Studies on the biosynthesis and heterologous expression of complex secondary metabolites from streptomycetes

Dissertation

zur Erlangung des Grades des

Doktors der Naturwissenschaften (Dr.rer.nat.)

der Naturwissenschaftlich-Technischen Fakultät III

Chemie, Pharmazie, Bio- und Werkstoffwissenschaften

der Universität des Saarlandes

von

Tina Maria Binz

Saarbrücken

Februar 2009

Publications

Wenzel, S.C., Meiser, P., **Binz, T.M.**, Mahmud, T., Müller, R., 2006. Nonribosomal peptide biosynthesis: point mutations and module skipping lead to chemical diversity. *Angew Chem Int Ed Engl.; 45(14):2296-301.*

Binz, T.M., Wenzel, S.C., Schnell, H.-J., Bechthold, A., Müller, R., 2008. Heterologous expression and genetic engineering of the phenalinolactone biosynthetic gene cluster using Red/ET recombineering. *ChemBioChem; 15;9(3):447-445*

Conference contributions

Binz, T.M., Wenzel, S.C., Mahmud, T., Müller, R., October 2005. Comparison of the myxochromide A and S biosynthetic gene cluster from *Myxococcus xanthus* and *Stigmatella aurantiaca*; International workshop: Biology of Bacteria Producing Natural Products, Dresden, Germany (poster presentation)

Binz, T.M., Wenzel, S.C., Bechthold, A., Müller, R., September 2006. Heterologous expression of the Phenalinolactone biosynthetic gene cluster in different *Streptomyces* sp. via Red/ET recombineering; Chemical Biology: Directing Biosynthesis, University of Cambridge, UK (poster presentation)

Binz, T.M., Wenzel, S.C., Bechthold, A., Müller, R., October 2006. Heterologous expression of the Phenalinolactone biosynthetic gene cluster in different *Streptomyces* sp. via Red/ET recombineering; International workshop: Biology of Bacteria Producing Natural Products, Tübingen, Germany (oral presentation)

Binz, T.M., Wenzel, S.C., Bechthold, A., Müller, R., October 2007. Heterologous expression of the Phenalinolactone biosynthetic gene cluster in different *Streptomyces* sp. via Red/ET recombineering; European Bioperpectives, Cologne, Germany (poster presentation)

Binz, T.M., Wenzel, S.C., Bechthold, A., Müller, R., October 2007. Heterologous expression of the Phenalinolactone biosynthetic gene cluster in different *Streptomyces* sp. via Red/ET recombineering; International workshop: Biology of Bacteria Producing Natural Products, Saarbrücken, Germany (poster presentation)

Binz, T.M., Sosio,M., Donadio, S. and Müller, R., September 2008. Identification of the GE81112 biosynthetic gene cluster; International workshop: Biology and Chemistry of Antibiotic-Producing Bacteria, Berlin, Germany (oral presentation)

Acknowledgements

First of all, I would like to express my sincere gratitude to my doctoral advisor Prof. Dr Rolf Müller, for his guidance, advice, and encouragement during the course of my Ph.D. thesis. His constant confidence and persistence in face of research uncertainties motivated me to continue my work and inspired me much.

At the same time I would like to thank my co-advisor Dr. Silke Wenzel for the very pleasant and successful team work, her support, advice, and motivation the entire time.

I would also like to thank Prof. Dr. Manfred Schmitt for gladly consenting to review my thesis.

I kindly thank my cooperation partners Margherita Sosio and Stefano Donadio from KtedoGen and Hans-Jörg Schnell and Andreas Bechthold from Freiburg University for their assistance and the successful cooperation.

I greatly thank Dr. Kira Weissman not only for the excellent proof-reading of the thesis but also for her skillful remarks and helpful discussions which inspired me much.

I especially wish to thank Daniel Krug for his unremitting help and assistance with the analytics and for his extremely valuable answers to all my questions.

Of course (!) I also thank all members from our research institute Pharmazeutische Biotechnologie and especially all members of my office. I am grateful for all their encouragements, futile discussions, and jokes; I really enjoyed the friendly work atmosphere!

Last but not least my heartfelt thanks to my family and my friends for always believing in me and their joyful and constant support.

Table of contents

Figures

Tables

Abbreviations

Abbreviations

A domain	Adenylation domain
ADP	Adenosine diphopshate
AMP	Adenosine monophosphate
Amp	Ampicillin
Apra	Apramycin
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosome
bp	Base pair
BSA	Bovine serum albumine
BPC	Base peak chromatogram
C	Carbon
C domain	Condensation domain
Cm	Chloramphenicol
CoA	CoenzymeA
Cos	Cosmid
Da	Dalton
DIG	Digoxigenin
DMSO	Dimethyl sulfoxide
DNA	Desoxyribonucleic acid
dNTP	Desoxyribonucleosid triphosphate
DTT	1,4-Dithiothreitol
E domain	Epimerization domain
EDTA	Ethylenediamine tetraacetic acid
EIC	Extracted ion chromatogram
h	Hour
HPLC	High performance liquid chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kan	Kanamycin
kb	Kilo base pairs
kV	Kilovolt
l	Liter
LB	Luria Bertani
M	Molar
Mbp	Mega base pairs
Me	Methyl
Mg	Magnesium
mg	Milligram
min	Minutes
ml	Milliliter
mM	Millimolar
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
m/z	Mass-to-charge-ratio
N	Nitrogen
NRPS	Nonribosomal peptide synthetase
O	Oxygen

OD	Optical density
Orf	Open reading frame
oriT	Origin of transfer
P	Phosphor
PAGE	Polyacrylamid gelelectrophoresis
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
PKS	Polyketide synthase
Ppant	Phospopanthetein
PP_i	Pyrophosphate
RBS	Ribosome binding site
RNA	Ribonucleic acid
Rpm	Revolutions per minute
Rt	Retention time
RT	Room temperature
S	Sulfur
SAP	Shrimps alkaline phopsphatase
Sec	Seconds
SDS	Sodium dodecyl sulfate
sp.	Species
TE	Thioesterase
Tet	Tetracycline
U	Units
UV	Ultraviolet
V	Volt
v/v	Volume to volume
WT	Wild type
Zeo	Zeocin
°C	degree Celcius
μF	micro Farad
μ	micro
Ω	Ohm

Zusammenfassung

Die heterologe Expression komplexer Naturstoff-Biosynthesewege spielt eine immer größer werdende Rolle bei der Entdeckung und Modifikation von Wirkstoffen aus der Natur und gewinnt somit zunehmend an Bedeutung in der biotechnologischen Forschung. Diese Methode ermöglicht die Produktion komplexer Metabolite in leicht handhabbaren Wirtsorganismen und bildet so die Grundlage für die Entwicklung neuer Naturstoffderivate durch genetische Manipulation der Biosynthesewege. Die vorliegende Arbeit beschreibt eine innovative Strategie für die heterologe Expression des komplexen Phenalinolacton-Biosynthesewegs aus dem Streptomyceten Tü6071. Mit Hilfe der Red/ET Rekombination wurde dieses Gencluster, das auf 2 Cosmiden vorlag, kloniert und in verschiedenen Streptomyceten-Stämmen heterolog exprimiert. Das somit etablierte Expressionssystem bietet nun die Möglichkeit, durch gezielte genetische Manipulation der Biosynthesewege in *Escherichia coli* und anschließende heterologe Expression neue, möglicherweise wirksamere Naturstoffe und Naturstoffderivate zu entwickeln.

Der zweite Teil dieser Arbeit beschäftigt sich mit Untersuchungen zur Biosynthese einer ungewöhnlichen Klasse von Sekundärstoffen: den GE81112 Tetrapeptiden. Um die zugrundeliegende Biosynthese dieser Naturstoffe nun detailliert zu untersuchen, wurde eine Strategie für die Klonierung und Identifizierung des entsprechenden Genclusters entwickelt, das schließlich auf zwei überlappenden Cosmiden identifiziert wurde. Gleichzeitig konnten Methoden zur genetischen Manipulation des Stammes entwickelt werden, die es ermöglichten das Gencluster durch Geninaktivierungsexperimente zu identifizieren. Mit Hilfe der erhaltenen Ergebnisse konnte schließlich ein erstes Model für die GE81112 Biosynthese erarbeitet werden. Die in dieser Arbeit erhaltenen Einblicke in die Biosynthese können in Zukunft einen wichtigen Beitrag zur Entwicklung neuer GE81112 Derivate leisten.

Abstract

The heterologous expression of natural product biosynthetic pathways is of increasing interest in biotechnology and drug discovery. This approach enables the production of complex metabolites in more amenable host organisms and provides the basis for the generation of novel analogs through genetic engineering. In this thesis, we describe a straightforward strategy for the heterologous expression of the highly complex phenalinolactone biosynthetic pathway, which was recently cloned from *Streptomyces* sp. Tü6071. Using Red/ET recombineering, the phenalinolactone pathway was reconstituted from two cosmids and heterologously expressed in several *Streptomyces* strains. The established expression system now provides a convenient platform for functional investigations of the biosynthetic genes and the generation of novel analogs, by genetic engineering of the pathway in *Escherichia coli*.

The second part of this thesis describes work on a distinct class of secondary metabolites, the GE81112 tetrapeptide family. We developed a strategy for the cloning and identification of the GE81112 biosynthetic gene cluster, in order to investigate the biosynthetic pathway in detail. Generation of a cosmid library enabled us to identify the corresponding biosynthetic gene cluster on two overlapping cosmids. In parallel, we established methods to manipulate the strain genetically, allowing us to verify the identity of the GE81112 gene cluster by gene inactivation experiments. In addition, we characterized several proteins from the pathway using enzymatic assays *in vitro*. Taken together, these data have enabled us to propose a preliminary model for GE81112 biosynthesis. The results also open the door to developing new derivatives of these promising compounds by genetic engineering.

1   Introduction

1.1   Natural products as a robust source of new drugs and drug leads

Microorganisms produce a vast number of secondary metabolites with useful biological activities for both human and veterinary medicine, including compounds which function as antibiotics, immunosuppressive and antitumor agents, as well as herbizides, insecticides and antiparasitic agents. Figure 1.1 exemplifies a further characteristic feature of these so-called 'natural products': their significant structural diversity. The recruitment of many different building blocks, as well as the activity of diverse modifying enzymes, leads to the construction of highly complex chemical structures.

Bleomycin
Antitumor
(Streptomyces verticillus)

Vancomycin
antibacterial
(Amycolatopsis orientalis)

Gramicidin
Antibacterial
(Bacillus brevis)

Rapamycin
Immunosuppressive
(Streptomyces hygroscopicus)
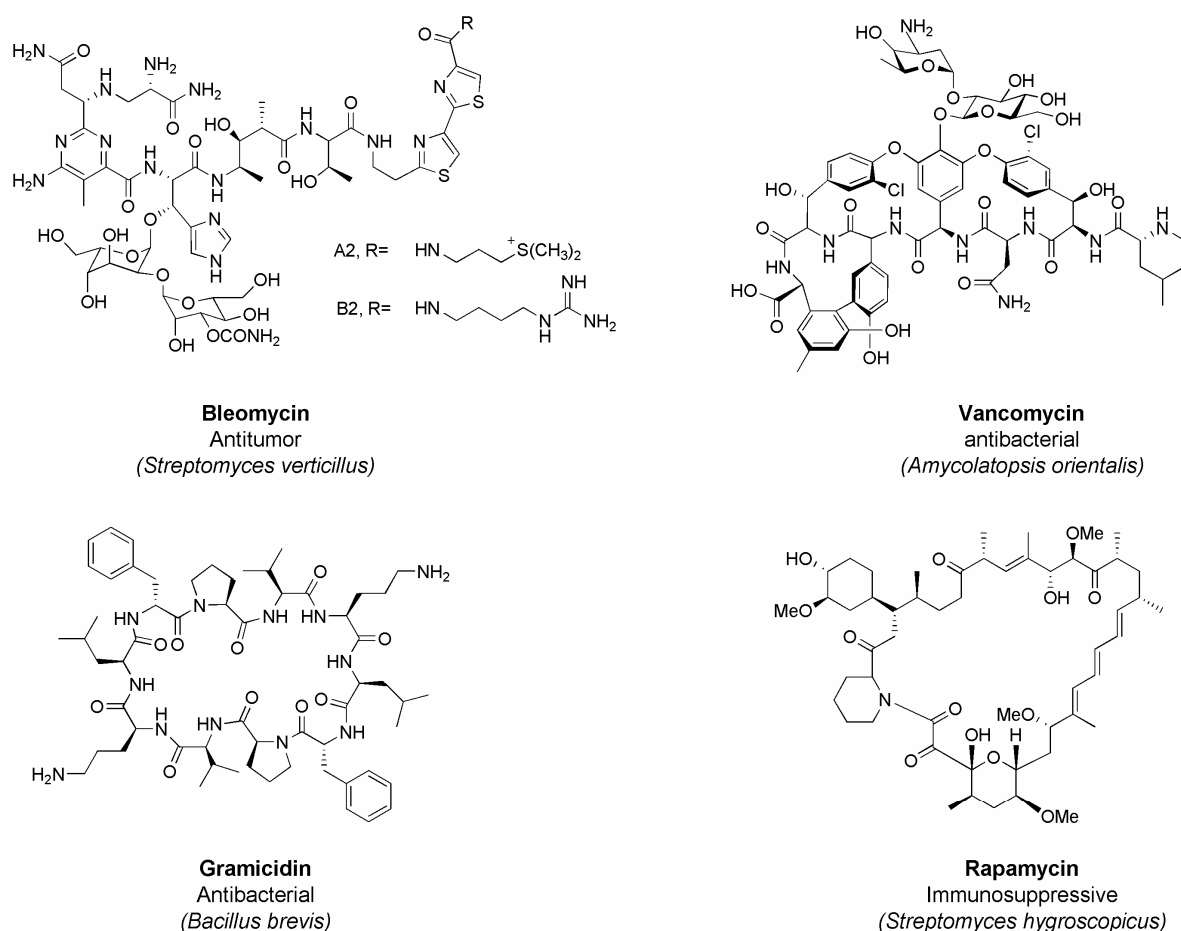
Figure 1.1:   **Prominent natural products used in clinical applications.**

Natural product research, as we know it today, began with the serendipitous discovery of penicillin in 1929 by Alexander Fleming – an event which inaugurated a new era in the treatment of bacterial diseases [1]. Since then, many natural products and synthetically modified natural product derivatives have been successfully developed for clinical use to treat human diseases, across a broad range of therapeutic areas [2]. Even today, natural products continue to play a dominant role in the discovery of leads for the development of new drugs. In fact, approximately 50% of the pharmaceuticals currently in clinical use are of natural product origin .

A special focus in the field of natural product research is the development of new antibiotics [3]. Motivating these efforts is the fact that infectious diseases remain the second-leading cause of death worldwide and the third-leading cause of mortality in economically-advanced countries [4]. In order to deal with growing bacterial resistance to existing drugs, there is a pressing need for the discovery and development of new antibiotics [5]. Despite the fact that large pharmaceutical companies have significantly de-emphasized natural product discovery over the past several years, many new and interesting molecules with unique scaffolds and/or novel modes of action have been found during this time by research efforts in academia and small pharma companies [6]. Smaller biotechnology companies, in particular, have increased their focus on natural product research, mainly in the infectious diseases but also in anticancer areas (Figure 1.2) [6]. These findings emphasize that natural product research continues to provide significant value in the discovery of novel chemical structures and bioactive leads for clinical development.

Platensimycin
antibacterial

Sorbicillactone
anticancer

Cyanosporaside A
anticancer
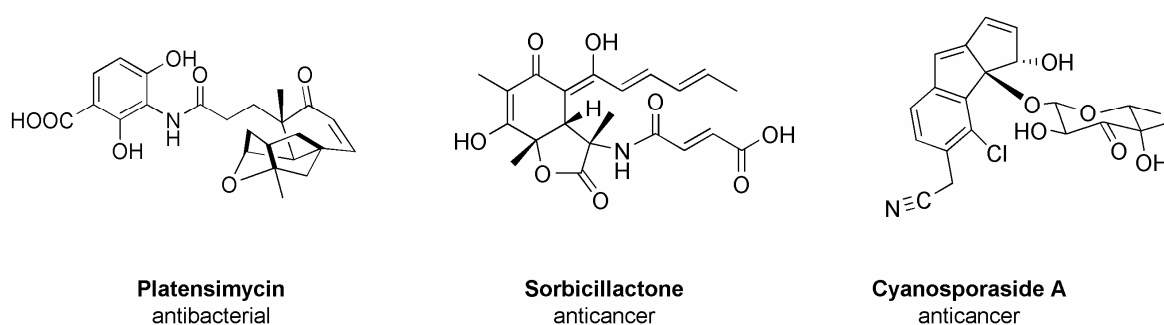
Figure 1.2: New substances derived from natural products.

Among the natural producers of secondary metabolites, beside plants, microorganisms are a particularly promising source of novel bioactive entities. In nature, the diversity of

microorganisms is enormous and only a very minor portion of bacterial and fungal species have been cultured to date and examined for secondary metabolite production [7]. The majority of bioactive compounds were isolated from, non-filametous bacteria, fungi and actinomycetes [7;8].

With respect of secondary metabolism, actinomycetes are the most-intensively studied group; among them, the genus *Streptomyces* has been proven to be the most proficient producer of secondary metabolites, at least under standard laboratory conditions [9]. A wide array of bioactive molecules has been isolated from streptomycetes, leading to their historical importance in natural product discovery [3].

1.2 Streptomycetes as important natural product producers

Streptomycetes are soil-dwelling bacteria and ubiquitous in nature. Their ability to colonise the soil is greatly facilitated by growth as a vegetative hyphal mass which can differentiate into spores that assist in spread and persistence (Figure 1.3). The spores are a semi-dormant stage in the life-cycle, and can survive in soil for long periods (several years).
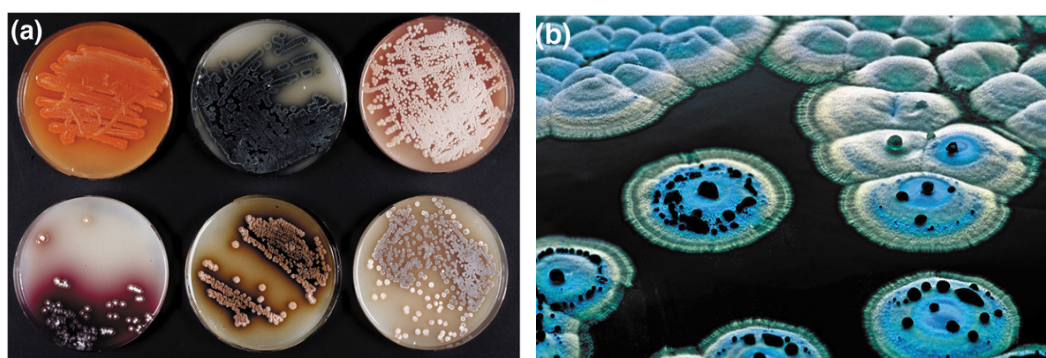
Figure 1.3:   **Streptomycetes colonies.**

(a) Typical colony morphologies of *Streptomyce sp.* isolated from the soil. (b) A panoramic view of the colony morphology of *Streptomyces coelicolor*. Both peripheral and aerial mycelia develop from the central mass of the colony. Metabolites, including the blue antibiotic actinorhodin, are excreted into the medium and into aqueous droplets on the hydrophobic surface of the colony. (Figures reproduced from reference [9]).

Streptomycetes produce many extracellular enzymes, such as cellulases, xylanases, amylases, maltases, and others. These catalysts make them important players in soil biodegradation processes, as they are able to decompose complex mixtures of polymers in dead plants, animal and fungal material [10]. Further characteristics of these strains are their high G+C content (72% on average), linear chromosomes and and large genome sizes. The sequenced genome of *Streptomyces coelicolor* comprises 8.6 Mbp [11], whereas the *S. avermitilis*

genome contains 9 Mbp [12]. However, their most notable characteristic is their capability to produce biologically active secondary metabolites, a feature which has made them the focus of natural product research.

Streptomycetes were first recognized as natural product producers some 60 years ago, with the discovery of streptomycin [13]. Since then, the majority of known antibiotics have been isolated from streptomycetes. Of the 12000 or so antibiotics known by 1995, 55% were derived from streptomycetes and 11% from other actinomycetes [14]. Among these antibiotics are drugs which have been in clinical use as antibiotics for many years, including the macrolide antibiotic erythromycin A synthesized by the actinomycete *Saccharopolyspora erythraea* [15] and the aromatic polyketide tetracycline [16] (Figure 1.4).

Erythromycin A
Antibacterial
(Saccharopolyspora erythrea)

Tetracycline
Antibacterial
(Streptomyces aureofaciens)

Doxorubicin
Anticancer
(Streptomyces peutceticus)

Daptomycin
Antibacterial
(Streptomyces roseosporus)
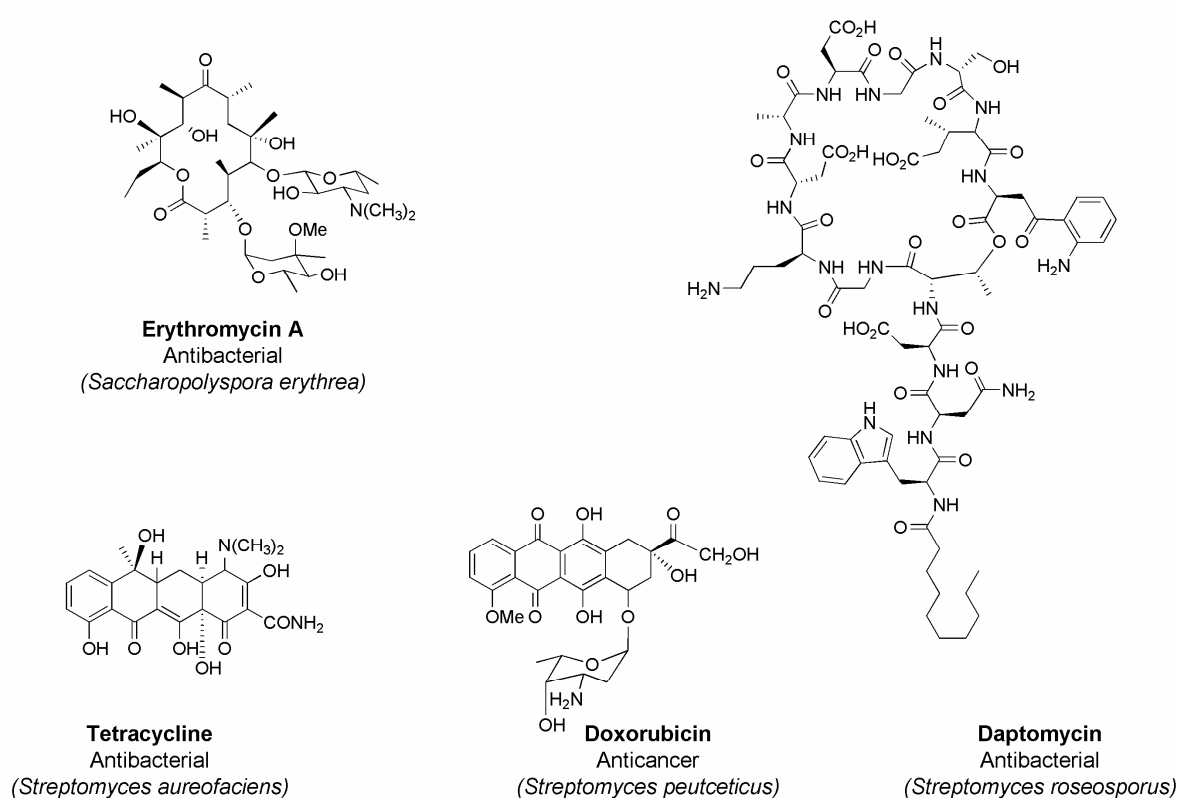
Figure 1.4: **Natural compounds from actinomycetes and streptomycetes which serve as important medicines.**

Another prominent example of a therapeutically relevant streptomycete natural product is the cyclic lipopeptide antibiotic daptomycin (Cubicin®) from *S. roseosporus*. Daptomycin was approved in the United States in 2003 and in Germany in 2006 for the treatment of skin infections caused by Gram-positive pathogens. In addition to antibiotics, streptomycetes have

also yielded important anticancer compounds, such as the cytostatic doxorubicin [17], which is widely used in chemotherapy (Figure 1.4).

As the vast majority of therapeutically important secondary metabolites are produced from *Streptomyces sp.*, making them a proven source for therapeutically-useful compounds, they should be the continued focus of research to identify new bioactive substances. In the last years intense efforts have also been made to discover new groups of microbial producers with the assumption that an increase to discover new organisms correlates with the potential to discover chemical diversity (new structures). These investigations dealt with marine bacteria, pathogenic bacteria (pseudomonads) or organisms that are difficulte to cultivate (cyanobacteria) [18-20]. Among the bacteria discovered in these screenings especially the group of myxobacteria have attracted attention as proficient producers of novel bioactive compounds [21].

In the majority of the cases, these secondary metabolites are the end products of complex, multistep biosynthetic processes. Understanding the mechanisms involved in the biosynthesis of the compounds is an essential prerequisite for optimizing production and yield. Furthermore, directed manipulation of the genes governing secondary metabolism offers a promising alternative to total or semi-synthesis, to generate altered natural products. The success of this approach depends on the cloning and subsequent genetic and biochemical characterization of the corresponding biosynthetic pathways.

Fortunately, in the majority of cases, the compounds are produced by sets of genes which are clustered in the bacterial chromosome; this feature presumably allows the microorganisms to co-regulate the expression of the many genes. The first proof that the genes for antibiotic biosynthesis are clustered was provided by Malpartida and Hopwood in 1984 [22]. The researchers isolated and cloned a continuous DNA fragment from *S. coelicolor* which apparently carried the complete genetic information required for the biosynthesis of the antibiotic actinorhodin. Indeed, the cloned DNA fragment could complement several actinorhodin non-producing mutants. Furthermore, introduction of the DNA fragment into the heterologous host *S. pavulus* directed actinorhodin biosynthesis – providing proof that the cluster of genes located on this DNA fragment was responsible for the production of the antibiotic actinorhodin. The advantage for natural product researchers is that if a single gene within the gene cluster is located, the others can mostly be identified fairly straightforwardly by chromosomal walking.

Many molecular tools have been developed for targeted genetic manipulation in streptomycetes over the last several decades [10] and hundreds of biosynthetic gene clusters

have been identified to date. Together, these developments have enabled detailed analysis of the molecular principles that govern multiple pathways [23]. Recently, this field has been revolutionized by advances in sequencing technologies, which now allow the rapid sequencing of entire bacterial genomes in a matter of days. The first *Streptomyces* strain to be sequenced was *S. coelicolor* [11]. Before the genome of *S. coelicolor* was sequenced, it was known that the strain produces three antibiotics, actinorhodin, prodiginin and the nonribosomal peptide calcium-dependent antibiotic (CDA) [10;11]. However, the genome sequence revealed a further 23 gene clusters (comprising approximately 5% of the total genome), which are likely to be dedicated to secondary metabolism.

Thirty gene clusters were identified in the second sequenced strain, *S. avermitilis*, which could be correlated to secondary metabolism, corresponding to 6.5% of the genome. Although antibiotic production by the *Streptomcyes* has been studied intensively over the last century, mathematical models predict that only 3% of antibacterial compounds synthesized by these bacteria have been discovered to date [24]. These facts indicate that the genomic capacity of these microorganisms to synthesise natural products is much higher then originally anticipated, a hypothesis which is supported by the available genome sequencing data. In order to identify, isolate and evaluate these as yet uncharacterized biosynthetic gene clusters and exploit them for analogue production, techniques must be developed to allow the straightforward genetic manipulation of the strains. Genetic manipulation in natural producer strains is straightforward in some model organisms (for example *S. coelicolor*) but for many other *Streptomyces* strains it remains more challenging, although there has been some progress [10]. Many of these strains are slow-growing or not yet culturable, which further hinders the genetic manipulation of the biosynthetic pathways.

One very useful approach is to express entire gene clusters in genetically more tractable production hosts, which facilitates the genetic manipulation of the secondary metabolite pathways. This so-called 'heterologous expression' technique represents a promising tool for both obtaining deeper insights into the biosynthetic pathways, and for generating novel derivatives of natural compounds for evaluation as drug leads.

1.3   Heterologous expression of natural product pathways

There has been considerable recent interest in both the natural product research and drug discovery communities, in developing heterologous expression techniques for complete secondary metabolite pathways. Furthermore, heterologous expression can also be used to

screen genomic DNA obtained from microbial niches (metagenomes) for the production of novel secondary metabolites [25;26].

Efforts towards achieving heterologous expression have met with considerable success during the last decade [27-29]. However, in most cases, only small biosynthetic gene clusters were expressed in hosts which were closely-related to the producing strains [30]. To date, the heterologous expression of secondary metabolite pathways from *Streptomyces* has mainly been achieved in related actinomycete species, and the relatively small biosynthetic gene clusters (less than 30 kbp) were often located on a single cosmid or bacterial artificial chromosome (BAC) within a genomic library [31-33]. Expression of larger and more complex pathways remains more demanding, but can be facilitated by constructing several plasmids that harbor subsets of the biosynthetic genes [34].

As this approach requires time consuming classical cloning steps, however, a new method for the expression and engineering of large biosynthetic pathways was developed in our working group by Wenzel *et al.* [35]. In this case, Red/ET (λ-mediated) recombineering technology [36-38] was used to reconstitute the hybrid nonribosomal peptide synthetase/polyketide synthase (NRPS/PKS) myxochromide S pathway from the myxobacterium *Stigmatella aurantiaca* on one expression construct and to integrate genetic elements for expression in *Pseudomonas putida.* Heterologous expression was successful, and the construct yielded higher amounts of myxochromide in *P. putida* than in the natural producer *Stigmatella aurantiaca* [35;39;40].

Red/ET is a novel technique for DNA manipulation, which is precise and independent of the presence of restriction sites and the size of the DNA molecule. These features make it ideal for the manipulation of large DNA fragments. DNA molecules targeted by Red/ET recombineering are precisely altered by homologous recombination in strains of *E. coli* which express phage-derived protein pairs, either RecE/RecT from the Rac prophage, or Redα/Redβ from λ phage. These protein pairs are functionally and operationally equivalent. RecE and Redα are 5'→3' exonucleases, while RecT and Redβ are DNA annealing proteins (Figure 1.5).
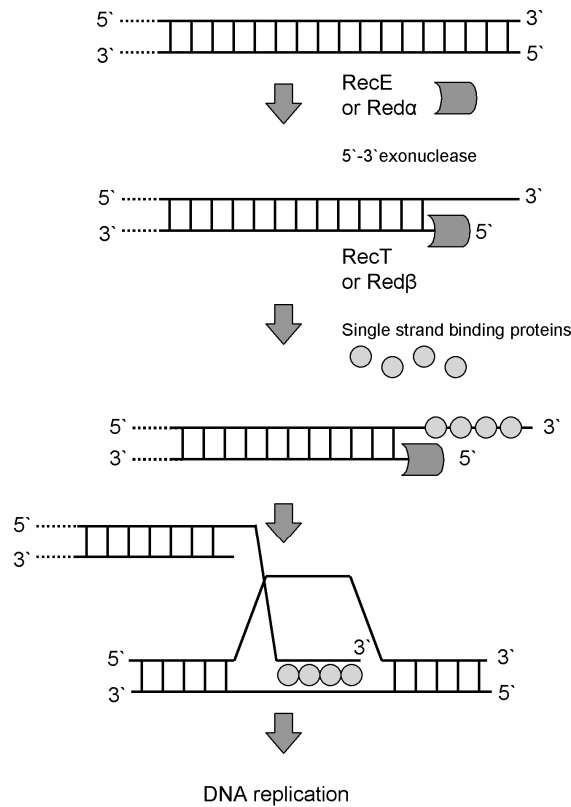
Figure 1.5: **Mechanism of Red/ET recombineering.**

The target double strand DNA is digested by a 5′→3′ exonuclease (RecE or Redα), after which the single-strand binding protein (RecT or Redβ) can bind to the DNA to stabilize the single strand. Homologous recombination can then take place. Diagram reproduced from www.genebridges.com.

