking strain-specific NRPS system.

To overcome difficulties resulting from limited promoter strength, a second arabinose inducible araBAD promoter system was incorporated downstream of the MOX homologue for subsequent independent control of modifying enzymes. Induction with L-arabinose allows simultaneous gene expression of respective BGC parts. Again, working with MiC21, all BGC-encoded genes located downstream of MiC21 MOX were put under the control of the second promoter, with the generated expression plasmid ultimately bearing all genes also expressed in the long version of the single promoter construct. Furthermore, the mix & match toolbox was used to combine genes of three different clusters by inserting the back part of PAO1 cluster

29 known to be involved in biosynthesis of **48** and **49** downstream of the *X. stockiae* DSM17904-derived NRPS PxaA and the MOX gene originating in *M. auratinigra* DSM44815.

One distinct compound signal was detected by HPLC at t_R = 11.2 min in the extract of test A1 that was not found in any other extract. With *m/z* = 235.2 the molecular ion possesses a weight within the size range of PA derivatives. In contrast to expression of the core system only, pyrrolizixenamides **50** - **52** were not identified in noticeable abundances during any time point, suggesting divergent secondary metabolite production. A mass adduct of 235.2 was detected a second time at a retention time of 14.89 min, letting one assume that metabolite molecules by chance share a similar if not identical molecular weight but exhibit different chemical properties such as polarity, causing one compound to be retained on the column 4 min longer than the other (Figure 53). Assigning the generated molecular ions as unspecific background signal could be eliminated as this specific mass-to-charge ratio was neither detected at other runtimes nor was it found in the LC-MS data of test expressions A2 to A5. Thus, closer characterization of the unique molecule seemed to be of interest.

Whereas heterologous expression of *Micromonospora* cluster 21 showed enhanced production of more than one secondary metabolite, the HPLC expression profile of clones possessing enzymes involved in PA-biosynthesis of three different native bacterial strains were comparatively empty. Independent of the presence of the second araBAD promoter system (B1 and B2), extracts of *E. coli* BAP1 cells carrying the core enzymes as well as the PAO1C29 back part exhibited an intense metabolite signal at t_R = 13.0 min, which can be found in all expression profiles including the empty vector negative control. Peak distribution for both tests B1 and B2 was identical to the empty vector, hinting at unsuccessful translation of all enzymes encoded on the plasmid. Experiments must be repeated for clarification purposes and to gain information for future test series.

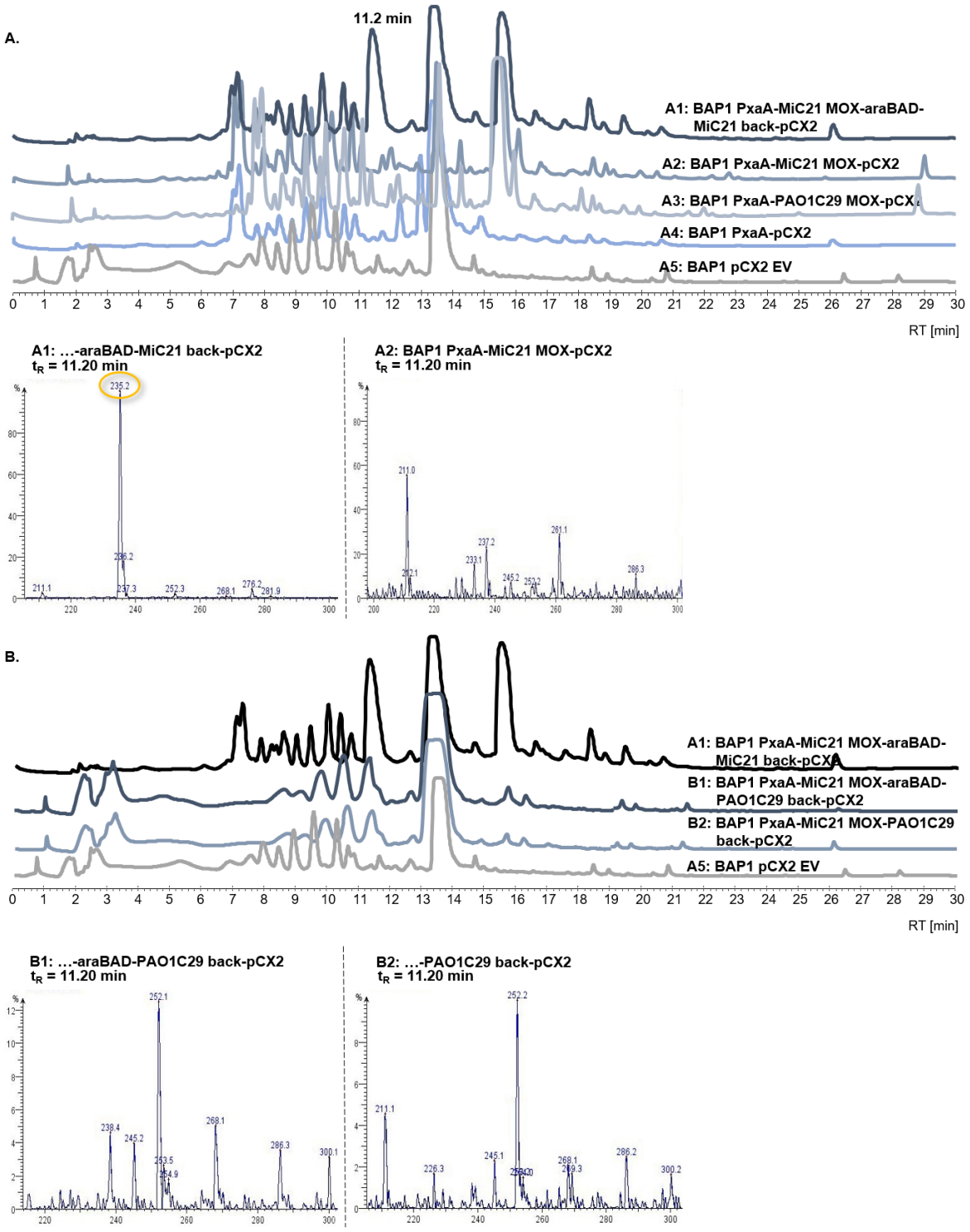


Figure 53. LC-MS data displaying heterologous expression of mixed BGC parts from multiple native bacterial strains using a double promoter expression system to generate novel PA derivatives. **A.** Chromatogram of *E. coli* extract holding full-length MiC21 BGC in comparison with shorter BGC variants. Presence of unique compound peak at t_R = 11.2 min verified by mass analysis. **B.** Mix and match of BGC parts from three different strains, including PAO1C29 back part active in **48** and **49** biosynthesis in single and double promoter expression plasmid.

Taken together, the mix and match biomolecular toolbox was utilized to combine cluster pieces and single genes and to build a functional construct. Conservation of MOX homologues was proven upon pyrrolizixenamide biosynthesis where even the distantly related BraC showed high substrate promiscuity. Control of reaction product outcome, however, could not be influenced by feeding of different building blocks. Based on successful expression of the minimal BGC resulting in production of **50** - **52**, tests using the mix and match vector system were conducted to address expression of additional modifying enzymes leading to putative novel PA derivatives. Insertion of either a single gene for functional studies on the specific enzyme or multiple genes of a respective BGC for monitoring overall BGC final product outcome was performed in single or multi-promoter systems. When putting all genes of interest under the control of only one araBAD promoter, no significant changes in secondary metabolite distribution were observed and for most experiments performed, overall protein translation was most likely abolished by insertion of additional genes. Establishment of a double promoter system where all genes encoding tailoring enzymes were put under the control of second araBAD promoter led to an alteration in the secondary metabolite profile of an extract containing the back part of MiC21. With the mass-to-charge ratios of the most abundant molecular ion in hand, further characterization of the compound of interest is a work in progress (Figures S9).

3.3 Reconstitution of PA total biosynthesis *in vitro*

3.3.1 Biosynthetic pathway studies

A central aspect addressed within this thesis was the catalytic mechanism of PA biosynthesis. To gain insight into the assembly line, we aimed at dissecting the biosynthetic pathway and performing step-by-step reconstitution of the unique reactions taking place, with the focus being set to the construction of the PA core molecule. It is proposed in literature that an NRPS enzyme initiates PA establishment by coupling of amino acids L-serine and L-proline with a ketoacyl thioester followed by action of an unusual TE-domain which permits cyclization and subsequent formation of a heterocyclic intermediate. The second enzyme involved in catalysis is a BVMO which converts the intermediate by a sequence of reactions, including Baeyer-Villiger oxidation, hydrolysis, decarboxylation, and ring contraction, to obtain the final PA core structure (see chapter 1.3.3).[79] To date, little is known about the MOXs exact mechanism of action, especially concerning substrate specificity. We therefore aspired to express both NRPS systems and MOX enzymes from selected bacteria in high quantity and quality, thereby retaining catalytic activity of the enzymes for *in vitro* testing as well as protein characterization purposes.

The use of *E. coli* as host organisms has proven to be the method of choice concerning the expression of recombinant proteins, with key qualities such as rapid growth, high density cultures, fast and easy transformation of exogenous plasmid DNA as well as easy manipulation, making protein expression highly efficient. Various types of *E. coli* vector systems have been developed over the years, which exhibit different feature combinations (promoter, origin of replication, affinity tags, selection markers etc.), making this well-studied organism suitable for almost any need.[143]

Multiple vector systems were tested during this project to not only make recombinant NRPS and MOX protein expression possible, but also to optimize expression conditions.

3.3.2 Cloning of NRPS systems for recombinant protein expression

In general, vector systems pET28b(+)-SUMO and pHis8C(+)-TEV were chosen for the expression of NRPS systems involved in PA-biosynthesis. When inserting the nucleotide sequence of choice at the multiple cloning site of the pET28b(+)-SUMO vector via the BamHI restriction site, the protein is designated to carry an N-terminal His_6 peptide tag in combination with a SUMO fusion tag. The His_6 tag was used for protein purification using Ni^{2+}NTA whereas

the SUMO tag intended to function as a stabilizing agent for the C-terminally linked expression protein, thereby enhancing solubility. Furthermore, this vector system enables high protein production levels. As an alternative expression system, the pET-based vector pHis8C(+)-TEV exhibiting a His_8 affinity tag at the C-terminal end of the inserted NRPS gene sequence was used, which is proposed to support proper folding and tertiary structure formation at high production levels. In cases where peptide tag and fusion partners were in N-terminal position to the gene of interest, in-frame cloning was ensured by single or multiple base pair insertion upstream of the 5' end of the target gene. An overview of all cloned NRPS systems is given in Table 2.

Table 2. Successfully assembled plasmid constructs for recombinant NRPS protein expression.

Bacterial origin	Gene # within BGC/ Name	Size [bp]	Expression vector	Final construct designation
DSM44815	MIMO_2654	7101	pET28b(+)-SUMO pHis8(+)C-TEV	MiC21 NRPS-pET28b(+)-SUMO MiC21 NRPS-pHis8(+)C-TEV
DSM40593	SFUL_137	7230	pET28b(+)-SUMO	SFC2 NRPS- pET28b(+)-SUMO
DSM17904	PxaA	7224	pET28b(+)-SUMO pHis8(+)C-TEV pGS21a(+)-TEV pCX2 pCX2-N-His_6	PxaA-pET28b(+)-SUMO PxaA-pHis8(+)C-TEV PxaA NRPS-pGS21a(+)-TEV PxaA-pCX2 PxaA-N-His_6
DSM15199	VY86_08980	7212	pET28b(+)-SUMO pETDuet pGS21a(+)-TEV pCX2 pBAD-His_6	PTC12 NRPS- pET28b(+)-SUMO PTC12 NRPS-pETDuet PTC12-pGS21a(+)-TEV PTC12 NRPS-pCX2 PTC12 NRPS-pBAD-His_6
PAO1	PA_3327	7059	pET28b(+)-SUMO pHis8(+)C-TEV pACYC-Duet1	PAO1C29 NRPS- pET28b(+)-SUMO PAO1C29-pHis8(+)C-TEV PAO1C29-pACYCDuet1
SMg1	M444_02650	7203	pET28b(+)-SUMO	SMg1C5 NRPS-pET28b(+)-SUMO
HKI0770	Pys PCP-domain	3894 210	pET28a(+) pET28b(+)-SUMO	Pys-pET28a(+) Pys-PCP-pET28b(+)-SUMO

The NRPS gene PxaA, originally encoded within the XSC19, was obtained from the Bode group (University of Frankfurt) together with the corresponding MOX gene PxaB in one plasmid and normally used for heterologous BGC expression *in vivo*. PCR-amplification of the target gene was therefore based on plasmid template DNA. In addition to the generally used pET-based vector systems, PxaA protein production was also tested in the pGS21a(+)-TEV vector

system. Besides the His_6 tag, the expressed protein carries an N-terminal GST fusion protein that enhances target protein stability and solubility.

Like PxaA, the NRPS system originating in *Photorhabdus temperata* DSM15199 was cloned into the pCX2 vector, which uses the $araP_{BAD}$ promoter as a regulatory element for gene transcription. The intension of using an alternative promoter system was to avoid leaky expression, as it is the case for the T7 promoter. Being a positive control regulator, the $araP_{BAD}$ promoter reduces basal expression of genes and thus produces lower background. Unfortunately, the pCX2 vector was originally designed without the presence of an affinity tag, therefore a modified version bearing an N-terminal His_6 tag was constructed. In parallel, cluster # 12 NRPS was inserted into the tag-bearing pBAD-C-His_6 expression system.[143]

Duet vector systems such as pETDuet-1 (PTC12) and pACYC-Duet-1 (PAO1C29 NRPS) were used for co-expression tests, where NRPS and MOX target genes were introduced on one plasmid but under the control of two independent promoters. By introducing core PA genes into these Duet systems, protein interactions and putative protein complex formation as well as the effect of co-expression on protein product yield, solubility and retainment of catalytic activity were to be studied.

Apart from the bimodular NRPS systems associated with PA production, the monomodular Pys NRPS was expressed to study the enzymes' substrate specificity in the related pyreudione biosynthetic pathway. The fully assembled Pys-N-His_6-pET28a(+) construct was provided by our collaboration partners at the Hans Knöll Institute in Jena (Stallforth group). Catalytic activity of recombinantly expressed Pys enzyme was measured based on activation and loading of the PCP-domain. Therefore, the isolated Pys PCP-domain was also cloned into the pET28b(+)-SUMO vector system.

Protein expression plasmid construction was achieved by SLIC or HiFi DNA Assembly. Target NRPS genes were amplified using KOD Xtreme™ Hot Start DNA Polymerase from respective genomic DNA, with annealing primers exhibiting 20 bp nucleotide overhangs complementary to the 5' and 3' ends of insertion site on the vector backbone (Figure 54). In cases of difficult to amplify insert DNA due to high GC content and repetitive sequences, leading to hairpin and primer dimer formation, homology arms were attached to the vector backbone instead.

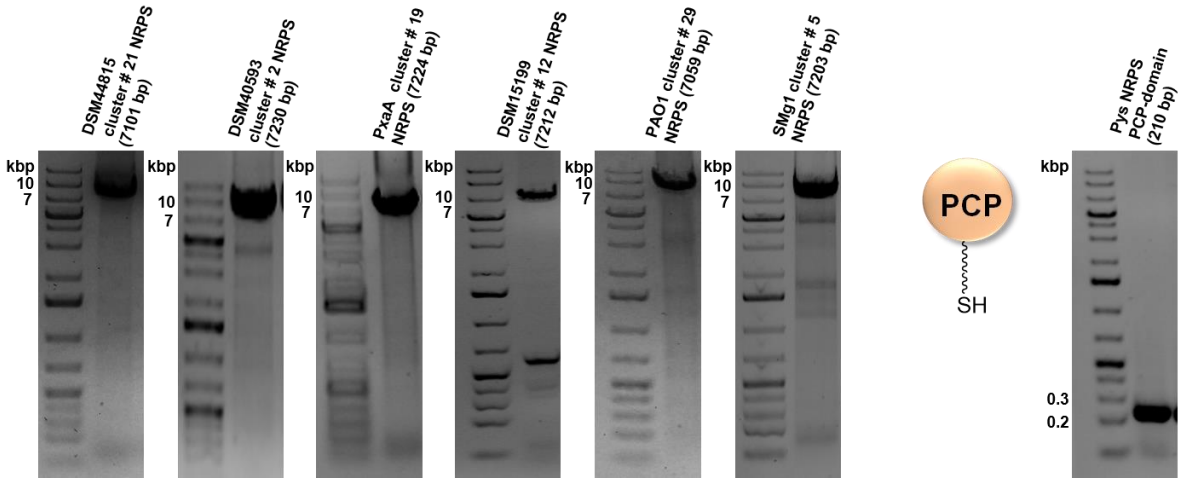


Figure 54. PCR-amplification of NRPS genes for recombinant protein expression. Target NRPS genes (7000 – 8000 bp) captured from selected bacterial genomes were successfully amplified. Cloning of the Pys NRPS PCP-domain (252 bp) was performed for catalytic activity and substrate loading assays.

3.3.3 Cloning of BVMOs into *E. coli* protein expression systems

Like NRPS cloning, MOX genes were preferably inserted into pET-based vector systems, either the N-terminally His_6-tagged pET28b(+)-SUMO or the C-terminally His_8-tagged pHis8C(+)-TEV. The application of expression systems which attach different types of fusion protein tags (e.g., GST or MBP tags) helped increase protein solubility and stability as well as protein yield[144]. Cloning into suitable vectors such as pGS21a(+)-TEV or pMAL-His_6-TEV was therefore carried out for some of the selected MOXs. In addition, setting target gene regulation under the control of a non-leaky promoter ($araP_{BAD}$) was seen as an alternative to the commonly used leaky T7-promoter system.

An overview of the selection of MOX genes that were cloned as well as the vector systems they were introduced into, can be seen in Table 3.

Table 3. Overview of established MOX-containing expression constructs.

Bacterial origin	Gene # within BGC/ Name	Size [bp]	Expression vector	Final construct designation
DSM44815	MIMO_2655	1164	pET28b(+)-SUMO pHis8(+)C-TEV	MiC21 MOX-pET28b(+)-SUMO MiC21 MOX-pHis8(+)C-TEV
DSM17904	PxaB	1176	pET28b(+)-SUMO pCX2	PxaB-pET28b(+)-SUMO PxaB-pCX2
DSM15199	VY86_08975	1161	pET28b(+)-SUMO pCX2 pGS21a(+)-TEV pMAL-His_6-TEV pBAD-His_6	PTC12 MOX- pET28b(+)-SUMO PTC12 MOX-pCX2 PTC12 MOX- pGS21a(+)-TEV PTC12 MOX- pMAL-His_6-TEV PTC12 MOX- pBAD-His_6
PAO1	PA3328	1167	pET28b(+)-SUMO pHis8(+)C-TEV pGS21a(+)-TEV	PAO1C29 MOX- pET28b(+)-SUMO PAO1C29 MOX- pHis8(+)C-TEV PAO1C29 MOX- pGS21a(+)-TEV
SMg1	M444_02655	2523	pET28b(+)-SUMO	SMg1C5 MOX-pET28b(+)-SUMO
F613-1	BB341_06570	1194	pET28b(+)-SUMO	SclC22 MOX-pET28b(+)-SUMO
	BB341_16995	1200	pET28b(+)-SUMO	SClC39 MOX- pET28b(+)-SUMO
Psp. SH-C52	BraC	1185	pET28b(+)-SUMO	BraC- pET28b(+)-SUMO

All the selected MOX genes, ranging in nucleotide lengths between 1100 and 2500 bp, were successfully amplified with Q5® High-Fidelity DNA Polymerase and inserted at the respective restriction site of the vector of choice (Figure 55). As previously mentioned, the MOX gene PxaB originating in *Xenorhabdus stockiae* DSM17904 was amplified from plasmid DNA as template instead of genomic DNA. The BraC-pET28b(+)-SUMO construct, harboring a MOX gene from the PA-related brabantamide-producing biosynthetic pathway, had previously been assembled in our laboratory by Dr. Françoise Schaefers. Plasmid establishment was achieved by SLIC cloning.

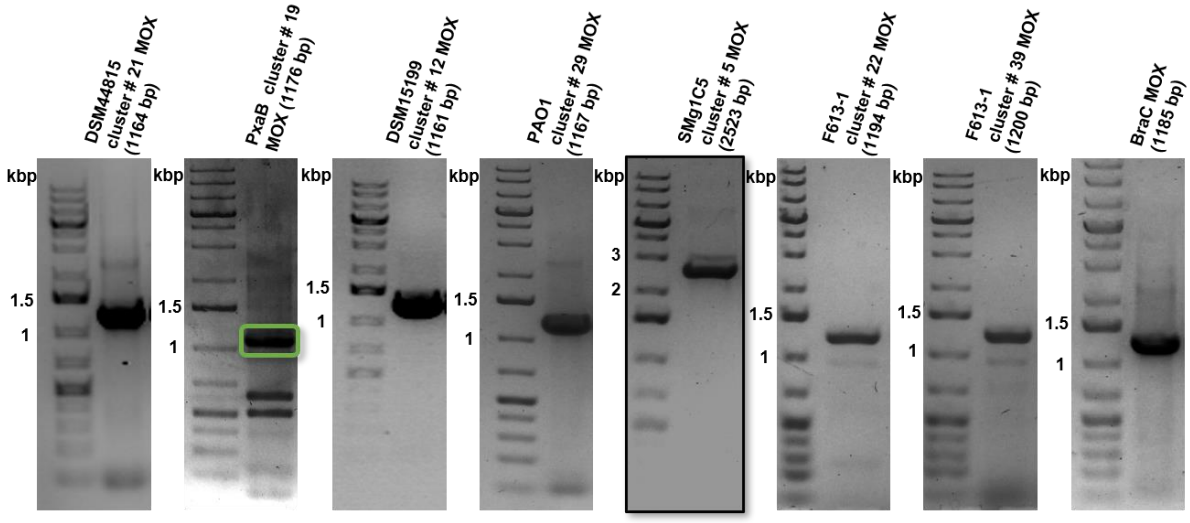


Figure 55. Capturing of MOX genes from genomic or plasmid template DNA by PCR. Except for PxaB insert amplification, designed annealing primers were specific for the nucleotide sequence of interest and thus amplicons were obtained in high amount and purity. PxaB insert DNA was purified by gel extraction. MOX genes were of 1100 – 1200 bp in length. Solely SMg1C5 MOX was double in size (black boxed agarose gel).

3.3.4 Cloning of accessory genes involved in PA precursor synthesis

A central aspect for enabling successful *in vitro* reconstitution of secondary metabolite pathways, such as PA biosynthesis, are the assay conditions. *In vivo*, enzyme-catalyzed reactions linked to one pathway take place in the same environment, sometimes even within the same cellular compartment. This leads to suitable presence of salts, a certain pH that guarantees protein functionality as well as the presence of co-factors, co-enzymes, and many other small molecules.[145] Furthermore, *E. coli* cellular pathways can be used to provide precursor molecules or substrates necessary for heterologous expression of biosynthetic pathways *in vivo*. It is impossible to reconstitute the complexity of *in vivo* cellular conditions *in vitro*. However, some criteria can be adjusted, such as the pH and ionic strength of assay buffers, supplementation of co-factors and co-enzymes as well as providing substrates and precursor molecules.

As previously mentioned, NRPS systems in particular are responsible for catalyzing PA intermediate production (such as **43**) by incorporating and coupling L-amino acids with activated fatty acid thioesters to create the bicyclic pyrrolizidine ring system typical for PAs. Under *in vitro* conditions, these precursor molecules must be supplied in their activated forms

so that the NRPS system can catalyze the target reaction. Heterologous expression of recombinant accessory proteins that provide these substrates is therefore necessary. Generation and supplementation of fatty acyl-CoA thioesters was intended to be catalyzed by the acyl-CoA ligase FadD, a protein involved in fatty acid degradation processes and thus produced by all *E. coli* strains. The FadD gene was PCR-amplified from *E. coli* BL21DE3 genomic DNA and inserted into the pHis8C(+)-TEV vector system (Figure 56A.). As described in chapter 3.2.1.2, the SAM-dependent enzyme AzeJ encoded within the azetidomonamide-producing BGC of *Pseudomonas aeruginosa* PAO1 is responsible for generating the non-proteinogenic amino acid L-AZC out of methionine-derived SAM. Since L-AZC is the key precursor molecule differentiating PA biosynthesis in PAO1 from other bacterial pathways, AzeJ was cloned into the pGS21a(+)-TEV vector system (Figure 56B.) for co-expression experiments with NRPS and MOX core genes AzeB and AzeC.[115]

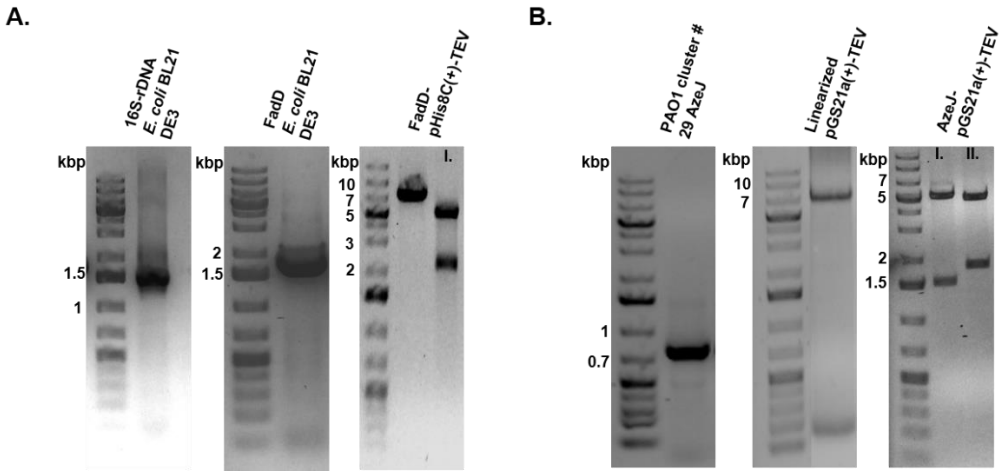


Figure 56. Introduction of target accessory genes for recombinant protein expression. **A.** Isolated *E. coli* BL21DE3 genomic DNA was verified by 16s-rDNA screening PCR using general 27F/1492R primers. FadD insert DNA (1686 bp) was successfully introduced into the pHis8C(+)-TEV vector at the BamHI restriction site. Analytical restriction digest with AflIII (2101, 4906 bp) was performed to confirm plasmid assembly (overall plasmid size: 7007 bp). **B.** SAM-dependent enzyme AzeJ (753 bp) was inserted into the restriction digested and dephosphorylated pGS21a(+)-TEV vector (6200 bp) via HiFi DNA Assembly to yield the AzeJ-pGS21a(+)-TEV plasmid. Analytical restriction digest with I. SacI/NdeI (1533, 5305 bp) and II. SphI/HindIII (1840, 5085 bp) validated correctness of the construct.

Overall, out of ten bacterial strains chosen for in-depth investigations, enzymes of seven strains involved in PA core structure production were accessed. NRPS systems of interest were captured from corresponding gDNA with the help of KOD Xtreme™ Hot Start DNA Polymerase whereas MOX genes were successfully amplified using Q5® High-Fidelity DNA Polymerase. Target DNA was inserted into a variety of different vector systems exhibiting

alternate regulatory elements (e.g., promoter type, origin of replication) or tags (His_6, MBP, GST) which can contribute to protein solubility and stability upon expression. Finally, cloning of genes encoding for proteins involved in PA precursor and substrate synthesis was achieved.

3.3.5 Recombinant protein expression and purification

3.3.5.1 Bimodular NRPS systems

In initial protein expression attempts with NRPS systems cloned into the pET28b(+)-SUMO vector, most proteins from different bacterial strains could not be expressed. Exceptions were the SFC2 NRPS and SMg1C5 NRPSs, where the full-sized enzyme still linked to the SUMO fusion tag could be detected in SDS gels after protein purification (Figure 57). Protein translation was induced using 0.1 mM IPTG and cells were cultured at 16 °C over night.

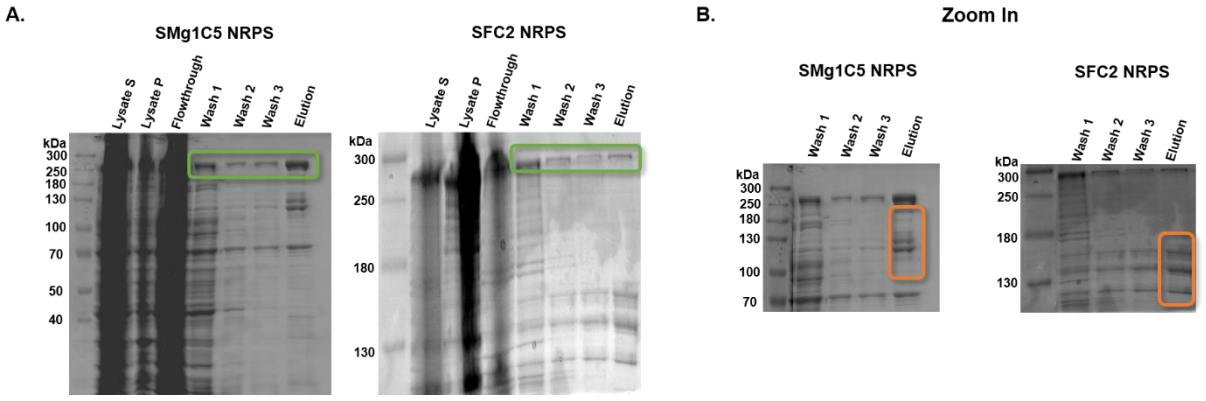


Figure 57. Protein expression profile of NRPS systems. **A**. SDS-gel documentation of complete NRPS protein purification process. NRPS system is highlighted in green. **B.** Protein eluate of both SMg1C5 and SFC2 NRPS contained impurities associated with truncated versions of the respective protein (orange boxes).

When comparing the overall cellular expression of the *E. coli* BL21DE3 expression strain (lysate supernatant and pellet fractions) with the purified protein eluate, the quantity of NRPS protein produced is low although protein yield can reach up to 50 % of total cell volume in this type of vector and host expression system.[146] Low yields could be linked to the large protein size of approximately 250 kDa and its complexity, as it consists of multiple functional domains (C, A, PCP, TE). Due to the high GC content, mRNA secondary structure formations could cause premature termination during the translation process, leading to truncated versions of the NRPS system, as can be seen in the elution fraction (Figure 57B). These truncated versions effect the purity of the full-sized protein since they also carry an N-terminal His_6-tag and some also sizes over 100 kDa, making them difficult to separate by Ni^{2+}NTA column and

centrifugation using a molecular weight cut off (MWCO). Furthermore, the N-terminal 5'-untranslated region as well as the initial codons of the target gene have an impact on translation efficiency. In particular the GC-content of the gene nucleotide sequence which can result in secondary structure formation of mRNA and prevent translation initiation.[147] The choice of restriction site for gene insertion can also influence translation efficiency if it makes unfavorable secondary structures when in combination with a specific codon sequence.[148] With 50 – 60 % GC-content of the first 50 bps and some NRPS genes starting with a GTG instead of an ATG start codon (e.g., MiC21 NRPS), these factors could have had an influence on NRPS protein yield. Furthermore, the difference between non-expressing and expressing pET28b(+)-SUMO constructs was an alteration of bp insertion upstream of the start codon for in frame cloning (two bps: protein expresses, or five bps: no protein expression).

NRPS systems from *Pseudomonas aeruginosa* PAO1 (PAO1C29 NRPS) and *Xenorhabdus stockiae* DSM17904 (PxaA) were successfully expressed in the pHis8C(+)-TEV vector. Analogous to the functional NRPS-pET28b(+)-SUMO vector system, the minimum number of two bps were inserted for the target gene to be in frame with the promoters' starting codon, leading to the assumption that the distance between SD-sequence and gene start influences protein translation success. The use of a C-terminal His_8-tag improved protein purity after purification from *E. coli* BL21DE3 because only fully translated protein contains the affinity tag. Thus, copurification of truncated proteins was reduced (Figure 58).

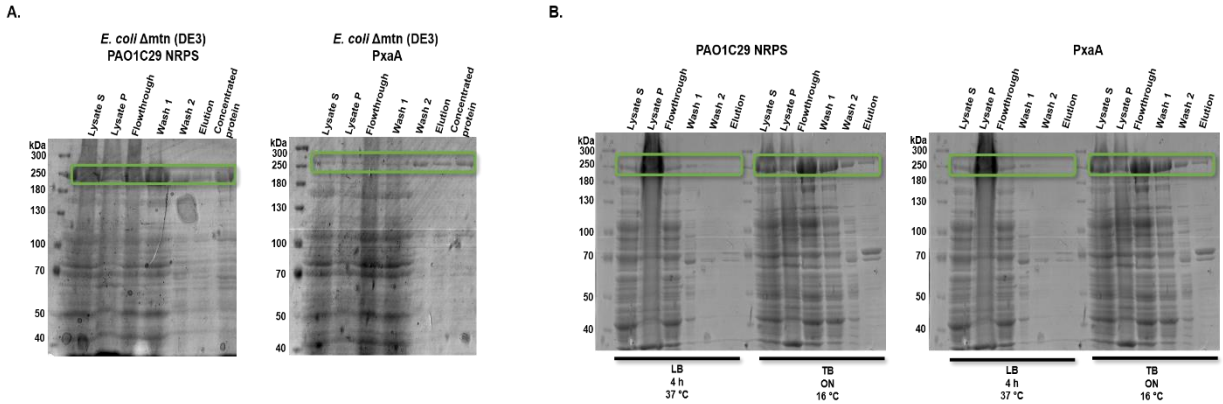


Figure 58. SDS gels of recombinant NRPS expression tests performed under specific expression conditions. **A.** Expression level in *E. coli* Δmtn (DE3) after induction with 0.05 mM IPTG was like that of *E. coli* BL21DE3 when induced with 0.1 mM IPTG (figure part B). **B.** Optimization of NRPS protein expression by varying medium composition, growth period and cultivation temperature. S: supernatant fraction; P: pellet fraction.

Based on recombinant protein expression successes of the C-terminally tagged NRPS systems, expression parameters as well as purification steps were varied to further optimize protein quantity and quality. In many cases, the choice of the host expression strain can make the difference on whether a protein is expressed, for example by expressing additional co-chaperones that support the folding process of the target protein and enhance solubility. Furthermore, strains such as BL21(DE3) have high protein expression levels but simultaneously also possess high levels of basal expression. We therefore intended to use an alternative *E. coli* Δmtn (DE3) host strain, derived from the KEIO collection mutant JW0155-1 and originally designed to study radical SAM-related pathways, which reduces background protein production significantly while maintaining a similar efficiency as *E. coli* BL21(DE3).[149] However, heterologous expression of both PxaA and PAO1C29 NRPS in *E. coli* Δmtn did not increase protein yield or protein purity significantly. Although expression of PAO1C29 NRPS seems to be elevated to some extent, most of the produced protein remained insoluble (Figure 58). The concentration of IPTG used for induction of target gene transcription was another aspect considered to lower the amount of insoluble protein lost in the pellet fraction of previous tests (Figure 58). Since IPTG is not metabolized by the cell, a constant level of induction persists throughout the entire expression process. Thus, lowering of induction levels from 0.1 mM to 0.05 mM in our expression experiments may result in slower protein production, promotion of correct folding and receival of quantitatively higher amounts of soluble NRPS.[150] Comparison of protein expression profiles by SDS gel electrophoresis revealed no remarkable difference between the two IPTG induction amounts. In line with the decrease of IPTG to obtain more soluble protein, a change of medium (LB instead of TB) in combination with altered growth temperatures and times was tested to enhance production of (soluble) protein in a higher yield. Media composition supposedly influences cell survival, growth, and type of metabolite production as individual components (salts, carbon sources etc.) can trigger different cellular processes, including protein expression.[151] Exchanging the nutrient richer TB for the less nutrient-containing LB was intended to investigate whether LB components support NRPS production in a more efficient way than those present in TB medium. Protein expression was performed in LB medium for 4 h at 37 °C before harvesting the cells whereas standard approaches included recombinant NRPS production in TB medium over night at 16 °C. Both PxaA and PAO1C29 NRPS were expressed (Figure 58) under the new conditions. Unfortunately, most of the protein was found to be insoluble, as it was only present in the pellet fraction and the overall yield was significantly reduced. A shorter growth period could not compensate the high expression temperature in terms of generating more soluble protein and reduction of the expression temperature to 16 °C did not change the protein expression profile.

In conclusion, the standard procedure was the most efficient for protein expression and was applied to produce sufficient amounts (3-5 mg/L) of NRPS for enzyme activity assays.

SDS gels in Figures 57 and 58 showed that decisive amounts of protein were lost during the purification process, especially in the pellet fraction, due to solubility issues and upon column loading for subsequent Ni^{2+}NTA chromatography (flowthrough fraction). To reduce NRPS loss to a minimum, the obtained fractions were processed individually. The inclusion body-containing pellet fraction was treated with the non-denaturing solubilization agent *N*-lauroyl sarcosine (NLS), an ionic detergent that promotes micelle formation and overcomes the protein solubility barrier.[152] In this way, a significant amount of NRPS protein was recovered. Furthermore, protein unable to stay attached to the Ni^{2+}NTA resin of the column during loading via its His_6-tag was collected in the flowthrough and incubated a second time with fresh beads before repeating the purification procedure (Table S5). Lastly, fractions collected when washing with buffer 2 (40 mM imidazole) did not differ in their protein elution profile from the final elution step (250 mM imidazole) and showed no major impurities, therefore the two fractions were often combined and concentrated as one.

3.3.5.2 Monomodular NRPS system pys

Together with our collaboration partners at the HKI Jena (Martin Klapper) we set out to study the substrate scope of the PA derivative-producing monomodular NRPS system Pys found in the bacterial *Pseudomonas fluorescens* HKI0770 strain. Unlike biosynthesis of **37** and **40**, where a bimodular NRPS system in combination with a MOX is needed to establish the PA core, the presence of a MOX enzyme is not necessary for production of pyrrolizidine diones. Since all reaction steps in pyreudione biosynthesis (**38**) are catalyzed by the monomodular Pys NRPS system, the mechanism of action differs from bimodular NRPSs. The A-domain initiates amino acid loading (L-serine and L-proline) followed by C-domain catalyzed coupling to a β-ketoacyl thioester and release of the heterocyclic product by Dieckmann cyclization at the TE-domain. To study substrate promiscuity of the C-domain-catalyzed reaction towards different β-ketoacyl thioesters, full functionality of the recombinantly expressed Pys protein had to be ensured. As previously mentioned, we received the assembled pET28a(+)-Pys plasmid from our collaboration partners. With a size of 142 kDa, Pys protein could be obtained in high yields of up to 10 mg/L and with high purity after concentrating with a MWCO of 100 kDa to remove remaining impurities (Figure 59A, Table S5).

Posttranslational modification of the NRPS PCP domain at a conserved serine residue converts the carrier protein from the *apo*- to the *holo*-form. Therefore, availability of the *holo*-protein can be used as an indicator for correct folding of the expressed protein and subsequent NRPS enzyme activity. A coenzyme A-derived phosphopantetheine (ppan) co-factor is covalently attached to the serine moiety by a specific PPTase, which is often co-expressed with the NRPS system *in vivo*.[53] For *in vitro* assays, the standalone Pys-PCP domain was expressed using the pET28b(+)-SUMO vector and purified via its N-terminal His_6-tag (Figure 59B). Alongside the target domain, an unidentified contamination with a protein of higher and one of lower molecular weight were co-purified. Unfortunately, these impurities could not be removed using MWCO-containing concentrators and remained in the protein eluate for enzyme activity testing.

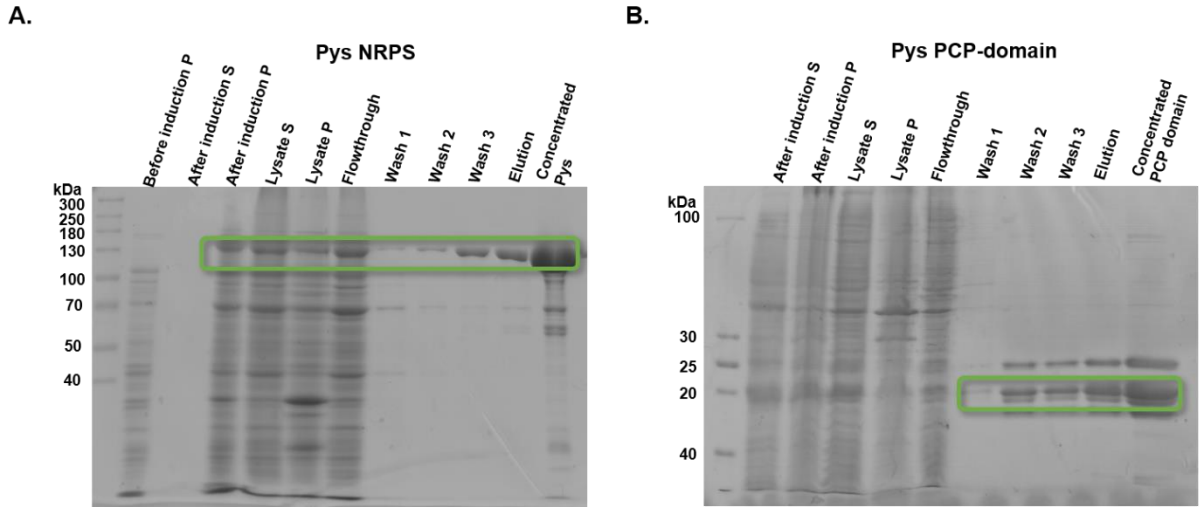


Figure 59. Full protein Pys and PCP-domain expression. **A.** Standard protocol purification led to receival of relatively pure recombinantly expressed Pys NRPS protein. **B.** Heterologous expression of Pys PCP-domain in *E. coli* BL21(DE3).

3.3.5.3 FAD-dependent BVMOs

In contrast to NRPS proteins, which were preferably expressed with C-terminal fusion tags, all selected bacterial MOX systems were producible bearing N-terminal tags such as SUMO, GST or MBP. pET28b(+)-SUMO was the primary vector of choice (Figure 60) and expression in alternative vector systems was solely to optimize protein production as well as purification.

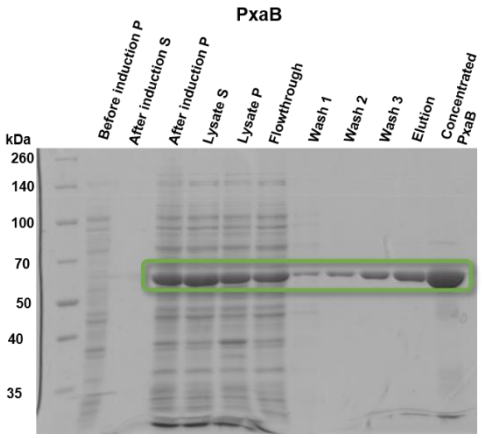


Figure 60. SDS-gel illustrating the protein purification process of MOX enzyme PxaB. Addition of IPTG induced overexpression of protein which was separated from *E. coli* host cell native proteins using multiple washing steps by increasing buffer imidazole concentration. PxaB was eluted in a final step and concentrated for further utilization in enzyme activity assays.

In combination with the terminally-linked fusion protein of choice, the selected MOXs displayed molecular weights of approximately 60 – 70 kDa. Often, native *E. coli* proteins possess similar sizes, making it difficult to discriminate between target protein and background signal. Despite the lack of a fusion tag, these specific *E. coli* contaminants interact via clustered histidine residues or metal binding sites with the Ni^{2+}NTA resin and thus remain bound to the column during washing steps and co-elute with the protein of interest.[153] Native *E. coli* BL21(DE3) production was monitored by expressing the empty vector simultaneously to the MOX-containing expression construct and comparing elution profiles after protein purification to discriminate target protein from impurities based on SDS gel electrophoresis (Figure 61). In cases of doubt, proteins were further characterized by MS-MS analysis to confirm identity (Figures S10 - S11).

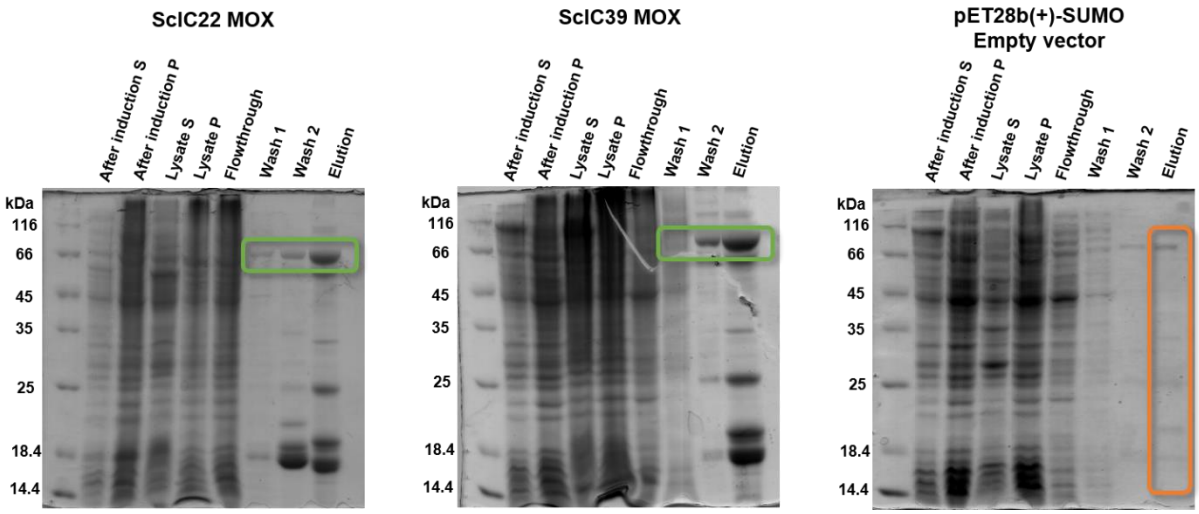


Figure 61. Expression of MOXs from *Streptomyces clavuligerus* F613-1 clusters # 22 and # 39. His_6-tag purification of pET28b(+)-SUMO constructs led to elution of overexpressed MOX proteins still bearing SUMO-tag (marked in green). Comparison with expression and purification profile of empty pET28b(+)-SUMO vector enabled discrimination between target and native proteins (marked in orange).

Vectors with GST and MBP tags were intended to further enhance solubility and stability, to minimize the occurrence of side products in the final eluate fraction, and to improve overall product yield and purity. The MBP-tag was of particular interest for aiding in target protein solubilization as it shows intrinsic chaperone activity. On the contrary, GST was attached at the N-terminus primarily to protect target proteins from proteolytic degradation.[154] PTC12 MOX was successfully expressed when linked to an MBP-fusion tag as can be seen in Figure 62A. Using Ni^{2+}NTA column chromatography for purification, enough target protein with acceptable quality for activity testing was obtained. Here, the concentration of inducer used for translation initiation did not significantly affect product outcome. However, when purifying PTC12 MOX over a Dextrin Sepharose packed column where protein is eluted by increasing maltose buffer concentration, the MBP- tag broke off the protein fusion construct, causing the MOX of interest to be eluted in the flowthrough fraction with all other proteins soluble in the cell lysate and receiving the standalone MBP-tag in the final elution step (Figure 62B). In cases where column chromatography is necessary, it is therefore advisable to pre-purify MBP-tagged proteins via the N-terminal His_6-tag prior to applying them to an MBP-Trap™ HP column. The issue of fusion tag cleavage and consequent purification of tag instead of target protein was also observed for PTC12 MOX bearing a GST-tag. The 42 kDa MOX with the 26 kDa GST protein attached was expressed successfully. However, due to the instability of the fusion construct and thus loss of the affinity tag during the purification process, the final eluate contained only

a minimal amount of target protein compared to the amount of GST-tag detected by SDS gel electrophoresis (Figure 62C).

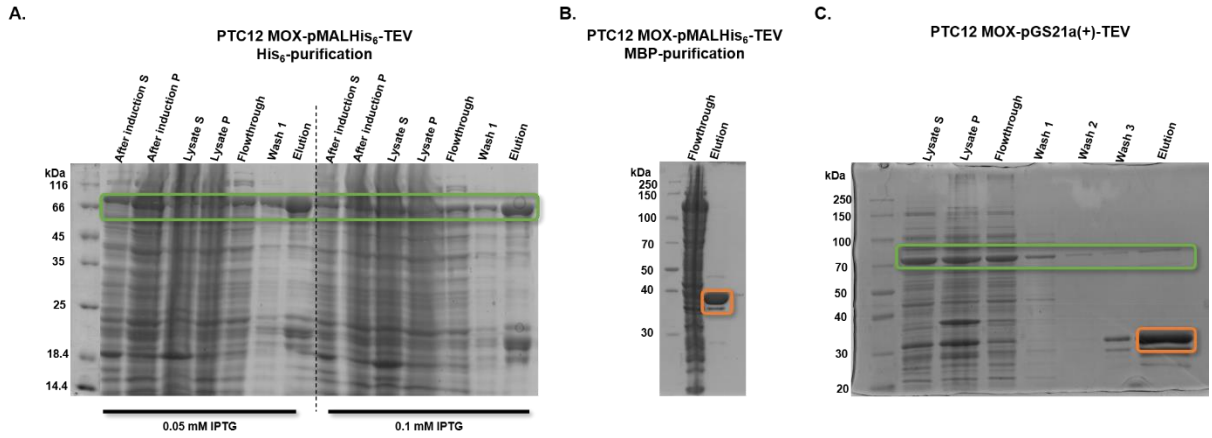


Figure 62. Purification of N-terminally tagged MOX protein using different fusion tags and column chromatography methods. **A.** His_6-tag purification yielded soluble protein independent of inducer concentration (green box). **B.** Separation of target MOX protein from native *E. coli* proteins via the MBP-tag protein resulted in elution of the MOX with all other proteins due to undesired tag cleavage (eluted maltose binding protein marked in orange). **C.** MOX protein expression of construct fused to N-terminal GST-tag was possible (marked in green). However, the main component in final eluate was broken-off GST-tag (orange box).

Further protein expression studies aimed at enhancing solubility and retrieval of protein from the pellet fraction by examining protein behavior in the presence of various additives such as NLS, mannitol, L-arginine, L-glutamic acid, ethanol, or glucose. Different types of additives contribute to target protein stability and increase of expression levels in various ways, for example, by serving as ligands which stabilize existing protein conformations and protect proteins from structure alteration such as misfolding or unfolding. Furthermore, additives can either reduce protein-protein interactions to omit aggregation (chaotropic agents) or they can stabilize intramolecular bonds to maintain protein tertiary structure (kosmotropic agents). In addition, sugars, salts or charged amino acids often aid protein stability by altering pH or buffer composition.[155] Agents such as mannitol, L-arginine, L-glutamic acid[156] and NLS were supplemented to the cell lysate during the initial protein purification step to prevent protein aggregation and thermally stabilize soluble protein, whereas ethanol and glucose addition was performed simultaneous to induction with IPTG. Application of ethanol was of particular interest as it is proposed to function in multiple ways: on the one hand, increasing expression of recombinant proteins by altering membrane fluidity and the cellular environment leading to enhanced DNA synthesis, on the other hand, mimicking a heat-shock response which in turn

can support target protein solubility e.g., by expression of co-chaperones.[157] The intension of glucose addition was to delay protein synthesis by repressing lac promoter induction and “leaky” expression, thus resulting in less insoluble protein as well as reduced background protein production.[146]

Unfortunately, except for the additive NLS, the efficacy of supplementary molecules to enhance protein solubility and stability could not be proven for selected recombinantly expressed MOX enzymes. Neither 3 % EtOH nor 1 mM glucose added to the cultivation broth improved overall expression yields (Figure 63A). Supplementation with amino acids L-arginine and L-glutamic acid (50 mM each) did not affect protein stability. Even more, the expected increase of solubility not only remained unfulfilled but instead protein could only be detected in the cell lysate pellet fraction (Figure 63C). An alteration of buffer composition concerning pH, salts, and buffering substances in combination with amino acid or sugar availability also did not have an impact on protein characteristics. Treatment of lysed cell pellet with NLS for 4 – 6 h led to recovery of MOX protein into the soluble state, as can be seen in Figure 63B upon comparison of supernatant and pellet fractions before and after detergent treatment. Extension of the incubation time (e.g., to 12 – 24 h) or repetition of the procedure with the same cell pellet would most likely result in further resolubilization of aggregated target protein and an increase in overall protein yield.

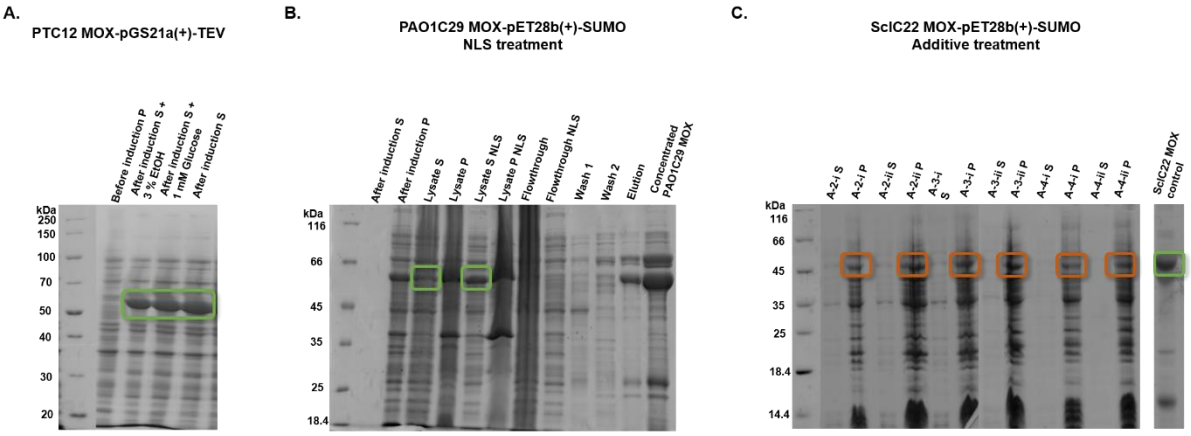


Figure 63. Improvement of protein expression, stability, and solubility by addition of supplements. **A.** EtOH or glucose was added to the culture broth upon protein induction to increase product outcome. **B.** Incubation of lysed cell pellet for 4 h with NLS led to resolubilization of insoluble protein (comparison lysate supernatant before and after treatment marked in green). **C.** Addition of either mannitol (2), L-arginine (3) or L-glutamic acid (4) to the standard lysis buffer (i: NaH_2PO_4, NaCl, imidazole, glycerol, pH 8.0) or an alternative buffer (ii: NaCl, HEPES, pH 7.0) could not improve protein solubility since target protein is absent in supernatant fractions and only present in pellet fractions (orange boxes) but not in increased yields as desired.

3.3.5.4 Proteins involved in PA precursor biosynthesis

As mentioned in chapter 3.3.4, both the native *E. coli* fatty acyl CoA ligase FadD and the SAM-dependent enzyme AzeJ participating in pyrrolizidine alkaloid derivative production in *Pseudomonas aeruginosa* PAO1 are essential to activating and delivering precursor molecules for product synthesis. Whereas FadD catalyzes conversion of an unreactive fatty acid to a reactive thioester species, AzeJ provides nonnatural amino acid substrate for subsequent incorporation catalyzed by the core NRPS enzyme. Protein expression of both enzymes was essential for *in vitro* preparation of activated substrate molecules needed for step-by-step pathway reconstitution in following *in vitro* tests. The 62 kDa large FadD was obtained by heterologous expression in *E. coli* BL21(DE3) or BAP1 cells using the pHis8C(+)-TEV vector system, whereas AzeJ protein was stabilized by fusion with an N-terminally linked GST-tag. As can be seen in Figure 64, certain amounts of soluble FadD protein were most likely not so tightly bound to the Ni^{2+}NTA resin and thus were already eluted during the second washing step with 60 mM imidazole. Since wash fraction two and the elution fraction did not differ in their protein profile, they were combined to increase overall FadD yield and concentrated using a 30 kDa MWCO Vivaspin® tube to remove low molecular weight protein contaminants. AzeJ protein was obtained in smaller amounts and less purity than FadD. Due to its size of 27 kDa and the attached GST-tag (26 kDa) it was difficult to remove native protein contaminants of similar or higher molecular weight.

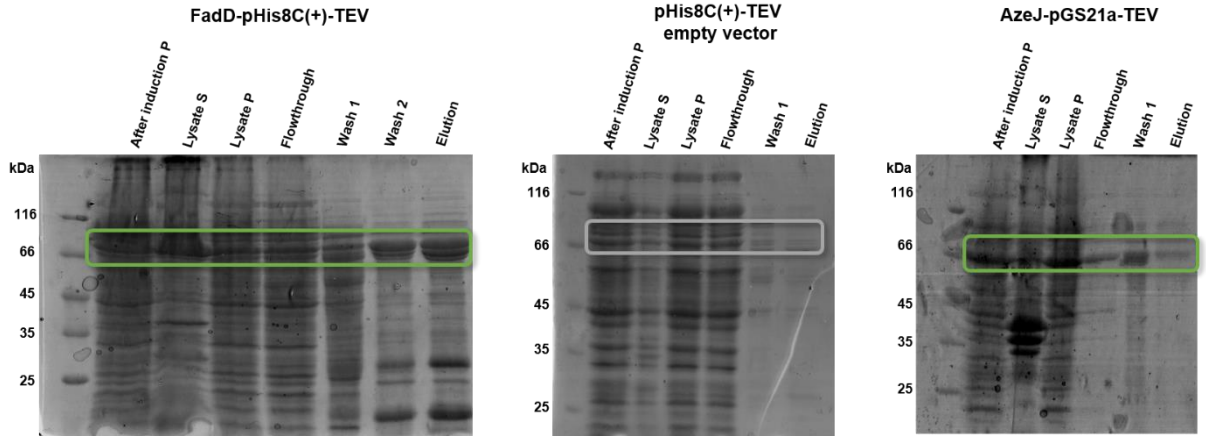


Figure 64. SDS gels of recombinantly expressed reaction-relevant proteins. Direct comparison of FadD and AzeJ expression profiles with an empty vector control confirmed target protein production.

In summary, multiple representatives of both NRPS systems and BVMOs from selected bacterial BGCs were obtained by recombinant production in *E. coli* hosts to study the mechanism of PA biosynthesis *in vitro*. Due to their high molecular weight of more than 250 kDa and complexity resulting from multimodularity, NRPS yields ranged from 3 – 5 mg per Liter of cultivation broth. Especially proteins bearing an N-terminal fusion tag often showed contamination of the final eluate with truncated protein. Attachment of the affinity tag to the C-terminus eliminated a great amount these contaminants, resulting in a relatively pure final protein. The monomodular Pys NRPS system expressed very well and with high purity.

A selection of MOX enzymes was successfully expressed in high yields of up to 25 mg per Liter culture broth. Most of the proteins were stabilized by an N-terminally linked SUMO-tag (Table S5). To further enhance solubility of the 45 kDa sized proteins, a variety of optimization approaches were tested, including fusion to alternative protein tags such as MBP and GST. Supplementation with specific additives, either during expression induction to control DNA synthesis and produce more soluble protein (EtOH, glucose), or during protein purification to regain aggregated protein (NLS) or prevent aggregation (L-arginine, L-glutamic acid, mannitol) was tested. Whereas most optimization attempts did not significantly improve MOX stability and solubility, application of NLS made it possible to resolubilize aggregated protein and thus improve overall target protein yield.

In addition to the successful expression of multiple homologues of both NRPS and MOX core enzymes, FadD as well as the SAM-dependent enzyme AzeJ were obtained in amounts sufficient for utilization in *in vitro* assays.

3.3.6 Enzyme activity assays for stepwise *in vitro* pathway reconstitution

3.3.6.1 FadD-catalyzed fatty acid activation

The initial reaction taking place in PA biosynthesis comprises the incorporation and coupling of activated fatty acid and amino acid substrates. *In vivo*, carboxylic acid activation is processed in several metabolic routes. However, the complexity of cellular process cannot easily be mimicked or reconstructed *in vitro*. Nevertheless, precursor molecules equivalent to those available *in vivo* can be generated by enzymatic catalysis inside a test tube. Preparation of fatty acid thioesters, in particular acyl-CoA, which are much better acyl group donors than carboxylic acids, was studied using the acyl-activating enzyme FadD from *E. coli* BL21(DE3).[158] Conversion of undecanoic (U, C_{11}), lauric (L,C_{12}) and palmitic acids (PL,C_{16}) in the presence of FadD was tested with purified protein, lysate supernatant after centrifugation

and crude lysate still containing cell debris after sonication with simultaneous variation of solvent for dissolving fatty acid substrate. In addition, reaction times and temperatures were altered to find optimal substrate conversion conditions. Reaction assay setup to test acyl-CoA synthetase activity was based on the protocol published by Morgan-Kiss and Cronan[159] (Table 4) with slight modifications. For lysate supernatant and crude lysate, all reaction components were added as a master mix to 0.3 mL of the respective lysate.

Table 4. Enzyme assay setup to test FadD acyl-CoA synthetase activity.

Components	Stock concentration [mM] or [µM], [%]	Final concentration [mM] or [µM], [%]
ATP	1 mM	50 µM
$MgCl_2$	100 mM	8 mM
CoA-SH	10 mM	0.5 mM
Fatty acid*	100 µM	10 µM
FadD	50 – 150 µM	10 µM
EDTA	100 mM	2 mM
KF	1000 mM	20 mM
Triton X-100	100 %	0.1 %
Tris-buffer	50 mM Tris-HCl, 10 % glycerol, pH 7.5	

*U or L dissolved in acetone, DMSO or DMF

Initial attempts to receive, lauroyl- (**68**), (un)decanoyl- (**71**,**72**) or palmitoyl-CoA (**73**) with purified FadD remained without success for all reaction types (Figure 65). Identical repetition of enzyme activity assays as well as tests using shorter fatty acid educts (octanoic (O) and hexanoic (H) acid) and an alternative HEPES buffering system did not resolve these issues. Publications thematizing FadD-catalyzed acyl-CoA synthesis executed *in vitro* reactions at either 30 °C, 37 °C or at room temperature (20 °C) which is why test reactions were conducted at both room temperature (lower limit) and 35 °C (intermediate temperature of upper limits). A lack of conversion at room temperature could possibly result from low enzyme activity, since most proteins have optimal activity temperatures of 28 °C to 37 °C. In contrast, the absence of fatty acid thioesters in tests performed at 35 °C hints at catalytic inactivity of FadD. Yoo et al.[160] experienced FadD only being fully functional in the presence of a lipid surface and that *in vitro* supplementation of Triton X-100 detergent in combination with bacterial membranes is necessary to ensure protein activity. Although Triton X-100 was applied for the test reactions, FadD activation could have been unsuccessful due to missing bacterial membrane components. Fatty acids were dissolved in either acetone, DMSO or DMF and added to the assay with a final proportion of 10 % (v/v) since solubility tests performed prior to reaction setup had indicated protein solubility in all three solvents up to this solvent percentage. Despite testing protein solubility, most samples showed critical amounts of protein aggregation and

degradation at the end of the reaction time which went hand in hand with a lack of protein activity in all measured assays. A negative control solution without purified FadD remained clear even after overnight incubation, thereby confirming the instability of FadD under the given reaction conditions.