A functional interaction between RecE and RecT, or between Redα and Redβ, is also required in order to catalyze the homologous recombination reaction. Recombination is mediated by homology regions, which are stretches of identical DNA sequences found on both molecules to be recombined (Figure 1.6). As the sequence of the homology regions can be generated in a chemically synthesized oligonucleotide, any position on a target molecule can be specifically altered. Red/ET recombination thus allows, at least in principle, every type of possible DNA modification (insertions, deletions, substitutions, fusions, point mutations, direct cloning and subcloning) regardless of the size of the target molecule (e.g. cosmids or BACs) [41].
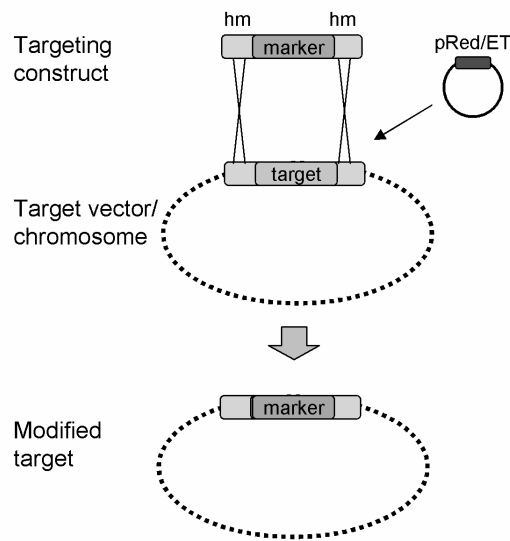
Figure 1.6:   Red/ET recombineering.

The central step in Red/ET recombineering is the crossover between a targeting construct containing homology arms (hm) and the target, which can be a gene locus on the *E. coli* chromosome or any other stretch of DNA in a BAC or plasmid vector contained within the cell. Diagram reproduced from www.genebridges.com.

Perlova, *et al.* [42] also used Red/ET to stitch together a myxobacterial biosynthetic pathway. Here, Red/ET recombineering was employed to modify two cosmids containing the entire myxothiazol pathway. In a final ligation step, the whole pathway was reconstituted on one expression construct, which was subsequently expressed in *Myxococcus xanthus*. Production of myxothiazol was comparable to that of the natural producer *Stigmatella aurantiaca* [42].The same pathway was also expressed in *Pseudomonas putida* [43]. These and subsequent examples [44;45] demonstrate how this promising technique can now be used for the reconstitution of large and complex biosynthetic pathways in *E. coli* and their subsequent heterologous expression in related or non-related heterologous hosts. Figure 1.7 gives a general overview of the strategy for cloning and heterologous expression of large natural product assembly lines [35;39]
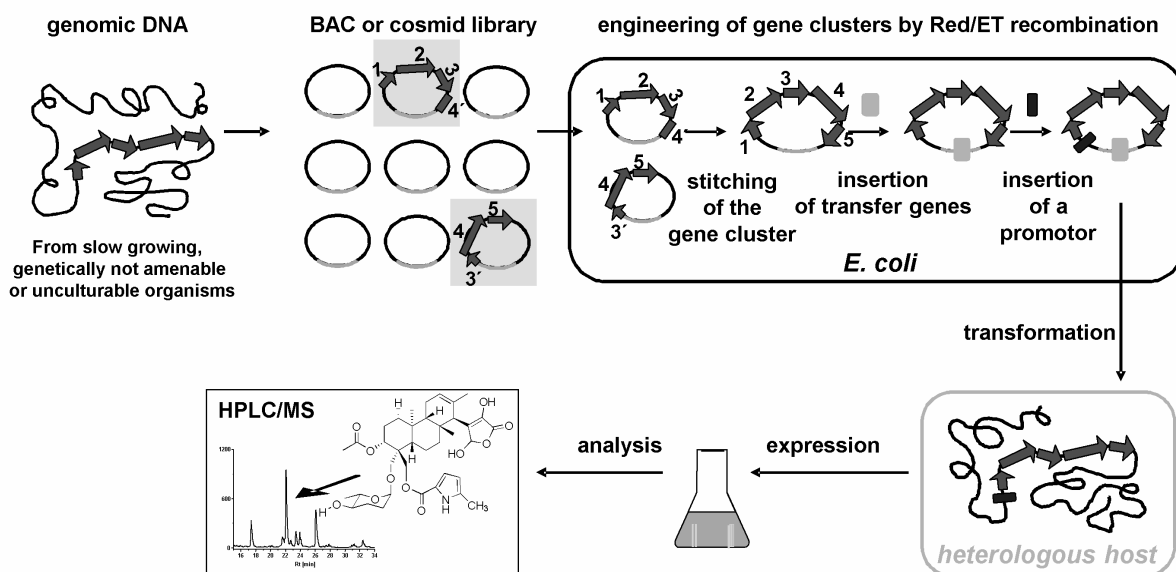
Figure 1.7:   **General strategy for the heterologous expression of large biosynthetic pathways.**

After screening of a BAC or cosmid library for constructs containing parts of the investigated natural product biosynthetic gene cluster, overlapping fragments can be stitched together in *E. coli* using Red/ET recombineering. The resulting construct harboring the complete natural product assembly line can be further modified in *E. coli* by recombineering, e.g., by inserting genes needed for transformation into the heterologous host and/or insertion of promoter region(s). The final construct can be transferred into a suitable heterologous host for expression of the biosynthetic pathway. Natural product formation can ultimately be analyzed by conventional techniques, e.g., HPLC-MS and/or TLC [35].

The successful heterologous expression of several classes of natural products, including the polyketides (PKs) and non-ribosomal polypeptides (NRPs) and their hybrids, presented a special challenge, because the biosynthesis of these complex structures typically requires large gene sets [46;47]. The core structures in each case are assembled by gigantic multienzymes called polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs), respectively. In both systems, the proteins have to be posttranslationally modified in order to be active [48]. These multifunctional enzymes can reach remarkable sizes, for example mycolactone A and B from the the bacterium *Mycobacterium ulcerans* are synthezised by PKS multienzymes housing eight and seven extension modules containing 16990 amino acids and 14130 amino acids, respectively [49].

Once the proteins are expressed as functional proteins within the cell, they require a specific pool of substrates, including but not limited to short-chain fatty acids, proteinogenic and non-proteinogenic amino acids, mevalonate, shikimate pathway intermediates, and sugars to enable natural product formation. Therefore, the required substrates must be available in the heterologous hosts in adequate amounts, at the correct time point.

1.4   The biosynthetic logic of nonribosomal peptide synthetases

NRPS multienzymes are composed of repetitive, multicatalytic units called modules which carry out several reactions in a coordinated manner. Typically, each module in the assembly line performs one cycle of chain extension (incorporation of one amino acid into the growing peptide chain). One cycle of chain extension comprises three essential enzymatic reactions: 1) recognition of an amino acid as substrate and its activation as its aminoacyl-adenylate; 2) covalent binding of the amino acid as a thioester to the multienzyme; and 3) condensation with the amino acyl or peptidyl group tethered to the neighbouring module [50]. This process is schematized in Figure 1.8. Overall, peptide synthesis proceeds in an N- to C-terminal direction.

Each of the steps in chain extension is carried out by a specific domain within the module. A typical module consists minimally of an adenylation (A) domain, a peptidyl carrier protein (PCP) domain (also referred to as a thiolation (T) domain) and a condensation (C) domain [47]. The A domain is responsible for recognizing and activating the starter/extender unit, which is then covalently linked to the PCP. The attachment point is the terminal sulfhydryl of a phosphopantetheine arm attached to a highly conserved serine residue [51]. This 'swing arm' is added post-translationally to the PCP by a phosphopantetheinyl transferase (PPTase). The Ppant arm (20 Å in length) is characterized by its flexibility, which presumably facilitates the transport of the intermediates to the reaction centers [51]. The condensation domain then catalyses peptide bond formation between the aminoacyl-S-PCP and the incoming aminoacyl or peptidyl-S-PCP (see Figure 1.8). The release of the intermediates in a linear or cyclic form is normally performed by a thiosterase domain, which is located at the end of the last module of the assembly line [52;53]
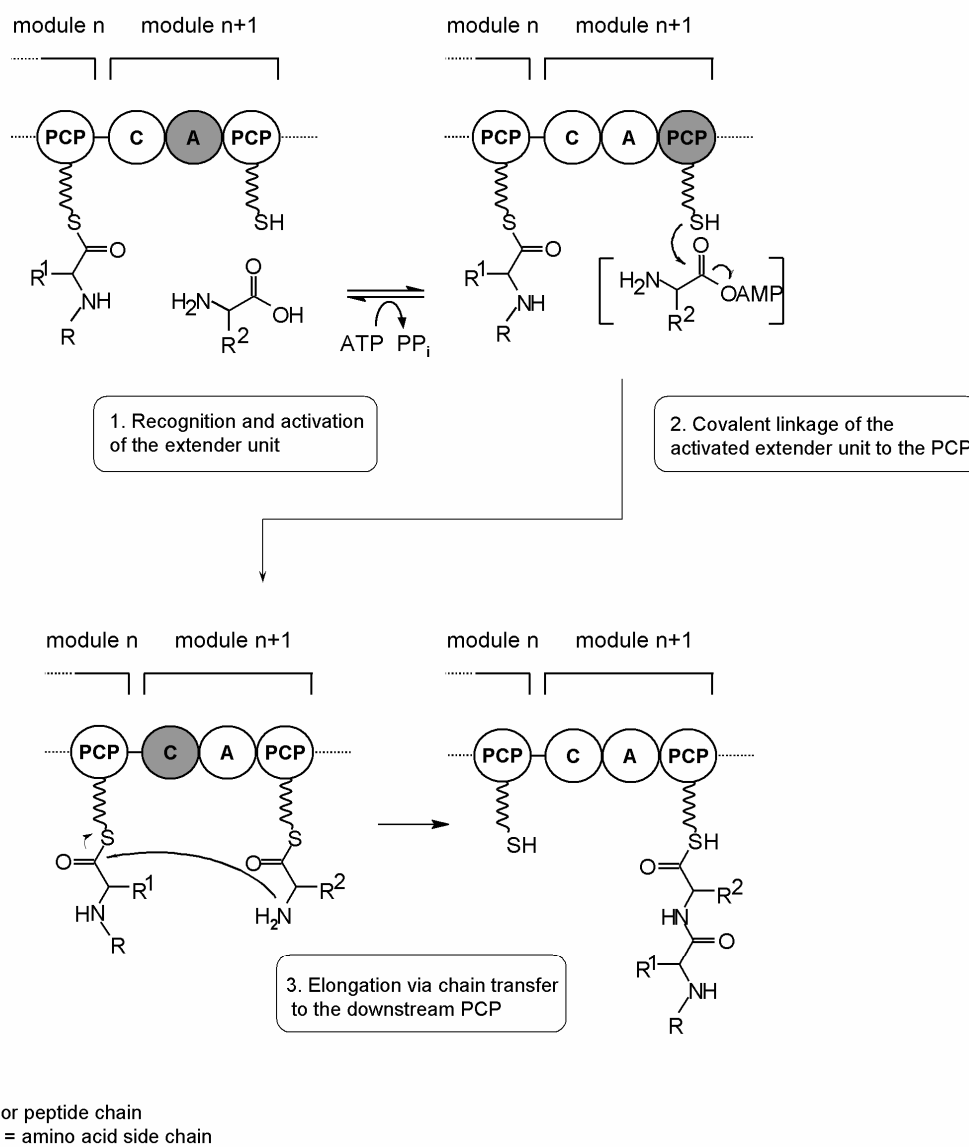
Figure 1.8: **Reactions catalyzed by the essential domains in NRPS multimodular assembly lines.**

Domains that are involved in the particular reactions are highlighted in grey. A: activation domain ; PCP: peptidyl carrier domain; C: condensation domain.

A variety of optional domains increase the structural diversity of the products, including methyl transferase (MT), epimerase (E), oxidation (Ox) and heterocyclization (HC) activities [54]. These tailoring enzymes are embedded in the NRPS assembly lines and modify the elongating chains while they are still covalently tethered to the proteins as peptidyl-*S*-enzyme intermediates. Other tailoring enzymes act in *trans*, recognizing either or both the peptidyl chains or the PCP domains. A third group of tailoring enzymes act after the assembly-line stage of the biosynthesis, and further transform the released peptides (e.g., by introducing cross-linking or addition of glycosyl groups) [55].

In general, nonribosomal peptide synthesis is a linear process in which each module is responsible for one particular chain elongation step and the specific order of the modules defines the sequence of the incorporated amino acids. However, recent characterization of many NRPS systems has revealed several examples where the arrangement of modules and domains within the enzyme is not co-linear with the sequence of transformations required to generate the observed product. Therefore, Mootz *et al.* proposed a classification of NRPS systems into three groups, according to their biosynthetic logic [56]: linear NRPSs, iterative NRPSs and nonlinear NRPSs. Representative examples for each type of biosynthesis are shown in Figures 1.9–1.11.
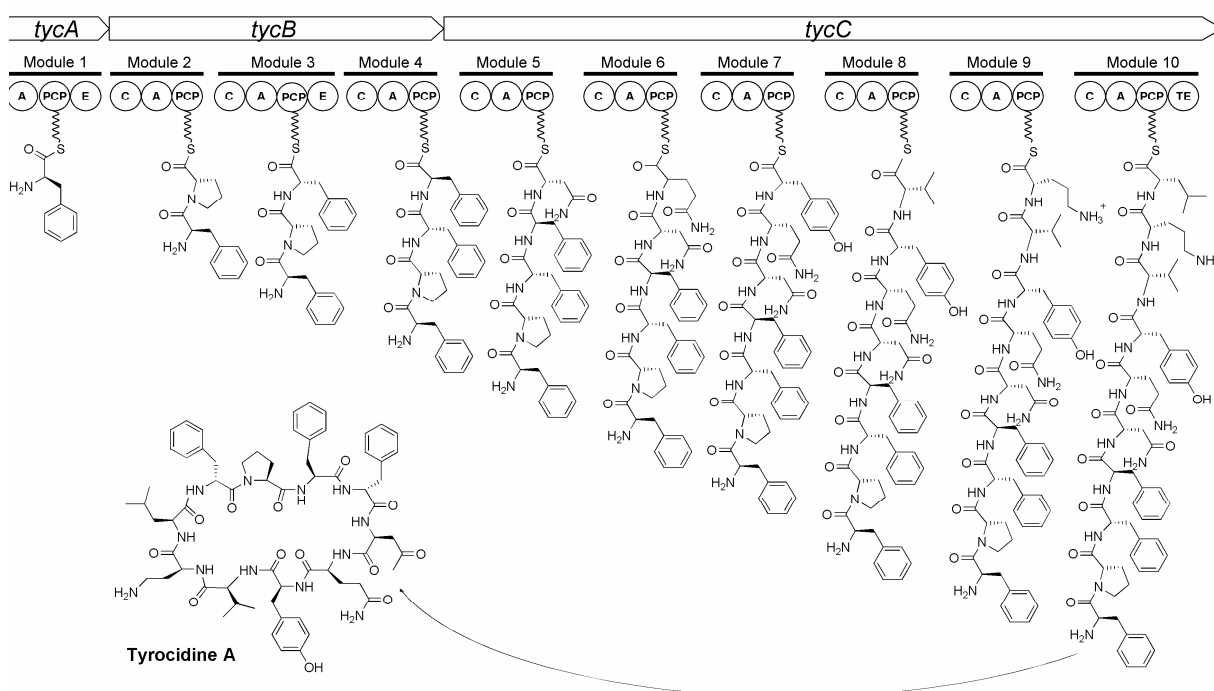
Figure 1.9: **The multiple carrier thiotemplate mechanism illustrated with the example of tyrocidine A synthesis.**

Three peptide synthetases encoded by the genes *tycA*, *tycB* and *tycC* act together to assemble the cyclic decapeptide. Diagram reproduced from [57].

Figure 1.9 illustrates a linear NRPS, which is involved in the biosynthesis of the decapeptide tyrocidine [58]. In a linear NRPS, the core domains are arranged in the order C-A-PCP. The initiation module, which lacks the C domain, is responsible for the incorporation of the first amino acid. The terminal module normally contains a TE domain for release of the full-length peptide chain. Other examples of linear NRPSs are those responsible for biosynthesis of surfactin [59] and pristinamycin [60].
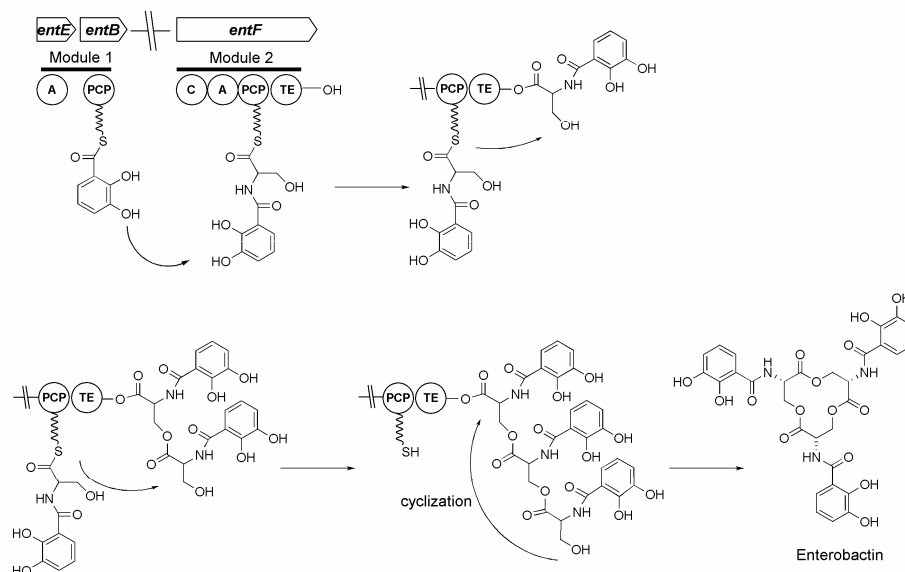
Figure 1.10:   **The enterobactin assembly line as example for iterative NRPSs.**

Three dihydroxybenzoylserine-*S*-PCP intermediates are generated on the two modules which are oligomerized and cyclized to the final product on the TE domain. Diagram reproduced from [56].

In iterative NRPSs, modules or domains are used more than once for the assembly of a particular product. This strategy is employed to construct peptide chains that consist of repeated smaller units. The key step is then the oligomerization and cyclization of the monomers by the thioesterase domain. For example, in enterobactin biosynthesis (Figure 1.10), the complete NRPS assembly line acts iteratively to generate a common structural unit (trimer of dihydroxybenzoylserine units) [61], which is then oligomerized to the final product. Bacillibactin of *B. subtilis*, a cyclotrimerized dihydroxybenzoyl-glycyl-threonine-peptide, is structurally very similar [62]. An iterative NRPS is also involved in the assembly of gramicidin [63].
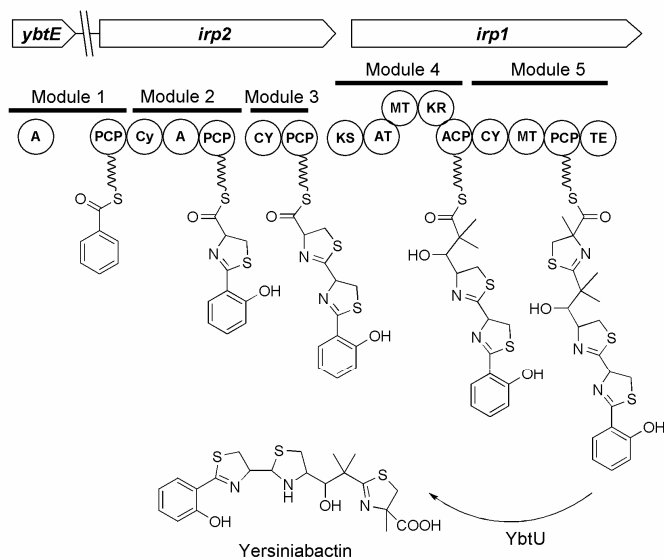
Figure 1.11: **The non-linear yersiniabactin NRPS.**

Yersiniabactin biosynthesis proceeds on a mixed NRPS/PKS system. Cy: cyclization domain, MT: methylation domain. Diagram reproduced from [56].

Over the last several years, it has become clear that a number of biosynthetic gene clusters deviate in their domain organization from the standard C-A-PCP architecture. In fact, these nonlinear NRPS clusters comprise a major proportion of the NRPS systems found in nature [64]. The main characteristics of these systems are the unusual arrangement of the core domains C, A and PCP. Such a non-linear NRPS is depicted in Figure 1.11. The NRPS responsible for the siderophore yersiniabactin is a striking example of nonlinear NRPS assembly, where one A domain is used to load three different PCPs [65]. Due to the unusual domain organization of nonlinear NRPS biosynthetic gene clusters, it is very difficult to predict the product structures with any confidence. Therefore, detailed biochemical studies are needed to understand the function and the interplay between the enzymes, which makes research with nonlinear NRPS very challenging.

1.5   Outline of the dissertation

Studies on the biosynthesis and the heterologous expression of complex secondary metabolites from streptomycetes

The present thesis deals with experiments to heterologously express the phenalinolactone biosynthetic gene cluster from *Streptomyces* Tü6071, and to investigate the biosynthesis of the GE81112s from *Streptomyces* 14386. Figure 1.12 shows the GE81112 family and the phenalinoactones, of which the main metabolites are PL A and PL D.

GE factor A	R=O	R₁=H
GE factor B	R=O	R₁=NH2
GE factor B1	R=NH	R₁=NH2

PLA	R₁=OH	R₂=	R₃=
PLD	R₁=H	R₂=	R₃=
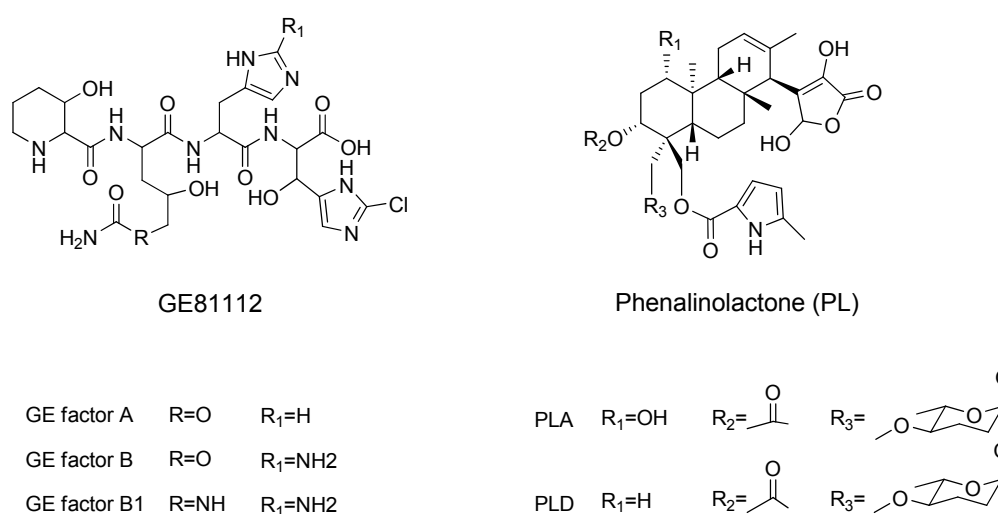
Figure 1.12:   Structures of the GE81112 metabolites and the phenalinolactones.

The phenalinolactones are structurally intriguing tricyclic terpene glycosides which are produced by *Streptomyces* sp. Tü6071. The compounds exhibit promising antibacterial activity [66]. In addition to a highly oxidized γ-butyrolactone ring, which is most likely derived from pyruvate, the oxidatively functionalized terpenoid backbone is decorated with a 5-methylpyrrole-2-carboxylic acid as well as the rare deoxyhexose, 4-*O*-methyl-L-amicetose. The recent cloning and sequence analysis of the *pla* biosynthetic gene cluster from *Streptomyces* Tü6071 [66] revealed a highly complex genetic architecture that consists of 35 *orfs*, which are organized into 11 putative operons.

Our goal was to heterologously express the cluster using Red/ET recombineering, as the technique had already been shown to be suitable for the reconstitution and expression of natural product pathways. However, at the outset of this work, the approach had not been applied to non-PKS/NRPS gene clusters from streptomycetes. Initially the PL cluster was

located on two different cosmids. Therefore, the first aim of the work was to collect the biosynthetic genes together onto one expression construct, and then to introduce genetic elements for expression in different heterologous hosts. The established expression system was then expected to provide a convenient platform for functional investigations of the biosynthetic genes and the generation of novel analogues, by genetic engineering of the pathway in *E. coli*.

The second part of this thesis describes work on a second class of secondary metabolites, the GE81112 tetrapeptide family. The GE81112 compounds were isolated from a *Streptomyces* sp., and to date, three variant GE factors have been identified (GE81112 factors A, B1 and B). The metabolites were found in the course of microbial product screening experiments aimed at discovering novel antibiotics acting on bacterial protein synthesis. Extensive NMR and MS studies revealed that the GEs incorporate 4 uncommon amino acids: the nonproteinogenic amino acids hydroxypipecolic acid and hydroxypentanoic acid, as well as an (amino)histidine and a hydroxychlorohistidine [67] (Figure 1.12). Although the core structure of the GEs could be predicted straightforwardly to arise from a nonribosomal peptide synthetase (NRPS) multienzyme, the origin of several amino acids and functional groups, including the aminohistidine as well as the chlorination and carbamoylation, was not obvious from considerations of classical assembly-line biosynthesis. As the biosynthetic gene cluster had not been identified, the first objective was to clone the GE81112 biosynthetic gene cluster in order to allow studies of the underlying biosynthesis. Once the cluster was identified, it would then be annotated, as the first step towards deciphering the biosynthetic pathway. In parallel, we aimed to develop methods to genetically manipulate the strain, as they were not available. A final goal was to investigate GE biosynthesis by heterologous expression.

2 Material and Methods

2.1 Chemicals

All chemicals used in this work were obtained from the manufacturers, at analytical grade.

Table 2.1: Chemicals used and their sources

Chemical product	Manufacturer
Acetone Acetonitrile Chloroform Ethanol Ethyl acetate Hexane Methanol Magnesium sulfate 2-Propanol Succinic acid	Sigma Aldrich
Bacto agar Casitone Casaminoacids Malt extract Tryptone Peptone Yeast extract	Becton Dickinson and Co., USA
Ammonium acetate Sodium-EDTA Potassium hydroxide	Fluka
Acetic acid Boric acid Bromphenolblue Calcium chloride-di-hydrate Glucose monohydrate Glycerin Magnesium chloride Magnesium sulfate hepta-hydrate Potassium acetate Potassium hydrogen phosphate Potassium sulfhate Sodium chloride Sodium citrate-di-hydrate Sodium hydroxide	Merck
BSA (Bovine Serum Albumin)	New England Biolabs
Ammonium persulfate 5-Bromo-4-chloro-3-indolyl- -D-galactopyranoside Coomassie Brilliantblue 250R Dithiothreitol Ethidium bromide solution (1%) Formamide Maleic acid Sodium dodecyl sulfate (SDS) Rotiphorese®Gel 30 TEMED Triton X-100	ROTH

2.2 Enzymes, Kits and Markers

All enzymes used in this work are listed in Table 2.2 and were used according to the manufacturers' protocols.

Table 2.2: **Enzymes, kits and markers and their manufacturers**

Product	Manufacturer
Page Ruler™ Prestained Protein Ladder	MBI Fermentas
Page Ruler™ Protein Ladder	
1 kb DNA ladder	
T4 DNA ligase	
Shrimp alkaline phosphatase	
Restriction endonucleases	
Plasmid purification kit	
Nucleotides	
Cre recombinase	New England Biolabs
Phusion DNA Polymerase	Finnzymes
Triple Master Polymerase	Eppendorf
HotStarTaq Ploymerase	Quiagen
Topo TA Cloning® Kit	Invitrogen
Gel-Dry™ Drying Kit	
Nucleospin® Extract	Macherey & Nagel
Ribonuclease A	Roth
Proteinase K	
Lysozyme	
DIG labeleld DNA Molecular Weight Marker III	Roche
DIG-Labeling and Detection Kit	
DIG-Blocking solution	
DIG EasyHyb	
DIG PCR Labeling Mix	
BugBuster® Protein Extraction Reagent	Novagen
Gigapack® III Gold Packaging Extract	Stratagene

2.3 Buffers and stock solutions

2.3.1 Solutions and buffers for molecular biology applications

Table 2.3:   Solutions and buffers for molecular biology

Buffer	Constituents	
Cell Lysis buffer (P1)	NaOH	20 mM
	10% SDS	10 ml
	H_2O	to 100 ml
Cell suspension buffer (P2)	Glucose	5 mM
	Tris	2.5 mM
	EDTA	1 mM
	H_2O	to 100 ml
DNA Loading buffer (6×)	Glycerin (87%)	3 ml
	Bromphenol blue	25 mg
	Xylene cyanol	25 mg
	H_2O	to 10 ml
0.5 M EDTA	EDTA	186 g
	NaOH	20 g
	H_2O	to 1L
Glycerol (50%)	Glycerol	115 ml
	H_2O	to 1L
Glycerol (20%)	Glycerol	57.5 ml
	H_2O	to 100 ml
0.1 M IPTG	IPTG	240 mg
	H_2O	to 10 ml
5 M Sodium chloride	NaCl	580 g
	H_2O	to 100 ml
Neutralisation buffer (P 3)	3 M Potassium acetate	60 ml
	Acetic acid	11.5 ml
	H_2O	to 100 ml
10% SDS solution	SDS	10 g
	H_2O	to100 ml
SET buffer	Tris-HCl (pH 7.5)	20 mM
	NaCl	75 mM
	EDTA (pH 8)	25 mM
		To 100 ml

Buffer	Constituents	
TBE buffer (1x)	TRIZMA base	242 g
	Boric acid	57.1 ml
	Na-EDTA	100 ml
	H_2O	to 1L
SM-buffer	Tris-HCl (1 M, pH 7.4)	50 ml
	NaCl	5.4 g
	$MgSO_4 . 7 H_2O$	2 g
	Gelatine-solution (2%)	5 ml
	H_2O	to 1 L
1 M Tris-HCl	TRIZMA Base	60.5 g
	H_2O	to 500 ml
P (protoplast) buffer	Sucrose	103 g
	K_2SO_4	0.25 g
	$MgCl_2 . 6 H_2O$	2.02 g
	Trace element solution	2 ml
	H_2O	to 800 ml
	Add after autoclaving:	
	KH_2PO_4 (0.5%)	1 ml
	$CaCl . 2H_2O$ (3.68%)	10 ml
	TES buffer (5.73%, pH 7.2)	10 ml

2.3.2 Solutions and buffers for Southern Blot and colony hybridization

Table 2.4: Solution and buffers for Southern Blot and colony hybridization

Buffer	Constituents	
Depurination buffer	HCl (1 M)	250 ml
	H_2O	to 1L
Denaturation buffer	NaOH	20 g
	NaCl (5 M)	333 ml
	H_2O	to 1 L
Neutralisation buffer	Tris HCl (1 M, pH 7.4)	500 ml
	NaCl	175.5 g
	H_2O	to 1L
SSC (20×)	NaCl	155.3 g
	Sodium citrate.2 H_2O	88.2 g
	H_2O	to 1L

Buffer	Constituents	
Hybridization buffer	DIG Easy Hyb Granules	
	H_2O	to 64 ml
2× wash solution	SSC (20×)	100 ml
	10% SDS	10 ml
	H_2O	to 1L
1× wash solution	SSC (20×)	50 ml
		10 ml
		to 1 L
Maleic acid buffer	Maleic acid	11.61 g
	NaCl	8.8 g
	H_2O	to 1L
Blocking solution	DIG blocking reagent	50 g
	Maleic acid buffer	to 500 ml
Detection buffer	Tris-HCl (1 M)	100 ml
	NaCl	5.8g
	H_2O	to 1 L
Stripping buffer	NaOH	8 g
	SDS (10%)	10 ml
	H_2O	to 1L

2.3.3 Solutions and buffers for enzymatic applications

Table 2.5:   Solution and buffers for enzymatic applications

Buffer	Constituents	
10x buffer	Tris-HCl	750 mM
	NaCl	1 M
	$MgCl_2$	100 mM
	H_2O	to 50 ml
Binding buffer	Tris-HCl (pH 7.8)	20 mM
	NaCl	200 mM
	Glycerol	100 ml
	Imidazole	60 mM
	H_2O	to 1L
Borate buffer	Boric acid	200 mM
	H_2O	to 1L

Buffer	Constituents	
Elution buffer	Tris-HCl (pH 7.8)	20 mM
	NaCl	200 mM
	Glycerol	100 ml
	Imidazole	500 mM
	H_2O	to 1 L
Stop-mix	Activated charcoal	6 g
	Tetrasodium pyrophosphate	50 ml
	Perchloric acid	19.2 ml
	H_2O	to 500 ml
Stripping buffer	Sodium phosphate (pH 7.4)	20 mM
	NaCl	500 mM
	EDTA	50 mM
	H_2O	to 1 L
Nickel solution	$NiSO_4$	100 mM
Coomassie Brilliant Blue	Brilliant Blue 250	2 g
	H_2O	450 ml
	Methanol	450 ml
	Acetic Acid	100 ml
4× Protein loading buffer	Tris (1 M, pH 6.8)	5 ml
	87% Glycerol	4 ml
	Bromophenol blue	10 mg
	SDS	1 g
	H_2O	to 10 ml
	DTT 500 mM	800 μl
	The DTT is added directly before use.	
PBS buffer	1 PBS tablet	
	H_2O	to 500 ml
Wash buffer	Tetrasodium pyrophosphate	50 ml
	Perchloric acid	19.2 ml
	H_2O	to 500 ml

2.4 Media

2.4.1 Media for *E. coli* cultivation

Table 2.6: Media for *E. coli* cultivation (for agar plates 16 g/L agar were added to the medium)

Medium	Constituents	
LB medium	Tryptone	10 g
	Yeast extract	5 g
	NaCl	5 g
	H_2O	to 1 L
2YT medium	Tryptone	16 g
	Yeast extract	10 g
	NaCl	5 g
	H_2O	to 1L
Freezing solution	$MgSO_4$ x 7 H_2O	760 mg
	Na-Citrat x 2 H_2O	4.5 g
	NH42 SO4	9 g
	Glycerol	440 g
	H_2O	to 800 ml
	Add after autoclaving:	
	K_2HPO_4	47 g
	KH_2HPO_4	18 g
	H_2O	to 200ml

2.4.2 Media for cultivation of pseudomonads

Table 2.7: Media for cultivation of pseudomonads

Medium	Constituents	
LB medium	Tryptone	10 g
	Yeast extract	5 g
	NaCl	5 g
	H_2O	to 1 L
PMC medium	K_2HPO_4	6 g
	KH_2PO_4	5 g
	$NH_4(SO_4)_2$	1 g
	Na-Citrate	8.8 g
	$MgSO_4$	800 mM
	H_2O	to 1 L

2.4.3 Media for cultivation of streptomycetes

Table 2.8: Media for cultivation of streptomycetes

Medium	Constituents	
Tryptone soy broth medium (TSB)	Tryptone soy broth	30 g
	H_2O	to 1L
Yeast extract-malt extract medium (YEME)	Yeast extract	3 g
	Peptone	5 g
	Malt extract	3 g
	Glucose	10 g
	Sucrose	340 g
	H_2O	to 1L
Soya mannitol medium (SM)	Mannitol	20 g
	Soya flour	20 g
	H_2O	to 1L
INA5 medium	Glycerol	30 g
	Soya flour	15 g
	$CaCO_3$	5 g
	NaCl	2 g
	H_2O	to 1L
T6 medium	Glycerol	45 g
	Soya flour	25 g
	$CaCO_3$	2g
	H_2O	to 1 L
R2YE medium	Sucrose	51.1 g
	K_2SO_4	0.125 g
	$MgCl_2.6\,H_2O$	5.06 g
	Glucose	5 g
	Casaminoacids	0.05 g
	H_2O	to 400 ml
	Add after autoclaving:	
	KH_2PO_4 (0.5%)	5 ml
	$CaCl_2.2H_2O$ (3.68%)	40 ml
	L-Proline (2 %)	7.5 ml
	TES-Puffer (5.73 %, pH 7.2)	50 ml
	Trace element solution [10]	1 ml
	1 N NaOH	2.5 ml
	Yeast extract (10%)	25 ml

Medium	Constituents	
V6 medium	Glucose	20 g
	Meat extract	5 g
	Yeast extract	5 g
	Peptone	5 g
	Casein	3 g
	NaCl	1.5 g
		to 1L
	pH 7.5	
NBG	Peptone	10 g
	Meat extract	3 g
	Glucose	10 g
	NaCl	5 g
	H_2O	to 1L
NL111	Meat extract	20 g
	$CaCO_3$	10 g
	Malt extract	100 g
	H_2O	to 1L
	pH 7.2	

2.5 Antibiotics

Table 2.9: Antibiotics used in this work

Antibiotic	Stock solution	End concentration
Ampicillin	100 mg/ml	100 µg/ml
Chloramphenicol	34 mg/ml	34 µg/ml
Kanamycin	60 mg/ml	60 µg/ml
Tetracyline	12.5 mg/ml	12.5 µg/ml
Zeocin	100 mg/ml	25µg/ml
Apramycin	60 mg/ml	60 µg/ml
Nalidixic acid	25 mg/ml	25 µg/ml

2.6   Instruments and materials

- **Agarose gel electrophoresis**

Electrophoresis chamber SUB-CELL® GT	(Biorad)
Electrophoresis chamber MINI-SUB-CELL® GT	(Biorad)

- **Centrifuges**

Cool centrifugee 5805R	(Eppendorf)
Tabletop centrifuge 5415D	(Eppendorf)
Centrifuge Avanti JE	(Beckmann Coulter)
Biofuge Pico	(Heraeus)

- **Electroporator**

Gene pulser XCell	(Biorad)
Electroporation cuvette 0.1 cm	(Biorad)

- **Electro blotter**

Trans-Blot SD Semi Dry Transfer Cell	(Biorad)

- **French Press**   (SLM Aminco)

- **HPLC**

DAD-coupled HPLC	(Dionex)

- **HPLC-MS**

Agilent 110 series HPLC system	(Agilent)
Bruker HCTplus	(Bruker)
Bruker micrOTOF	(Bruker)
LTQ-Orbitrap	(Thermo Finnigan)

- **Image documentation**

N-1000 Darkroom	(Peqlab)
Camera biovision	(Peqlab)

- **Incubators**

Incubators	(Binder)
Hybridization oven APT Line Series BFED	(Binder)
Multitron Shakers	(Infors)

- **Laminar flow**

HeraeusLaminAir®	(Kendro)
Hera Safe	(Kendro)

- **NMR**

  Bruker Advance 500 (Bruker)

- **PCR**

  Mastercycler Gradient (Eppendorf)

  PCR Sprint Thermocycler (Thermo Electron Corp)

- **pH-measurements**

  pH-Meter 766 Calimatic (Knick)

- **Photometer**

  Helios Epsilon Spectrophotometer (Thermo)

- **Protein purification**

  Äkta Prime system (GEHealthcare)

- **TLC**

  Alugram®SL G/UV$_{254}$ TLC Platten (Macherey & Nagel)

- **Thermomixer**

  Thermomixer compact (Eppendorf)

  Thermomixer comfort (Eppendorf)

- **Sterilisation**

  Systec VX-150 (Systec)

- **Ultrasonic bath**

  Sonarex (Bandelin)

- **Water processing**

  Milli-Q water purification system (Millipore)

  PURELAB *ultra* (ELGA)

- **Others materials**

  Membrane filter 0.22 μm (Millipore)

  Nylon membrane, positively charged (Roche)

  Microspin Columns (Amersham Pharmacia)

2.7 Bacterial strains, oligonucleotides and plasmids

2.7.1 Bacterial strains

Table 2.10: Bacterial strains used in this work

Strain	Genotype/characteristics	Reference
E.coli DH10B	F-*mcrA*Δ(*mrr-hds*RMS-*mcr*BC) *Ø80dlacZ*Δ*M15*Δ*lacX74 deoR recA1 endA1 araD 139*Δ *(ara,leu) 7696 galU galK rpsL nupG*	[68]
E. coli HS 996	*F-mcrA*Δ*(mrr-hdsRMS-mcrBC) Ø80dlacZ*Δ*M15*Δ*lacX74 deoR recA1 endA1 araD 139*Δ *(ara,leu) 7696 galU galK rpsL nupG fhuA::IS*	Invitrogen
E. coli ET 12567/pUZ8002	*dam13::Tn9 dcm-6 hsdM hsdR recF143 zjj-201::Tn10 galK2 galT22 ara14 lacY1 xyl-5 leuB6 thi-1 tonA31 rpsL136 hisG4 tsx-78 mtlI glnV44,* carrying the oriT mobilisating plasmid pUZ8002	[10]
E. coli GB2005	HS996 Δ*recT* and *red*α	Genebridges
E. coli GB2005 red	GB2005 insertion of the *pBAD recE recT γ recA* operon	Genebridges
E. coli SCS110	*rpsL (Strr) thr leu endA thi-1 lacY galK galT ara tonA tsx dam dcm supE44 Δ(lac-proAB) [F′ traD36 proAB lacIqZΔM15]*	Stratagene
E. coli BL21(DE3)	*F⁻ ompT gal dcm lon hsdS_B(r_B⁻ m_B⁻) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*	[69]
E. coli Rosetta BL21 (DE3) pLysS/RARE	*F⁻ ompT hsd S_B(R_B⁻ m_B⁻) gal dcm λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])* pLysSRARE (Cm^R)	Novagen [69]
E.coli SURE	*endA1 glnV44 thi-1 gyrA96 relA1 lac recB recJ sbcC umuC::Tn5 uvrC e14⁻ Δ(mcrCB-hsdSMR-mrr)171 F′[proAB⁺ lacI^q lacZΔM15 Tn10(tet^R)]*	Stratagene [70]
*P. putida KT*2440	wild type	[71]
S. lividans TK24	Derived from *S. lividans* TK66	[72]
S. coelicolor A3(2)	wild type	[73]
S. coelicolor M512	Δ*redD*, Δ*actII-ORF4* SCP1⁻SCP2⁻	[74]
*S. Tü*6071	phenalinolactone producer	[66]
S. 14386	GE81112 producer	[75]

2.7.2 Oligonucleotides

2.7.2.1 Primers used in the phenalinolactone project

Table 2.11: Primers used in the phenalinolactone project, DST= desalted

Primer	Sequence (5′→3′)	Purification
PhlET6	CGAGCGGGGATGCGCCAACTTCGCCCGTTCTTCTTCCTAA GGGCGAATTCGGATCTGATCAGCACGTGTT	HPLC
PhlET7	GTAGAGGATCTCTAGATTAATTGGGAGTGATTTCCCTTGTT TAAAGGATCTCAGTCCTGCTCCTCGGCCA	HPLC
PhlET10	GTAGAGGATCTCTAGATTAATTGGGAGTGATTTCCCTTGTT TAAAGGATCAGTACTCCACTAGTGTGAATTGTAATACGAC TCACTATAGGGCGATTTAAAGGATCCGGCCAGCCTCGCAG AGCAG	HPLC
PhlET11	GGGCGCTCTGGTTGTGGAGCAGGTGCTCGTACCAGTGGCC GACGACGTACTTACTTGGCGGAAGTCTGCT	HPLC
PhlET18	CAGCGCGATGGTGTACGAGAGGCGGTCGTCCGCCCATGCG TGAGCTGAAGGCTGCTCTA GTACACCTGA	HPLC
PhlET19	CGGGGGATGGGATTGCGTAGGTCTTGGGTTGCTGCCGTCT CGGCATGCACGTCGTTGCTGTATCTCCAG	HPLC
PhlET20	TCTCGGATCTGGTCGTGCGAAGGCGCTCCGACAACGCCGT TTGAAAAGGGGCTGCTCTA GTACACCTGA	HPLC
PhlET21	CGAATGCGCCGGATCCGTTGTTGCTGTTCCTCGGTGAGGTC AAACTCCACGTCGTTGCTGTATCTCCAG	HPLC
PhlET22	GAGGCCACGCTGGGGGCCCACCGCCACATCTGCTGAAGAG AATGGTCGGGGAAGTTCCTATTCTCTAGAAAGTATAGGAA CTTCTGCAGCAGCACGTGTTGAC	HPLC
PhlET23	TCCGGCTGCACGACGGCCTCGTCGGCGTGGCGGACGCGCT CTCTTCTTGAGAAGTTCCTATACTTTCTAGAGAATAGGAAC TTCAAGCGCCGGTTCGTACGGCG	HPLC
PhlET28	CGGCCCCCAGAAGTGAGGGTACTGCCCGGCGAACAGGGG TGATTTAAATCCTGGTGTCCCTG	HPLC
PhlET29	AGAACGCCAACGGGATCGACTGGCTCTGAATTCTCCGTGA CCATGCCCATCAGAACCTTCCTACCAACCGGCACGATTGT GCCCACAACAGCATCTTACGCCCCGCCCTGCCACTC	HPLC
Apra_for	GTGCAATACGAATGGCGAAA	DST
Apra_rev	TCAGCCAATCGACTGGCGAG	DST
ermE CPhl19_for	GCCAGGCTTCTTGCCCATCT	DST
ermE CPhl19_rev	ACGTGCGGCGTCCGCGCCGA	DST

2.7.2.2 **Primers used in the GE81112 project**

Table 2.12: Primers used in the GE81112 project

Primer	Sequence (5′→3′)	Purification
Cyclo_14386_for	AGCGCCAGGAGAGACCCTTG	DST
Cyclo_14386_rev	GCGTGCAACTGGGTGACGGC	DST
Cyclo Probe_for	CCCCGGGCCGGAACGCCATG	DST
Cyclo Probe_rev	GAACGCCATGACCATGCGGG	DST
Tubz_up	GGAGGTGGCCGTGCAGAGGATGTC	DST
TubZ_down	CTGCACGCGCTGATGGATGAGGTC	DST
RapL_up	GCGCGTCGGGCATCGGTGTG	DST
RapL_down	CCTTGGGCAGCTCGGTCTTG	DST
Cyclo KO_for	ACTGCAGAATTCGCGACGAACGCCCGCACCG	DST
CycloKO_rev	TACTGAGAATTCGGTGCCTACGCCGGGCTC	DST
NRPS-A1-up	CGGCTCCACCGGCACNCCNAARGGNG	DST
NRPS-H1-dn	CGGCCGAGGTCGCCNGTNCKRTA	DST
RevA3	CCTCCGGSCCSACCGGSMCGCCSAAGG	DST
PSLGG 2000	GCCGCCSAGSCYGAAGAA	DST
BI11-T7end-2-for	TGGAGCGTCAGGCGCATCCC	DST
BI11-T7end-2-rev	GGACCGTCTTCACCGAGGCC	DST
pGEX_Cyclo_for_EcoRI	CCTTGGGAATTCACAAGCCCTAACACTGAGCG	DST
pGEX_Cyclo_for_NotI	GCCCGTGCGGCCGCTCACCGCCGCGCTCCCTG	DST
pET28_Cyclo_for_NdeI	CCTTGGCATATGACAAGCCCTAACACTGAGCG	DST
pET28_Cyclo_for_BamHI	GCCCGTGGATCCTCACCGCCGCGCTCCCTG	DST
BI11A1_pip_for	CCGTACATATGCTCGTTCCTGCCGGAACCGG	DST
BI11A1_pip_rev	CGCGGGGATCCCTACCGCGAGGCGTTCGAGT	DST
BI11A2_for	GCCGCCATATGTCTCCCCGGCACCCGGTCGC	DST
BI11A2_rev	GGTCAGGATCCCCGGCGCGGGGCGGGGTGCC	DST
BI11A3_for	CCGACCATATGCCCTCGGTCCCGGTCGGCAG	DST
BI11A3_rev	ACCGGGGATCCACGATCGAGCACCCGCGACC	DST
FD10_KO_for	GCCAGTCAGCGACGGCCAGTGCGAGCGCGGC GAGCAGCACGTCGTTGACAATTTAAATCCTGG TGTCCCT	HPLC
FD10_KO_rev	GCGTATGTGATTTATACGTCGGGTTCGACGGG GCGGCCGAAGGGTGTGGTTTACGCCCCGCCCT GCCACT	HPLC
Sonde BI11 SEQ1	CATGTTCGAGCTGTCCTTCC	DST
Sonde BI11 SEQ2	CAGGTCGGCCACCTCCAACG	DST
Sonde BI11 SEQ3	GCTGGACGCGAACCTGACGG	DST

Primer	Sequence (5′→3′)	Purification
Sonde BI11 SEQ4	GCGCCCTGATCTCCTCCACGG	DST
Sonde BI11 SEQ5	TCGGGCAGTGCGTCGAGCATG	DST
Sonde BI11 SEQ6	GCCGTCCGCCGGAGACCTTG	DST
Sonde BI11 SEQ7	CATCAGCTCCCGGGTCGTGGT	DST
Sonde BI11 SEQ7_rev	GTCAGCGCCACGACCGACGAC	DST
Sonde BI11 SEQ6_rev	ATGATCGCGAAGCGCGGCTGA	DST
Sonde BI11 SEQ5_rev	TCGGCGCGAGCGGCACCGGCC	DST
Sonde BI11 SEQ4_rev	AGTTCTGGCGGGCCACCGAGG	DST
Sonde BI11 SEQ 3_rev	CCGCGTGCTGCAACTGCCTTC	DST
Sonde BI11 SEQ 2_rev	GGCGCGGGCGTTCGTCGCCGA	DST
T7	TAATACGACTCACTATAGGGC	DST
T4	CACCTGTGGCGCCGGTGATG	DST
LacZ1	CTTGGGCTGCAGGTCGAC	DST
LacZ2	GTGTGGAATTGTGAGCGG	DST

2.7.3 Plasmids

2.7.3.1 General plasmids

Table 2.13: General plasmids used in this work

Plasmids	Relevant characteristics	Reference
pSC101-BAD-γβαA-Apra	ET cloning vector	Genebridges
pSC101-BAD-γβαA-Amp	ET cloning vector	Genebridges
pCR2.1-TOPO	Cloning vector, *lacZα, t7 , f1ori, neoR, bla, pUC origin*	Invitrogen
pKC1132	Knockout vector for *Streptomyces, oriT* RK2, *lacZα, aac(3)IV*	[10]
pKC1139	Knockout vector for *Streptomyces*, *oriT* RK2 *lacZα, aac(3)IV, ori* psG5	[10]
pRK2013	*Tn903*, RK2	[76]
pGEX-6-P-1	Expression vector, for expression with N-terminal GST Tag	GE Healthcare
pET28b(+)	Expression vector, for expression with N-terminal His Tag	Novagen
pOJ436	Cosmid vector, *oriT* RK2 *aac(3)IV*, λ*cos attP* ØC31	[10]

2.7.3.2 **Plasmids used in the phenalinolactone project**

Table 2.14: Plasmids used in the phenalinolactone project

Plasmids	Vector	Relevant characteristics
CPhl7	pOJ436	ET cloning product, Cos 3-1O12 + zeocin resistance gene, $apra^R$, zeo^R
CPhl8	pOJ436	ET cloning product, complete PL biosynthetic gene cluster containing the zeocin resistance gene and the *oriT-tet-trpE* cassette, $apra^R$, tet^R, zeo^R
CPhl9	pOJ436	ET cloning product, Cos 10-4D08 + *oriT-tet-trpE* cassette, tet^R, $apra^R$
CPhl10	pOJ436	ET cloning product, CPhl8 + P_m-promotor in front of the gene *plaM2*, tet^R, $apra^R$,cmR
CPhl13	pOJ436	ET cloning product, complete PL biosynthetic gene cluster containing *loxP-kan*-cassette, *apraR, tetR, zero, kanR*
CPhl14	pOJ436	ET cloning product, complete PL biosynthetic gene cluster, knockout of *plaP5, apraR, tetR, zero, kanR*
CPhl15	pOJ436	ET cloning product, complete Phl biosynthetic gene cluster containing *loxP-kan*-cassette, *apraR, tetR, zero, kanR*
CPhl16	pOJ436	ET cloning product, complete PL biosynthetic gene cluster, knockout of *plaP2, apraR, tetR, zero, kanR*
CPhl17	pOJ436	ET cloning product, complete PL biosynthetic gene cluster containing *frt-kan*-cassette, *apraR, tetR, zero, kanR*
CPhl18	pOJ436	ET cloning product, complete PL biosynthetic gene cluster, knockout of *plaA6, apraR, tetR, zero, kanR*
CPhl19	pOJ436	ET cloning product, complete PL biosynthetic gene cluster containing bidirectional ermE promoter in front of the genes *plaM2* and *plaO2*
Cos3-1O12	pOJ436	Cosmid containing part of the PL biosynthetic gene cluster, $apra^R$
Cos10-4D08	pOJ436	Cosmid containing part of the PL biosynthetic gene cluster, $apra^R$

2.7.3.3 **Plasmids used in the GE81112 project**

Table 2.15: Plasmids used in the GE81112 project

Plasmids	Vector	Relevant characteristics
Cos FD10	pOJ436	Cosmid from *Streptomyces* 14386 cosmid library, pOJ436 backbone, *apra^R*
Cos AI6	pOJ436	Cosmid from *Streptomyces* 14386 cosmid library, pOJ436 backbone, *apra^R*
Cos BI11	pOJ436	Cosmid from *Streptomyces* 14386 cosmid library, pOJ436 backbone, *apra^R*
Cos BA23	pOJ436	Cosmid from *Streptomyces* 14386 cosmid library, pOJ436 backbone, *apra^R*
Cos EB22	pOJ436	Cosmid from *Streptomyces* 14386 cosmid library, pOJ436 backbone, *apra^R*
pKC1132_FD10	pKC1132	pKC1132 derivative containing an internal 1.3 kb fragment of NRPS from FD10, *apra^R*
pKC1132_AI6	pKC1132	pKC1132 derivative containing an internal 1.3 kb fragment of NRPS from AI6, *apra^R*
pKC1132_BI11 _PipA	pKC1132	pKC1139 derivative containing an internal 0.6 kb fragment of pipecolic acid incorporating A-domain from BI11, *apra^R*
pKC1139_FD10	pKC1139	pKC1132 derivative containing an internal 1.3 kb fragment of NRPS FD10, *apra^R*
pKC1139_AI6	pKC1139	pKC1132 derivative containing an internal 1.3 kb fragment of NRPS AI6, *apra^R*
pKC1139_ BI11_PipA	pKC1139	pKC1132 derivative containing an internal 0.6 kb fragment of pipecolic acid incorporating A-domain from BI11, *apra^R*
BI11-PipA-Topo	pCR2.1-TOPO	Topo derivative containing an internal 0.6 kb fragment of pipecolic acid incorporating A-domain from BI11, *apra^R*
Cyclo-pGEX	pGEX6-P-1	pGEX-6P-1 derivative for expression of the cyclodeaminase gene with N-terminal GST Tag, *amp^R*
BI11A1	pET28b(+)	pET28b derivative for expression of BI11 A1 with N-terminal His tag*, kan^R*
BI11A2	pET28b(+)	pET28b derivative for expression of BI11 A2 with N-terminal His tag, *kan^R*
BI11A3	pET28b(+)	pET28b derivative for expression of BI11 A3 with N-terminal His tag, *kan^R*

2.8   Streptomycetes

2.8.1 Mutants of *S. lividans* TK24

Table 2.16: Mutants of *S. lividans* TK24 generated in this work

Mutant	Relevant characteristics
S. lividans TK24::CPhl8	Integration of CPhl8 at the *attP* site for heterologous expression, *apra*[R]
S. lividans TK24::CPhl14	Integration of CPhl14 at the *attP* site for heterologous expression, *apra*[R]
S. lividans TK24::CPhl16	Integration of CPhl16 at the *attP* site for heterologous expression, *apra*[R]
S. lividans TK24::CPhl18	Integration of CPhl18at the *attP* site for heterologous expression, *apra*[R]
S. lividans TK24::CPhl19	Integration of CPhl19 at the *attP* site for heterologous expression, *apra*[R]
S. lividans TK24::FD10	Integration of FD10 at the *attP* site for heterologous expression, *apra*[R]
S. lividans TK24::BI11	Integration of BI11 at the *attP* site for heterologous expression, *apra*[R]
S. lividans TK24::BA23	Integration of BA23 at the *attP* site for heterologous expression, *apra*[R]

2.8.2 Mutants of *S. coelicolor* A3(2)

Table 2.17: Mutants of *S. coelicolor* A3(2) generated in this work

Mutant	Relevant characteristics
S .coelicolor A3(2)::CPhl8	Integration of CPhl8 at the *attP* site for heterologous expression, *apra*[R]
S .coelicolor A3(2)::CPhl14	Integration of CPhl14 at the *attP* site for heterologous expression, *apra*[R]
S .coelicolor A3(2)::CPhl16	Integration of CPhl16 at the *attP* site for heterologous expression, *apra*[R]
S .coelicolor A3(2)::CPhl18	Integration of CPhl18 at the *attP* site for heterologous expression, *apra*[R]
S .coelicolor A3(2)::CPhl19	Integration of CPhl19 at the *attP* site for heterologous expression, *apra*[R]

2.8.3 Mutants of *S. coelicolor* M512

Table 2.18: Mutants of *S. coelicolor* M512 generated in this work

Mutant	Relevant characteristics
S .coelicolor M512::CPhl8	Integration of CPhl8 at the *attP* site for heterologous expression, *apra*^R
S .coelicolor M512::CPhl14	Integration of CPhl14 at the *attP* site for heterologous expression, *apra*^R
S .coelicolor M512::CPhl16	Integration of CPhl16 at the *attP* site for heterologous expression, *apra*^R
S .coelicolor M512::CPhl18	Integration of CPhl18 at the *attP* site for heterologous expression, *apra*^R
S .coelicolor M512::CPhl19	Integration of CPhl19 at the *attP* site for heterologous expression, *apra*^R

2.8.4 Mutants of *S.* 14386

Table 2.19: Mutants of *S.* 14386 generated in this work

Mutant	Relevant characteristics
S. 14386::pKC1132_BI11_PipA	Insertion of pKC1132_BI11_PipA in *S.* 14386 to create a knockout of the GE81112 cluster

2.9 Cultivation and conservation of strains

2.9.1 *E. coli/Pseudomonas putida*

3 ml of sterile LB medium were inoculated with a few cells from a single colony and cultivated at 30 °C (*Pseudomonas*) or 37 °C (*E. coli*) at 180 rpm overnight. For storage at −80 °C, 1 ml of the culture was mixed with 1 ml of 50% glycerol. To restart cultivation, a few cells from the frozen cultures were inoculated into fresh LB medium with an inoculating loop. Antibiotics were added to the growth medium as appropriate.

2.9.2 Streptomycetes

2.9.2.1 Spores

1 ml of sterile 20% glycerol was added to a fresh well-sporulating plate. The surface of the culture was scraped vigorously with an inoculating loop to resuspend the spores. The crude suspension was then poured into a sterile falcon tube and diluted with 20% glycerol to obtain

a homogenous suspension. 1 ml of this spore suspension was then aliquoted into 1.5 ml eppendorf tubes and frozen at –20 °C.

2.9.2.2 Mycelia

A square from a well-sporulating plate was used for inoculation of 30 or 50 ml medium (typically TSB) as preculture. The primary growth cultures were inoculated 1:100 with the preculture, and incubated at 28 °C or 30 °C at 180 rpm. For storage at –80 °C, 1 ml of the culture was mixed with 1 ml of 50% glycerol. To restart cultivation, a few cells were inoculated on fresh agar plates.

2.10 Isolation of prokaryotic DNA

2.10.1 Isolation of genomic DNA

Genomic streptomycetes DNA was isolated using the salting out procedure [10]. Mycelia from a 30 ml culture were harvested at 3000 rpm for 10 min. The mycelia were resuspended in 5 ml SET buffer, and then 100 µl lysozyme solution (1 mg/ml lysozyme in P-buffer) was added. The mixture was incubated at 37 °C for 30–60 min. 140 µl proteinase K solution (20 mg/ml in H_2O) was then added, and then the solution was mixed by inversion. 600 µl of 10% SDS was added and again the sample was mixed by inversion. The mixture was incubated at 55 °C for 2 h. 2 ml 5M NaCl were added, and again mixed by inversion. After cooling to 37 °C, 5 ml chloroform were added, and the sample was mixed by inversion for 30 min at 20 °C. The sample was then centrifuged for 15 min at 3000 rpm at room temperature. The supernatant was transferred to a fresh tube, 0.6 vol isopropanol were added, and then the sample was mixed by inversion. After ca. 3 min, the DNA could be spooled onto a sealed Pasteur pipette. The DNA was rinsed in ca. 2 ml 70% ethanol, air dried and dissolved in 200 µl H_2O by heating at 55 °C.

2.10.2 Isolation of plasmid and cosmid DNA by alkaline lysis

An overnight culture of 1–3 ml of *E. coli* was harvested by centrifugation in 1.5 ml eppendorf tubes. The supernatant was removed and the pellet was resuspended in 250 µl buffer P1 + RNAse (100µg/ml). The cells were then lysed in 250 µl P2 (NaOH/SDS). The lysate was then neutralised by adding 250 µl potassium acetate. The samples were then centrifuged for 10 min at 13000 rpm. The plasmid DNA remained in the supernatant was precipitated with 600 µl

isopropanol. After 10 min of centrifugation at 13000 rpm, the DNA pellets were washed with 700 µl 70% EtOH, air dried, and then dissolved in 20 µl H_2O.

2.10.3 Isolation of plasmid DNA with the NucleoSpin® Plasmid Kit

Isolation of plasmid DNA with the NucleoSpin® Plasmid Kit was performed according to the manufacturer's instructions.

2.11   Separation and purification of DNA

A 0.8% agarose gel containing ethidium bromide (EtBr, ~0.25 µg/ml) was prepared and the DNA samples, prepared in 6× DNA loading dye, were loaded into the slots of the gel. A size standard (GeneRuler^{TM} 1 kb DNA ladder, Fermentas) was used to determine the length of DNA samples from PCR or specific fragments from restrictions. Electrophoresis was performed in 1× TBE buffer with a voltage of approximately 10 V per cm gel length until the dye front reached the edge of the gel. DNA was visualized under UV light (245 nm) using a gel documentation (peqlab).

2.11.1 Extraction of DNA from agarose gels

After separation by gel electrophoresis, desired DNA fragments were excised from the gel under UV light. Extraction from the gel was carried out with the NucleoSpin® Extract Kit, according to the manufacturer's protocol.

2.11.2 DNA precipitation

DNA dissolved in H_2O was mixed with 1/10 vol 3M NaOAc solution. After addition of 1 volume isopropanol, the sample was incubated 20 min at –20 °C. After centrifugation for 10 min at 13000 rpm, the pellet was washed with 700 µl 70% EtOH, air dried and redissolved in the desired volume of H_2O.

2.11.3 Phenol/chloroform extraction

An equal volume of phenol/chloroform was added to the DNA sample, followed by vortexing. After centrifugation the upper aqueous phase was separated cleanly by decanting. Phenol was quantitatively removed by subsequent precipitation with isopropanol (see part 2.10.2).