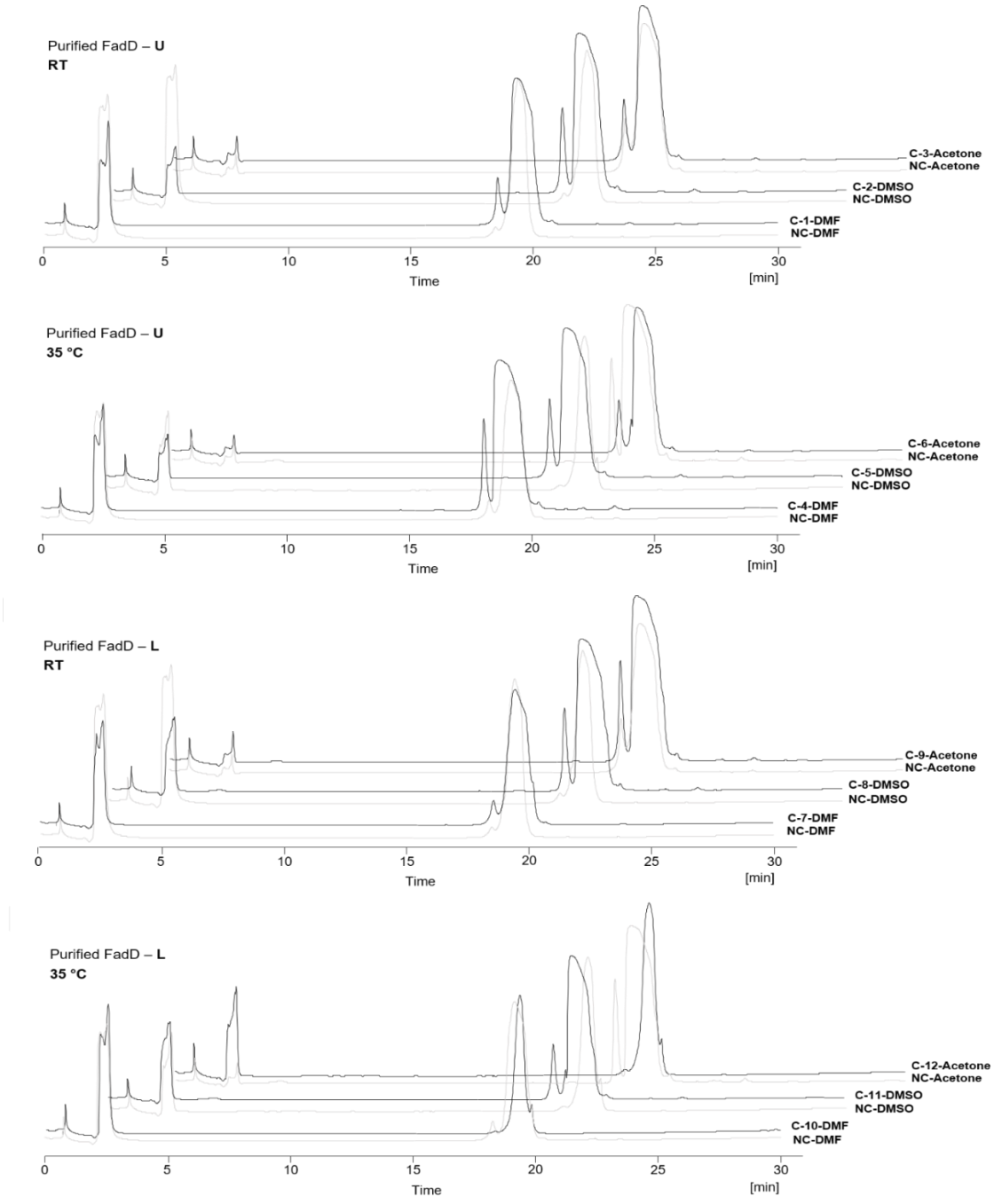


Figure 65. Enzymatic synthesis of acyl-CoA derivatives. Purified acyl-CoA ligase FadD from *E. coli* BL21(DE3) was intended to catalyze undecanoyl and lauroyl group transfer onto CoA for fatty acid activation. Test assays were performed at room temperature (RT) and at 35 °C, with fatty acids dissolved in solvents acetone, DMSO or DMF.

Exchange of the buffering agent Tris for HEPES and pH elevation from 7.5 to 7.8 and reduction of the amount of total solvent present in the test assay (5 – 7 %) did not overcome insolubility and protein nonfunctional issues. The intention of testing acyl-CoA biosynthesis upon overnight incubation in addition to 15 – 45 min time periods was to enable full conversion. However, elongated incubation times could have possibly led to decrease in enzyme kinetic rates at later time points of the reaction causing less and slower conversion of substrate and in turn leading to a shift in equilibrium and decreased temperature optima.[161] Thus, longer reaction times might not have been beneficial for precursor production in our case which could explain the absence of desired acyl-CoA product after 12 h.

Since multiple attempts to conduct acyl-CoA biosynthesis remained unsuccessful using purified FadD protein, follow-up experiments were tested with cell lysates. Pure lysate samples were prepared by cell lysis of the *E. coli* expression host cells by ultrasonification with subsequent centrifugation to eliminate cellular debris from the solution. It was thought that native *E. coli* proteins, for example chaperones or putative co-factors which are also soluble in lysis buffer, can have a stabilizing effect on FadD and furthermore aid in correct protein folding. Alternatively, assays with sonicated crude lysate still containing cellular debris were performed in case non-soluble components contribute to FadD folding and functionality. In *E. coli*, FadD functions as an acyl-CoA synthetase and together with the outer membrane protein FadL is crucial for transport and processing of exogenous long chain fatty acids. Whereas FadL binds fatty acids and shuttles them through outer membrane, periplasmatic space and the inner membrane, FadD associated with the inner membrane uses the long chain fatty acids as substrates and converts them to thioesters. The cell makes use of the generated acyl-CoA thioesters either as a source of energy or as substrates for various molecules.[162] Thus, assays using crude lysate were considered as promising since cell membranes as well as membrane-associated proteins remained in the sample.

Unfortunately, neither pure lysate nor crude lysate assays yielded acyl-CoA molecules, independent of the fatty acid hydrocarbon chain length (Figure 66). As during experiments with purified FadD protein, only two prominent peaks were detected by HPLC at retention times of 18 and 19 min, which exhibited *m/z* ratios ranging from 500 to 900 with high relative ion abundances. However, none of these ion peaks could be linked to **68** ($[M+H]^+$ = 950.8 g/mol), **71** ($[M+H]^+$ = 936.8 g/mol) or alternative acyl-CoA formation. Since acyl-CoA precursor activation could not be achieved by enzymatic catalysis, preparation of decanoyl-CoA (**72**) and **68** was executed by chemical synthesis (section 3.4.1).

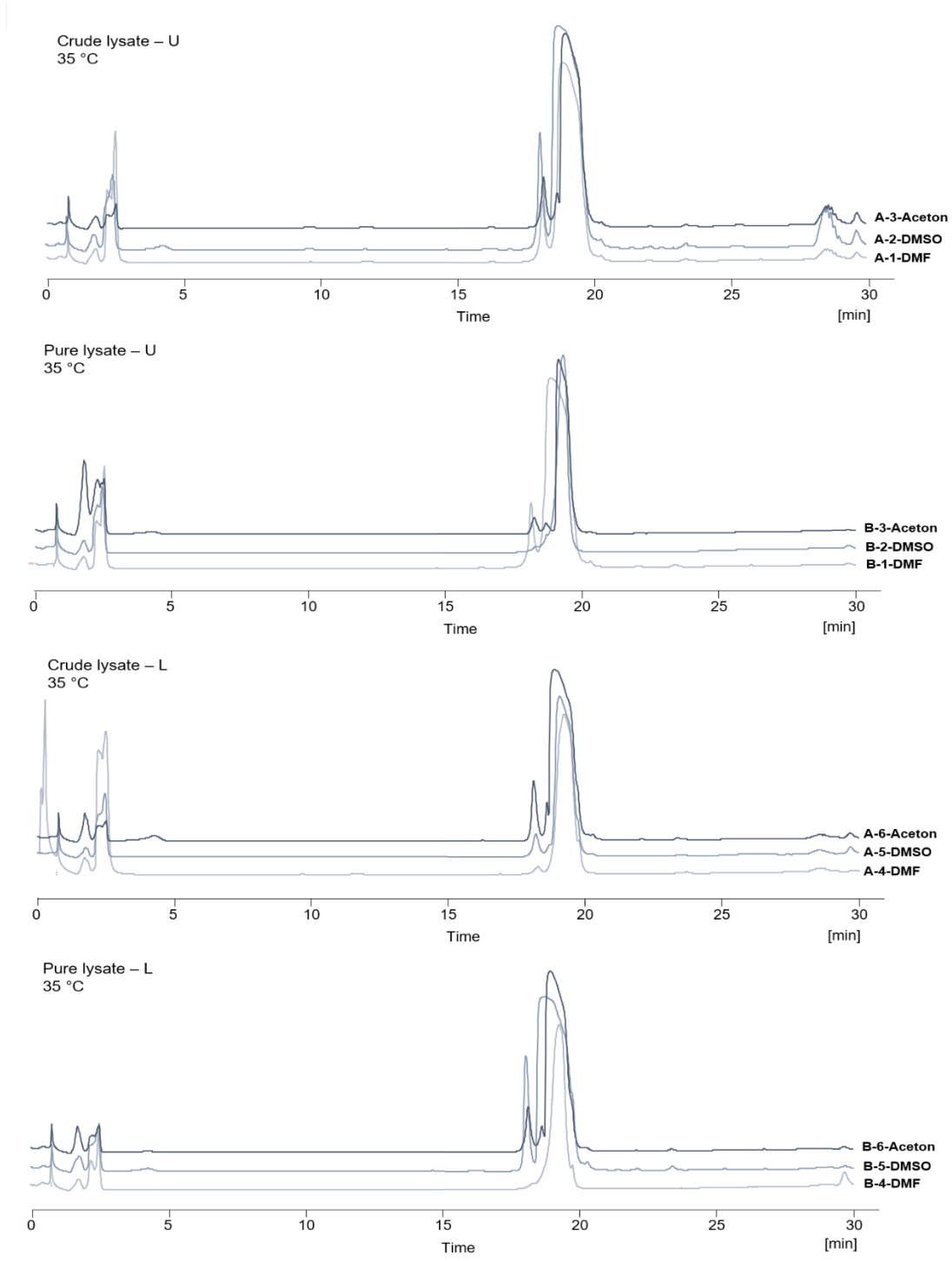


Figure 66. Performance of enzyme catalyzed acyl-CoA biosynthesis using FadD-containing lysate. Tests were executed using a variety of fatty acids like lauric (L) or undecanoic (U) acid as presented. Alteration of solvent for substrate dilution and impact on assay outcome was tested. A: Crude lysate. B: purified lysate.

In summary, testing of FadD acyl-CoA synthetase activity based on literature-published *in vitro* experiments could not be reproduced in multiple attempts and despite alteration of various reaction parameters (fatty acid type, reaction temperature, buffer and pH, incubation time, protein sample type, cosolvent). Acyl-CoA production was not detected by LC-MS analysis in any case, most likely due to nonfunction of FadD. MS-MS analysis validated the presence of recombinantly expressed characteristic FadD since *E. coli* produce additional enzymes in this size range of 60 kDa and a false protein could have been tested by accident (Figure S12).

3.3.6.2 Generation of PA pathway intermediates by NRPS catalysis

The initial reaction step in several bacterial pathways involved in PA biosynthesis is catalyzed by multimodular NRPS systems. As previously mentioned, these megasynthetases are extraordinarily complex and within each module they are further subdivided into domains. The essential domains (A, C, PCP) of each module work together to identify, activate, modify, and incorporate one specific building block (an amino acid) into the growing polypeptide chain. NRPS systems involved in PA biosynthesis are often bimodular systems which couple two amino acids and a fatty acid thioester, thereby leading to formation of [5,6]-bicyclic systems, such as intermediates **63** – **67** for pyrrolizixenamide and brabantamide production. Characterization of entire NRPS systems has proven to be difficult since these large, multimodular enzymes are expressed only in small amounts and under specific conditions by their native producers. To overcome native host limitations, recombinant protein overexpression in heterologous hosts such as *E. coli* has become an attractive alternative to obtain increased protein yields, thereby enabling detailed analysis of these enzymes.[163] Nevertheless, heterologous expression attempts to produce large amounts of protein have remained challenging which could also be observed for the selected bacterial NRPSs expressed within the course of this thesis. Expression tests showed successful protein production only in the presence of a C-terminal fusion tag (section 3.3.4.1). Furthermore, issues concerning NRPS activation *in vitro* have been reported, such as a highly inefficient *apo*- to *holo*-conversion (1-2 % of total protein) and thus lack of catalytic activity in conducted enzyme activity assays.[163] 4'-PPTases such as Sfp from *Bacillus subtilis* posttranslationally modify the *apo*-NRPS system by loading of the ppan co-factor onto the serine residue of the PCP domain. *In vivo*, this priming reaction is much more efficient than *in vitro*, especially when the modifying 4'-PPTase is co-expressed with the NRPS of choice, for instance in *E. coli* BAP1 cells. Additionally, for assay performance and concentration calculation it is important to keep in mind that the actual proportion of active enzyme is significantly lower than the inactive

protein amount within the whole protein sample. Depending on the protein batch, proportions can vary by a factor of three.[163]

Based on these findings described by Linne and Marahiel[163], NRPS enzyme activity assays were performed using either recombinantly expressed NRPS systems primed by *in vivo* co-expression with Sfp or by *in vitro* activation with purified Sfp protein. The prepared reaction mixtures contained essential co-factors and co-substrates needed for full functionality of the single domains. ATP was needed to activate substrates and permit substrate loading catalyzed by the A domains whereas $MgCl_2$ was supplied to ensure full functionality of 4'-PPTase at the PCP domain (Table 5).[164]

Table 5. Reaction assay setup for NRPS-catalyzed PA intermediate biosynthesis.

Components	Stock concentration in [µM] or [mM]	Final concentration in [µM] or [mM]
Substrate*	100 mM	10 mM
NADPH**	100 mM	10 mM
FAD^+**	20 mM	2 mM
$MgCl_2$	100 mM	7.5 mM
ATP	100 mM	10 mM
L-Serine	100 mM	2 mM
L-Proline	100 mM	2 mM
Coenzyme A	100 mM	2 mM
NRPS	20 – 40 µM	10 µM
Sfp	100 µM	5 µM
Tris	50 mM Tris, 10 % glycerol, pH 7.5	

* Lauroyl-SNAc (**74**), **68** and **72** dissolved in DMSO or acetone. ** Not obligatory for catalytic activity.

In section 3.3.4.1 protein expression and optimization of NRPS systems such as PxaA, PAO1C29 NRPS, SFC2 NRPS and SMg1C5 NRPS had been thematized with the result of respective proteins being expressed in low amounts of 3 – 5 mg per Liter expression culture and still containing impurities due to presence of native *E. coli* proteins or truncated protein versions after purification. MS-MS analysis of the purified protein was performed prior to utilization in experiments to confirm the identity of the respective NRPS system (Figure S13). Enzyme activity assays were executed using recombinant proteins in combination with chemically synthesized lauroyl-SNAc (**74**) (section 3.4.2.4) as substrate.

Enzyme activity assays performed with each of the four purified NRPSs failed to couple amino acid and fatty acid thioester substrates and generate the [5,6]-bicyclic PA intermediate as only the presence of the educt **74** was seen in LC-MS measurements (Figure 67). Whereas multiple

molecular ion adducts of the substrate mass 301.21 g/mol were detected in all enzyme assay samples, not a single assay confirmed formation of intermediate **67** with a mass of 348.26 g/mol. Negative assay results could be linked to multiple issues arising from enzyme preparation as well as assay setup and conduction. As mentioned earlier, *in vitro* posttranslational modification of the PCP domain serine residue catalyzed by Sfp has proven to be inefficient, thus the main proportion of present NRPS would have stayed in the inactive *apo*-form and was incapable of processing **74**. Furthermore, the SNAc moiety may cause problems concerning substrate binding in the NRPSs' active site due to its significantly different spatial orientation compared to the natural substrate acyl-CoA thioester.[165] Lastly, initial test assays may have been unsuccessful because of stability issues of the protein as a decisive amount of protein aggregation could be observed after overnight incubation at 28 – 30 °C. Protein misfolding and the presence of the C-terminal His_8-tag were also aspects to consider for full functionality of the native protein.[166-167]

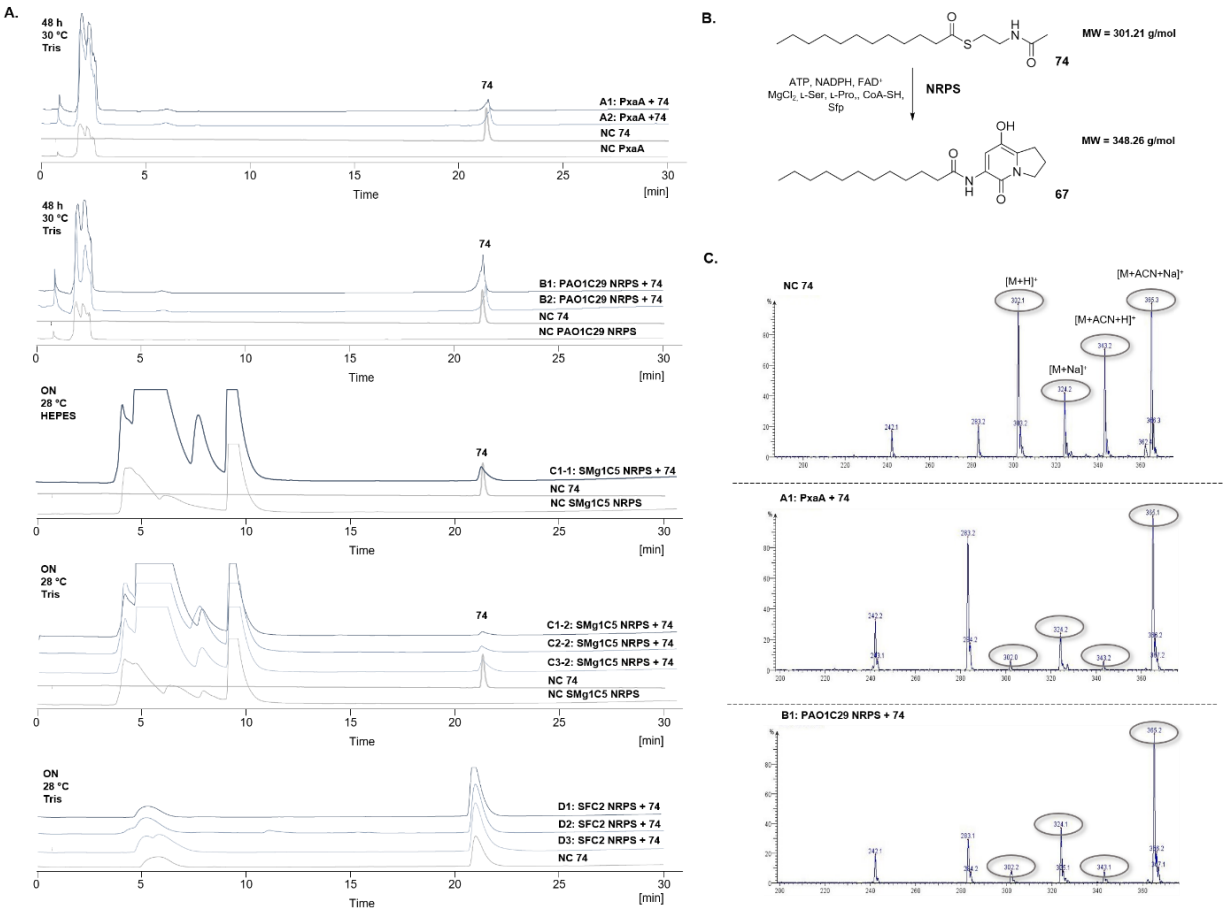


Figure 67. Overview of conversion assays with **74** catalyzed by a selection of NRPS homologues. **A.** HPLC chromatograms of performed catalytic activity tests show a single peak at t_R=21.3 min corresponding to the SNAc-thioester substrate **74**. Both HEPES and Tris buffer were tested when using SMg1C5 NRPS. Optimal assay reaction temperatures were chosen between 28 °C and 30 °C. **B.** Scheme illustrating *in vitro* conversion of **74** to PA intermediate **67** by NRPS megasynthetase in the presence of co-factors and co-substrates. Molecular weight of both **74** and expected product **67** are given. **C.** Representative mass spectra of processed enzyme assay samples of PxaA and PAO1C29 NRPS-catalyzed reactions after overnight incubation shown in comparison with negative control lacking NRPS protein.

Several alternative approaches were tested to overcome insufficient posttranslational modification, including *in vivo* co-expression of NRPS systems with Sfp in *E. coli* BAP1 cells, *in vitro* NRPS priming in a separate reaction (Table 6) prior to the standard reaction assay as well as combinations of *in vivo-in vitro* and *in vitro-in vitro* priming. Attachment of the ppan-arm before initiating NRPS-catalysis in an extra reaction setup was thought to increase *holo*-conformation abundance compared to simultaneous priming and catalysis since additional reaction components may impair Sfp function and thus block activation activity.

Table 6. Assay setup for Sfp-catalyzed NRPS priming reaction.

Components	Stock concentration in [µM] or [mM]	Final concentration in [µM] or [mM]
$MgCl_2$	100 mM	10 mM
DTT	100 mM	5 mM
Coenzyme A	1 mM	11– 24 µM*
Sfp	4 µM	0.1 µM
NRPS	20 – 40 µM	10 – 20 µM
HEPES	75 mM HEPES, pH 7.0	

* Coenzyme A (**13**) added in 10 – 20 % excess compared to NRPS protein

Besides unsuccessful posttranslational modification, the issue of substrate acceptance based on thioester type was addressed. In Nature, fatty acids are made reactive by coupling them with coenzyme A (**13**) to form an acyl-CoA thioester bearing variable hydrocarbon chain lengths. Thus, chemical synthesis of **68** and **72** (see section 3.4.1) and subsequent utilization in place of the SNAc mimic **74** for NRPS activity assays was tested to eliminate possible difficulties which could have arisen here.

Regarding acquired test data, none of the adjustments could help achieve a successful assay outcome resulting in biosynthesis of the PA intermediate bearing a C_{12} hydrocarbon moiety. Although *m/z* ratios of some molecular ion adducts of desired intermediates **66** and **67** were detected (e.g., $[M+NH_4]^+$, $[M+CH_3OH+H]^+$, $[M+2ACN+H]^+$), comparison with the pure acyl-CoA substrates revealed the presence of ions carrying similar molecular weights. Thus, the detected ions most likely result from the reaction mixture instead of synthesized product (Figure 68). Even application of freshly expressed Sfp protein R4-4 with a 300-fold increased PCP modification rate compared to wildtype Sfp[168] did not help to overcome these priming issues. In addition, a novel problem may have occurred during *in vivo* co-expression. Most likely, attachment of the ppan-moiety was successful, however in the following, *apo*-NRPS loaded intracellular acetyl-CoA, a molecule involved in many metabolic pathways of the cell, and the PCP domain was therefore blocked for subsequent peptide biosynthesis, a process called mispriming.[169] In this case, removal of acetyl-CoA would be necessary prior to *in vitro* testing when repeating protein expression using in *vivo* posttranslational modification.

Apart from inefficient priming, non-functional protein could have been obtained after protein purification due to misfolding. Presence of the C-terminal His_8 affinity tag may influence NRPS stability and furthermore prevent native protein conformation needed for full activity.[170-171] Based on these literature-published observations, removal of the His_8-tag by TEV protease cleavage following protein purification and concentration was intended to obtain the native, functional NRPS conformation.

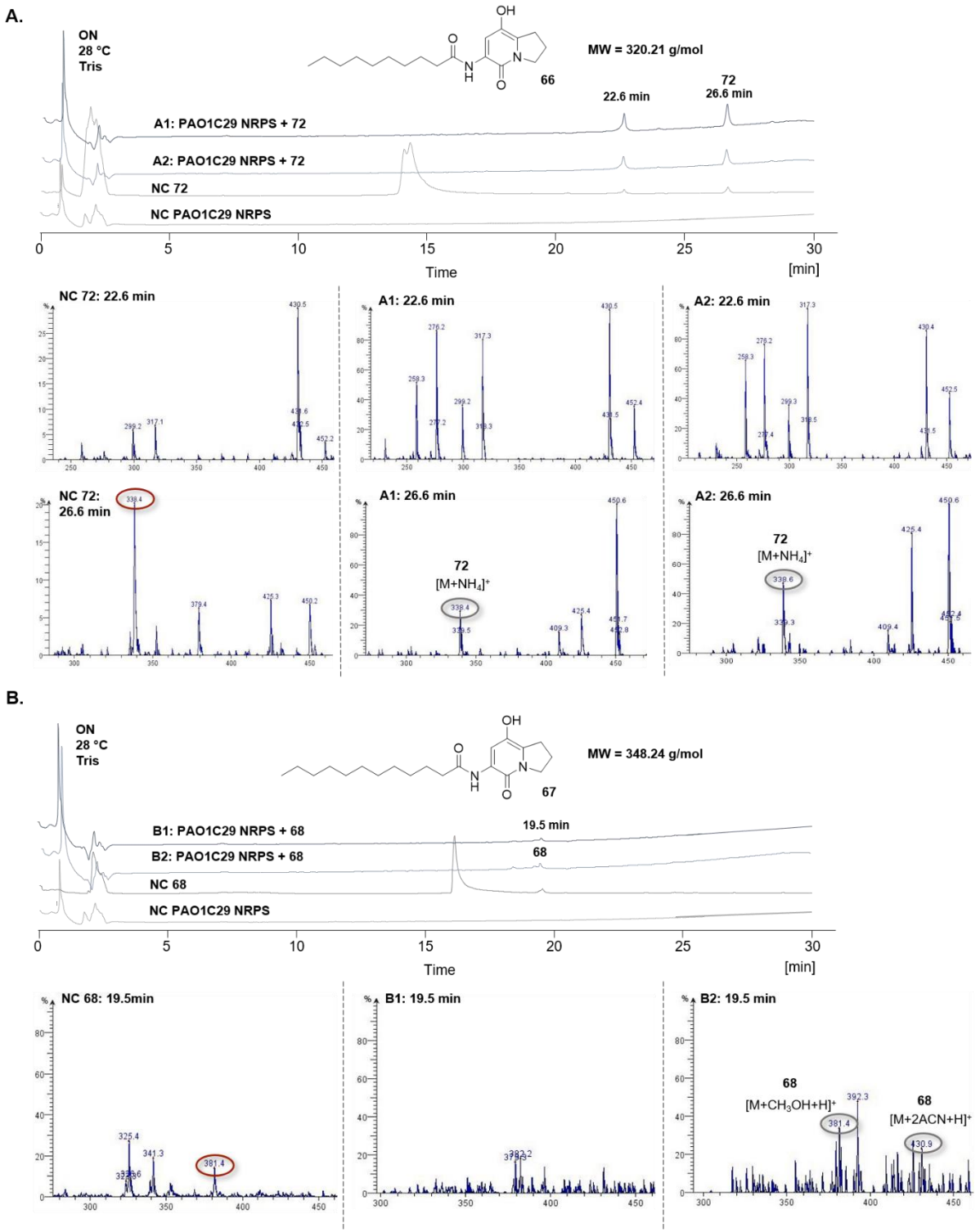
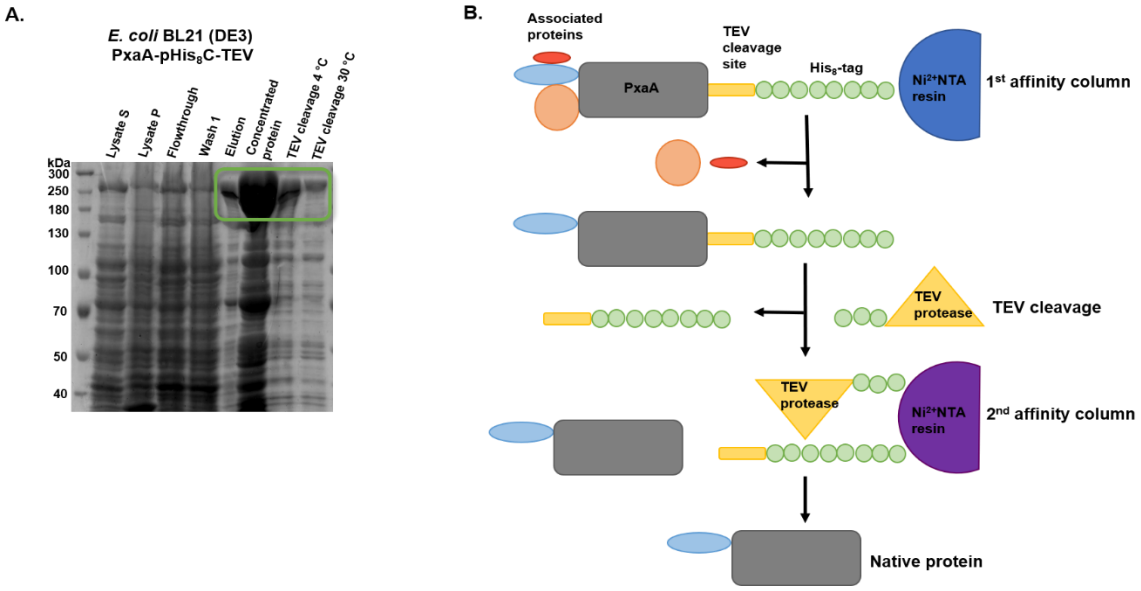


Figure 68. *In vivo* plus *in vitro* Sfp-catalyzed NRPS priming to achieve PA intermediate biosynthesis. Expression construct was first co-expressed with Sfp in *E. coli* BAP1 and additionally primed *in vitro* after protein purification. LC-MS data presented for **A.** duplicate tests of PAO1C29 NRPS-catalyzed conversion of **72** and **B.** biosynthesis of PA intermediate **67** coming from substrate **68**. Molecular structure and weight of desired intermediate products **66** and **67** noted.

According to the SDS-gel presented in Scheme 16A, C-terminal cleavage of the His_8-tag using TEV protease did not have an impact on overall protein solubility. Affinity tag removal was tested both at 4 °C and 30 °C, with the low temperature preventing high activity rates of TEV protease but simultaneously maintaining protein thermal instability during cleavage. TEV protease possesses highest catalytic efficiency at 30 °C. Therefore, removal of the His_8-tag was also performed at high temperatures with the risk of NRPS degradation to occur due to thermal instability. After incubation with TEV protease for 1 ½ hours at the respective incubation temperature, in both cases TEV protease and standalone His_8-tag were successfully removed using Ni^{2+}NTA resin. Whereas the affinity tag as well as the commercially available TEV protease bearing an N-terminal His_6-tag were retained on the column, native NRPS protein was eluted with the flowthrough fraction and directly concentrated for enzyme assay usage (Scheme 16B).



Scheme 16. SDS-gel of purified PxaA protein before and after removal of His_8-tag via TEV protease cleavage. **A.** Last three lanes show PxaA fraction after initial purification as well as after TEV protease-mediated tag removal with subsequent purification and concentration at two different cleavage temperatures. **B.** Schematic illustration of TEV cleavage with subsequent removal of TEV protease and affinity tag via column chromatography.

Since measured HPLC chromatograms were empty for all performed test reactions, PA intermediate biosynthesis had once more failed (Figure 69). Although still soluble after three hours incubation time, protein components had denatured and aggregated overnight, thus hinting at stability issues and a non-optimal reaction environment. Lack of detecting acyl-CoA substrate was due to the sample preparation method for LC-MS analysis. Whereas acyl-CoA

derivatives are soluble in water, the [5,6]-bicyclic PA intermediates require an organic solvent such as MeOH or EtOAc for extraction from the reaction mixture. Accordingly, only reaction product should have been present in the measured organic phase of all test assays and product absence was therefore linked to reaction failure.

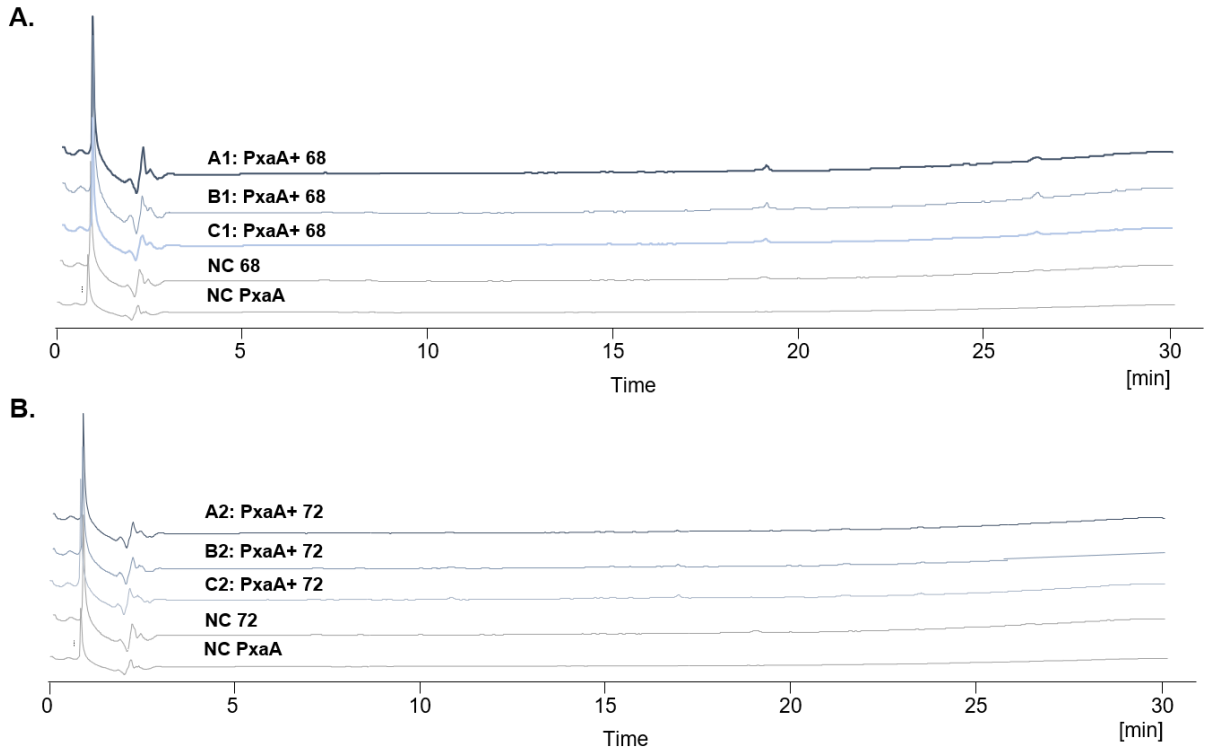


Figure 69. Comparison of enzyme assay outcome before and after affinity tag removal. Reactions were conducted with either purified protein still bearing C-terminal His_8-tag (**A**) or purified native protein (**B**: 4 °C; C: 30 °C, see scheme 14A lanes 7 and 8) after TEV-protease mediated affinity tag cleavage. Test assays were supplemented with **68** or **72**.

Addressing the most important challenges one by one and in combination with each other unfortunately did not result in obtaining a fully functional NRPS protein and therefore subsequent *in vitro* conversion of adequate substrates remained unsuccessful using purified protein. Further tests were performed using cleared cell lysate or lysate containing cell debris to examine whether the presence of additional proteins or cellular components could have a stabilizing effect and assist in formation of the native NRPS system (Figure 70).

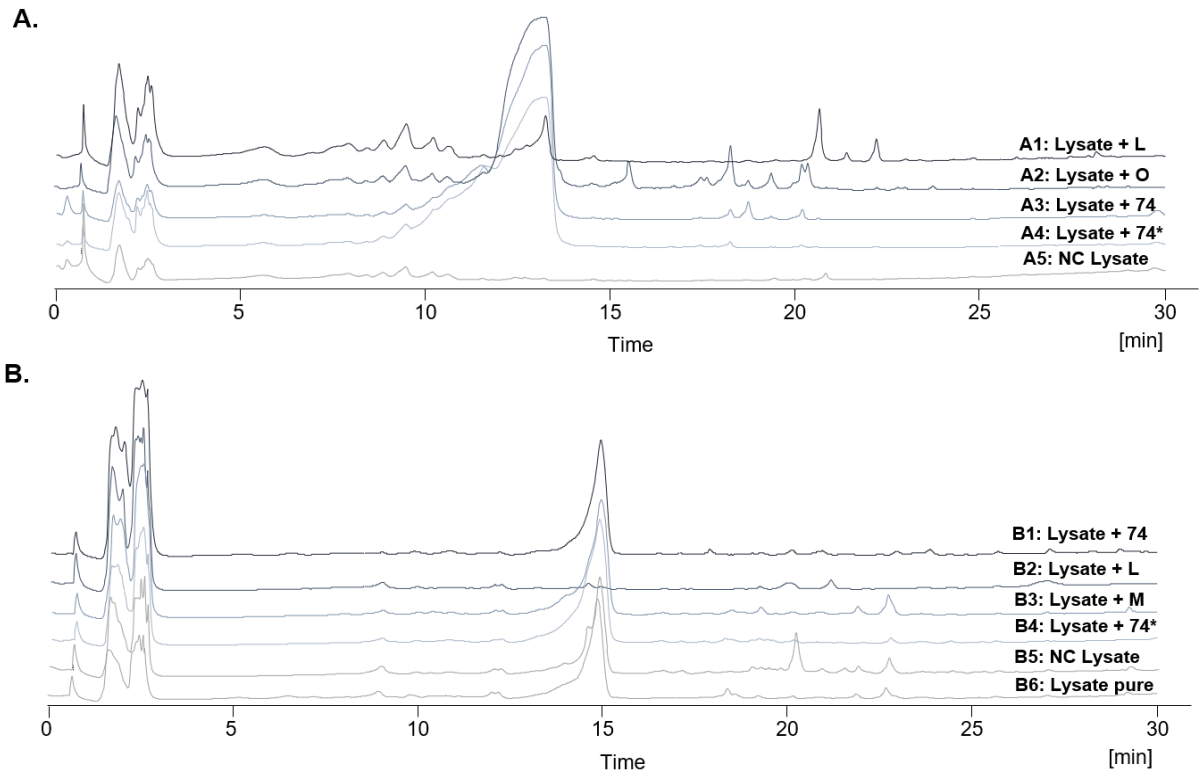


Figure 70. Testing of NRPS-catalyzed reaction using crude cell lysate. **A.** Cleared cell lysate was supplemented with co-factors, co-substrates, and substrates (L-serine, L-proline, fatty acid/ ketoacyl-thioester). Nonactivated fatty acids lauric acid (A1; L) and octanoic acid (A2; O) were added as an alternative to the activated **74** (A3 and A4). Furthermore, a test reaction was performed without co-factors and co-substrates but instead only providing amino acid and fatty acid thioester educts (A4). Cleared lysate supplemented with above mentioned additives lacking the fatty acid component was used as a negative control. **B.** Crude cell lysate still containing cell debris after lysis was treated with all reaction components. Alongside **74** (B1) and lauric acid (B2), β-hydroxy myristic acid (B3; M) was added. One test (B4) was performed using only the minimal set of supplements reduced to direct reaction substrates. Negative controls were prepared without fatty acid components, with one sample containing all other reaction components (B5) and another consisting of only crude cell debris lysate (B6). Assays were incubated over night at 28 °C while shaking.

As for the enzyme activity tests using purified protein, neither cleared lysate nor cell debris assays led to biosynthesis of the desired [5,6]-bicyclic intermediate. Both series of measurements showed one broad peak at retention times of 13 to 15 min which can be associated with the elution of diverse cellular proteins and components. Putative remnants of fatty acid substrates were detected as less intense peaks eluting at retention times later than 17 min with over 58 % ACN percentage in addition to further lysate elements. Whereas acyl-CoA and acyl-SNAc thioesters were necessary substrates for conversion assays using purified protein, feeding of nonactivated fatty acids was considered an alternative in crude assays since

native *E. coli* enzymes bearing an acyl-CoA synthetase function are likely to be present and active. However, the formation of either **68** (A1) or octanoyl-CoA (**75**; A2) from inactive fatty acids was not observed and therefore the hypothesis of *in vitro* fatty acid activation for subsequent incorporation by the NRPS system could not be validated because the test reactions in general did not work.

All in all, *in vitro* reconstitution of the initial NRPS-catalyzed reaction in the PA-related biosynthetic pathway was not achieved. Although recombinant protein was obtained in sufficient amounts for testing, protein purification and tag removal did not help to receive a fully functional NRPS system capable of executing the designated catalytic steps. Attempts to tackle these challenges included optimization of PCP-domain priming to receive higher amounts of *holo*-protein, chemical synthesis of thioester substrates also occurring in Nature as well as crude lysate testing to increase chances of obtaining correctly folded, thermally robust, and active protein. Hence, to overcome these difficulties, extensive investigations on the NRPS protein itself as well as finding optimal reaction conditions need to be made in the future.

3.3.6.3 Expanding the substrate scope of the monomodular NRPS pys

Klapper et al. identified a set of molecules produced by *Pseudomonas fluorescens* HKI0770 which exhibit amoebicidal activity.[172] Apart from a tetramic acid functionality, these so-called pyreudiones (**38**) exhibit the typical pyrrolizidine core structure common to all members of the PA class. Compared to the bimodular NRPS systems encoded in PA-related BGCs predicted to produce pyrrolizixenamide-like molecules, variants of **38** are synthesized by the standalone monomodular NRPS pys. The A-domain adenylates L-proline and promotes aminoacyl adenylate loading onto the ppan moiety of the PCP domain where C-domain catalyzed condensation with a β-ketoacyl thioester results in amide bond formation. Finally, a terminal TE-domain facilitates pyreudione product release by Dieckmann cyclization.[172] *In vivo* expressions have proven to be successful, with PCP-domain priming being achieved through co-expression of the PPTase sfp from *Bacillus subtilis* alongside the protein of interest as described in section 3.2.2.3.1. Furthermore, experiments showed that substrate acyl precursors are metabolites of fatty acid primary metabolism and there is no need for heterologous expression of additional enzymes *in vivo*.[173]

As for the recombinantly expressed bimodular NRPS systems presented in section 3.3.5.2, *in vitro* reconstitution of synthesis of **38** was expected to be of a challenge, in particular concerning posttranslational modification of the PCP-domain serine side chain hydroxyl group. Sfp-induced PCP-domain priming experiments performed during this thesis were highly inefficient since most of the protein remained in its inactive *apo*-form. To ensure efficient ppan-loading, pys PCP-domain posttranslational modification was visualized by site-specific labeling of the PCP domain at the serine hydroxyl function with a fluorescent dye-conjugated ppan moiety (Table 7). The fluorescent probe was prepared by chemical conjugation of **13** with the thiol reactive tetramethylrhodamine-5-(and -6)-*C*2 maleimide (TMR-Mal; **103**) as described in section 3.4.2.5 to obtain TMR-Mal-CoA (**76**).[174]

Table 7. Pys PCP-domain loading assay using TMR-Mal-CoA fluorescent probe (**76**).

Components	**Stock concentration [µM] or [mM]**	**Final concentration [µM] or [mM]**
Pys *apo*-PCP-domain	200 µM	100 µM
TMR-Mal-CoA	420 µM	150 µM
Sfp	50 µM	0.5 µM
$MgCl_2$	100 mM	10 mM
Tris	50 mM Tris, pH 7.4	

Domain labeling was monitored by LC-MS analysis. However, empty HPLC chromatograms and absence of mass ions hinting at TMR-Mal-CoA or TMR-Mal-ppan-PCP-domain presence for all test samples were unsatisfactory and did not confirm a successful sfp-initiated priming reaction. The lack of signals can be associated with sample preparation for LC-MS measurement, since the reaction sample was filtered to remove particles and aggregates which could clog both column and ESI source, presumably also holding back labeled PCP-domain. With no fluorescence imaging devices in hand, a simple but efficient method for visualizing TMR-Mal-CoA-ppt-PCP-domain conjugation was established. Total test reaction volumes were applied to prepacked Ni^{2+}NTA spin columns (2 mL, Qiagen) to allow binding of the N-terminally His_6-tagged PCP-domain. Excess unbound **76** and other assay components were removed with wash buffers 1 and 3 also used for standard protein purification procedures. Pink coloring of Ni^{2+}NTA resin after column washing was evidence for successful labeling of the colorless PCP-domain because **76** is only prevented from elution when bound to the His_6-tagged protein. Detachment of protein from the resin with elution buffer resulted in a pink colored eluate, giving further evidence of TMR-Mal-CoA.ppt-PCP-domain occurrence.

Nevertheless, MS-MS measurements of the primed PCP-domain were executed to confirm these findings (Figures S14 – S15). Looking at the secondary structure of the pys PCP-domain, two serines at positions 984 and 992 may be subject to posttranslational modification. Not only did MS^2 measurements support visual evidence of successful PCP-domain loading, but site-specific labeling identified serine 984 to be the position where posttranslational modification takes place.

With studies validating efficient pys NRPS activation by sfp-catalyzed *apo*- to *holo*-conversion, a focus was set to investigating β-ketoacid substrate selectivity of the C-domain. Due to biosynthesis of multiple pyreudione derivatives of varying alkyl chain length by the same monomodular system, the pys C-domain was predicted to exhibit a broad substrate tolerance.[172] Chemical synthesis of multiple β-ketoacyl thioesters with altering alkyl chain length (C_6-C_{10}) and degree of saturation (diene and alkyne) as well as differences in thioester functional group (SNAc, thiophenol, *tert*-butylthiol) was done in our laboratory by Manuel Einsiedler (Figure 71).

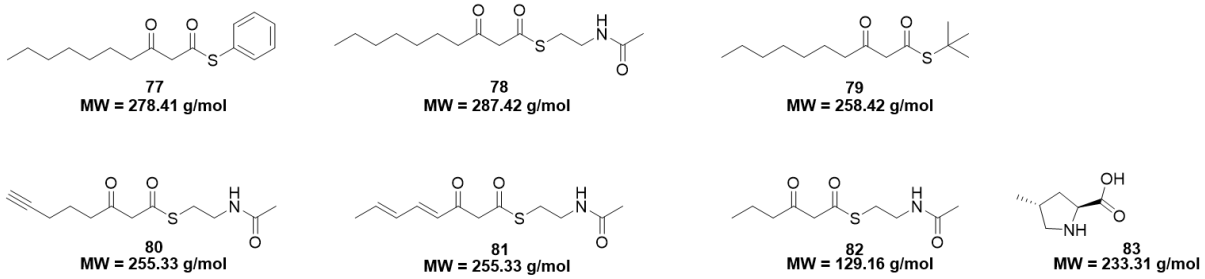


Figure 71. Chemically synthesized β-ketoacyl thioester substrates as well as modified amino acid L-proline precursor for studies on biosynthesis of **38** derivatives.

As for the performed bimodular NRPS enzyme activity assays, which included all necessary substrates as well as co-factors and co-substrates, Pys protein activity assays were set up as described in Table 8.

Table 8. Test reaction setup for pys NRPS-mediated derivative biosynthesis of **38**.

Components	Stock concentration in [µM] or [mM]	Final concentration in [µM] or [mM]
Substrate*	100 mM	2 mM
NADPH**	100 mM	10 mM
FAD^{+}**	20 mM	2 mM
$MgCl_2$	100 mM	7.5 mM
ATP	100 mM	10 mM
L-Serine	100 mM	2 mM
L-Proline	100 mM	2 mM
Coenzyme A	100 mM	2 mM
NRPS	20 – 40 µM	10 µM
Sfp	100 µM	5 µM
HEPES	100 mM HEPES, 1.5 % glycerol, pH 7.8	

* Substrates **77** – **83.** ** Not obligatory for catalytic activity.

Enzyme activity assays were incubated overnight at 28 °C – 30 °C while shaking at 300 rpm. Reactions were quenched with methanol and the supernatant was analyzed by LC-MS measurements in search of pyreudione products (Figure 72).

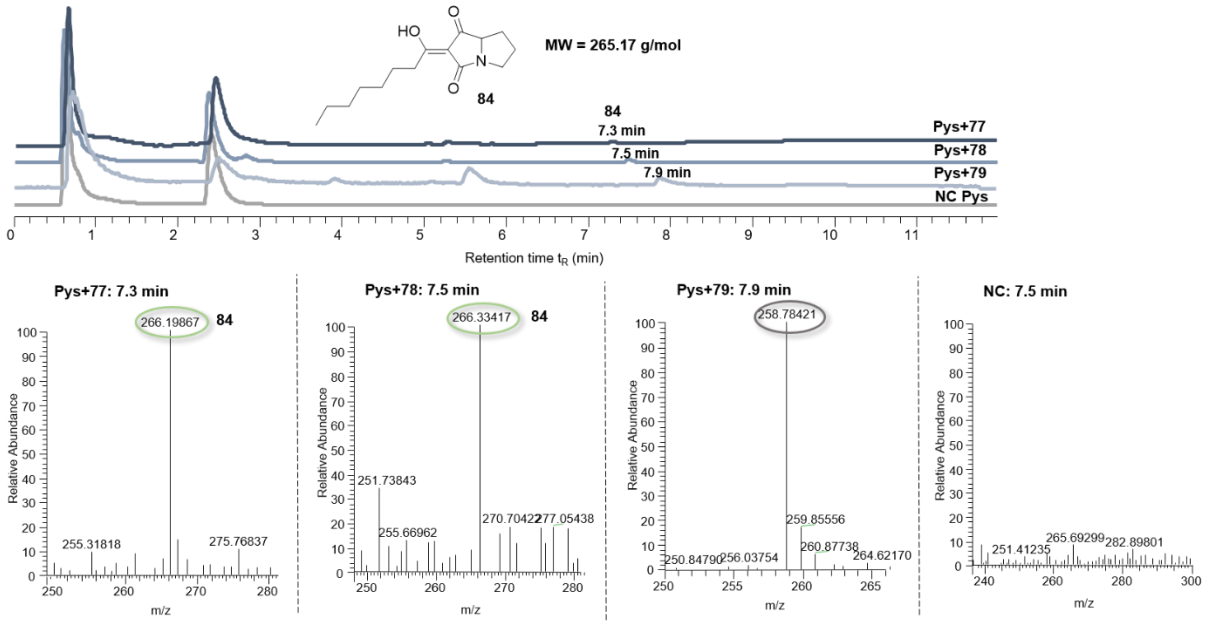


Figure 72. Pys-catalyzed pyreudione derivative biosynthesis using precursor molecules **77**, **78** and **79** bearing different reactive thioester moieties. LC-MS data display formation of product **84** upon conversion of **77** and **78**, whereas only educt was detected when applying substrate **79**.

Precursors possessing C_{10}-saturated acyl side chain moieties were incorporated by the NRPS system when they were present as reactive thiophenol (**77**) or SNAc (**78**) thioesters (Figure 72). In contrast, formation of **84** was not observed for substrate **79** activated as a *tert*-butylthiol, most likely due to the bulkiness of the thioester group which is known to exhibit steric hindrance and lead to conformational change of its binding partners (Figure 71).[175] The large areal extent may have prohibited substrate binding in the enzymes' active site or alternatively binding was possible, however upon binding the active site protein conformation was decisively changed and thus may have initiated loss of function.

Conversion of substrate **82** carrying a saturated but shorter C_6 hydrocarbon moiety yielded pyreudione E (**85**), a derivative lacking amoebicidal activity presumably due to its shorter side chain.[173] The combination of **78** with the amino acid precursor **83** was tested to investigate whether C-domain catalyzed condensation of the aminoacyl adenylate with the β-ketoacyl thioester is restricted by simultaneous modification of both components. Incorporation of proline analogues has previously been studied for the pys A-domain, revealing successful pyreudione formation based on L-pipecolic acid and L-hydroxyproline, whereas all other L-amino acids were not accepted as alternative substrates.[173] Analogous to standard incorporation of L-proline, **83** was successfully coupled to the other basic building blocks to generate the methylated pyreudione product **86** (Figure 73).

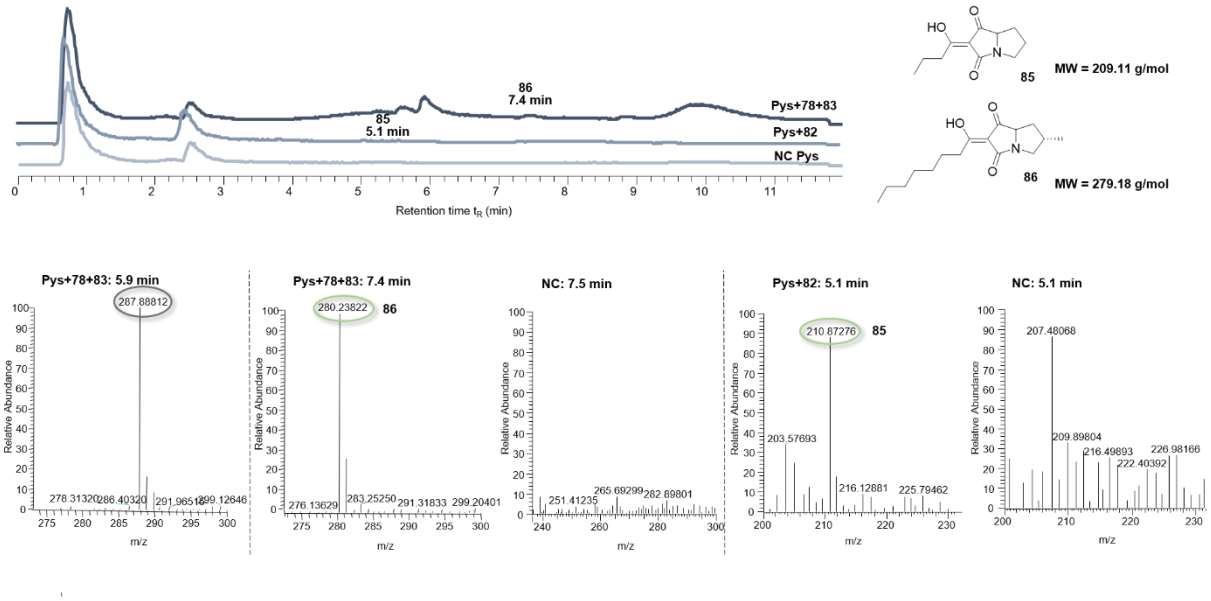


Figure 73. Generation of novel pyreudione derivatives from modified amino acid and β-ketoacyl precursors. Comparison to negative control LC-MS data confirmed product formation in both cases.

These results show that the pys NRPS system has promiscuity towards substrates of different acyl side chain length as well as a modified version of the amino acid building block. Furthermore, spatial orientation and steric hindrance of the activating thioester seem to play a role for substrate binding in the enzymes' catalytic pocket and subsequent conversion. The substrate scope was expanded towards alteration of the side chain moiety based on these preliminary data.

In contrast to precursor molecules **77** - **79** and **83**, substrates carrying unsaturated alkyl side chains were not incorporated into pyreudiones (products **87** and **88**; Figure 74). Although corresponding product ion masses of *m/z* = 234 were detected for both alkyne (**80**) and diene (**81**) alongside the substrates at *m/z* = 256, the negative control run also exhibited these product ions in even higher relative abundance, thus concluding that the desired product was absent in the test reactions. Lack of conversion can presumably be explained by alteration of the interaction between binding pocket residues and the respective substrates. Due to the presence of double/triple bonds, less hydrogen bridge interactions can be formed, which play a role in positioning of the substrate inside the enzyme pocket for subsequent conversion. Furthermore, higher electron densities within the substrate molecule may alter interactions with active site regions and influence the conformation of the substrate in relation to the PCP-domain where the aminoacyl adenylate is attached. To enable condensation and amide bond formation, spatial proximity of the molecules to be coupled is of importance, a criterium which might remain unfulfilled for alkene and alkyne substrates.

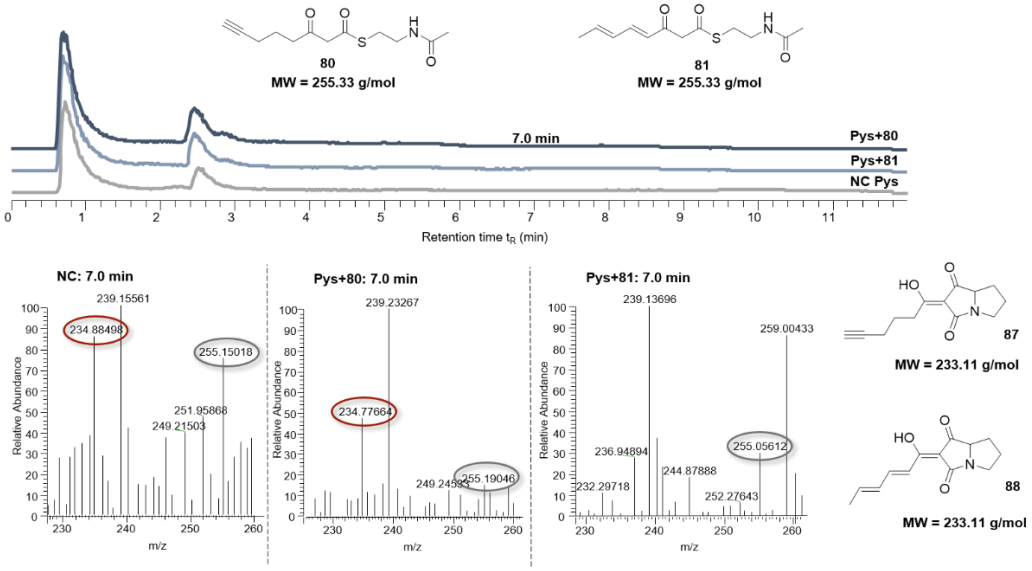


Figure 74. LC-MS analysis of pys NRPS-catalyzed reaction using substrates with unsaturated hydrocarbon side chains. Both diene and alkyne conversions were compared to a negative control lacking protein.

Taken together, it was proven that the monomodular pys NRPS system originating from *Pseudomonas fluorescens* HKI0770 can be heterologously expressed in *E. coli*. In addition, conversion into the fully functional *holo*-form can be achieved *in vivo* as well as *in vitro*. Enzyme activity assays displayed efficient conversion of saturated β-ketoacyl thioesters of varying hydrocarbon chain length. Furthermore, L-proline derivatives were integrated into the pyreudione backbone without difficulties. Steric hindrance, as in the case of bulky *tert*-butylthiol (**79**), and putative changes in substrate positioning inside the catalytic pocket, as observed for substrates **80** and **81** possessing alkyne or alkene moieties, limited biosynthesis of **38**. Further investigations on pys substrate scope, including chemical synthesis of substrates bearing branched alkyl side chains or carrying additional functional groups (e.g., hydroxyl groups, sugars), will give more information on the mode of substrate binding and interaction with active site residues. Crystallization of the NRPS system with and without bound substrate is work in progress executed by our collaboration partners and will give further insights on the mechanism of action.

3.3.6.4 Conversion of PA pathway intermediates to 37 or 40

Studies addressing *in vitro* reconstitution of the NRPS-catalyzed primary reaction in PA core molecule biosynthesis unfortunately did not render novel insights into the mechanism of action, most likely due to nonfunctional recombinant NRPS protein (see section 3.3.5.3). Accessibility to pathway intermediates needed for studying the BVMO-catalyzed reaction mechanism was therefore facilitated by *in vivo* heterologous expression of PxaA NRPS enzyme and isolation of target intermediate molecules from crude extract by preparative HPLC. Heterocyclic ring structures carrying C_6, C_7 and C_8 side-chain moieties were obtained this way (see section 3.2.2.3.1), with intermediates **63** (C_6) and **65** (C_8) available in sufficient amounts to be subsequently applied in MOX conversion assays (Figure 75).

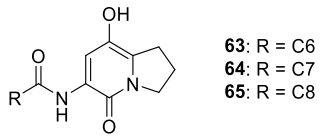


Figure 75. PA biosynthetic pathway intermediates assembled by bimodular NRPSs in an initial reaction step.

3.3.6.4.1 BVMO mechanism of action

To date, quite a few details are known about the BVMO mechanism of action. BVMOs belong to the enzyme class of oxidoreductases, which oxidize all sorts of ketones (linear, cyclic, aromatic) to their corresponding esters or lactones. The enzymes' catalytic activity depends on the presence of a flavine-containing co-factor such as FAD or FMN as well as a regeneratable electron donor system like NADPH/$NADP^+$.[176-177] In addition, aerobic conditions are necessary to initiate Baeyer-Villiger oxidation via incorporation of one oxygen atom into an existing carbon-carbon bond and simultaneous reduction of a second oxygen atom to water. BVMOs play an important role in secondary metabolism by constructing carbon scaffolds or catalyzing oxidative rearrangements and tailoring reactions.[118]

Most MOXs encoded within BGCs involved in PA-biosynthesis are type O or atypical BVMOs, multifunctional enzymes that catalyze not only a single oxidation reaction but instead are engaged in entire enzyme cascades. Like the BVMO LgnC involved in legonmycin (**39**) biosynthesis[91], enzymes PxaB (pyrrolizixenamide production) and BraC (brabantamide production) are predicted to catalyze a four-step enzymatic process. Initially, a carbamate is formed when Bayer-Villiger oxidation of the intermediate leads to incorporation of an oxygen atom, thereby promoting ring expansion. Subsequent hydrolysis of the carbamate is followed by elimination of carbon dioxide, resulting in ring contraction and generation of a novel nitrogen-carbon bond and formation of the pyrrolizidine core structure. Lastly, hydroxylation takes place to yield the respective PA derivative.[91]

To validate the BVMO-catalyzed mechanism of action leading to production of pyrrolizixenamides, conversion of intermediates **63** and **65** by PxaB as well as two additional homologues originating in *Pseudomonas aeruginosa* PAO1 cluster # 29 and *Streptomyces clavuligerus* F613-1 cluster # 22 was tested (Table 9).

Table 9. Enzyme assay reaction setup for pyrrolizixenamide production.

Components	Stock concentration [µM] or [mM]	Final concentration [µM]
Intermediate **63** or **65**	650 µM	200 µM
NADPH	1 mM	100 µM
FAD^+	1 mM	50 µM
BVMO	100 – 250 µM	10 - 20 µM*
Tris	50 mM Tris-HCl, 10 % glycerol, pH 7.5	

* final concentration dependent on protein stock concentration

In initial tests using PxaB, reactions were executed at 30 °C while shaking at 450 rpm on a tabletop shaker for 1.5 h, 3 h and overnight. Analysis of LC-MS data revealed full conversion of both intermediates **63** and **65** in all cases independent of the incubation time, yielding products **50** and **52**. Henceforth, experiments with PxaB homologues were incubated overnight out of convenience (Figure 76).

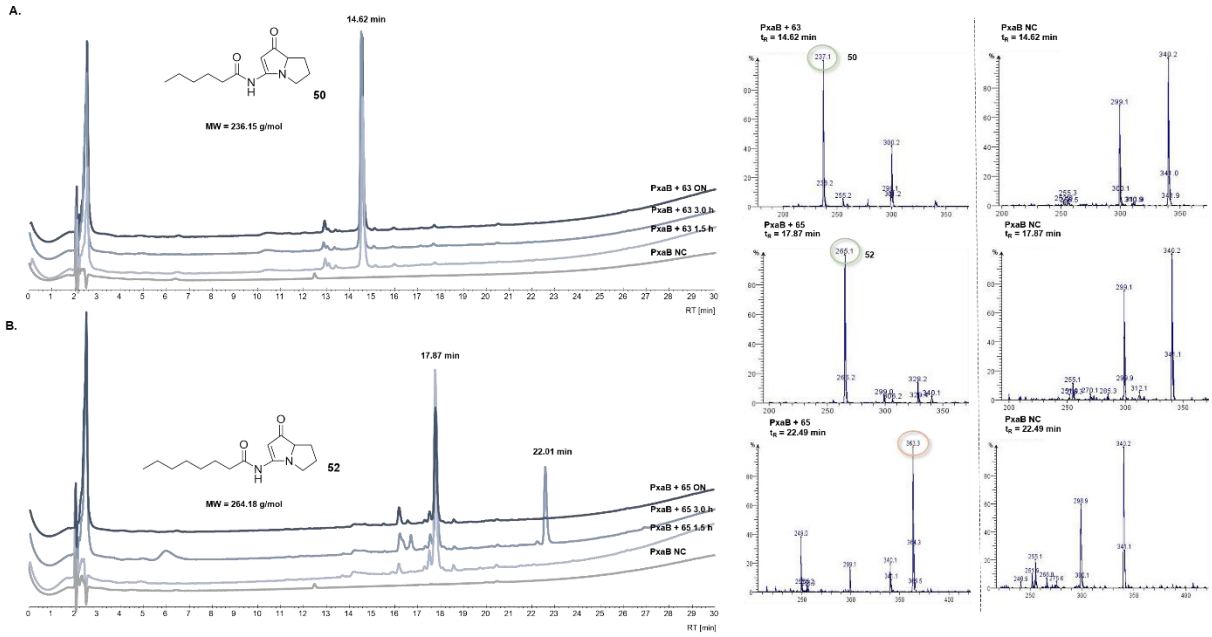


Figure 76. LC-MS confirmation of intermediate conversion by BVMO PxaB to obtain **50** and **52**. **A.** Product **50** was detected in all reaction setups independent of the reaction time for incubating at 30 °C while shaking. Comparison of sample mass spectra with a negative control lacking intermediate **63**. **B.** *In vitro* conversion of intermediate **65** yielded detectable amounts of **52**. Mass spectra support observed PxaB enzyme activity. MS measured using an Advion mass spectrometer. A 30 min HPLC run applying an ACN gradient from 5 – 100 % was chosen for measurements.

Conversion of heterocyclic intermediates was quantitative as no traces of either **63** or **65** were detected after assay workup in LC-MS measurements. Recorded *m/z* ratios of peaks representing the main component of the respective assay at retention times of t_R= 14.62 min (**50**) and t_R= 17.87 min (**52**) matched the predicted $[M+H]^+$ masses of 237.16 g/mol (C_6) and 265.19 g/mol (C8) (Figure 76). These findings validate the successful recombinant expression of a fully functional, catalytically active PxaB protein. A third peak of interest was detected for the PxaB + **65** sample left to incubate for 3 h (Figure 76B). Mass analyses indicate that a molecule with an *m/z* ratio of 363.3 was eluted after 22.49 min. Referring to intermediates **63** - **65**, these molecules simply differ in the length of their fatty acid side chain moiety by a single CH_2 group. Theoretical elongation of the side chain to C_{15} predicts an identical *m/z* ratio of

363.3 for the $[M+H]^+$ mass of the odd-chain pentadecanoic acid (C_{15})-containing PA derivative. Occurrence of this putative C_{15} side chain residue may be linked to dimerization events between two intermediate molecules **65** including decarboxylation to eliminate a single carbon atom following the final hydroxylation step in the atypical BVMO enzyme cascade, resulting in the odd-numbered side chain. Since this second peak with an unknown product was only detected once amongst the multiple reaction repetitions executed throughout the course of this thesis, insufficient amounts (only traces) of material were available for in-depth characterization. Alternatively, contamination may have occurred explicitly to the respective sample during reaction setup.

PAO1C29 MOX enzyme activity test assays were prepared analogously to PxaB tests as described in Table 9. BVMO identity was validated by MS-MS (Figures S10 – S11). The selected BVMO shares 53.43 % sequence identity with PxaB (see section 3.1.3). Therefore, it was interesting to find out whether the enzymes perform catalysis according to a common mechanism of action when given the same substrate to process. Quantitative conversion of both PA pathway intermediates **63** and **65** to pyrrolizixenamides **50** and **52** was observed when adding purified PAO1C29 MOX to the reaction mix, showing that both enzymes function identically and catalyze the same reactive steps in their respective biosynthetic pathways (Figure 77).

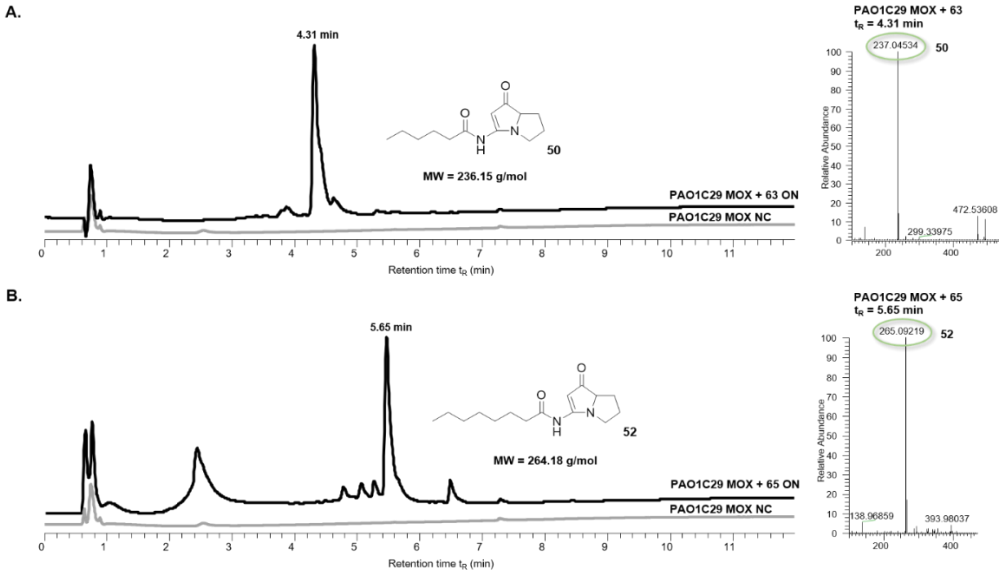


Figure 77. PAO1C29 MOX-catalyzed *in vitro* biosynthesis of **37** variants. Selected PA-intermediate molecules were successfully converted by the catalytically active BVMO from *Pseudomonas aeruginosa* PAO1, with predicted products representing the major component in the processed tests (**A:** 237.2 g/mol; **B:** 265.2 g/mol). LC-MS data was recorded with an LCQ Fleet Ion Trap mass spectrometer. A 12 min method establishing an ACN gradient of 5 – 100 % was used for product detection.

Despite differences concerning general cluster complexity, both BVMOs PxaB and PAO1C29 MOX seem to follow an identical mechanism of action when using the same intermediate molecules as substrates for the reaction they catalyze. The obtained results support published proposals which hypothesize BVMO-catalyzed conversion of PA-pathway intermediates in a multi-step assembly line to yield a variety of PA derivatives depending on the structure of the intermediate. To our knowledge, the work presented here is the first functional and mechanistic validation of BVMOs from PA pathways *in vitro*.

3.3.6.4.2 Pyrrolizixenamide production efficiency dependent on BVMO availability

As previously mentioned, remnants of intermediates **63** and **65** could not be detected by LC-MS after termination and workup of the reaction. Therefore, complete conversion can be assumed. Nevertheless, subsequent experiments were performed using varying concentrations of MOX enzyme to test concentration dependency of the reaction and collect further information considering potential scaling up of the reaction to obtain sufficient material amounts for qualitative product analysis. In the original assay 20 µM of respective protein was enough for full transformation. Concentrations of 30, 40, 50 and 100 µM protein were tested and reaction success was measured by relative abundance of the resulting product within each sample.

As can be seen in Figure 78, the relative abundance of the pyrrolizixenamide **50** ion peak (4.31 min) was monitored and compared when using varying concentrations of catalytic enzyme. Clearly, alteration of PxaB input at lower concentrations has a significant impact on product abundance based on an increase of 35 % when adding 40 µM protein instead of 30 µM. Further elevation of PxaB concentration above 40 µM resulted in comparatively small abundance increments of up to 10 %, even when doubling protein concentration, leading to the assumption that the catalytic reaction nears a state of saturation.

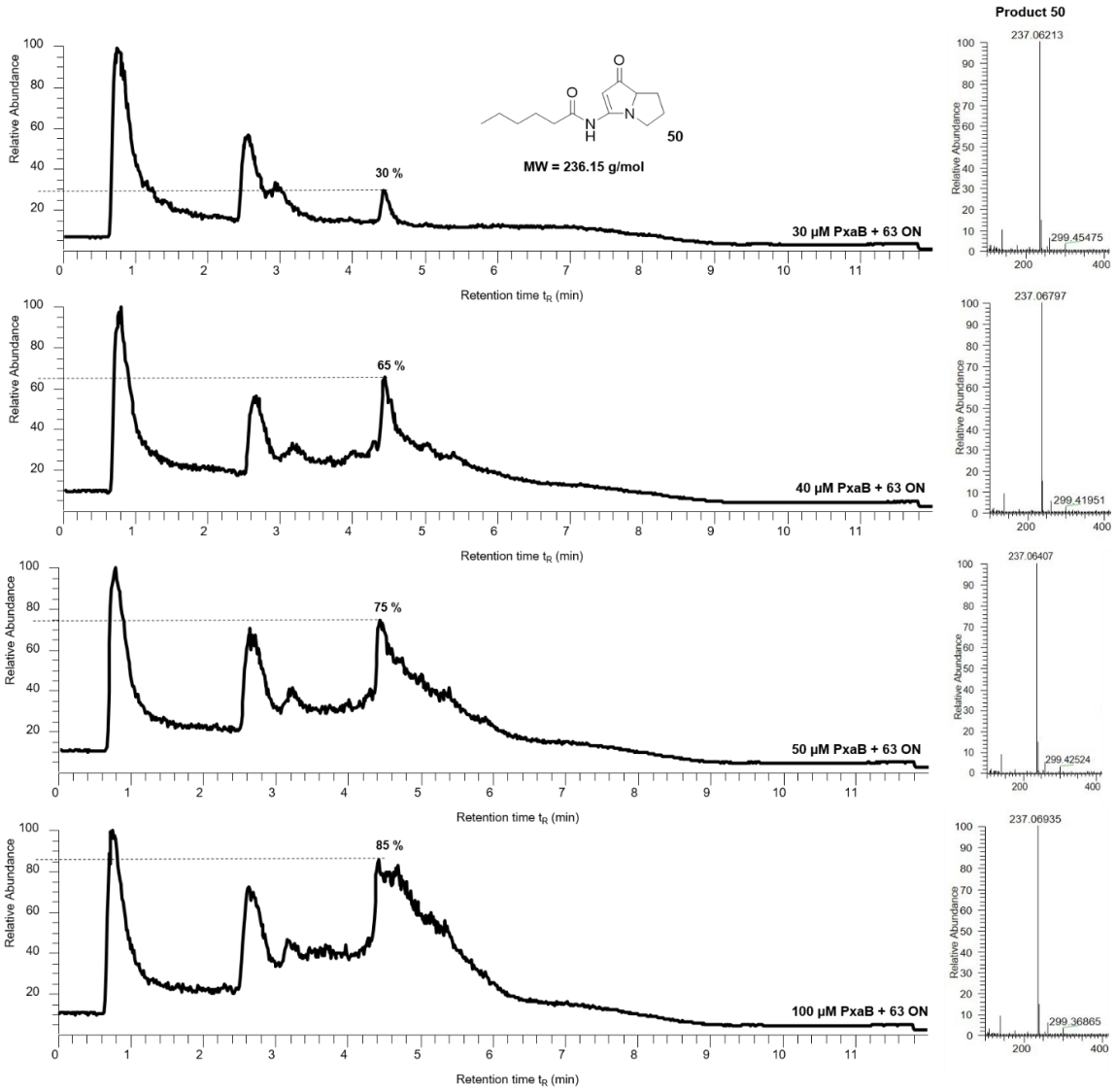


Figure 78. Quantitative analysis of reaction conversion efficiency dependent on the concentration of PxaB. MS-spectra show successful biosynthesis of **50** (m/z: 237.2) in all test assays. Relative abundance of product molecule was used as a measure for turnover efficiency.

Hypotheses to explain these findings may be based on theoretical knowledge on enzyme kinetics considering the single substrate-enzyme model described below (equation 1).

$$E + S \rightleftharpoons ES \rightleftharpoons E + P$$

Equation 1. Equation describing a first order one enzyme-one substrate enzyme kinetic model. Taken and adapted from Berg et al..[178]

Initial test reactions had proven quantitative conversion of intermediates **63** and **65** by PxaB after 1.5 h reaction time (3.3.5.4.1), leaving no traces of substrate molecule to be detected by mass spectrometry. From this, it can be assumed that an equilibrium between enzyme and substrates and the formation of an enzyme-substrate complex as well as between enzyme-substrate complex and product release is reached during overnight incubation of the reaction mixtures at low enzyme concentrations (< 10 µM). An increased protein concentration could have the effect of shifting the equilibrium in the direction of enzyme-substrate complex formation since more protein is available for binding the same amount of substrate. The equilibrium shift towards bound substrate in turn would benefit molecule conversion and product outcome, resulting in higher abundance of product ions just as detected in LC-MS measurements (Figure 78). In the case of PxaB, concentrations of 40 µM seem to alter the equilibrium in noticeable amounts towards product formation whereas the enzymatic system converges to a saturated state with concentrations of over 50 µM. However, to support these hypotheses, the mechanism of PxaB BVMO action needs to be studied in more detail in the future, not only by performing kinetic assays with substrate variations but also reviewing and elaborating the kinetic model concerning co-factor-dependent activity.[177, 179] Regarding NADPH and FAD^+, the BVMO follows a sequential ordered mechanism, where both co-factors must bind to the enzyme prior to substrate binding and must still be bound for product release to occur, thus bringing a high level of complexity into the kinetic model (equation 2).[178]

$$E + S_1 + S_2 \rightleftharpoons ES_1S_2 \rightleftharpoons E + P_1 + P_2$$

Equation 2. Sequential ordered mechanism model describing biomolecular reactions including multiple substrates. Taken and adapted from Berg et al..[178]

3.3.6.4.3 NADPH as an essential component to BVMO function

Numerous publications have studied the BVMO mechanism of action and found evidence that enzyme activity is dependent on the co-factors NADPH and FAD^+ and that the reaction follows a sequential ordered mechanism. Crystal structures have revealed a double domain architecture, with one domain involved in FAD^+-binding and the other responsible for NADPH-association. Elucidation of intermediate states has shed light on the multistep mechanism, including the necessity of NADPH to be present in the complex to reduce FAD^+ via hydride transfer. Reduced FAD species can subsequently react with molecular oxygen to form a peroxyflavine intermediate. A conformational shift of the NADPH-binding domain is needed to alter the active site and enable substrate binding.[180]

To confirm the NADPH-dependence of selected bacterial MOXs involved in PA-biosynthesis, test reactions with and without supplementation of NADPH as co-factor were prepared (Table 10).

Table 10. Reaction assay setup for testing NADPH-dependency of expressed BVMOs. NC: Negative control.

Components	**Stock concentration [µM] or [mM]**	**Final concentration [µM]**	**A**	**NC A**	**B**	**NC B**
Intermediate **63***	650 µM	200 µM	✓	×	✓	×
NADPH	1 mM	100 µM	✓	✓	×	×
FAD^+	1 mM	50 µM	✓	✓	✓	✓
BVMO	100 - 250 µM	20 µM	✓	✓	✓	✓
Tris	50 mM Tris-HCl, 10 % glycerol, pH 7.5		✓	✓	✓	✓

* dissolved in DMSO

As predicted, NADPH is obligatory to full catalytic functionality of all expressed and tested MOXs from *Xenorhabdus stockiae* DSM17904, *Pseudomonas aeruginosa* PAO1, *Streptomyces clavuligerus* F613-1 and *Pseudomonas sp.* SH-C52 bacterial strains. LC-MS analysis of reactions that were conducted either in the presence or absence of NADPH showed successful conversion of **63** whereas a lack of product formation was observed without reducing agent (Figure 79).

In pyrrolizixenamide biosynthesis, intermediate **63** and product pyrrolizixenamide **52** have an identical *m/z* ratio and corresponding molecular weights of 264.3 g/mol. Retention times in HPLC data were compared to confirm a lack of **63** conversion in the absence of NADPH and simultaneously eliminate false negative production of **52** which might have been possible if longer chain intermediates had been present as impurities of the main substrate (Figure 79A). Despite an identical *m/z* ratio, the two molecules differ in retention times on the C_{18}-column. Intermediate **63** possesses the same C_6 side chain moiety as **50** but shows minimally increased polarity due to an additional keto group and is therefore eluted earlier (t_R: 4.17 min) than the pyrrolizixenamide product (t_R: 4.31 min). On the contrary, product **52** is retained more than one minute longer on the column (t_R: 5.65 min) compared to **50** since it bears two additional CH_2 groups which make the molecule less polar (Figure 79B). Overall, discrimination between intermediate **63** and product **50** was possible due to varying molecule polarity and thus elution at different time points from the C_{18}-column.

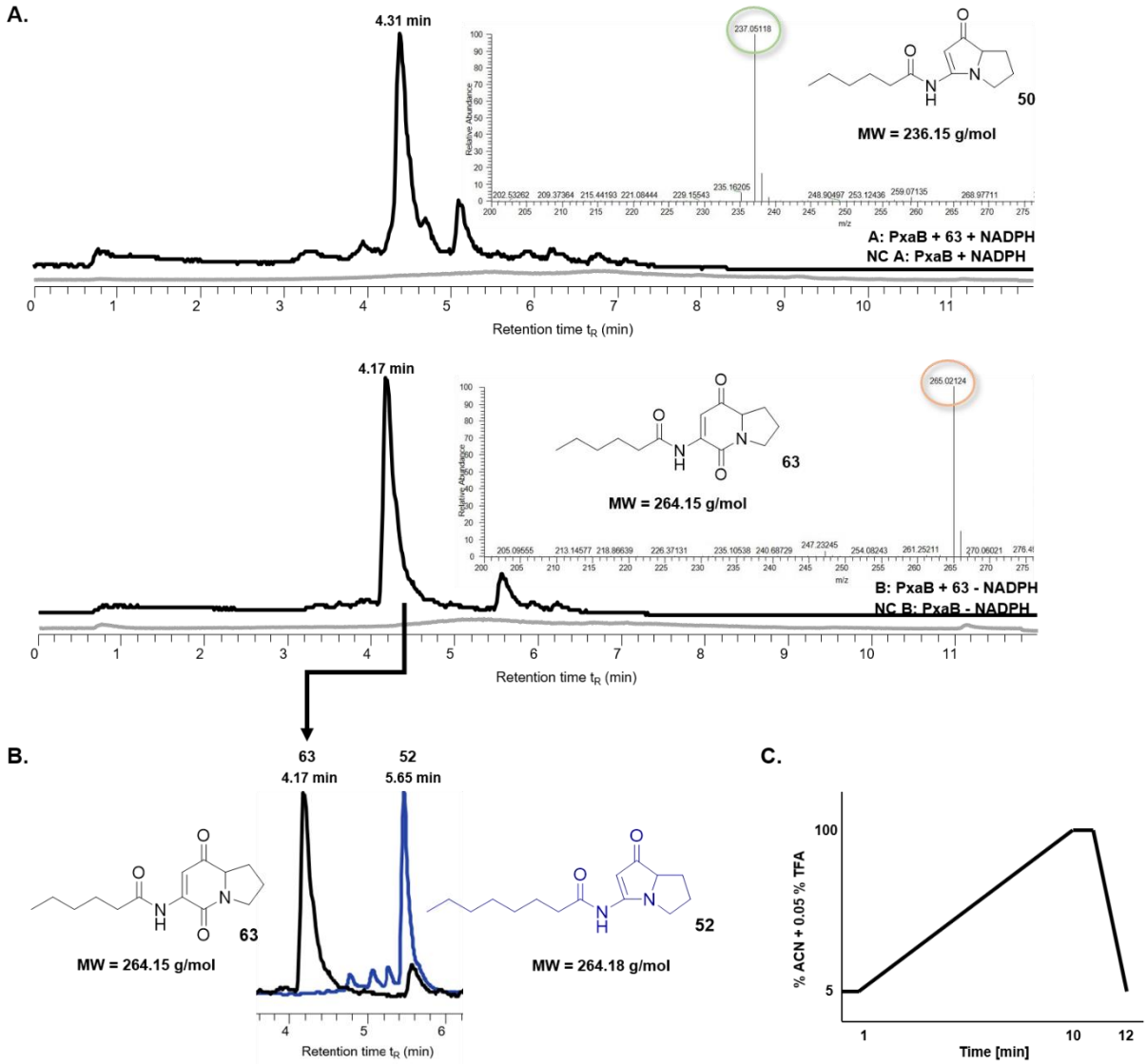


Figure 79. Studies on NADPH-dependency of the BVMO-catalyzed conversion reaction. **A.** Absence of co-factor NADPH prevents conversion of intermediate **63** (t_R: 4.17 min) to pyrrolizixenamide **50** (t_R: 4.31 min). **B.** Comparison of molecule retention times t_R highlights differences in chemical properties of intermediate **63** and product **52** (t_R: 5.65 min) which possess identical molecular weights (264 g/mol). **C.** Separation of intermediate and product PAs is possible using the HPLC-gradient method for measuring enzyme test assays shown. Gradient starts with 5 % ACN + 0.05 % TFA and 95 % water + 0.05 % TFA, then goes to 100 % ACN + 0.05 % TFA before re-equilibrating column for subsequent runs.

3.3.6.4.4 Molecular oxygen as a limiting factor for Baeyer-Villiger oxidation

In previous assays, the catalytic mechanism of action of selected bacterial BVMOs was studied and it was found that common characteristics such as NADPH-dependence, turnover efficiency dependent on the concentration of components taking part in the reaction cascade and aspects of enzyme kinetics coincide with published data on other BVMOs. Furthermore, it is known that the biochemical Baeyer-Villiger oxidation takes place under aerobic conditions since molecular oxygen is needed to incorporate an oxygen atom during ring expansion and use of the second oxygen atom to eliminate water in later stages of the reaction cascade. While trying to perform upscaling of the original 100 µL total volume test tube reaction, the availability of oxygen turned out to be a limiting factor for executing assays in 1.5 - 2 mL Eppendorf tubes. Whereas conversion of intermediate **63** to **50** was still detected for total volumes of 400 µL per tube, functionality of the respective MOX was apparently impaired due to oxygen deficiency at higher reaction volumes (Figure 80).

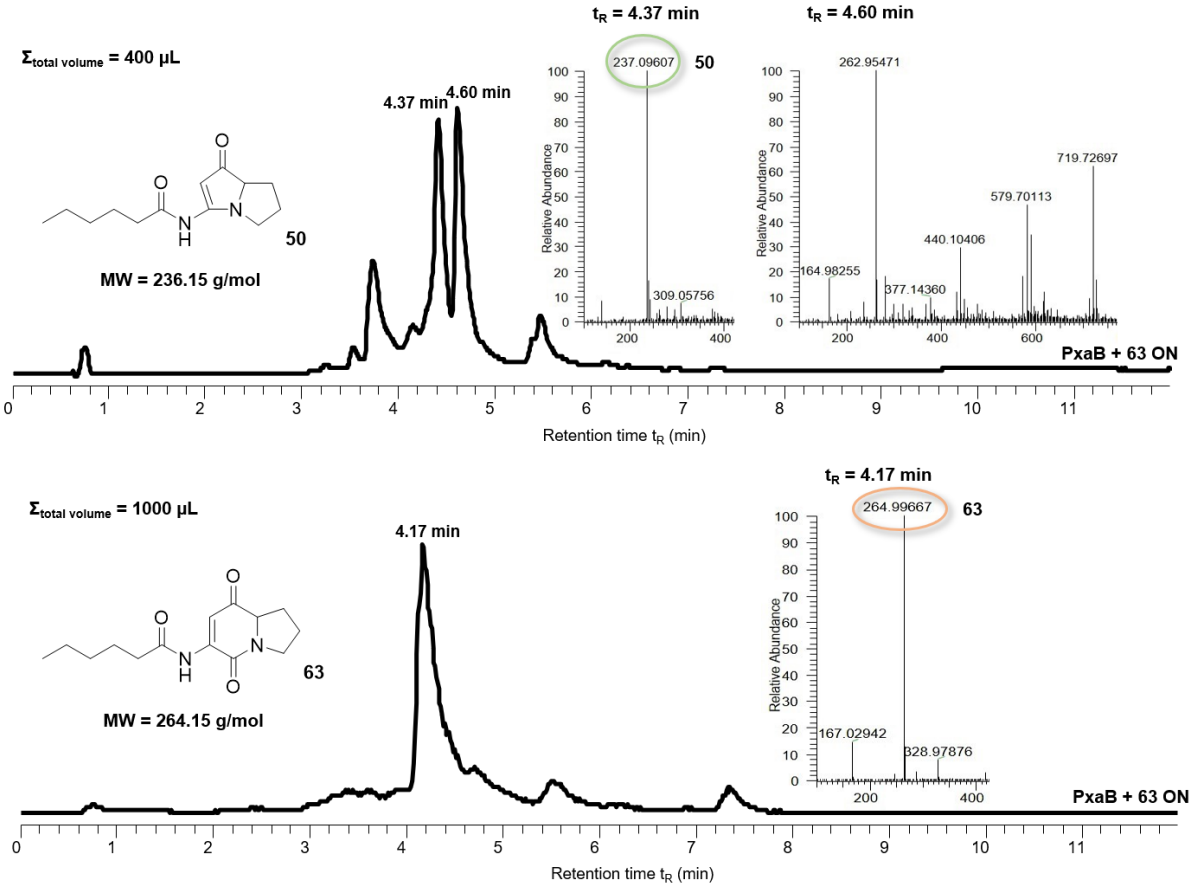


Figure 80. Upscaling of BVMO-catalyzed reaction. Production of **50** was monitored by LC-MS when total assay volumes were increased to 400 µL with maximum shaking on tabletop shaker (600 rpm). Oxygen availability in 2 mL Eppendorf tubes was insufficient for intermediate **63** conversion in 1 mL reactions volumes.

Most likely, the diameter of the test tube was too small to allow sufficient gas exchange at the air-liquid boundary, thereby preventing BVMOs to bind molecular oxygen and initiate catalysis. To overcome this obstacle in reaction scale-up, *in vitro* aeration conditions were optimized. Two approaches considering an alternative assay setup were developed to ensure adequate oxygen availability during the full duration of the assay. One experiment consisted of a petri dish filled with 5 mL of reaction assay that was shaken over night at 28 °C and 70 rpm, a setup designed to increase the liquid surface for enhanced gas exchange. An alternative experiment aimed at guaranteeing a constant oxygen flow during reaction time by pumping air into the Eppendorf tube via a hole in the lid while shaking at maximum speed on a tabletop shaker at 28 °C over night (Figure 81).

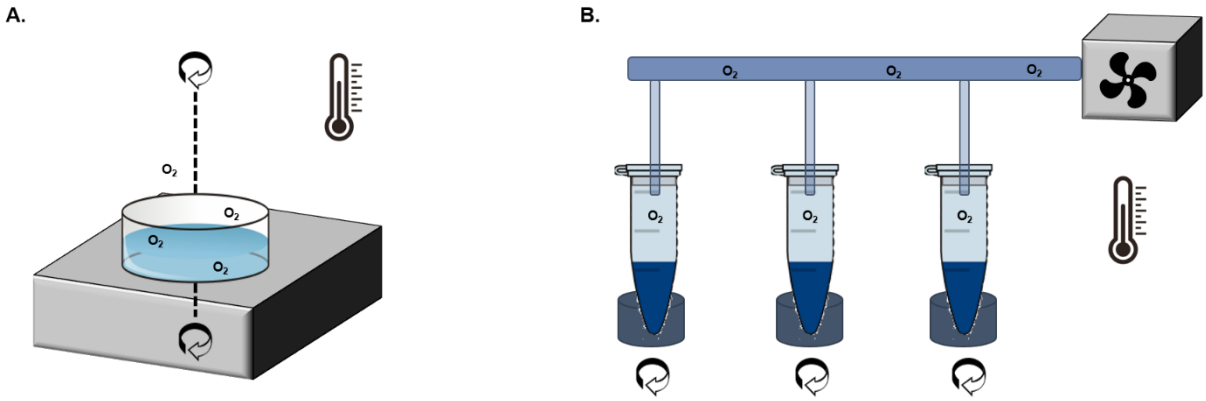


Figure 81. Experimental setups to improve *in vitro* aeration. **A.** A petri dish containing the 5 mL total volume including all necessary reaction components was incubated over night at 28 °C and 70 rpm shaking. **B.** Test volumes of 1 mL were continuously purged with air through a hole inside the tube lid while shaking at 600 rpm and 28 °C in a tabletop shaker.

The developed method A for upscaling of the MOX-catalyzed reaction proved to be an efficient solution for pyrrolizixenamide production at a 50 x bigger scale than initially applied (Figure 82). This leads to the conclusion that aeration issues present before in the Eppendorf tube can be solved by use of an alternative reaction container which allows better accessibility to oxygen.

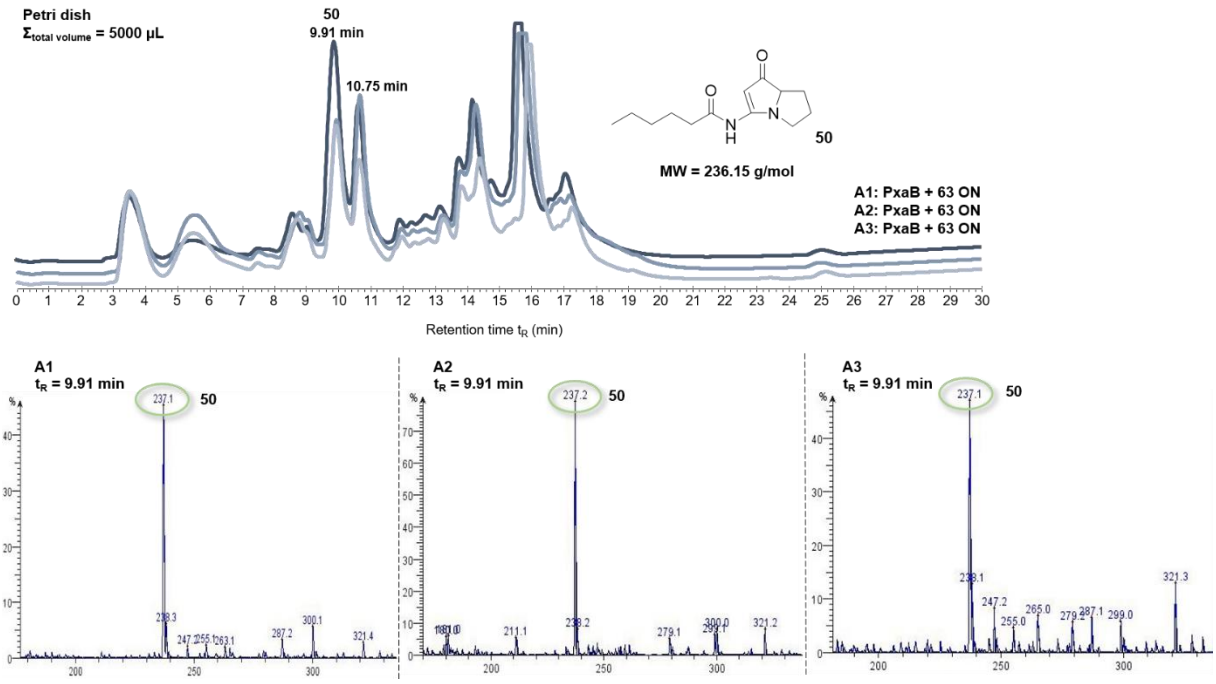


Figure 82. *In vitro* upscaling of pyrrolizixenamide production using petri dishes. Triplicate reactions were prepared to validate assay outcome. For assay analysis, a 30 min HPLC run was chosen (5-100 % ACN). A 30 min HPLC method was chosen for optimal separation purposes: 0-2 min: 5 % solvent B (ACN + 0.05 % TFA), 2-28 min: 5-100 % B, 28-29 min: 100 % B, 29-30 min: 100-5 % B.

According to the recorded mass spectra, the peak detected at a retention time of 9.91 min was clearly assigned to compound **50**. Retention times differ from observations made previously in Figure 79 due to running of a longer HPLC method with an altered gradient. However, elution of **50** occurs at ca. 37 % ACN in both cases. In addition, a second peak at t_R= 10.75 min with a detected *m/z* ratio of 279.1 is specific for the petri dish reaction. Considering solely the *m/z* value it can be assumed that the molecule eluted is also a derivative of **37** containing three additional CH_2 groups compared to **50**. However, not only was the assay performed with purified intermediate **63** which has a shorter chain length than the detected putative pyrrolizixenamide derivative is proposed to have, but also the difference in retention time of less than one minute makes this thought unlikely.

LC-MS data collected for assays performed with method B proved that direct aeration into an Eppendorf tube allowed a certain amount of intermediate conversion, but was not as efficient as method A. The prominent peak at t_R= 9.91 min confirmed production of **50**. However, mass detection showed a higher substrate than product abundance (Figure 83).

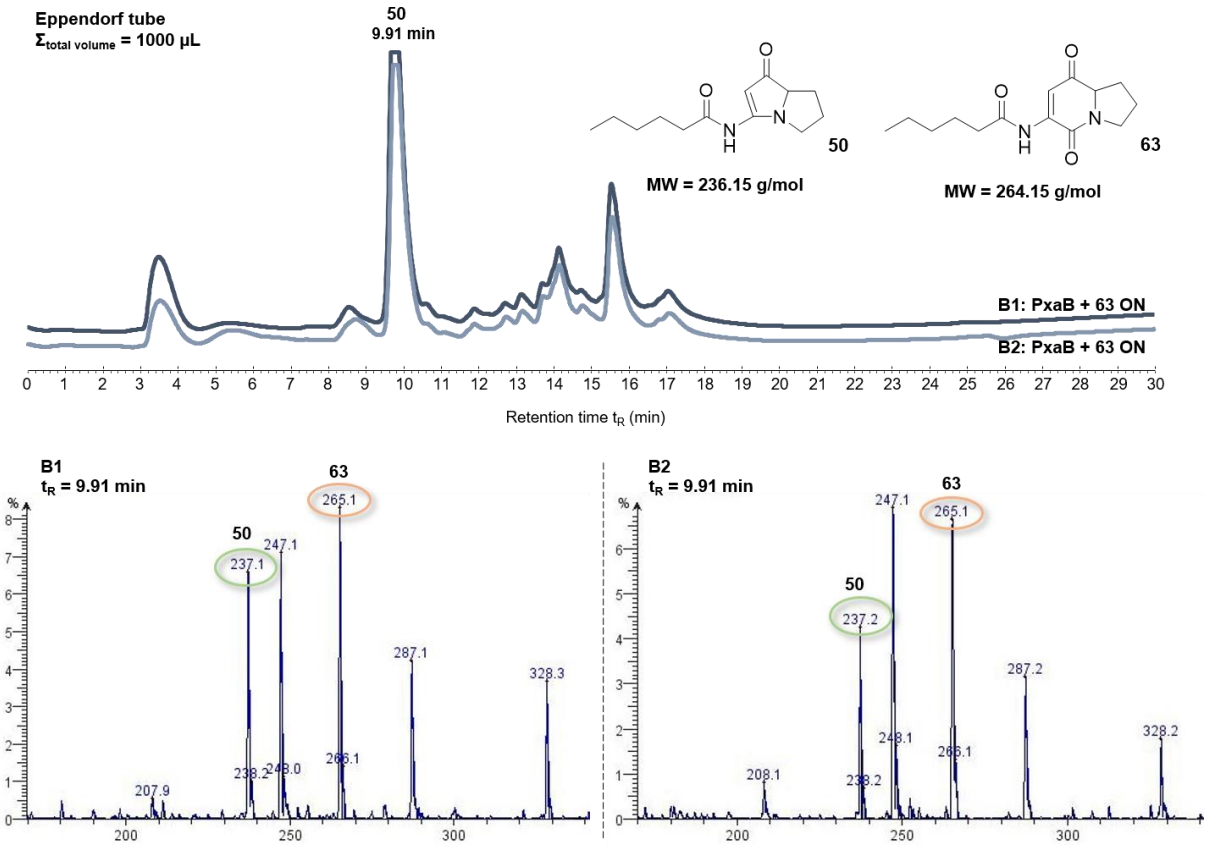


Figure 83. *In vitro* upscaling of pyrrolizixenamide production by enabling constant aeration inside Eppendorf tubes. Duplicate reactions were prepared to make assay outcome representative.

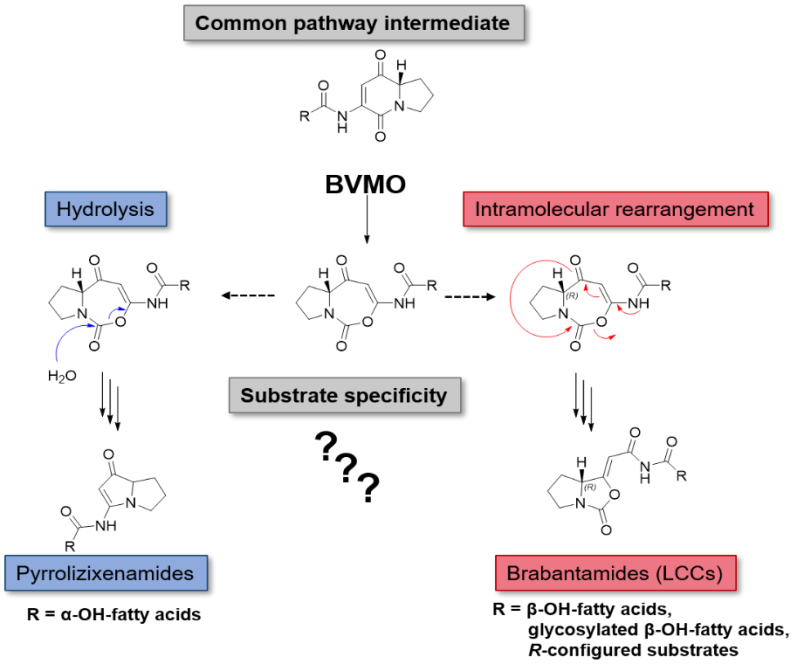
The PxaB-catalyzed reaction cascade was incomplete since intermediate **63** was still present in the sample after overnight incubation. Compared to the small-scale assays (section 3.3.5.4.1) where full substrate conversion was already observed after 1.5 h incubation time, continuous aeration of a 4x reaction volume scale-up remained inefficient, leading to the assumption that the container size and diameter available for oxygen exchange is one of the limiting factors for upscaling of the reaction.

3.3.6.4.5 Substrate specificity of BVMOs from related pathways

Publications thematizing several PA-related biosynthetic pathways have revealed similar mechanisms of action for NRPS systems in terms of generating bicyclic intermediates applicable for further processing by BVMOs. Structures of the respective intermediate molecules differ from one another based on the amino and fatty acids which are incorporated. For example, the NRPS system of the *Pseudomonas aeruginosa* PAO1 BGC preferably incorporates an azetidine 2-carboxylic acid (L-AZC) instead of a proline, which in turn is

provided by the SAM-dependent enzyme AzeJ encoded within the BGC of interest. Thus, a bicyclic intermediate consisting of a four- and a six-membered ring is formed as an alternative to the five-six-membered bicyclic structure found in several prominent PA-derivatives such as **37** or **46** etc..[103, 115] Furthermore, the attachment of fatty acid moieties, varying in saturation degree, branching, length and modification, to the core ring structure leads to great PA-diversity. Whereas variants of **37** possess saturated long chain fatty acid residues, PA-derivatives such as legonmycins[91] (**39**) and jenamidines[181] (**41**) bear unsaturated, branched and odd chain moieties.

Regarding the great diversity of bacterial PAs that have been found so far, it can be assumed that the BVMO homologues active in these respective biosynthetic pathways show high tolerance towards different substrates. MOXs involved in **37**- and the **40**-generating pathways were of particular interest during this thesis as it was previously postulated that the final products emerge from the same core ring system.[79] Our hypothesis for pathway divergence at the Baeyer-Villiger oxidation step is based on the type of fatty acid residue attached to the bicyclic intermediate structure. We thus set out to test substrate specificity of both pyrrolizixenamide-producing BVMO PxaB and MOXs from the BGCs of *Pseudomonas aeruginosa* PAO1, *Streptomyces clavuligerus* F613-1 and *Micromonospora auratinigra* DSM88415 as well as the brabantamide-producing BVMO BraC (Scheme 17).



Scheme 17. Emergence of PA molecules belonging either to the class of pyrrolizixenamides or brabantamides (LCCs) from a common precursor is predicted to be substrate dependent. So far, the mechanism driving the reaction in favor of one compound class has not been elucidated.

Analogous to testing of pyrrolizixenamide production by PxaB-catalyzed conversion of intermediates **63** - **65**, BraC substrate specificity was probed *in vitro* using the same substrates. Figure 84 shows biosynthesis of a molecule with an *m/z* ratio of 237.2 after overnight reaction incubation at t_R= 15.09 min which goes in line with production of **50**. In contrast, the predicted brabantamide derivative with the respective C_6 side chain moiety and a mass of 280.14 g/mol (**89**) was not detected in either of the samples. These findings suggest that a substrate molecule containing a saturated and rather short fatty acyl moiety promotes BraC-induced biosynthesis of **37** over **40**, thus representing “unnatural” BraC reactivity.

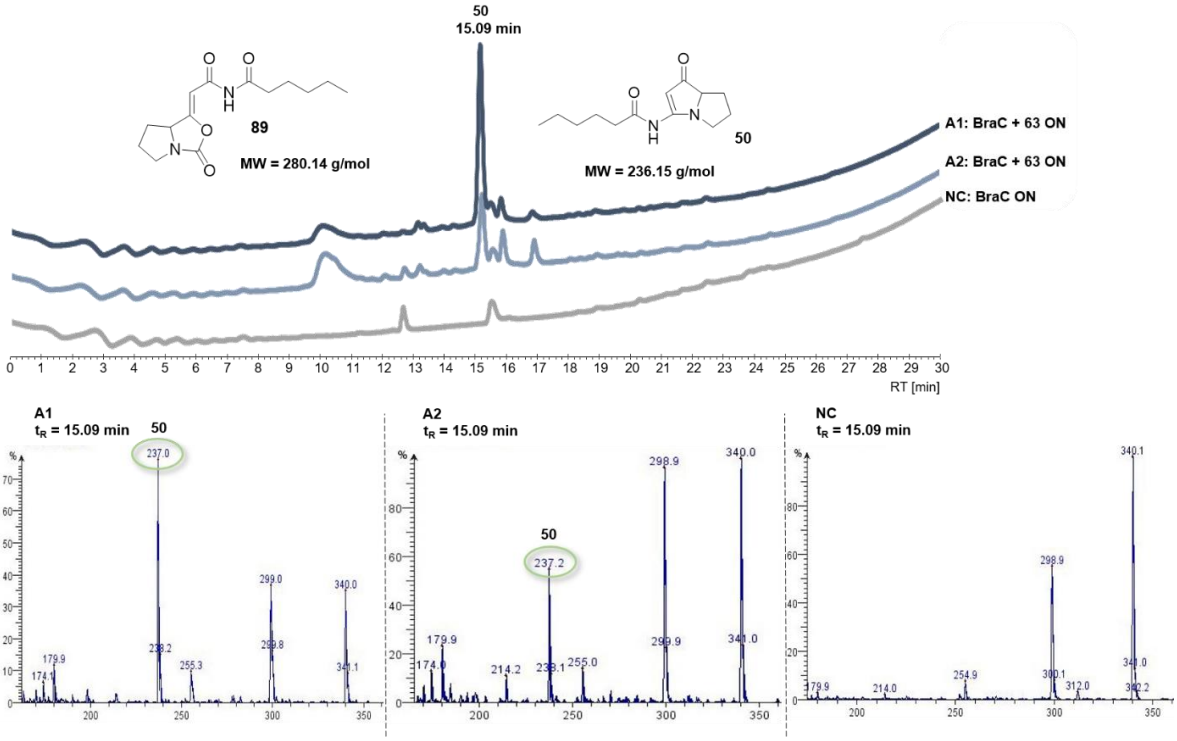


Figure 84. BraC-catalyzed conversion of intermediate **63**. Enzyme assays were performed at 30 °C over night while shaking. A main peak at a retention time of t_R= 15.09 min (51 % ACN) corresponds to the product also detected in PxaB-catalyzed assays (section 3.3.5.4.3). Negative control without substrate lacked product formation, thereby confirming protein activity.

With the previous results in hand, BraC enzyme assays performed with intermediate **65** bearing a hydrocarbon chain elongated by C_2H_4 compared to intermediate **63** were likely to yield **52** instead of the corresponding brabantamide derivative (**90**). Indeed, LC-MS revealed the successful biosynthesis of the PA-derivative since the main peak at t_R= 18.20 min exhibited an *m/z* ratio of 265.1 which coincides with the molecular weight of **52** (Figure 85). Interestingly, a second, less prominent peak at a retention time of 17.73 min indicated the presence of another molecule that had been formed during enzyme catalysis. With the help of mass

analysis, an isotope peak at *m/z* = 309 was identified, which matches the $[M+H]^+$ mass of the predicted brabantamide derivative (Figure 85). These specific molecules **89** and **90** share the core structure including the β-ketoacyl moiety with the published variants of **40** but are not hydroxylated at the γ-position and thus do not carry a sugar moiety like L-rhamnose.

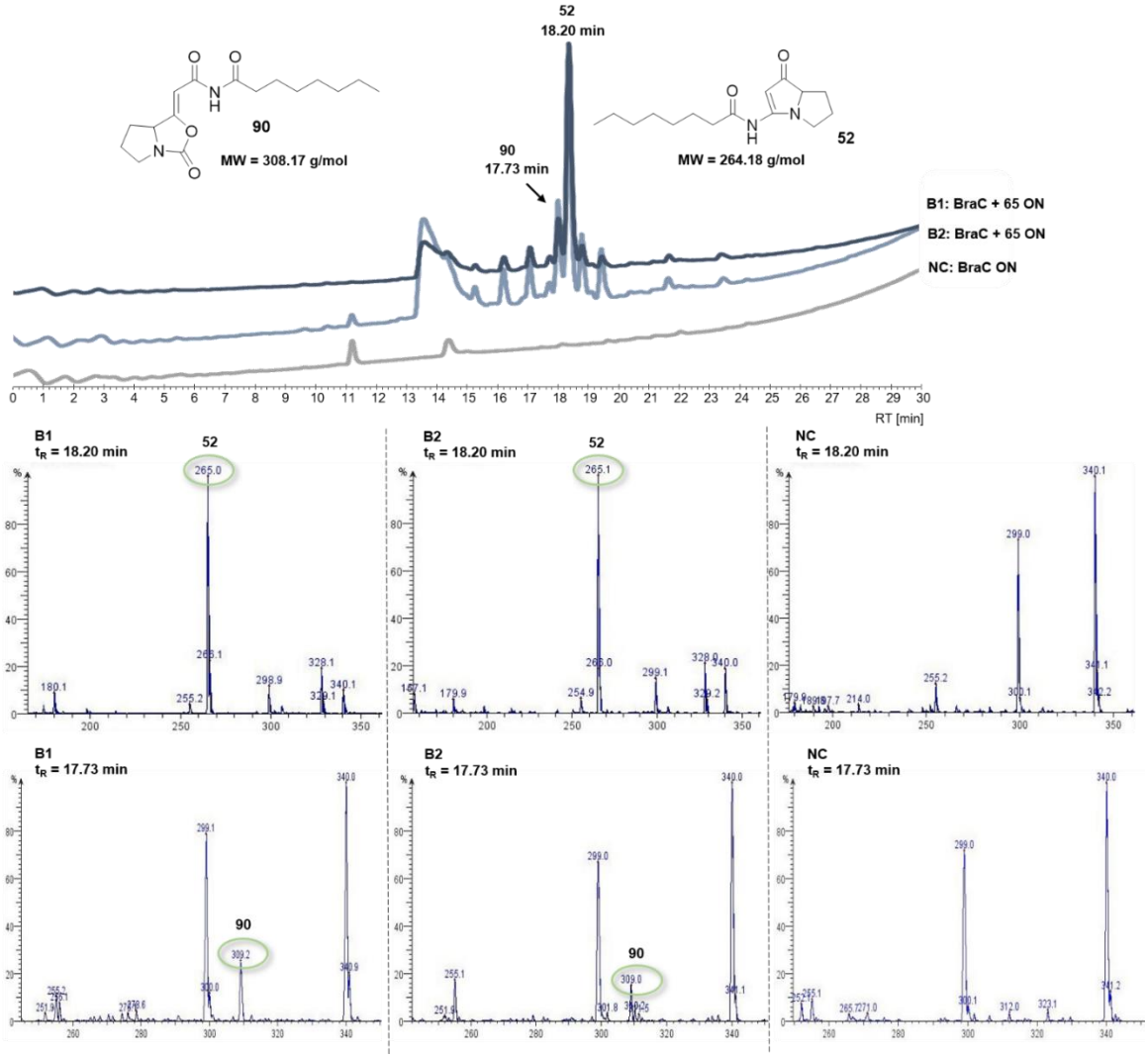


Figure 85. Transformation of intermediate **65** by BraC-mediated catalysis into two distinct products. The main reaction product was detected at t_R= 18.2 min with an *m/z* ratio of 265. In addition, a less abundant ion peak with *m/z* = 309.2 was found after running the gradient for 17.73 min (60 % ACN). Additional peaks correspond to enzyme assays components also soluble in EtOAc.

Thus, it seems as if the hydrocarbon chain length plays an important role in directing product outcome. This is an exceptional finding, as it indicates for the first time that substrate conversion to lipocyclocarbamates by BraC is indeed dictated by substrate structure alone. In consecutive experiments it would therefore be of interest to further elongate the fatty acid side chain moiety and investigate whether the reaction can be further shifted towards brabantamide

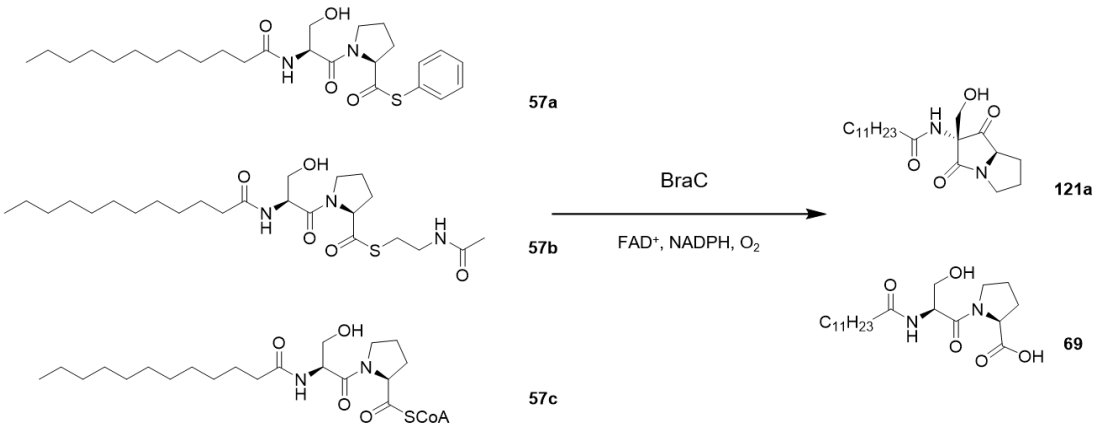
production. Although intermediate **67** with a C_{12} chain was one of the four intermediates produced *in vivo*, the amounts obtained were insufficient for further tests. Chemical synthesis of the C_{12} intermediate as well as derivatives thereof was still a work in progress in the laboratory at the end of this thesis.

After formation of the [5,6]-bicyclic intermediate, biosynthesis pathways of pyrrolizixenamides and brabantamides diverge. Whereas an intramolecular allylic 1,3-transposition causes rearrangement of the ring system for brabantamide production, the pyrrolizixenamide [5,5]-bicyclic core ring system is generated by hydrolysis.[79] The observed formation of both molecules in parallel hints at a MOX enzyme capable of switching between the two mechanisms, BraC. Regarding crystal structures of published BVMOs not involved in PA biosynthesis, two critical steps, both dependent on NADPH-binding and reorientation, include generation of a reactive flavine peroxyanion species and opening of the binding pocket to allow substrate entry and subsequent formation of the Criegee intermediate. NADPH conformational change not only enables substrate access to the enzymes' catalytic site, but it also prevents substrate from diffusion back into the solvent and at the same time blocks bulk solvent entry.[177, 180] The binding pocket itself is predicted to be relatively large which explains the high substrate scope of BVMOs. Yet, examinations concerning how substrates of varying size and structural complexity exactly bind inside the pocket and interact with specific amino acids remain open. Furthermore, the structure of a BVMO specifically involved in PA biosynthesis has not been published so far, thus making it difficult to explain the shift from hydrolysis of the [5,6]-bicyclic intermediate to allylic 1,3-transposition observed when elongating the substrates' side chain moiety from C_6 to C_8. Since the hydrocarbon chain is not involved in the reaction, it may be assumed that it influences the exact position of the substrate inside the binding pocket and thus alter interactions with amino acid binding partners. In addition, a coordinated water molecule inside the catalytic site that contributes to substrate-enzyme interaction may influence product outcome.[180, 182]

Recapitulating, *in vitro* assays have shown an identical catalytic mechanism for [5,6]-bicyclic pathway intermediate conversion for all tested MOX homologues, thereby underlining their common mode of action. Relative product yield was increased by reaction parameter optimization, starting with the amount of enzyme used for testing and continuing with overcoming reaction limitations including oxygen and co-factor availability. Although traces of a **52** brabantamide-type molecule **90** were discovered alongside the main pyrrolizixenamide product in a single assay, feeding of alternative substrates could not direct the reaction towards brabantamide biosynthesis significantly.

3.3.6.4.6 Substrate conversion using simplified precursor molecules of 40

In previous research conducted by Dr. Françoise Schaefers in our laboratory, a great effort has been made to shed light into the mechanism of brabantamide biosynthesis. Within her dissertation, Ms. Schaefers set a focus on elucidating transformation of the PCP-bound linear di-lipopeptide to the [5,6]-bicyclic intermediate, a reaction postulated to be catalyzed by the TE-domain of the corresponding NRPS system.[183] She chemically synthesized simple thioesters **57a-c**, activated versions of lauric acid bearing different functional groups, and subsequently tested formation of novel products **69** and **121a** catalyzed by the TE-domain or the TE-PCP didomain (Scheme 18). Unfortunately, none of the expected products were detected during LC-MS analyses. Furthermore, **57a-c** were used as substrates for conversion assays with the BVMO BraC in the presence and absence of the PCP-domain, with results similar to TE-domain testing and thus proving mechanistic studies to be difficult for brabantamide biosynthesis.



Scheme 18. Hypothetical BraC-catalyzed reaction of chemically synthesized simple precursor molecules **57a** – **57c**.

Based on these preliminary findings, a test series was performed with the available substrates **57a** and **57b** using alternative MOX systems such as PxaB as well as enzymes derived from *Streptomyces clavuligerus* F613-1, *Pseudomonas aeruginosa* PAO1 and *Micromonospora auratinigra* DSM44815. In addition, tests were repeated with BraC to have a direct comparison. Assays were prepared analogously to the previously conducted enzyme studies using PxaB, containing NADPH and FAD^+ as co-factors and ensuring sufficient oxygen availability in a buffered system at neutral pH (Table 10).

LC-MS data of test reactions revealed the presence of three dominant products, with two of them being absent in the negative control lacking PxaB protein. The peak detected at t_R=7.59 min is identical to the substrate-containing negative control and mass analysis confirmed the identity of **57a**. Peaks detected at 6.18 min and 7.07 min retention time showed similar intensities in the HPLC-chromatogram and corresponding ions were equally abundant in the mass spectra (Figure 86). Interestingly, the observed *m/z* ratio of 385.0 and 367.0 are in line with the $[M+H]^+$ masses of predicted reaction products **121a** and **69**, respectively, leading to the assumption that these molecules were indeed formed due to PxaB activity. Moreover, these findings proved the establishment of a pyrrolizidine core in the absence of a [5,6]-bicyclic intermediate. The bulky thiophenol moiety of the substrate seemingly does not impair the enzymes' reactivity. Whereas **121a** most likely results from a nucleophilic attack from the C_α of serine onto the thioester, a new side reaction of BraC, **69** is a product of hydrolytic activity. Identical results were obtained for assays using purified MOX enzymes from selected bacterial strains that validate the occurrence of the observed reaction.

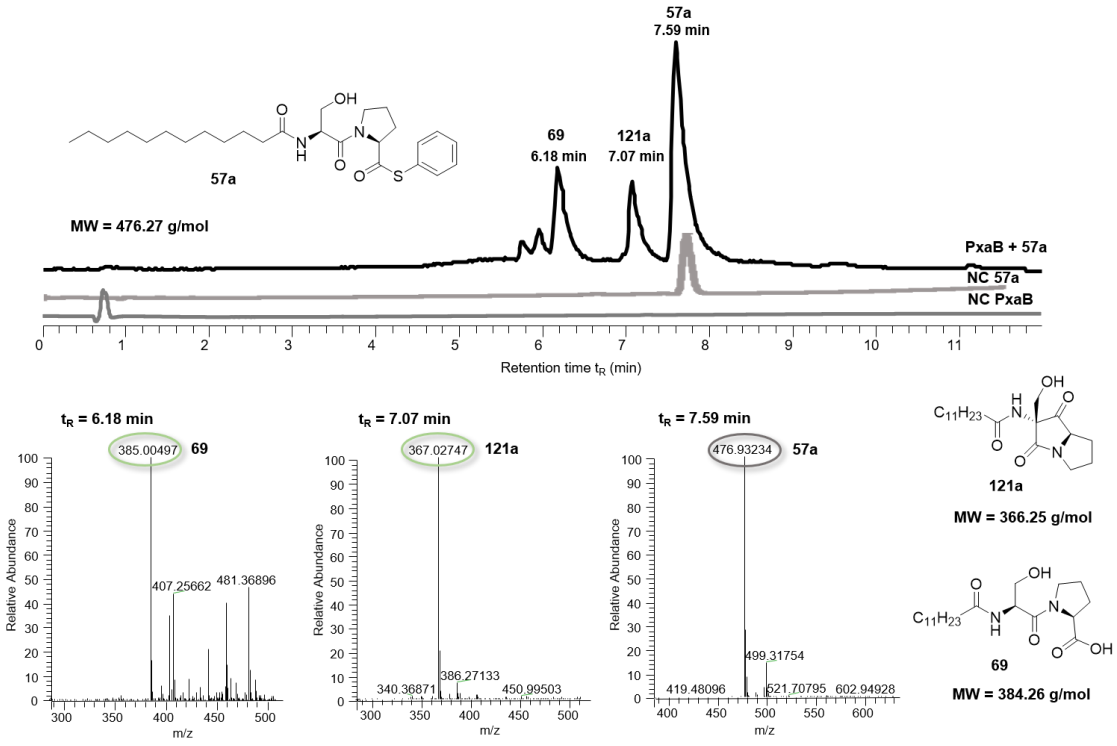


Figure 86. *In vitro* turnover of brabantamide precursor **57a** by PxaB. LC-MS data revealed presence of multiple molecules. Structural and molecular weight data given for educt as well as predicted reaction products.

As previously mentioned, substrate **57b** contains *N*-acetylcysteamine (SNAc) and therefore exhibits an altered spatial occupation compared to the thiophenol group in **57a**. It was therefore of interest to see if the type of functional group coupled to lauric acid influenced product outcome or even prohibited MOX-catalyzed substrate conversion.

Figure 87 displays a similar reaction outcome as found for **57a**. Again, three peaks were visible in the chromatogram recorded at 220 nm, with two of them coinciding with products **121a** and **69.** Upon comparison of the sample with a negative control, the highest peak at t_R= 6.37 (*m/z* = 485.3) could once again be associated with the unreacted **57b**. Due to successful conversion of the simple brabantamide precursor to the predicted products by the PAO1C29 MOX enzyme, the exchange of the reactive thioester group does not have an impact of catalysis.