2.12 Polymerase chain reaction

For polymerase chain reaction or PCR different polymerases were used which are listed in Table 2.20.

2.12.1 Composition of standard PCR setup

Different annealing temperatures were tested for the PCR reactions

Table 2.20: PCR protocols

		Phusion-Polymerase	Triple Master Polymerase	Hot Star Taq Polymerase
Buffers		4 µl 5× GC buffer or 5× HF buffer	5 µl 10× Tuning buffer	2.5 µl 10× HST buffer
MgCl$_2$ (25 mM)		–	–	1.5 µl
dNTPs (1.2 mM/dNTP)		4 µl	8 µl	4 µl
DMSO		0.6 µl	–	0.75 µl
Glycerol (50%)		–	–	4 µl
Primers 1/2 (50 pmol/µl)		0.5 µl	0.5 µl	0.5 µl
Polymerase		0.2 µl	0.3 µl	0.1 µl
Template		0.5 µl	0.5 µl	0.5 µl
Sterile water		10.2 µl	32.5 µl	10.65 µl
Total volume		20 µl	50 µl	25 µl
Primary denaturation	Temp.	98 °C	95 °C	95 °C
	Time	30 min	3 min	15 min
Denaturation	Temp.	98 °C	95 °C	94 °C
	Time	8 s	1 min	20 s
Annealing	Temp.	60 °C	58 °C	58 °C
	Time	30 s	1 min	30 s
Extension	Temp.	72 °C	72 °C	72 °C
	Time	0.25 min/kb	1 min/kb	1 min/kb
Final extension	Temp.	72 °C	72 °C	72 °C
	Time	10 min	10 min	10 min
Cycles		32	30	30

2.12.2 Colony PCR

Cells were removed from an agar plate with an inoculating loop or 10 µl of cell suspension was taken from a liquid culture and transferred into an eppendorf tube containing 10 µl H_2O. The cell mixture was then boiled at 95 °C for 10 min in an eppendorf thermomixer. 1 µl of this mixture was used as template in the PCR reaction. Hot Star Taq Polymerase was used for all colony PCRs following the protocol in Table 2.20.

2.13 Enzymatic manipulation of DNA

2.13.1 Restriction endonucleases

Reactions were performed in a total volume of 20 µl containing 2 µl 10×buffer, 5 µl DNA, 0.5µl restriction enzyme and 12.5 µl H_2O. The restriction mix was incubated for 2–4 hours or overnight at 37 °C (or at other incubation temperatures recommended by the manufacturer). All restriction enzymes and corresponding buffers were purchased from Fermentas GmbH.

2.13.2 Ligation

Ligations were carried out in a total reaction volume of 10 µl, containing 1 µl vector, 7 µl insert DNA, 1 µl 10×ligation buffer and 1 µl T4 DNA ligase. The reaction was incubated overnight at 16 °C, and then heat-inactivated for 20 min at 65 °C (Thermomixer comfort) following verification of the successful ligation by agarose gel electrophoresis.

2.13.3 Dephosphorylation

Dephosphorylation was carried out within the restriction reaction. 1 µl shrimp alkaline phosphatase (SAP) was added to the restriction digestion. Incubation was carried out at 37 °C for 1 h. Heat-inactivation of the enzyme was then performed for 15 min at 65 °C.

2.13.4 Cloning with TOPO-TA-Cloning kit

Direct cloning of PCR products was carried out using the TOPO TA Cloning® kit (Invitrogen). TOPO TA Cloning® provides a highly efficient, 5-minute, one-step cloning strategy ("TOPO® Cloning") for the direct insertion of Taq polymerase-amplified PCR products into a plasmid vector. Taq polymerase has a nontemplate-dependent terminal transferase activity that adds a single deoxyadenosine (A) to the 3´ ends of PCR products. The linearized vector supplied in this kit has single, overhanging 3´ deoxythymidine (T) residues.

This allows PCR inserts to ligate efficiently with the vector. If PCR reactions were carried out with other then Taq polymerase, addition of 3' A-overhangs is needed. All reactions were performed according to the manufacturer's protocol.

2.14 Red/ET recombination

For Red/ET recombination the target cosmid/plasmid was first electroporated into competent cells containing the Red/ET catalyzing plasmid GB2005/pSC101-BAD-γβαA-amp. Selection was carried out using ampicillin (50 µg/ml) and the appropriate antibiotic from the incoming target construct. Cells were grown at 30 °C overnight and used as preculture. 1.4 ml LB medium were inoculated with 30 µl of the cells and cells grown at 30 °C until an $OD_{600} = 0.2$ was reached. Expression of Red/ET proteins was induced by adding 25 µl of 10% L-arabinose. Further incubation was carried out at 37 °C until an $OD_{600} = 0.4$ was obtained. Cells were then harvested by centrifugation for 1 min at 13000 rpm and washed twice with 1 ml chilled H_2O. After the last washing step, the cells were resuspended in 30 µl H_2O and electroporated with purified PCR product. Electropoartion conditions were as follows: 1350 V, 10 µF, 600 Ohms. After electroporation, 1 ml pre-warmed LB was added and the cells were cultivated for 1 h at 37 °C without antibiotic. The cells were then plated out on LB-agar containing the appropriate antibiotics. Recombination products were checked by restriction digest with several enzymes (selected on the basis of *in silico* digests of the expected products).

2.15 Transformation of bacteria

2.15.1 Electroporation of *E. coli*

To obtain electrocompetent cells, cells were grown at 37 °C to an $OD_{600} = 0.6–0.8$ in 1.5 ml Eppendorf tubes. The cell suspension was then centrifuged and washed twice with chilled H_2O. Cells were then resuspendend in 30 µl H_2O and used for electroporation with a 0.1 cm cuvette. Electroporation parameters were as follows: 1250 V, 10 µF, 600 Ohms, and the expected time constant was 5 ms. After electroporation, 1 ml of pre-warmed LB was added and the cells were cultivated for 1 h at 37 °C without antibiotic. The cells were then plated out on LB-agar containing the appropriate antibiotics.

2.15.2 Conjugation of streptomycetes with *E. coli*

Electrocompetent cells of *E. coli* ET12567/pUZ8002 grown in the presence of kanamycin (kan) (25 µg/ml) and chloramphenicol (cm) (25 µg/ml) to maintain selection for pUZ8002, were prepared as described previously (2.15.1). The competent cells were transformed with the *oriT*-containing vector, and selected for the incoming plasmid only. A colony was inoculated in 1 ml of LB-medium containing cm and kan and the antibiotic used to select for the *oriT*-containing plasmid, and grown overnight. The overnight culture was diluted 1:100 in fresh LB containing antibiotics, and grown at 37 °C to an OD_{600} of 0.4–0.6. The cells were washed twice with an equal volume of LB and resuspended in 0.1 volume of LB. During the washing of the *E. coli* cells, for each conjugation approximately 10^8 *Streptomyces* spores were added to 500 µl 2YT broth, heat shocked at 50 °C for 10 min and then cooled. 500 µl *E. coli* cells were then added to 500 µl of the heat-shocked spores or mycelial fragments. The mixture was spun briefly and most of the supernatant was poured off. The pellet was then resuspended in the residual liquid. The cells were plated out on SM agar plates without antibiotic. After 16–20 h incubation at 30 °C, cells were scraped from the plate and resuspended in 500 µl LB medium. The mixture was then plated in different dilutions on SM agar plates containing nalidixic acid (25 µg/ml) and apramycin (60 µg/ml). Incubation at 30 °C was continued until the appearance of potential exconjugants.

2.15.3 Protoplast transformation

25 ml YEME medium were inoculated with 0.1 ml spore suspension and incubated for 36–40 h in a 30 °C shaker at 180 rpm. Cultures of *S. lividans* and *S. coelicolor* were poured into a 50 ml falcon tube and centrifuged for 10 min at 1000×g. The supernatant was discarded and the pellet was resuspended in 15 ml of 10.3% sucrose. After 10 min of centrifugation, the supernatant was again discarded and the washing step was repeated. The mycelia were resuspended in 4 ml lysozyme solution (1 mg/ml in P-buffer) and incubated at 30 °C for 15–60 min. 5 ml P-buffer were added and the protoplasts were sedimented gently by spinning at 1000×g for 7 min. The supernatant was discarded and the protoplasts were resuspended in 1 ml P-buffer. 300 µl PEG (sterile filtered) were mixed with 900 µl P-buffer. 400 µl of this solution were mixed with 100 µl of the freshly-prepared protoplasts and 10 µl DNA. Dilutions of this mixture were plated out on R2YE agar plates, and the plates incubated at 30 °C. After 16–20 h, the plates were overlaid with 1 ml H_2O containing 1 mg/ml apramycin. Resistant colonies typically appeared after 3 days.

2.16 Construction of a cosmid library from *Streptomyces* 14386

This method was used to clone large genomic DNA fragments (35–45 kb) into cosmid vectors [77] using phage transfection.

2.16.1 Preparation of the ligation reaction

Genomic DNA of *Streptomyces* 14386 (see part 2.10.1) was partially digested with *Sau*3A. After separation by gel electrophoresis, fragments showing an average size between 35 and 45 kb were pooled and extracted with phenol/chloroform (2.11.3) and precipitated. The DNA was subsequently dephosphorylated (2.10.3) and extracted with phenol/chloroform again. The aqueous phase was later precipitated together with the vector pOJ436.

The *E. coli/Streptomyces* shuttle vector pOJ436 was digested with the restriction endonuclease *Pvu*II and subsequently dephosphorylated with SAP. The DNA was extracted with phenol/chloroform and precipitated. The vector was subsequently digested with *Bam*HI, extracted with phenol/chloroform again and precipitated together with the partially-digested genomic DNA.

2.16.2 Ligation and phage infection

The precipitated DNA pellet from 2.16.1 was resuspended in 15 µl H_2O and ligated at 16 °C (2.13.2). The DNA was precipitated again and resuspended in 20 µl H_2O. 4 µl from the ligation reaction were mixed with the GigapackIII Gold Packaging Extract (Stratagene), incubated for 1.5 h at room temperature and subsequently mixed with 500 µl SM-buffer and 50 µl chloroform. The mixture was centrifuged briefly and stored at 4 °C.

In parallel *E. coli* SURE was grown in LB medium supplemented with 10 mM Mg_2SO_4 and 0.2% maltose, until an $OD_{600} = 0.7$ was reached. Various amounts of packaging extract were then mixed with *E. coli* SURE as follows:

2 µl packaging extract + 23 µl SM buffer + 25 µl *E. coli* SURE

5 µl packaging extract + 20 µl SM buffer + 25 µl *E. coli* SURE

10 µl packaging extract + 15 µl SM buffer + 25 µl *E. coli* SURE

The mixtures were incubated for 30 min at 37 °C. 200 µl LB were then added and incubation was continued at 37 °C while shaking 2 h at 165 rpm. The cells were plated out on LB agar containing 60 µg/ml apramycin. The titer of the packaging extract was determined through counting the colonies. Typically, 2 µl of packaging extract yielded 2275 colonies.

2.16.3 Generation of the cosmid library

To obtain a 10-fold coverage from an estimated genome size of 8 Mbp, ca. 2000 cosmid clones were needed. In order to ensure that enough clones were obtained, 5 µl packaging extract were used and prepared as described in 2.16.2. After incubation, the cells were plated on LB agar containing 60 µg/ml apramycin in 22 cm × 22 cm NUNC plates (Nalgene). The plates were incubated at 32 °C overnight and the colonies were picked using the Qbot-robot (Gentrix) in the Department of Genome Analysis at the Helmholtz-Centre for Infection Research in Braunschweig in six 384-well microtiterplates containing 50 µl 2YT medium (60 µg/ml apramycin).

2.16.4 Robotically produced high-density colony arrays

The plates were incubated at 37 °C overnight and the growing colonies were spotted with the Qbot-robot (Gentrix) on fresh NUNC-LB-agar plates (60 µg/ml apramycin) which were covered with a 22 cm × 22 cm nylon membrane. 50 µl freezing solution (Table 2.6) was subsequently added to each well of the microtiter plates were subsequently amended with 50 µl freezing solution (2.4.1) and frozen at -80 °C.

.

Altogether 12 membranes were spotted with the complete cosmid library. To enable an explicit assignment of the transferred colonies to the six 384er microtiterplates each colony of the microtiterplates was spotted twice in a definend 4x4 array (Figure 2.1).
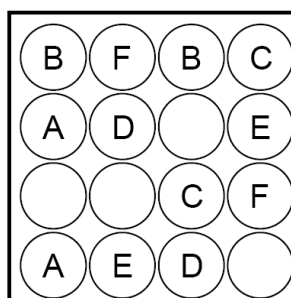
Figure 2.1: Spotting array of the high-density colony filters.

The NUNC-LB-agar plates with the nylon membranes were incubated at 32 °C for 16–20h. For the subsequent lysis of the cells, the membranes were processed as follows: Whatman paper was soaked with denaturation solution, neutralization solution and 2× SSC solution (Table 2.4). The membranes (bottom side) were transferred onto the Whatman paper, initially using denaturation solution. After 15 min incubation, the membranes were briefly transferred

to dry Whatman paper in order to remove remaining denaturation solution. Subsequent incubation in neutralization solution (15 min) and 2× SSC (10 min) was carried out, followed by drying of the membranes on fresh Whatman paper. The colonies were fixed on the membranes through incubation at 80 °C for 2 h. The membranes were treated with 14 ml Proteinase K solution (2 mg/ml) for 1 h at 37 °C. To remove cell debris, the membranes were treated with H_2O soaked tissues. After drying at room temperature, the processed membranes could be used for hybridization.

2.17 DNA Hybridization

2.17.1 Southern Blot

5 μg genomic DNA were digested with different restriction enzymes (2.13.1) and separated by gel electrophoresis (2.11). After gel documentation, the gel was depurinated for 10 min in 0.25 M HCl, with gentle shaking. The gel was subsequently incubated twice for 20 min in denaturation buffer (Table 2.4) and neutralization buffer (Table 2.4) at room temperature. After treatment of the gel, the DNA was blotted onto a membrane by vacuum blot.

2.17.2 Vacuum Blot

The membrane placed onto Whatman paper soaked with 10× SSC (Table 2.4) on the vacuum blotter, in order to prevent the formation of air bubbles. A plastic mask was placed on the membrane leaving a window for placing the gel onto it. The gel was placed onto the membrane and the vacuum pump was switched on to immobilize the gel onto the membrane. The vacuum blot chamber was filled with 10× SSC and blotting was carried out for 90 min, with a vacuum of 5 mm Hg.

2.17.3 Generation of DIG-labeled probes

For the generation of DIG-labeled probes the 'DIG PCR labeling mix' from Roche was used. Digoxigenin labeled desoxynucleotides were employed in a standard PCR reaction (2.12). The obtained probe was stored at –20 °C, until it was used for hybridization.

2.17.4 Hybridization and detection with the DIG labeling system

The DIG systems protocol from Roche was used. The membrane was pre-hybridized in hybridization buffer for 2 h at 42 °C. The DIG labeled probe was denaturated for 10 min at 95

°C and directly cooled on ice. The probe was then added to the hybridization buffer to obtain a final concentration of DIG labeled probe of 7.5–15 ng/ml. The pre-hybridization solution was removed and fresh hybridization buffer containing the DIG-labeled probe was added to the membrane. Hybridization was performed at 42 °C overnight. Non stringent hybridization (heterologous) was performed at 38–40 °C. After hybridization, the membrane was washed twice with 2× SSC washing buffer (Table 2.4) at room temperature, followed by stringent washing with 0.5× SSC (Table 2.4) at 68 °C (58 °C for colony hybridization). The membrane was subsequently equilibrated in maleic acid buffer (Table 2.4) for 10 min and then incubated with blocking solution (Table 2.4) for 30 min at room temperature. DIG antibody (Anti digoxigenin-AP conjugate, Fab fragment) was added to fresh blocking solution (1:10000) and the membrane was incubated for 30 min in this solution. The membrane was then washed twice with maleic acid buffer for 15 min. The membrane was then briefly equilibrated with detection buffer (Table 2.4) (5 min) and subsequently covered with CDP-Star solution (Roche) (1 ml/100 cm^2). After 1 min of incubation, the membrane was shrink-wrapped and the chemoluminescence was detected by gel documentation (Peqlab).

2.17.5 Stripping of membranes

Membranes were treated with stripping buffer (Table 2.4) for 20 min at room temperature. After a brief equilibration in 2× SSC, the membranes were either used again for hybridization or stored at −20 °C.

2.17.6 Colony Hybridization

Hybridization of the high-density colony filters from *S.* 14386 was carried out with the DIG system from Roche as described in 2.17.4.

2.17.7 Analysis of DNA- and protein sequences

Sequence data were analyzed with the program Vector NTI (InforMax). Additional programs for sequence analysis were freely accessible via the internet.

BLAST   http://www.ncbi.nlm.nih.gov/blast/

Frame plot   http://www.nih.go.jp/~jun/cgi-bin/frameplot.pl

PKS/NRPS Analysis   http://www.tigr.org/jravel/nrps/

ClustalW	http://www.ebi.ac.uk/clustalw/
PFAM	http://www.sanger.ac.uk/Software/Pfam/search.shtml
Npbiogene	http://www.npbiogene.com/
NRPS predictor	http://www-ab.informatik.uni-tuebingen.de/toolbox/ index.php?view=domainpred

2.18 Protein and enzymatic analysis

Different proteins from the GE biosynthetic gene cluster were overexpressed as GST- or His-tagged fusion proteins for biochemical characterization. The genes were cloned into the expression vectors pGEX-6P-1 (N-terminal GST-tag) or pET-28b(+) (N-terminal His-tag) (Table 2.13). The genes are expressed under the control of IPTG-inducible promoters (T7 promoter in the pET system, P_{tac} promoter in the pGEX system). The pET-28b(+) plasmid carries kanamycin resistance, the pGEX-6P-1 plasmid ampicillin resistance. The genes were cloned into the expression vectors, sequenced and transformed into *E. coli* BL21(DE3) or *E. coli Rosetta* BL21 (DE3) pLysS/pRARE.

2.18.1 Cultivation and cell disruption

A single colony of *E. coli* BL21 or *E. coli Rosetta* BL21 (DE3) pLysS/pRARE containing the plasmid with the expression vector construct was used for inoculation of a LB culture (with the appropriate antibiotic). This preculture was grown overnight and used to inoculate the expression culture (1:100 dilution). The culture was then grown until an OD_{600} = 0.8 at 30 °C and 180 rpm, and then induced with IPTG to a final concentration of 0.1 M. Incubation was continued at 16 °C overnight, or at 30 °C for additional 2 h. Cells were harvested by centrifugation for 10 min at 10000 rpm and were resuspended in an appropriate buffer (e.g. PBS buffer or binding buffer for Äkta-purification) (Table 2.5)). All further steps were carried out on ice or at 4 °C.

To check for the presence of soluble protein, 1 ml of the cell culture was centrifuged, the pellet was resuspended in 100 µl BugBuster® Protein Extraction Reagent (Novagen) and then the mixture was incubated for 20 min at room temperature. The cell lysate was then centrifuged for 10 min at 4 °C and 10 µl of the supernatant and the remaining pellet were analyzed by SDS-PAGE. If the protein was soluble, the residual culture was used for cell disruption at larger scale. For this, the cells were lysed using a French press at 700 psi (typically two passages). *E. coli Rosetta* BL21 (DE3) pLysS/pRARE strains were induced to

carry out self-lysis by freezing and thawing, which caused the production of lysozyme. In this case, the cell lysate was centrifuged for 45 min at 10000 rpm, and the remaining supernatant was used for further purification.

2.18.2 Protein purification

2.18.2.1 GST-tagged proteins

GST fusion proteins were purified using GST MicroSpin columns (GE Healthcare). The lower caps were removed from the columns, and each column was placed into a clean 2 ml eppendorf tube and spun for 1 min at 3500 rpm. The flow-through was discarded and the lower caps were replaced. 600 µl of the cell lysate (see part 2.18.1) were applied to the column. The columns were recapped at the top and mixed gently by inversion for 20 min at room temperature. The top and bottom caps were removed and each column was placed into a clean 2 ml eppendorf tube. The tube was centrifuged for 1 min at 3500 rpm to collect the flow-through. Each tube was placed into a new eppendorf tube and 600 µl 1× PBS wash buffer was applied to each column. This step was repeated once. 150 µl cleavage buffer containing 10 µl PreScission (GE Healthcare) protease were then added to the column and the top and bottom caps were replaced; the column was then incubated at 4 °C overnight. On the following day, the columns were placed into eppendorf tubes and the eluate containing the protein without GST tag was collected and checked using SDS-PAGE.

2.18.2.2 His-tagged proteins

Prepacked HisTrap™ HP columns were used for preparative purification of histidine-tagged recombinant proteins by immobilized metal ion affinity chromatography (IMAC) on the Äkta prime. 15 ml protein lysate were filtered through a sterile filter and loaded onto the 1 ml HisTrap column. Purification was performed as recommended in the GE Healthcare manual (HisTrap HP, Instructions 71-5027-68 AF). The following gradient was applied:

Table 2.21: **Imidazol gradient applied for Äkta system.**

Volume [ml]	Eluent A: binding buffer [%]	Eluent B: elution buffer [%]
20	100	0
40	100	0
60	100	0
70	88	12
80	80	20
90	60	40
100	40	60
110	0	100

Pressure limit: 0.55 MPa; Flow 1ml/min

To check the fractions for the presence of protein, 150 µl of each fraction were precipitated with an equal volume of acetone at –20 °C for 20 min. After centrifugation, the precipitated protein pellet was resuspended in 10 µl of 4× protein loading buffer and analyzed by SDS-PAGE.

2.18.3 Bradford assay

Bradford assays were carried out according to the manufacturer's manual (Biorad). In each case, a standard curve was measured with bovine serum albumin (BSA).

2.18.4 Protein concentration by centrifugation

Concentration of protein extracts was performed using Amicon Ultra-30 K centrifugal filters (Millipore), according to the manufacturer's instructions. The nominal molecular weight limit (NMWL) of the filters was 30,000 Da.

2.18.5 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis, is a technique to separate proteins according to their molecular weight [78]. The following proteins markers were applied for size determination:

Marker PageRuler™, Fermentas: 10, 15 25, 30, 40, 50, 60, 70, 80, 100, 120, 170 kDa

Marker PageRuler™ Prestained, Fermentas : 10, 15 25, 35, 40, 55, 70, 120, 170 kDa

2.18.5.1 Preparing of SDS gels and electrophoresis

For detection of proteins, a 10% separating and a 5% stacking polyacrylamide gel were prepared. Ca. 20 µg protein was loaded into each lane of the gel. The protein samples were mixed with 10 µl of 4× protein loading buffer and boiled for 10 min at 95 °C. After loading the protein sample and a prestained protein marker (Fermentas) onto the gel, electrophoresis was performed in 1× running buffer at 100 V. Electrophoresis was stopped when the bromophenol blue front reached the edge of the gel (approximately 1.5 h).

10% Separating gel:

H_2O	5.9 ml
30% Bis-/Acrylamide mix	5 ml
1.5 M Tris (pH 8.8)	3.8 ml
10% SDS	150 µl
10% APS	150 µl
TEMED	6 µl

The solution was poured between two glass plates, overlaid with isopropanol and allowed to polymerize.

5% Stacking gel:

H_2O	7 ml
30% Bis-/Acrylamide mix	1.7 ml
1.5 M Tris (pH 6.8)	1 ml
10% SDS	100 µl
10% APS	100 µl
TEMED	10 µl

The solution was poured onto the separating gel and allowed to polymerize.

2.18.5.2 Detection of protein bands

Protein gels were stained with Coomassie Brilliant Blue. The staining solution contained methanol/acetic acid/water (10:10:80) and 1% Coomassie Brilliant Blue R250 (Roth). Staining was carried out for 1–2 h, or was accelerated by heating the gel in the microwave for

a couple of seconds. Destaining was performed in the same mixture without Coomassie. Gel-Dry™ Drying Kit (Invitrogen) was used to conserve gels.

2.18.6 Protein identification by MALDI-MS

The desired protein band was excised from the SDS-PAGE gel and washed three times with water. 500 µl of a 50% acetonitrile (ACN) mixture in H_2O were added until the gel piece was destained. The gel piece was then air dried and 500 µl ACN were added for 10 min. The gel piece was air dried again and 20 µl trypsin were added to each sample. After complete rehydration of the gel piece, 20 µl 40 mM NH_4HCO_3 were added and the gel piece was incubated overnight. TFA was added to a final concentration of 0.1%. The sample was then stored at –20 °C prior to analysis by MALDI-MS.

2.18.7 Cyclodeaminase assay

To determine the cyclodeaminase acitivity one millimole L-or D-Lysine was incubated at 30 °C with 10 µM protein, 5 mM Tris (pH 8.0), 1 mg/ml BSA, and 100 µM NAD^+. H_2O was added at a total volume of 100 µl and the reaction was incubated for 16 h. The reaction was quenched with two reaction volumes of acetonitrile, stored at -20 °C for at least 10 min and spun at 13000 rpm for 10 min. The supernatant was vacuum-dried and solved in 200 µl borate buffer and used for derivatization.

For derivatization, the reactant solution constisted of FMOC-Cl dissolved in acetone (1.5 mg/ml). Two hundred microliters of FMOC solution was added to the 200 µl of the reaction/borate buffer mixture and incubated 10 min at room temperature. Termination of the reaction and removal of excess reagent (Fmoc-Cl), its hydrolysis product FMOC-OH, and acetone were performed by extraction with 600 µl hexane. After two additional extractions, 400 µl 25% acetonitrile in 0.1 M borate buffer was added. A 50 µl aliquot of the diluted sample was injected into the HPLC system. FMOC-derivatized compounds were separtated with a standard gradient (see section 2.10.1.). The peak corresponding to FMOC pipecolic acid was detected with a mass of $m/z = 571.2$ $[M-H]^-$.

2.18.8 ATP-PP$_i$ exchange assay

This assay was carried out in order to determine the substrate specificity of adenylation domains. Therefore 2 µM protein, 5 µl dATP (40 mM), 10 µl 10× buffer and 5 µl amino acid (40 mM) were mixed together. The reaction was started by addition of 0.1 µCi [^{32}P]-

pyrophosphate (Perkin Elmer). Incubation was done for 30 min and the reaction was stopped by the addition of 500 µl stop-mix. After centrifugation for 1 min at 13000 rpm the pellet was washed twice with 1 ml 0.1M wash buffer (Table 2.5). The charcoal pellet was then resuspended in 0.5 ml H_2O. 3.5 ml scintillation fluid was then added to the mixture and the radioactivity was determined by scintillation counting.

2.19 Analysis of secondary metabolites in natural producer strains and recombinant *Streptomyces* strains

Streptomyces strains used in this work were *Streptomyces* Tü6071 [66], *S. lividans* TK24 [72], *S. coelicolo*r A3(2) [11], *S. coelicolor* M512 (Δ*redD*, Δ*actII-ORF4* SCP1⁻SCP2⁻) [73] and *Streptomyces* 14386 [67] (Table 2.10). For production and heterologous expression of secondary metabolites different cultivation conditions were used for the different strains.

2.19.1 Cultivation in production medium

2.19.1.1 Phenalinolactones

For phenalinolactone production, the cultures were cultivated under standard conditions at 28 °C and 180 rpm on a rotary shaker, and harvested after 6–8 days by centrifugation. NL111 (Table 2.8) liquid medium was used for production. Precultures were grown in the same medium and used to inoculate production cultures (1:100 dilution). Studies on the optimization of PL production were performed under standard cultivation conditions in NL111 medium, but with growth at 37 °C and also in NBG, SM and INA5 medium. Apramycin (60 µg/ml) was used for cultivation of recombinant *Streptomyces* strains.

2.19.1.2 GE81112

Precultures were grown in V6 medium and used to inoculate production cultures (1:100 dilution). INA5 or T6 (Table 2.8) production medium was used for production of GE81112. The cultures were maintained at 30 °C and 180 rpm on a rotary shaker and harvested after 6–8 days. Apramycin (60 µg/ml) was used for cultivation of recombinant *Streptomyces* strains.

2.19.2 Feeding experiments

Precultures were grown in V6 medium at 30 °C for 3 days and then used to inoculate a 50 ml production culture in T6 medium (1:100 dilution). Stock solutions of deuterium-labeled L-

and D-pipecolic acid and uniformely labeled (U^{13}C) L-histidine (500 mM) were prepared and introduced into the cultures after 72, 96, 120, 144 and 168 h, to a final concentration of 1 mM. Cells were harvested after 196 h. As a control, the same cultivation was performed without feeding of labeled substances.

2.19.3 Extraction

After cultivation in production medium, the cells were harvested by centrifugation at 5000 rpm for 10 min. The culture supernatant was extracted twice with ethyl acetate. The solvent was then removed by evaporation, redissolved in methanol (1% of the original culture volume) and centrifuged for 10 min. The supernatant was transferred to a fresh vial. The cell pellet was extracted with a mixture of acetone/methanol (1:1). The solvent was then removed by evaporation redissolved in methanol (1% of the original culture volume) and centrifuged for 10 min. The supernatant was transferred to a fresh vial.

2.19.4 Purification of GE compounds

The GE81112 producer *S*. 14386 was cultivated at 30 °C in 5 L baffled shake flasks at 180 rpm in GE production medium (see part 2.19.1.2). 13 L were harvested after 6 days of cultivation. Cells were extracted with acetone/methanol (1:1) and the supernatant was extracted with ethyl acetate (1:1), as described in 2.19.3. HPLC-MS analysis revealed that the desired compounds could only be found in the supernatant. 13 L of supernatant were then extracted twice with ethyl acetate (1:1). After evaporation of the solvent, the extract was resuspended in methanol and extracted with heptane (1:1). The extract was then fractionated by preparative RP-HPLC (Dionex P680 HPLC equipped with a PDA-100 Photodiode Array Detector; Waters xBridge™ Prep C18 column (5µm OBD™ 19 × 100 mm)) using a stepwise gradient, (0–2 min, 10% acetonitrile; 2–30 min, 10–95% acetonitrile; 10 ml/min; detection at 280 nm; injection volume, 250 µl) coupled to a mass spectrometer (Bruker HCTplus ion trap, operating in positive-ionization mode at a scan range of *m/z* 100–1100).

Fractions were checked by HPLC-MS analysis, and combined when they contained the desired compounds. The solvent was evaporated, and the compounds were further enriched by gel permeation chromatography (Sephadex LH-20 in methanol). Pure compound was obtained by sequential RP-HPLC using a stepwise gradient (solvent B: acetonitrile containing 0.01% formic acid, 0–10 min, 30% B; 10–20 min, 30–50%; 20–25 min, 50–95% B; 25–27 min, 95%B, injection volume 200 µl).

2.19.5 Analysis of the secondary metabolites

2.19.5.1 HPLC-UV/Vis, HPLC-MS, HPLC-MS/MS, LTQ-Orbitrap-MS

10 μl of the extracts (see part 2.19.3) were analyzed by HPLC-MS analysis. Analysis was carried out using an Agilent 1100 series solvent delivery system equipped with a photodiode array detector and coupled to a Bruker HCTplus ion trap mass spectrometer. Chromatographic separation was carried out on a Nucleodur C18/3 μm RP column (125 × 2 mm; Macherey & Nagel) equipped with an C18/5 precolumn (8 × 3 mm) using a mobile phase system consisting of H_2O (A) and acetonitrile (B), each containing 0.1% formic acid. The following gradient was applied: 0–2 min, 25% B; 2–22 min linear from 25–95 %B; 22–24 min isocratic at 95% B. Detection was carried out in negative or positive ionization mode at a scan range of $m/z = 100$–1100.