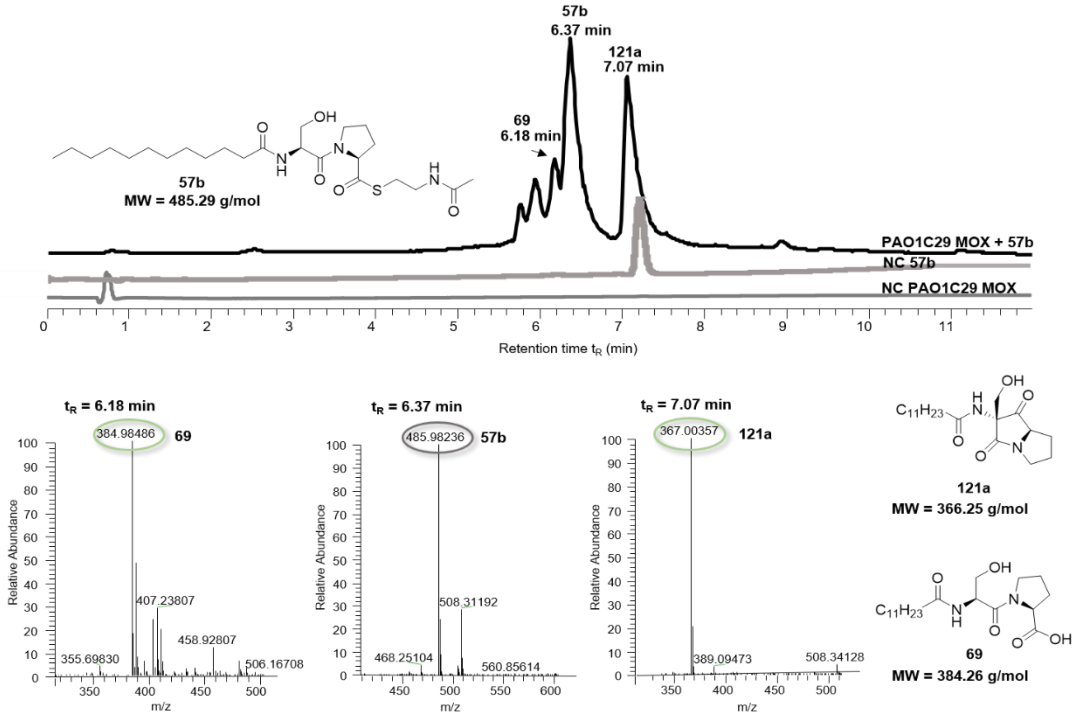


Figure 87. PAO1C29 MOX-catalyzed conversion of brabantamide precursor **57b** *in vitro*. The substrate **57b** was detected by LC-MS as the main component alongside two equally distributed components corresponding to the postulated reaction products.

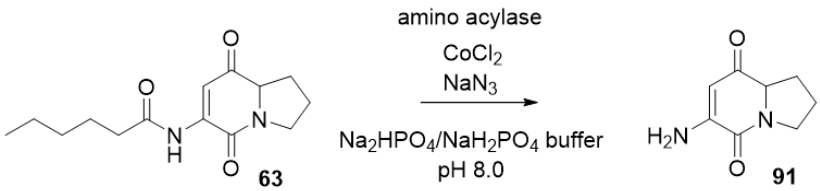
In conclusion, chemically synthesized simple brabantamide precursor molecules **57a** and **57b** which had previously failed to be converted by BraC, were successfully processed by PxaB and its homologues. Although both substrates were initially prepared for testing TE-domain activity, the common [5,6]-bicyclic intermediate structure was not necessary for BVMO-catalyzed reaction cascades to take place but instead the presence of a single pentacyclic ring was sufficient for transformation into products **121a** and **69**. The necessity for one of the two rings already being present in the intermediate was underlined by failure of conversion assays

with substrate **70** for all MOXs (Figure S16). For **70**, the absence of a thioester functionality may also have restricted processing. Finally, substrate conversion catalyzed by BraC could be achieved under the given reaction conditions since identical results presented in Figures 86 and 87 were obtained (Figures S17 – S18). Thus, it can be concluded that PxaB, BraC and additionally tested BVMO homologues share a similar substrate scope and additionally perform a new side reaction.

3.3.6.4.7 Enzymatic deacetylation of pyrrolizixenamide intermediates

Within the course of studying BVMO substrate promiscuity and the influence of substrate structural complexity, one objective was to achieve direct control of the reaction towards brabantamide biosynthesis. As demonstrated by previously executed experiments, the MOX BraC responsible for brabantamide establishment has so far converted all available intermediate molecules to derivatives of **37** instead of the desired brabantamide alternatives (**40**). The results shown in Figure 85 strongly indicate catalytic activity of BraC towards lipocyclocarbamates, with substrates bearing longer fatty acid side-chains. Unfortunately, such substrates were not available by biotechnological approaches. Synthesis of analogues resembling brabantamide-like intermediates was therefore planned as a combination of enzymatic deacetylation (Scheme 19) coupled to the attachment of chemically synthesized side-chain moieties (Scheme S1).

Pyrrolizixenamide pathway intermediates such as **63** were available in sufficient amounts from *in vivo* heterologous expressions and thus could be used as substrates for the initial enzyme-catalyzed reaction step. Aminoacylases belong to the class of hydrolases and characteristically catalyze cleavage of an existing amide bond between an amino acid and an acyl residue. This feature was to be utilized to deacetylate intermediate **63** and generate a free amine **91**. The aim was then to synthetically attach alternatively modified hydrocarbon side-chain moieties to this amine at the [5,6]-bicyclic core.



Scheme 19. Attempted deacetylation reaction catalyzed by aminoacylase enzyme specific for *N*-acyl-L-amino acid derivatives.

Being metalloenzymes, aminoacylases require a co-factor such as zinc (Zn^{2+}) or cobalt (Co^{2+}) for full functionality. Furthermore, catalytic activity is stated to be optimal at pH 8.0 and temperatures of 50 °C according to the supplier. An alternative reaction protocol described by Blaser et al.[184] performed the reaction at 37 °C for 48 h. Furthermore, test reactions included buffer exchange from a phosphate basis to 50 mM Tris, 10 % glycerol buffer (pH 7.5). Due to the removal of the hydrophobic acyl side chain the deacetylated reaction product was predicted to be significantly more polar than intermediate **63** and thus, elution from the C_{18}-column at earlier retention times and solubility in the aqueous phase during extraction were assumed.

After overnight incubation of the reaction mixture at 50 °C and shaking and subsequent extraction and washing procedures, the expected deacetylated compound **91** seemingly was detected in both the organic and the aqueous phase of the raw extract (Figure 88). In comparison to a negative control test assay lacking aminoacylase (A4), the most abundant molecular ion detected at t_R = 12.43 min carried a mass-to-charge ratio of *m/z* = 167.0 corresponding to the $[M+H]^+$ molecular weight of the deacetylated product (A1 and A3). However, the mass spectrum of A4 clearly shows a second prominent mass adduct with the same ratio, leaving the question open whether relative abundance of the ion with *m/z* = 167.0 resulted from a successful deacetylation event or was just generated as part of the fragmentation pattern of **63** by electron spray ionization. Hypotheses, that formation of the respective adduct is a result of the analytical method measurement setup was supported by a similar retention time of reaction product compared to **63**. Unlike observations made for molecules soluble in EtOAc, a shift in retention time of more than two min was expected since loss of the complete C_6 acyl side chain leads to a significant increase in molecular polarity. Previously, elongation of the hydrocarbon chain of the acyl group by two CH_2 groups had already caused a two min shift, therefore C_6 removal would have caused an even higher retention time shift. However, this was not the case in the measured deacetylation assay, leaving the question open, why any kind of shift was observed between products of aqueous and EtOAc phases (Figure 88: A3 compared to A4). NMR-analysis of the obtained compound was performed, revealing only presence of the starting material in the solution, thus making enzymatic deacetylation catalyzed by aminoacylase unsuccessful. Despite adaptation of multiple reaction parameters and use of more than one purchased aminoacylase (Amano Acylase, CAS Nr. 9012-37-7 and D-Aminoacylase recombinant from *E. coli*, CAS Nr. 65979-42-2, both from Sigma), the amide bond could not be hydrolyzed. Thus, to obtain promising analogues that promote brabantamide biosynthesis, development of a full synthetic route seems to be the most promising alternative to tackle this task.

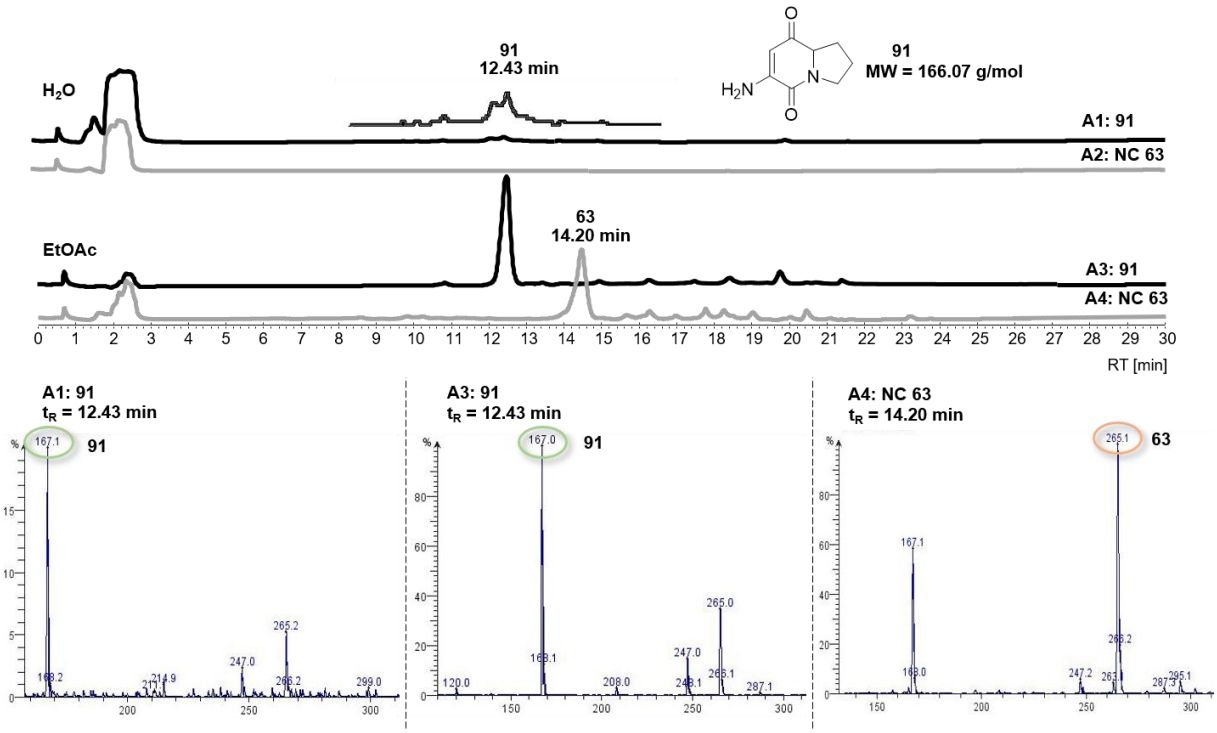


Figure 88. LC-MS profiles of enzymatic deacetylation reaction. Organic and aqueous phases of extraction were prepared separately for analytical measurements.

3.4 Chemical synthesis of precursor molecules

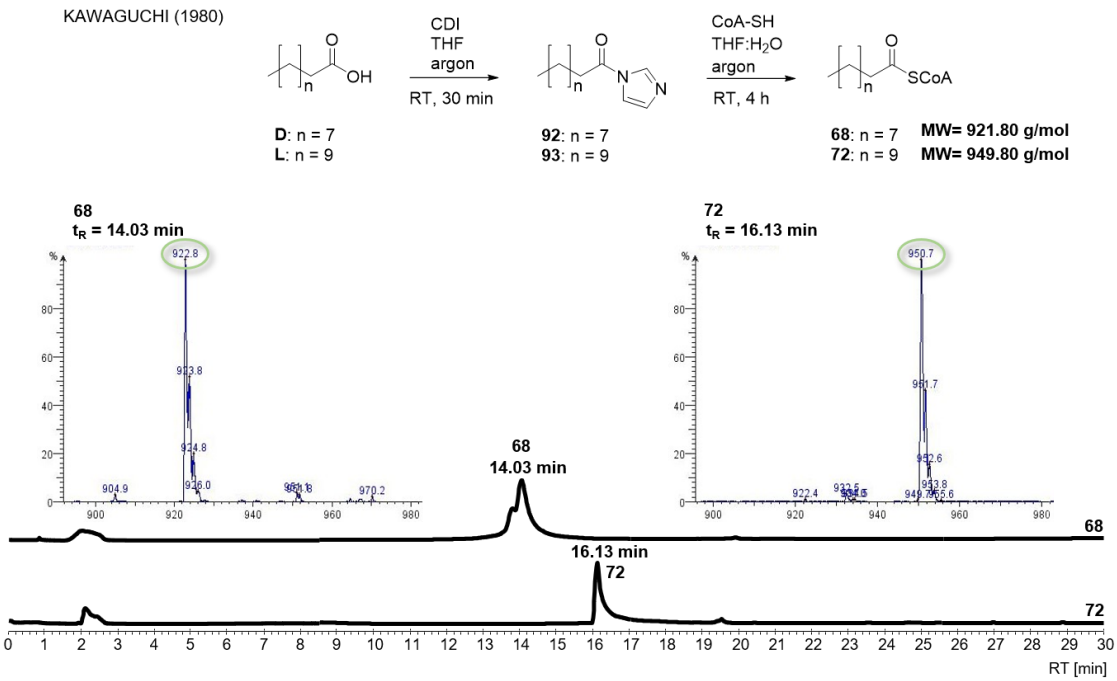
Successful PA biosynthesis *in vivo* can take place due to availability of diverse enzymes involved in transforming basic building blocks, such as fatty acids, into forms that are reactive enough to be directly engaged in the NRPS-initiated catalytic pathway. Whereas the endogenous fatty acid metabolism as well as *E. coli*-own enzymes such as FadD or its homologues provide acyl-CoA precursor variants that can subsequently be incorporated into the PA backbone by peptide bond formation *in vivo*, no substitute mechanism exists *in vitro*. For *in vitro* reconstitution of the PA biosynthetic pathway, it was therefore inevitable to prepare activated fatty acid precursors as substrates for the enzyme test assays to be performed.

Fatty acid activation was executed by thioesterification with either SNAc, t-BuSH or CoA-SH thiols.

3.4.1 Synthesis of acyl-CoA thioesters

Activation of fatty acids by coupling to CoA-SH resulting in an acyl-CoA thioester is a common reaction taking place in all living organisms. Products of acyl-CoA metabolism regulate many metabolic enzymes and cell signaling pathways, thus assigning them a major role concerning maintenance of cell functionality.[185] Synthesis of acyl-CoA thioesters was therefore desired to mimic native conditions *in vitro* as closely as possible in the absence of additional functional enzymes taking over fatty acid activation.

A modified protocol of Kawaguchi et al.[186] was applied to generate acyl-CoA precursors from decanoic (D) and lauric (L) acids. In an initial reaction performed in tetrahydrofuran (THF) and under argon atmosphere, the carboxylic acid was converted into an active acyl imidazole species (**92**, **93**) by carbonyldiimidazole (CDI)-catalyzed anhydride formation followed by CO_2 elimination. A second reaction step was necessary for coupling of the acyl side chain to thiol CoA-SH to yield products **68** and **72** in quantitative amounts (Scheme 20). Acyl-CoA precursors were subsequently supplemented to *in vitro* assays for testing NRPS activity by incorporation into the expected PA [5,6]-bicyclic pathway intermediate (see section 3.3.5.2).



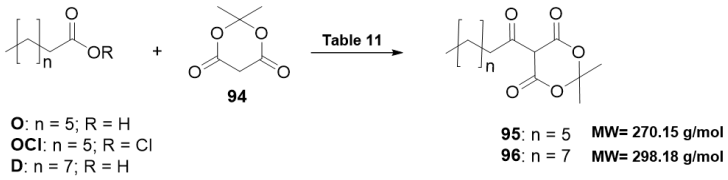
Scheme 20. Two-step reaction mechanism for acyl-CoA generation. Both products **68** and **72** were obtained in quantitative amounts. $[M+H]^+$ molecular weights of corresponding compounds displayed as most abundant mass adducts in LC-MS measurements.

3.4.2 Synthesis of biomimetic thioesters

As previously mentioned, biosynthesis of diverse natural products in Nature relies on the availability of reactive CoA-conjugates, which are commonly generated by coupling of a carboxylic acid to the thiol CoA-SH. However, both *in vivo* and *in vitro* studies on metabolic pathways have been hampered by the chemical properties and behavior of CoA, including its size and structural complexity, long-term instability, and thus difficulty of purification and storage under laboratory conditions. In addition, commercially available CoA and acyl-CoA derivatives are expensive. Chemical tools, such as synthetic CoA-mimics, have been established to overcome these hurdles and simplify study efforts. SNAc, a truncated analogue with increased cell membrane permeability and higher stability, has become a valuable alternative for carboxylic acid activation.[165] Multiple protocols for acyl-SNAc thioester synthesis were tested throughout this thesis, using not only long chain fatty acids but also employing carboxylic acids with short and medium length hydrocarbon moieties.

3.4.2.1 Generation of β-ketoacyl intermediates using Meldrum's acid

In initial attempts to obtain β-ketoacyl-SNAc derivatives by chemical synthesis to investigate Pys-catalyzed pyreudione biosynthesis, establishment of a reactive acyl-containing intermediate by coupling of a carboxylic acid to Meldrum's acid (**94**) in a first reaction step was performed (Scheme 21). Meldrum's acid, also known as 2,2-dimethyl-1,3-dioxane-4,6-dione, is an acidic compound which reacts with carbonyls, preferably in the presence of a catalytic agent such as pyridine. Generation of acyl-Meldrum's acid has been described extensively in literature, but only a small selection of protocols was tested (Table 11).



Scheme 21. Activation of a variety of fatty acids via Meldrum's acid. O = octanoic acid; OCl = octanoyl chloride; D = decanoic acid.

Table 11. Test reaction conditions for synthesis of acyl-Meldrum's acid using substrate **O**, **OCl** and **D**.

Entry	Substrate	Reagents	Solvent	Reaction conditions	Product	Reference	Yield [%]
1	O	Pyridine	DCM	1 h 0°C, 1 h RT	**95**	OIKAWA[187]	40 - 57
2	O	Pyridine, DMAP	DCM	ON RT	**95**	SOEHANO[188]	51
3	O	Pyridine, DMAP	DCM	1 h -15 °C	**95**		49
4	O	Pyridine, DMAP	DCM	ON -15 °C	**95**		54
5	O	TEA, DMAP	DCM	1 h -15 °C	**95**		45
6	O	TEA, DMAP	DCM	ON -15 °C	**95**		53
7	OCl	DMAP	$CHCl_3$	1 h 0°C, 1 h RT	**95**	YAMAMOTO[189]	51
8	OCl	Pyridine, DMAP, EDC	DCM	15 min RT, 5.5 h RT	**95**	STEGLICH[190]	54
9	OCl, D	DMAP, DCC	DCM	ON RT	**95**, **96**	BYCROFT[191]	49
10	OCl, D	DMAP, EDC	DCM	ON RT	**95**, **96**	RUYSBERGH[192]	56 - 66

Relatively pure acyl-Meldrum's acid intermediates **95** and **96** in yields of up to 66 % were obtained after preparative HPLC purification when using a slightly modified reaction setup (10) as described by Ruysbergh et al. in 2016.[192] When using equal equivalents of fatty acid and **94**, ^{1}H-NMR analysis showed signals belonging to keto- and enol-forms of the intermediate as well as remnants of unreacted substrate (Figure S19). Washing with the milder formic acid (10 mM and 100 mM FA tested) compared to original use of 2 M HCl led to a reduction in the number of detected signals that were most likely related to intermediate decomposition in the presence of a strong acid. With product yields of up to 66 %, chemical synthesis of **95** and **96** was most efficient when using either reaction condition 1 analogous to the description by Oikawa (1978) or reaction condition 10 as described previously.[187] The obtained yellow-brownish oily liquids were further used as substrates for acyl-SNAc synthesis. Formation of the reaction intermediate out of substrates over time was monitored by analytical HPLC as exemplarily shown in Figure 88. Bond formation between the more polar **94** and the less polar fatty acid (O) leads to generation of intermediate **95** which is more polar than the fatty acid on its own. This leads to elution of **95** ca. 1 min earlier than substrate **94** from the column (Figure 89).

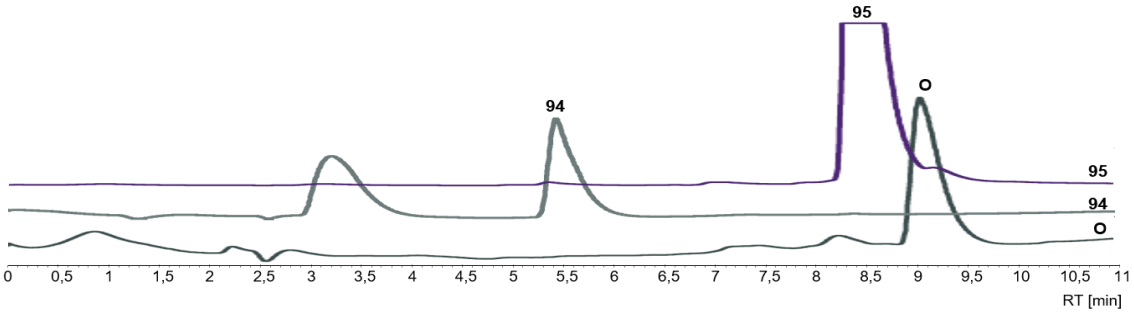
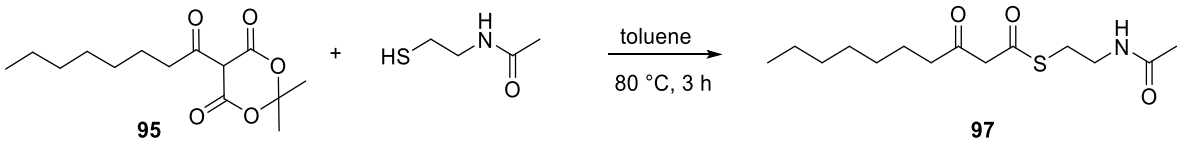


Figure 89. HPLC chromatograms of pure reaction substrates **94** and **O** compared to the acyl-Meldrum's acid intermediate **95**, a coupling product of the two substrates. A one-minute shift in retention time indicates bond formation between carbonyl and Meldrum's acid and results in a change of compound polarity.

3.4.2.2 Synthesis of β-ketoacyl-SNAc thioesters from 95 and 96

Acyl Meldrum's acid derivatives readily react with nucleophiles and electrophiles, making them attractive substrates for a broad range of synthetic applications.[193] Synthesized intermediates were further used to generate β-ketoacyl thioesters, molecules needed for testing catalytic activity of the monomodular NRPS system Pys involved in pyreudione biosynthesis (Scheme 22). Furthermore, the resulting products have the potential of being applied as simplified

substrate mimics for studies addressing directed establishment of **40**. Reactivity with thiols, such as SNAc, has been shown and heating of the intermediate in the presence of toluene promotes (thio)ester exchange and decarboxylation, resulting in the formation of *S*-(2-acetamidoethyl)3-oxodecanethioate (**97**, Table 12).



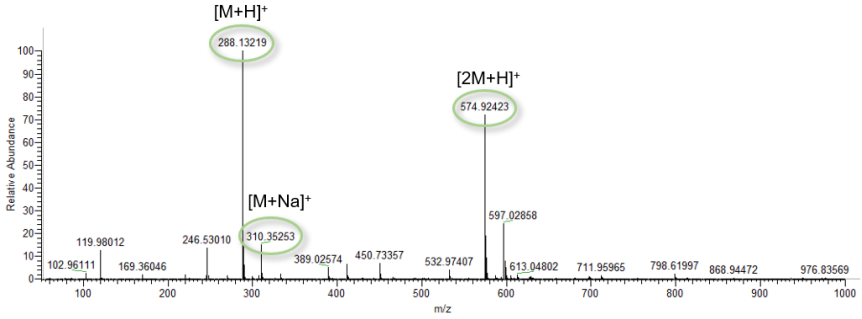
95 97

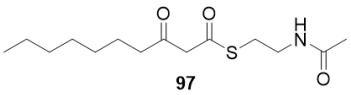
Scheme 22. Transfer of the activated β-ketoacyl moiety onto an SNAc thioester for precursor establishment.

Table 12. Tested reaction conditions for β-ketoacyl-SNAc synthesis.

Entry	Substrate	Reagents	Solvent	Reaction conditions	Product	Reference	Yield [%]
1	**95**	SNAc	Toluene	3 h, 80 °C	**97**	LIU[194]	64
2	**95**	SNAc	ACN	3 h, reflux	**98**		53

The yellow oily liquid acyl-Meldrum's acid intermediate **95** was successfully converted into an orange solid β-keto thioester **97** with a yield of 64 % after silica column chromatography. Mass spectra analysis of the two main fractions clearly stated presence of the desired SNAc-thioester as the main product, since the mass to charge ratios of three characteristic molecular ions were detected, including the $[M+H]^+$, $[M+Na]^+$ and $[2M+H]^+$ adduct masses (Figure 90).



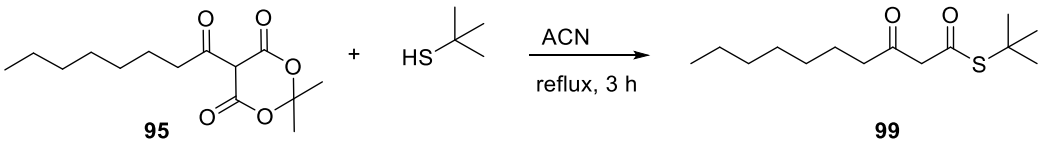


97

Adduct	Predicted m/z ratio	Measured m/z ratio
M	287.423	-
$[M+H]^+$	288.430276	288.132190
$[M+Na]^+$	310.412218	310.352530
$[2M+H]^+$	575.853276	574.924230

Figure 90. Mass spectrum of main fraction of β-ketoacyl-SNAc thioester made from octanoyl chloride (OCl). Comparison of calculated with measured *m/z* ratios confirmed identity of most abundant adduct **97**.

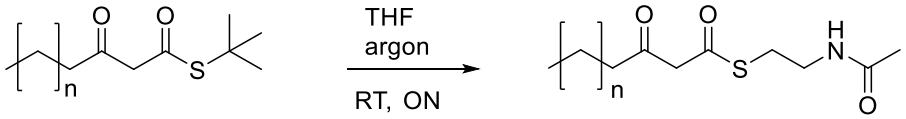
Alternatively, acyl-Meldrum's acid derivatives were mixed with *tert*-butylmercaptan (*t*BuSH) in acetonitrile for 3 h under reflux and purified by silica column chromatography (DCM 97: MeOH 3) to obtain a viscous, orange-colored liquid for example of S-(tert-butyl) 3-oxodecanethioate (**100**) with 53 % overall yield (Scheme 23).



Scheme 23. Chemical synthesis of β-ketoacyl-*t*BuSH product using altered conditions as presented in Table 12, entry 2.

3.4.2.3 Transthioesterification of β-ketoacyl thioesters

For rapid generation of multiple acyl-CoA biomimetics, the mechanism of transthioesterification was used for one step acyl group transfer between *t*BuSH and SNAc thiols. Substitution of the bulky *t*BuSH by the less space demanding SNAc functional group was of particular interest to study enzyme substrate promiscuity. The reaction was carried out in an oxygen-free environment with overnight incubation (Scheme 24). Since SNAc is a stronger nucleophile compared to *t*BuSH, addition of excess nucleophile (five equivalents) was sufficient for acyl group transfer to proceed at room temperature after initial heating with a final yield of 55 % (Table 13).



99: n = 5
100: n = 7

97: n = 5
98: n = 7

Scheme 24. Transthioesterification reaction for rapid thioester group exchange.

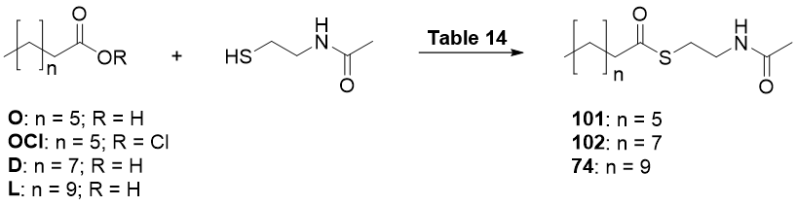
Table 13. Reaction conditions for single step generation of β-ketoacyl thioester variants by transthioesterification.

Entry	Substrate	Reagents	Solvent	Reaction conditions	Product	Yield [%]
1	**99**, **100**	-	THF	15 min, 50 °C ON, RT	**97**, **98**	49 - 55

Taken together, several β-ketoacyl thioesters of different side chain lengths (C_8 – C_{10}) bearing either the *t*BuSH or the SNAc functional group as CoA mimic were successfully synthesized in a two-step reaction mechanism. Initial generation of an acyl-Meldrum's acid intermediate was necessary for carbonyl activation followed by thioester formation. Overall product yields between 49 % and 55 % demand for further reaction optimization. However, obtained amounts were satisfactory for application in *in vitro* enzyme assays.

3.4.2.4 Synthesis of α-ketoacyl-SNAc thioesters as precursors for production of 37

Synthesis of β-ketoacyl thioesters by generation of an acyl-Meldrum's acid intermediate was necessary for installing the β-keto function essential to pyreudione biosynthesis and further of importance for attachment of the rhamnose sugar moiety as part of the brabantamide molecule. In contrast, activated carboxylic acid substrates used by bimodular NRPS systems for buildup of **37** are α-ketoacyl thioesters, therefore alternative synthetic routes were chosen for precursor preparation (Scheme 25, Table 14).



Scheme 25. Conversion of a selection of fatty acids to α-ketoacyl thioesters.

Table 14. Test reaction conditions for synthesis of α-ketoacyl thioesters with varying hydrocarbon chain length.

Entry	Substrate	Reagents	Solvent	Reaction conditions	Product	Reference	Yield [%]
1	O, L	DIC, DMAP	DCM	1 h, 0 °C, 3 h, RT	**74**, **101**	COCHRANE[195]	79
2	O, OCl, L	PyBOP, DIPEA	DMF	1 h, 0 °C, 3 h, RT	**74**, **101**		55 – 75

A first chemical route was set up according to a protocol published by Cochrane et al. in 2016[195], in which the carbodiimide coupling reagent DIC in combination with the nucleophilic catalyst DMAP was applied for single step α-ketoacyl thioester synthesis. Alternatively, activation of the carboxyl moiety and subsequent acyl transfer onto the thiol group of SNAc was catalyzed by PyBOP in the presence of the base DIPEA. Formation of the corresponding acyl-SNAc compound was monitored over time by LC-MS, the product peak progressively growing and the substrate signal simultaneously decreasing with ongoing reaction period. The high relative abundance of several mass adducts corresponding to the desired reaction product confirmed successful thioester bond generation, as exemplarily shown for *S*-(2-acetamidoethyl)-dodecanethioate (**74**, Figure 91). Final yields of purified products ranged from 55 % (**101**) to 79 % (**74**), with total amounts sufficient for extensive enzyme tests.

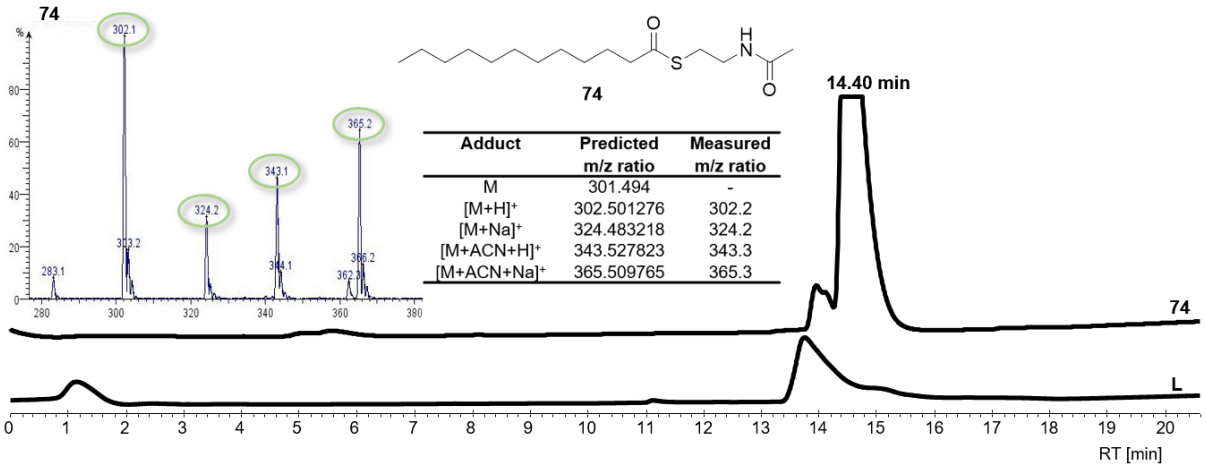
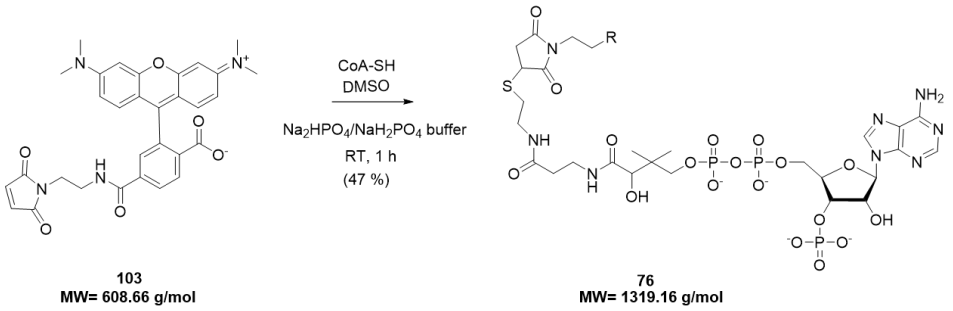


Figure 91. Monitoring of synthesis **74** by LC-MS. Multiple molecular ions were detected with *m/z* ratios corresponding to mass adducts of the desired compound.

3.4.2.5 Synthesis of small molecule probe-CoA conjugates

Full catalytic activity of heterologously expressed and subsequently purified proteins is an absolute requirement for enzyme activity assays to be successful. To eliminate this serious source of errors leading to test assay disfunction, it must be verified that heterologously expressed NRPS enzymes can be primed at the PCP domain active site serine residue by PPTases such as Sfp *in vitro*. Since PPTases catalyze transfer of the ppan moiety of CoA onto the PCP-domain, labeling of the ppan-arm with a detectable probe can be performed to study *in vitro* priming activities. For this, **13** was labeled with tetramethylrhodamine-5-(and-6) C_2 maleimide (TMR-Mal; **103**) in a Michael condensation reaction, resulting in thioether bond formation between fluorophore and CoA thiol group (Scheme 25).[174] TMR-Mal-CoA synthesis (**76**) was performed according to the protocol published by Yin et al.[174], progression of the reaction was followed by analytical HPLC and the resulting product was purified by preparative HPLC, giving a pink solid (47 % yield).



Scheme 25. Reaction conditions employed for CoA-labeling with TMR-Mal fluorophore.

Reaction product formation was followed by chromatography by measuring at λ = 556 nm, a wavelength close to the emission maximum λ = 572 nm of **103** and conjugates. The main TMR-Mal-CoA product eluted after 18.80 min and was separated from additionally occurring side products and substrate remnants by preparative HPLC (Figure 92).

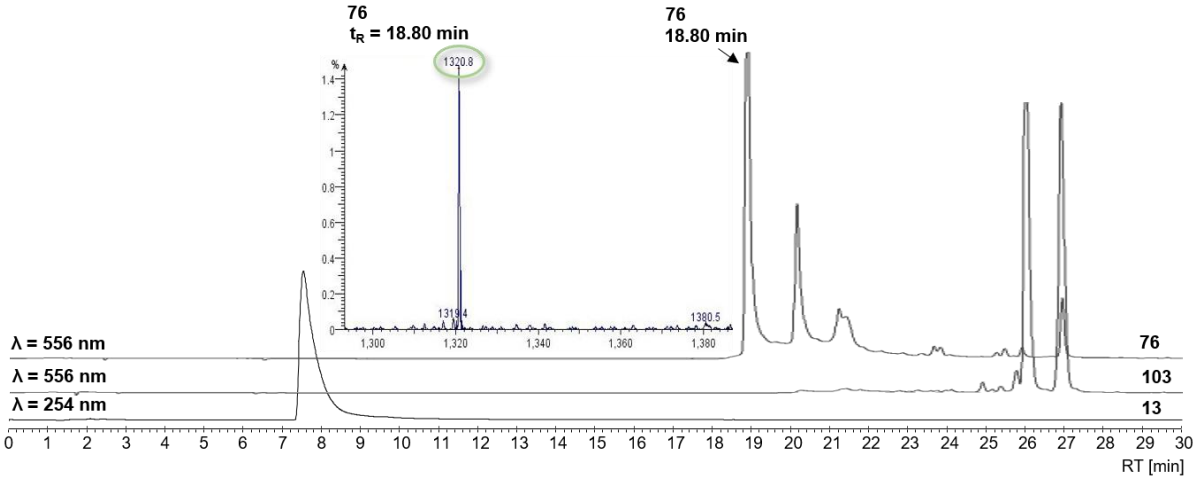


Figure 92. HPLC chromatograms were recorded at wavelengths of 254 nm to detect **13** and at λ = 556 nm for visualization of TMR-Mal-conjugate **76** (excitation maximum: λ = 544 nm; emission maximum: λ = 572 nm). Since pure TMR-Mal was purchased as a mixture of 5- and 6- isomers, two distinct peaks were detected in the HPLC chromatogram.

Purified compound **76** was successfully applied for site-specific PCP-domain labeling of the monomodular Pys system involved in pyreudione biosynthesis catalyzed by Sfp PPTase (see section 3.3.5.3).

Chemical synthesis of several substrates and precursor molecules for *in vitro* experiments was needed throughout the course of this thesis, molecules that are usually supplied by endogenous host cell metabolism *in vivo*. For studies on the initial reaction step in pyrrolizixenamide biosynthesis, fatty acids needed to be available as thioesters to be incorporated into the pathway intermediate by the NRPS. Apart from a variety of CoA-conjugates of varying acyl side chain length that represent the naturally occurring active form in Nature, a selection of SNAc- and *t*BuSH-containing biomimetics were synthesized. Biomimetics may be advantageous due to increased stability and smaller size, therefore α- and β-ketoacyl thioesters were synthesized carrying either of the three thiol groups. Lastly, CoA was labeled with the fluorophore TMR-Mal to generate a probe for monitoring Sfp-initiated PCP-domain loading with the ppan-moiety.

4. Summary and Outlook

4.1 Conclusion of project outcome

Within the framework of this thesis, a great variety of experiments were conducted to gather insights into the biosynthesis of bacterial PAs. Bioinformatic analyses showed that BGCs encoding for enzymes putatively involved in PA biosynthetic mechanisms are present in bacteria of diverse environments and places on Earth, highlighting the potential and relevance of the produced PAs for bacterial secondary metabolism and survival. From the many strains exhibiting BGCs of interest, a selection of commercially available candidate bacterial strains was made for elaborate study.

Despite originating in different phyla and differing in the overall gene cluster complexity of the PA-related BGCs of interest, evolutionary conservation of the core NRPS and MOX enzymes was underlined. Gene homologies of more than 75 % for NRPS systems and up to 65 % for MOX enzymes support their importance for PA-biosynthesis in bacteria (Figure 93).

Underlying tight regulation, many of these gene clusters remain transcriptionally silent when native hosts are cultivated under standard laboratory conditions. Thus, heterologous expression in well-studied and commonly used hosts *E. coli* and *Streptomyces* was performed to activate BGCs and examine gene function and product outcome.

Heterologous expression of BGC # 12 from *Photorhabdus temperata* DSM15199 was achieved in *E. coli*, thereby producing already published PAs pyrrolizixenamides A (**50**), B (**51**) and C (**52**)[79] in amounts of up to 20 mg per Liter of cultivation broth. Furthermore, three yet unknown pyrrolizixenamide derivatives bearing longer hydrocarbon side chains were identified (**53** – **55**). Evidence of PA derivative biosynthesis was also observed upon expression of *Streptomyces sp.* Mg1 BGC # 5 in *E. coli*. However, detailed compound identification and structural elucidation was still work in progress at the end of this thesis.

Expressing BGCs of interest in *Streptomyces* hosts proved to be more challenging and multiple optimization steps were necessary to enable gene cluster activation. Modifications at the cloning (adjustment of spacer regions, refactoring, cluster splitting, etc.) and expression steps (variation of host organism and cultivation parameters) led to expression of promising PA derivative candidates in low titers. Presence of PAs **50** – **52** in extracts of *Streptomyces clavuligerus* F613-1 BGC # 39-containing *Streptomyces* hosts was verified by LC-MS. In addition, multiple extracts of BGCs with higher complexity, such as *Streptomyces fulvissimus* DSM40593 BGC # 2, contained metabolites displaying characteristic retention times (t_R = 15 -

22 min) and *m/z* ratios (*m/z* = 237 – 324), some directly in line with variants of **37** and all within the size range of PA derivatives.

The sensitivity of BGC transcriptional regulation in the native bacterial strain was visualized by cultivation of *Micromonospora auratinigra* DSM44815 in different media, leading to unique metabolic profiles for each culture broth. One compound found in the extract of NZA medium-cultivated *Streptomyces* was of particular interest, as it indicated similar properties to PAs in initial analyses.

All in all, fast and comparatively easy access to target bacterial BGCs involved in PA biosynthesis was enabled by heterologous expression in suitable host systems under laboratory conditions (Figure 93).

A biomolecular mix and match toolbox was established to study the function of the proposed core enzymes encoded within the BGCs of interest in detail. Multiple full-sized (270 kDa), multi-domain NRPS enzymes responsible for biosynthesis of the bicyclic PA intermediate were successfully expressed in yields of up to 4 mg/L. The substrate scope of the monomodular NRPS system Pys involved in the related pyreudione biosynthetic pathway was tested by supplementing the reaction with multiple substrates **77** – **83**. Alongside the literature-known pyreudione A (**84**),[172] two novel pyreudione derivatives **85** and **86** were obtained.

Reconstitution of the second step of PA biosynthesis, the Baeyer-Villiger oxidation-initiated reaction cascade, was achieved using purified MOX enzyme (yields of up to 45 mg/L) and underlined the necessity of cofactors NADPH, FAD^+ and O_2 for full catalytic activity. PAs **50** – **52** were found to be the most abundant molecules when supplementing the *in vitro* assay with intermediates **63** – **65**. Exchange of literature-known MOX PxaB with homologues from related clusters such as PAO1C29 MOX, MiC21 MOX or BraC all led to biosynthesis of compounds **50** – **55**, thereby highlighting a common mechanism of action and high gene conservation among this group of enzymes. Studies addressing the point of divergence during the MOX-reaction, leading either to formation of pyrrolizixenamides or brabantamides, strongly hint at a substrate-driven directionality. With the available intermediates, compounds **50** – **55** were biosynthesized by both PxaB and BraC in all assays, whereas evidence of a brabantamide-like derivative **90** was observed in only one case. Since feeding of basic building blocks, such as rhamnose or β-hydroxy fatty acids, did not support reaction outcome in favor of production of derivatives of **40**,[97] chemical synthesis of brabantamide-like precursors was commenced as an alternative testing strategy.

Taken together, *in vitro* step-by-step reconstitution of the PA biosynthetic pathway has given insights on the function of both core enzymes NRPS and MOX and their role in PA formation and furthermore has shed light on enzyme-substrate specificity and limiting factors (Figure 93).

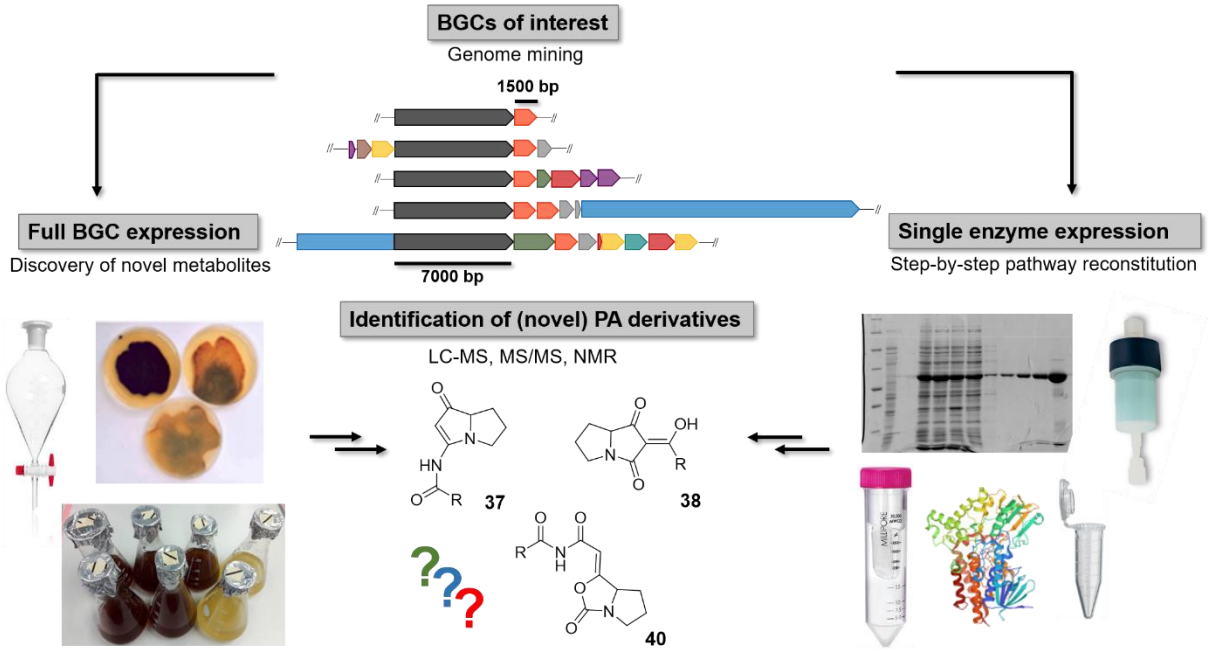


Figure 93. Workflow applied within this thesis for elucidation of PA biochemistry. Genome mining efforts led to the identification of several bacterial BGCs putatively involved in PA biosynthesis. Heterologous expression of full-length BGCs led to isolation of both literature-known and novel derivatives. Expression and purification of core enzymes involved in PA biosynthesis with subsequent testing of catalytic activity using substrate variations yielded both, published compounds as well as novel modified analogues.

The data and results collected within the course of this thesis, including heterologous expression of full-sized BGCs as well as step-by-step dissection of the biosynthetic pathway, contribute to the basic understanding on the biosynthesis of bacterial PAs. Furthermore, optimization attempts made to experimental setups to enable or enhance successful test outcome provide valuable information considering future handling and discovery of novel PAs.

4.2 Thoughts thematizing future project work

Experiments performed within the course of this thesis gave insights into the structural diversity of PA natural products based on a common mechanism for pyrrolizidine core establishment, which is highly conserved amongst many bacterial strains bearing PA producing BGCs. Based on the collected information, new questions and ideas arise and pending challenges are yet to be tackled.

Using homologous recombination in *E. coli* or DiPaC, large-sized clusters (> 25 kbp) with complex gene organization could not be captured and expressed, thus alternative methods should be tested in the future to overcome these restrictions. TAR cloning (transformation associated recombination) in yeast is predicted to be a promising option, since it follows a similar principle as LLHR/LCHR but can capture DNA of up to 300 kbp length from both simple and complex genomes. Even for genomes with high GC contents, the limitation for DNA recovery lies at 100 kbp, therefore accessibility of selected PA producing BGCs in the size range of 25 – 31 kbp should be possible.[196] Otherwise, application of a BAC (bacterial artificial chromosome) system may be considered for target BGC capturing.[197]

For successfully captured PA-producing BGCs and established plasmids, optimization of heterologous expression conditions is necessary for unambiguous compound identification and characterization. Expression of selected BGCs in *E. coli* systems already allows large-scale production of molecules of interest. To further simplify target compound detection, reduction of background noise in the HPLC expression profile caused by media components or molecules resulting from endogenous *E. coli* metabolism needs to be addressed. Alteration of pore size (Å) and resin material (polystyrene versus polyacrylic resin) may increase affinity of product molecules to the sorbent, making it easier to collect them from the cultivation broth.[198] Additionally, preparative purification protocols need to be adapted to receive highly pure target compounds, for example by using isocratic HPLC solvent distribution to separate multiple molecules hidden underneath a single peak in the recorded chromatogram. For some of the expressed BGCs of interest, initial attempts have been made and characterization is still a work in progress. Directed manipulation of the *E. coli* expression system to further increase cell permeability and enable uptake and processing of a variety of substrates is another aspect to consider in upcoming experimental setups.

In vivo heterologous expression of target BGCs in *Streptomyces* must be thoroughly revised concerning plasmid arrangement as well as cultivation conditions. Finding ways to activate BGCs of interest under laboratory conditions, for example by varying the spacer region between RBS and start codon of the initial gene or inserting additional promoters, must be

prioritized. The foundation for the establishment of a mix & match biomolecular toolbox analogous to the one already applied in *E. coli* has been made, with two consecutive promoters controlling simultaneous translation of multiple genes. Apart from finding ways to promote activation of silent BGCs, guaranteeing reproducibility of promising results upon repetition of tests and in this context enabling compound production on a larger scale is a central aspect for future investigations. Characterization of secondary metabolites will only be possible if enough pure product is available for in-depth LC-MS and NMR analysis.

Efforts have been made towards the elucidation of mechanistic details of PA backbone biosynthesis by stepwise reconstitution of enzyme-catalyzed reactions *in vitro*. Despite extensive testing, the initial NRPS-catalyzed reaction could not be simulated so far. With only low amounts of target protein available for experimental setups, protein expression must be revised to achieve significant increases in protein yields per Liter, to enhance protein solubility and reduce the percentage of incompletely translated protein. Alteration of expression parameters, co-expression of stabilizing enzymes such as chaperones or supplementation with solubilizing agents either during expression or as part of the protein purification procedure are aspects to consider. Optimization of the purification process can also help to remove confounding factors which falsify protein concentration and possibly interfere with target protein activity in the *in vitro* assay. Following affinity chromatography, ion exchange (Resource™Q) and size exclusion (SEC) chromatography should be applied to further purify NRPS systems, optimally using automated processes such as the Äkta pure systems. Finally, full functionality of the purified NRPS system must be ensured for *in vitro* assays to have a chance at success. If solutions to these challenges can be found, reconstruction of total biosynthesis of PAs *in vitro* can be completed.

The reaction cascade catalyzed by BVMOs was reconstituted with success and gave insights into the mechanism of action, supporting the hypothesis of a substrate-driven mode of action. Expansion of the MOX substrate pool using chemical synthesis to generate a collection of derivatives will be inevitable to understand substrate limitations to greater extent. It is the idea to manipulate the reaction in a way that alternative molecules such as brabantamides are produced instead of the readily available pyrrolizixenamides. Finally, crystallization of MOX homologues in complex with cofactors and/or substrates will facilitate structure elucidation efforts.

5. Material & Methods

5.1 Measuring devices

Autoclave. For the sterilization of media, glass ware and any consumable material or to kill (contaminated) cell cultures after finishing experiments, either an autoclave Tuttnauer 3170ELV (biomedis; TUM) or the autoclave VWR VAPOUR-Line (VWR; TUD) was used. The bench top Sanoclav autoclave (Wolf; TUD) was utilized to sterilize small volumes or single glass flasks.

Äkta pure. Fast protein liquid chromatography was conducted with the fully automated Äkta pure protein purification system (GE Healthcare) using a 10 mL superloop for lysate loading onto the HisTrap HP 5 mL chromatography column (GE Healthcare).

Cell disruption. For protein purification, cells were lysed via ultrasonification with the SONOPULS ultrasonicator HD 2070 (Bandelin).

Centrifuges. The bench top centrifuge 5415D (Eppendorf) was used for most standard applications involving tubes of 1.5 – 2 mL size. A cooling centrifuge 5418R (Eppendorf) was used for volumes up to 2 mL that had to be kept cool throughout the entire centrifugation process. Samples of up to 50 mL were centrifuged in the Allegra™ X-22R centrifuge (Beckmann-Coulter) with or without cooling. In general, competent cell preparation as well as protein concentration was carried out in a Heraeus Multifuge X3R (Thermo Scientific) whereas the harvesting of cells (big rotor) or the removal of cell debris after cell lysis (small rotor) was done in the Sorval RC6+ centrifuge (Thermo Scientific).

Clean bench. All working procedures including sterile material and bacterial handling were carried out under a Heraeus HERAsafe® clean bench (Thermo Fisher). Before starting work, the clean bench was decontaminated with EtOH and sterilized via UV light.

Electrophoresis. Agarose gel electrophoresis was performed by connecting the Mini-Sub® cell GT electrophoresis chamber (BioRad) to the electrophoresis power supply EPS 3011/601 (Amersham pharmacia biotech) and applying a voltage of 120 V for 25 - 35 min. SDS gels were installed into a MiniPROTEAN® electrophoresis chamber filled with buffer and an electric current of 35 mA/gel (200 V, 55 Watt) was applied by connecting the chamber to the PowerPac Basic power supply (BioRad).

Electroporator. Transformation of plasmids into electrocompetent *E. coli* cells was executed using the MicroPulser™ Electroporation Apparatus (BioRad). Used electroporation cuvettes were rinsed with EtOH and H_2O and then exposed to UV light in a CL1000S UV crosslinker cabinet (VWR).

Gel documentation. DNA-containing agarose gels were imaged using the GeneGenius Bio Imaging System Gel Documentation Transilluminator (Syngene) coupled to the GeneSnap data analysis software. Visualization of protein purification results was performed by scanning of the gels in a commercial scanning device.

Heating cabinets. Bacterial cultivation on solid medium as well as the incubation of restriction digests took place in INCU-line incubation cabinets (VWR) heated to either 30 °C or 37 °C.

Incubating shakers. Liquid media bacterial cultivation was performed at different temperatures and shaking velocities in Multitron and Ecotron incubating shakers (Infors HT). Small volumes up to 2 mL, for example enzyme assays, were left to shake and incubate on a bench top ThermoMixer C or the Thermomixer comfort 5355 (both Eppendorf).

Lyophilization. Residual H_2O after solvent evaporation was removed from the product by lyophilizing the sample using the Alpha 2-4 device (Christ) connected to a chemistry hybrid pump RC6 (Vacuubrand).

Mass spectrometer. An electro spray ionization (ESI) source was used to generate detectable ions for all MS measurements. For full functionality, the Advion expression CMS single quadrupole mass spectrometer was combined with a Genius N118LA nitrogen gas generator (Peak Scientific) and a vacuum pump RV12 (Edwards). TLC-MS or LC-MS measurements were performed by coupling the mass analyzer with a PlateExpress™ TLC-MS interface (Advion) and the Jasco PU-1580 Intelligent HPLC Pump. Alternatively, low concentration samples were measured with the more sensitive LCQ Fleet™ Ion Trap mass spectrometer (Thermo Fisher) which was combined with the UltiMate 3000 HPLC system (Thermo Fisher) for LC-MS analysis.

NMR devices. Standard ^{1}H-NMR spectra were generated either in a Bruker Avance™ III HD 300 MHz or Bruker Avance™ I 500 MHz device in the analytics department at TUM. Analogous instruments were utilized at TUD. For ^{13}C-NMR, samples were measured with the Bruker Avance™ III 500 MHz C spectrometer. Full spectral analyses were carried out on a Bruker Avance™ III 600 MHz device. Samples to be measured were preferably prepared in deuterated MeOD, $CDCl_3$ or DMSO-d6 solvents. Recorded spectra were analyzed using MestReNova

software (MestreLab Research S.L.), with the chemical shifts δ presented in parts per million (ppm) and coupling constants J given in Hertz (Hz).

Nanophotometer. Determination of DNA and purified protein concentrations was achieved with the NanoPhotometer® P330 (Implen).

PCR cyclers. DNA multiplication and colony screening were performed in T100 Thermal Cyclers (BioRad).

Photometer. Bacterial cell culture growth was monitored by measuring the optical density at 600 nm wavelength (OD_{600}) with a BioPhotometer (Eppendorf).

Rotary evaporator. Solvent from extraction or synthesis procedures was evaporated using Hei-VAP Core (Heidolph) or Rotavapor® R-300 glass (Büchi) evaporator apparatuses combined with heated water baths and coupled to the SCC950 vacuum pump system (KnF).

Vacuum concentrator. The SpeedVac centrifugal vacuum concentrator 5301 (Eppendorf) was used to remove solvents, in particular H_2O, from sample sizes up to 2 mL volume, for example to concentrate DNA or to prepare enzyme assay reactions for LC-MS analysis.

5.2 Chemicals, solvents and kits

Chemicals. All chemicals used throughout this thesis were obtained from one of the following companies: Sigma-Aldrich, Carl Roth, Alfa Aeser, VWR, Fisher Scientific, AppliChem, Merck, Acros Organics and Carbolution. Unless designated otherwise, all chemicals were of high pureness and could be applied without prior purification or filtration steps necessary.

Solvents. For this thesis, solvents were needed for many experimental procedures, including ethyl acetate (EtOAc), methanol (MeOH), acetone (mainly for extraction), ethanol (EtOH), isopropanol (*iso*-Prop; as buffer components), HPLC-grade acetonitrile (ACN) and HPLC-grade methanol (for LC-MS and MS measurements), all bought from Fisher Scientific. Solvents used for HPLC analysis were supplemented with 0.05 % TFA or 0.1 % FA (Fisher Scientific) to enhance sample retention time and make peaks sharper. Ultrapure distilled deionized water used for buffers, LC-MS measurements etc. was prepared with a TKA GenPure ultrapure water system (JWT GmbH). Deuterated solvents such as $CDCl_3$, CD_3OD and DMSO-d_6 for NMR sample preparation were obtained from Sigma-Aldrich or Acros Organics.

Gases. All gases used were obtained from Linde or Praxair/Nippon Gases.

Kits. Commercially available kits for DNA isolation and purification were obtained from NEB (Monarch® DNA Gel Extraction Kit), Jena Bioscience (PCR Purification Kit) and VWR (peqGOLD Plasmid Miniprep Kit I).

Enzymes. Restriction enzymes and their corresponding buffers were ordered from NEB or Thermo Scientific, preferably the high fidelity (HF) versions. Q5® High-Fidelity DNA Polymerase and Phusion® High-Fidelity DNA Polymerase were obtained from NEB, KOD Xtreme™ Hot Start Polymerase was received from Merck.

5.3 Buffers & Media

5.3.1 Media

All prepared media were autoclaved for sterilization before use.

5.3.1.1 Liquid media

- Luria Betrani Broth (LB)

Component	Amount
Tryptone	10 g
Yeast extract	5 g
NaCl	10 g
ddH_2O	add to 1 L

- Terrific Broth (TB)

The 900 mL medium mixture was autoclaved before adding 100 mL of sterile filtered phosphate buffer (10x).

Component	Amount
Tryptone	12 g
Yeast extract	24 g
Glycerol	4 mL
ddH_2O	900 mL

10x phosphate buffer	
Component	**Amount**
K_2HPO_4 x 3 H_2O	164.325 g
KH_2PO_4	23.125 g
ddH_2O	add to 1 L

- 2x YT

After dissolving all media components, the pH was adjusted to 7.0 with NaOH.

Component	Amount
Tryptone	16 g
Yeast extract	10 g
NaCl	5 g
ddH_2O	add to 1 L

- SOC

The 971 mL medium mixture was autoclaved for sterilization purposes before adding sterile filtered salt and sugar solutions.

Component	Amount
Tryptone	20 g
Yeast extract	5 g
NaCl	0.58 g
KCl	0.19 g
ddH_2O	971 mL

Component	Stock	Amount
MgCl2 x 6 H2O	1 M	10 mL
MgSO4 x 7 H2O	1 M	10 mL
D-glucose	40 % (w/v)	9 mL

- CASO

After all media components were dissolved, the pH was adjusted to 7.3.

Component	Amount
Casein peptone	17 g
K_2HPO_4	2.5 g
NaCl	5 g
Soy peptone	3 g
D-glucose	2.5 g
ddH_2O	add to 1 L

- GYM

The pH was adjusted to 7.2 with KOH after all components have been dissolved in ddH_2O.

Component	Amount
Glucose	4 g
Yeast extract	4 g
Malt extract	10 g
ddH_2O	add to 1 L

- YEME

The pH of the 1 L liquid medium was adjusted to pH 7.0 with KOH before autoclaving the solution. Afterwards, sterile filtered sugar and amino acid components.

Component	Amount
Yeast extract	3 g
Malt extract	3 g
Peptone	5 g
D-glucose	10 g
Sucrose	340 g
ddH_2O	add to 1 L

Component	Stock	Amount
$MgCl_2$ x 6 H_2O	1 M	2 mL
Glycine	20 % (w/v)	25 mL

- R5A

All media components were added to 980 mL of ddH_2O and the pH was adjusted to 6.8 with KOH before filling up to 1 L. After autoclaving, 2 mL/L of trace elements solution were added.

Component	Amount
Saccharose	103 g
K_2SO_4	0.25 g
$MgCl_2$ x 6 H_2O	10.12 g
Glucose	10 g
Casamino acids	0.1 g
Yeast extract	5 g
MOPS	21 g
ddH_2O	add to 1 L

trace elements solution

Component	Amount
ZnCl2	40 mg
$FeCl_3$ x 6 H_2O	200 mg
$CuCl_2$ x 2 H_2O	10 mg
$MnCl_2$ x 4 H_2O	10 mg
$Na_2B_40_4$ x 10 H_2O	10 mg
$(NH_4)_6Mo_7O_{24}$ x 4 H_2O	10 mg
ddH_2O	add to 1 L

- M9 mineral medium

As a basis for the medium, 867 mL of ddH_2O were autoclaved prior to adding the following components.

Component	Stock	Amount
M9 salt solution	10 x	100 mL
Glucose	20 % (w/v)	20 mL
$MgSO_4$	1 M	1 mL
$CaCl_2$	1 M	0.3 mL
Biotin	1 mg/mL	1 mL
Thiamin	1 mg/mL	1 mL
Trace elements solution	100 x	10 mL
ddH_2O		867 mL

10x M9 salt solution

Component	Amount
Na_2HPO_4 x 2 H_2O	75.2 g
KH_2PO_4	30 g
NaCl	5 g
NH_4Cl	5 g
ddH_2O	add to 1 L

100x trace elements solution

Component	Amount
EDTA	5 g
$FeCl_3$ x 6 H_2O	0.83 g
$ZnCl_2$	84 mg
$CuCl_2$ x 2 H_2O	13 mg
$CoCl_2$ x 2 H_2O	10 mg
H_3BO_3	10 mg
$MnCl_2$ x 4 H_2O	1.6 mg
ddH_2O	add to 1 L

For the 10 x M9 salt solution, all salts were dissolved in 800 mL of ddH_2O and the pH was set to 7.2 with NaOH. Once the pH was adjusted the salt solution was filled up to a total volume of 1 L and sterilized in the autoclave.