Phenalinolactones were identified by comparison to the retention times (rt) and the MS data of authentic standards (phenalinolactone A: rt = 13.5 min, $[M–H]^- = 714.5$; phenalinolactone D: rt = 18.1 min, $[M–H]^- = 698.3$) in negative ionization mode. Phenalinolactone E was identified by comparison to the retention times and the MS data of phenalinolactone D (phenalinolactone E: rt = 13.0 min, $[M–H]^- = 570.3$). LC-coupled FT-Orbitrap-MS analysis was carried out with an Accella UPLC system (Thermo Electron Corporation) coupled to a LTQ Orbitrap mass spectrometer (Thermo Fisher Scientific) operating in negative or positive ionization mode at a scan range of m/z 100–2000. A Hypersil Gold column (2.1 × 50 mm; Thermo Fisher Scientific) was used for separation with a solvent system consisting of H_2O (A) and acetonitrile (B), each containing 0.1% formic acid. A gradient of 5–95 % B was applied over 10 min. Measurements were carried out in single ion mode (SIM). GE8112 compounds were identified by comparison to the retention times (rt) and the MS data of authentic standards (GE factor A: $[M+H]^+ = 644.21858$; GE B: $[M+H]^+ = 659.22953$ and GE factor B1: $[M+H]^+ = 658.2458760$) in the negative ionization mode.

2.19.5.2 NMR

The structure of the isolated compounds 631, 645 and 659 were elucidated by using 1D (^{1}H-, ^{13}C-NMR) and 2D NMR analysis (^{1}H-^{1}H-COSY, HSQC heteronuclear sequential quantum coherence, HMBC heteronuclear multiple bond correlation) on Bruker DRX 500 or Bruker Advance 500 spectrometers using several solvents ([d6]DMSO and CD_3OD). Calculation of theoretical NMR-spectra was performed with ACD/Labs 7.00 software.

3   Results

3.1   Heterologous expression and genetic engineering of the phenalinolactone biosynthetic gene cluster using Red/ET recombineering

3.1.1   Goals

The aim of the thesis was the development of heterologous expression techniques for complex secondary metabolite pathways from streptomycetes using Red/ET recombineering. The biosynthetic gene cluster which was chosen for this work was the phenalinolactone biosynthetic gene cluster from *Streptomyces* Tü6071. The gene cluster was already cloned and sequenced in the group of Andreas Bechthold from Freiburg University, from whom the strain was obtained. As the complete gene cluster was located on two different cosmids, we aimed in the first step to reconstitute the complete PL biosynthetic pathway on one expression construct using Red/ET recombineering. After the successful cloning, we planned to express the pathway in different heterologous host strains such as pseudomonads and streptomycetes. The successful heterologous expression would then open the door for generation of novel analogues by genetic engineering.

3.1.2   Reconstitution of the complete *pla* biosynthetic gene cluster

We employed Red/ET recombineering in *E. coli* [36-38] to rebuild the entire *pla* biosynthetic pathway on an integrative *E. coli-Streptomyces* shuttle vector (pOJ436), and to introduce additional genetic elements for heterologous expression in pseudomonads and streptomycetes.

3.1.2.1   Cloning strategy

The biosynthetic gene cluster comprises 11 transcriptional units harboring 35 genes. Two pOJ436-derived cosmids (Cos3-1O12 and Cos10-4D08 [66]) containing overlapping regions of the *pla* gene cluster were used as starting points. The cloning strategy is described in Figure 3.1, and consists of three different cloning steps that are each discussed in detail below.
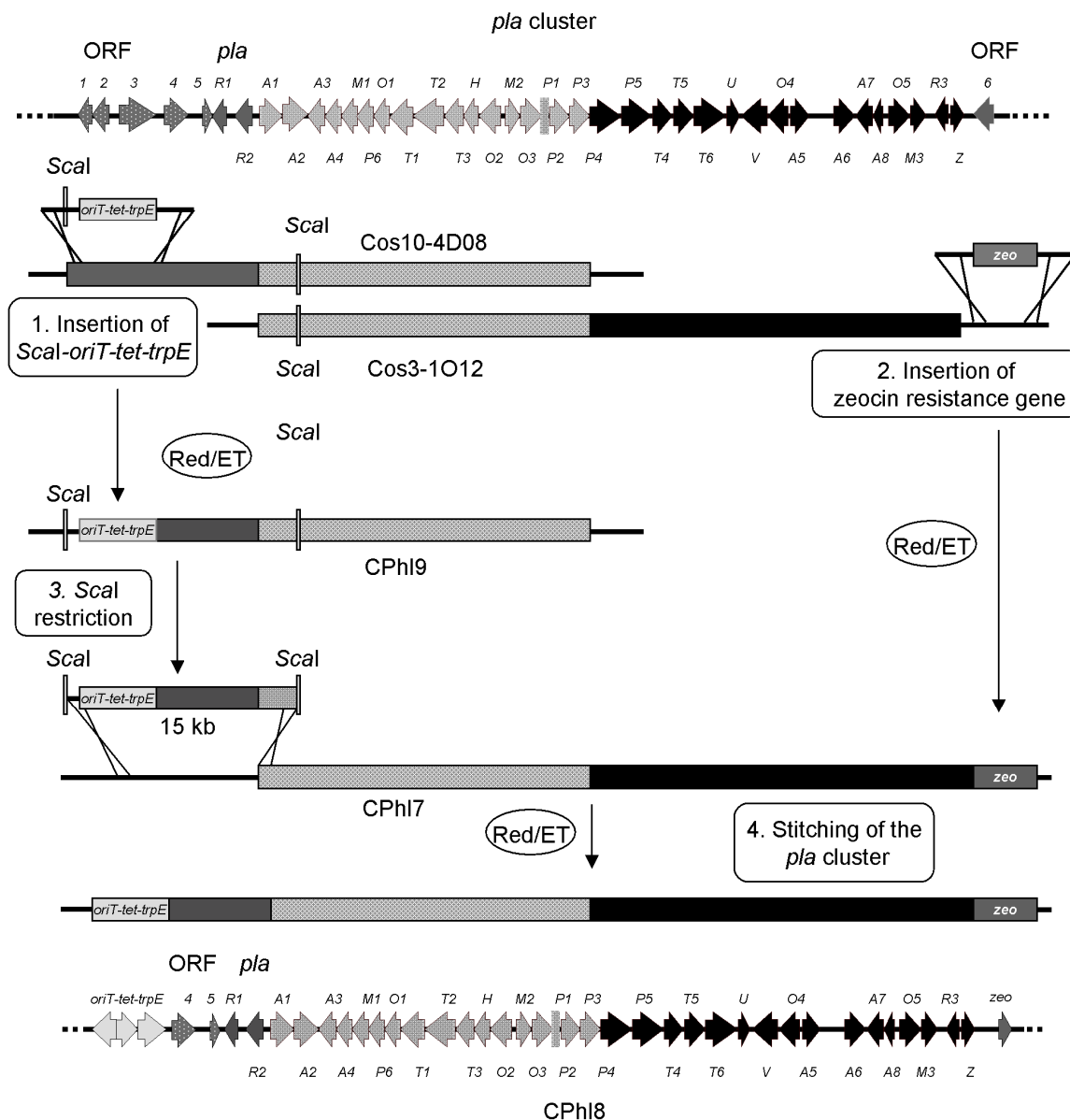
Figure 3.1: **Description of the cloning strategy.**

The phenalinolactone biosynthetic gene cluster consists of 35 orfs, and was located on two overlapping cosmids (Cos3-1O12 and Cos10-4D08) carrying the pOJ436 backbone. In the first step, a *Sca*I-*oriT-tet-trpE* cassette was inserted via Red/ET recombineering into Cos10-4D08 at the upstream end of the gene cluster, generating the construct CPhl9. In the next step, introduction of the zeocin resistance gene at the downstream end of the cluster led to the construct CPhl7. After restriction digest of CPhl9 with *Sca*I, the resulting 15 kb fragment was used to reassemble the whole cluster using Red/ET recombineering. The final expression construct containing the whole cluster was designated CPhl8.

3.1.2.2 **Cloning step 1: introduction of the *Sca*I-*oriT-tet-trpE* box at the 5′-end of the *pla* biosynthetic gene cluster**

Cos10-4D08 was modified by insertion of a gene cassette into the upstream end of the *pla* gene cluster, to generate construct CPhl9. In addition to a *Sca*I restriction site and a tetracycline (*tet*) resistance gene, the introduced cassette also contained an *oriT* as well as a portion of the *trpE* gene from *Pseudomonas putida* which should in future enable the transfer of the expression construct into pseudomonads. The *oriT* was needed for conjugation and the *trpE* gene for integration into the *Pseudomonas putida* genome. This step also resulted in the deletion of *orfs* 1–3, which we expected not to be involved in PL biosynthesis. The ca. 4 kb gene cassette was amplified by PCR from template CMch37 [35] which contained the *oriT-tet-trpE* box. The primers PhlET10 and PhlET11 that were used in the PCR contained three homology arms in total (H1, H2 and H3). Homology arms 1 and 2 were used to enable the integration of the *oriT-tet-trpE* box into Cos10-4D08, via double homologous recombination (Figure 3.2)
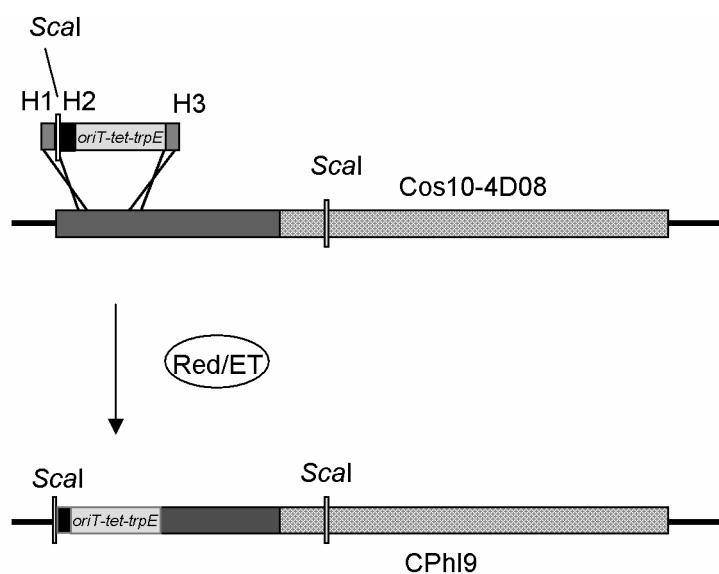
Figure 3.2:   **Red/ET Cloning step 1.**

Integration of *oriT-tet-trpE* box into cosmid 10-4D08 via double homologous recombination resulted in the construct CPhl9. H1, H2 and H3: homology arms 1, 2, and 3

Homology arm 3 and the *Sca*I restriction site were needed for the subsequent cloning steps. Selection for correct recombinant mutants was carried out on LB-agar containing tetracycline (12.5 µg/ml) and apramycin (60 µg/ml). To confirm the individual cloning steps, each of the constructs generated by recombination was digested with a diagnostic set of restriction enzymes (Figure 3.5).

3.1.2.3 Cloning step 2: introduction of the zeocin resistance gene at the 3′ end

In the next cloning step we aimed to reconstruct the whole phenalinolactone biosynthetic gene cluster on one expression construct. Therefore, construct CPhl9 was digested with *Sca*I to generate a linear fragment which could then be recombined with the target construct Cos3-1O12 containing the remaining part of the cluster. As the whole *Sca*I restriction digest was to be used for transformation, it was formally possible that there could be still some undigested construct CPhl9 in the transformation mixture. Selection on tetracycline (*oriT-tet-trpE box*) and apramycin (target molecule) would then result in mutants carrying the constructs CPhl9 and Cos3-1O12, without any recombination taking place. To avoid this problem, we inserted the zeocin resistance gene in Cos3-1O12 to create CPhl7. The introduction of a new resistance gene enabled selection on tetracycline and zeocin, forcing homologous recombination and avoiding replication of the undigested construct CPhl9 in the mutants. The zeocin resistance gene was inserted into Cos3-1O12 at the opposite end of the gene cluster (downstream of *plaZ*), an alteration which simultaneously deleted *orf6* which was not needed for phenalinolactone biosynthesis.
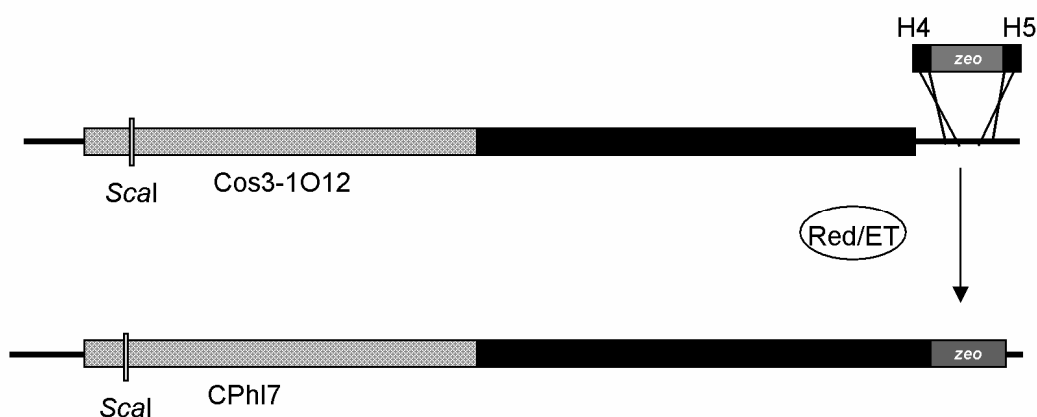
Figure 3.3:   **Red/ET Cloning step 2.**
Integration of the zeocin resistance gene into cosmid 3-1O12 via double homologous recombination resulted in the construct CPhl7. H4, H5: homology arms 4 and 5.

The ca. 600 kb zeocin cassette was amplified by PCR, again using MCh37 as template. Two homology arms (H4 and H5) were introduced with primers PhlET6 and PhlET7. Double homologous recombination via these homology arms enabled the insertion of the zeocin cassette into Cos3-1O12. Selection for correct recombinant mutants was carried out on LB-agar containing zeocin (25 µg/ml) and apramycin (60 µg/ml). To confirm the cloning step, the

construct CPhl7 generated by recombination was digested with a diagnostic set of restriction enzymes (Figure 3.5).

3.1.2.4 Cloning step 3: stitching of the *pla* biosynthetic gene cluster

Construct CPhl9 was digested with *Sca*I to generate a linear 15 kb fragment. A *Sca*I site had been inserted into the construct with the primer PhlET10 in the first cloning step described in chapter 3.1.2.2. A second *Sca*I restriction site was already present in the overlapping region of the cosmids (Figure 3.4).
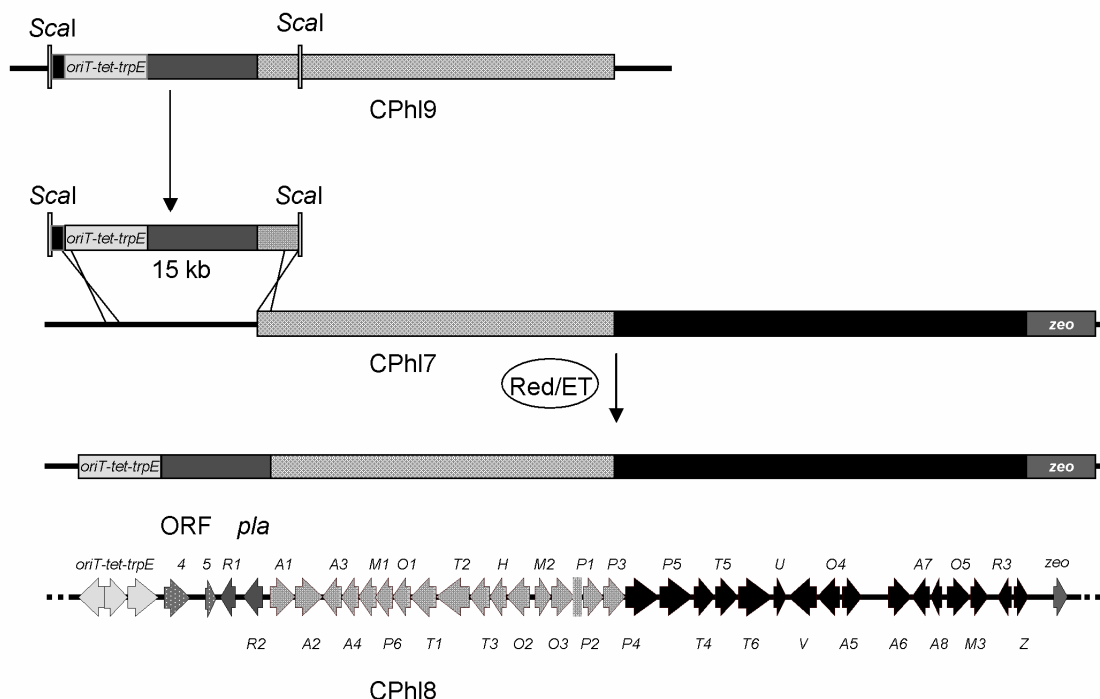
Figure 3.4:   Red/ET Cloning step 3.

Digestion of CPhl9 with *Sca*I and recombination of the 15 kb linear fragment with CPhl7, resulted in the final construct CPhl8 containing the complete *pla* biosynthetic pathway.

This 15 kb fragment was then transformed into *E. coli* GB2005/pSC101-BAD-γβαA-amp in order to achieve recombination with CPhl7. Recombination of the linear fragment with CPhl7 containing the missing end of the gene cluster resulted in the final construct CPhl8 containing the complete phenalinolactone pathway flanked by the *oriT-tet-trpE* box at the 5′-end and the zeocin resistance gene at the 3′-end of the cluster (Figure 3.4). Selection for correct recombinant mutants was carried out on LB-agar containing zeocin (25 μg/ml) and tetracycline (12.5 μg/ml). The individual cloning steps were confirmed by restriction analysis with the enzymes *EcoR*I, *Bgl*II and *Pvu*II, as shown in Figure 3.5.
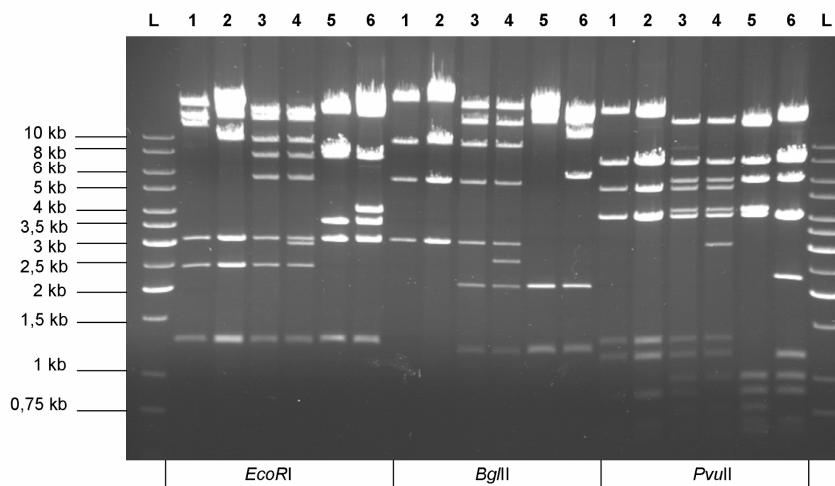
Figure 3.5: **Restriction analysis of the Red/ET recombination constructs.**

Lane 1: Cos3-1O12, 2: CPhl7, 3: CPhl8, 4: CPhl10, 5: CPhl9, 6: Cos10-4D08. Lanes 1 and 6 show the restriction pattern of the starting constructs Cos3-1O12 and Cos10-4D08. Lane 2 shows the restriction fragments of CPhl7 into which the zeocin resistance gene was inserted, and lane 5 shows the restriction pattern of CPhl9 with the introduced *oriT-tet-trpE* box. Comparing the recombination constructs with the starting constructs enabled us to confirm the identity all of the constructs. Lane 8 shows the final construct CPhl8 containing all the fragments combined from CPhl7 and CPhl9. Lane 4 shows the restriction pattern of the construct CPhl10, which contains a promoter exchange. All restriction patterns showed the expected fragments.

In summary, using this cloning strategy, we were able to reconstitute the complete phenalinolactone biosynthetic pathway on one expression construct containing the pOJ436 backbone. At this point, the expression construct was ready for transformation into various heterologous host strains.

3.1.3 Transformation of the biosynthetic gene cluster into heterologous host strains

Two genera of bacteria, *Pseudomonas* and *Streptomyces,* were selected as hosts, as they had already been proven as suitable for heterologous expression of natural products [79].

3.1.3.1 **Triparental conjugation of the expression construct CPhl8 into pseudomonads**

The final expression construct CPhl8 was transferred into *Pseudomonas putida* KT2240 by triparental mating. Exconjugants were screened on PMM agar containing tetracycline (12.5 µg/ml) for the selection of *P. putida*::CPhl8 mutants. The transformation efficiency was very high, yielding several thousand tetracycline-resistant exconjugants. Single colonies were used for inoculation of liquid cultures.

3.1.3.2 **Conjugation of the expression construct CPhl8 into streptomycetes**

The expression construct CPhl8 carried the vector backbone of the *E. coli/Streptomyces* shuttle vector pOJ436, enabling integration at the chromosomal *attP* site in streptomycetes genomes. Three different *Streptomyces* strains were used for transformation: *Streptomyces lividans* TK24, *Streptomyces coelicolor* A3(2) and *Streptomyces coelicolor* M512. The construct was transformed using biparental conjugation, and selection was carried out at 30 °C on SM agar plates containing apramycin (60 µg/ml) and nalidixic acid (25 µg/ml). In all cases, the transformation efficiency was very high, yielding several thousands apramycin-resistant exconjugants. After 3–5 days, single colonies were transferred to fresh SM agar plates containing apramycin and nalidixic acid. When sporulation was observed, the colonies were used to inoculate liquid cultures.

3.1.3.3 **Verification of the mutants**

After transformation of the expression construct CPhl8 into pseudomonads and streptomycetes, many antibiotic-resistant colonies were obtained in all cases. To verify the integration of the constructs into the genome of the different strains, primers apra_for and apra_rev were designed to amplify the apramycin resistance gene from the vector backbone of CPhl8. For the verification of the *Pseudomonas* mutants, a colony PCR was performed with 2 randomly chosen colonies. 1 µl of an overnight culture was used as template for the PCR reaction. The *Pseudomonas putida* KT2240 wild type strain was used as a negative control, while DNA from the expression construct CPhl8 was used as a positive control. PCR products with the expected size (777 bp) were obtained for all mutants and for the positive control, whereas the negative control did not yield a PCR product (Figure 3.6). These result indicated the successful integration of the expression construct into the *Pseudomonas putida* genome.
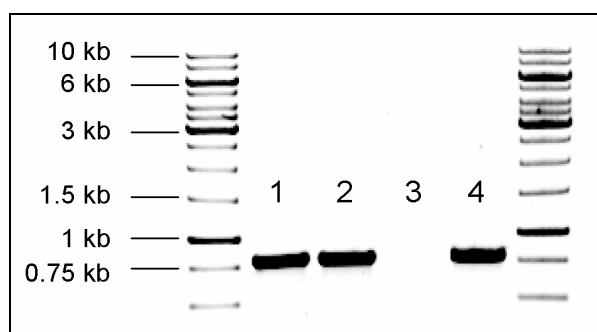
Figure 3.6:   **Verification of *P. putida* KT2240::CPhl8 mutants by colony PCR.**

Amplification of the apramycin resistance gene. Lane 1: *P. putida* KT2240::CPhl8#1; Lane 2: *P. putida* KT2240::CPhl8#2; Lane 3: *P. putida* WT; Lane 4: CPhl8 DNA.

For the verification of the *Streptomyces* mutants, genomic DNA was isolated from the mutants and from the wild type strain. A PCR was then performed using the apramycin primers to amplify the apramycin resistance gene from the vector backbone. Figure 3.7 exemplifies one PCR reaction using four randomly chosen *S. lividans* TK24 mutants and wild type as template. The expected PCR product with a size of 777 bp was present in all mutants and in the positive control (CPhl8 DNA), but not in the wild type. PCR reactions were performed for all mutants. Integration of the expression construct into the genome of the different *Streptomyces* strains was confirmed in all mutants.
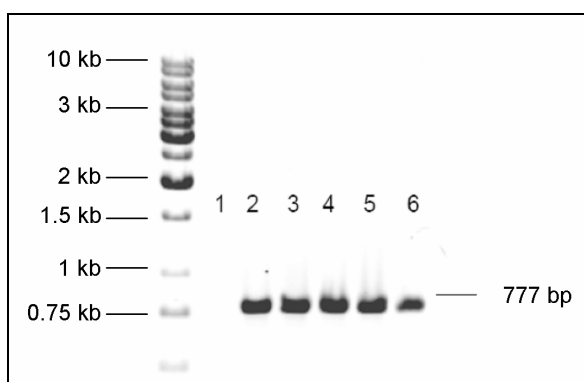
Figure 3.7:   **Verification of *S. lividans* TK24::CPhl8 mutants by PCR.**

Amplification of the apramycin resistance gene. Lane 1: *S. lividans* TK24 WT; Lanes 2–5: S. *lividans* TK24::CPhl8 mutants #1–4; Lane 6: CPhl8 DNA.

3.1.4 Heterologous expression of the phenalinolactone biosynthetic gene cluster

In order to heterologously express the phenalinolactone pathway in pseudomonads and streptomycetes, cultivation conditions had to be established. Different media and temperatures were tested, as described below.

3.1.4.1 **Cultivation of pseudomonads**

Two mutants of *P. putida* KT2240::CPhl8 were inoculated into a 50 ml culture LB medium containing 1 ml XAD and tetracycline (12.5 µg/ml). In parallel, the wild type *P. putida* KT2240 was inoculated into a 50 ml culture LB containing 1 ml XAD without antibiotic. Cultivation was carried out at 16 °C and 30 °C for 3 days. The cultures were then harvested by centrifugation and the XAD was extracted with MeOH/acetone and submitted to HPLC-MS analysis.

3.1.4.2 **Cultivation of streptomycetes**

Streptomyces lividans TK24::CPhl8, *S. coelicolor* A3(2)::CPhl8 and *S. coelicolor* M512::CPhl8 were grown in 100 ml NL111 production medium containing apramycin (60 µg/ml) at 28 °C for 6–8 days. In parallel, the corresponding wild type strains as well as the natural phenalinolactone producer Tü6071 were grown in production medium without antibiotic. The cultures were harvested by centrifugation and the supernatant was extracted with ethyl acetate and submitted to HPLC-MS analysis.

3.1.4.3 **HPLC-MS analysis**

3.1.4.3.1   Production profile of the natural producer *Streptomyces* TÜ6071

Initially the production profile of an extract of the natural producer Tü6071 was analyzed by HPLC-MS to establish HPLC conditions for the detection of phenalinolactones. Dürr *et al.* [80] demonstrated previously that production of phenalinolactones in NL111 production medium produced the highest yields. The primary metabolites produced during cultivation in NL111 medium were phenalinolactones (PLs) A and D, whereas PLs B and C were not detected [80] (Figure 3.8).

PL	R_1	R_2
A	OH	CH_3
B	OH	H
C	OH	CH_2OCH_3
D	H	CH_3

Phenalinolactone (PL)
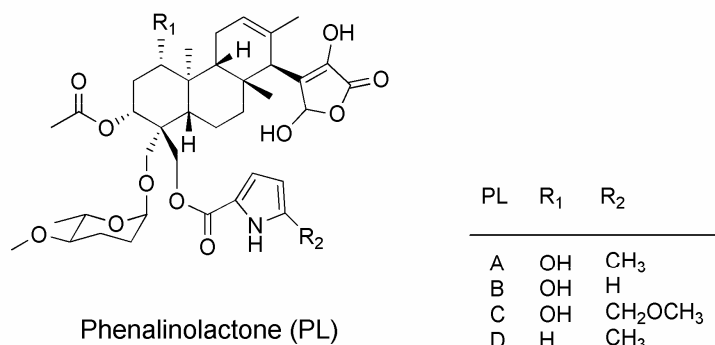
Figure 3.8:   Chemical structures of phenalinolactones A, B, C and D.

Figure 3.9 shows the typical production profile of the natural producer *Streptomyces* Tü6071. Phenalinolactones were identified by comparison to the retention times (rt) and the MS data of authentic standards. Analysis was performed in negative ionization mode enabling detection of the phenalinolactone A and D parent ions (m/z [M–H]$^-$ = 714.3 (RT: 13.8 min) and m/z [M–H]$^-$ = 698.3 (18.1 min), respectively).
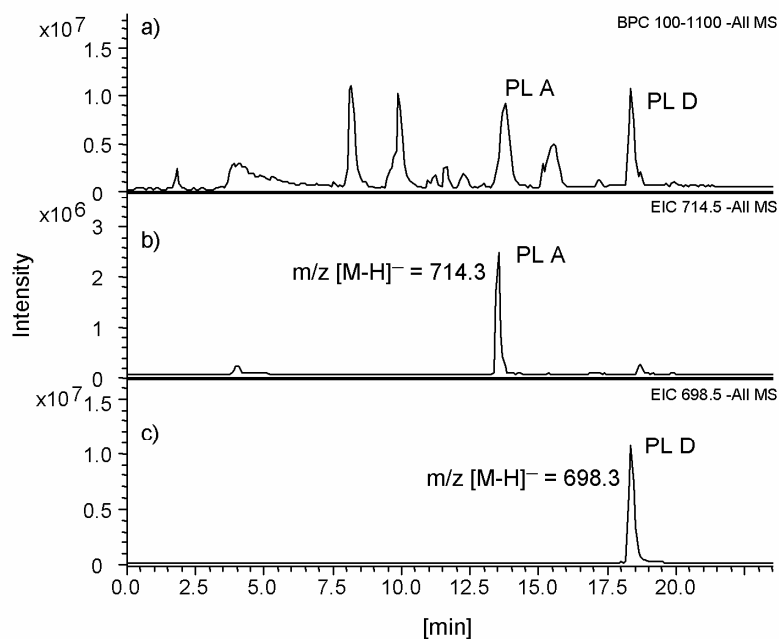
Figure 3.9: **HPLC-MS analysis of the natural producer *Streptomyces* Tü6071.**

a) Base peak chromatogram (BPC) in negative ionization mode within a mass range of 100–1100 B) Extracted ion chromatogram (EIC) of PL A with a mass of m/z [M–H]$^-$ = 714.3 c) Extracted ion chromatogram (EIC) of PL D with a mass of m/z [M–H]$^-$ = 698.3.