The 100 x trace elements solution was prepared by initial dissolving of EDTA in 800 mL of ddH_2O followed by pH adjusting to 7.5 using NaOH. The other salt components were then added and the total volume of 1 L was sterile filtered.

- ISP-4

All media components except for the trace salts solution were dissolved in ddH_2O and autoclaved. The trace salts solution was made sterile by filtration and then added to the sterile ISP-4 medium.

Component	**Amount**
Soluble starch	10 g
K_2HPO_4 x 3 H_2O	1 g
$MgSO_4$	1 g
NaCl	1 g
$(NH_4)_2SO_4$	2 g
$CaCO_3$	2 g
Trace salts solution	1 mL
ddH_2O	add to 1 L

Trace salts solution

Components	**Amount**
$FeSO_4$ x 7 H_2O	0.1 g
$MnCl_2$ x 4 H_2O	0.1 g
$ZnSO_4$ x 7 H_2O	0.1 g
ddH_2O	100 mL

- N-Z-Amine

All media components were dissolved in 950 mL of ddH_2O and the pH was adjusted to 7.2. Afterwards, ddH_2O was added to a total volume of 1 L.

Component	**Amount**
Glucose	10 g
Soluble starch	20 g
Yeast extract	5 g
N-Z-Amine	5 g
ddH_2O	add to 1 L

5.3.1.2 Solid media

- LB agar

500 mL of agar were put into a 1 L bottle to avoid overflowing of medium upon autoclaving.

Component	Amount
Tryptone	10 g
Yeast extract	5 g
NaCl	10 g
Agar	15 g
ddH_2O	add to 1 L

- MS agar

The main media components were sterilized by autoclaving before sterile filtered salt solutions were added.

Component	Amount
Mannitol	10 g
Soya flour	10 g
Agar	10 g
ddH_2O	475 mL

Component	Stock	Amount
$MgCl_2$	0.5 M	10 mL
$CaCl_2$	2 M	15 mL

- CASO agar

The pH was adjusted to 7.2 with KOH. 500 mL of agar were put into a 1 L bottle to avoid overflowing of medium upon autoclaving.

Component	Amount
Peptone from caseine	15 g
Peptone from soy	5 g
NaCl	5 g
Agar	15 g
ddH_2O	add to 1 L

- GYM agar

The pH was set to 7.2 with KOH before sterilizing by autoclaving. 500 mL of agar were put into a 1L bottle to avoid overflowing of medium upon autoclaving.

Component	Amount
Glucose	4 g
Yeast extract	4 g
Malt extract	10 g
$CaCO_3$	2 g
Agar	12 g
ddH_2O	add to 1 L

- ISP-4 agar

All media components except for the trace salts solution were dissolved in ddH_2O and autoclaved. The trace salts solution was made sterile by filtration and then added to the sterile ISP-4 medium. 500 mL of agar were put into a 1L bottle to avoid overflowing of medium upon autoclaving.

Component	Amount
Soluble starch	10 g
K_2HPO_4 x 3 H_2O	1 g
$MgSO_4$	1 g
NaCl	1 g
$(NH_4)_2SO_4$	2 g
$CaCO_3$	2 g
Trace salts solution	1 mL
Agar	20 g
ddH_2O	add to 1 L

Trace salts solution

Components	Amount
$FeSO_4$ x 7 H_2O	0.1 g
$MnCl_2$ x 4 H_2O	0.1 g
$ZnSO_4$ x 7 H_2O	0.1 g
ddH_2O	100 mL

5.3.2 Buffers

5.3.2.1 Buffers used for cloning and DNA handling

- gDNA preparation

Lysis buffer

Component	Amount
EDTA	25 mM
Tris-HCl	25 mM
Sucrose	0.3 mM
pH 7.5	

TE buffer

Component	Amount
Tris-HCl	10 mM
EDTA	1 mM
pH 7.5	

CTAB solution

Component	Amount
CTAB	10 % (w/v)
NaCl	0.7 M

- Colony screening PCR

After dissolving all components, the buffer pH was adjusted with HCl to 8.8.

10 x Taq reaction buffer

Component	Amount
Tris-HCl	4.85 g
$(NH_4)_2SO_4$	2.64 g
KCl	1.49 g
$MgSO_4$	0.48 g
TritonX-100	2 mL
ddH_2O	add to 200 mL

- Agarose gel composition and sample preparation

TAE buffer

1 x TAE buffer used for running agarose gels was made by dilution of a 50x stock with ddH_2O. A ready-to-use ROTIPHORESE® 50x TAE buffer concentrate was obtained from Roth and diluted 1:50 for electrophoresis.

50x TAE buffer

Component	Amount
Tris-HCl	2 M
Acetic acid	1 M
EDTA	50 mM
pH 8.5	

DNA loading dye

Components were dissolved by vigorous vortexing to yield a green liquid. Aliquots of 1 mL were made and stored at -20 °C.

10 x DNA loading dye

Component	Amount
Sucrose	20 g
Orange G	0.075 g
Xylen cyanol	0.025 g
Bromphenol blue	0.025 g
ddH_2O	add to 50 mL

- Bacterial cell cultivation and storage

Antibiotic stock solutions

All water-soluble antibiotics were sterile filtered before use. Antibiotic stock solutions were applied in dilution ratios of 1:1000.

Antibiotic	Abbreviation	Stock concentration	Solvent
Kanamycin	Kan	50 mg/mL	ddH_2O
Ampicillin	Amp	100 mg/mL	50 % EtOH
Chloramphenicol	Cam	25 mg/mL	absolute EtOH
Tetracyclin	Tet	10 mg/mL	70 % EtOH
Apramycin	Apr	30 mg/mL	ddH_2O
Nalidixic acid	NA	25 mg/mL	0.3 M NaOH

- Chemocompetent cells

The pH of the transformation buffer was set to 6.7 with KOH before autoclaving. Afterwards, 5 mL of the sterile filtered $MnCl_2$ solution were added.

Transformation buffer

Component	Amount
HEPES	0.238 g
$CaCl_2$ x 2 H_2O	0.221 g
KCl	1.864 g
ddH_2O	95 mL

$MnCl_2$ solution

Component	Amount
$MnCl_2$ x 2 H_2O	1.78 g
ddH_2O	10 mL

- Cryo stock solutions

Streptomyces cryo solution

Component	Amount
Sucrose	20 g
Glucose	10 g
ddH_2O	70 mL
or	
Glycerol	50 % (v/v)

E. coli cryo solution

Component	Amount
Glycerol	50 % (v/v)

5.3.2.2 Buffers for protein expression and enzyme assays

- Ni-NTA protein purification buffers for His_6-tagged proteins

Manual protein purification

Lysis buffer

Component	Amount	
$NaH_2PO4 \times 2\ H_2O$	50 mM	7.80 g
NaCl	300 mM	17.53 g
Imidazole	**10 mM**	0.68 g
Glycerol	10 %	100 mL
ddH_2O	add to 1 L	
pH 8.0		

Elution buffer

Component	Amount	
$NaH_2PO4 \times 2\ H_2O$	50 mM	7.80 g
NaCl	300 mM	17.53 g
Imidazole	**250 mM**	17.02g
Glycerol	10 %	100 mL
ddH_2O	add to 1 L	
pH 7.5		

Wash buffer 1

Component	Amount	
$NaH_2PO4 \times 2\ H_2O$	50 mM	7.80 g
NaCl	300 mM	17.53 g
Imidazole	**20 mM**	0.68 g
Glycerol	10 %	100 mL
ddH_2O	add to 1 L	
pH 7.5		

Wash buffer 2

Component	Amount	
$NaH_2PO4 \times 2\ H_2O$	50 mM	7.80 g
NaCl	300 mM	17.53 g
Imidazole	**40 mM**	17.02 g
Glycerol	10 %	100 mL
ddH_2O	add to 1 L	
pH 7.5		

Wash buffer 3

Component	Amount	
NaH_2PO4 x 2 H_2O	50 mM	7.80 g
NaCl	300 mM	17.53 g
Imidazole	**80 mM**	0.68 g
Glycerol	10 %	100 mL
ddH_2O	add to 1 L	
pH 7.5		

Wash buffer 3A

Component		Amount
NaH_2PO4 x 2 H_2O	50 mM	7.80 g
NaCl	300 mM	17.53 g
Imidazole	**50-70 mM**	3.40 – 4.77 g
Glycerol	10 %	100 mL
ddH_2O	add to 1 L	
pH 7.5		

Äktapure automated His_6-tagged protein purification

Prepared buffers were degassed for 30 min using an ultrasonic bath and filtered with a filtration device before use in the automated chromatography system.

Buffer A

Component	Amount	
NaH_2PO4 x 2 H_2O	50 mM	7.80 g
NaCl	300 mM	17.53 g
Imidazole	**80 mM**	0.68 g
Glycerol	10 %	100 mL
ddH_2O	add to 1 L	
pH 7.5		

Buffer B

Component		Amount
NaH_2PO4 x 2 H_2O	50 mM	7.80 g
NaCl	300 mM	17.53 g
Imidazole	**500 mM**	34.04 g
Glycerol	10 %	100 mL
ddH_2O	add to 1 L	
pH 7.5		

- Pys protein purification

Lysis buffer

Component	Amount
NaH2PO4	50 mM
NaCl	300 mM
pH 8.0	

- Protein solubility enhancement

Peternel buffer

Component	Amount
Tris	50 mM
NaCl	150 mM
N-lauroylsarcosine	0.2 % (w/v)
pH 8.0	

- Preparative protein excision from gel

Electroelution buffer native PAGE

Component	Amount
Tris	0.25 M
pH 6.8	

- Protein storage and enzyme assays

General buffers

Tris-HCl buffer

Component	Amount
Tris	50 mM
Glycerol	10 %
pH 7.5	

HEPES buffer

Component	Amount
HEPES	100 mM
Glycerol	1.5 %
pH 7.8	

Pys protein buffer

Pys protein reaction buffer

Component	Amount
Tris	100 mM
$MgCl_2$	25 mM
EDTA	2.5 mM
pH 7.5	

- SDS-PAGE buffers

Gel preparation

4x separating buffer

Component	Amount
Tris	90.855 g
SDS	0.8 % (w/v)
ddH_2O	500 mL

4x stacking buffer

Component	Amount
Tris	30.285 g
SDS	0.8 % (w/v)
ddH_2O	500 mL

Separating gel

Components	Amount*			
% gel	8	10	12	15
AA (40 % v/v)	1.1 mL	1.4 mL	1.7 mL	2.1 mL
4x separating buffer	1.4 mL			
ddH_2O	3.1 mL	1.8 mL	2.5 mL	2.1 mL
APS (10 % w/v)	37.5 µL			
TEMED	5.625 µL			

Stacking gel

Components	Amounts*
% gel	4
AA (40 % v/v)	0.2 mL
4x stacking buffer	0.5 mL
ddH_2O	1.3 mL
APS (10 % w/v)	20 µL
TEMED	2 µL

*Amounts are given for the making of 1 SDS gel consisting of separating and stacking gel components.

- Sample buffer

Components were mixed by vigorous vortexing and aliquoted before adding 5 µL of SDS sample buffer to 20 µL of protein sample.

5x SDS sample buffer

Component	Stock	Amount	
Tris	1 M, pH 6.8	25 % (v/v)	3.75 mL
β-Mercaptoethanol		5 % (v/v)	0.75 mL
SDS		10 % (w/v)	1.5 g
Bromphenol blue		0.2 % (w/v)	0.03 g
Glycerol		50 % (v/v)	7.5 mL
ddH_2O		add to 15 mL	

- Electrophoresis buffer

1 x SDS buffer used for running SDS gels was made by dilution of a 10x stock with ddH_2O. A ready-to-use ROTIPHORESE® SDS-PAGE 10x concentrate was obtained from Roth and diluted 1:10 for electrophoresis.

10x SDS running buffer

Component	**Amount**
Tris	30.28 g
Glycine	187.7 g
SDS	10 g
ddH_2O	add to 1L

- Native PAGE buffers

Gel preparation

4x separating buffer

Component	**Amount**
Tris	18.2 g
1 M HCl	adjust to pH 8.9
ddH_2O	to 100 mL

7x stacking buffer

Component	**Amount**
Tris	5.7 g
1 M HCl	adjust to pH 6.7
ddH_2O	to 100 mL

Separating gel (7.5 %)

Components	**Amount***
4x separating buffer	0.75 mL
AA (30 % v/v)	0.75 mL
ddH_2O	1.5 mL
TEMED	2 µL
APS (10 % w/v)	21 µL

Stacking gel (4.3 %)

Components	**Amount***
7x stacking buffer	0.25 mL
AA (30 % v/v)	0.25 mL
ddH_2O	1.25 mL
TEMED	2 µL
APS (10 % w/v)	10 µL

*Amounts are given for the making of 1 native gel consisting of separating and stacking gel components.

Sample loading buffer

All components were mixed by vortexing before applying sample loading buffer to a final concentration of 1x.

Native PAGE sample loading buffer (3x)

Components	**Amount**
Glycerol	3 mL
Running buffer (50x)	0.6 mL
Bromphenol blue	tip of spatula
ddH_2O	to 100 mL

Electrophoresis buffer

For electrophoresis, 50x running buffer was diluted to a final concentration of 1x with ddH_2O.

Native PAGE 50x running buffer

Components	**Amount**
Tris	7.5 g
Glycine	36 g
ddH_2O	to 250 mL

- PAGE gel staining and destaining

Coomassie staining solution

Component	**Amount**	
Coomassie Brilliant Blue	0.02 % (w/v)	0.2 g
Acetic acid	10 % (v/v)	100 mL
Isopropyl alcohol	25 % (v/v)	250 mL
ddH_2O	add to 1 L	

Destaining solution

Component	**Amount**
EtOH	400 mL
Acetic acid	100 mL
ddH_2O	500 mL

5.4 Organisms, vectors, plasmids and primers

5.4.1 Organisms

- *E. coli* host strains

Strain designation	Genotype	Competency	Reference/Origin
E. coli DH5α	*F– φ80lacZΔM15 Δ(lacZYA-argF)U169 recA1 endA1 hsdR17(rK–, mK+) phoA supE44 λ– thi-1 gyrA96 relA1*	chemo	Lab stock NEB
E. coli DH10β	*F– mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara-leu)7697 galU galK λ– rpsL(StrR) nupG*	chemo	Lab stock NEB
E. coli DH10β::mtaA		chemo	Lab stock[79]
E. coli Top10	*F- mcrA (mrr-hsdRMS-mcrBC) 80lacZ M15 lacX74 recA1 ara 139 (ara-leu)7697 galU galK rpsL (StrR) endA1 nupG*	chemo	Lab stock Thermo Scientific
E. coli BL21DE3	*F– ompT hsdSB (rB–, mB–) gal dcm (DE3)*	chemo	Lab stock Thermo Scientific
E. coli SoluBL21DE3	*F- ompT hsdSB (rB-mB-) gal dcm (DE3)†*	chemo	Genlantis
E. coli BAP1	*F−ompT hsdSB(rB− mB−) gal dcm (DE3) ΔprpRBCD (sfp)*	chemo	Lab stock[199]
E. coli Arctic Express	*F– ompT hsdS(rB–mB–) dcm+ Tetr gal λ(DE3) endA Hte [cpn10cpn60 Gentr]*	chemo	Agilent
E. coli EPI300™	*F– λ– mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 Δ(lac)X74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK rpsL (StrR) nupG' trfA dhfr*	electro	Epicentre
E. coli ET12567pUZ8002	*F- dam-13::Tn9 dcm-6 hsdM hsdR zjj-202::Tn10 recF143 galK2 galT22 ara-14 lacY1 xyl-5 leuB6 thi-1 tonA31 rpsL136 hisG4 tsx-78 mtl-1 glnV44*	electro	Lab stock Prof. Dr. Leonard Kaysser (University Tübingen)

† The SoluBL21 strain contains uncharacterized mutations obtained through special selection criteria. These mutations make the strain able to express insoluble proteins in soluble form, fully or partially, in most tests conducted.

- *Streptomyces* host strains

Strain designation	Genotype	Reference/Origin
Streptomyces coelicolor M1154	*Δact Δred Δcpk Δcda rpoB [C1298T] rpsL [A262G]*	Lab stock
Streptomyces lividans TK24	*str-6 SLP2− SLP3−*	Lab stock
Streptomyces albus B2P1	Based on *S. albus* J1074 wild type but with deleted BGCs	Prof. Dr. Andriy Luzhetzkyy (University Saarland)[131]
Streptomyces albus B3P1		
Streptomyces albus B4		
Streptomyces albus Del14		

- Bacterial strains exhibiting PA-producing gene clusters

Strain designation	Origin
Xenorhabdus stockiae DSM 17904	Prof. Dr. Helge Bode (University Frankfurt)
Photorhabdus temperata DSM 15199	DSMZ
Micromonospora auratinigra DSM44815^T	ZIEL institute (TUM)
Streptomyces fulvissimus DSM40593	DSMZ
Streptomyces sp. Mg1	Prof. Dr. Paul Straight (Texas A&M, U.S.A)
Streptomyces cattleya DSM46488	DSMZ
Streptomyces clavuligerus F613-2	Prof. Dr. Guangxiang Cao (Shandong Academy of Medical Sciences, China)
Actinosynnema mirum DSM43827	DSMZ
Pseudomonas aeruginosa PAO1	Prof. Dr. Stephan Sieber (TUM)
Pseudomonas fluorescens HKI0770	Prof. Dr. Pierre Stallforth (HKI Jena)

5.4.2 Vectors

All listed vectors were used as backbone structures for cloning efforts conducted within this thesis.

Vector	Size [bp]	Function	Resistance/Inducer	Origin
pACYC-Duet-1	4008	PE	Cam^R	Lab stock
pBAD-His6	4636	PE	Amp^R/ L-arabinose	Prof. Dr. Katrin Lang (TUM)
pCC1FOS	8139	LCHR	Cam^R	Lab stock
pCX2	6338	SE	Kan^R/ L-arabinose	Prof. Dr Helge Bode (University Frankfurt)
pCX2-NHis6	6542	PE / SE	Kan^R/ L-arabinose	This work
pETDuet-1-RBS2M	5438	PE / SE	Amp^R / IPTG	Lab stock
pET28a(+)	5369	PE	Kan^R / ITPG	Prof. Dr. Pierre Stallforth (HKI Jena)
pET28b-SUMO	5650	PE	Kan^R / ITPG	Lab stock
pET28bpTetGFPv2	6552	SE	Kan^R / tetracyclin	Lab stock
pGS21a-TEV	6175	PE	Amp^R / IPTG	Lab stock
pHis8C(+)-TEV	5323	PE	Kan^R / ITPG	Lab stock
pMal-6His-TEV-MCS	5609	PE	Amp^R / IPTG	Lab stock
pSET152ermE-rev	5830	SE	$Apra^R$	Lab stock

PE: protein expression, SE: substrate expression

5.4.3 Plasmids

The following plasmids were constructed throughout the course of this work for protein or substrate expression purposes. Listed plasmids represent a selection, not all constructs established within the course of this thesis.

Plasmid	Size [bp]	Function	Origin
BraC-pHi8CTEV	6502	PE	Françoise Schaefers
MiC21-pSET152ermE*rev	19451	SE	This work
MiC21 NRPS-pHis8CTEV	12422	PE	This work
MiC21 MOX-pET28bSUMO	6812	PE	This work
PAO1C29s-pSET152ermE*rev	14060	SE	This work
PAO1C29-pSET152ermE*rev	21349	SE	This work
PAO1C29s-pETpTetGFPv2	14779	SE	This work
PAO1C29 NRPS-pHis8CTEV	12380	PE	This work
PAO1C29 MOX-pET28bSUMO	6815	PE	This work
PAO1C29 AzeJ-pGS21aTEV	6925	PE	This work
PTC12-pCX2	14718	SE	This work
PTC12-pETpTetGFPv2	14933	SE	This work
PTC12 NRPS-pGS21aTEV	13387	PE	This work
PTC12 MOX-pET28bSUMO	6812	PE	This work
PF Pys-PCP-pET28bSUMO	5858	PE	This work
PF Pys NRPS-pET28a	9261	PE	Martin Klapper
SclavuC39-pSET152ermE*rev	14368	SE	This work
SclavuC39 NRPS-pSET152ermE*rev	13170	SE	This work
SclavuC39 NRPS-pSET152ermE*rev-kasOP	13286	SE	This work
SclavuC39 NRPS-pSET152ermE*rev-SF14	13272	SE	This work
SclavuC22-pET28bSUMO	6842	PE	This work
SclavuC39-pET28bSUMO	6848	PE	This work
SFC2-pSET152ermE*rev	20396	SE	This work
SFC2 NRPS-pSET152ermE*rev	13170	SE	This work
SFC2 NRPS-pET28bSUMO	12878	PE	This work
SMg1C5 T1-pSET152ermE*rev	8701	SE	This work
SMg1C5 T1+T2-pSET152ermE*rev	18455	SE	This work
SMg1C5-pSET152ermE*rev	19385	SE	This work
SMg1C5-pETpTetGFPv2	20098	SE	This work
SMg1C5 NRPS-pET28bSUMO	12851	PE	This work
SMG1C5 MOX-pET28bSUMO	8171	PE	This work
PxaA-pCX2	13555	SE	Helge Bode

Plasmid	Size [bp]	Function	Origin
PxaB-pCX2	7510	PE	This work
PxaA-pHis8CTEV	12545	PE	This work
PxaB-pET28bSUMO	6824	PE	Françoise Schaefers
PxaA-PxaB-pCX2	14730	SE	This work
PxaA-BraC-pCX2	14763	SE	This work
PxaA-MiC21 MOX-pCX2	14721	SE	This work
PxaA-PAO1C29 MOX-pCX2	14718	SE	This work
PxaA-MiC21 MOX-MiC21s-pCX2	17359	SE	This work
PxaA-MiC21 MOX-MiC21-pCX2	20067	SE	This work
PxaA-MiC21 MOX-PAO1C29-pCX2		SE	This work
PxaA-MiC21 MOX-SCDH-pCX2	15600	SE	This work
PxaA-MiC21 MOX-araBAD-MiC21-pCX2	21422	SE	This work
PxaA-MiC21 MOX-T7-MiC21-pCX2	20327	SE	This work
PxaA-MiC21 MOX-araBAD-PAO1C29-pCX2	23352	SE	This work

PE: Protein expression; SE: Substrate expression

5.4.4 Primers

Lists of used primers.

- *In vivo* heterologous expression systems

Name	Sequence 5' → 3'
MiC21 F + 1bp spac pSET152ermE*rev GA	agcggagcaacggaggtacggaaggtGTGGGTCTCTCCCATCGCAT
MiC21 F + 8bp spac pSET15ermE*rev GA	agcaacggaggtacggaaggttgtattcGTGGGTCTCTCCCATCGCAT
MiC21 R + pSET152ermE*rev GA	ctgcaggtcgactctagaggaggCTAGCGGTTGAGGTAGGCCG
PAO1C29s F + pETpTetGFP GA	tgatagagaagaggatcgaccATGGTTCGTTTCGCTCGCTTGCCGCTATC
PAO1C29s R + pETpTetGFP GA+	agttcttcacctttgctaacCATAAGCTTTCATTCGACTCCTGGGAAGGTCAGTGCCA
PAO1C29 NRPS R + pETpTetGFP GA+	gttcttcacctttgctaacCATAAGCTTTCATGGGCGGGTCCGTTCGCTG
pSETFr1 F + PAO1C29 Fr2 GA	aagatcgttccgacaatggactgatgcatatgCCTCTCTAGAGTCGACCTGCAGCCCAAGC
pSETFr1 R + PAO1C29 Fr2 GA	gggcgagagaggtttgctggtgaataacatgtgTCATTGCTCCTCGAGGGCGGATCGTT
pSETFr2 F + PAO1C29 Fr3 GA	taggcggggcgctgggccgctgatgtgCCTCTCTAGAGTCGACCTGCAGC
pSETFr2 R + PAO1C29 Fr3 GA	tggtcttccatgtcgttgggcatgtgTCAGTCCATTGTCGGAACGATCTT
PAO1C29 Fr3 F + GA pSET152ermE*rev F12	gctggccgaagatcgttccgacaatggactgacacATGCCCAACGACATGGAAGACCATCTCCTC
PAO1C29 Fr3 R + GA pSET152erme*rev F12	gacggccagtgccaagcttgggctgcaggtcgactctagagaggcacaTCAGCGGCCCAGCGCCC
PTC12 F + pCX2 GA	tttcacacaggaggctagcaatgaaacaattacaaaaagtaatatctc
PTC12 R + pCX2 GA	tatccaattgagatctgccatcaactatccagggtaaatatc
pETpTetGFP F + PTC12 GA	gatatttaccctggatagttgaATGGTTAGCAAAGGTGAAGAACT
pETpTetGFP R + PTC12 GA	atattacttttgtaattgtttcatGGTCGATCCTCTTCTCTATCACTG
SclavuC39 F + pSET152ermE*rev GA	agcaacggaggtacggaaggatgtattcGTGGCCGTTCGCCGCACG

Name	Sequence 5’ → 3’
SclavuC39 R + pSET152ermE*rev GA	gcaggtcgactctagagaggTCACACGGTGTGGATGTCCCG
linear pSET152ermE*rev-SclavuC39 F	GGCAGGGGACCAAGGCG
linear pSET152ermE*rev-SclavuC39 R	ACGGTGAACGTGTGCAGCC
SFC2 F	ATGGAAGGTGTCGCAGTGACCCCC
SFC2 R	TTAGTGCCTTGCGCCGCGAGTGGT
pSET152ermE*rev F + SFC2 GA	accactcgcggcgcaaggcactaaCCTCTCTAGAGTCGACCTGCAGC
pSET152ermE*rev R + SFC2 GA	gggtcactgcgacaccttccatgaatacaaCCTTCCGTACCTCCGTTGCT
pSET152-SFC2-HR-L-F + GA	caacggaggtacggaaggggccgccgggctcggac
pSET152-SFC2-HR-L-R + GA	ccgtcgcccgcggccgccgaagagcgaggacgccgcct
pSET152-SFC2-HR-R-F + GA	cgctcttcggcggccgcgggcgacggaggggtaga
pSET152-SFC2-HR-R-R + GA	aggtcgactctagagaggttcctgtgcagggaagaggag
pCC1FOS-SFC2-CAP L F	GAGACTATAGAATACTCAggccgccgggctcggac
pCC1FOS-SFC2-CAP L R	CTCCGTCGCCCAAGCTTcgaagagcgaggacgccgcct
pCC1FOS-SFC2-CAP R F	CTCGCTCTTCGAAGCTTgggcgacggaggggtaga
pCC1FOS-SFC2-CAP R R	TCGACCTGCAGGCATGCAttcctgtgcagggaagaggag
linear pSET152 SFC2 new F	GGGCGACGGAGGGGTAGA
linear pSET152 SFC2 new R	CTCCGCCCACAGCTGGA
SMg1C5 T1 F + pETpTetGFP GA	tgatagagaagaggatcgaccGTGACCGCTGCGGCGCCGGA

Name	Sequence 5’ → 3’
SMg1C5 T2 F + T1 GA	aggaaaccccaacagttgaaATGAACCACCACATGCCGTATCCCCTG
SMg1C5 T2 R + T1 GA	cttcacctttgctaaccataTCAGCTGCCGGTGGCGTTCC
SMg1C5 + pSET152ermE*rev GA T1 F	agcaacggaggtacggaaggttgtattcGTGACCGCTGCGGCGCC
SMg1C5 + pSET152ermE*rev GA R	ctgcaggtcgactctagagaggTCAGCTGCCGGTGGCGTTCC
SMg1C5 + pSET152ermE*rev GA T2 F	agcaacggaggtacggaaggttgtattcATGAACCACCACATGCCGTATCCCCT
SMg1C5 + pSET152ermE*rev GA + StuI T1 R	ctgcaggtcgactctagagagg**aggcct**TCAACTGTTGGGGTTTCCTTCGAGCC
BraC F + pXst4 GA	gatcgctgacgtcggtacccATGACGGAACCGCGTAAAGCTA
BraC R + pXst4 GA	cagcggtttctttaccagacGTCGTTGACATACGCCGCCAC
MiC21 MOX F + pXst4 GA	gatcgctgacgtcggtacccGTGAAGGCGATCATCATTGGC
MiC21 MOX R + pXst4 GA	cagcggtttctttaccagacTCAATGGCCATGATCATGCAC
MiC21 MOX R + StuI + pXst4 GA	cagcggtttctttaccagacaggcctTCAATGGCCATGATCATGCAC
PAO1C29 MOX F + pXst4 GA	gatcgctgacgtcggtacccATGAGCAGACATCCCCTGAAG
PAO1C29 MOX R + pXst4 GA	cagcggtttctttaccagacTCATTCGACTCCTGGGAAGGT
PxaB F + pXst4 GA	gatcgctgacgtcggtacccATGAGCCGCCCGCTGAAAGTGATTATTGC
PxaB R + pXst4 GA	cagcggtttctttaccagacCAGGGTCTGCAGGGTCAGATCGCCCA
MiC21 Rev + pXst4 GA	cagcggtttctttaccagacCTAGCGGTTGAGGTAGGCCG
MiC21 short Rev + pXst4 GA	cagcggtttctttaccagacCTATGCGGCGTCGATCTGCT
MiC21 SCDH F + pXst4-MOX GA	gtgcatgatcatggccattgaaggGTGTTCAGAAGGCGAAGACCGT
MiC21 SCDH R + pXst4-MOX GA	gcagcggtttctttaccagacaggTCAGTCGGTCGAGACAAGTGG

Name	Sequence 5’ → 3’
MiC21 back StuI F + pXst4-MiMOX GA	gtgcatgatcatggccattgaaggcctATGGCCATTGATGCCCCGGC
MiC21 back StuI R + pXst4-MiMOX GA	cagcggtttctttaccagacaggCTAGCGGTTGAGGTAGGCCGCC
T7 Prom F + pXst4-MiC21 MOX GA	gcatgatcatggccattgaaggTAATACGACTCACTATAGGGGAATTGTG
T7 Prom R + pXst4-MiC21 MOX GA	cggtttctttaccagacaggACCAGACTCGAGGGTACCGA
MiC21 back F +T7 GA	ataagaaggagatatacatatggcaATGGCCATTGATGCCCCGGC
MiC21 back R +T7 GA	gtggccggccgatatccaattgaCTAGCGGTTGAGGTAGGCCGCC
MiC21 back Fwd + T7 NotI GA	agctccgtcgacaagcttgcaATGGCCATTGATGCCCCGGC
MiC21 back Rev + T7 NotI GA	gggtggtggtggtggtggtgctcgagtgcCTAGCGGTTGAGGTAGGCCGCC
araBAD F + pXst4 MiC21 long BglII GA	gtgcatgatcatggccattgaaggaTTATGACAACTTGACGGCTA
araBAD R + pXst4 MiC21 long BglII GA	gccggggcatcaatggccattaaCATATGCTAGCCTCCTGTGT
MiC21 back BglII F + pXst4-MiMOX GA	gtgcatgatcatggccattgaaggagatctATGGCCATTGATGCCCCGGC
pXst4-MiC21 long-StuI lin Rev	CCTTCAATGGCCATGATCATGCACCG
pXst4-MiC21 long-StuI lin Fwd	CCTATGGCCATTGATGCCCCGGC
PAO1C29 back BglII F + pXst4 MiMOX GA	gtgcatgatcatggccattgaaggagatctATGATGGCCGAGATACGACGCCCGCTGTCC
PAO1C29 back BglII R + pXst4 MiMOX GA	cagcggtttctttaccagacaggTCAGCGGCCCAGCGCCCCGC
araBAD R + pXst4-PAO1C29 back GA	cgggcgtcgtatctcggccatcattaaCATATGCTAGCCTCCTGTGT
kasOP + StuI Fwd	aggcctTGTTCACATTCGAACGGTC
kasOP + XbaI Rev	tctagaGTCCGTACCTCCGTTGCT
SF-14 + StuI Fwd	aggcctCCTATCCAGGAGATATTATGAGTTACG
SF-14 + XbaI Rev	tctagaGTCCGTACCTCCGTTGC
gapdh + StuI Fwd	aggcctGCTGCTCCTTCGGTCGGA

Name	Sequence 5’ → 3’
gapdh + XbaI Rev	tctagaGTCCGTACCTCCGTTGCTCG
SclavuC39 NRPS + StuI Rev	gcaggtcgactctagaggaggcctTCACTCCCCGATCAGTTCGGTGACG
SclC39 MOX Fwd + kasOP GA	gttcgagcaacggaggtacggactctagaATGAGCAGGCGCAGGGCCCTT
SclC39 MOX Fwd + SF14 GA	ttacgagcaacggaggtacggactctagaATGAGCAGGCGCAGGGCCCTT
SclC39 MOX Rev + Promoter GA	gtgccaagcttgggctgcaggtcgactTCACACGGTGTGGATGTCCCGGG

- Protein expression primers.

Name	Sequence 5’ → 3’
MiC21 NRPS F + pHisCTEV GA	gatataccatggcgggtaccgcaGTGGGTCTCTCCCATCGCATTCC
MiC21 NRPS R + pHisCTEV GA	caagcttgtcgacgagctcgaattcgTCACGCCGCCTCTTCCCG
PAO1C29 NRPS F + pHis8CTEV GA	gatataccatggcgggtaccgcaATGGTTCGTTTCGCTCGCTTGCCGCTATC
PAO1C29 NRPS R + pHis8CTEV GA	caagcttgtcgacgagctcgaattcgTCATGGGCGGGTCCGTTCGCTG
PTC12-NRPS F + pGS21aTEV GA	ttcagggagccatggctgatATGAAACAATTACAAAAAGTAATATCTC
PTC12-NRPS R + pGS21aTEV GA	agctcgaattcggatccgatTTAATTGATGTTGCGGTTAG
PxaA F + pHisCTEV GA	tataccatggcgggtaccgcaATGAAAACTTCACAATTAGTACCTCTTAC
PxaA R + pHisCTEV GA	caagcttgtcgacgagctcgaattcgTCAATGGACGGGACATATTT
SFC2 NRPS F + pET28bSUMO GA	tcacagagaacagattggtggaATGGAAGGTGTCGCAGTGACCCCC
SFC2 NRPS R + pET28bSUMO GA	actggcggccgttactagtgTCAATGGCCGATCAGCTCGTTGAC
pET28bSUMO F + SMg1C5 NRPS GA	ccgaaccggccgggagggctgaCACTAGTAACGGCCGCCAGTGTGC
pET28bSUMO R + SMg1C5 NRPS GA	cggcatgtggtggttcattgcACCAATCTGTTCTCTGTGAGCCTCAA

Name	Sequence 5’ → 3’
BraC F	ATGACGGAACCGCGTAAAGCTA
BraC R	GTCGTTGACATACGCCGCCAC
pET28bSUMO F + MiC21 MOX GA	gtgcatgatcatggccattgaCACTAGTAACGGCCGCCAGTGTGC
pET28bSUMO R + MiC21 MOX GA	caatgatgatcgccttcactgcACCAATCTGTTCTCTGTGAGCCTCAA
pET28bSUMO F + PAO1C29 MOX GA	accttcccaggagtcgaatgaCACTAGTAACGGCCGCCAGTGTGC
pET28bSUMO R + PAO1C29 MOX GA	cttcaggggatgtctgctcattgcACCAATCTGTTCTCTGTGAGCCTCAA
pET28bSUMO F + PTC12 MOX GA	gatatttaccctggatagttgaCACTAGTAACGGCCGCCAGT
pET28bSUMO R + PTC12 MOX GA	gactttcaaaggtttgttcatGGATCCACCAATCTGTTCTCTG
PTC12 MOX F + pMALHis6TEV GA	gaatctttattttcagggccatatgggaATGAACAAACCTTTGAAAGTC
PTC12 MOX R + pMALHis6TEV GA	tcaagcttgaattcctcgagggtaccgTCAACTATCCAGGGTAAATATC
PxaB Fwd	ATGAGCCGCCCGCTGAAAGTGATTATTGC
PxaB Rev	CAGGGTCTGCAGGGTCAGATCGCCCA
pET28bSUMO F + SclC22 MOX GA	cacggcctgacggcgtgaCACTAGTAACGGCCGCCAGTGTGC
pET28bSUMO R + SclC22 MOX GA	accggccacgaggaccttcattgCACCAATCTGTTCTCTGTGAGCCTCAA
pET28bSUMO F + SclC39 MOX GA	cccgggacatccacaccgtgtgaCACTAGTAACGGCCGCCAGTGTGC
pET28bSUMO R + SclC39 MOX GA	aagggccctgcgcctgctcattgCACCAATCTGTTCTCTGTGAGCCTCAA
pET28bSUMO F + SMg1C5 MOX GA	tgtccgtcctcaccgagtgaCACTAGTAACGGCCGCCAGTGTGC
pET28bSUMO R + SMg1C5 MOX GA	accgatgaccagggctttcattgCACCAATCTGTTCTCTGTGAGCCTCAA
MiC21 SCDH F	GTGTTCAGAAGGCGAAGACCGT
MiC21 SCDH R	TCAGTCGGTCGAGACAAGTGG

Name	Sequence 5’ → 3’
FadD F + pHis8CTEV GA	gatataccatggcgggtaccgcaTTGAAGAAGGTTTGGCTTAAC
FadD R + pHis8CTEV GA	caagcttgtcgacgagctcgaattcgTCAGGCTTTATTGTCCACTT
PAO1 AzeJ F	GTGAGTCAGAACATGGATCTCACGAT
PAO1 AzeJ R	TCATGCCGCCGACGAACC
pET28bSUMO + AzeJ GA F	ggttcgtcggcggcatgaCACTAGTAACGGCCGCCAGTGTGC
pET28bSUMO + AzeJ GA R	gagatccatgttctgactcactgcACCAATCTGTTCTCTGTGAGCCTCAA

- General screening and sequencing primers,

Name	Sequence 5’ → 3’
KanR seq	AACTCACCGAGGCAGTTC
LacI seq	ATCAACTGGGTGCCAGC
Lac operator part Fwd	GGAATTGTGAGCGGATAAC
Lac promoter part Rev	CTTTATGCTTCCGGCTCG
M13 Fwd	TGTAAAACGACGGCCAGT
M13 Rev	CAGGAAACAGCTATGACC
M13 Rev long	CAGGAAACAGCTATGACATGATTA
pET Rev	CTAGTTATTGCTCAGCGG
pRSET Rev new	GGGTTATGCTAGTTATTGC
pRSET-Rev new long	CCAAGGGGTTATGCTAGTTATTGC
T7 Promoter Fwd	TAATACGACTCACTATAGGG
T7 Terminator Rev	CAAAAAACCCCTCAAGAC
T7 Terminator Rev Long	CAAAAAACCCCTCAAGACCCGTTTA
TetR Tn10 screen Fwd	GTGAGTATGGTGCCTATCTA
TEV Site Fwd	GAGAATCTTTATTTTCAGGGA
BraC screen Fwd	GTCATGACCGAAGGCTAT
MiC21 plasmid screen Fwd	CATCATCGATCGTTCCCC
MiC21 plasmid screen Rev	GTAGTCGATGTCGATGACG
MiC21 screen end Fwd	ATGACGGCGGTATCGGGC

Name	Sequence 5’ → 3’
MiC21 NRPS screen Fwd_gapdhP	CGAGGATCCAGCAAGTGT
MiC21 NRPS screen Fwd	CAGCCTGACCTACGACC
PAO1C29 plasmid screen Fwd	CTATGCGGTGATGGTCGC
PAO1C29 plasmid screen Rev	GTGTGGATCAGTTCCTCG
PAO1C29 plasmid screen F2	GCATCGAGCAGTTGGAGCAT
PAO1 MOX screen Rev	CAGTCCTTCGCCTGTTC
PTC12 NRPS screen Fwd 1	ATGAAACAATTACAAAAAGTAATATCTC
PTC12 NRPS screen Fwd 2	GTTATGACCACGGCACTTG
PTC12 NRPS screen Fwd 3	ACTATCGAGTGCTAAACAG
PTC12 NRPS screen Fwd 4	CTGATACGCTGCGTGACCG
PTC12 NRPS screen Fwd 3'end out	TCCACAAGCGAGCTCAAG
PTC12 NRPS screen Rev 1	TTAATTGATGTTGCGGTTAG
PTC12 NRPS screen Rev 2	CATATCGAACAGATTCTGA
PTC12 NRPS screen Rev 3	CAACATAAGGCAGATAAAT
PTC12 NRPS screen Rev 4	GTAGGCATCAATAAACATA
NRPS screen Rev 5'end out	CAGCCATACAATGATTGATAC
PxaA screen Fwd.2	GAATAGAGTGGCTGAGTACT
PxaA screen Rev.1	TATGACACATCTACGGTTGCC
Pys plasmid screen Rev	GTCATGCGTTGCAGATAG
SCC37 plasmid screen Rev	CGGTTGAACAGCGGTAC
SclavuC22 plasmid screen Rev	GAAGCCGACGAAGATGCT
SclavuC39 plasmid screen Rev	GACAGGGTCACGTCGTAG
SclavuC39 NRPS screen Fwd	GAACGTGCTCTGCTATGTG
SFC2 plasmid screen Rev	CAACGGCATGCTGTTCG
SFC2 long seq Fwd 1	CGCACGTCGATGTGGTC
SFC2 long seq Rev 1	GCATAGCAGGCGTTCGGA
SMg1C5 plasmid screen Rev	ATCGACGATCACCACCT

Name	Sequence 5’ → 3’
SMg1C5 T1 screen Rev	GTAGTCGCTGACACTGACA
SMg1C5 T2 screen Fwd2	TGGAACTTCAGCTACGTCGAG

5.5 Bioinformatics & computational applications

- Genome mining

NCBI database

The Basic Local Alignment Tool (BLAST) provided by the NCBI database was used to find homologous genes to those described to be involved in PA biosynthesis in the publication of Schimming et al.[79]. Based on the known DNA sequences of both the NRPS and the FAD-dependent monooxygenase, a comparison to sequence databases identified several genes encoding regions of similar sequences which could be involved in vuPA construction. Approximately ten bacterial strains with varying sequence identity percentage compared to the published reference genes were chosen and their whole genome sequences were used for further analyses[200].

AntiSMASH (Antibiotics & Secondary Metabolites Analysis Shell)

Whole genome sequences obtained from the NCBI database were used as input files in AntiSMASH versions 3.0 and 4.0[201-203] to identify, locate, annotate, and analyze secondary metabolite biosynthetic gene clusters within each of the selected bacterial genomes. Screening through all annotated BGCs, including those responsible for the synthesis of polyketides, non-ribosomal peptides, terpenes, aminoglycosides, aminocoumarins, indolocarbazoles, lantibiotics, bacteriocins, nucleosides, beta-lactams, butyrolactones, siderophores, melanins and so on, led to the identification of single BGCs encoding the genes of interest predicted to catalyze PA formation. In addition, the NRPS domain setup as well as a putative core structure assembly was predicted.

Gene homology studies

Online services such as the Multiple Sequence Alignment Viewer 1.18.0 (NCBI) and MAFFT (Multiple Alignment using Fast Fourier Transformation) tools (EMBL-EBI) as well as the visualizing tool Evolview v3 used for displaying phylogenetic relationships of bacterial strains encoding PA-producing BGCs were applied for studies on gene homology and evolutionary relations.

- Cloning

Geneious version R8.1.9 (Biomatters)[204] and SnapGene Viewer® 5.0.7 (snapgene.com) data analysis tools were used during the entire cloning process. By importing whole genomes or contigs from the NCBI and AntiSMASH databases, exact boundaries of BGCs or single genes

could be assigned. Furthermore, the software was applied to design oligonucleotides for PCR amplification of vector and insert DNA, to assemble plasmids *in silico*, to identify restriction enzyme cutting sites and to perform sequence alignments, e.g., after DNA sequencing.

NEB tools

Several NEB interactive tools were applied for cloning purposes. The NEB T_m calculator was used to determine the optimal oligonucleotide annealing temperature for PCR reactions depending on the polymerase and buffers present in the reaction setup. Primer design and *in silico* plasmid assembly was modeled with the NEBuilder assembly tool and the NEBcutter V2.0 was used to identify restriction sites within the cloned plasmid DNA. For the performance of analytical restriction digests for plasmid assembly validation (see section 5.5.3), the Double Digest Finder helped to find compatible restriction enzymes.

Primer design

Oligonucleotides with compatible annealing temperatures, similar lengths, and a low tendency to form hairpins or primer dimers (self- and hetero-) were designed using multiple DNA analysis tools such as NCBI Primer BLAST, the IDT OligoAnalyzer, and the Sigma-Aldrich OligoEvaluator™.

- Protein analysis

A variety of online protein analysis tools were used, including the ExPASY ProtParam (calculation of protein significant parameters such as molecular weight, theoretical pI, extinction coefficient, estimated half-life etc. based on a given amino acid sequence), ExPASY translate (DNA to amino acid sequence conversion) and the ProteinCalculator tool (sourceforge).

- Protein purification

For automated protein purification using the Äkta Pure chromatography system (GE Healthcare) the Unicorn 6 software (GE Healthcare) was used for method development and monitoring of the protein purification process. Furthermore, automated fraction collecting was controlled by the application.

- Imaging systems

DNA imaging

The image acquisition software GeneSnap (Syngene) was used to capture electrophoretic gel images of PCR reactions, plasmid assemblies, restriction digests and colony screening procedures.

Protein imaging

SDS gels were imaged and saved by scanning them with a commercial scanner device and processing them with the IrfanView picture editor.

- LC analysis

For system control, real time data acquisition of separation processes and data evaluation, either Chrompass chromatography (JASCO; used at TUM) or Clarity Chrom chromatography (Knauer; used at TUD) software was applied. All established running methods are listed in section 5.9.1.

- MS analysis

The Advion Data Express software (Advion) was used to import and process mass spectra obtained directly from the Advion expression CMS single quadrupole mass spectrometer. Coupled LC-MS measurements executed with the LTQ Fleet™ mass spectrometer (Thermo Scientific) were analyzed with the Xcalibur™ Qual Browser 2.2 SP1.48 software (Thermo Scientific).

- NMR analysis

All obtained NMR spectra were evaluated using the MestReNova (MestreLab Research S.L.) application.

5.6 Molecularbiological methods

5.6.1 Isolation of genomic DNA

The cultivated bacterial strains were centrifuged to remove the growth medium and the pellet was washed with 0.9 % NaCl. After removal of the washing solution the cell pellet was resuspended in 5-10 mL of lysis buffer by vortexing. The solution was then freeze-thawed three times, alternating between liquid nitrogen and a 55 °C water bath to assist in cell lysis. Lysozyme at a final concentration of 10 mg/mL and RNase A were added before incubating at 37 °C for 1 h with occasional inversion of the tube. This step was followed by the addition of Proteinase K (0.5 mg/mL final concentration) and SDS (2 % v/v final) and ½ h incubation at 37 °C, then ½ h at 55 °C to ensure protein denaturation and digestion as well as to complete cell lysis. Polysaccharides and cell debris were subsequently precipitated with NaCl (1.3 M) and CTAB solution (1 % v/v) at 65 °C for 10 minutes. Removal of any remaining proteins or cellular debris was conducted by incubation of the pellet solution mixed with an equal volume of chloroform isoamylalcohol (24:1) on ice for ½ h before centrifuging (10 min, 10000x g). The aqueous phase was next separated from the rest of the solution and transferred into a fresh tube where genomic DNA (gDNA) extraction was performed by extracting twice with phenol chloroform isoamylalcohol (25:24:1). The obtained aqueous phases were again collected in a fresh tube and the gDNA was recovered by addition of 0.6 volumes of absolute isopropanol, centrifugation (1 min, 17000x g) and careful liquid removal to yield a white pellet. After washing the gDNA pellet with absolute ethanol, the pellet was dried at 55 °C on a bench top shaker with an open lid, resuspended in TE buffer and stored at -20 °C until further use.

5.6.2 Polymerase Chain Reaction (PCR)

- Screening PCR

To confirm the isolation of the correct gDNA, screening PCR was performed using primers binding specifically to the target sequence. Furthermore, the successful transformation of plasmids into host bacteria was monitored using this method. Total reaction volumes of 25 µL consisted of: 2.5 µL Taq buffer, 1 µL DMSO, 0.25 µL dNTPs (10 mM), 0.25 µL forward primer (20 µM), 0.25 µL reverse primer (20 µM), 0.125 µL Taq polymerase, 15.625 µL double-distilled water (ddH_2O) and 5 µL of template (either gDNA or bacterial cell material). The general cycling protocol is given in table 15.

Table 15. Cycling conditions for screening PCR.

Step	Temperature	Time	Cycles
Initial denaturation	95 °C	3 min	
Denaturation	95 °C	20 s	
Hybridization	45 – 72 °C	20 s	30
Elongation	72 °C	60 s/kb	
Final elongation	72 °C	5 min	
Storage	12 °C	∞	

- PCR amplification of target DNA from gDNA, linear or plasmid DNA

Routine PCR amplification was performed using Q5® High-Fidelity DNA Polymerase (NEB), with a 25 µL reaction volume made up of: 5 µL 5x Q5 reaction buffer, 5 µL 5x Q5 high GC enhancer, 0.5 µL dNTPs (10 mM), 0.625 µL forward and reverse primers (20 µM each), 0.125 µL Q5 DNA polymerase, 20 – 50 ng of template DNA filled up to 25 µL with ddH_2O. The standard cycling procedure is given in table 16.

Table 16. Cycling conditions for standard Q5 polymerase PCR.

Step	Temperature	Time	Cycles
Initial denaturation	98 °C	3 min	
Denaturation	98 °C	30 s	
Hybridization	50 – 72 °C	30 s	34
Elongation	72 °C	30 s/kb	
Final elongation	72 °C	5 min	
Storage	12 °C	∞	

- Amplification of long, difficult and GC-rich targets

For some of the selected targets, DNA sequences were not only in the range of 13 – 30 kb, most also exhibited over 70 % in GC content. Here, a more powerful polymerase, the KOD Xtreme™ Hot Start DNA Polymerase (Merck) was necessary to amplify genes of interest. One PCR reaction contained: 25 µL 2x Xtreme™ buffer, 10 µL dNTPs (2 mM), 0.75 µL forward and reverse primers (20 µM each), 20 – 50 ng DNA template, 1 µL polymerase, all filled up with ddH_2O to give a total reaction volume of 50 µL per tube. Thermocycling conditions for this touchdown PCR procedure are shown in table 17.

Table 17. Cycling conditions for KOD polymerase PCR.

Step	Temperature	Time	Cycles
Initial denaturation	94 °C	2 min	1
Denaturation	98 °C	1 s	
Hybridization	74 °C	1 min/kb	5
Denaturation	98 °C	1 s	
Hybridization	72 °C	1 min/kb	5
Denaturation	98 °C	10 s	
Hybridization	70 °C	1 min/kb	5
Denaturation	98 °C	10 s	
Hybridization	68 °C	1 min/kb	15
Final elongation	68 °C	1 min/kb	1
Storage	12 °C	∞	

- Isolation and preparation of PCR amplified plasmid or linear DNA

Successfully amplified DNA of interest was isolated and purified using either the PCR Purification Kit (Jena Bioscience) or the Monarch® DNA Gel Extraction Kit (NEB) according to the manufacturer's protocol. The purification kit was applied when little or no side products occurred whereas the DNA of interest was excised using the gel extraction kit when the reaction product was impure.

5.6.3 Enzymatic DNA processing and modification

- Linearization and dephosphorylation of circular DNA

To insert target genes into suitable vector systems, vector backbones needed to be linearized using restriction enzymes. Single cutting enzymes obtained from NEB or Thermo Scientific were applied in a 50 µL preparative restriction digest consisting of 1 µg of circular template DNA, 2 µL of the selected restriction enzyme, 5 µL of the respective restriction digest buffer, and ddH_2O. The reaction was incubated at 37 °C for three to five hours, before heat-inactivating the enzyme if necessary.

To prevent recircularization of the linearized vector backbone ends during ligation, removal of 5' phosphate groups was necessary. After preparative restriction digest, linearized DNA was directly dephosphorylated by addition of the Antarctic Phosphatase (NEB; 2 µL) together with Antarctic Phosphatase buffer (6 µL) and ddH_2O (2 µL) to the reaction mixture. After overnight incubation at 37 °C, the enzyme was heat-inactivated (70 °C, 5 min) and linear DNA was purified either using the PCR purification or DNA gel extraction kit (see section 5.2).

- DpnI digest for template DNA removal

Especially when using plasmids as template DNA for PCR-mediated vector backbone linearization, removal of the template after PCR was essential to obtain a pure product. The endonuclease DpnI (NEB) was utilized to degrade methylated and hemi-methylated template DNA as described in the following: 1 µL DpnI alongside with 5 µL of CutSmart® buffer (NEB) per 50 µL PCR sample were mixed and incubated over night at 37 °C, before applying all to an agarose gel and performing gel extraction and purification with the DNA gel extraction kit.

- Enzymatic excision of gDNA

For LLHR and LCHR experiments, the DNA sequence of interest was excised from gDNA by enzymatic digestion. Restriction enzymes were chosen to cut as close as possible to the borders of the desired DNA piece to minimize its' size as much as possible. Isolated gDNA (50 µg at 200 ng/µL) was mixed with the appropriate restriction enzymes (320 U final amount; 20 U/ µL), RNAse (20 µg at 10 µg/µL), 10x digestion buffer (final concentration 1x) and ddH_2O. The total sample volume of 400 µL was then incubated for 4 h at 37 °C before extracting the DNA with an equal amount of phenol chloroform isoamylalcohol (25:24:1). Following thorough inversion and centrifugation (9400 x g, 10 min at RT) the aqueous phase was transferred to a fresh tube. Addition of 1/15 volume of 3 M sodium acetate (pH 7.5) as well as 2.5 volumes of ethanol with gentle mixing was used to precipitate (9400 x g, 2 min) the digested genomic DNA. After discarding the supernatant and removing all the liquid in the tube the DNA pellet

was washed with 1 mL of 70 % (v/v) ethanol (9400 x g, 1 min) before drying it in a bench top thermal shaker at 45 °C until all the liquid was evaporated. The gDNA was then resuspended in ddH_2O, concentration and purity were measured (see section 5.5.5) and the DNA was stored at -20 °C until further use.

- Identity confirmation of plasmid or vector DNA

Analytical restriction digest was performed to confirm plasmid DNA identity after isolation and purification. Enzymes from NEB or Thermo Scientific were chosen according to their specific cutting abilities (one or multiple times) on the target plasmid. A 10 µL reaction setup contained 240 ng of plasmid DNA, 2 µL of each of the restriction enzymes, 1 µL of a respective restriction digest buffer and ddH_20. Incubation took place at 37 °C for minimum one hour before the reaction was applied to an agarose gel.

5.6.4 Plasmid assembly

Both Sequence and Ligation Independent Cloning (SLIC) and Gibson Assembly (GA) were used to assemble at least two fragments to generate a plasmid. Both methods are sequence-independent, they commence with the same starting material and end with an identical final product, however they differ in their mechanism of connecting insert and vector backbone. PCR amplification of both insert DNA and linear vector DNA with subsequent PCR-cleanup or optional preparative restriction digest of circular vector backbone (see section 5.5.3) yielded the starting material for plasmid assembly. Primers for target insert amplification were designed to carry 20 – 30 bp homology arms at their 5’ sides whose sequence is identical to the linearized vector ends using the NEBuilder tool. Ratios between two- and five-fold excess of insert over vector DNA were tested for optimal plasmid assembly outcome.

- Seamless Ligation Independent Cloning (SLIC)

Single as well as multiple DNA fragments were inserted into a vector system of choice using the SLIC cloning method based on the 3’ exo activity of T4 DNA polymerase. The 10 µL sample consisting of linear insert DNA and linear vector backbone at a ratio of 2:1/3:1/4:1 or 5:1 with a minimal amount of 0.02 pmol, T4 DNA polymerase (0.5 µL; NEB) and NEBuffer 2.1 (1 µL) was kept at room temperature for 2 ½ min before incubating on ice for 10 min. Half of the reaction volume was subsequently transformed into competent cells, where repair of occurring single stranded nicks took place.

- Gibson/HiFi DNA Assembly (GA)

Gibson assembly combines 5’ exonuclease and ligase activity in one reaction, thus making plasmid assembly more efficient. NEBuilder® HiFi DNA Assembly Mix (10 µL) containing all necessary reaction components, was added to 10 µL of insert-vector DNA mixture (ratios 2:1 or 3:1) and incubated for 1 h at 50 °C before transforming 10 µL of the total reaction volume into competent *E. coli* cells. The optimal insert-vector ratio was determined using an in-house calculation tool.

- Ligation

As an alternative to GA and SLIC assembly, the standard ligation approach was employed to insert genes of interest into a suitable vector backbone. Both vector and fragment to be inserted were designed to possess corresponding restriction sites either at the insertion site (vector) or at both ends of the gene sequence (insert) so that treatment with the appropriate restriction enzyme resulted in complementary sticky ends. For the generation of compatible linear vector and insert DNA, reaction mixtures were set up following the protocol for linearization and dephosphorylation of circular DNA (section 5.5.3) with subsequent DNA purification via gel extraction (section 5.2).

For ligation, a total reaction volume of 20 µL, consisting of linearized vector and insert DNA, 1 µL of T4 DNA Ligase (Jena Bioscience), 1.5 µL 10x T4 DNA Ligase buffer and ddH_2O was prepared. A minimum amount of 0.02 pmol was used for both vector and insert, with a preferable insert:vector ratio of 3 to 5. After gentle mixing, the ligation reaction was incubated at 16 °C over night. Enzyme inactivation was achieved by 10 min incubation at 65 °C before transforming 10 µL into chemo- or electrocompetent cells (section 5.5.9).

- Homologous recombination (LLHR/LCHR)

Especially long and rather difficult gene cluster sections from a complex DNA source to be cloned were subject to either Linear plus Linear or Linear plus Circular homologous recombination (LLHR and LCHR). This way, not only large BGCs could be directly inserted into a vector but plasmid backbone exchange was also possible.

Alongside the BGC of interest which was excised from gDNA via enzymatic digest (see section 5.5.1), so-called capturing vectors needed to be assembled to enable homologous recombination to take place. These plasmids contain homology regions of about 500 bp flanking the BGC of interest at both the 3’ and the 5’ end which are separated by a specific restriction site. Linearization of the capturing vector allowed LLHR with the excised piece of gDNA to take place. For LCHR, the linear DNA fragment to be inserted contained 50 bp

homology arms on both ends with sequences identical to the parts directly adjacent to the insertion site of the circular vector or plasmid DNA. In addition, the template DNA and capturing plasmid must exhibit different antibiotic resistance cassettes for selection. All primers, capturing arms and vectors needed were designed using the Geneious and SnapGene bioinformatics tools (see section 5.5).

For LLHR to take place, an overnight culture of *E. coli* GB05dir cells was grown and 300 µL of the culture were inoculated into 15 mL of fresh, antibiotic-free LB medium the next day. Cell growth of the main culture at 37 °C and 200 rpm shaking was monitored by measuring the OD_{600} on a regular basis until a value of approximately 0.3 was reached (after about 1.5 h). Production of RecE/RecT proteins necessary for homologous recombination was induced with 1 M L-arabinose (300 µL) and further incubation at 37 °C up to an OD_{600} of 0.6 – 0.7 (after about 45 min). Next, 1.5 mL of cells was transferred to a precooled tube, centrifuged (9000 x g, 30 s, 4 °C) and the supernatant discarded. Washing of the cell pellet was performed three times using ice cold ddH_2O. Finally, the washed pellet was resuspended in 40 µL of ice cold ddH_2O and 500 ng of linearized vector/plasmid DNA as well as 5 µg of digested genomic DNA were added to GB05dir cells. Introduction of the DNA into the cells was accomplished via electroporation (see section 5.5.9). Transformed cells were incubated at 37 °C for 2 h while shaking before plating them on agar containing the appropriate antibiotics.

Plasmid backbone exchange was achieved using LCHR. Electrocompetent *E. coli* GB08red cells were prepared in advance according to the protocol given in section 5.5.9. The circular template plasmid was then transformed into GB08red cells, positive clones were validated via screening PCR and 5 mL precultures of the transformants were grown overnight in LB medium and the respective antibiotics at 37 °C upon shaking. For induction of recombinase production in GB08red cells (Redα/Redβ), 15 mL of fresh LB medium containing the antibiotic the plasmid carries a resistance for were inoculated with 150 µL of preculture and cultured to an OD_{600} of 0.3. Induction with 150 µL L-arabinose was followed by further cell growth until a value of 0.6 – 0.8 was reached. Next, cells were made electrocompetent according to the standard protocol (see section 5.5.9), before electroporating them with the linearized template DNA. After transformation and incubation, the *E. coli* suspension was plated on agar containing the antibiotic of the linearized template DNA.

5.6.5 DNA concentration and purity

The NanoPhotometer® P330 (Implen) was used to determine the concentration and quality of purified linear or plasmid DNA. For this, a blank measurement with ddH_2O was done before applying 1 µL of DNA sample to the NanoPhotometer® P-Class Submicroliter Cell. Default parameter values were as listed: a dilution factor of 1.000, the background correction on, multiplication factor 50 (for dsDNA) and a lid factor of 10 with a pathlength of 1 mm (unless concentrations are too high, then lid factor 50 with pathlength 0.2 mm). High A_{230} and A_{280} values indicate impure samples with a large amount of background signal, e.g., resulting from buffer salts and protein components whereas high A_{260} values are a sign of DNA purity. Since background correction was on for all measurements, the DNA concentration was calculated as in the following equations 3 - 5.

$$DNA\ concentration = (Abs_{260} - Abs_{320}) \times multiplicationfactor$$

$$Abs.ratio_1 = (Abs_{260} - Abs_{320}) \div (Abs_{280} - Abs_{320})$$

$$Abs.ratio_2 = (Abs_{260} - Abs_{320}) \div (Abs_{230} - Abs_{320})$$

Equations 3 – 5. Calculation of DNA concentration based on measured absorption values.

5.6.6 Sequencing of DNA products

Highly pure plasmid DNA was sequenced at GATC/Eurofins using overnight LightRun tube processing. Sample preparation included 450 – 500 ng of purified DNA combined with a specific sequencing primer (2.5 µM final concentration) to give 10 µL of probe. The obtained sequencing results were analyzed using the sequencing alignment tool of the Geneious software (see section 5.5).

5.6.7 Plasmid isolation and purification

Positive clones after successful transformation of plasmids were grown as 5 mL precultures overnight at 37 °C, 200 rpm shaking to obtain high amounts of plasmid. Plasmid isolation and purification was performed using the peqGOLD Plasmid Miniprep Kit I (VWR) according to the manufacturer's instructions. Purified plasmid DNA was eluted in 35 - 60 µL ddH_2O and the concentration was measured as mentioned above in section 5.5.5.

5.6.8 Agarose gels for DNA visualization

Confirmation of successful PCR amplification or plasmid assembly, visualization of an analytical restriction digest or screening for positive clones after transformation was possible by applying samples to agarose gels. In general, 1 % agarose gels were prepared out of 0.4 g / 0.8 g of agarose (Roth) and 40 mL / 80 mL of 1x TAE buffer (composition see section 5.3.2.1). The agarose was dissolved in TAE buffer (microwave) and immediately cooled under running water to omit any boiling retardation. DNA Stain Clear G (SERVA) was added to the molten agarose (1.7 µL / 3.4 µL) for nucleic acid detection before pouring it inside a gel casket. A comb containing the pockets for sample application was dipped into the liquid agarose and everything was left to solidify. Samples were prepared by adding 1/10 volume of 10x DNA loading dye. GeneRuler™ 1 kb Plus DNA Ladder (Thermo Fisher Scientific) was used as a marker substance. Separation of nucleic acids was achieved by running the gel in a Mini-Sub Cell GT Cell gel chamber (BioRad) connected to a standard Power Pack P25 low voltage supply (Biometra) for 25 – 30 min at 120 V. DNA visualization and analysis was made possible with the GeneGenius Bio Imaging System Gel Documentation Transilluminator (Syngene).

5.6.9 Transformation, conjugation and cultivation of *E. coli* and *Streptomyces*

- Preparation of chemocompetent *E. coli* cells

In advance, a freshly grown overnight culture of the respective cells was prepared in SOC medium without any antibiotics. A 250 mL main culture was inoculated with 2.5 – 5 mL of preculture and left shaking overnight at 18 °C and 200 rpm. After reaching an OD_{600} of 0.4 – 0.6, cells were cooled on ice (10 min), transferred to precooled centrifuge flasks and the medium was removed by centrifugation (10 min, 4000 rpm, 4 °C). The remaining pellet was resuspended in 80 mL of transformation buffer and stored for 10 min on ice. Centrifugation and medium removal steps were repeated, and the washed pellet was again resuspended in 18.6 mL transformation buffer supplemented with 1.4 mL DMSO. The solution was kept on ice for further 10 min, afterwards aliquots of 100 µL each were made, frozen in liquid nitrogen and chemocompetent cells were stored at -80 °C.

- Preparation of electrocompetent *E. coli* cells

Overnight cultures were prepared using LB medium and optionally antibiotics according to the cell type used (e.g., kanamycin and chloramphenicol for *E. coli* ET12567 pUZ8002). A main culture of 250 mL was inoculated with 2.5 mL of preculture and kept at 37 °C up to an OD_{600}

between 0.4 and 0.7. Cells were pelleted (5 min, 4000 rpm, 4 °C) and washed three times with 10 % (v/v) glycerol (40 / 20 / 10 mL) before resuspending them in 500 µL of 10 % (v/v) glycerol. Aliquots of 50 µL were frozen in liquid nitrogen and electrocompetent cells stored at -80 °C until further use.

- Heat shock transformation into chemocompetent cells

Transfer of SLIC/GA assembled plasmid DNA was achieved via heat shock. Initially, an aliquot of chemocompetent cells was thawed on ice before adding 150 – 200 ng of pure plasmid DNA or 5 – 10 µL of assembly mix. After a 30 min incubating period on ice, cells were heat shocked at 42 °C for 1 ½ in a bench top thermal shaker to introduce plasmid DNA. 1 mL of SOC medium was added immediately after heat shock and the transformed cells were subsequently incubated at 37 °C while shaking for minimum 1 h. Cells were then centrifuged (2 min, 6000 x g), the majority of the supernatant discarded, the pellet resuspended in remaining medium and plated on agar containing the appropriate antibiotics. Successfully transformed *E. coli* grown overnight at 37 °C were validated via screening PCR.

- Electroporation into electrocompetent cells

Assembled plasmids as well as LLHR and LCHR assembly reactions (see section 5.5.4) were introduced into *E. coli* via electroporation. For already assembled plasmids, 150 – 200 ng of DNA was added to an aliquot of previously thawed electrocompetent cells. The mixture was transferred to an electroporation cuvette and put into the MicroPulser™ Electroporation Apparatus (Bio-Rad). The program Ec2 was used to apply an electric pulse and successful electroporation was indicated by pulse values of 4.5 – 5.5 ms. SOC medium was added immediately after the pulse, cells were gently resuspended, and the mixture was transferred to a fresh tube. After 1 h shaking at 37 °C, the medium was removed, the pellet resuspended in 200 µL of supernatant and plated onto antibiotic-containing agar. Growth of transformed *E. coli* cells was monitored after overnight incubation at 37 °C via screening PCR.

- Cultivation of *E. coli* cells

In general, precultures were grown in LB or SOC medium containing the respective antibiotics at a dilution of 1:1000 (e.g., 5 mL preculture plus 5 µL antibiotic stock solution). Cultures were prepared in test tubes or Erlenmeyer flasks and incubated at 37 °C and 200 rpm shaking (Infors) unless designated otherwise.

Main cultures were cultivated in a selection of different media, predominantly LB, TB or 2xYT medium, according to the purpose of cell cultivation. LB and TB medium were mostly used for protein overexpression and production whereas TB and MM medium found application when

expressing gene clusters for secondary metabolite expression. For protein production, cultivation temperatures ranged from 16 – 20 °C with overnight shaking (200 rpm). Natural product expression was either conducted for 2 days at 28 °C or 5 - 7 days at 20 °C.

- Conjugation and cultivation of *Streptomyces*

Bacterial conjugation from *E. coli* to Actinomycetes was used to introduce plasmid DNA for secondary metabolite expression. Electrocompetent *E. coli* ET12567 pUZ8002 cells were transformed with the plasmid to be expressed in *Streptomyces* strains, plated on LB agar containing kanamycin (50 µg/µL), apramycin (30 µg/µL) and chloramphenicol (25 µg/µL), grown overnight at 37 °C and screened for positive clones. Precultures of validated colonies harboring the plasmid of interest were cultivated in LB medium plus apramycin and 400 – 1000 µL were used to inoculate 20 mL main cultures containing all three antibiotics. Cells were grown to an OD_{600} of 0.4 – 0.6 and subsequently washed by repeated centrifugation (5 min, 4000 rpm, 4 °C), discarding of the supernatant and resuspension of the cell pellet in fresh LB medium. After washing, ET12567 pUZ8002 cells carrying the plasmid of interest were resuspended in 500 µL LB without antibiotics. Aliquots of *Streptomyces* spores (10 µL each) were prepared for conjugation by mixing them with 500 µL 2xYT medium followed by a 10 min activation step at 50 °C and cooling on ice. Ready-to-use *E. coli* cells were incubated together with the active spores before centrifuging both together for 2 min at 4000 rpm. After discarding the supernatant, cells were resuspended and plated on MS agar supplemented with 10 mM $MgCl_2$ and 60 mM $CaCl_2$. Plates were incubated for 16 – 30 h at 30 °C and covered with 1 mL sterile ddH_2O containing 1 mg nalidixic acid and 1.25 mg apramycin. After drying, plated cells were left to grow for 7 days at 30 °C. Streptomyces clones were then transferred to fresh MS agar plates containing apramycin (30 µg/µL) and nalidixic acid (25 µg/µL) for renewed growth. Simultaneously, precultures in CASO medium containing the respective antibiotics were prepared for liquid culture cultivation (see section 5.5.9).

- *Streptomyces* spore preparation

Cryo cultures of *Streptomyces* host strains were thawed and 50 µL containing at least one colony were resuspended in 200 µL of ddH_2O. The mixture was plated on MS agar and incubated at 30 °C until spores formed. Spores were removed with a cotton swab, transferred into cryo tubes containing 20 % glycerol and stored at -80 °C.

- Long-term *Streptomyces* cultivation

In general, optimal Actinomycete growth is guaranteed at 28 – 30 °C. Precultures were prepared in CASO medium containing apramycin (30 µg/µL) and nalidixic acid (25 µg/µL) and

left to shake for 2 – 3 days until spherical colonies became visible. Colonies were then transferred to Erlenmeyer flasks filled with ISP4 medium for 5 – 7-day incubation at 28 °C upon shaking at 180 rpm.

After conjugation and plating on MS agar with 10 mM $MgCl_2$ and 60 mM $CaCl_2$, single clones could be picked after about 2 weeks. Colonies were picked with a toothpick, resuspended in 50 µL ddH_2O and 25 µL were streaked onto a fresh MS agar plus apramycin plus nalidixic acid plate whereas the other 25 µL were used to inoculate 5 mL precultures for liquid medium expression.

- Long-term storage of *E. coli* and *Streptomyces*

Both *E. coli* and *Streptomyces* strains containing plasmids of interest were kept as cryo stocks for long-term storage at -80 °C. Cryo stocks consisted of 700 µL bacterial preculture mixed with 300 µL of a glycerol-containing buffer (see section 5.3.2.1).

5.7 Heterologous expression of biosynthetic pathways

5.7.1 Production and purification of proteins in *E. coli*

- Protein expression

Standard host strains used for protein overproduction included *E. coli* BL21DE3 and BAP1, *E. coli* SoluBL21 or Arctic Express cells if proteins turned out to be insoluble. 5 mL Overnight cultures of successfully transformed clones were used at a 1:100 dilution to inoculate main cultures supplemented with the appropriate antibiotics.

Initial small-scale test expressions were conducted to determine optimal growth and expression conditions for each protein to be studied. Test cultures comprised 50 mL of medium and growth was influenced by the variation of multiple parameters, including medium type (LB or TB), cultivation temperature and amount of inductor added for initiation of protein expression. Protein solubility was tested by adding a variety of tags fused to the C- or N-terminal ends of the target protein (His_6, GST, MBP tags). Once the optimal expressions were found, protein overproduction was scaled up. The expression protocol was identical for both small- and large-scale heterologous expression.

After inoculation of main cultures in Erlenmeyer flasks, each containing 1 L of LB or TB medium supplemented with the respective antibiotics (1:1000 dilution), cultures were grown to an OD_{600} of 0.4 – 0.8 (LB) or 0.6 – 1.0 (TB) at 37 °C while shaking. As soon as the OD was reached,

cultures were put on ice for 30 min before inducing protein expression by addition of an inductor such as IPTG or L-arabinose (0.05 – 0.1 mM). During expression, the temperature was reduced to 16 – 20 °C to prevent too much protein production and in consequence protein aggregation as well as to enable correct protein folding. Cells were harvested by transferring the culture broth to centrifugal flasks and centrifuging (10 min, 6000 x g) to remove any medium. The remaining cell pellet was transferred into a falcon tube and washed with 0.9 % NaCl solution. After removal of any supernatant, the cell pellet was stored at -80 °C until further use.

- Cell lysis via sonification

The thawed cell pellet was resuspended in lysis buffer (4 mL/g pellet) by vigorous vortexing. Once a homogenous solution was obtained, cell disruption was achieved with the SONOPULS ultrasonicator. Sonication was performed on ice using repeated 10 s pulses followed by 10 s breaks with 30 – 50 % pulse intensity. In the following, cell debris was removed by centrifugation for 30 min at 13 000 x g and 4 °C and the protein-containing supernatant was transferred to a fresh tube for purification. For evaluation and monitoring of protein expression, 20 µL of supernatant as well as 20 µL of pellet resuspended in 100 µL ddH_2O were taken aside for SDS-PAGE analysis (see section 5.6.1).

- Manual protein purification of His_6-tagged proteins

After sonification and removal of cell debris by centrifugation, the supernatant was transferred to a fresh tube, treated with nickel NTA resin (1 mL/ 1 L expression culture; Jena Bioscience or CubeBiotech) and shaken on ice for 1 to 1 ½ h to allow protein binding via its His_6-tag to the resin. The lysate was then poured into a pre-equilibrated Ni column (GE Healthcare Life Sciences) to catch the resin and let the liquid pass through. Cellular proteins as well as remaining cell debris were removed in the following washing steps by application of washing buffers containing increasing amounts of salts (20, 40, and 50-80 mM imidazole) before eluting the protein of interest with 250 mM imidazole-containing buffer. 20 µL of flow through, washing steps and the final elution were removed for SDS-PAGE analysis.

- Äkta Pure automated protein purification

As an alternative to manually packed Ni-NTA columns, a prepacked HisTrap HP 5 mL column (GE Healthcare) was used together with an automated Äkta Pure chromatography system (GE Healthcare) to obtain proteins as pure as possible. The cell lysate supernatant was injected onto the column via a super loop and washing as well as elution of protein was achieved by running an imidazole gradient over a designated amount of time. Buffers A and B were mixed by fully automated valves to enable a gradual increase of imidazole from 0 to 250 mM in the

buffer solution running through the column. Samples of the collected protein fractions were taken for SDS-PAGE analysis.

The same chromatography system and purification principle was used for recombinant proteins bearing an MBP fusion tag where an MBP Trap HP (5 mL; GE Healthcare) and Buffers C and D were used.

- Enhancement of protein solubility using N-lauroylsarcosine

Proteins remaining in the cell pellet fraction after cell disruption and centrifugation were solubilized by treatment with N-lauroylsarcosine. The cell pellet was resuspended in Peternel buffer (section 5.3.2.2) and incubated on ice with shaking for 2 – 4 h. Centrifugation for 30 min at 13000 rpm was performed to separate protein-containing supernatant and cell debris. The supernatant fraction was subsequently treated with NiNTA resin and protein purification proceeded following the standard protein purification protocol (see previous bullet points).

- Protein desalting using PD-10 columns

Exchange of the elution buffer with an assay or storage buffer and removal of the high imidazole concentration was achieved using a PD-10 desalting column prepacked with Sephadex G-25 resin (GE Healthcare). After equilibrating the resin with the exchange buffer, the eluted protein was applied step-by-step (2.5 mL at a time) to the column. Larger volumes were applied with intermediate washing steps to retain full Sephadex binding capacity. The eluted desalted protein was subsequently concentrated.

- Protein concentration via UV-Vis

Desalted protein was transferred to Vivaspin® columns with an appropriate volume capacity and molecular weight cutoff (MWCO). For monooxygenases, a MWCO of 10 000 Da was used, NRPS systems were concentrated using columns with a MWCO of 30 000 Da or 100 000 Da. For removal of excess buffer liquid, spin columns were centrifuged at 4 °C and 6500 rpm until the desired end volume of protein sample was reached (0.2 – 1.0 mL). To omit aggregation of protein, the sample was carefully mixed in between centrifugation runs. The final protein sample was transferred to a fresh 1.5 mL tube and the concentration was measured. A 20 µL sample was put aside for SDS-PAGE analysis.

Protein concentrations were measured using the NanoPhotometer® P330 (Implen) at 280 nm wavelength and calculated using the following equation:

$$c_{Protein} = (Abs_{280} - Abs_{320}) \times A_{280}factor \times lid\ factor \times dilution\ factor$$

Equation 6. Determination of protein concentration.

Molar extinction coefficients and molecular weights of studied proteins are given in table 18.

Table 18. Protein data necessary for concentration determination.

Protein	# Amino acids	Theoretical pI	Molecular weight [Da]	Molar extinction coefficient ε_{280nm} [$M^{-1}cm^{-1}$] S-S bonds
PxaA NRPS	2407	5.62	270673.38	261140
PxaB MOX	392	5.18	43018.91	48150
PTC12 NRPS	2402	6.12	269233.53	282640
PTC12 MOX	387	8.73	41973.12	12263
MimoC21 NRPS	2366	5.72	260321.93	254580
MimoC21 MOX	388	8.00	41677.28	53525
SMg1C5 NRPS	7302	6.03	259932.12	235140
SMg1C5 MOX	846	9.45	92735.46	92735
SclC22 MOX	397	7.79	42337.94	53190
SclC39 MOX	399	6.82	43097.82	54430
PAO1C29 NRPS	2352	5.46	259406.41	267640
PAO1C29 MOX	388	8.37	42498.32	57660
Pys NRPS	1298	5.91	142016.11	86990
Pys NRPS PCP	70	4.80	76665.78	125
BraC MOX	395	7.73	42714.71	64775
FadD Ligase	561	6.22	62331.99	53540
Sfp PPTase	224	5.51	26167.69	29130

- Protein concentration via Bradford method

Molar extinction coefficients could not be calculated for some proteins, e.g., the standalone PCP domain, therefore the Bradford method using Roti®-Nanoquant solution (Roth) was applied for protein quantification. A standard dilution series was prepared according to the manufacturer's manual with 0, 2.5, 10, 50 and 100 µg/mL bovine serum albumin to establish a calibration curve, using ddH_2O as a blank value. 200 µL dilutions of purified protein were prepared with ddH_2O which were then mixed with 800 µL of 1x Roti®-Nanoquant working solution. Absorption at wavelengths of 590 nm and 450 nm was measured with the NanoPhotometer® P330 (Implen) for all calibration samples and the quotient A_{590}/A_{450} was plotted against the amount of protein to obtain a calibration curve. Identical measurements

were repeated for all protein dilutions and upon determining the quotient and regarding the dilution factor, the protein concentration could be directly extracted from the calibration curve.

- SDS-PAGE for confirmation of protein purification

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed to monitor the successful expression of protein as well as the process of purification and to show the final state of pure protein.

A reducing agent like DTT or β-ME was used to break up secondary and tertiary structures by reducing disulfide bonds whereas SDS was added to denature proteins by cleaving hydrogen bonds, thereby giving them a linear structure and masking protein charge. This way, proteins were separated only based on their molecular weight. Both components were part of the 5x SDS loading buffer, which was added in a 1:5 ratio to the collected and prediluted 20 µL samples of lysis, washing, desalting, and concentrating steps. To make samples comparable to each other, the same number of cells needed to be present in each sample. This was achieved by calculating the necessary volume of cultured *E. coli* cells using equation 7. Boiling for 5 min at 95 °C initiated denaturing of prepared samples. After cooling at room temperature, samples were loaded onto the gel.

$$V = \frac{1}{(2 \times OD_{600})}$$

Equation 7. Determination of sample volume needed as input for SDS probe.

Discontinuous gels consisted of a separating gel at the bottom and a stacking gel at the top. According to the protein size, separating gels of acrylamide/N, N'-methylene bisacrylamide content ranging from 8 – 15 % were prepared, with the acrylamide percentage increasing the smaller the proteins got (see section 5.3.2.2). In addition, gels components included 4x SDS stacking or separating buffer and ddH_2O. Polymerization was induced by the addition of the radical starter APS in combination with the catalyst TEMED. The MiniPROTEAN® Electrophoresis System (BioRad) was used to make gels and perform electrophoresis. Gel material was poured into molds, starting with the separating gel at the bottom. The gel was covered with isopropanol until the acrylamide/N, N'-methylene bisacrylamide solidified to avoid air bubble formation at the interface to the separating gel which was poured afterwards. Combs containing pockets for sample application were inserted into the liquid stacking gel material and left to harden. The ready-to-use gel was then inserted into the electrophoresis chamber,

the chamber was filled with 1x SDS running buffer, the combs were removed and 2- 8 µL of prepared samples loaded onto the gel. As a standard, marker solutions of different size composition were applied for protein size determination (2 µL marker). Electrophoresis was conducted at 200 V, 35 mA/gel, 55 W for 1 h 5 min or until the samples completely ran through the gel. Proteins were then visualized by staining of the gel with CoomassieBrilliantBlue solution followed by destaining with a solution containing acetic acid. For equal distribution of staining and destaining solutions, the gel-bearing tray was put on a shaker and rotated slowly.

- Native PAGE for protein purification and preservation of enzyme activity

Protein purification via native, non-denaturing PAGE was performed to increase protein purity and retain proteins in their natural active configuration. Gels were prepared in the same manner as described in section 5.3.2.2 for SDS-PAGE gels but using native separating and stacking buffers without any reducing agents. Protein samples were cooled throughout the entire preparation time and sample loading buffer was added to a 1x final concentration immediately before sample loading onto the gel. Electrophoresis was performed for 2 h (35 mA/gel, 180 V, 50 W) with constant cooling of the electrophoresis chamber to 4 °C. Subsequent staining with CoomassieBrilliantBlue and destaining with acetic acid was done by gentle shaking on a bench top shaker at 4 °C.

Recovery of the protein of interest from the polyacrylamide gel was performed as follows. Gel slices containing the protein were excised using a scalpel, cut into smaller pieces, and transferred to a 2 mL Eppendorf tube. Protein was eluted from the gel slices by adding 500 µL of protein elution buffer, vortexing, incubating for 10 min on ice followed by vortexing and removal of the supernatant. The elution step was repeated once more to obtain as much protein as possible. Eluted protein was further concentrated in a Vivaspin® 500 centrifugal concentrator (MWCO 100 kDa; Sartorius) before applying to test reactions.

- MS-MS analysis for protein identity confirmation

Tandem mass spectrometry was used to confirm the identity of heterologously expressed proteins in cases where end stage protein purity was still low or cell-own proteins produced by the host organism showed similar molecular weights. Sample preparation was dissected into three steps: protein preparation, protein desalting and filtering for MS-MS. Optimal protein concentrations for tandem MS analysis range from 1 – 10 mg/mL and the according volumes were pipetted into Lobind microcentrifuge tubes (Eppendorf). Protein denaturation was initiated by adding 7 M urea and 2 M thiourea and vortexing the sample until all solid components were completely dissolved. The sample was then reduced with 10 mM DTT at 37 °C for 1 h in a

tabletop shaker (500 rpm). Alkylation was promoted by the addition of 30 mM iodine acetamide and 30 min incubation at 25 °C, 500 rpm shaking in the dark. After quenching the alkylation reaction with DTT (30 mM; 30 min, 25 °C, 500 rpm) the sample was diluted 1:4 with 50 mM TEAB. Overnight protein digestion at 37 °C without shaking with trypsin (1:40 enzyme/protein mass ratio) was quenched by lowering the pH to 3 with 1 % formic acid (FA) and vigorous vortexing. For desalting of the sample, a C18 filter membrane was prepared by washing with once with MeOH and 80 % ACN containing 0.5 % FA followed by three wash runs using 0.5 % FA (each round of washing 30 sec at 1000x g). The trypsin digested sample was then loaded onto the equilibrated C18 filter membrane, washed three times with 0.5 % FA and eluted into a fresh Lobind tube in 60 µL 80 % ACN containing 0.5 % FA. The eluting solvent was subsequently evaporated in the SpeedVac concentrator 5301 (Eppendorf) and the sample stored at -80 °C until the day of analysis. Prior to MS-MS analysis, the digested and desalted protein sample was filtered. Tandem mass spin columns (Sartorius) were washed with 1 % FA (16 000x g, 2 min) before applying the protein sample also dissolved in 1 % FA (vortexing, 15 min sonification, vortexing). The sample was then filtered into a fresh Lobind tube by centrifugation (16 000x g, 2 min) and consequently transferred to a certified MS vial for analysis using the Orbitrap Fusion or Orbitrap XL (Thermo Fisher).

5.8 Secondary metabolite expression and isolation

5.8.1 Expression of natural products

- Expression in *E. coli* and isolation with Amberlites XAD-16

Precultures of *E. coli* BAP1 cells harboring arabinose inducible pCX2 constructs were incubated overnight at 37 °C and 200 rpm shaking in LB medium supplemented with kanamycin (50 µg/µL). Grown cells were then used to inoculate main cultures at a 1:100 dilution in TB medium containing kanamycin (1:1000). Cultures were left shaking at 28 °C and 180 rpm for 19 h before gene cluster expression was induced with 0.1 mM L-arabinose. In addition, the basic building blocks L-serine, L-proline as well as a fatty acid of choice for example lauric acid needed for PA biosynthesis were fed at 100 mg/L. After 10 – 12 h incubation (28 °C, 180 rpm), 1 % (w/v) of washed Amberlites XAD-16 (Alpha Aeser) was added to the expression cultures for metabolite binding. Incubation was continued for further 12 h before the culture broth was centrifuged (6000 x g, 10 min) to remove the medium. Cell pellet together with the polymeric adsorption material were subject to extraction. Secondary metabolites were released from Amberlites XAD-16 by extensive washing with MeOH for 2 h,

including MeOH exchange in between. All MeOH phases were pooled, and the solvent was evaporated using a rotavapor apparatus (Heidolph). The remaining substance was then dissolved in a 1:1 solvent mixture of H_2O and EtOAc and extracted according to the standard procedure, however leaving out the acidification step (see section 5.7.2).

- Expression in *E. coli* using GFP

Successful gene transcription and protein production for substrate expression was tracked for gene clusters inserted into the pET28pTetGFPv2 expression vector via the production of the fluorescent GFP protein at the 3' end of the cluster. Precultures of *E. coli* BAP1 variants were used at a 1:100 dilution to inoculate LB/TB/M9 medium main cultures containing kanamycin. After incubation for 2 – 4 h at 37 °C with 180 rpm shaking and reaching an OD_{600} of 0.6/0.8/1.0, cultures were cooled on ice for 30 min before gene cluster expression was induced with 0.5 µg/mL tetracyclin (10 mg/mL stocks). Cultures were incubated in the dark at 20 °C, 180 rpm from then on for 3 – 7 days. At the end of the expression duration, cells were harvested by centrifugation (6000x g, 10 min), the resulting supernatant and pellet fractions were separated for individual extraction according to the standard protocol (see section 5.7.2).

- Expression in *Streptomyces* in liquid medium

Spherical colonies from 2 – 3-day old precultures were used to inoculate ISP4 medium containing apramycin (30 µg/µL) and nalidixic acid (25 µg/µL) in 1:1000 dilutions. All gene clusters to be expressed were cloned into the pSET152ermErev vector and thus under the control of the constitutive ermE promoter, therefore no inducer was of need. Erlenmeyer flasks filled with 50 mL of culture broth were left to shake at 180 rpm and 28 °C for 5 – 7 days before centrifuging (4500 x g, 15 min) to separate pellet and supernatant and continue with standard extraction (see section 5.7.2).

- Expression in *Streptomyces* on solid medium

Streptomyces clones were plated onto GYM and ISP4 agar containing respective antibiotics and left to grow at 30 °C for either 5 – 10 or 14 days. The agar covered with grown streptomycetes was then cut into small pieces and transferred to Erlenmeyer flasks filled with different combinations and ratios of solvents. Extraction was performed using mixtures of EtOAc:MeOH:acetic (80:15:5), EtOAc:MeOH (80:20) and pure MeOH to test optimal conditions. The solvent-agar mix was put into an ultrasonication bath for 10 min before removing the solvent via filtration and evaporation. Extraction was performed two more times using the same solvent ratios. After evaporating the extracting solvents, the crude extract remained in a water remnant which was removed by lyophilization.

5.8.2 Extraction of secondary metabolites

In general, both *E. coli* and *Streptomyces* derived natural product expression cultures were extracted following a standard protocol unless designated otherwise. Following the separation of supernatant from pellet after centrifugation, the resulting fractions were treated differently. The cell pellet was resuspended via vortexing in a 1:1 mixture of acetone and MeOH, cells were disrupted in an ultrasonication bath (30 min) and subsequent centrifugation (4500 x g, 20 min) led to the separation of cell debris from a solvent mixture containing the secondary metabolites of interest. While cell debris was discarded, the solvent was evaporated to yield the crude pellet fraction extract. The culture broth supernatant fraction was acidified with 1 M HCl to a pH = 4 – 5 prior to extraction. An equal amount of EtOAc was added, the organic phase separated after initial extraction and the watery phase extracted with EtOAc for further three times. Afterwards, collected organic phases were combined, dried with saturated NaCl solution as well as dry $MgSO_4$ and filtered. EtOAc was evaporated to obtain the crude supernatant fraction extract. All obtained crude extracts, products of heterologous BGC expression, were prepared analogously for LC-MS analysis. Approximately 0.1 mg of both pellet and supernatant fractions was dissolved in 100 µL of MeOH and filtered via a Fisherbrand™ syringe filter 0.2 µm (Fisher Scientific) into an inlet-containing MS vial. Samples were then stored at 4 °C for minimum 30 min to remove any remaining insoluble components. Volumes of 15 – 25 µL supernatant sample and 20 – 45 µL pellet fraction were injected per measurement.

Lyophilization

Residual H_2O which can destroy extracted products or acids that would be concentrated upon evaporation of H_2O and lead to product destruction were removed via lyophilization. Round bottom flasks filled to maximum 1/3 of the total volume with crude extract plus residual H_2O were connected to the lyophilization apparatus for minimum 24 h to yield dry and powdery substances.

5.9 *In vitro* enzyme activity assays

5.9.1 CoA-labeling for PCP-domain loadings assays

To ensure that NRPS proteins were successfully activated either *in vivo* (in *E. coli* BAP1 cells) or *in vitro* by the PPTase Sfp, the ppan carrier CoA was labeled with the fluorescent dye tetramethylrhodamine-5-(and-6)*C*2-maleimide (TMR-Mal) at the thioester. The overall protocol was adapted from Yin et al. (Q98) on site-specific protein labeling catalyzed by Sfp. A TMR-Mal-CoA conjugate was synthesized combining 1 mg of TMR-Mal (dissolved in 0.25 mL DMSO) with 1.6 mg of coenzyme A trilithium salt in 0.75 mL of sodium phosphate buffer (100 mM, pH 7.0) and stirred at room temperature for 1 h in the dark. Afterwards the product was analyzed with analytical HPLC and purified using preparative HPLC with a 30 min gradient run (see section 5.9.1). Solvent was evaporated from the purified TMR-Mal-CoA fractions and the compound was lyophilized to give a powdery pink colored product.

The standalone Pys PCP domain was cloned into the pET28bSUMO vector and thus exhibited a His_6-tag for protein purification. Protein concentration was determined with the Bradford assay (see section 5.6.1). The Sfp-TMR-Mal-CoA loading assay was set up in a total reaction volume of 500 µL Tris buffer (50 mM, 10 mM $MgCl_2$, pH 7.4), including the key components TMR-Mal-CoA (150 µM), Sfp R4-4 (0.5 µM) and the purified apo PCP domain (100 µM). For ppan transfer to take place, the assay was incubated for 1 h at 30 °C (no shaking). The reaction volume was then directly loaded onto prepacked Ni-NTA spin columns (Qiagen), washed with protein purification buffers 1 and 3 and eluted with elution buffer (see section 5.3.2.2) to obtain the optically visible holo PCP domain carrying the TMR-Mal labeled ppan moiety.

5.9.2 NRPS activation via Sfp

Ab initio phosphopantetheinylation of PCP domains was performed to ensure an active NRPS enzyme for following substrate conversion assays. A reaction mixture in HEPES buffer (75 mM, pH 7.0) comprising 10 mM $MgCl_2$, 5 mM DTT, 10 - 20 µM of *apo* NRPS protein as well as 11 - 24 µM coenzyme A was prepared and the priming reaction was initiated by the addition of 0.1 µM of Sfp R4-4 and incubation at 37 °C for 30 min. The reaction mixture was then used as a basis for substrate conversion assays without needing to purify the *holo* NRPS before.

- Multistep total synthesis of PA derivatives

PA intermediate production using purified NRPS systems

The NRPS' mode of action was analyzed by setting up an assay where PA basic building blocks were to be incorporated and linked together to form a PA intermediate structure. Activated fatty acid thioesters were synthesized in advance (see section 5.10). Purified NRPS protein PCP domains were activated with Sfp R4-4 in *vivo* (expression in *E. coli* BAP1 cells) or *in vitro* beforehand. Co-factors including NADPH (10 mM), FAD (2 mM) and coenzyme A (2 mM) as well as $MgCl_2$ (7.5 mM), ATP (10 mM) and Sfp (5 µM) were mixed with 10 µM NRPS in either HEPES (100 mM, 1.5 % glycerol, pH 7.8) or Tris-buffer (50 mM, 10 % glycerol, pH 7.5). Amino acids L-serine and L-proline and the activated fatty acid (dissolved in DMSO, DMF or acetone) were supplemented at 2 - 10 mM concentration. Assays of 100 µL total volume were put into a bench top shaker (450 rpm) overnight at 28 - 30 °C. The reaction was quenched by acidification with 1 M HCl followed by EtOAc extraction, solvent removal using the SpeedVac concentrator 5301 (Eppendorf) and LC-MS preparation by dissolving the product in 100 µL of MeOH and transferring it to a glass MS vial.

5.9.3 Substrate-driven PA derivative production via monooxygenases

MOX homologues were tested in their activity in an assay proving the successful conversion of pure PA intermediates into their corresponding PA product molecules. Purified MOX proteins were used at a concentration of 20 µM together with the cofactors NADPH (100 µM) and FAD (50 µM) to convert 200 µM of PA intermediate (in DMSO) to the final product, using Tris as a buffering system (50 mM Tris, 10 % glycerol, pH 7.5). After overnight incubation in a tabletop shaker (28 °C, 450 rpm) acidification caused precipitation of the protein and quenching of the reaction. EtOAc extraction and receival of a dry product was followed by LC-MS analysis preparation, including dissolving in MeOH and subsequent transfer into an MS-vial. 50 µL of sample were used per measurement.

5.9.4 One pot total synthesis of PA derivatives

Purified proteins

Pyrrolizixenamide production in a one pot reaction was tested e.g., using the NRPS and MOX enzymes of *Xhenorhabdus stockiae* DSM17904. The total reaction volume of 150 µL contained the following components: 2 mM L-serine and L-proline each, 2 mM fatty acid thioester (SNAcylated), 2mM coenzyme A and FAD, 7.5 mM $MgCl_2$, 10 mM of ATP as well as 10 mM freshly prepared NADPH, all in Tris buffer (50 mM, 10 % glycerol, pH 7.5). Catalyzing enzymes were added at 5 µM (NRPS) and 10 µM (MOX) alongside the NRPS-activating Sfp R4-4 (5 µM). The test tube was shaken (350 rpm) overnight at 28 °C before quenching the reaction by acidification with 1 M HCl. For product isolation, an equal volume of EtOAc was pipetted to the assay and vigorous vortexing followed by centrifugation (16000 x g, 2 min) allowed the separation of proteins from the supernatant. The organic phase was transferred to a clean tube and the EtOAc solvent was removed by evaporation using a SpeedVac concentrator 5301 (Eppendorf). For LC-MS analysis, dried sample was resuspended in 100 µL MeOH, transferred to an LC-MS vial and 50 µL were injected for the measurement with the “enzyme assay” run (see section 5.9.1).

Cell lysates

For *in vitro* production of pyrrolizixenamides or brabantamides, lysate assays were conducted by combination of NRPS and MOX enzymes from different bacterial strains. Separate protein expression cultures were prepared, induced with 0.1 mM IPTG or L-arabinose and cultivated overnight at 16 °C, 180 rpm shaking. Cells were harvested (6000 x g, 10 min), lysed according to the standard protocol and the supernatants were transferred to fresh falcon tubes. In a first approach NRPS and MOX lysates were combined and left to shake on ice for 1 and 4 h before basic building blocks (L-serine, L-proline and activated fatty acid) were added and incubation extended for further 2 – 4 h at 30 °C and 180 rpm. Lysates were kept separate and combined only directly for the assay in a second approach.

The substrate conversion assay using cell lysates was performed in HEPES buffer (100 mM, 1.5 % glycerol, pH 7.8) and consisted of 2 mM FAD, 2 mM coenzyme A, 7.5 mM $MgCl_2$, 10 mM ATP, freshly prepared NADPH (10 mM) as well as the PPTase Sfp (5 µM). The basic building blocks L-serine, L-proline and the fatty acid thioester were supplemented once more at 2 mM concentrations each. The 200 µL enzyme assays were incubated overnight at 28 °C on a tabletop shaker (500 rpm). Addition of 1 M HCl and vortexing followed by EtOAc extraction was done to isolate compounds. The SpeedVac concentrator 5301 (Eppendorf) was used to

dry the products and remove residual solvent. Samples were then prepared for LC-MS analysis by dissolving them in 100 µL of MeOH.

5.9.5 Single step total synthesis of pyreudione derivatives (38)

Enzyme activity of the NRPS system Pys from *Pseudomonas fluorescens* HKI0770 was tested with a substrate conversion assay to yield pyreudione natural product derivatives. The purified Pys protein was added at 10 µM to a reaction mixture in HEPES buffer (10 mM, 1.5 % glycerol, pH 7.8) containing 2 mM L-proline, 2 mM coenzyme A, 2 mM FAD, 7.5 mM $MgCl_2$, 10 mM ATP, 10 mM NADPH and 2 mM of the fatty acid thioester to be incorporated into the pyreudione structure which was dissolved in DMSO. To ensure that the Pys PCP domain was loaded with the ppan moiety, 5 µM of Sfp R4-4 (Q97) were included in the 100 µL of total reaction volume. The assay was set up on ice, most components in stock were thawed directly in advance and NADPH was freshly prepared. The reaction tube was transferred into a tabletop shaker and incubated at 30 °C and 350 rpm over night. Enzyme activity was quenched by the addition of an equal amount of MeOH and vigorous vortexing followed by 2 min centrifugation (16 000x g) to remove protein aggregates. The supernatant was then transferred to a fresh tube and 50 µL were taken for LC-MS analysis using the 30 min run (using own LC-MS) or the LCQ-Fleet 12 min run (see section 5.9.1).

5.9.6 Deacetylation of PA intermediates for brabantamide intermediate synthesis

N-(5,8-dioxo-1,2,3,5,8,8a-hexahydroindolizin-6-yl) hexanamide (**63**, 0.01g, 0.0379 mmol, 1 eq) was dissolved in a solution of $CoCl_2$ (38 µL, 2.5 mM), NaN_3 (0.85 µL, 154 µM) and phosphate buffer (0.232 mL, 0.1 M) and the pH was adjusted to 8.0. A solution of L-Acylase I (Sigma-Aldrich; 17 µL, 25 mg/mL, CAS 9012-37-7) was added and the reaction was left to shake overnight at 50 °C on a bench top thermal shaker. The reaction was quenched with 1 M HCl and acidified to pH 3.0, before washing with EtOAc (3x) to remove educt and lyophilization of the product-containing watery phase. The final product was measured via analytical HPLC on a C18 column (30 min method; A = H_2O + 0.05 % TFA, B = ACN + 0.05 % TFA; 0-2 min 5 % B, 2-28 min 5-100 % B, 28-29 min 100 % B, 29-30 min 5 % B). The product was obtained as a yellow-orange solid.

ESI-MS (+): calculated for $C_8H_{10}N_2O_2$ = 166.07, found 167.0 $[M+H]^+$.

5.10 Analytical methods

5.10.1 High Performance Liquid Chromatography (HPLC)

Analytic HPLC was used as a general method for the analysis of enzyme assay products, as a reaction control for synthetic routes or to analyze the composition of crude extracts and identify compounds of interest. A C18 column (Eurospher II 100 RP C-18, 5 µm, 250 x 16 mm; Knauer) including a precolumn (30 x 16 mm; Knauer) was used for separating problems in combination with the solvents H_2O (A) and ACN (B), both containing 0.05 % TFA, which functioned as the mobile phase. The separating process was controlled by a Knauer pump and UV detection system. Filtered samples were injected with volumes ranging between 10 and 50 µL. Semi-preparative HPLC was used for the purification of target compounds in amounts large enough to perform further characterization analyses. Crude extract samples were prepared analogously to those used for analytical HPLC, with 1 g of material being dissolved in 1 mL of MeOH. Before applying 200 – 700 µL of sample to the C18 column, the solution was filtered through a 0.22 µm filter (Fisher Scientific) to omit clogging of the column by occurring precipitate. Jasco System control and data acquisition was monitored using Chrompass chromatography software (JASCO). The sample was injected manually via a 1 mL loop system and fractions containing the protein of interest were collected in test tubes. Fractions containing the peaks of interest were combined and ACN was removed using the rotavapor apparatus (Heidolph). The remaining product left in ddH_2O containing TFA was lyophilized to obtain a powdery substance which was further analyzed by LC-MS and NMR.

Table 19. Compilation of most frequently used HPLC program runs used throughout this thesis.

30 min[1]			38 min			45 min		
Time	**Gradient [%**		**Time**	**Gradient [%]**		**Time**	**Gradient [%]**	
[min]	**A**	**B**	**[min]**	**A**	**B**	**[min]**	**A**	**B**
0	95	5	0	95	5	0	95	5
0 - 2	95	5	0 - 2	95	5	0 - 2	95	5
2 - 28	0	100	2 - 36	0	100	2 - 43	0	100
28 - 29	0	100	36 - 37	0	100	43 - 44	0	100
29 - 30	95	5	37 - 38	95	5	44 - 45	95	5

60 min			30 min[2]			30 min[3]		
Time	**Gradient [%]**		**Time**	**Gradient [%]**		**Time**	**Gradient [%]**	
[min]	**A**	**B**	**[min]**	**A**	**B**	**[min]**	**A**	**B**
0	95	5	0	95	5	0	95	5
0 - 2	95	5	0 - 2	75	25	0 - 2	60	40
2 - 57	0	100	2 - 36	0	100	2 - 43	0	100
58 - 59	0	100	36 - 37	0	100	43 - 44	0	100
59 - 60	95	5	37 - 38	95	5	44 - 45	95	5

[1]: 5 - 100 %; [2]: 25 - 100 %; [3]: 40 – 100 %

Table 20. Isocratic HPLC runs for sister peak isolation.

30 min[4]		
Time	**Gradient [%**	
[min]	**A**	**B**
0	95	5
0 - 2	40	60
2 - 28	40	60
28 - 29	0	100
29 - 30	95	5

[4]: isocratic 60 %

5.10.2 Liquid Chromatography coupled with Mass Spectrometry (LC-MS)

For direct mass detection of substance peaks identified in the recorded chromatogram, liquid chromatography was coupled with mass spectrometry. For HPLC-MS analysis, the Knauer chromatography analysis system described in section 5.9.1 was combined with an Advion expression CMS single quadrupole mass spectrometer using ESI for molecule ionization (Advion). Enzyme assays were measured with the more sensitive LCQ Fleet™ Ion Trap mass spectrometer (Thermo Fisher) together with the UltiMate 3000 HPLC system (Thermo Fisher). MS data analysis was performed using either Advion Data Express or Thermo Scientific Xcalibur software (see section 5.5)

Table 21. LCQ-Fleet program run used for sensitive molecule detection.

12 min

Time [min]	Gradient [% A	B
0	95	5
0 - 1	95	5
1 - 10	0	100
10 - 11	0	100
11 - 12	95	5

5.10.3 ^{1}H and ^{13}C-NMR and 2D-NMR

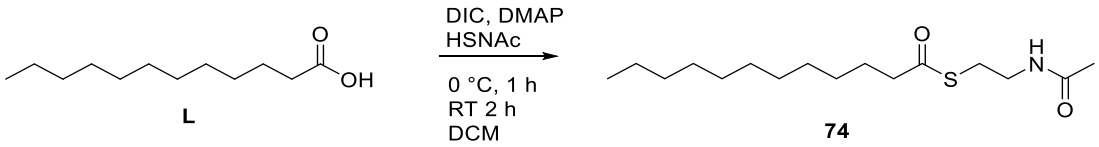
Structure determination of a purified molecule was achieved with ^{1}H and ^{13}C-NMR measurements executed in Bruker NMR spectrometers. For standard ^{1}H-NMR analysis, the Bruker Avance III HD 300 MHz was used, whereas longer and more complex measurements, including ^{13}C as well as COSY and NOESY, were executed with either Bruker Avance III 500 MHz C or Bruker Avance I 500 MHz spectrometer. Spectra were analyzed with MestReNova software (see section 5.5).

5.11 Chemical Synthesis

5.11.1 Synthesis of SNAc and *t*BuSH α-ketoacyl thioesters

Generation of α-ketoacyl thioester variations **74**, **101** and **102** was performed following the principle synthetic route described by COCHRANE et al.[195] for the preparation of S-2-acetamidoethyl decanethioate. A stirred solution of the desired carboxylic acid (0.5 g, 2.9 mmol) in dry DCM (25 mL) was prepared under argon atmosphere and cooled to 0 °C prior to addition of coupling reagents DIC (0.659 g, 3.19 mmol) and DMAP (0.039 g, 0.32 mmol). Following a 10 min preincubation, SNAc/*t*BuSH (0.450 g, 3.77 mmol) was given to the mixture, allowed to stir first at 0 °C for 1 h before slowly bringing the reaction back to RT and further stirring for 2 h. Precipitate formed during the reaction was separated from the solvent by filtration, washed with 1 M HCl as well as saturated $NaHCO_3$ and finally dried using $MgSO_4$ and via solvent evaporation. Silica column chromatography (95 DCM: 5 MeOH) yielded 55 - 78 % of a relatively pure white powder.

Representative: **S-(2-acetamidoethyl) dodecanethioate**



Scheme 24. Synthesis of **74**.

Chemical formula: $C_{16}H_{31}NO_2S$

Molecular weight: 301.49 g/mol

ESI-MS: calculated for $C_{16}H_{32}NO_2S$ [M+H]: 302.4973 and $C_{16}H_{31}NaNO_2S$ [M+Na]: 324.4792; measured [M+H]: 302.2 and [M+Na]: 324.2.

^{1}H-NMR (300 MHz in $CDCl_3$): δ 6.05 (s, 1H), 3.45 (d, J = 5.0 Hz, 2H), 3.03 (t, J = 6.4 Hz, 2H), 2.57 (t, J = 7.5 Hz, 2H), 2.00 (s, 3H), 1.68 (p, J = 7.8 Hz, 2H), 1.27 (m, 16H), 0.90 – 0.85 (t, J = 7.1,3H) (Figure S19).

Alternatively, a slightly modified method was tested. Carboxylic acids or acyl chlorides (1 eq, 0.693 mmol) were dissolved in DMF (3.5 mL/ mmol) under argon atmosphere and cooled to 0 °C using an ice bath. DIPEA (3 eq, 0.354 mL, 2.080 mmol) was added dropwise while stirring and incubated for 15 min. Supplementation with PyBOP and SNAc/*t*BuSH was followed by a 30 min incubation period at 0 °C and continuous stirring for 3 h at RT. Reaction product

formation was monitored via analytical HPLC after 30 min and 3 h. The reaction was diluted with EtOAc (15 mL), washed with saturated salt solutions of NH_4Cl (2x20 mL), $NaHCO_3$ (20 mL) and NaCl (20 mL), organic phases were combined, dried over $MgSO_4$, the solvent was evaporated, and the final products obtained in yields of over 50 %.

Representative: **S-(2-acetamidoethyl) octanethioate**

Chemical formula: $C_{12}H_{23}NO_2S$

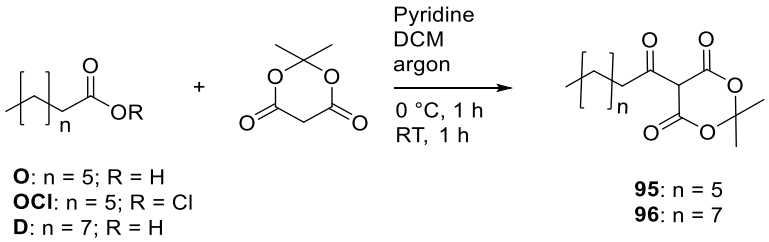
Molecular weight: 245.38 g/mol

ESI-MS: calculated for $C_{12}H_{23}NO_2S$ [M+H]: 246.147276 measured [M+H]: 246.5.

^{1}H NMR (400 MHz in $CDCl_3$): δ 6.11 (s, 1H), 3.47 (q, J = 5.9 Hz, 2H), 3.04 (t, J = 6.6 Hz, 2H), 2.61 – 2.55 (t, J = 6.6 Hz, 2H), 2.03 (s, 3H), 1.68 (m, J = 7.4 Hz, 2H), 1.31 – 1.26 (m, J = 6.3 Hz, 8H), 0.89 (t, J = 6.8 Hz, 3H) (Figure S20).

5.11.2 Synthesis of 5-alcanoyl-2,2-dimethyl-1,3-dioxane-4,6-diones

Following the protocol of OIKAWA et al.[187], a solution of Meldrum's acid (**94**, 0.177 g, 1.230 mmol) in DCM was prepared under argon atmosphere. Pyridine (2 eq, 0.198 mL, 2.459 mmol) was supplemented and the mixture was cooled to 0 °C before adding 1 equivalent of acyl chloride, followed by 1 h incubation each at 0 °C and later at RT. The reaction mixture was washed with 0.1 M HCl as well as H_2O (2x), the watery phases were discarded, the organic phases combined, dried with Na_2SO_4, filtered and the solvent removed from the product via evaporation to obtain the acyl Meldrum's acid derivative in 57 % yield.



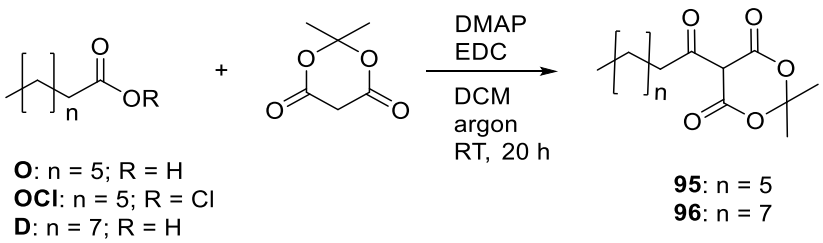
O: n = 5; R = H
OCl: n = 5; R = Cl
D: n = 7; R = H

95: n = 5
96: n = 7

Scheme 25. Synthesis of **95** and **96**.

Corresponding to the synthetic route described by RUYSBERGH et al.[192], the carboxylic acid of choice (1 eq, 4 mmol) was dissolved in dry DCM (40 mL) before adding DMAP (2.1 eq, 1.0262 g, 8.4 mmol), EDC (1.1 eq, 0.8435 g, 4.4 mmol) and Meldrum's acid (2 eq, 0.5765 g, 2

mmol) under Schlenk reaction conditions. The reaction mixture was stirred overnight at RT, the solvent was evaporated, the remaining raw product redissolved in EtOAc, washed (2 M HCl), dried ($MgSO_4$) and filtered to yield the acylated Meldrum's acid derivative with 48 % product outcome. The resulting product was used in subsequent reactions without further purification.



Scheme 26. Alternative synthetic route to generate **95** and **96**.

Representative: **2,2-dimethyl-5-octanoyl-1,3-dioxane-4,6-dione**

Chemical formula: $C_{14}H_{22}O_5$

Molecular weight: 270.33 g/mol

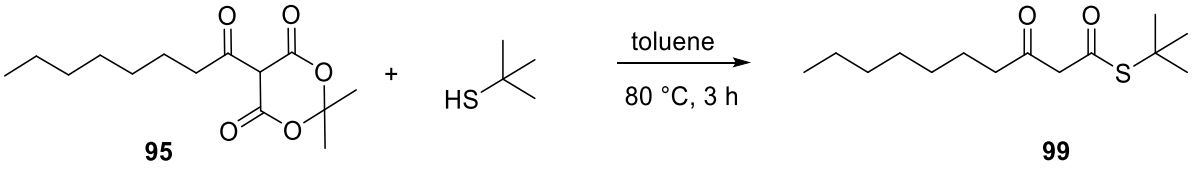
ESI-MS: calculated for $C_{14}H_{23}O_5$ [M+H]: 271.337276 and $C_{14}H_{22}NaO_5$ [M+Na]: 293.319218; measured [M+H]: 271.4 and [M+Na]: 293.7.

^{1}H-NMR (500 MHz in $CDCl_3$): δ 3.07 – 3.02 (t, J = 6.2 Hz, 2H), 1.72 (s, 6H), 1.70 – 1.64 (m, 2H), 1.42 – 1.35 (m, 2H), 1.33 – 1.24 (m, 7H), 0.86 (t, J = 7.0 Hz, 3H) (Figure S21).

5.11.3 Synthesis of 3-oxoacyl-SNAcs or 3-oxoacyl-*t*BuSHs

Acyl group transfer onto the CoA biomimetic SNAc/*t*BuSH thiol functional group was performed following the general procedure published by LIU et al.[194]. Acyl Meldrum's acid (0.270 g, 1 mmol) was prepared in 5 mL of toluene and mixed with a solution of SNAc (0.119 g, 1 mmol) or *t*BuSH dissolved in an equal amount of toluene. The reaction was stirred in a round-bottom flask for 3 h at 80 °C and a reddish-brown solid was obtained after solvent evaporation. The crude product was dissolved in EtOAc and purified via silica column chromatography to receive an orange solid with up to 47 % overall yield.

Representative: **S-(tert-butyl) 3-oxodecanethioate**



95 **99**

Scheme 27. Synthesis of **99** from **95.**

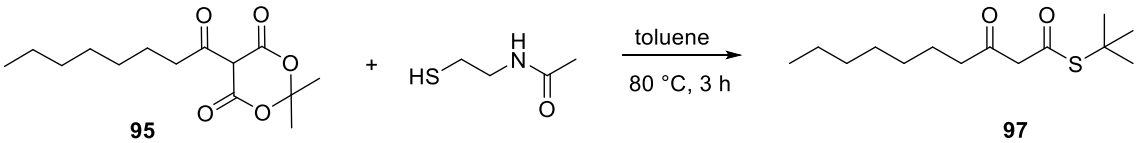
Chemical formula: $C_{14}H_{26}O_2S$

Molecular weight: 258.42 g/mol

ESI-MS: calculated for $C_{14}H_{27}O_2S$ [M+H]: 259. 427276 and $C_{14}H_{27}NaO_2S$ [M+Na]: 281.409218; measured [M+H]: 259.8 and [M+Na]: 281.7.

^{1}H-NMR (300 MHz in $CDCl_3$): δ 3.53 (s, 2H), 2.50 (t, *J* = 7.3 Hz, 2H), 1.53 (m, 2H), 1.45 (s, 9H), 1.28 – 1.22 (m, 9H), 0.86 (t, J = 7.14 Hz, 3H) (Figure S22).

Representative: **S-(2-acetamidoethyl) 3-oxodecanethioate**

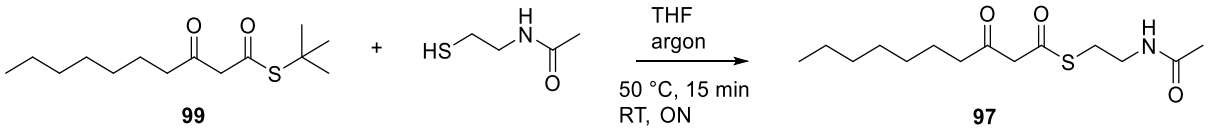


95 **97**

Scheme 28. Synthesis of **97** from **95.**

5.11.4 Transthioesterification of β-ketoacyl thioesters

S-(2-acetamidoethyl) 3-oxodecanethioate: Exchange of the thiol functionality mimicking CoA in selected conjugates was executed in an oxygen-free environment under argon atmosphere. S-(tert-butyl) 3-oxodecanethioate (1 eq, 0.4662 g, 1.806 mmol) was dissolved in THF, SNAc (5 eq, 0.960 mL, 9.029 mmol) was added to the round-bottom flask and the mixture was stirred for 15 min at 50 °C. The reaction was brought back to RT by removal of the oil bath and incubated overnight before workup including washing and drying steps as well as solvent evaporation and reaction control by analytical HPLC. The desired S-(2-acetamidoethyl) 3-oxodecanethioate was obtained with 51 % yield.



Scheme 29. Synthesis of **97** from **99**.

Chemical formula: $C_{14}H_{25}NO_3S$

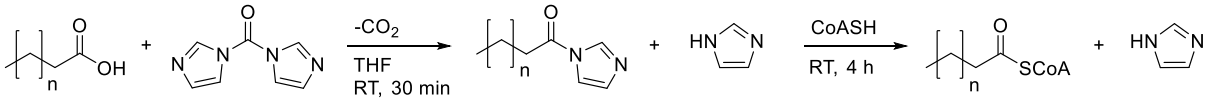
Molecular weight: 287.423 g/mol

ESI-MS: calculated for $C_{14}H_{26}NO_3S$ [M+H]: 288.430276 and $C_{14}H_{25}NaNO_3S$ [M+Na]: 310.412218 measured [M+H]: 288.20251 and [M+Na]: 310.40968.

^{1}H-NMR (300 MHz in $CDCl_3$): δ 5.88 (s, 1H), 3.46 – 3.41 (t, J = 5.9 Hz, 2H), 3.17 (s, 2H), 3.04 – 2.99 (t, J = 6.4 Hz, 2H), 2.58 - 2.53 (t, J = 6 Hz, 2H), 1.96 (s, 3H), 1.68 – 1.60 (m, 2H), 1.29 – 1.23 (m, 8H), 0.86 (t, J = 6.0 Hz, 3H). (Figure S23).

5.11.5 Synthesis of acyl-CoA thioesters

In principle, reaction procedures were performed as described by KAWAGUCHI et al.[186]. Initially, 5 µmol of fatty acid (decanoic, lauric acid) were dissolved in 0.1 mL THF under argon atmosphere. Carbonyl diimidazole (CDI; 6 µmol) was dissolved in 0.1 mL THF and added to the fatty acid, leaving the reaction to stir at room temperature for 30 min. After solvent removal, the intermediate was taken up in a 2:1 mixture of THF:H_2O (0.2 mL). CoASH (5 µmol), also dissolved in 0.5 mL of THF:H_2O (2:1), was added, the pH was set to 7 – 7.5 with NaOH and left to stir for 4 h at room temperature under argon atmosphere. THF was then removed by evaporation whereas the remaining water-soluble product was first acidified with dilute acetic acid and then lyophilized to yield the acyl-CoA thioester product.



O: n = 5
D: n = 7
U: n = 8
L: n = 9
P: n = 13

75: n = 5
72: n = 7
71: n = 8
68: n = 9
73: n = 13

Scheme 30. Multi-step synthetic route for establishment of **68**, **71** – **73** and **75**.

Chemical formula: I. Decanoyl-CoA: $C_{13}H_{54}N_7O_{17}P_3S$; II. Lauroyl-CoA: $C_{33}H_{58}N_7O_{17}P_3S$

Molecular weight: I. Decanoyl-CoA: 921.8 g/mol; II. Lauroyl-CoA: 949.8 g/mol

ESI-MS: I. Decanoyl-CoA: calculated for $C_{13}H_{55}N_7O_{17}P_3S$ [M+H]: 322.7972 and $C_{13}H_{54}NaN_7O_{17}P_3S$ [M+Na]: 344.7792; measured [M+H]: 922.8; II. Lauroyl-CoA: calculated for $C_{33}H_{59}N_7O_{17}P_3S$ [M+H]: 950.8072 and $C_{33}H_{58}NaN_7O_{17}P_3S$ [M+Na]: 972.7892; measured [M+H]: 950.7

6. Bibliography

1. Shen, B., A New Golden Age of Natural Products Drug Discovery. *Cell* 2015, *163* (6), 1297-1300.
2. Osbourn, A., Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends genet. : TIG* 2010, *26* (10), 449-57.
3. Roze, L. V.; Chanda, A.; Linz, J. E., Compartmentalization and molecular traffic in secondary metabolism: a new understanding of established cellular processes. *Fungal Genet. Biol. : FG & B* 2011, *48* (1), 35-48.
4. Buchanan, B.; Gruissem, W.; Jones, R., *Biochemistry & Molecular Biology of Plants*. *Am. Soc. Plant Physiol*.: 2000.
5. Dias, D. A.; Urban, S.; Roessner, U., A historical overview of natural products in drug discovery. *Metab.* 2012, *2* (2), 303-36.
6. Montaser, R.; Luesch, H., Marine natural products: a new wave of drugs? *Future Med. Chem.* 2011, *3* (12), 1475-89.
7. Krause, J.; Tobin, G., *Using Old Solutions to New Problems - Natural Drug Discovery in the 21st Century*. IntechOpen: 2013.
8. Schmitt, E. K.; Hoepfner, D.; Krastel, P., Natural products as probes in pharmaceutical research. *J. Ind. Microbiol. Biotechnol.* 2016, *43* (2-3), 249-260.
9. Harvey, A. L., Natural products in drug discovery. *Drug Discov. Today* 2008, *13* (19-20), 894-901.
10. Fleming, A., On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzæ. *Br. J. Exp. Pathol.* 1929, *10* (3), 226-236.
11. Lobanovska, M.; Pilla, G., Penicillin's Discovery and Antibiotic Resistance: Lessons for the Future? *Yale J. Biol. Med.* 2017, *90* (1), 135-145.
12. Levine, D. P., Vancomycin: a history. *Clinical infectious diseases : ACS Infect. Dis.* 2006, *42 Suppl 1*, S5-12.
13. Chopra, I.; Roberts, M., Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol. Mol. Biol. Rev. :* 2001, *65* (2), 232-60 .
14. Grossman, T. H., Tetracycline Antibiotics and Resistance. *Cold Spring Har. Perspect. Med.* 2016, *6* (4), a025387.