Fragmentation of the parent ions leads to a characteristic fragmentation pattern (MS/MS fingerprint) that could be used to unambiguously identify PL A and PL D. Figure 3.10 shows the detected parent ions of PL A and PL D, and the corresponding MS/MS fragmentation patterns. The MS/MS fragmentation pattern of PL A exhibited characteristic fragmentation ions with masses m/z [M–H]$^-$ = 582.7, 600.5, 642.4, 670.4 and 686.4. The corresponding fragments for PL D were m/z [M–H]$^-$ = 584.7, 626.6, 654.4, 670.4. On the basis of these fragmentation patterns, it was possible to clearly identify phenalinolactones in the extracts of the heterologous host strains.
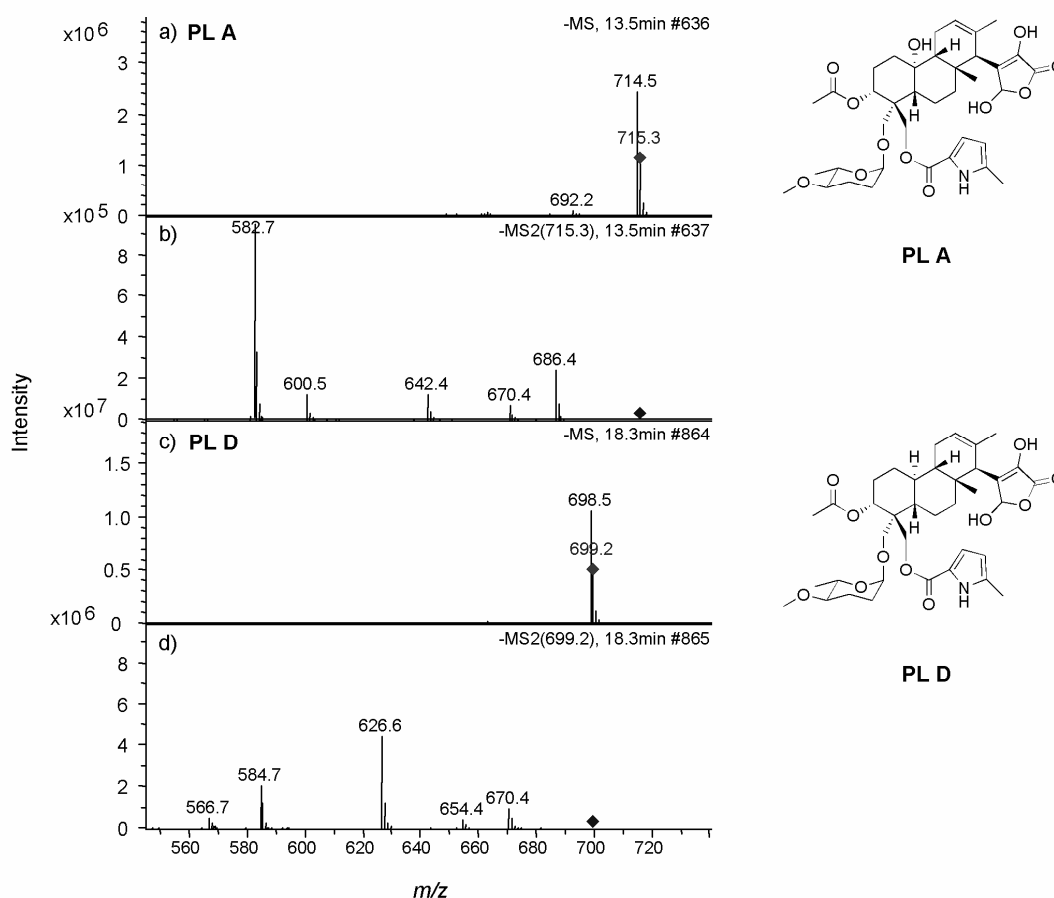
Figure 3.10:   **ESI-MS/MS fragmentation patterns of PL A and PL D.**

a) Parent ion of PL A with a mass of m/z [M–H]$^-$ = 714.3 detected at a retention time of 13.5 min. b) MS/MS fragmentation pattern of PL A showing characteristic fragmentation ions. c) Parent ion of PL D with a mass of m/z [M–H]$^-$ = 698.3 detected at a retention time of 18.3 min. d) MS/MS fragmentation pattern of PL D showing characteristic fragmentation ions.

3.1.4.3.2   Production profile of the *Pseudomonas* mutants

The basepeak chromatograms were analyzed for the presence of masses corresponding to PL A and PL D, but neither compound was detected in the extracts. Therefore, a new, higher selectivity and sensitivity HPLC-MS method was used (developed by Daniel Krug). This method detected the CID (collision-induced dissociation) fragmentation of compounds with masses 714.5 and 698.3 eluting between retention times of 12–14 min and 18–20 min, and was validated by analysis of the phenalinolactone A and D standards. However, analysis using this new method still failed to reveal the presence of phenalinolactones A and D.

3.1.4.3.3 Production profile of the *Streptomyces* mutants

We next investigated the extracts of the *Streptomyces* mutants, using the previously-described method. Phenalinolactones A and D were detected in extracts of the natural producer *Streptomyces* Tü6071, and in the recombinant *S. lividans* and *S. coelicolor* strains harboring the *pla* gene cluster. As expected no phenalinolactones were present in extracts of the corresponding wild type *S. lividans* and *S. coelicolor* strains. This result clearly demonstrated the successful heterologous production of the terpene glycosides (Figure 3.11).
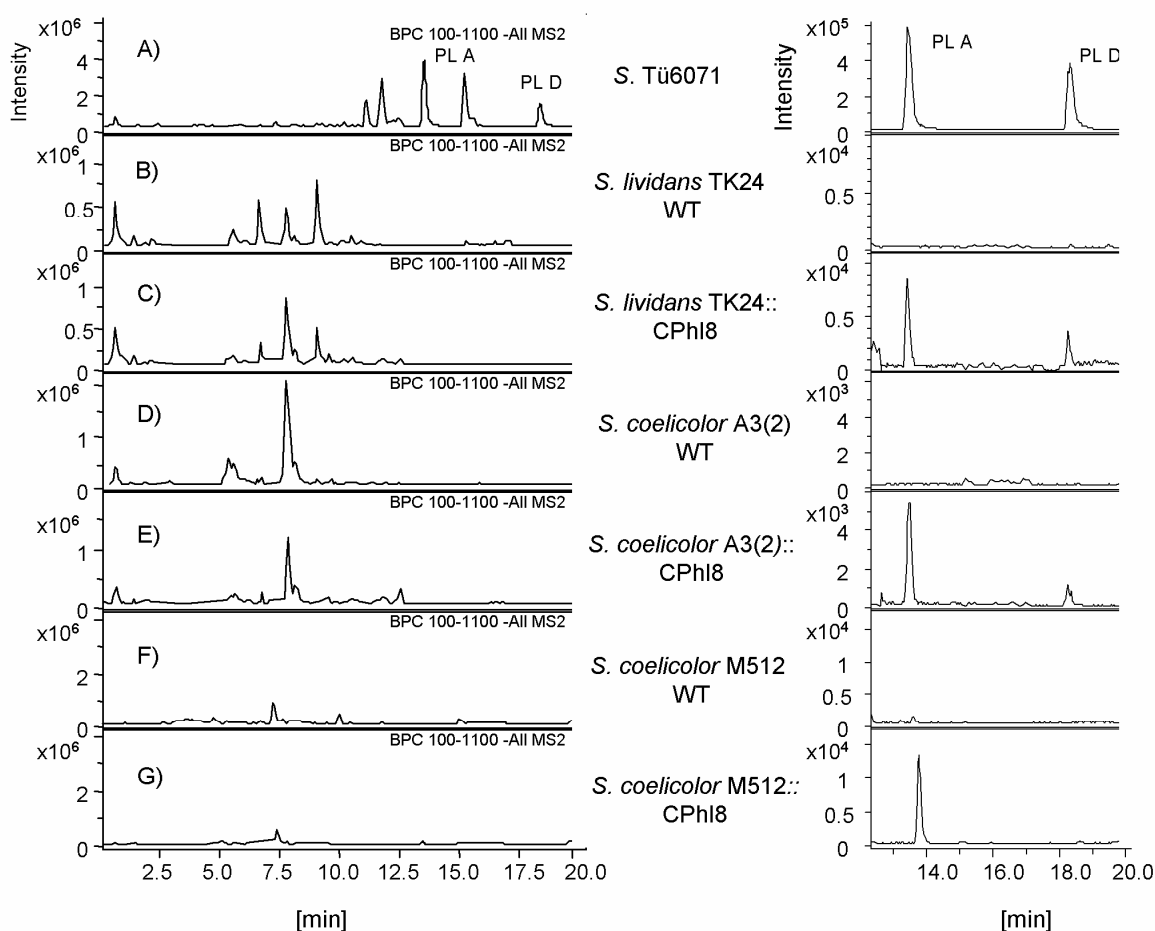
Figure 3.11: **HPLC/MS analysis of streptomycetes mutants carrying the PL pathway.**

A) BPC of natural phenalinolactone producer strain *S.* sp. Tü6071, *S. lividans* TK24 WT and *S. lividans::CPhl8*. B) Natural producer strain *S.* sp. Tü6071, *S. coelicolor* A3(2) WT and *S. coelicolor::CPhl8*. C) Natural producer strain *S.* sp. Tü6071, *S. coelicolor* M512 WT and *S. coelicolor*::CPhl8. The chromatograms show the negative ions within a mass range of 100–1100 in auto MS^n mode. CID fragmentation between retention times of 12–14 min and 16–18 min was monitored for the masses 714.5 and 698.3, corresponding to the phenalinolactone A and D standards. The section in each chromatogram between 12–18 min is shown in expanded form on the right. PL A and/or D were detected in the extracts of the recombinant strains.

3.1.4.4 **Optimization of heterologous PL production in *Pseudomonas***

3.1.4.4.1 Promoter exchange

As there was no PL detected in extracts of the *Pseudomonas* mutants harboring the *pla* biosynthetic pathway, we decided to exchange one of the native *Streptomyces* promoters against the inducible Pm-promoter. Importantly, the Pm-promoter had already been validated during the heterologous expression of the myxochromide biosynthetic pathway [35]. As the cluster is composed of 11 different transcription units (operons), we intended to introduce the Pm-promoter at a position which would drive transcription of the genes for the core biosynthetic machinery. Based on biosynthetic considerations, enzymes PlaT1−T6 comprise the machinery to biosynthesize the phenalinolactones scaffold [66]. The first precursor of PL biosynthesis is geranylgeranyl diphosphate (GGDP), which is generated by PlaT4 from 4 monomers of isopentenyl pyrophosphate (IPP) (see Discussion Figure 4.2). PlaT6 and PlaT5 are expected to be involved in IPP production. So these first three enzymes should be able to catalyze the formation of GGDP. PlaT1-PlaT3 then catalyze the formation of PL precursor 1 and 2 (shown in Figure 4.2, Discussion part). Unfortunately, the corresponding genes (*plaT1−3* and *plaT4−6*) are located in different transcription units and exhibit different orientations (Figure 3.12). Therefore, we introduced the promoter in front of the gene *plaM2* where the initial genes for biosynthesis are located (*plaT4-6*), expecting that overexpression of the early genes in the pathway would lead to the initiation of PL production or to the production of GGDP or PL precursors. For this, a cassette containing the Pm-promoter was amplified by PCR using CMch37 [35] as template, the *xylS* gene and the chloramphenicol resistance gene, as well as 50 bp homology arms to enable double homologous recombination. This cassette was introduced by Red/ET recombineering and exconjugants were selected on LB agar containing chloramphenicol (cm) (34 µg/ml) and apramycin (60 µg/ml).
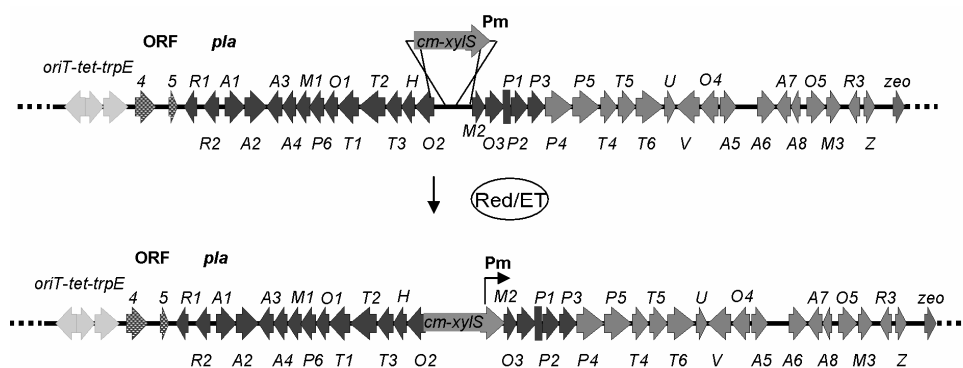
Figure 3.12: Introduction of the inducible Pm-promoter in front of *plaM2* by Red/ET recombineeering.

The new construct CPhl10 was verified by restriction analysis (Figure 3.5) and subsequently transformed into the *P. putida* KT2240 by triparental conjugation. Integration of the construct into the chromosome was confirmed by colony PCR analysis, amplifying the apramycin resistance gene as described above (3.1.2.3.). Expression was induced by adding 1 mM toluic acid to the cultures after 2 h of cultivation. We expected at least that GGDP would be expressed when the core biosynthetic genes (*plaT4-6*) were under the control of the Pm-promoter. However, HPLC-MS analysis still did not reveal production of PL precursors or PL production in extracts of the *Pseudomonas* mutants.

3.1.4.5 Optimization of heterologous PL production in *Streptomyces*

3.1.4.5.1 Changing cultivation conditions

Production of PL in the heterologous *Streptomyces* strains was quantified using authentic PL standards. Production under standard conditions (NL111 medium at 28 °C; underlined in grey, Table 3.1) was much lower in the heterologous host strains (maximum 0.13 µg/L in *S. lividans* TK24) than in the native host *S.* Tü6071 (Table 3.1). In an attempt to optimize production in the heterologous hosts, different cultivation conditions (additional media and temperatures) were tested (Table 3.1), with growth at 28 °C. Production was highest under standard conditions in NL111 production medium (Table 3.1). In INA5 and SM medium, PL production was typically lower in the heterologous host strains then in the natural producer, while the same media yielded less phenalinolactones from the native host relative to cultivation in NL111 medium. Growth in NBG medium completely abolished production by all strains. These results indicated that PL production in NL111 medium was optimal relative to cultivation in INA5, SM and NBG media.

Table 3.1: **Phenalinolactone production by *Streptomyces* Tü6071, *S. lividans* TK24, *S. coelicolor* A3(2) and *S. coelicolor* M512 during cultivation in different media at 28 °C.** Highest production yield underlined in grey.

Strain	Production of phenalinolactone (µg/L)			
	NL111	INA5	SM	NBG
S. Tü6071	500	200	200	-
S. lividans TK24::CPhl8	**0.13**	0.10	0.05	-
S. coelicolor A3(2)::CPhl8	0.01	0.01		-
S. coelicolor M512::CPhl8	0.5	0.03	0.01	-

We also evaluated cultivation at 37 °C. Under these conditions, phenalinolactone production by *S. lividans* was increased 100-fold compared to growth at 28 °C (Table 3.2). However PL production by the *S. coelicolor* mutants was completely abolished at this higher temperature.

Table 3.2: **Phenalinolactone production by *Streptomyces* Tü6071, *S. lividans* TK24, *S. coelicolor* A3(2) and *S. coelicolor* M512 during cultivation in NL111 medium at 28 °C and 37 °C.**

Strain	Production of phenalinolactone (µg/L)	
	28 °C	37 °C
S. Tü6071	500	200
S. lividans TK24::CPhl8	0.13	**10**
S. coelicolor A3(2)::CPhl8	0.01	-
S. coelicolor M512::CPhl8	0.5	-

3.1.4.5.2 Precursor feeding

We also attempted to improve production by supplementation with a putative biosynthetic precursor. Because the core terpenoid backbone in phenalinolactone biosynthesis is derived from isoprenoid building blocks [66], we administered mevalonolactone to the culture medium. Mevalonolactone is the lactonized form of mevalonate, a direct precursor in the mevalonate pathway for the construction of isoprenoid units [74]. Most *Streptomyces* strains employ the non-mevalonate pathway for the formation of isoprenoids, but some strains

additionally utilize the mevalonate pathway to produce terpenoid antibiotics [74]. However, feeding of 1 mM mevalonolactone did not increase phenalinolactone production in *Streptomyces* sp. Tü6071 nor in the heterologous host strains

3.1.4.5.3   Promoter exchange

We also evaluated the effect of exchanging a number of the native promoters against the strong constitutive *ermE* promoter [81]. We intended to introduce the promoter at the same position (between *plaO2* and *plaM2*) as in the construct for the PL expression in the pseudomonads (3.1.4.4.1), as this transcriptional unit contained the genes for the core biosynthetic gene cluster. We expected that at least GGDP or precursor 1 or 2 should be expressed (Figure 4.2, Discussion part). For this, we constructed a bidirectional promoter so that transcription in both transcription units containing the core genes would be driven (Figure 3.13). We therefore amplified the ampicillin (amp) resistance gene with primers containing 50 bp homology arms to enable homologous recombination. The *ermE* promoter was also introduced using the primers, as it is only ca. 60 bp in size. The amplified cassette was introduced by Red/ET recombineering, and recombinant mutants were selected on LB agar containing amp (100 µg/ml) and apra (60 µg/ml). The new construct CPhl19 was verified by restriction analysis, and subsequently transformed into the three heterologous host strains. In each case, integration of construct CPhl19 into the genome was verified by PCR, as described earlier (3.1.3.3.).
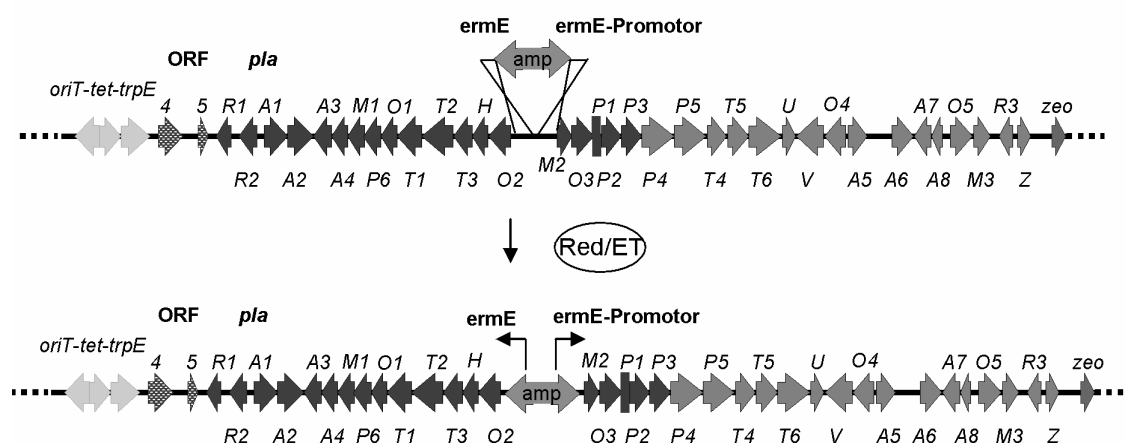
Figure 3.13:   Introduction of two *ermE* promoters between *plaO2* and *plaM2* by Red/ET recombineering.

HPLC-MS analysis of the extracts from the mutants revealed that neither PL nor any PL precursors were produced in the mutants. Instead PL production was completely abolished.

3.1.5 Genetic modification of the *pla* biosynthetic gene cluster

The established heterologous expression system enabled attempts to engineer the PL biosynthetic pathway to generate novel analogs. In theory, using Red/ET recombineering, we could perform single or multiple in-frame deletions at any desired position within the expression construct. As proof-of-principle, we aimed to modify PL biosynthesis by deletion of selected genes from the CPhl8 expression construct.

3.1.5.1 **Deletion of *plaA6***

The first aim was to modify the *pla* gene cluster by deletion of the glycosyl transferase-encoding gene *plaA6* from the CPhl8 expression construct, to generate a phenalinolactone derivative lacking the L-amicetose moiety. For this, a kanamycin resistance cassette was amplified by PCR to incorporate flanking flippase (FLP) recombinase target sites (FRT sites), as well as 50 bp homology arms to enable double homologous recombination (Figure 3.14). After replacement of the target gene (*plaA6*) by the selection marker (*kan*) using Red/ET recombineering, the *kan* gene was excised from the expression construct by FLP-recombinase-catalyzed site-specific recombination, to create a markerless in-frame deletion. The modified *pla* gene cluster was again verified by restriction analysis. The construct CPhl18 was subsequently transformed into the three heterologous *Streptomyces* strains by conjugation, and integration of the construct into the chromosome was confirmed by PCR analysis by amplifying the apramycin resistance gene, as described earlier.
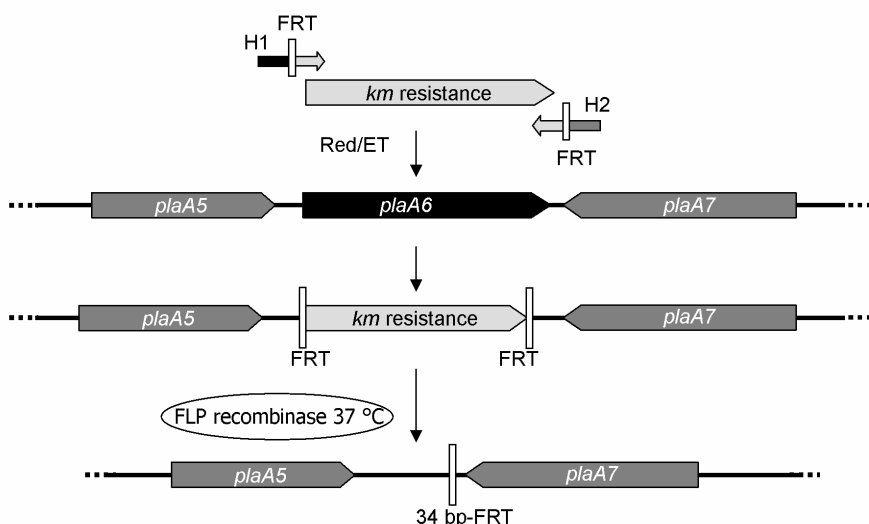
Figure 3.14: **Deletion of the glycosyl transferase encoding gene *plaA6*.**

Red/ET recombineering was used to replace *plaA6* with the kanamycin resistance cassettewhich was flanked by FRT recognition sites. Incubation with FLP recombinase at 37 °C led to a markerless deletion of *plaA6* in the *pla* cluster.

3.1.5.2 Deletion of *plaP2* and *plaP5*

By deletion of the acyltransferase encoding gene *plaP2*, we aimed to generate a derivative lacking the 5-methylpyrrole-2-carboxylic acid moiety. The gene *plaP5* is proposed to be responsible for the methylation of the pyrrole carboxylic acid [66], and thus deletion of this gene was expected to give an unmethylated PL derivative. To create these deletions, we exploited the Cre/lox system instead of the FLP/FRT system. Using the Cre system would enable us to do two deletions in parallel using the FRT system too. Therefore a kanamycin resistance cassette was amplified by PCR to incorporate Cre recombinase target sites (loxP sites), as well as 50 bp homology arms to enable double homologous recombination (analogous to Figure 3.14). After replacement of the target gene (*plaP2* or *plaP5*) by the selection marker (*kan*) using Red/ET recombineering, the *kan* gene was excised from the expression construct by Cre-recombinase-catalyzed site-specific recombination, to create a markerless in-frame deletion. The modified *pla* gene cluster was subsequently transformed into three heterologous host strains by conjugation, and integration of the construct (CPhl14 (Δ*plap2*) and CPhl16 (Δ*plap5*) into the chromosome was confirmed by PCR analysis by amplifying the ampramycin resistance gene, as described earlier.

3.1.5.3 **Verification of the new PL derivatives by high-resolution Orbitrap MS analysis**

Nine mutant strains were obtained after conjugation with the different constructs listed in Table 3.3. The production profiles of the mutant strains were analyzed by HPLC-MS, and compared to that of the respective wild type strains, and the natural phenalinolactone producer *Streptomyces* Tü6071. The expected masses of the new derivatives are also listed in Table 3.3.

Table 3.3:   *Streptomyces* **mutants obtained after conjugation with the modified constructs**

Construct	Modification in the biosynthetic pathway	Heterologous host	Sum formula of expected PL derivatives	Expected mass [M–H]⁻
CPhl14	Δ*plaP2*	*S. lividans* TK24	$C_{37}H_{50}NO_{11}$	684.3
CPhl14	Δ*plaP2*	*S. coelicolor* A3(2)	$C_{37}H_{50}NO_{11}$	684.3
CPhl14	Δ*plaP2*	*S. coelicolor* M512	$C_{37}H_{50}NO_{11}$	684.3
CPhl16	Δ*plaP5*	*S. lividans* TK24	$C_{25}H_{35}O_8$	463.3
CPhl16	Δ*plaP5*	*S. coelicolor* A3(2)	$C_{25}H_{35}O_8$	463.3
CPhl16	Δ*plaP5*	*S. coelicolor* M512	$C_{25}H_{35}O_8$	463.3
CPhl18	Δ*plaA6*	*S. lividans* TK24	$C_{31}H_{40}NO_9$	570.3
CPhl18	Δ*plaA6*	*S. coelicolor* A3(2)	$C_{31}H_{40}NO_9$	570.3
CPhl18	Δ*plaA6*	*S. coelicolor* M512	$C_{31}H_{40}NO_9$	570.3

As expected, phenalinolactones A and D (*m/z* [M–H]⁻ = 714.5 at retention time 13.8 min and *m/z* [M–H]⁻ = 698.5 at 18.1 min) were detected in the extract of the natural producer, but not in extracts of the wild type heterologous hosts. In the extract of the strain *S. coelicolor* M512::CPhl18, we detected a new compound designated as PL E (*m/z* [M–H]⁻ = 570.3 at 13.0 min) which differed from PL D by 128 atomic mass units, consistent with the anticipated loss of the 4-*O*-methyl-L-amicetose moiety (Figure 3.15). MS² analysis confirmed that the new compound had a similar fragmentation pattern to the PLs (shown in Figure 3.16).
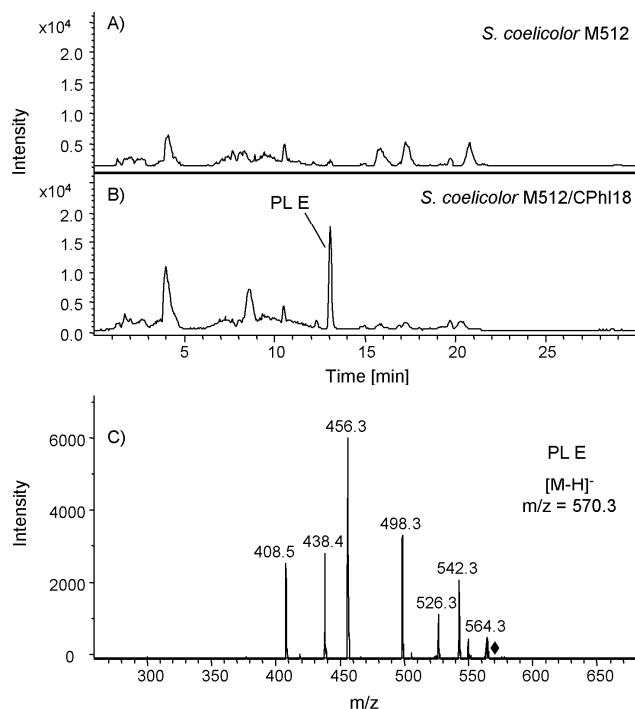
Figure 3.15:   **HPLC/MS analysis of *S. coelicolor*::CPhl18.**

A) BPC of parental strain *S. coelicolor* M512 and B) *S. coelicolor* M512::CPhl18 (*plaA6* deletion). A new compound, designated PL E, was detected in the extract of *S. coelicolor* M512::CPhl18 at a retention time of 13.0 min. C) Fragmentation pattern of PL E (molecular ion, *m/z* [M–H]$^-$ = 570.3).

The compound was not detected in extracts of the *S. lividans*::CPhl18 and *S. coelicolor*::CPhl18 mutants, nor in the extracts of the other heterologous host strains carrying the constructs CPhl14 and CPhl16. No expected derivatives could be detected in the extracts of the mutants carrying constructs CPhl14 and CPhl16 instead PL production was completely abolished in these mutants.

To verify the identity of PL E from *S. coelicolor* M512::CPhl18, we analyzed the compound by high resolution LC-coupled Fourier transform-Orbitrap MS. This analysis confirmed the expected elemental composition of $C_{31}H_{40}O_9N_1$ (measured *m/z* [M–H]$^-$ = 570.26733; calculated *m/z* [M–H]$^-$ = 570.27086; Δ = –3.52166 amu). Additional structural information was obtained using tandem-MS and subsequent comparison of the MS^2 fingerprints with those obtained for the glycosylated compound PL D. The principal collision induced dissociation (CID)-fragmentations for the standard substance PL D and the new derivative PL E were assigned as summarized in Table 3.4. Fragmentation of the terpenoid backbone as well as of the sugar and the pyrrole-carboxylic acid groups was not observed in the MS^2

spectra of PL D, indicating that despite the loss of the sugar group, the fragmentation pattern of phenalinolactones D and E should be very similar. Indeed, comparison of the fragmentation pattern of both molecules showed that they exhibited the same set of peaks, but offset by a constant mass difference corresponding to L-amicetose (Table 3.5 and Figure 3.16). The peaks are derived from the parent molecules by expulsion of CO, CO_2 and H_2O, although the precise locations of these losses remain to be elucidated. The expulsion of $C_4H_2O_4$, which occurred from both compounds, can be attributed to the γ-butyrolactone. These data strongly suggest that the PL E is the desired unglycosylated derivative of PL D.
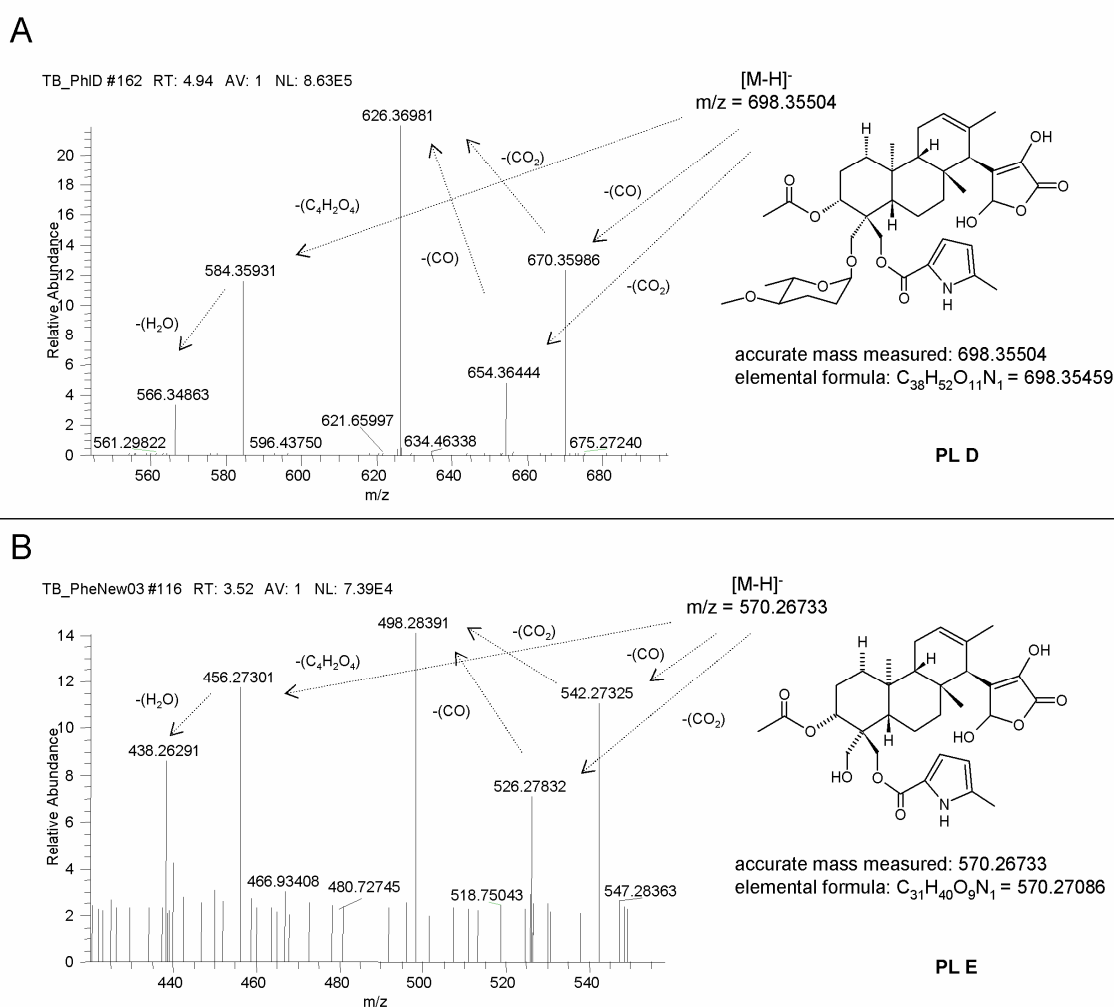
Figure 3.16: **Verification of the unglycosylated derivative PL E by LTQ-Orbitrap MS analysis.**

Accurate mass determination of the standard substance PL D (A) and the unglycosylated compound PL E (B) with high resolution LC-coupled Orbitrap-MS yielded molecular ions of m/z [M–H]⁻ = 698.35504 and m/z [M–H]⁻ = 570.26733, respectively, suggestive of the elemental formulas $C_{38}H_{52}O_{11}N_1$ and $C_{31}H_{40}O_9N_1$. Additional support for structure assignment was provided by MS² studies, which showed that PL D and PL E have the same fragmentation patterns, but with the peaks offset by a constant mass corresponding to the L-amicetose moiety missing in PL E.

Table 3.4: Principal fragments of the standard substance PL D and the unglycosylated derivative PL E

Accurate mass of fragments	Formula proposal and calculated mass	deviation ΔmDa	Neutral loss	Δm/z	Group	Theor.	ΔmDa
PL D							
698.35504							
$C_{38}H_{52}O_{11}N_1 =$							
698.35459							
0.45613 mDa							
670.35986	$C_{37}H_{52}O_{10}N_1 = 670.35967$	0.19	$698.35504 \rightarrow 670.35986$	27.99518	CO	27.99437	0.001
654.36444	$C_{37}H_{52}O_9N_1 = 654.36476$	0.32	$698.35504 \rightarrow 654.36444$	43.9906	CO_2	43.98928	0.0013
626.36981	$C_{36}H_{52}O_8N_1 = 626.36984$	0.3	$670.35986 \rightarrow 626.36981$	43.99005	CO_2	43.98928	0.0008
584.35931	$C_{34}H_{50}O_7N_1 = 584.35928$	0.04	$698.35504 \rightarrow 584.35931$	113.99573	$C_4H_2O_4$	113.99476	0.001
566.34863	$C_{34}H_{48}O_6N_1 = 566.34871$	0.08	$584.35931 \rightarrow 566.34863$	18.01068	H_2O	18.01002	0.0007
PL E							
570.26733	$C_{30}H_{40}O_8N_1 = 542.27594$	2.69	$570.26733 \rightarrow 542.27325$	27.99408	CO	27.99437	0.0003
$C_{31}H_{40}O_9N_1 =$	$C_{30}H_{40}O_7N_1 = 526.28103$	2.71	$570.26733 \rightarrow 526.27832$	43.98901	CO_2	43.98928	0.0003
570.27086	$C_{29}H_{40}O_6N_1 = 498.28611$	2.21	$542.27325 \rightarrow 498.28391$	43.98934	CO_2	43.98928	0.0001
-3.52166 mDa	$C_{27}H_{38}O_5N_1 = 456.27555$	2.54	$570.26733 \rightarrow 456.27301$	113.99432	$C_4H_2O_4$	113.99476	0.0004
	$C_{27}H_{36}O_4N_1 = 438.26428$	1.37	$456.27301 \rightarrow 438.26291$	18.0101	H_2O	18.01002	0.0001

The left-hand "Accurate mass of fragments" column also lists the fragment accurate masses: 542.27325, 526.27832, 498.28391, 456.27301, 438.26291 (PL E fragment masses).

3.2   Identification of the GE81112 biosynthetic gene cluster

3.2.1 Goals

As the GE81112 biosynthetic gene cluster had not been identified, the first objective was to clone the cluster in order to allow studies of the underlying biosynthesis. (For this, we aimed to generate a cosmid library based on the *E. coli/Streptomyces* shuttle vector pOJ436 [10]). We then hoped to use *in vitro* protein expression and targeted gene inactivation to obtain deeper insights into the biosynthesis. To support this work, we aimed to establish genetic manipulation methods for the strain, starting with optimization of production and cultivation conditions for the GE81112 compounds.

3.2.2 Production profile of *Streptomyces* 14386

The GE81112 producing *Streptomyces* strain 14386 was provided by the company KtedoGen (Italy). Initial experiments for cultivation optimization to produce GE compounds in shake flasks were carried out by Margherita Sosio (KtedoGen). Sosio tested two different media, INA5 and T6, for both cultivation of the strain and GE production. Precultures were grown in V6 medium for 96 h, and then used to inoculate INA 5 and T6 production medium (1:100). The GE81112 production results are shown in Table 3.5.

Table 3.5:   **GE81112 compounds produced under different cultivation conditions.**

Medium/cultivation time	LC/MS [M+H]$^+$
INA 5 (96 h)	ND
INA 5 (140 h)	ND
T6 (96 h)	ND
T6 (140 h)	644, 659

ND: Not Detected

Detection of GE81112 factor A and B by LC/MS was only possible after cultivation in T6 medium for 140 h. These results showed that T6 (cultivation time: 140 h) is the best medium identified to date for production of GE compounds. To reproduce these results in our lab, we cultivated the strain under the same conditions and analyzed the extracts for the presence of the GE metabolites. Overall, our results were nearly identical to those obtained by KtedoGen. For example, during cultivation in INA5 medium, GE compounds could never detected by

MS. During cultivation in T6 medium for 140 h, low amounts of the main GE compound B (*m/z* [M+H]$^+$ = 659.22953) were detected by high resolution MS analysis on an Orbitrap LTQ mass spectrometer with the sum formula $C_{24}H_{36}O_{10}N_{10}Cl_1$ (see Figure 3.17).
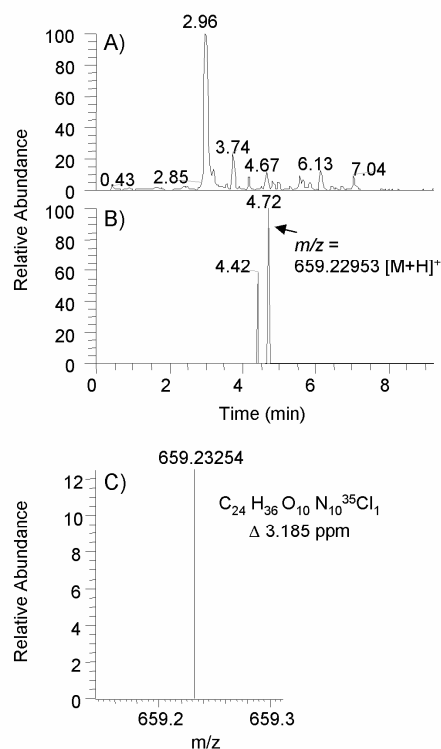
Figure 3.17: **Analysis of the extracts of strain *S*. 14386 by high-resolution MS on an LTQ Orbitrap mass spectrometer.**

A) Base peak chromatogram (BPC) with B) extracted ion chromatogram (EIC) showing a molecular ion of *m/z* [M+H]$^+$ = 659.23254, as well as C) the elemental formula.

Despite trying several different cultivation conditions, production of the GE compounds remained low or even undetectable by MS. Due to the poor yields, it was not possible to obtain an MS/MS spectrum of the substances from the culture extract. These findings are in full agreement with the results obtained by KdetoGen. Although production was low in T6 medium detection of the GE81112 B compound was reliable so that we used these cultivation conditions for future experiments.

During analysis of the culture extracts from *S*. 14386, it became apparent that the strain was producing additional metabolites. These substances first attracted attention because one of them (659) had nearly the same mass by HPLC-MS analysis as the corresponding GE81112 compound B and also incorporated chlorine as shown by the isotopic pattern (Figure 3.18).
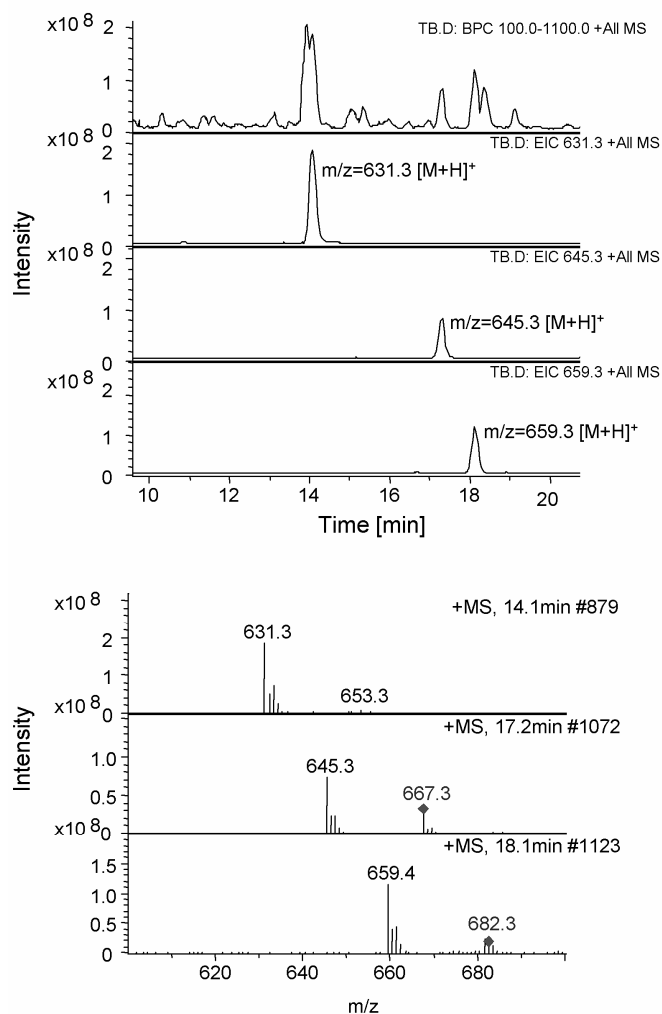
Figure 3.18: **HPLC/MS analysis of an extract of strain *S.* 14386.**

A) Base peak chromatogram (BPC) with extracted ion chromatograms (EIC) showing molecular ions of *m/z* [M+H]$^+$ = 631.3, 645.3 and 659.3. B) Mass spectrum of the three substances showing isotopic patterns characteristic of chlorinated compounds.

A detailed analysis of the extract revealed three likely related substances with masses of *m/z* [M+H]$^+$ = 631.3, 645.3, and 659.3, which all showed an isotopic pattern which is typical for chlorinated substances (Figure 3.18 B)). To obtain accurate mass information, the compounds were analyzed at high resolution using an Orbitrap LTQ mass spectrometer. Determination of the exact masses and the corresponding elemental formulae (Figure 3.19) confirmed that the compounds differed by 14 mass units from each other. Correspondingly, they share the same overall fragmentation patterns, but with peaks offset from each other by 14 mass units.
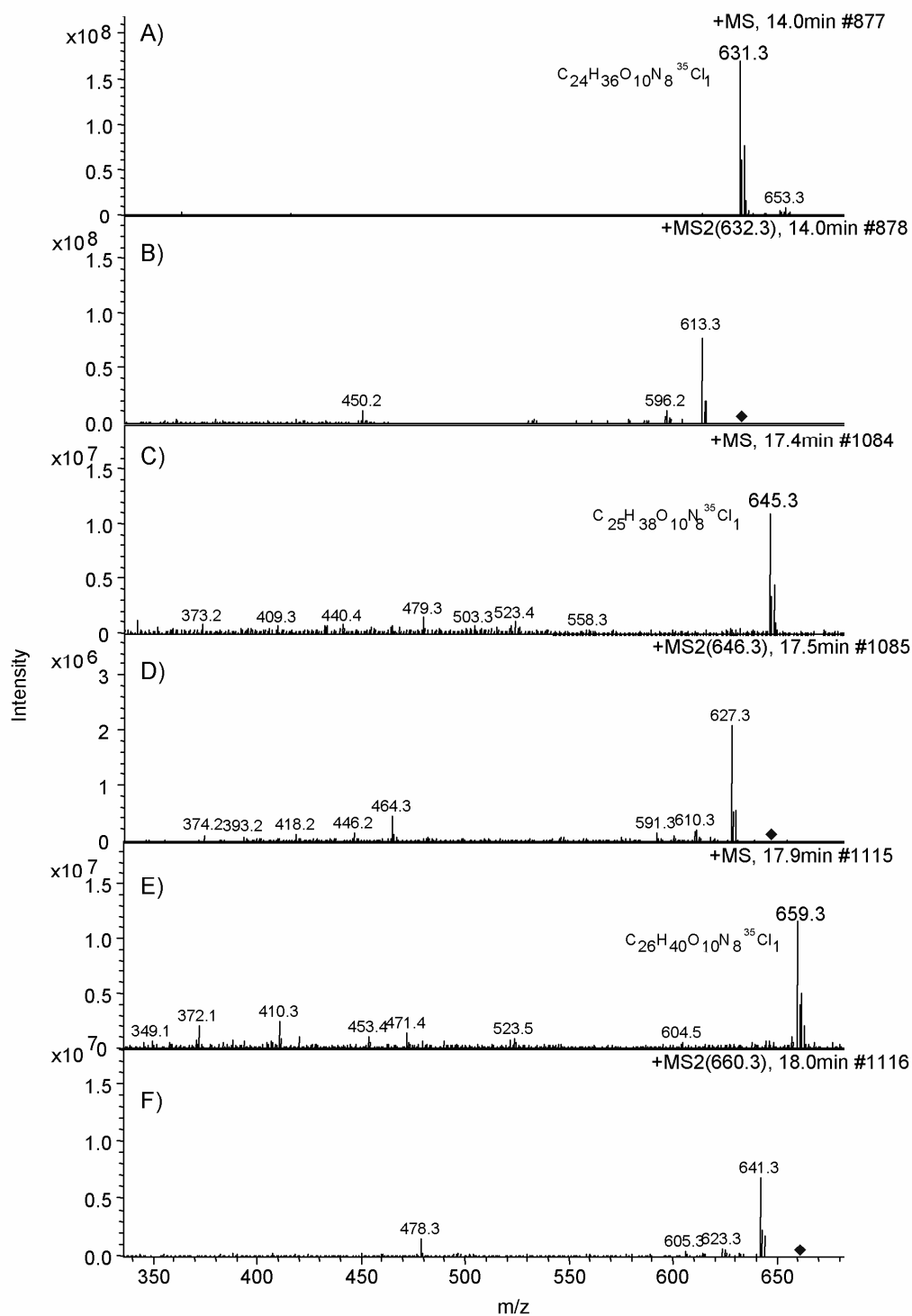
Figure 3.19: **Analysis by high resolution MS of the substances with molecular masses 631, 645, 659.**

The determined accurate masses and molecular formulae are shown. A) Pseudo-molecular ion of $[M+H]^+$=631.3, B) fragmentation pattern of molecular ion $[M+H]^+$=631.3, C) Pseudo-molecular ion of $[M+H]^+$=645.3, D) fragmentation pattern of molecular ion $[M+H]^+$=645.3, E) Pseudo-molecular ion of $[M+H]^+$=659.3 F) fragmentation pattern of molecular ion $[M+H]^+$=659.3

This analysis suggested that the substances are related to the known GE series of compounds, as they are chlorinated and have similar elemental formulas; however, as the fragmentation

patterns differed, the compounds are not identical but remain interesting to investigate. Finally the MS analysis of the production profile revealed that we could now detect the main GE81112 factor B compound. In the next step we aimed to generate a cosmid library to allow us to identify the corresponding biosynthetic gene cluster.

3.2.3 Generation of a cosmid DNA library

One strategy for cloning and identifying a gene cluster is to generate a gene library in *E. coli*. A gene library is a population of organisms, each of which carries a DNA molecule that was inserted into a cloning (e.g. cosmid) vector. Ideally, all of the cloned DNA molecules represent the entire genome of the organism meaning that the desired biosynthetic gene cluster should be also present there. Knowing the size of the genome of the organisms from which one would generate the library one can calculate the approximately number of cosmids containing the gene cluster. When we assume that the average size of a *Streptomyces* genome is 9 Mbp and a cosmid vector can carry ca. 40 kb of insert DNA then we need to generate ca. 225 cosmid clones to clone the whole genome. To get a 10-fold coverage of the genome 2250 cosmid clones need to be obtained. As biosynthetic gene clusters are normally between 30 and 60 kb approximately 30-60 cosmids are expected to harbor parts of the gene cluster. The cosmids containing the desired gene cluster can then be identified by screening of the cosmid library with labeled probes (Figure 3.20).
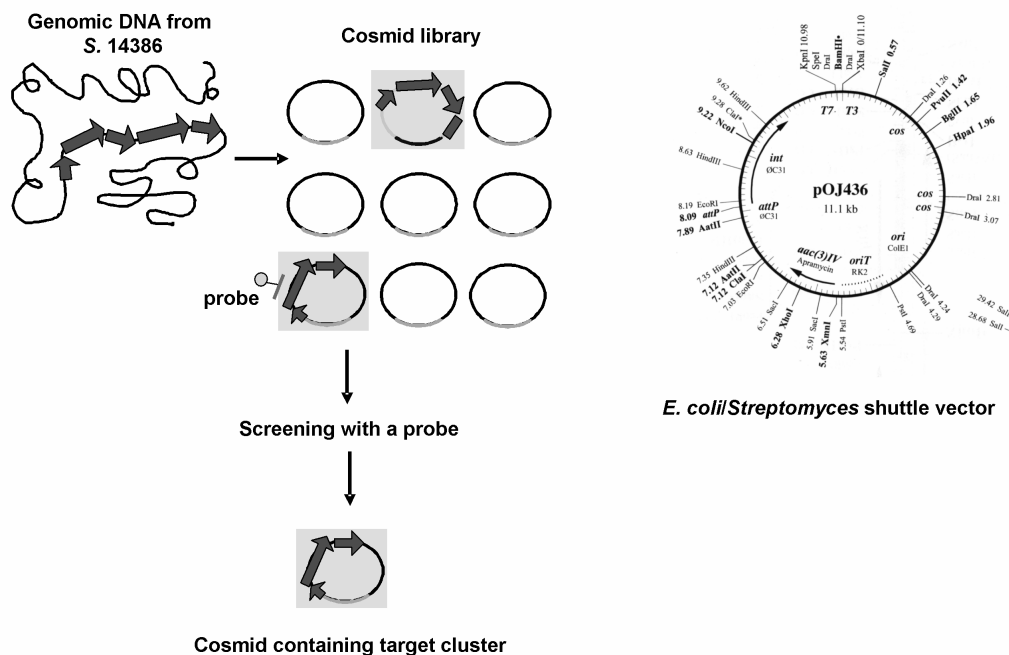
Figure 3.20: **General overview of the strategy for generation of a cosmid library.**

Genomic DNA of the desired organism is isolated and cloned into cosmid vectors. The desired cosmid(s) containing the whole biosynthetic gene cluster or portions of it can then be identified with a labeled probe. A common cosmid vector is the *E. coli/Streptomyces* shuttle vector pOJ436 [10].

As a cosmid vector, we used the *E. coli/Streptomyces* shuttle vector pOJ436 [10]. Genomic DNA from the GE81112 producing strain was isolated using the salting out procedure (described in section 2.10.1). The isolated DNA was then partially digested with *Sau*3A to obtain fragments 30–35 kb in size. Fragments of the correct size were pooled and used for ligation into the *Pvu*II/*Bam*HI-digested vector pOJ436. Figure 3.21 shows partially-digested DNA generated by using different amounts of enzymes. Fractions 6 and 7 were chosen for the ligation, as they contained the greatest proportion of fragments between 35–40 kb. The two fractions were pooled together and ligation was carried out as described in section 2.16.1. To determine the phage titer different amounts of infected *E. coli* were plated out yielding different amounts of cosmid clones (see section 2.16.1).
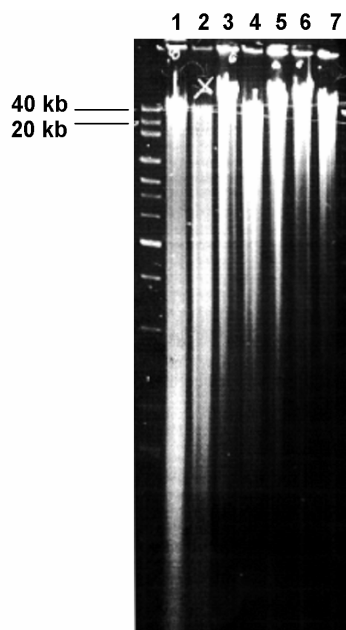
Figure 3.21: **Digestion of genomic DNA from *S.* 14386 with *Sau*3A.**
Lane 1: Ladder; Lane 2–7: variable digestion of genomic DNA obtained by using different amounts of enzyme.

We investigated whether 19 randomly choosen cosmid clones contained the correct insert size by digestion with *Bam*HI (Figure 3.22). In Figure 3.22 one can see that most of the clones showed different fragments larger then 10 kb indicating that most of the analyzed clones contained an insert with an average size of 30–40 kb. Based on this information, we created the cosmid library containing 2304 cosmid clones in order to reach a 10-fold coverage of the genome of *S.* 14386 (section 2.16.2).

The cosmid library was then ready for hybridization with a suitable probe to identify the GE81112 biosynthetic gene cluster.
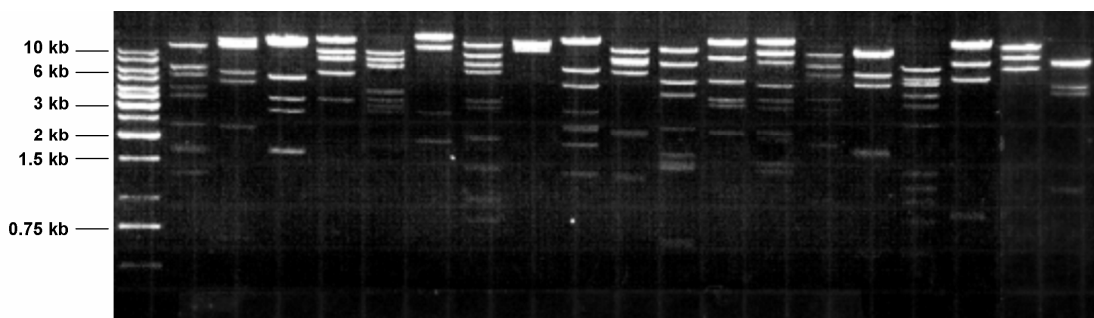
Figure 3.22: **Analysis of 19 randomly chosen cosmid clones.**

The clones were digested with *Bam*HI to determine whether they contained different inserts and inserts of the correct size.

3.2.3.1 **Retrobiosynthetic analysis of GE81112 to design suitable probes**

To find suitable probes for the hybridization of the cosmid library, we applied a "retrobiosynthetic" strategy which allowed us to predict some aspects of the genetic composition of the gene cluster from the structure of the natural product. As described previously, GE81112s were assumed to derive from a non-ribosomal peptide synthetase. It is known that for these systems there is a strong correlation between compound structure and biosynthetic organization (even for some domains in non-linear systems) [64]. For example, it

is possible to predict with some degree of confidence, the building blocks incorporated into the structure and therefore the number of modules and the complement of domains in the NRPS, as well as which functionalities are likely to arise by post assembly-line processing. Thus, according to the molecular structure of GE81112, we were able to propose a hypothesis for the complement of genes that was likely to be present in the GE81112 gene cluster. Assuming that the pathway will be co-linear (see section 1.4), the biosynthetic gene cluster should incorporate four NRPS chain extension modules which are responsible for activating and introducing the four observed amino acids (Figure 3.23). Each of the modules was expected to incorporate the three standard NRPS domains, C, A and PCP. We also anticipated that a halogenase would be involved in the chlorination of all three GE81112 compounds.
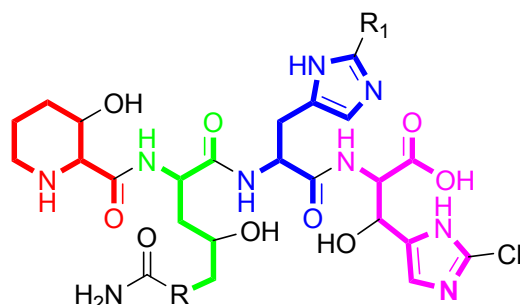
GE81112

GE factor A	R=O	R_1=H
GE factor B	R=O	R_1=NH$_2$
GE factor B1	R=NH	R_1=NH$_2$

Figure.3.23:   **Structures of GE81112 metabolites.**

The different building blocks were expected to be derived from different amino acids and are marked in different colors. Red: pipecolic acid, green: ornithine/glutamine, blue: histidine, purple: histidine

We also hypothesized that the starter unit for GE81112 biosynthesis would be pipecolic acid, or its hydroxylated derivative (shown in red in Figure 3.23). As the stereochemistry of the GE81112 compounds has not been solved yet, we did not know if L-or D-pipecolic acid would be incorporated. On the basis of published precedent [82], however, we assumed that the pipecolic acid would be formed from lysine via the action of a lysine cyclodeaminase (Figure 3.24).
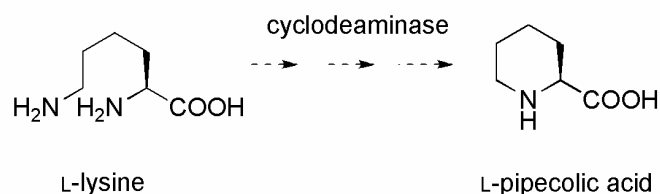
Figure 3.24: Example of conversion of L-lysine to L-pipecolic acid by the action of a lysine cyclodeaminase as it is known from rapamycin.

This proposal is in accordance with the *in vitro* characterization of RapL, a cyclodeaminase, which has been found in the rapamycin biosynthetic gene cluster cluster [82]. Other cyclodeaminases have been identified in the friulimicin [83], the virginiamycin S [84] and in the tubulysin biosynthetic gene clusters [85]. The *in vitro* reaction has also been reconstituted for LipE from the friulimicin gene cluster [83] and TubZ from the tubulysin biosynthetic gene cluster (Y. Chai, A. Sandmann, unpublished results). The pipecolate moiety in GE81112 differs from those in rapamycin and tubulysin, in that it is hydroxylated at position 3. This step was anticipated to be catalyzed by a β-hydroxylase, but the timing (i.e. before or after incorporation) could not be predicted. Another interesting feature of the structures is the amino acid which is incorporated second – aminohydroxypentanoic acid (building block 2, marked in green) – which we expected to derive from glutamine/glutamate or ornithine. We also hypothesized that the third and fourth amino acid building blocks (indicated in blue and purple) are likely to derive from histidine.

3.2.4 Design of the probes

On the basis of these predictions, we were able to design suitable probes for the screening of the cosmid library. For this, it is desirable to design a probe which is specific for an unusual feature of the cluster; as such a probe is likely to improve the chances of finding the correct cosmid. As described above, we predicted that a cyclodeaminase should be present in the genome. As such cyclodeaminases are relatively rare in biosynthetic pathways [86], this gene was judged to be a good candidate for trying to locate the GE81112 biosynthesis gene cluster in the genome of *S.* 14386. We therefore designed a probe based on cyclodeaminase sequences. To create the probe, we used specific primers to amplify two known cyclodeaminase genes – *tubZ* [85] (primers: TubZ_up and TubZ_down (section 2.7.2.2.)) from the tubulysin cluster and *rapL* [87] from the rapamycin cluster (primers: RapL_up and

RapL_down (2.7.2.2.)). The amplificates were pooled and used to probe the cosmid library at low stringency. To check if the probe worked, we hybridized the genomic DNA of *S.* 14386 (digested with different enzymes) with the probe which yielded several distinct hybridization signals, indicating that the cosmids containing cyclodeaminase genes were likely present, and therefore that the library was likely to contain the desired cluster.

As an alternative approach, we designed an additional probe to identify NRPS genes. For this, we used degenerate NRPS primers (RevA3 and PSLGG) (2.7.2.2.) which had already been employed successfully to identify NRPS genes in myxobacteria [88-90]. These primers amplify A domains between the structural regions A3 and A6 [91], where the amino acid binding pocket is located, giving a 1.3 kb PCR fragment. The PCR was carried out using a step-wise gradient with different annealing temperatures (46–52 °C) using the *S.* 14386 DNA as template. Several of the obtained PCR fragments were cloned into PCR 2.1 TOPO and sequenced. Different A domain binding pockets were obtained after sequencing (presumably specific for the incorporation of valine, threonine, glutamine) indicating that the probes had successfully amplified various A domains. As it was not possible to correlate the substrate specificities of the A domains to the GE81112 structure, we were unable to select a single A domain as a probe, and therefore decided to use the whole PCR reaction containing all of the amplified A domains. This approach was expected to ensure a high diversity of A domain hybridization so that as many as possible A domains would hybridize.

3.2.5 Identification of an unknown NRPS biosynthetic gene cluster

3.2.5.1 **Identification of two cosmids containing NRPS genes**

Using both the cyclodeaminase and A domain probes at low stringency, we identified several cosmids which hybridized with one or both of the probes (Figure 3.25). Ten of these cosmids were then chosen for a more detailed analysis. A PCR with the degenerate NRPS primers was performed to verify the hybridization results, yielding the expected 1.3 kb fragment in each case. PCR with the cyclodeaminase primers was not successful; this result was expected, however, as the primers used were not degenerate, and so would not have been expected to target the particular cyclodeaminase gene(s) present in the *S.* 14386 genome. The DNA of the ten cosmids was isolated, and digestion with different enzymes was performed to evaluate the restriction patterns and to check if similar restriction fragments for different cosmids were obtained. This result would indicate that the same DNA sequence (biosynthetic gene cluster) was present on the cosmids.
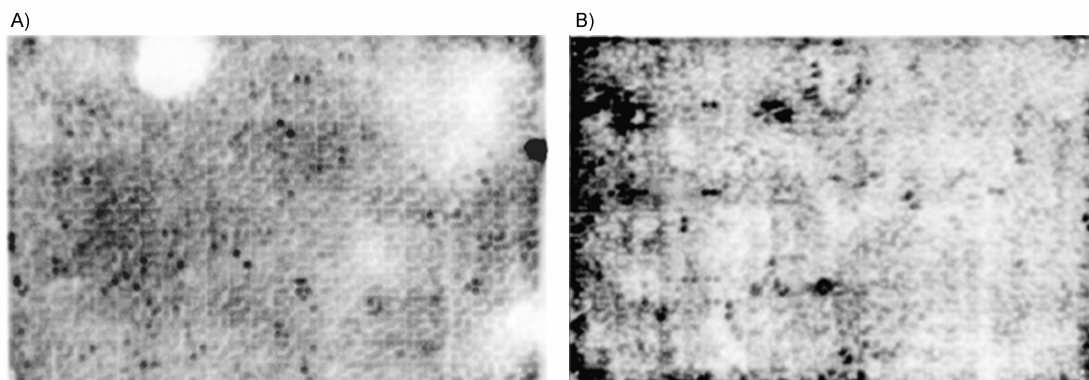
Figure 3.25: **High density colonies filters.**
A) Hybridization with the cyclodeaminase probe. B) Hybridization with the NRPS probe.

Two of the cosmids showed similar restriction patterns, indicating that they contained overlapping sequence. End sequencing (Primers T4 and T7 (Table 2.12)) revealed that one of the cosmids (AI6) contained NRPS sequences (Table 3.6), and so this cosmid and the overlapping one were sequenced in their entirety.