15. Zhang, H.; Wang, Y.; Wu, J.; Skalina, K.; Pfeifer, B. A., Complete Biosynthesis of Erythromycin A and Designed Analogs Using E. coli as a Heterologous Host. *Chem. Biol.* 2010, *17* (11), 1232-1240.
16. Ray, W. A.; Murray, K. T.; Meredith, S.; Narasimhulu, S. S.; Hall, K.; Stein, C. M., Oral erythromycin and the risk of sudden death from cardiac causes. *N. Engl. J. Med.* 2004, *351* (11), 1089-96.
17. Campbell, W. C.; Fisher, M. H.; Stapley, E. O.; Albers-Schönberg, G.; Jacob, T. A., Ivermectin: a potent new antiparasitic agent. *Science* 1983, *221* (4613), 823-8.
18. Meshnick, S. R., Artemisinin: mechanisms of action, resistance and toxicity. *Int. J. Parasitol.* 2002, *32* (13), 1655-60.
19. Zhang, Y.; Xu, G.; Zhang, S.; Wang, D.; Saravana Prabha, P.; Zuo, Z., Antitumor Research on Artemisinin and Its Bioactive Derivatives. *Nat. Prod. Bioprospect.* 2018, *8* (4), 303-319.
20. Tomasz, M., Mitomycin C: small, fast and deadly (but very selective). *Chem. Bio.I* 1995, *2* (9), 575-9.
21. Paz, M. M.; Zhang, X.; Lu, J.; Holmgren, A., A new mechanism of action for the anticancer drug mitomycin C: mechanism-based inhibition of thioredoxin reductase. *Chem. Res. Toxicol.* 2012, *25* (7), 1502-11.
22. Singla, A. K.; Garg, A.; Aggarwal, D., Paclitaxel and its formulations. *Int. J. Pharm.* 2002, *235* (1-2), 179-92.
23. Khanna, C.; Rosenberg, M.; Vail, D. M., A Review of Paclitaxel and Novel Formulations Including Those Suitable for Use in Dogs. *J. Vet. Int. Med.* 2015, *29* (4), 1006-12.
24. Feling, R. H.; Buchanan, G. O.; Mincer, T. J.; Kauffman, C. A.; Jensen, P. R.; Fenical, W., Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus salinospora. *Angew. Chem. Int. Ed.* 2003, *42* (3), 355-7.
25. Tobert, J. A., Lovastatin and beyond: the history of the HMG-CoA reductase inhibitors. *Nat. Rev. Drug Discov.* 2003, *2* (7), 517-26.
26. Bai, T.; Zhang, D.; Lin, S.; Long, Q.; Wang, Y.; Ou, H.; Kang, Q.; Deng, Z.; Liu, W.; Tao, M., Operon for biosynthesis of lipstatin, the Beta-lactone inhibitor of human pancreatic lipase. *Appl. Environ. Microbiol.* 2014, *80* (24), 7473-83.
27. Li, J. W.; Vederas, J. C., Drug discovery and natural products: end of an era or an endless frontier? *Science* 2009, *325* (5937), 161-5.
28. Burg, R. W.; Miller, B. M.; Baker, E. E.; Birnbaum, J.; Currie, S. A.; Hartman, R.; Kong, Y. L.; Monaghan, R. L.; Olson, G.; Putter, I.; Tunac, J. B.; Wallick, H.; Stapley, E. O.;

Oiwa, R.; Omura, S., Avermectins, new family of potent anthelmintic agents: producing organism and fermentation. *Antimicrob. Agents Chemother.* 1979, *15* (3), 361-367.

29. Ortholand, J. Y.; Ganesan, A., Natural products and combinatorial chemistry: back to the future. *Curr. Opin. Chem. Biol.* 2004, *8* (3), 271-80.
30. Newman, D. J.; Cragg, G. M., Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. *J. Nat. Prod.* 2012, *75* (3), 311-335.
31. Khaldi, N.; Seifuddin, F. T.; Turner, G.; Haft, D.; Nierman, W. C.; Wolfe, K. H.; Fedorova, N. D., SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fung. Gen. Biol. : FG & B* 2010, *47* (9), 736-41.
32. Skinnider, M. A.; Johnston, C. W.; Gunabalasingam, M.; Merwin, N. J.; Kieliszek, A. M.; MacLellan, R. J.; Li, H.; Ranieri, M. R. M.; Webster, A. L. H.; Cao, M. P. T.; Pfeifle, A.; Spencer, N.; To, Q. H.; Wallace, D. P.; Dejong, C. A.; Magarvey, N. A., Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. *Nat. Comm.* 2020, *11* (1), 6058.
33. Blin, K.; Medema, M. H.; Kottmann, R.; Lee, S. Y.; Weber, T., The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res.* 2017, *45* (D1), D555-d559.
34. Katz, L.; Baltz, R. H., Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* 2016, *43* (2-3), 155-76.
35. Luo, Y.; Cobb, R. E.; Zhao, H., Recent advances in natural product discovery. *Curr. Opin. Biotechnol.* 2014, *30*, 230-7.
36. Kunakom, S.; Eustáquio, A. S., Natural Products and Synthetic Biology: Where We Are and Where We Need To Go. *mSystems* 2019, *4* (3).
37. Fester, K.; Kutchan, T., Introduction to the Different Classes of Natural Products. 2009; pp 3-50.
38. Finking, R.; Marahiel, M. A., Biosynthesis of nonribosomal peptides1. *Annu. Rev. Microbiol.* 2004, *58*, 453-88.
39. Miller, B. R.; Drake, E. J.; Shi, C.; Aldrich, C. C.; Gulick, A. M., Structures of a Nonribosomal Peptide Synthetase Module Bound to MbtH-like Proteins Support a Highly Dynamic Domain Architecture *. *J. Biol. Chem.* 2016, *291* (43), 22559-22571.
40. Mach, B.; Reich, E.; Tatum, E. L., SEPARATION OF THE BIOSYNTHESIS OF THE ANTIBIOTIC POLYPEPTIDE TYROCIDINE FROM PROTEIN BIOSYNTHESIS. *Proc. Natl. Acad. Sci. U. S. A.* 1963, *50* (1), 175-181.

41. Kleinkauf, H.; Roskoski, R., Jr.; Lipmann, F., Pantetheine-linked peptide intermediates in gramicidin S and tyrocidine biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 1971, *68* (9), 2069-2072.
42. Gevers, W.; Kleinkauf, H.; Lipmann, F., The activation of amino acids for biosynthesis of gramicidin S. *Proc. Natl. Acad. Sci. U. S. A.* 1968, *60* (1), 269-276.
43. Gevers, W.; Kleinkauf, H.; Lipmann, F., Peptidyl transfers in gramicidin S bisoynthesis from enzyme-bound thioester intermediates. *Proc. Natl. Acad. Sci. U. S. A.* 1969, *63* (4), 1335-42.
44. Roskoski, R., Jr.; Gevers, W.; Kleinkauf, H.; Lipmann, F., Tyrocidine biosynthesis by three complementary fractions from Bacillus brevis (ATCC 8185). *Biochem.* 1970, *9* (25), 4839-45.
45. Sieber, S. A.; Marahiel, M. A., Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chemi. Rev.* 2005, *105* (2), 715-38.
46. Rausch, C.; Hoof, I.; Weber, T.; Wohlleben, W.; Huson, D. H., Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC. Evol. Biol.* 2007, *7*, 78-78.
47. Guenzi, E.; Galli, G.; Grgurina, I.; Gross, D. C.; Grandi, G., Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases. *J. Biol. Chem.* 1998, *273* (49), 32857-63.
48. Marahiel, M. A.; Essen, L. O., Chapter 13. Nonribosomal peptide synthetases mechanistic and structural aspects of essential domains. *Meth. Enzymol.* 2009, *458*, 337-51.
49. Marahiel, M. A.; Stachelhaus, T.; Mootz, H. D., Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem. Rev.* 1997, *97* (7), 2651-2674.
50. Miller, B.; Gulick, A., Structural Biology of Non-Ribosomal Peptide Synthetases. 2016; Vol. 1401, pp 3-29.
51. Felnagle, E. A.; Barkei, J. J.; Park, H.; Podevels, A. M.; McMahon, M. D.; Drott, D. W.; Thomas, M. G., MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochemistry* 2010, *49* (41), 8815-8817.
52. Conti, E.; Stachelhaus, T.; Marahiel, M. A.; Brick, P., Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.I* 1997, *16* (14), 4174-83.
53. Tufar, P.; Rahighi, S.; Kraas, F. I.; Kirchner, D. K.; Löhr, F.; Henrich, E.; Köpke, J.; Dikic, I.; Güntert, P.; Marahiel, M. A.; Dötsch, V., Crystal structure of a PCP/Sfp

complex reveals the structural basis for carrier protein posttranslational modification. *Chem. Biol.* 2014, *21* (4), 552-562.

54. Kittilä, T.; Mollo, A.; Charkoudian, L. K.; Cryle, M. J., New Structural Data Reveal the Motion of Carrier Proteins in Nonribosomal Peptide Synthesis. *Angew. Chem. Int. Ed.* 2016, *55* (34), 9834-40.
55. Koglin, A.; Mofid, M. R.; Löhr, F.; Schäfer, B.; Rogov, V. V.; Blum, M. M.; Mittag, T.; Marahiel, M. A.; Bernhard, F.; Dötsch, V., Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science* 2006, *312* (5771), 273-6.
56. Felnagle, E. A.; Jackson, E. E.; Chan, Y. A.; Podevels, A. M.; Berti, A. D.; McMahon, M. D.; Thomas, M. G., Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Mol. Pharmaceut.* 2008, *5* (2), 191-211.
57. Stachelhaus, T.; Mootz, H. D.; Bergendahl, V.; Marahiel, M. A., Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. *J. Biol. Chem.* 1998, *273* (35), 22773-81.
58. Bloudoff, K.; Schmeing, T. M., Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochim. Biophys. Acta Proteins Proteom.* 2017, *1865* (11 Pt B), 1587-1604.
59. Roche, E. D.; Walsh, C. T., Dissection of the EntF condensation domain boundary and active site residues in nonribosomal peptide synthesis. *Biochemistry* 2003, *42* (5), 1334-44.
60. Kohli, R. M.; Trauger, J. W.; Schwarzer, D.; Marahiel, M. A.; Walsh, C. T., Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases. *Biochemistry* 2001, *40* (24), 7099-108.
61. Bruner, S. D.; Weber, T.; Kohli, R. M.; Schwarzer, D.; Marahiel, M. A.; Walsh, C. T.; Stubbs, M. T., Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure (London, England : 1993)* 2002, *10* (3), 301-10.
62. Walsh, C. T.; Chen, H.; Keating, T. A.; Hubbard, B. K.; Losey, H. C.; Luo, L.; Marshall, C. G.; Miller, D. A.; Patel, H. M., Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr. Opin. Chem. Biol.* 2001, *5* (5), 525-34.
63. Balibar, C. J.; Vaillancourt, F. H.; Walsh, C. T., Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. *Chem. Biol.* 2005, *12* (11), 1189-200.

64. Dowling, D. P.; Kung, Y.; Croft, A. K.; Taghizadeh, K.; Kelly, W. L.; Walsh, C. T.; Drennan, C. L., Structural elements of an NRPS cyclization domain and its intermodule docking domain. *Proc. Natl. Acad. Sci. U. S. A.* 2016, *113* (44), 12432-12437.
65. Ali, H.; Ries, M. I.; Lankhorst, P. P.; van der Hoeven, R. A.; Schouten, O. L.; Noga, M.; Hankemeier, T.; van Peij, N. N.; Bovenberg, R. A.; Vreeken, R. J.; Driessen, A. J., A non-canonical NRPS is involved in the synthesis of fungisporin and related hydrophobic cyclic tetrapeptides in Penicillium chrysogenum. *PLoS One* 2014, *9* (6), e98212.
66. Peypoux, F.; Bonmatin, J. M.; Wallach, J., Recent trends in the biochemistry of surfactin. *Appl. Microbiol. Biotechnol.* 1999, *51* (5), 553-63.
67. Raafat El-Gewely, M., *Biotechnol. Annual Rev.* 2003; Vol. 9.
68. Hoyer, K. M.; Mahlert, C.; Marahiel, M. A., The iterative gramicidin s thioesterase catalyzes peptide ligation and cyclization. *Chem. Biol.* 2007, *14* (1), 13-22.
69. Mootz, H. D.; Schwarzer, D.; Marahiel, M. A., Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *Chembiochem.* 2002, *3* (6), 490-504.
70. Marshall, C. G.; Hillson, N. J.; Walsh, C. T., Catalytic Mapping of the Vibriobactin Biosynthetic Enzyme VibF. *Biochemistry* 2002, *41* (1), 244-250.
71. Gokhale, R. S.; Tuteja, D., Biochemistry of Polyketide Synthases. In *Biotechnol. Set.* 2001; pp 341-372.
72. Shen, B., Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* 2003, *7* (2), 285-95.
73. Caffrey, P., Dissecting complex polyketide biosynthesis. *Comput. Struct. Biotechnol. J.* 2012, *3*, e201210010.
74. Robbins, T.; Liu, Y. C.; Cane, D. E.; Khosla, C., Structure and mechanism of assembly line polyketide synthases. *Curr.Opin. Struct. Biol.* 2016, *41*, 10-18.
75. Weissman, K. J., Chapter 1 Introduction to Polyketide Biosynthesis. In *Meth. Enzymol.* Academic Press: 2009; Vol. 459, pp 3-16.
76. Robbins, T.; Kapilivsky, J.; Cane, D. E.; Khosla, C., Roles of Conserved Active Site Residues in the Ketosynthase Domain of an Assembly Line Polyketide Synthase. *Biochemistry* 2016, *55* (32), 4476-4484.
77. Bush, L. P.; Fannin, F. F.; Siegel, M. R.; Dahlman, D. L.; Burton, H. R., Chemistry, occurrence and biological effects of saturated pyrrolizidine alkaloids associated with endophyte-grass interactions. *Agric. Ecosyst. Environ.* 1993, *44* (1/4), 81-102.

78. Fu, P. P.; Xia, Q.; Lin, G.; Chou, M. W., Pyrrolizidine alkaloids--genotoxicity, metabolism enzymes, metabolic activation, and mechanisms. *Drug Metab.Rev.* 2004, *36* (1), 1-55.
79. Schimming, O.; Challinor, V. L.; Tobias, N. J.; Adihou, H.; Grün, P.; Pöschel, L.; Richter, C.; Schwalbe, H.; Bode, H. B., Structure, Biosynthesis, and Occurrence of Bacterial Pyrrolizidine Alkaloids. *Angew. Chem. Int. Ed.* 2015, *54* (43), 12702-5.
80. Robertson, J.; Stevens, K., Pyrrolizidine alkaloids: occurrence, biology, and chemical synthesis. *Nat. Prod. Rep.* 2017, *34* (1), 62-89.
81. Ruan, J.; Yang, M.; Fu, P.; Ye, Y.; Lin, G., Metabolic activation of pyrrolizidine alkaloids: insights into the structural and enzymatic basis. *Chem. Res. Toxicol.* 2014, *27* (6), 1030-9.
82. Fu, P.; Chou, M.; Xia, Q.; Yang, Y.; Yan, J.; Doerge, D.; Chan, P.-C., GENOTOXIC PYRROLIZIDINE ALKALOIDS AND PYRROLIZIDINE ALKALOID N-OXIDEs MECHANISMS LEADING TO DNA ADDUCT FORMATION AND TUMORIGENICITY. *J. Environ. Sci. Health, Part C* 2001, *19*, 353 - 385.
83. He, Y. Q.; Yang, L.; Liu, H. X.; Zhang, J. W.; Liu, Y.; Fong, A.; Xiong, A. Z.; Lu, Y. L.; Yang, L.; Wang, C. H.; Wang, Z. T., Glucuronidation, a new metabolic pathway for pyrrolizidine alkaloids. *Chem. Res. Toxicol.* 2010, *23* (3), 591-9.
84. Wiedenfeld, H., Plants containing pyrrolizidine alkaloids: toxicity and problems. *Food Addit. Contam. Part A - Chem.* 2011, *28* (3), 282-92.
85. Ober, D.; Kaltenegger, E., Pyrrolizidine alkaloid biosynthesis, evolution of a pathway in plant secondary metabolism. *Phytochemistry* 2009, *70* (15), 1687-1695.
86. Langel, D.; Ober, D.; Pelser, P., The evolution of pyrrolizidine alkaloid biosynthesis and diversity in the Senecioneae. *Phytochem. Rev.* 2011, *10*, 3-74.
87. Böttcher, F.; Ober, D.; Hartmann, T., Biosynthesis of pyrrolizidine alkaloids: Putrescine and spermidine are essential substrates of enzymatic homospermidine formation. *Can. J. Chem.* 2011, *72*, 80-85.
88. Blankenship, J. D.; Houseknecht, J. B.; Pal, S.; Bush, L. P.; Grossman, R. B.; Schardl, C. L., Biosynthetic precursors of fungal pyrrolizidines, the loline alkaloids. *Chembiochem.* 2005, *6* (6), 1016-22.
89. Schardl, C. L.; Grossman, R. B.; Nagabhyru, P.; Faulkner, J. R.; Mallik, U. P., Loline alkaloids: Currencies of mutualism. *Phytochemistry* 2007, *68* (7), 980-96.
90. Pan, J.; Bhardwaj, M.; Faulkner, J. R.; Nagabhyru, P.; Charlton, N. D.; Higashi, R. M.; Miller, A. F.; Young, C. A.; Grossman, R. B.; Schardl, C. L., Ether bridge formation in loline alkaloid biosynthesis. *Phytochemistry* 2014, *98*, 60-8.

91. Huang, S.; Tabudravu, J.; Elsayed, S. S.; Travert, J.; Peace, D.; Tong, M. H.; Kyeremeh, K.; Kelly, S. M.; Trembleau, L.; Ebel, R.; Jaspars, M.; Yu, Y.; Deng, H., Discovery of a Single Monooxygenase that Catalyzes Carbamate Formation and Ring Contraction in the Biosynthesis of the Legonmycins. *Angew. Chem. Int. Ed.* 2015, *54* (43), 12697-701.
92. Torres Pazmiño, D. E.; Winkler, M.; Glieder, A.; Fraaije, M. W., Monooxygenases as biocatalysts: Classification, mechanistic aspects and biotechnological applications. *J. Biotechnol.* 2010, *146* (1-2), 9-24.
93. Maier, T.; Förster, H. H.; Asperger, O.; Hahn, U., Molecular characterization of the 56-kDa CYP153 from Acinetobacter sp. EB104. *Biochem. Biophys. Res. Comm.* 2001, *286* (3), 652-8.
94. Murrell, J. C.; Gilbert, B.; McDonald, I. R., Molecular biology and regulation of methane monooxygenase. *Arch. Microbiol.* 2000, *173* (5-6), 325-32.
95. Mascotti, M. L.; Juri Ayub, M.; Furnham, N.; Thornton, J. M.; Laskowski, R. A., Chopping and Changing: the Evolution of the Flavin-dependent Monooxygenases. *J. Mol. Biol.* 2016, *428* (15), 3131-46.
96. van Berkel, W. J.; Kamerbeek, N. M.; Fraaije, M. W., Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *J. Biotechnol.* 2006, *124* (4), 670-89.
97. Schmidt, Y.; van der Voort, M.; Crüsemann, M.; Piel, J.; Josten, M.; Sahl, H. G.; Miess, H.; Raaijmakers, J. M.; Gross, H., Biosynthetic origin of the antibiotic cyclocarbamate brabantamide A (SB-253514) in plant-associated Pseudomonas. *Chembiochem.* 2014, *15* (2), 259-66.
98. Johnston, C. W.; Zvanych, R.; Khyzha, N.; Magarvey, N. A., Nonribosomal assembly of natural lipocyclocarbamate lipoprotein-associated phospholipase inhibitors. *Chembiochem.* 2013, *14* (4), 431-5.
99. Terness, P.; Oelert, T.; Ehser, S.; Chuang, J. J.; Lahdou, I.; Kleist, C.; Velten, F.; Hämmerling, G. J.; Arnold, B.; Opelz, G., Mitomycin C-treated dendritic cells inactivate autoreactive T cells: toward the development of a tolerogenic vaccine in autoimmune diseases. *Proc. Natl. Acad. Sci. U. S. A.* 2008, *105* (47), 18442-7.
100. Buechter, D. D.; Thurston, D. E., Studies on the pyrrolizidine antitumor agent, clazamycin: interconversion of clazamycins A and B. *J Nat Prod* 1987, *50* (3), 360-7.
101. Hori, M.; Naito, K.; Sakata, N.; Uehara, Y.; Umezawa, H., Inhibition of DNA replication and membrane transport of some nutrients by clazamycin in Escherichia coli. *J. Antibiot.* 1984, *37* (3), 260-6.

102. Jiang, B.; Zhao, W.; Li, S.; Liu, H.; Yu, L.; Zhang, Y.; He, H.; Wu, L., Cytotoxic Dibohemamines D-F from a Streptomyces Species. *J. Nat. Prod.* 2017, *80* (10), 2825-2829.
103. Liu, L.; Li, S.; Sun, R.; Qin, X.; Ju, J.; Zhang, C.; Duan, Y.; Huang, Y., Activation and Characterization of Bohemamine Biosynthetic Gene Cluster from Streptomyces sp. CB02009. *Org. Lett.* 2020, *22* (12), 4614-4619.
104. Subramanian, B.; Gao, S.; Lercher, M. J.; Hu, S.; Chen, W.-H., Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res.* 2019, *47* (W1), W270-W275.
105. Nah, H. J.; Pyeon, H. R.; Kang, S. H.; Choi, S. S.; Kim, E. S., Cloning and Heterologous Expression of a Large-sized Natural Product Biosynthetic Gene Cluster in Streptomyces Species. *Front. Microbiol.* 2017, *8*, 394.
106. Kouprina, N.; Larionov, V., TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. *Nat. Rev. Genet.* 2006, *7* (10), 805-12.
107. Fu, J.; Bian, X.; Hu, S.; Wang, H.; Huang, F.; Seibert, P. M.; Plaza, A.; Xia, L.; Müller, R.; Stewart, A. F.; Zhang, Y., Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* 2012, *30* (5), 440-6.
108. Greunke, C.; Duell, E. R.; D'Agostino, P. M.; Glöckle, A.; Lamm, K.; Gulder, T. A. M., Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential. *Metab. Eng.* 2018, *47*, 334-345.
109. D'Agostino, P. M.; Gulder, T. A. M., Direct Pathway Cloning Combined with Sequence- and Ligation-Independent Cloning for Fast Biosynthetic Gene Cluster Refactoring and Heterologous Expression. *ACS Synth. Biol.* 2018, *7* (7), 1702-1708.
110. Nolden, T.; Pfaff, F.; Nemitz, S.; Freuling, C. M.; Höper, D.; Müller, T.; Finke, S., Reverse genetics in high throughput: rapid generation of complete negative strand RNA virus cDNA clones and recombinant viruses thereof. *Sci. Rep.* 2016, *6* (1), 23887.
111. Naville, M.; Ghuillot-Gaudeffroy, A.; Marchais, A.; Gautheret, D., ARNold: a web tool for the prediction of Rho-independent transcription terminators. *RNA Biol.* 2011, *8* (1), 11-3.
112. Kallunki, T.; Barisic, M.; Jäättelä, M.; Liu, B., How to Choose the Right Inducible Gene Expression System for Mammalian Studies? *Cells* 2019, *8* (8), 796.
113. Li, M. Z.; Elledge, S. J., SLIC: a method for sequence- and ligation-independent cloning. *Methods Mol. Biol. (Clifton, N.J.)* 2012, *852*, 51-9.

114. Horbal, L.; Siegl, T.; Luzhetskyy, A., A set of synthetic versatile genetic control elements for the efficient expression of genes in Actinobacteria. *Sci. Rep.* 2018, *8*.
115. Hong, Z.; Bolard, A.; Giraud, C.; Prévost, S.; Genta-Jouve, G.; Deregnaucourt, C.; Häussler, S.; Jeannot, K.; Li, Y., Azetidine-Containing Alkaloids Produced by a Quorum-Sensing- Regulated Nonribosomal Peptide Synthetase Pathway in Pseudomonas aeruginosa. *Angew. Chem.* 2018, *131*.
116. Patteson, J. B.; Lescallette, A. R.; Li, B., Discovery and Biosynthesis of Azabicyclene, a Conserved Nonribosomal Peptide in Pseudomonas aeruginosa. *Org. Lett.* 2019, *21* (13), 4955-4959.
117. Fan, H.; Wang, J.; Komiyama, M.; Liang, X., Effects of secondary structures of DNA templates on the quantification of qPCR. *J. Biomol. Struct. Dyn.* 2019, *37* (11), 2867-2874.
118. Tolmie, C.; Smit, M. S.; Opperman, D. J., Native roles of Baeyer-Villiger monooxygenases in the microbial metabolism of natural compounds. *Nat. Prod. Rep.* 2019, *36* (2), 326-353.
119. Düzenli, Ö.; Okay, S., Promoter engineering for the recombinant protein production in prokaryotic systems. *AIMS Bioeng.* 2020, *7*, 62-81.
120. Lee, S. K.; Chou, H. H.; Pfleger, B. F.; Newman, J. D.; Yoshikuni, Y.; Keasling, J. D., Directed evolution of AraC for improved compatibility of arabinose- and lactose-inducible promoters. *Appl. Environ. Microbiol.* 2007, *73* (18), 5711-5.
121. Science, I. *SnapGene*, 5.2.4.
122. Gruss, A.; Moretto, V.; Ehrlich, S. D.; Duwat, P.; Dabert, P., GC-rich DNA sequences block homologous recombination in vitro. *J. Biol. Chem.* 1991, *266* (11), 6667-9.
123. Jones, K. L.; Kim, S. W.; Keasling, J. D., Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. *Metabol. Eng.* 2000, *2* (4), 328-38.
124. Rutledge, P. J.; Challis, G. L., Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* 2015, *13* (8), 509-23.
125. Cooper, G., *The Cell: A Molecular Approach*. Sinauer Associates: Sunderland (MA), 2000; Vol. 2.
126. Galm, U.; Shen, B., Expression of biosynthetic gene clusters in heterologous hosts for natural product production and combinatorial biosynthesis. *Expert Opin. Drug Discov.* 2006, *1* (5), 409-37.

127. Luo, Y.; Enghiad, B.; Zhao, H., New tools for reconstruction and heterologous expression of natural product biosynthetic gene clusters. *Nat. Prod. Rep.* 2016, *33* (2), 174-82.
128. Gaitatzis, N.; Hans, A.; Müller, R.; Beyer, S., The mtaA Gene of the Myxothiazol Biosynthetic Gene Cluster from Stigmatella aurantiaca DW4/3-1 Encodes a Phosphopantetheinyl Transferase that Activates Polyketide Synthases and Polypeptide Synthetases. *J. Biochem.* 2001, *129*, 119-24.
129. Gomez-Escribano, J. P.; Bibb, M. J., Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechnol.* 2011, *4* (2), 207-15.
130. Bekiesch, P.; Basitta, P.; Apel, A. K., Challenges in the Heterologous Production of Antibiotics in Streptomyces. *Arch. Pharm.* 2016, *349* (8), 594-601.
131. Myronovskyi, M.; Luzhetskyy, A., Heterologous production of small molecules in the optimized Streptomyces hosts. *Nat. Prod. Rep.* 2019, *36* (9), 1281-1294.
132. Jenkins, B.; West, J. A.; Koulman, A., A review of odd-chain fatty acid metabolism and the role of pentadecanoic Acid (c15:0) and heptadecanoic Acid (c17:0) in health and disease. *Molecules (Basel, Switzerland)* 2015, *20* (2), 2425-44.
133. Park, Y.-K.; Ledesma-Amaro, R.; Nicaud, J.-M., De novo Biosynthesis of Odd-Chain Fatty Acids in Yarrowia lipolytica Enabled by Modular Pathway Engineering. *Front. Bioeng. Biotechnol.* 2020, *7*, 484-484.
134. Struck, A. W.; Thompson, M. L.; Wong, L. S.; Micklefield, J., S-adenosyl-methionine-dependent methyltransferases: highly versatile enzymes in biocatalysis, biosynthesis and other biotechnological applications. *Chembiochem.* 2012, *13* (18), 2642-55.
135. Ma, J. F.; Shinada, T.; Matsuda, C.; Nomoto, K., Biosynthesis of phytosiderophores, mugineic acids, associated with methionine cycling. *J. Biolog. Chem.* 1995, *270* (28), 16549-54.
136. Greule, A.; Stok, J. E.; De Voss, J. J.; Cryle, M. J., Unrivalled diversity: the many roles and reactions of bacterial cytochromes P450 in secondary metabolism. *Nat. Prod. Rep.* 2018, *35* (8), 757-791.
137. Kavanagh, K. L.; Jörnvall, H.; Persson, B.; Oppermann, U., Medium- and short-chain dehydrogenase/reductase gene and protein families : the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell. Mol. Life Sci.* 2008, *65* (24), 3895-906.

138. Luengo, J. M.; García, J. L.; Olivera, E. R., The phenylacetyl-CoA catabolon: a complex catabolic unit with broad biotechnological applications. *Mol. Microbiol.* 2001, *39* (6), 1434-42.
139. Stover, C. K.; Pham, X. Q.; Erwin, A. L.; Mizoguchi, S. D.; Warrener, P.; Hickey, M. J.; Brinkman, F. S.; Hufnagle, W. O.; Kowalik, D. J.; Lagrou, M.; Garber, R. L.; Goltry, L.; Tolentino, E.; Westbrock-Wadman, S.; Yuan, Y.; Brody, L. L.; Coulter, S. N.; Folger, K. R.; Kas, A.; Larbig, K.; Lim, R.; Smith, K.; Spencer, D.; Wong, G. K.; Wu, Z.; Paulsen, I. T.; Reizer, J.; Saier, M. H.; Hancock, R. E.; Lory, S.; Olson, M. V., Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. *Nature* 2000, *406* (6799), 959-64.
140. Katsu, T.; Yoshimura, S.; Fujita, Y., Increases in permeability of Escherichia coli outer membrane induced by polycations. *FEBS Lett.* 1984, *166* (1), 175-8.
141. Vaara, M., Agents that increase the permeability of the outer membrane. *Microbiol. Rev.* 1992, *56* (3), 395-411.
142. Kallberg, Y.; Oppermann, U.; Persson, B., Classification of the short-chain dehydrogenase/reductase superfamily using hidden Markov models. *FEBS J.* 2010, *277* (10), 2375-86.
143. Rosano, G. L.; Ceccarelli, E. A., Recombinant protein expression in Escherichia coli: advances and challenges. *Front. Microbiol.* 2014, *5*, 172.
144. Zhao, X.; Li, G.; Liang, S., Several Affinity Tags Commonly Used in Chromatographic Purification. *J. Anal. Methods Chem.* 2013, *2013*, 581093.
145. García-Contreras, R.; Vos, P.; Westerhoff, H. V.; Boogerd, F. C., Why in vivo may not equal in vitro - new effectors revealed by measurement of enzymatic activities under the same in vivo-like assay conditions. *FEBS J.* 2012, *279* (22), 4145-59.
146. Pan, S. H.; Malcolm, B. A., Reduced background expression and improved plasmid stability with pET vectors in BL21 (DE3). *BioTechniques* 2000, *29* (6), 1234-8.
147. Jia, B.; Jeon, C. O., High-throughput recombinant protein expression in Escherichia coli: current status and future perspectives. *Open Biol.* 2016, *6* (8).
148. Mirzadeh, K.; Martínez, V.; Toddo, S.; Guntur, S.; Herrgård, M. J.; Elofsson, A.; Nørholm, M. H.; Daley, D. O., Enhanced Protein Production in Escherichia coli by Optimization of Cloning Scars at the Vector-Coding Sequence Junction. *ACS Synth. Biol.* 2015, *4* (9), 959-65.
149. Liao, C.; Seebeck, F. P., S-adenosylhomocysteine as a methyl transfer catalyst in biocatalytic methylation reactions. *Nat. Catal.* 2019, *2* (8), 696-701.

150. Donovan, R. S.; Robinson, C. W.; Glick, B. R., Review: optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter. *J. Ind. Microbiol.* 1996, *16* (3), 145-54.
151. Kram, K. E.; Finkel, S. E., Rich Medium Composition Affects Escherichia coli Survival, Glycation, and Mutation Frequency during Long-Term Batch Culture. *Appl. Environ. Microbiol.* 2015, *81* (13), 4442-50.
152. Peternel, S.; Grdadolnik, J.; Gaberc-Porekar, V.; Komel, R., Engineering inclusion bodies for non denaturing extraction of functional proteins. *Microb. cell factories* 2008, *7*, 34 (1-32).
153. Robichon, C.; Luo, J.; Causey, T. B.; Benner, J. S.; Samuelson, J. C., Engineering Escherichia coli BL21(DE3) derivative strains to minimize E. coli protein contamination after purification by immobilized metal affinity chromatography. *Appl. Environ. Microbiol.* 2011, *77* (13), 4634-46.
154. Costa, S.; Almeida, A.; Castro, A.; Domingues, L., Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system. *Front. Microbiol.* 2014, *5*, 63-63.
155. Leibly, D. J.; Nguyen, T. N.; Kao, L. T.; Hewitt, S. N.; Barrett, L. K.; Van Voorhis, W. C., Stabilizing additives added during cell lysis aid in the solubilization of recombinant proteins. *PLoS One* 2012, *7* (12), e52482-e52482.
156. Golovanov, A. P.; Hautbergue, G. M.; Wilson, S. A.; Lian, L.-Y., A Simple Method for Improving Protein Solubility and Long-Term Stability. *J. Am. Chem. Soc.* 2004, *126* (29), 8933-8939.
157. Chhetri, G.; Kalita, P.; Tripathi, T., An efficient protocol to enhance recombinant protein expression using ethanol in Escherichia coli. *MethodsX* 2015, *2*, 385-91.
158. Arora, P.; Vats, A.; Saxena, P.; Mohanty, D.; Gokhale, R. S., Promiscuous Fatty Acyl CoA Ligases Produce Acyl-CoA and Acyl-SNAC Precursors for Polyketide Biosynthesis. *J. Am. Chem. Soc.* 2005, *127* (26), 9388-9389.
159. Morgan-Kiss, R. M.; Cronan, J. E., The Escherichia coli fadK (ydiD) gene encodes an anerobically regulated short chain acyl-CoA synthetase. *J. Biolog. Chem.* 2004, *279* (36), 37324-33.
160. Yoo, J. H.; Cheng, O. H.; Gerber, G. E., Determination of the native form of FadD, the Escherichia coli fatty acyl-CoA synthetase, and characterization of limited proteolysis by outer membrane protease OmpT. *Biochem. J.* 2001, *360* (Pt 3), 699-706.
161. Daniel, R. M.; Danson, M. J., Temperature and the catalytic activity of enzymes: a fresh understanding. *FEBS Lett.* 2013, *587* (17), 2738-43.

162. Black, P. N., Characterization of FadL-specific fatty acid binding in Escherichia coli. *Biochim. Biophys. Acta* 1990, *1046* (1), 97-105.
163. Linne, U.; Marahiel, M., Reactions Catalyzed by Mature and Recombinant Nonribosomal Peptide Synthetases. *Meth. Enzymol.* 2004, *388*, 293-315.
164. Weber, T.; Marahiel, M. A., Exploring the Domain Structure of Modular Nonribosomal Peptide Synthetases. *Structure (London, England : 1993)* 2001, *9* (1), R3-R9.
165. Franke, J.; Hertweck, C., Biomimetic Thioesters as Probes for Enzymatic Assembly Lines: Synthesis, Applications, and Challenges. *Cell Chem. Biol.* 2016, *23* (10), 1179-1192.
166. Goh, H. C.; Sobota, R. M.; Ghadessy, F. J.; Nirantar, S., Going native: Complete removal of protein purification affinity tags by simple modification of existing tags and proteases. *Protein Expr. Purif.* 2017, *129*, 18-24.
167. Sabaty, M.; Grosse, S.; Adryanczyk, G.; Boiry, S.; Biaso, F.; Arnoux, P.; Pignol, D., Detrimental effect of the 6 His C-terminal tag on YedY enzymatic activity and influence of the TAT signal sequence on YedY synthesis. *BMC Biochem.* 2013, *14*, 28.
168. Zou, Y.; Yin, J., Phosphopantetheinyl transferase catalyzed site-specific protein labeling with ADP conjugated chemical probes. *J. Am. Chem. Soc.* 2009, *131* (22), 7548-9.
169. Kotowska, M.; Pawlik, K., Roles of type II thioesterases and their application for secondary metabolite yield improvement. *Appl. Microbiol. Biotechnol.* 2014, *98* (18), 7735-46.
170. Aslantas, Y.; Surmeli, N. B., Effects of N-Terminal and C-Terminal Polyhistidine Tag on the Stability and Function of the Thermophilic P450 CYP119. *Bioinorg. Chem. Appl.* 2019, *2019*, 8080697.
171. Booth, W. T.; Schlachter, C. R.; Pote, S.; Ussin, N.; Mank, N. J.; Klapper, V.; Offermann, L. R.; Tang, C.; Hurlburt, B. K.; Chruszcz, M., Impact of an N-terminal Polyhistidine Tag on Protein Thermal Stability. *ACS Omega* 2018, *3* (1), 760-768.
172. Klapper, M.; Götze, S.; Barnett, R.; Willing, K.; Stallforth, P., Bacterial Alkaloids Prevent Amoebal Predation. *Angew. Chem. Int. Ed.* 2016, *55* (31), 8944-7.
173. Klapper, M.; Braga, D.; Lackner, G.; Herbst, R.; Stallforth, P., Bacterial Alkaloid Biosynthesis: Structural Diversity via a Minimalistic Nonribosomal Peptide Synthetase. *Cell Chem. Biol.* 2018, *25* (6), 659-665.e9.
174. Yin, J.; Lin, A. J.; Golan, D. E.; Walsh, C. T., Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nat. Protoc.* 2006, *1* (1), 280-5.

175. Zorlu, Y.; Kumru, U.; İşci, Ü.; Divrik, B.; Jeanneau, E.; Albrieux, F.; Dede, Y.; Ahsen, V.; Dumoulin, F., 1,4,8,11,15,18,22,25-Alkylsulfanyl phthalocyanines: effect of macrocycle distortion on spectroscopic and packing properties. *Chem. Comm.* 2015, *51* (30), 6580-6583.
176. Torres Pazmiño, D. E.; Dudek, H. M.; Fraaije, M. W., Baeyer-Villiger monooxygenases: recent advances and future challenges. *Curr. Opin. Chem. Biol.* 2010, *14* (2), 138-44.
177. Malito, E.; Alfieri, A.; Fraaije, M. W.; Mattevi, A., Crystal structure of a Baeyer-Villiger monooxygenase. *Proc. Natl. Acad. Sci. U. S. A.* 2004, *101* (36), 13157-62.
178. Berg, J.; Tymoczko, J.; Stryer, L., *Biochemistry*. Freeman: New York, 2002; Vol. 5.
179. Fürst, M. J. L. J.; Gran-Scheuch, A.; Aalbers, F. S.; Fraaije, M. W., Baeyer–Villiger Monooxygenases: Tunable Oxidative Biocatalysts. *ACS Catal.* 2019, *9* (12), 11207-11241.
180. Yachnin, B. J.; Sprules, T.; McEvoy, M. B.; Lau, P. C. K.; Berghuis, A. M., The Substrate-Bound Crystal Structure of a Baeyer–Villiger Monooxygenase Exhibits a Criegee-like Conformation. *J. Am. Chem.Soc.* 2012, *134* (18), 7788-7795.
181. Hu, J.-F.; Wunderlich, D.; Thiericke, R.; Dahse, H.-M.; Grabley, S.; Feng, X.-Z.; Sattler, I., Jenamidines A to C: Unusual Alkaloids from Streptomyces sp. with Specific Antiproliferative Properties Obtained by Chemical Screening. *ChemInform* 2004, *35* (10).
182. Romero, E.; Gómez Castellanos, J. R.; Gadda, G.; Fraaije, M. W.; Mattevi, A., Same Substrate, Many Reactions: Oxygen Activation in Flavoenzymes. *Chem.Rev.* 2018, *118* (4), 1742-1769.
183. Schaefers, F. Herstellung und funktionelle Charakterisierung von neuen Biokatalysatoren für die Verwendung in der Naturstoffsynthese. Technical University of Munich, Garching, Germany, 2016.
184. Blaser, G.; Sanderson, J. M.; Batsanov, A. S.; Howard, J. A. K., The facile synthesis of a series of tryptophan derivatives. *Tetrahedron Lett.* 2008, *49* (17), 2795-2798.
185. Grevengoed, T. J.; Klett, E. L.; Coleman, R. A., Acyl-CoA metabolism and partitioning. *Annu. Rev. Nutr.* 2014, *34*, 1-30.
186. Kawaguchi, A.; Yoshimura, T.; Okuda, S., A new method for the preparation of acyl-CoA thioesters. *J. Biochem.* 1981, *89* (2), 337-9.
187. Oikawa, Y.; Sugano, K.; Yonemitsu, O., Meldrum's acid in organic synthesis. 2. A general and versatile synthesis of .beta.-keto esters. *J. Org. Chem.* 1978, *43* (10), 2087-2088.

188. Soehano, I.; Yang, L.; Ding, F.; Sun, H.; Low, Z. J.; Liu, X.; Liang, Z.-X., Insights into the programmed ketoreduction of partially reducing polyketide synthases: stereo- and substrate-specificity of the ketoreductase domain. *Org. Biomolec. Chem.* 2014, *12* (42), 8542-8549.
189. Yamamoto, Y.; Watanabe, Y.; Ohnishi, S., 1, 3-Oxazines and Related Compounds. XIII. Reaction of Acyl Meldrum's Acids with Schiff Bases Giving 2, 3-Disubstituted 5-Acy1-3, 4, 5, 6-tetrahydro-2<I>H</I>-1, 3-oxazine-4, 6-diones and 2, 3, 6-Trisubstituted 2, 3-Dihydro-1, 3-oxazin-4-ones. *Chem. Pharm. Bull.* 1987, *35* (5), 1860-1870.
190. Gilles, V.; Vieira, M. A.; Lacerda Jr., V.; Castro, E. V. R.; Santos, R. B.; Orestes, E.; Carneiro, J. W. M.; Greco, S. J., A New, Simple and Efficient Method of Steglich Esterification of Juglone with Long-Chain Fatty Acids: Synthesis of a New Class of Non-Polymeric Wax Deposition Inhibitors for Crude Oil. *J. Braz. Chem. Soc.* 2015, *26*, 74-83.
191. Chhabra, S. R.; Harty, C.; Hooi, D. S.; Daykin, M.; Williams, P.; Telford, G.; Pritchard, D. I.; Bycroft, B. W., Synthetic analogues of the bacterial signal (quorum sensing) molecule N-(3-oxododecanoyl)-L-homoserine lactone as immune modulators. *J. Med. Chem.* 2003, *46* (1), 97-104.
192. Ruysbergh, E.; Stevens, C. V.; De Kimpe, N.; Mangelinckx, S., Synthesis and analysis of stable isotope-labelled N-acyl homoserine lactones. *RSC Advances* 2016, *6* (77), 73717-73730.
193. Janikowska, K.; Rachon, J.; Makowiec, S., Acyl Meldrum's acid derivatives: Application in organic synthesis. *Russ. Chem. Rev.* 2014, *83*, 620.
194. Liu, Q.; Yao, F.; Chooi, Y. H.; Kang, Q.; Xu, W.; Li, Y.; Shao, Y.; Shi, Y.; Deng, Z.; Tang, Y.; You, D., Elucidation of Piericidin A1 biosynthetic locus revealed a thioesterase-dependent mechanism of α-pyridone ring formation. *Chem. Biol.* 2012, *19* (2), 243-53.
195. Cochrane, R. V. K.; Sanichar, R.; Lambkin, G. R.; Reiz, B.; Xu, W.; Tang, Y.; Vederas, J. C., Production of New Cladosporin Analogues by Reconstitution of the Polyketide Synthases Responsible for the Biosynthesis of this Antimalarial Agent. *Angew. Chem. Int. Ed.* 2016, *55* (2), 664-668.
196. Kouprina, N.; Larionov, V., Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. *Chromosoma* 2016, *125* (4), 621-632.
197. Shizuya, H.; Birren, B.; Kim, U. J.; Mancino, V.; Slepak, T.; Tachiiri, Y.; Simon, M., Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc. Natl. Acad. Sci. U. S. A.* 1992, *89* (18), 8794-8797.

198. Tomás-Barberán, F. A.; Blázquez, M. A.; Garcia-Viguera, C.; Ferreres, F.; Tomás-Lorente, F., A comparative study of different amberlite XAD resins in flavonoid analysis. *Phytochem. Anal.* 1992, *3* (4), 178-181.
199. Pfeifer, B.; Admiraal, S.; Gramajo, H.; Cane, D.; Khosla, C., Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of E. coli. *Science (New York, N.Y.)* 2001, *291*, 1790-2.
200. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic local alignment search tool. *J. Mol. Biol.* 1990, *215* (3), 403-10.
201. Medema, M. H.; Blin, K.; Cimermancic, P.; de Jager, V.; Zakrzewski, P.; Fischbach, M. A.; Weber, T.; Takano, E.; Breitling, R., antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* 2011, *39* (Web Server issue), W339-46.
202. Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Müller, R.; Wohlleben, W.; Breitling, R.; Takano, E.; Medema, M. H., antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* 2015, *43* (W1), W237-43.
203. Blin, K.; Wolf, T.; Chevrette, M. G.; Lu, X.; Schwalen, C. J.; Kautsar, S. A.; Suarez Duran, H. G.; de Los Santos, E. L. C.; Kim, H. U.; Nave, M.; Dickschat, J. S.; Mitchell, D. A.; Shelest, E.; Breitling, R.; Takano, E.; Lee, S. Y.; Weber, T.; Medema, M. H., antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* 2017, *45* (W1), W36-w41.
204. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; Thierer, T.; Ashton, B.; Meintjes, P.; Drummond, A., Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012, *28* (12), 1647-1649.

7. Supplements

Table S1. Compounds of the pyrrolizixenamide pathway: Identified masses during *in vivo* heterologous expression experiments.

Compound #	Name	Chemical formula	Calculated m/z	Measured m/z of $[M+H]^+$	Exact mass [g/mol]	Expressed by cluster #
50	Pyrrolizixenamide A	$C_{13}H_{20}N_2O_2$	236.32	237.1, 237.9	236.15	PTC12 SMg1C5
51	Pyrrolizixenamide B	$C_{14}H_{22}N_2O_2$	250.34	251.2, 251.3	250.17	PTC12
52	Pyrrolizixenamide C	$C_{15}H_{24}N_2O_2$	264.37	265.0, 265.1, 265.2 265.00263	264.18	PTC12 SclavuC39 MiC21
53	Pyrrolizixenamide D	$C_{16}H_{26}N_2O_2$	278.40	279.2, 279.1 279.03333	278.30	PTC12 PAO1C29s
54	Pyrrolizixenamide E	$C_{17}H_{28}N_2O_2$	292.42	293.1, 293.8 293.04690	292.22	PTC12 MiC21
55	Pyrrolizixenamide F	$C_{18}H_{30}N_2O_2$	306.45	307.0, 307.3	306.23	PTC12 PAO1C29s
56	Pyrrolizixenamide G	$C_{19}H_{32}N_2O_2$	320.48	321.3	320.25	Mix & Match
58	Pyrrolizixenamide I	$C_{21}H_{36}N_2O_2$	348.53	349.2, 349.4, 349.5, 349.13612	348.28	Mix & Match
63	N-(5,8-dioxo-1,2,3,5,8,8a-hexahydroindolizin-6-yl) hexanamide	$C_{14}H_{20}N_2O_3$	264.33	264.9, 265.1	264.15	XSC19
64	N-(5,8-dioxo-1,2,3,5,8,8a-hexahydroindolizin-6-yl) heptanamide	$C_{15}H_{22}N_2O_3$	278.35	278.8	278.16	XSC19
65	N-(5,8-dioxo-1,2,3,5,8,8a-hexahydroindolizin-6-yl) octanamide	$C_{16}H_{24}N_2O_3$	292.38	293.0	292.18	XSC19
66	N-(5,8-dioxo-1,2,3,5,8,8a-hexahydroindolizin-6-yl) decanamide	$C_{18}H_{28}N_2O_3$	320.43	320.6	320.21	XCS19
67	N-(5,8-dioxo-1,2,3,5,8,8a-hexahydroindolizin-6-yl) dodecanamide	$C_{20}H_{32}N_2O_3$	348.49	349.4	348.24	XCS19

Table S2. Unique molecular ion masses of unknown compounds detected in extracts of selected BGCs after *in vivo* heterologous expression within size range of PA derivatives (200 – 400 g/mol).

Expression by cluster #	m/z similar to 50	51	52	other
SclavuC39	236.3	247.2	-	287.9
SMg1C5		258.0, 264.0		324.8
SFC2	238.1	249.1	258.2	274.0
PAO1C29s		247.1, 249.1, 255.0	262.2, 266.2	283.3, 284.2, 297.2, 336.3, 364.2
MiC21			263.3	337.2, 367.2 (404.2, 564.4)

Table S3. Overview of generated pyreudiones from chemically synthesized substrates *in vitro*.

Substrate	Name	m/z	Product	Name	m/z predicted	m/z measured $[M+H]^+$
77	ME-190	278.13	**84**	Pyreudione A	265.17	266.3342, 265.35195
78	ME-097	287.16	**84**	Pyreudione A	265.17	265.6217, 265.77463
79	ME-087	258.17	**84**	Pyreudione A	265.17	-
80	ME-095	255.09	**87**	“Pyreudionine”	233.11	-
81	ME-099	255.09	**88**	“Pyreudiodiene”	233.11	234.71896, 234.21820
82	ME-103	231.09	**85**	Pyreudione E	209.11	210.04039, 210.09107
83	ME-133	129.08	**86**	Methyl-pyreudione A	279.18	280.632, 280.2

Table S4. Overview of biosynthetically obtained pyrrolizixenamide derivatives and brabantamide-like mimics *in vitro*.

Substrate	Name	m/z	Product	Name	m/z predicted	m/z measured $[M+H]^+$
43	N-(5,8-dioxo-1,2,3,5,8,8a-hexahydroindolizin-6-yl) hexanamide	264.33	**35**	Pyrrolizixenamide A	236.32	237.1, 237.04534, 237.09697, 237.05118
45	N-(5,8-dioxo-1,2,3,5,8,8a-hexahydroindolizin-6-yl) octanamide	292.38	**37**	Pyrrolizixenamide C	264.37	265.1, 265.09219
57a	-	476.27				367.02747
57b	-	485.29	**121a**		366.25	385.00497
57c	-	348.24	**69**		384.26	-
70	-	287.21				-

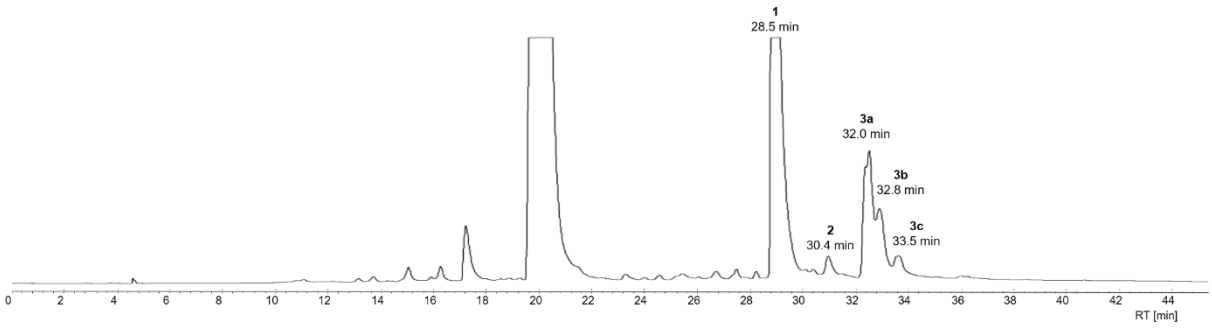


Figure S1. Preparative HPLC of *E. coli* BAP1 pCX2-PTC12 heterologous expression extract.

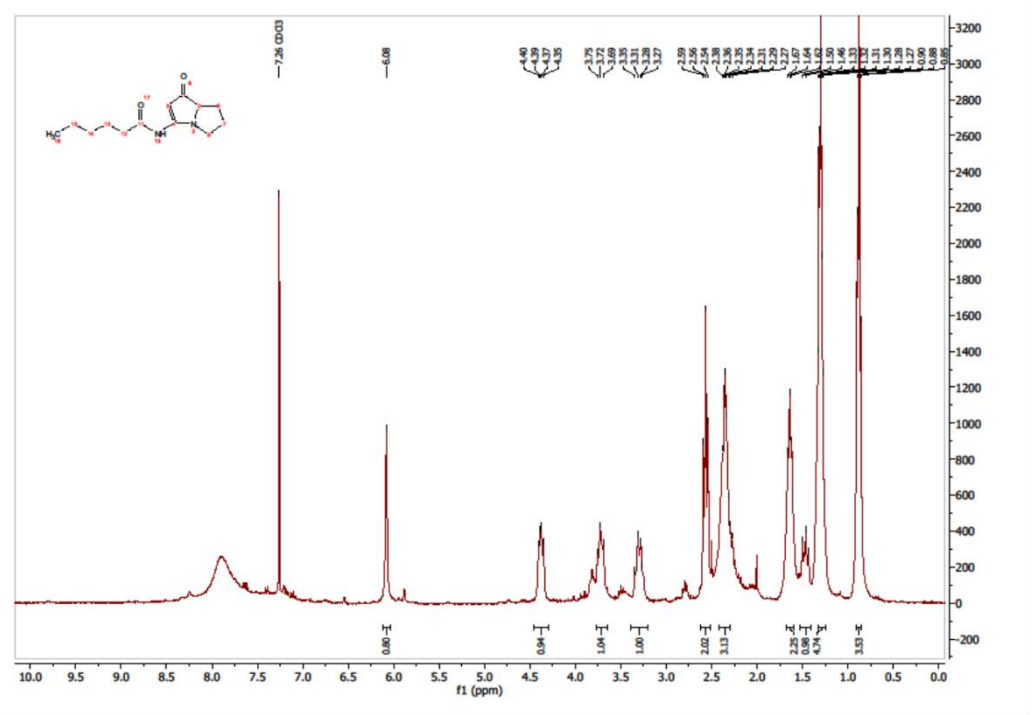


Figure S2. ^{1}H-NMR of pyrrolizixenamide product **50** isolated from *E. coli* BAP1 pCX2-PTC12.

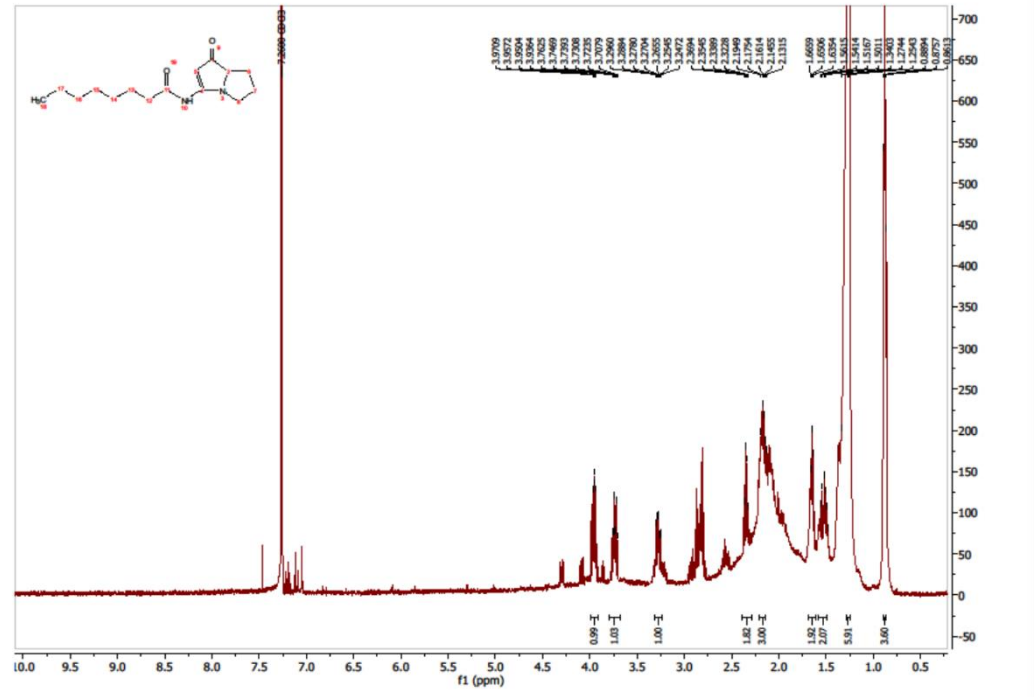


Figure S3. ^{1}H-NMR of pyrrolizixenamide product **51** isolated from *E. coli* BAP1 pCX2-PTC12.

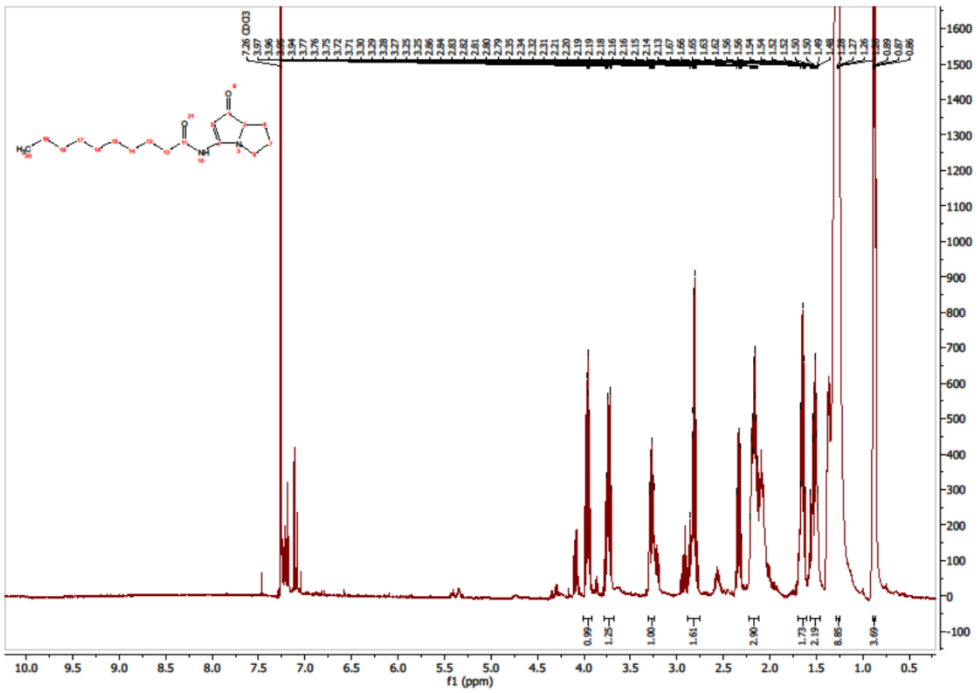


Figure S4. ^{1}H-NMR of pyrrolizixenamide product **52** isolated from *E. coli* BAP1 pCX2-PTC12.

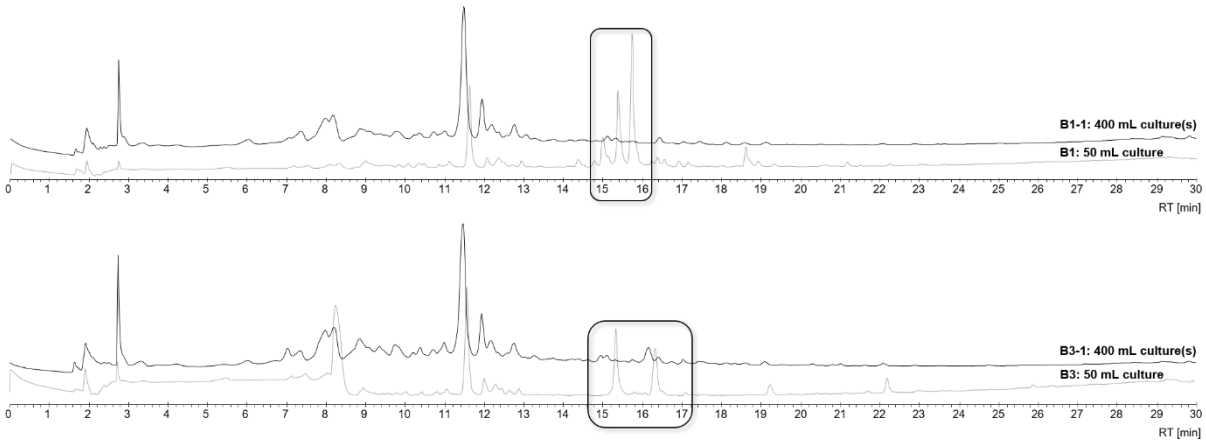


Figure S5. Comparison of small-scale *Streptomyces* heterologous expression attempts for M1154 SFC2-pSET152ermE*rev clones B1 and B3 with large-scale fermentation.

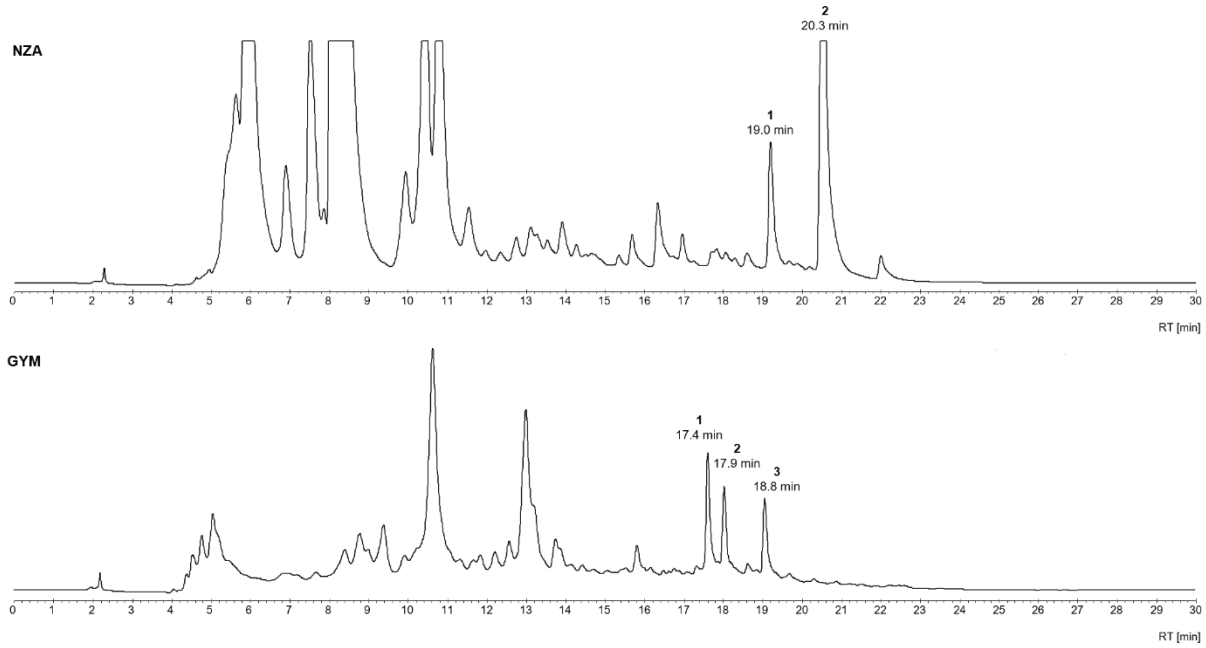


Figure S6. Preparative HPLC of extracts from native expression of *Micromonospora auratinigra* DSM44815 in different cultivation media.

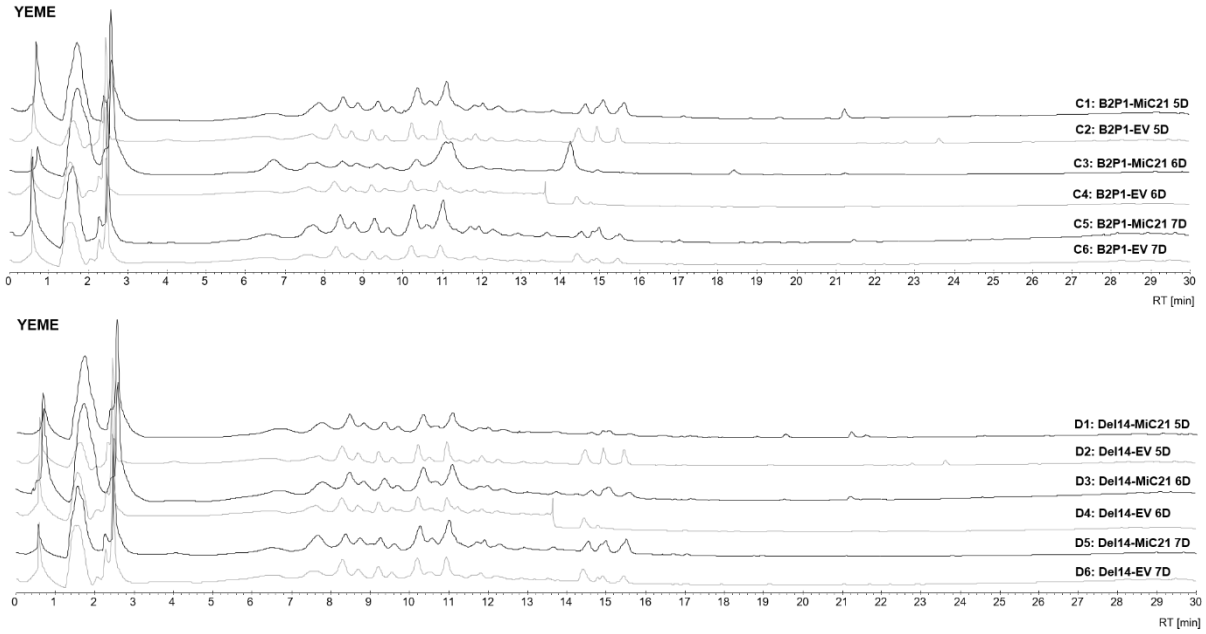


Figure S7. Heterologous expression of MiC21-pSET152ermE*rev in *Streptomyces albus* deletion strains B2P1 and Del14 in nutrient-rich cultivation medium YEME.

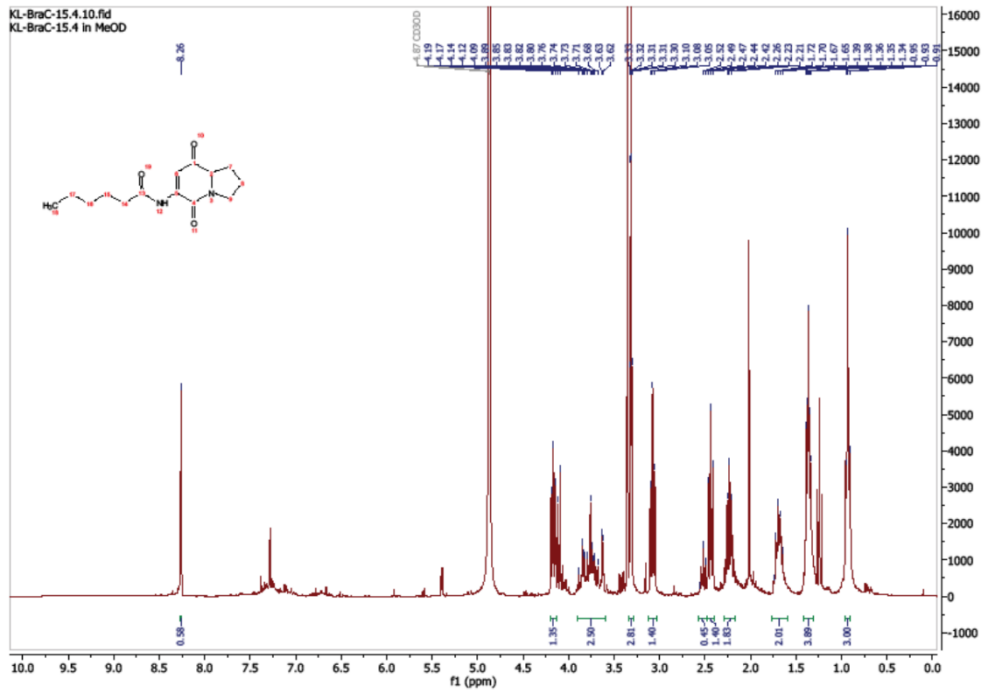


Figure S8. ^{1}H-NMR of pyrrolizixenamide pathway intermediate **63** isolated from mix & match expression construct pCX2-PxaA-BraC at t_R = 15.4 min.

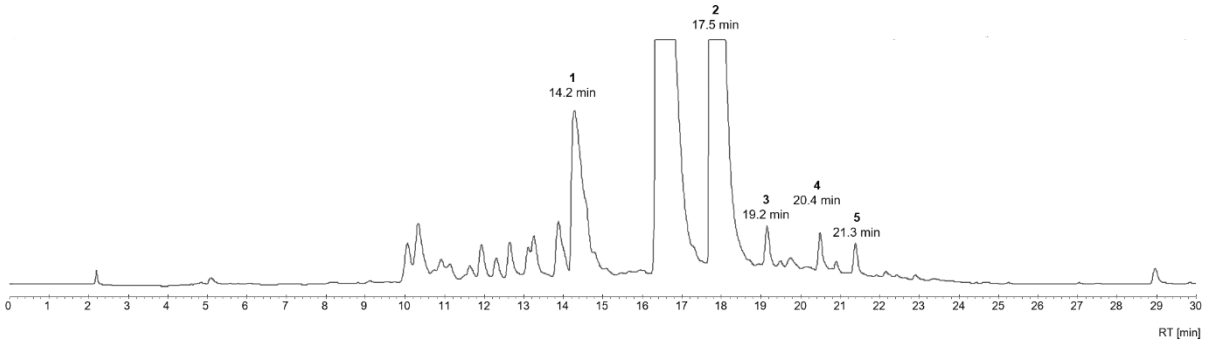


Figure S9. Preparative HPLC of *E. coli* BAP1-expressed mix & match construct pCX2-PxaA-MiC21 MOX-araBAD-MiC21back using a double promoter system for gene activation.

Table S5. Protein expression conditions applied to obtain highest protein yields possible.

Protein	Expression conditions
Bimodular NRPS systems	pHis8C(+)-TEV or pET28b(+)-SUMO
Final: max. 5 mg/L culture broth	2 bp insertion for in frame cloning
	TB medium
	Cultivation over night, 16 °C, 180 rpm shaking
	E. coli BL21DE3(+) or BAP1
	IPTG ≤ 0.1 mM
	NLS treatment to obtain soluble protein
	Lysate flowthrough incubated with Ni^{2+} beads twice to maximize total protein yield
Monomodular NRPS sytem	pET28a(+)-SUMO
Final: max. 10 mg/L culture broth	TB medium
	Cultivation over night, 16 °C, 180 rpm shaking
	E. coli BL21DE3(+) or BAP1
	IPTG: 0.1 mM
	Standard purification procedure
MOX systems	pET28b(+)-SUMO
Final: max. 25 mg/L culture broth	Cultivation over night, 16 °C, 180 rpm shaking
	E. coli BL21DE3(+) or BAP1
	IPTG: 0.1 mM
	Standard purification procedure

Figure S10. MS-MS analysis to confirm identity of PAO1C29 MOX enzyme. Nucleotide sequence coverage was 87 %.

Figure S11. MS-MS analysis validating identity of PxaB MOX enzyme. Sequence coverage was detected to be 86 %.

Figure S12. MS-MS analysis of FadD acyl-CoA ligase protein from *E. coli* BL21DE3.

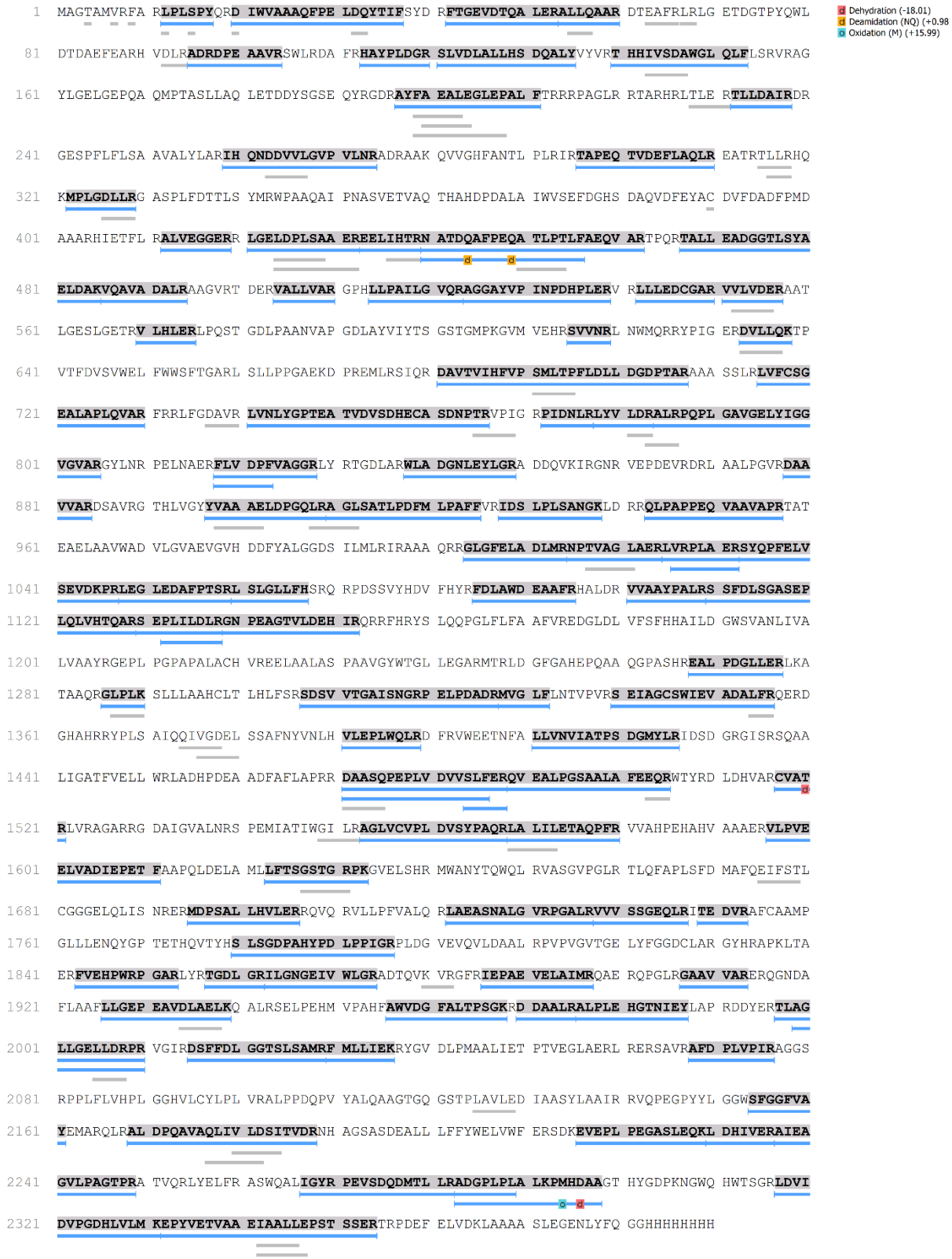


Figure S13. MS-MS analysis of bimodular NRPS system originating in *Pseudomonas aeruginosa* PAO1 C29 to confirm identity.

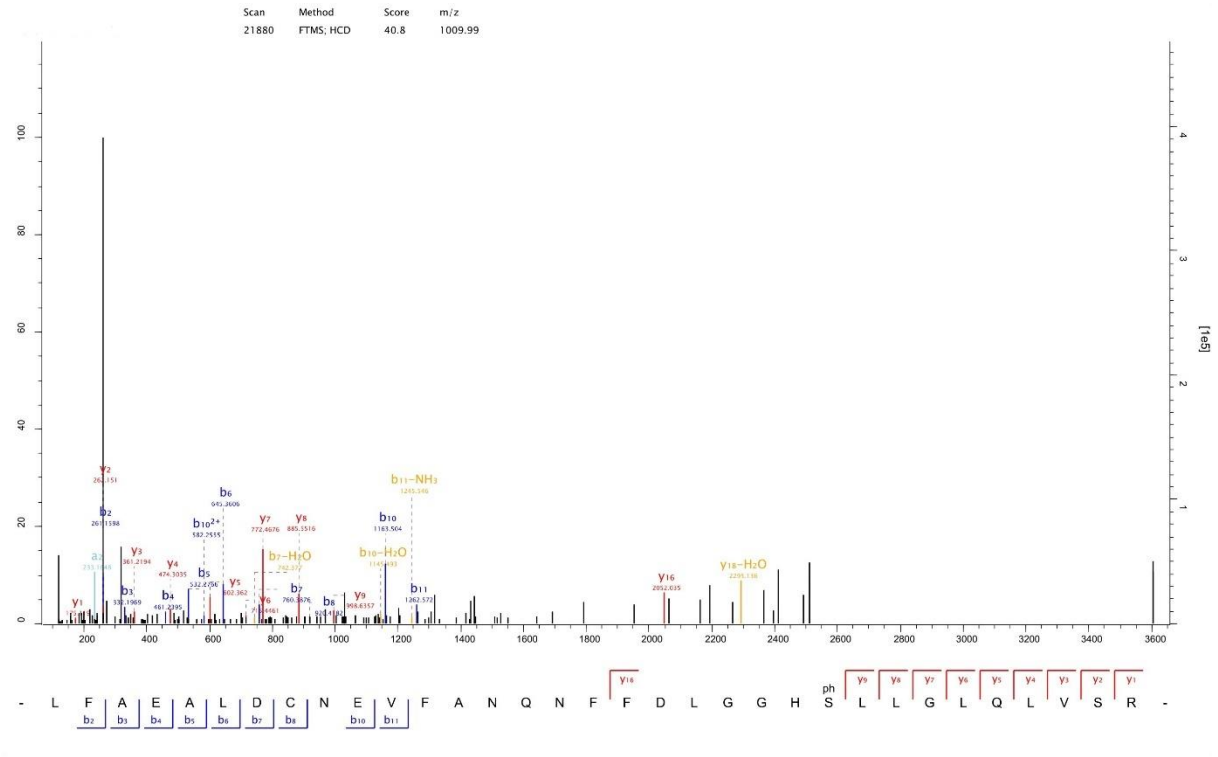


Figure S14. Tandem mass spectrometry for identification of active site serine within PCP domain of monomodular NRPS pys.

Figure S15. Tandem mass spectrometry for identity confirmation of monomodular NRPS pys involved in pyreudione biosynthesis. Nucleotide sequence coverage is 92 %.

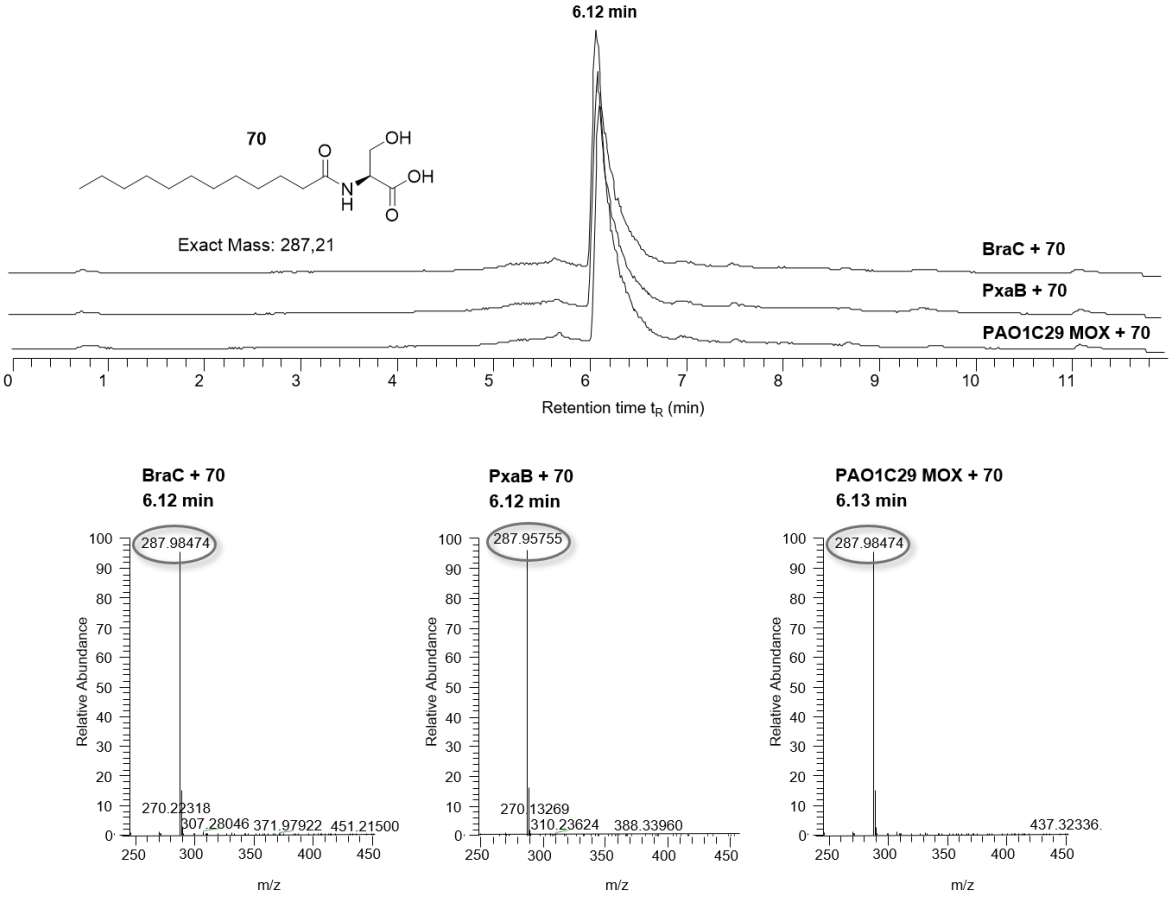


Figure S16. Conversion of precursor **70** to brabantamide derivative catalyzed by MOX homologues failed independent of enzymes' bacterial origin.

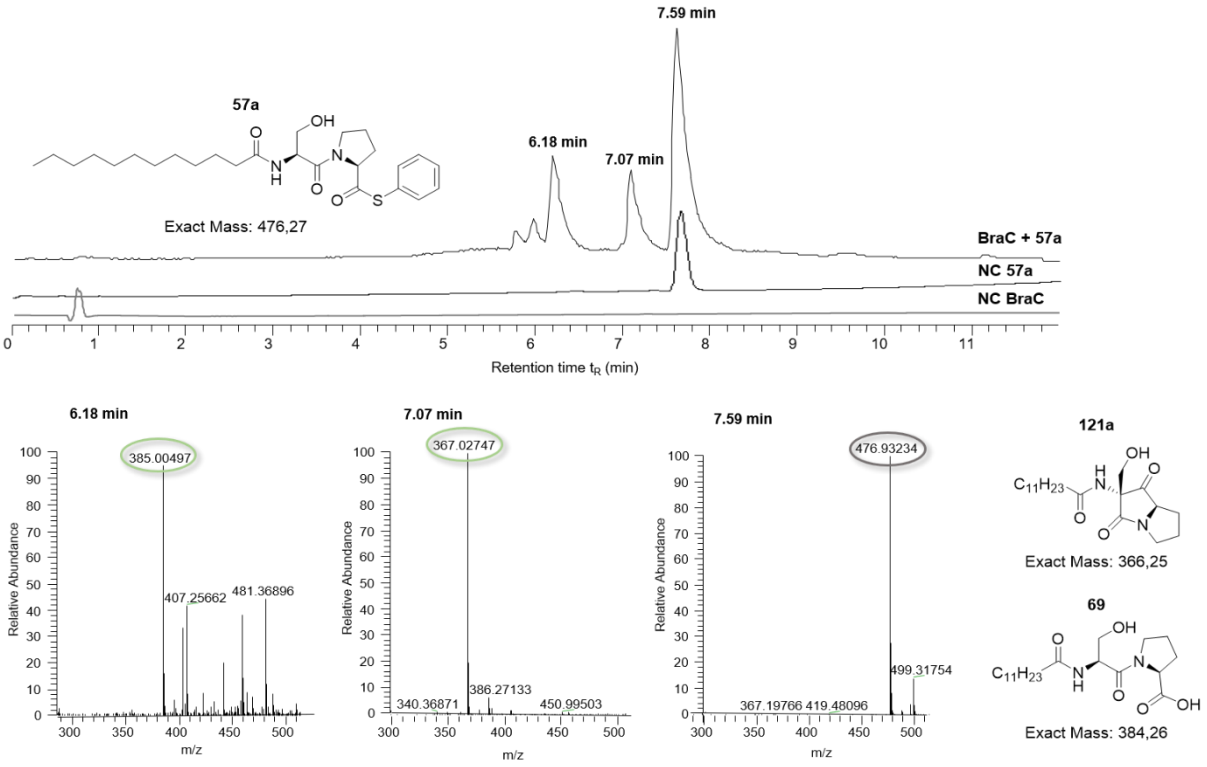


Figure S17. Successful conversion of simplified brabantamide precursor **57a** by BraC MOX.

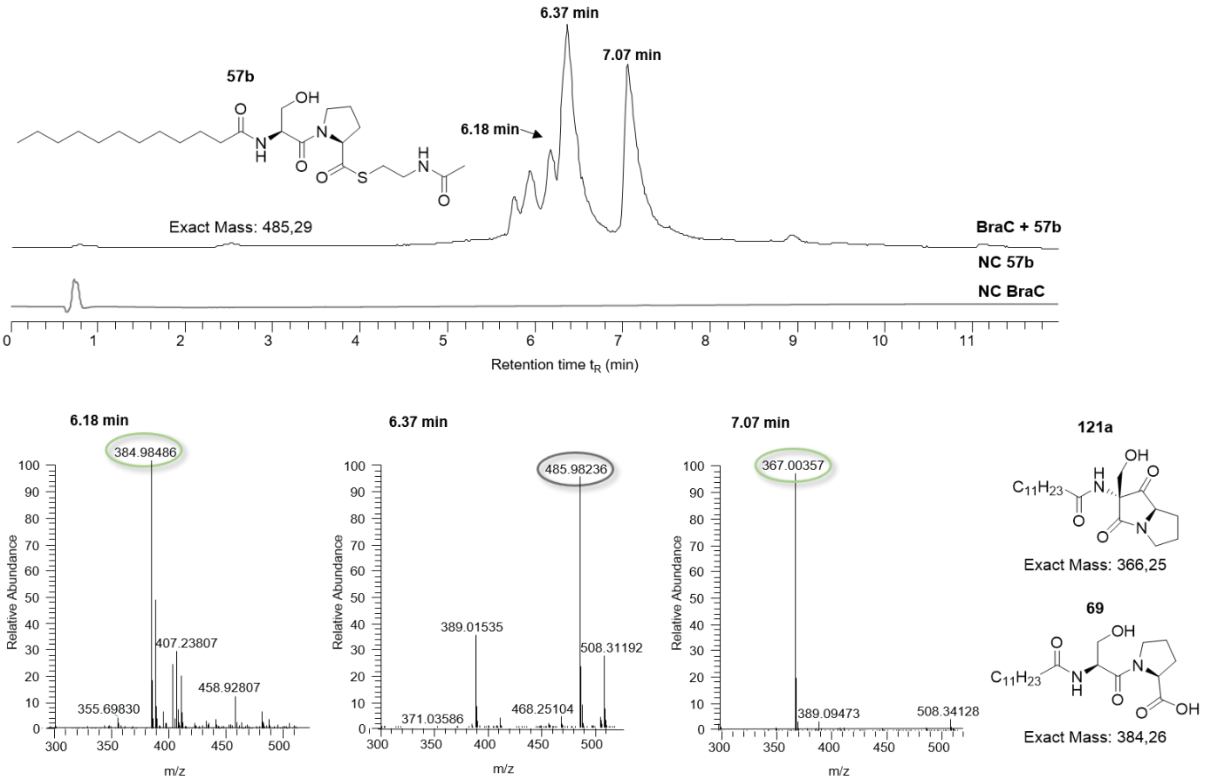
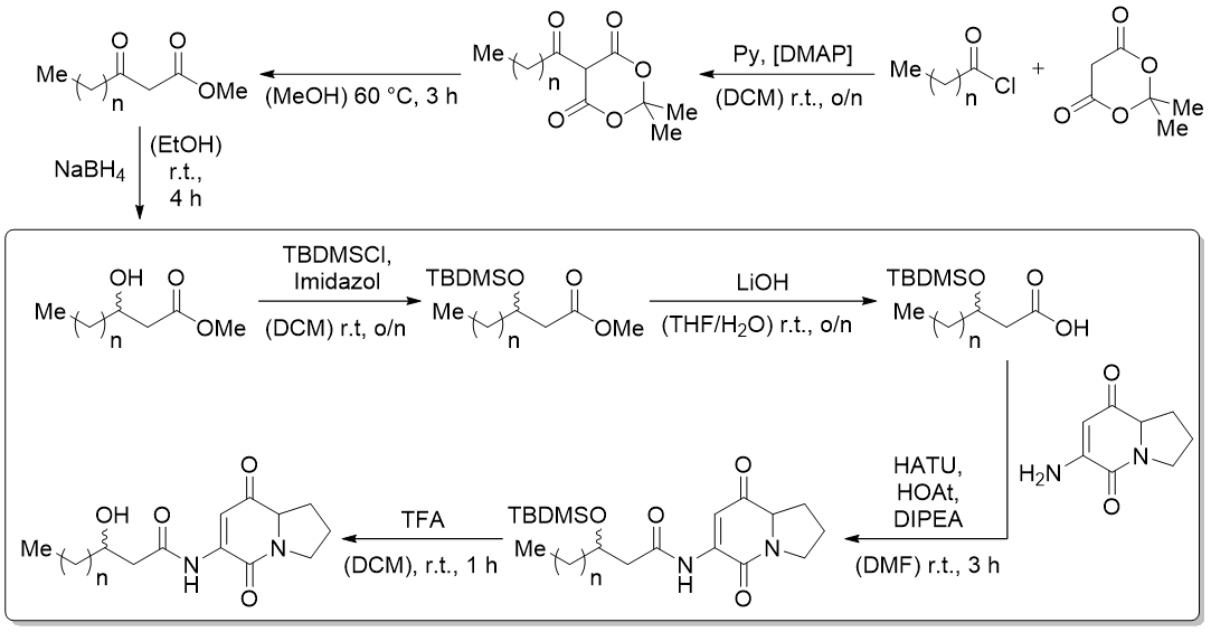


Figure S18. Successful conversion of simplified brabantamide precursor **57b** by BraC MOX.



Scheme S1. Chemical synthesis and attachment of modified acyl side chain moieties to deacetylated intermediate **63** for brabantamide intermediate generation.

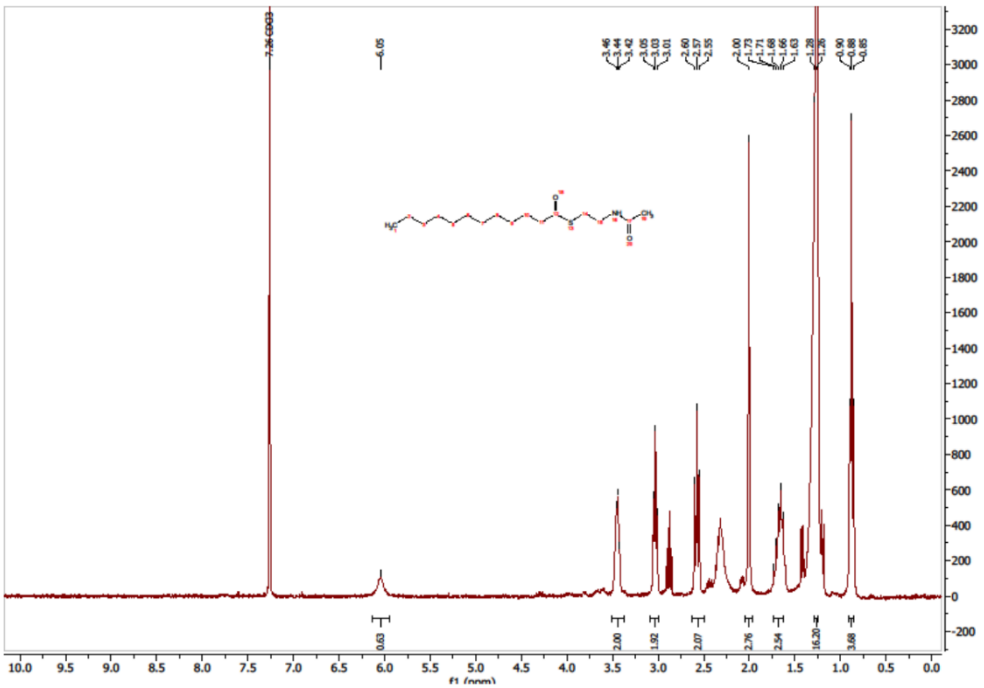


Figure S19. ^{1}H-NMR of purified SNAc thioester biomimetic **74**.

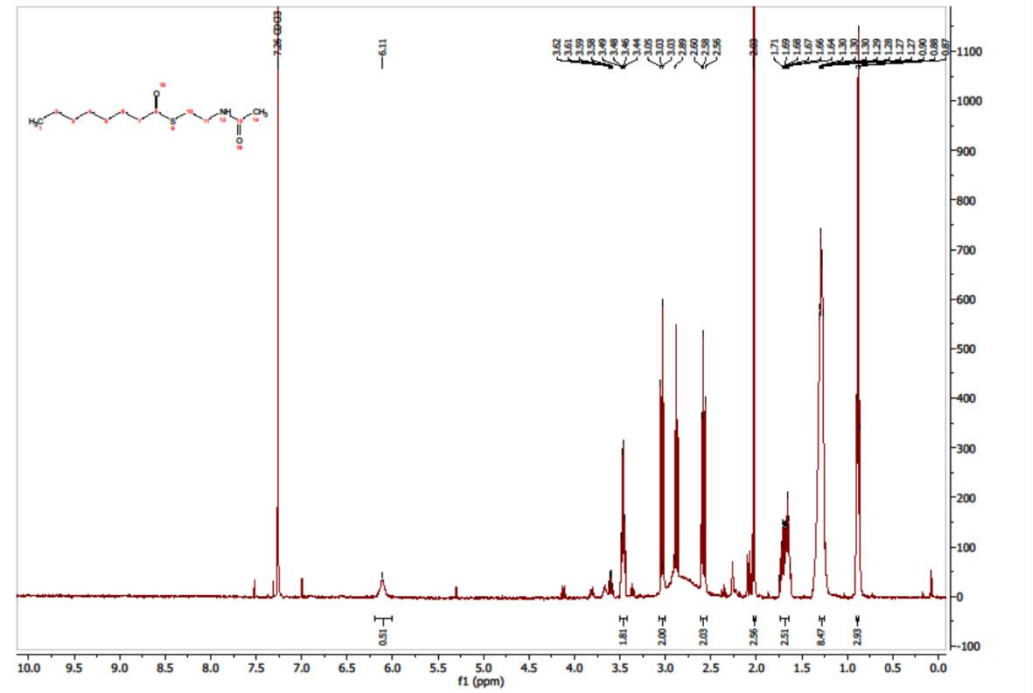
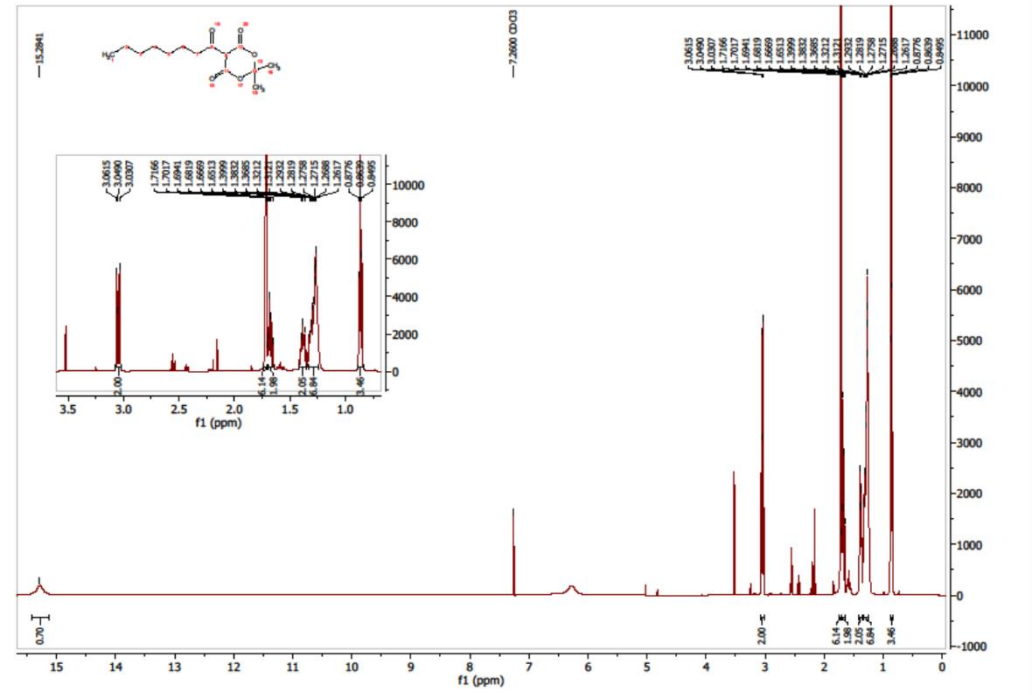
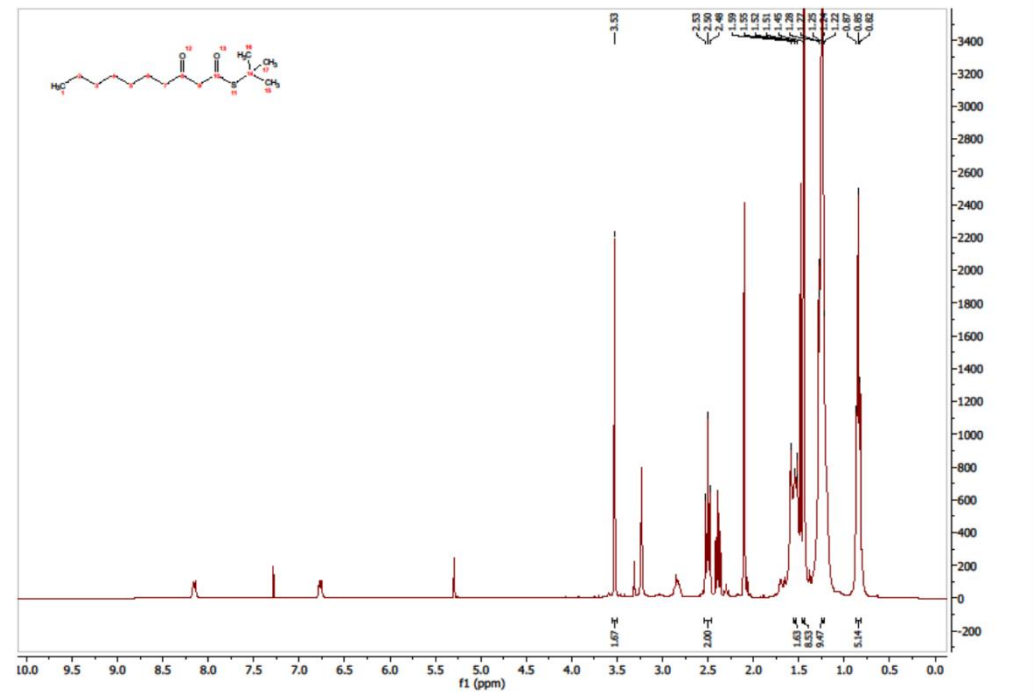


Figure S20. ^{1}H-NMR of purified SNAc thioester biomimetic **101**.



Figures S21. ^{1}H-NMR of purified octanoyl-Meldrum's acid **95**.



Figures S22. ^{1}H-NMR of purified acyl-*t*BuSH biomimetic **99**.

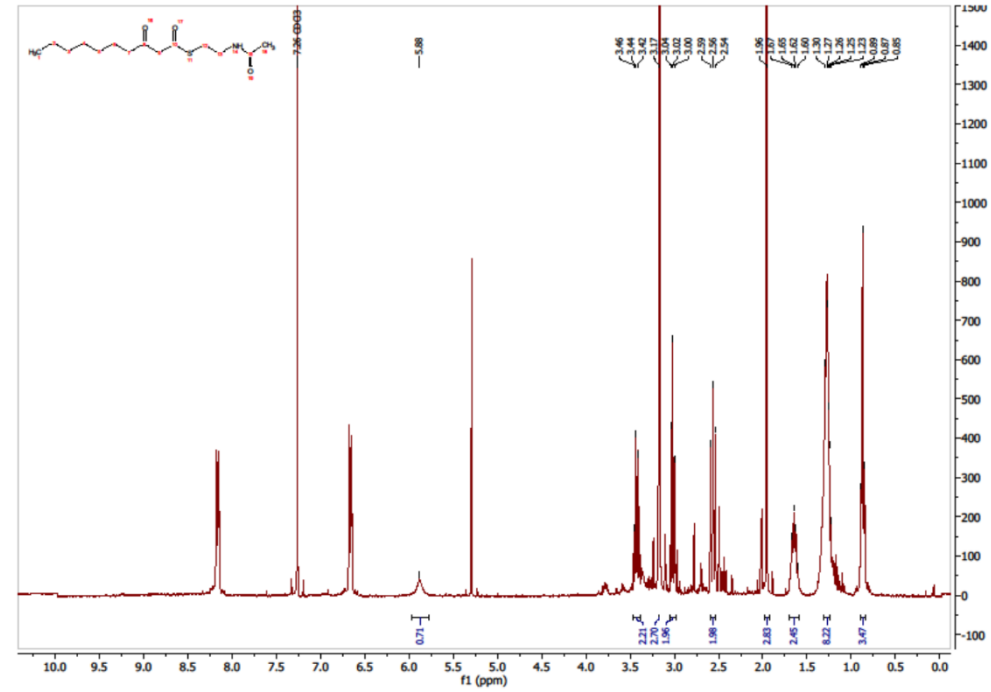


Figure S23. ^{1}H-NMR of purified β-ketoacyl thioester **97**.

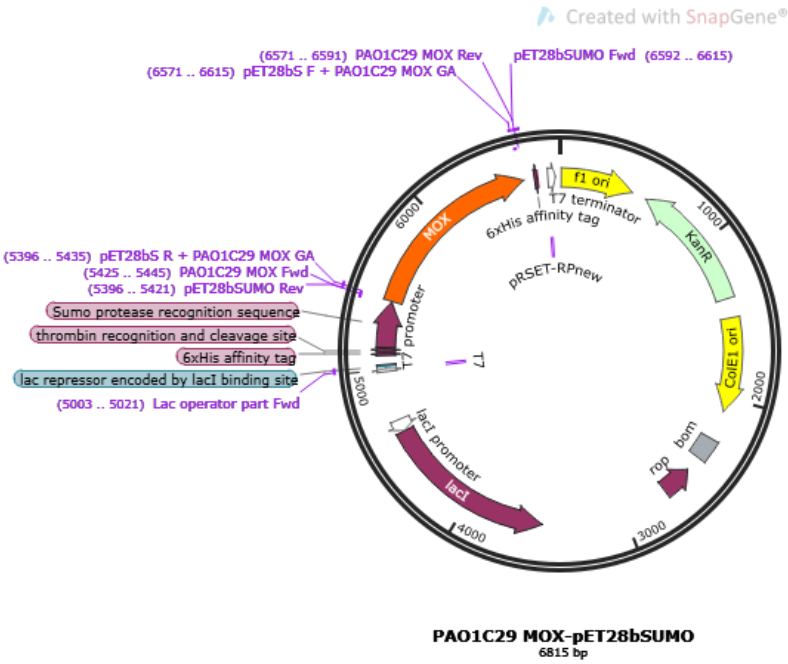


Figure S24. Representative plasmid map for MOX-containing recombinant protein expression construct. Shown: PAO1C29 MOX-pET28bSUMO.

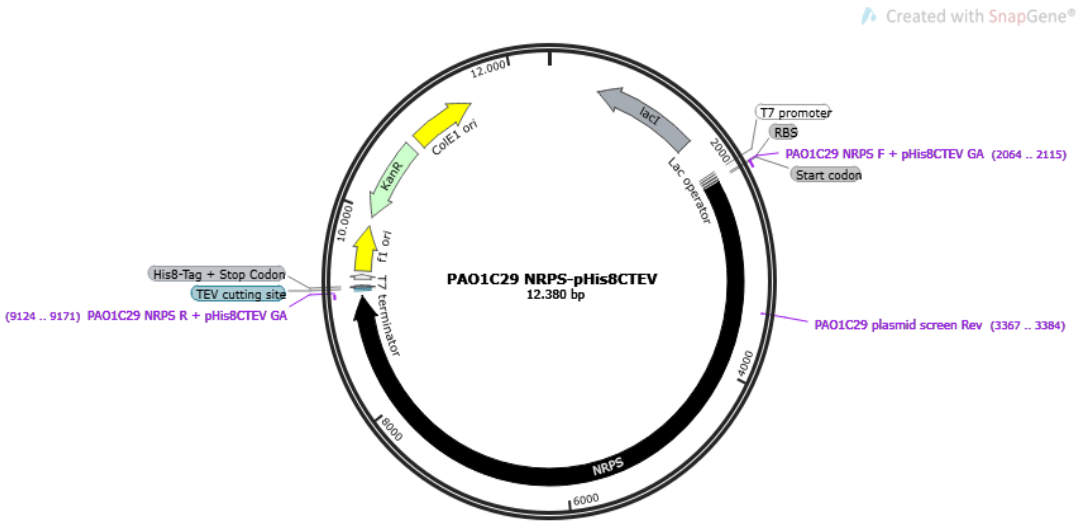


Figure S25. Representative plasmid map for NRPS-containing protein recombinant protein expression construct. Shown: PAO1C29 NRPS-pHis8C-TEV.

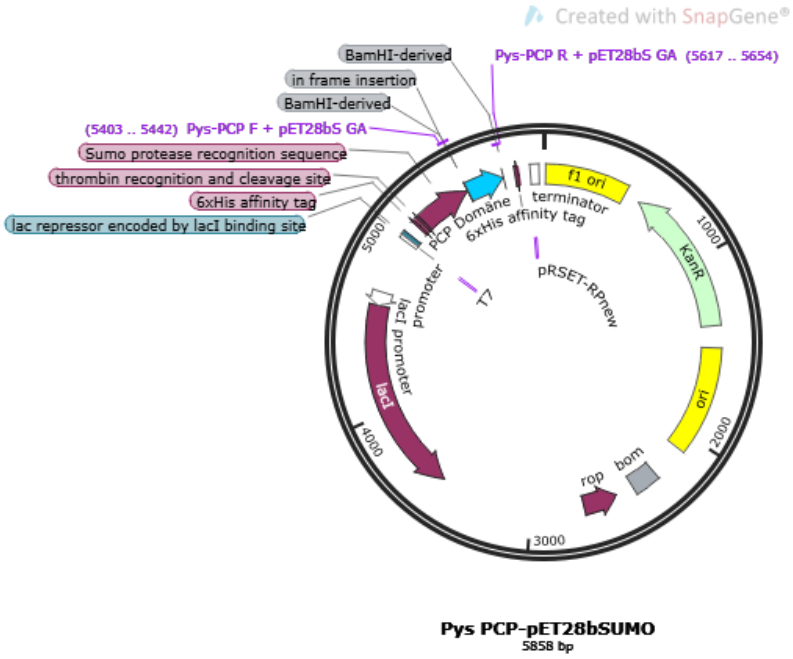


Figure S26. Cloning of pys PCP-domain for into pET28bSUMO protein expression vector for studies on domain activation.

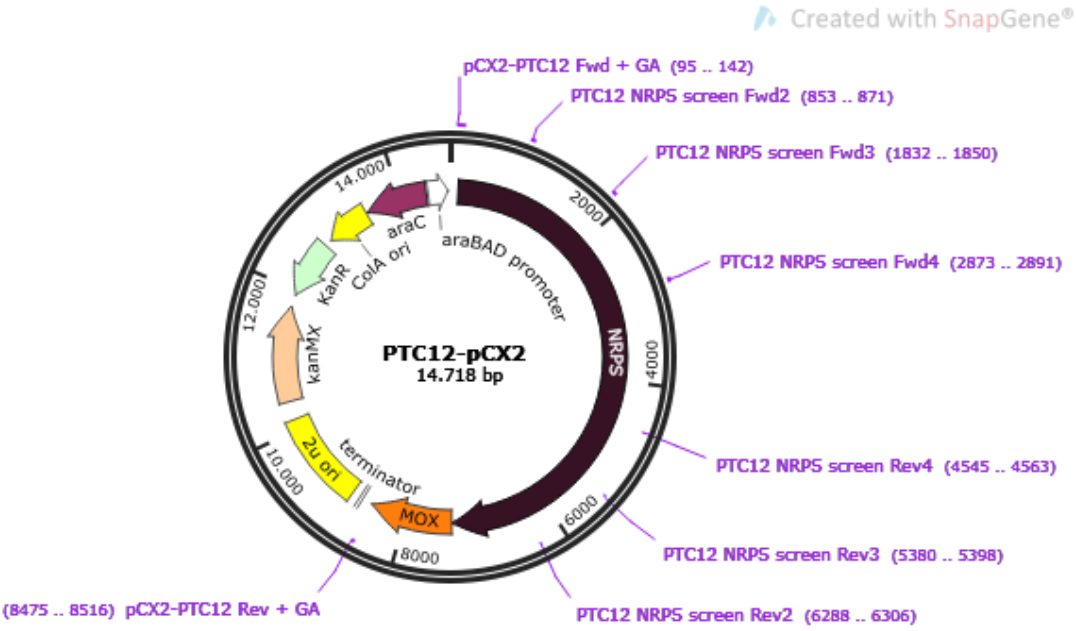


Figure S27. Representative plasmid map for DiPaC-assembled BGC in a single piece for heterologous expression in *E. coli*. Shown: PTC12-pCX2.

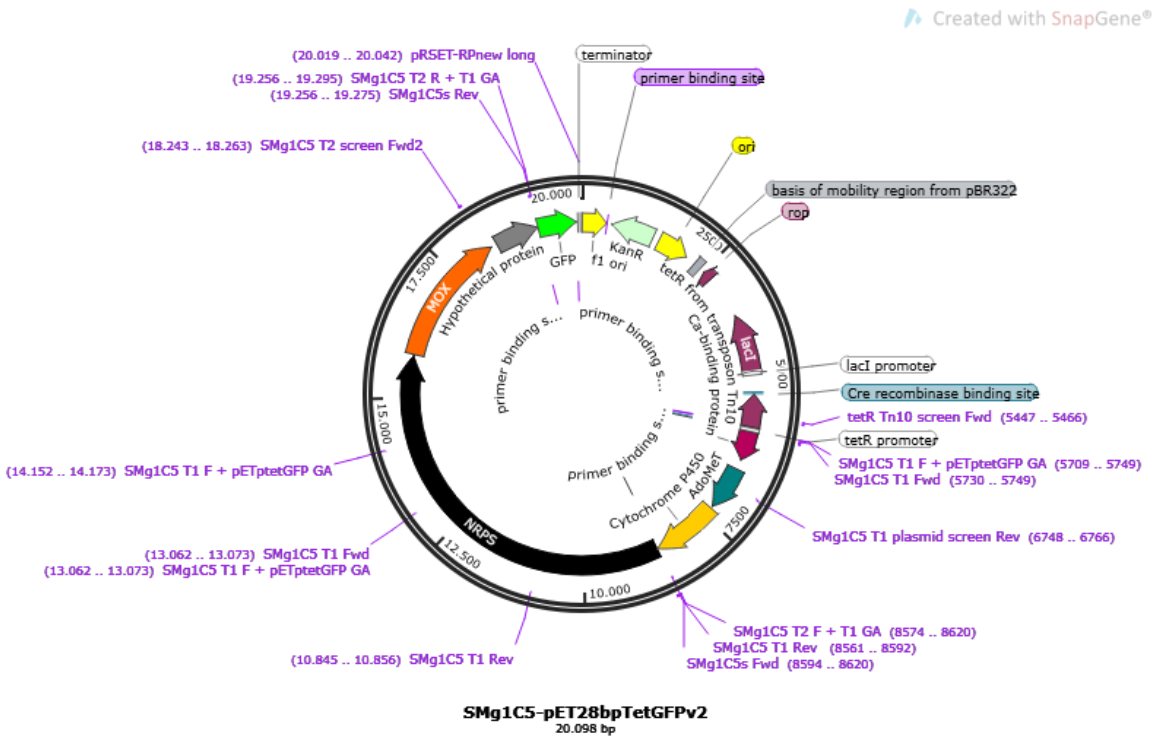


Figure S28. Representative plasmid map DiPaC-assembled BGC in multiple pieces for heterologous expression in *E. coli*. Shown: SMg1C5-pET28bpTetGFPv2.

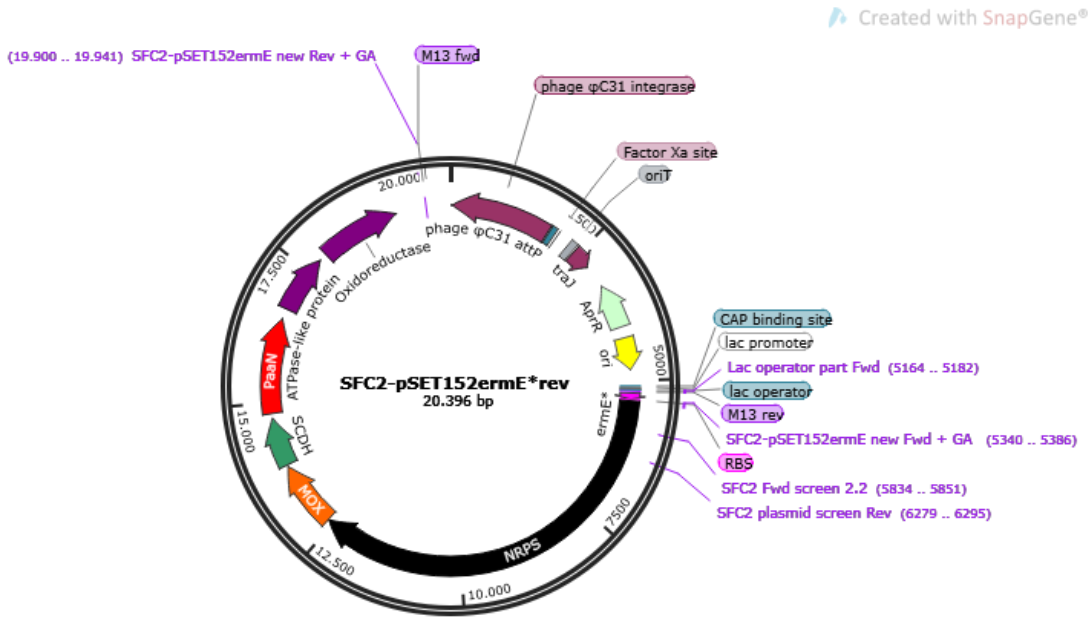


Figure S29. Representative plasmid map for construct assembled via LLHR/LCHR with full-length BGC for heterologous expression in *Streptomyces*. Shown: SFC2-pSET152ermE*rev.

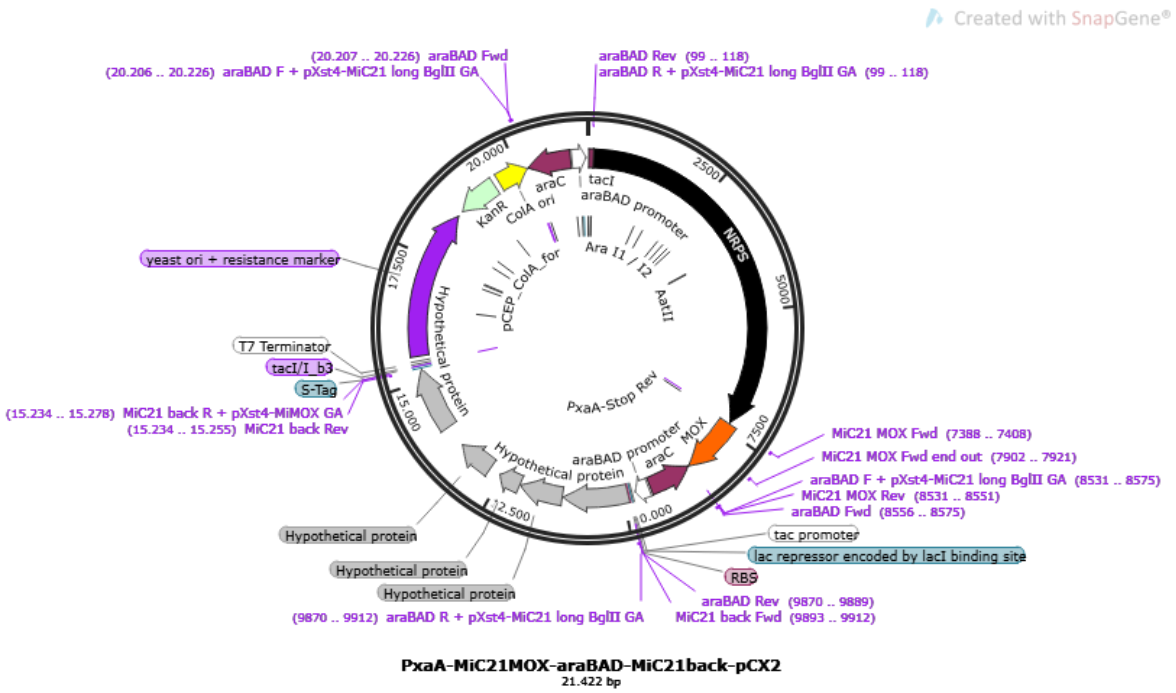


Figure S30. Representative plasmid map for combination of functional cluster parts via double promoter mix & match assembly to generate novel PA derivatives. Shiwn: PxaA-MiC21 MOx-araBAD-MiC21 back.

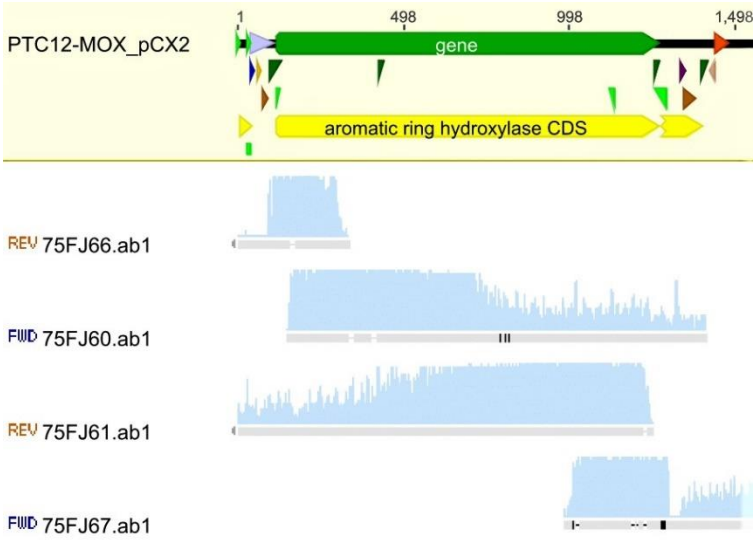


Figure S31. Full sequencing of PTC12 MOX gene representative for all BVMOs.

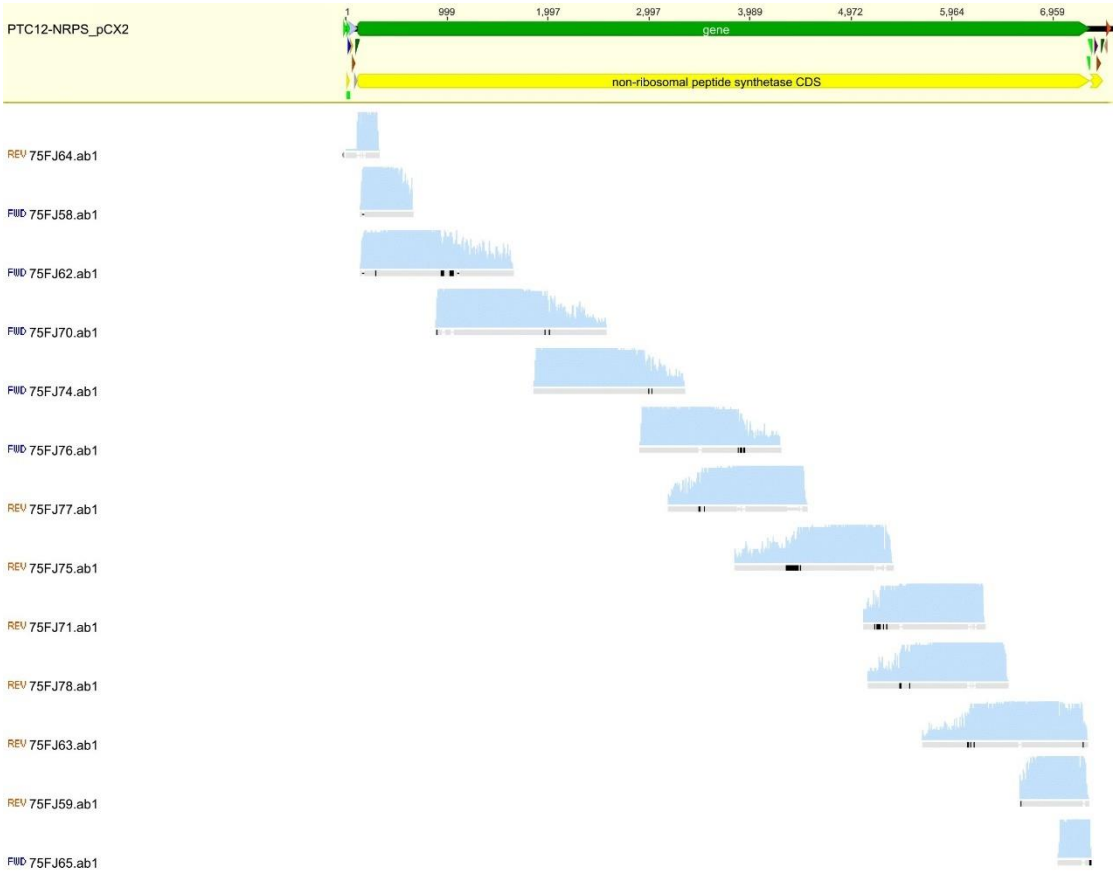


Figure S32. Full sequnecing of PTC12 NRPS gene representative for bimodular NRPS systems.

8. Acknowledgements

First, I would like to thank Prof. Dr. Tobias A.M. Gulder for giving me the opportunity to become part of his young and dedicated team, to grow with it and to develop into a valued member of the group. During the past years you have trusted me to work independently on my project and undergo a huge personal development, thereby always keeping an ear open for problems and challenges and if necessary, bringing me back on track. Every time I reached a dead end, you inspired me to keep going and broaden my horizon, all the time keeping up a healthy and casual working atmosphere. I really enjoyed the time in your group and all the off-work activities we did together!

A big thanks to the “original” TUM crew for welcoming me with open arms into the group, Mert, Paul, Christian and Elke, you made me feel comfortable from day one. Also, you gave me lots of valuable advice, showed me around and helped me whenever I needed it, I really appreciate it! Thanks for taking the time to read several reports and thesis despite having your own work to do! Paul, thanks for mentoring me and the great talks on football, Australia etc.

A special thanks goes to Anna G., my partner in crime. Endless train rides to work and back suddenly were not so endless and boring, standing in the cold and waiting was okay and before you knew it, I had to get off the train. Besides early morning singing sessions and discussions about bikes, mountain gear and running shoes, we were good at self-irony and “pöbeln”, I learned to bring that to perfection. I will always remember our winterly running rounds in the cold and dark :).

Coming to Dresden: To the Mensa crew Jonas, Mert, Max. At 11.30 a.m. all of you were ready to go, a thing to rely on, even during COVID-19. Politics, the stock market, cars – I even educated myself during the lunch break. Who needs a newspaper with you guys! Mert, thanks for bringing me to the Gulder Lab in the first place and the numerous hours of talks about God and the world, you have become a real friend during the years we have known each other. By the way, I have never driven through Munich using so many different routes and alternative ways to get to my destination :).

To my lab mates Manu and Tobi, thanks for bringing in a good mood almost every day, DJ Manu, nobody beats you when it comes to ending the day right! Tobi, we had interesting talks on mountain routes and you guys were great entertainment on the few weekends in Dresden. Hülya, thanks for tolerating the hundreds of times I came to you because of the HPLC or mass. I will definitely miss those cookies.

A big thanks also goes to Julia and Gesa, I am glad that the chemists and biologists got mixed up at TUD, because you two have become great friends during the past two years. Our cooking sessions, strawberry-picking adventures, SUP trips and Thursday evening hiking trips were one of a kind and made me enjoy my time in Dresden!

Thanks to Anke, Sabine and Monique for being so kind to take us all in and help with everything we needed!

Further thanks go to my friends from studying, Jackie, Isi and Tammie, it was nice to hear that we are all in the same situation, to get support and an open ear whenever needed.

Finally, a big thanks goes to my family for supporting me during my studies and my PhD, encouraging me to keep up the work, tolerating my endless talks on a topic in detail with words they have never heard before and always having my back. They always said: Do, what you want, but if you do it, do it right – a philosophy I have kept in mind during this entire time.

This work was supported by the Studienstiftung des deutschen Volkes with a three-year PhD scholarship.

9. Statement of authenticity

This thesis is submitted in fulfilment of the requirements to achieve the degree Dr. rer. nat. at the Technical University of Munich in Munich, Germany. The presented work in this thesis is, to the best of my knowledge, original except for sections that have been acknowledged in the text. I hereby declare that I have not previously submitted this material, either in whole or part, for a degree at this or any other institution. Unless otherwise indicated, all the data and observations presented are the results of my own work.

Signed:  (Katharina Lamm)

Date: May 5th, 2021