Table 3.6: **End sequencing results for cosmids FD10 and AI6 with T4 and T7 primers**

Cosmid	Homology (Name, Strain, Database no.)	Identity	Score	E-Value
FD10_T4	Polyketide synthase type I, *Streptomyces aizunensis*, gi/62737770/AAX98186186.1	40%	37.4 bits	0.26
FD10_T7	Glutamate synthase, small subunit, *Streptomyces avermitilis,*gi/29832732/ref/NP827366.1	89%	228 bits	2e-79
AI6_T4	Hypothetical protein SAV2451, *Streptomyces avermitilis* MA-4680, gi/29828993/ref/NP823627.1	83%	268 bits	3e-70
AI6_T7	Peptide synthetase 2, *Streptomyces filamentosus* gi/60650533/gb/AAX31558.1	51%	304 bits	6e-81

3.2.5.2 **Annotation of the unknown biosynthetic gene cluster**

The two cosmids were sequenced at the HZI (Helmholtz Center for Infection Research) in Braunschweig, in the Department of Genome Analysis. The obtained sequence was then analyzed for the presence of putative open reading frames (*orfs*) with FramePlot 2.3.2 [92], and preliminary functional assignments of individual *orfs* were made by comparison of the deduced gene products with proteins of known function in the BLAST database. Sequence analysis of cosmids AI6 and FD10 revealed three *orfs* for NRPS biosynthesis each containing four modules, which are all transcribed in the same direction (Figure 3.26 and Table 3.7). The modules include typical NRPS domains C, A and PCP domains, as well as E domains.
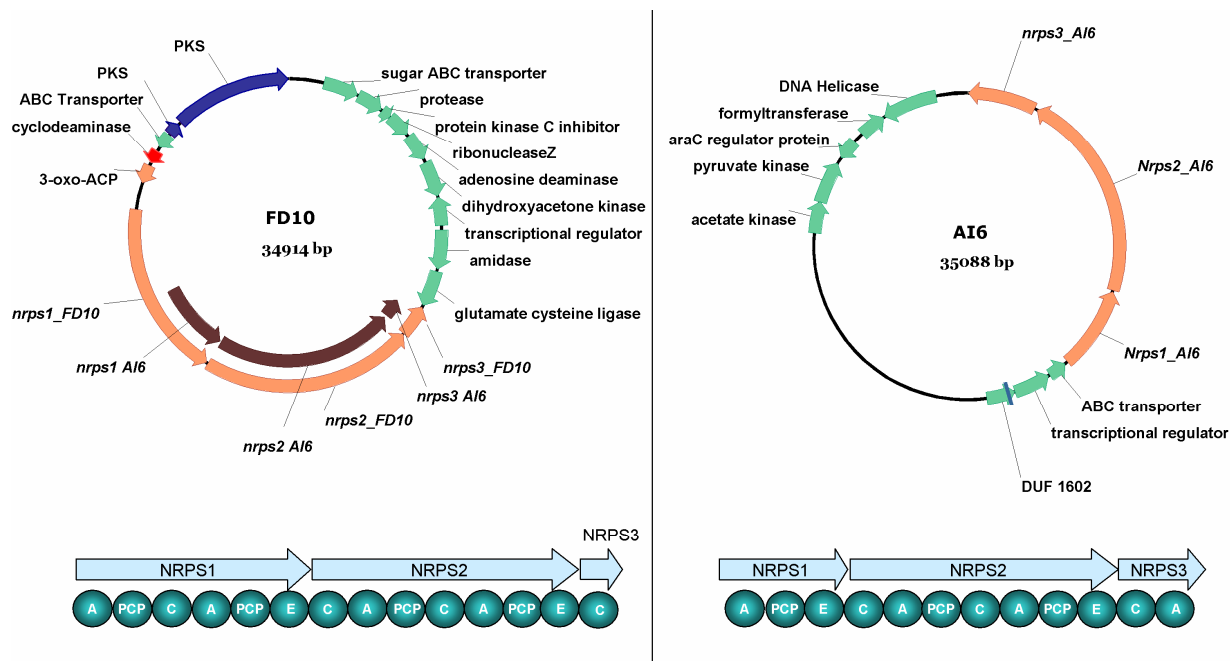
Figure 3.26: **Sequencing results from cosmids FD10 and AI6.**

Key: orange, NRPS and PKS genes; brown, regions which are shared between cosmids AI6 and FD10.

Table 3.7: **List of NRPS genes from cosmids FD10 and AI6**

Orf	Proposed function	Identity/similarity to protein/origin
Cosmid FD10		
nrps1_FD10	NRPS: A_1 PCP_1 C_2 A_2 PCP_2 E_2	52%, 64%: SACE4288, *Saccharopolyspora erythraea*
nrps2_FD10	NRPS: C_3 A_3 PCP_3 C_4 A_4 PCP_4 E_4	47%, 58%: CDA peptide synthetase I, *Streptomyces coelicolor* A3(2)
nrps3_FD10	NRPS: C_5	54%, 70%: pstC, *Actinoplanes friuliensis*
Cosmid AI6		
nrps1_AI6	NRPS: A_1 PCP_1 E_1	51%, 63%: SACE4288, *Saccharopolyspora erythraea*
nrps2_AI6	NRPS: C_2 A_2 PCP_2 C_3 A_3 PCP_3 E_3	47%, 58%: CDA peptide synthetase I, *Streptomyces coelicolor* A3(2)
nrps3_AI6	NRPS: C_4 A_4	50%, 65%: pstC, *Actinoplanes friuliensis*

The *nrps1_FD10* gene from cosmid FD10 most likely starts with an ATG which is located 9 bp downstream of a putative ribosome binding site (RBS) GAGG. *nrps1_FD10* appears to be part of an operon with *nrps_FD10_2-3*. *nrps2_FD10* represents another NRPS with a putative RBS (GGAG) upstream from the start codon ATG. The subsequent gene *nrps3_FD10*

encodes a NRPS protein, which only contains a C-domain. *nrps3_FD10 starts* with GTG, and a RBS (GGAG) is located 8 bp upstream. On the basis of sequence analysis and restriction mapping, it was expected that the two cosmids would contain portions of the same biosynthetic gene cluster, and would therefore share an overlapping region. However, annotation of cosmid AI6 revealed that only the *nrps2_AI6* gene showed exactly the same sequence as *nrps2_FD10*. The other two genes, *nrps1_AI6* and *nrps3_AI6*, exhibited some sequence identity to the corresponding FD10 genes, but were not identical along their full lengths. Specifically, *nrps1_AI6* is only half the size of *nrps1_FD10*, while only a portion of *nrps3_AI6* is present in *nrps3_FD10*. Additionally gene *nrps3_AI6* is interrupted at the end of the cosmid, so it is not clear how large this gene cluster actually is. Taken together, these observations indicate that the clusters might have arisen from gene duplication, as they are overall highly similar, though not identical.

At this point, it remained to establish which of the two clusters, if either, was responsible for the biosynthesis of GE81112. The presence of four modules in each gene cluster was consistent with the incorporation of four amino acids into the corresponding products, so no decision could be made on that basis. We therefore performed a detailed sequence analysis of the NRPS domains from each module.

3.2.5.3 **Analysis of the NRPS domains**

3.2.5.3.1 A domains

NRPS include a specific adenylation domain to select each amino to be incorporated into the peptide product; the sequence of A domains in the NRPS dictates the primary structure of the peptide product. A domains contain 10 highly-conserved core motifs (A1─A10), where the region between motifs A3 and A6 forms the amino acid binding pocket [91]. To determine the substrate specificity of the A domains in the putative GE81112 clusters, each of the A domains was aligned with A domains from the gramicidin, rapamycin and tubulysin synthetases. All ten core motifs were identified in each A domain from both cosmids, indicating that all A domains are functional.

Figure.3.27:   **Consensus core motifs of A domains from cosmid FD10.**

The ten core motifs A1–A10 of the NRPS A domains are shown. The A domains from cosmid FD10 were compared with known A domains from NRPSs of gramicidin (GrsA), rapamycin (RapP) and tubulysin (TubB). The amino acids marked in green match the consensus motif for each region. "x" can be any amino acid.

The substrate specificity of A domains can be predicted using the specificity-conferring code of adenylation domains [93;94]. The code was developed by two different groups [93;94] using the crystal structure of the phenylalanine-activating A domain PheA from the gramicidin synthetase (GrsA) from *Bacillus brevis*. The crystal structure led to the identification of the 8 residues that form the substrate binding pocket. By comparing the sequence of PheA, the structure of which was known, with the sequence of 160 other adenylation domains, eight residues that (putatively) form the amino acid binding pockets were identified. The signature sequences for different substrate specificities were then determined from groups of domains activating the same substrate. This code now allows the

substrates of uncharacterized or unknown A domains to be predicted from the primary sequences. Nevertheless some limitations of the code should be noted. First, there is only a limited amount of sequence information available for A domains, and thus it is currently impossible to discern with confidence consensus specificity sequences for all amino acids. Furthermore, NRPS are not restricted to proteinogenic substrates and more then 300 are known to be incorporated but the codes currently available are mostly restricted to proteinogenic amino acids. A clear shortcoming of the available biochemical data is that it is limited mostly to A domains from *E. coli*, *Bacillus sp.* and *Pseudomonas sp.* Actinomycetes, including notably *Streptomyces*, are largely missing from the essential database.

In order to predict the substrate specificity of the FD10 and AI6 A domains from the sequenced cosmids, the A domains were aligned against the A domain from the gramicidin A synthetase (GrsA), and the 8 residues that form the amino acid binding pocket were identified (Figure 3.28; shown for A domains from cosmid FD10).

```
                       coreA3                              coreA4
FD10A1      ---PVYVIYTSGSTGTPKGVVVPHGGLANQIPWMAAEWEITPRDVVLARHSVSFDAAGVE  57
FD10A4      DRHPVYVIYTSGSTGTPKGVVVTHGALTGQLRWLAAETELTPRDVVLARTPVSFDAAGAE  222
FD10A3      GLSAAYVIYTSGSTGRPKGVVVPHGNVVRLFSATDRWFDFGPDDVWTLFHSFAFDFSVWE  222
FD10A2      VRHPAYLIYTSGSTGLPKGVVVTHHGLASFSATEIDRFSVTPRSRILQFSSPSFDAAVLE  222
GrsA-M1-Phe ---LAYVIYTSGTTGNPKGTMLEHKGISNLKVFFENSLNVTEKDRIGQFASISFDASVWE  57


FD10A1      LWTPLASGAAMAMVSGEATRDPELLLDRIQRHGVTAMDMIPS----LLAAMPHDERGRS-  112
FD10A4      LWPPLVSGAAVALASAEVTRDPEQLLAFIGHHGVTVAQFVPS----LLAATPLDERGRG-  277
FD10A3      IWGPLLHGGRLVVVPFAVSRSPREFRALLVREGVTVLSQTPSAFYQLMAADAECGAGDGE  282
FD10A2      ICMALPAGAALVVPPPGPLAG-DVLARVLAEHSVTHALIPPA----ALASVPPVALPDF-  276
GrsA-M1-Phe MFMALLTGASLYIILKDTINDFVKFEQYINQKEITVITLPPT-----YVVHLDPERILS-  111

                                               coreA5
FD10A1      --VRVMFSGGEALTPALARQLIGAWD---LRLINTYGPTEATIQAT-AGDCAGLTDDAVS  166
FD10A4      --IRVLLMGGEALPAALAGQAAAAWG---AQVINVYGPTEATIQAT-SGRPDGSGDDATT  331
FD10A3      LVLRRVVFGGEALDLWRLGDWYARHGDAAPVLVNMYGITETTVHVSYVALDAGRVAGNAG  342
FD10A2      ---RTLIVGGEACSAELVDRWSAGRV-----MVNAYGPTESTVAATMSGPLAS---GGGV  325
GrsA-M1-Phe --IQTLITAGSATSPSLVNKWKEKVT-----YINAYGPTETTICAT--TWVATKETIGHS  162

                                  coreA6
FD10A1      VPIGLPVWNTTVHVLDEELRPVPAGTDGDLYIAGVQLARGYLNRPSLTAERFVADPYGPA  226
FD10A4      VPIGRPAWNTRVYVLDELLRPVPDGVAGEVYLAGGQLARGYLNRPGLTAERFVADPYGPA  391
FD10A3      SVVRGRGIADLRVYVLDGALRPVPGASGELYVAGAGLARGYLNRPGLTGERFVADPFGVP  402
FD10A2      PPIGRPVRGTEVFVLDEGLRLVPPGVPGELYIAGAGLARGYLRRGALTAERFVACPFGGP  385
GrsA-M1-Phe VPIGAPIQNTQIYIVDENLQLKSVGEAGELCIGGEGLARGYW-----------------  204
```

Figure.3.28: **Alignment of A domains from cosmid FD10 with the A domain of GrsA.**

grey: core motifs of A domain; black: 8 residues that form the amino acid binding pocket

Additionally the sequence was cross-checked with the bioinformatics programs (http://www.tigr.org/jravel/nrps/ and http://www-ab.informatik.uni-tuebingen. de/ toolbox/index.php?view=domainpred) which also allow substrate specificities to be predicted. Sequence analysis of the code residues from all of the A domains from both clusters allowed the prediction of substrate specificity in several cases (Table 3.8).

Table 3.8:   **Putative A domain 'code residues' within A domains from both clusters**

	Position of the amino acid within the A domain								Activated amino acid prediction	Identitiy to protein
	235	**236**	**239**	**278**	**299**	**301**	**322**	**330**		
FD10_A1	D	A	V	D	F	G	T	I	No hit	
FD10_A2	D	A	L	L	I	G	A	V	Phe/Trp	75%
FD10_A3	D	F	W	S	V	G	M	V	Thr	87%
FD10_A4	D	A	A	Q	L	G	V	I	No hit	
AI6_A1	D	A	L	L	I	G	A	V	Phe/Trp	75%
AI6_A2	D	F	W	S	V	G	M	V	Thr	87%
AI6_A3	D	A	A	Q	L	G	V	I	No hit	
AI6_A4	D	V	F	S	V	A	V	V	No hit	

When identity to known A domains could not be found in the database for the 8 amino acid residues, no hit was predicted from the bioinformatics program. We also compared the A domains manually with the known A domains which were published by Challis *et al.* [94]. Despite the use of multiple methods no substrate specificity could be predicted for FD10_A1, FD10_A4, AI6_A3 and AI6_A4.

Consistent with the observation that both clusters exhibited an overall high level of sequence similarity, three of the A domains from each cluster showed the same apparent substrate specificity (FD10_A2 and AI6_A1, FD10_A3 and AI6_A2 and FD10_A4 and AI6A3). FD10_A2 (which showed the same specificity as AI6_A1) exhibited 75% identity to the specificity pocket of the TycB-M2 A domain in *Bacillus brevis*, which selects and activates L-phenylalanine or L-tryptophan. Sequence analysis of FD10_A3 (same specificity as AI6_A2) revealed 87% identity to the specificity pocket of the SyrE-M7 A domain in *Pseudomonas syringae*, which is involved in selection and activation of L-threonine. Analysis of FD10_A1, FD10_A4 and AI6_A3 and AI6_A4 A-domains revealed that the identified code residues did not allow a prediction of the specificity.

3.2.5.3.2   PCP, C and E domains

The PCP, C and E-domains were analyzed to determine whether or not they were likely to be functional, which would support the hypothesis that the clusters act to biosyntheize a tetrapeptide. The peptidyl carrier protein (PCP) (or thiolation (T) domain) is the site of substrate attachment during chain elongation [95]. These proteins show a conserved signature sequence LGx(HD)S(LI), where the central S is the point of covalent attachment of the Ppant prosthetic group. The signature sequence including the active serine was found in all PCP domains from both cosmids FD10 and AI6, indicating that the domains should be activated to participate in chain extension.

```
                                     D I
        Consensus            DxFFxxLGGHSL

        FD10_PCP1            DSFFD-LGGHSL
        FD10_PCP3            DSFFE-LGGHSL
        CDA_PCP1             DGFFD-LGGHSL
        FD10_PCP2            DNFFD-LGGDSI
        FD10_PCP4            GDFFD-LGGDSI
        GrsB_PCP1            DNFFE-LGGHSL
        SrfA_PCP1            DDFFA-LGGHSL
        TycA_PCP_1           DNFYS-LGGDSI
```

Figure 3.29:   **PCP consensus motif.**

The PCP domains from cosmid FD10 were compared with known PCP domains from the gramicidin (GrsA), surfactin (SrfA), tyrocidine (TycA) and calcium-dependent antibiotic (CDA) NRPSs. The amino acids marked in grey match the consensus motif. "x" can be any amino acid. The highly conserved serine residue (site of Ppant attachment) is indicated in black.

C domains are responsible for the condensation of two amino acids activated on adjacent modules. These domains incorporate seven highly-conserved core regions, all of which were found in the C domains in both cosmids FD10 and AI6. We also analyzed the epimerase domains, which are responsible for substrate epimerization (formation of D-amino acids). Again, the seven characteristic conserved motifs for all E domains could be identified in the E domains on both cosmids. These results demonstrate that, in principle, all modules are active and so each system should be able to generate a tetrapeptide. Thus, these results did not allow us to prove or exclude that one or both of the clusters are responsible for the biosynthesis of GE81112 and further experiments/analysis had to be done. Therefore we continued our analysis, by investigating the cyclodeaminase from FD10.

3.2.6 Analysis of the cyclodeaminase from cosmid FD10

We also expected that the GE81112 cluster would include a cyclodeaminase, providing L-pipecolic acid from L-lysine. BLAST sequence analysis of both cosmids revealed that only cosmid FD10 contained a cyclodeaminase sequence (The protein showed 62% identity to the TubZ protein from *Angiococcus disciformis)*. To analyse the conserved regions of the protein which would indicate whether it was functional, we performed an alignment with three different cyclodeaminases TubZ [85], LipE [83] and RapL [87] (Figure 3.30).
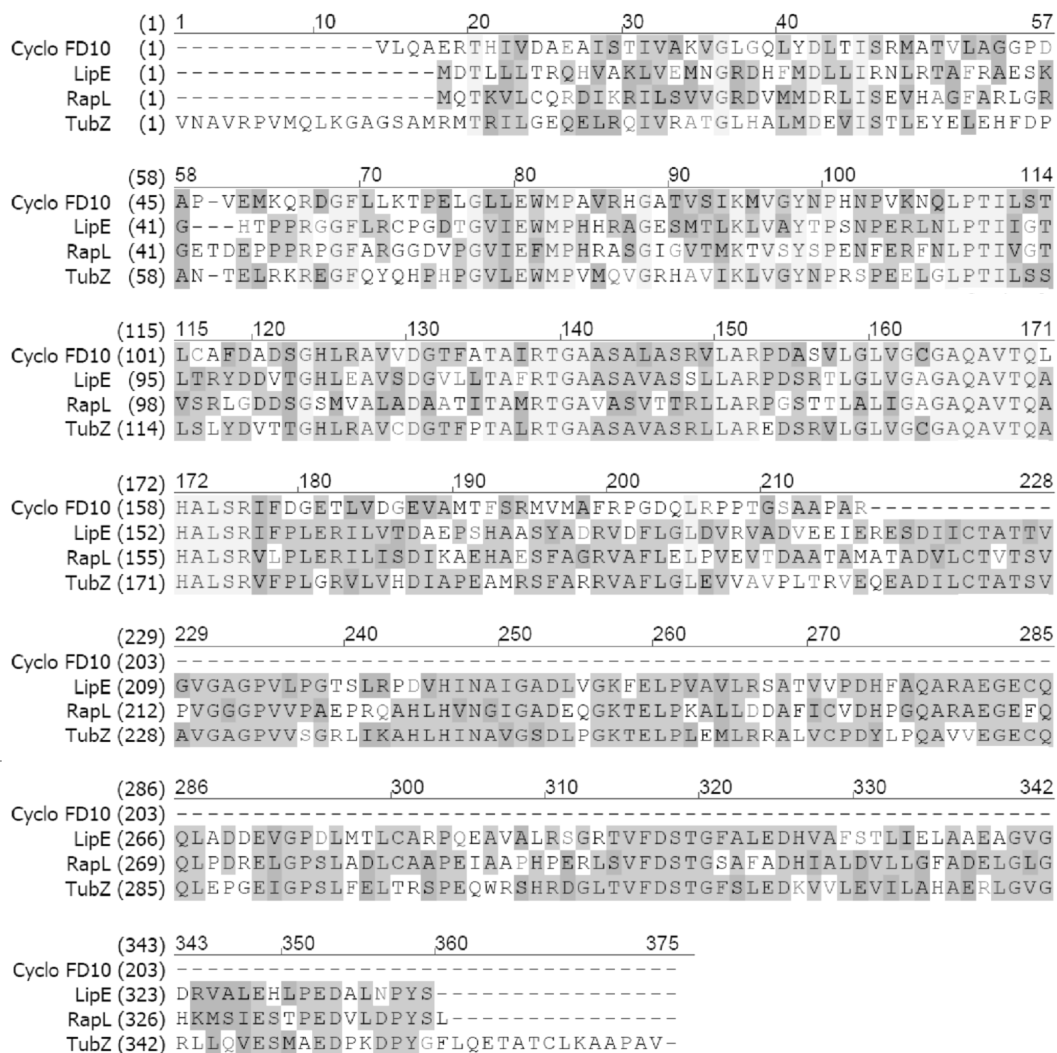
Figure.3.30:   Alignment of the cylodeaminase from cosmid FD10 with LipE, RapL and TubZ.

No color: Non-similar amino acid residues, blue: conservative, green: block of similar, yellow: identical

Cyclodeaminases are typically 1000–1100 kb in size (corresponding to approximately 350 amino acid residues). However, the cyclodeaminase present on cosmid FD10 was only 465 kb

in size, corresponding to 155 amino acid residues. Thus, more than half of the protein was apparently missing, relative to previously-characterized cyclodeaminases.

3.2.6.1 Heterologous expression of cosmids FD10 and AI6

In principle, the most straightforward method to verify the identity of the gene clusters would have been to knockout one of the NRPS genes, and then to determine if production of GE81112 was abolished. We attempted to generate this knockout by amplifying an internal region from the *nrps_2* gene, which was then cloned into the knockout vector pKC1132 [10], and introduced the vector into strain *S.* 14386 by conjugation. However, to date, the knockout has not been successful.

We therefore decided to use an alternative strategy to determine the identities of the cluster – heterologous expression of the biosynthetic gene clusters from both cosmids. The two cosmids were transformed into *S. lividans* TK24 and *S. coelicolor* A3(2) by conjugation. Exconjugants were verified by PCR (amplification of the apramycin resistance gene) and extracts were analyzed by high resolution Orbitrap-MS analysis (data not shown). We did not observe production of GE81112 in the extracts of the mutants harboring either cosmid AI6 or FD10. Instead, three new peaks with masses 631.2242 (retention time 3.86 min), 645.2389 (retention time 4.70 min) and 659.2545 (retention time 4.94 min) were detected in the extract of *S. lividans* TK24::FD10 (Figure 3.31), although the compounds were not present in *S. lividans TK24* wild type or in *S. lividans* TK24::AI6.
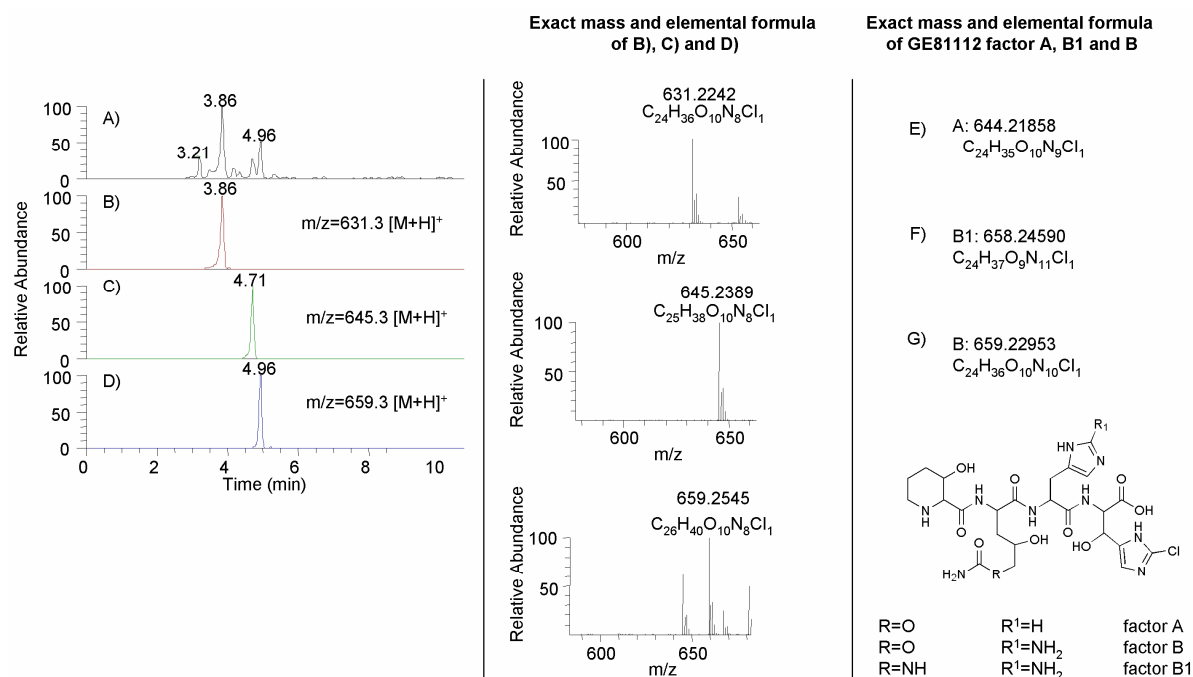
Figure 3.31: **Analysis of *S. lividans*::FD10 extract by high-resolution MS.**

A) Base peak chromatogram (BPC) of *S. lividans*::FD10 extract and extracted ion chromatograms (EICs), showing molecular ions and elemental formulas of B) *m/z*= 631.2242 [M+H]$^+$, C) *m/z* = 645.2389 [M+H]$^+$ and D) *m/z* = 659.2545 [M+H]$^+$. E), F) and G) show the exact masses and elemental formulas of GE81112 factor A, B1 and B for comparison.

Detailed MS/MS analysis revealed that all 3 of the substances exhibited exactly the same fragmentation pattern as the new chlorinated substances found during analysis of the GE producer strain (section 3.2.2). This result led us to conclude that the three substances 631, 645 and 659, had been heterologously expressed in *S. lividans* from cosmid FD10. No production of the same metabolites was observed from cosmid AI6, which is likely explained by the fact that the cluster is incomplete. In addition, no other new metabolites were detected in the extract of *S. lividans* TK24::AI6.

These results indicated that the biosynthetic gene cluster from cosmid FD10 is not responsible for biosynthesis of GE81112, but for the new substances which might be related to the GE compounds. Therefore we aimed to purify the novel substances and to elucidate their structures.

3.2.7 Isolation and attempted structure elucidation of three chlorinated substances

In previous experiments, we observed that the desired compounds were mostly localized to the culture supernatant, while only traces of the metabolites were present in the cell pellets. Cultivation with XAD resin did not lead to an increase of production, and only small amounts of the compounds were detected in the extract of the XAD. Therefore the strain was grown without XAD, and the compounds were isolated from the supernatant. *S.* 14386 was grown in GE production medium for 6 days and harvested by centrifugation.
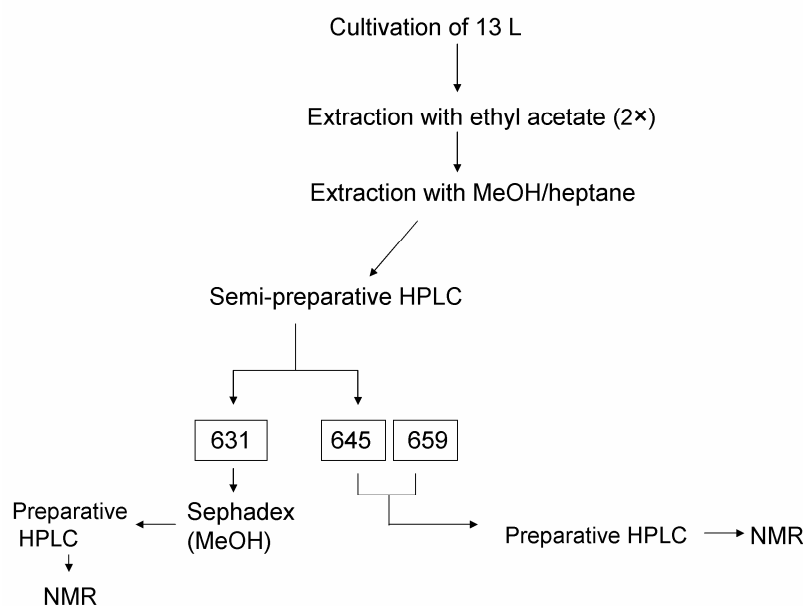
Figure.3.32: Purification procedure for substances **631, 645** and **659.**

The purification scheme is outlined in Figure 3.32. First, 13 L of the supernatant were extracted twice with the same volume of ethyl acetate. After the solvent was removed by evaporation, the extract was resuspended in 50 ml methanol and extracted with heptane for degreasing. The heptane and MeOH phases were subsequently analyzed by HPLC-MS.
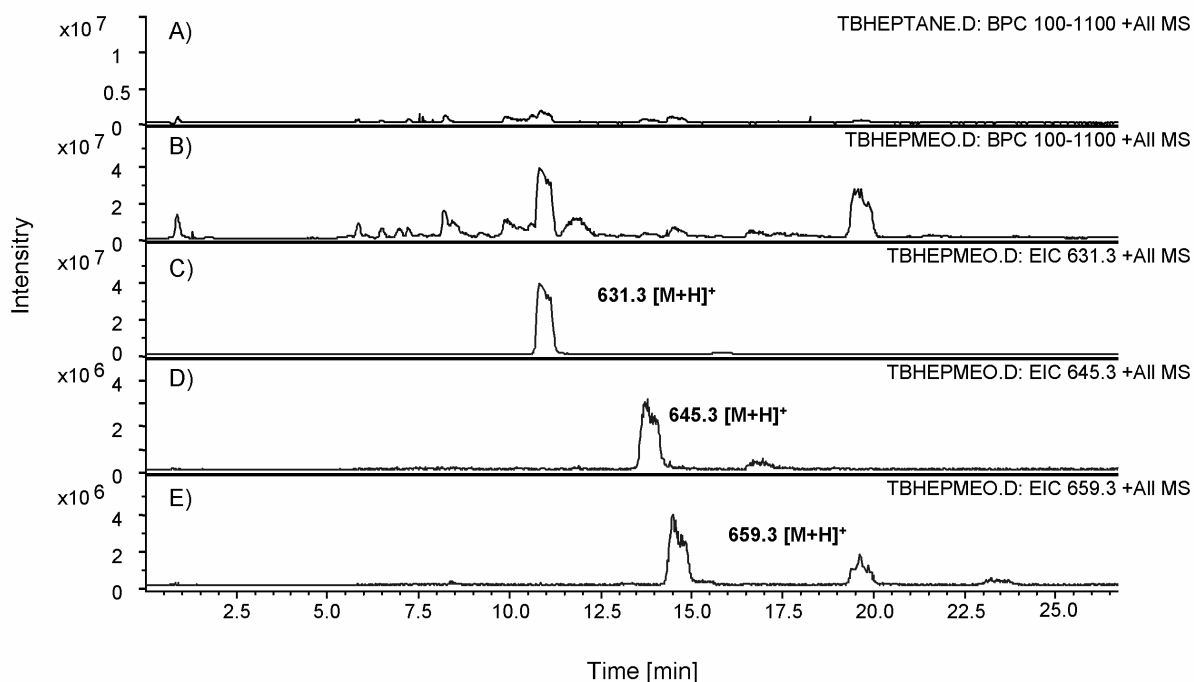
Figure 3.33: **HPLC-MS analysis of heptane and MeOH fractions.**

Base peak chromatogram of heptane A) and MeOH B) phases, and extracted ion chromatograms of MeOH fraction showing molecular ions of C) m/z $[M+H]^+$ = 631.3, D) 645.3 and E) 659.3.

Only traces of the compounds could be detected in the heptane phase, while all three compounds were present in the MeOH phase (Figure 3.33). The methanol phase was therefore evaporated and resuspended in a smaller volume of methanol (2 ml). This sample was then fractionated by semi-preparative HPLC using a stepwise gradient (solvent A: H_2O containing 0.01 % formic acid; solvent B: methanol containing 0.01% formic acid, 25%–90% B over 35 min, injection volume 250 µl). As the substances did not exhibit a significant UV signal, the HPLC was coupled to a mass spectrometer (Bruker HCT equipped with an ion trap mass spectrometer) for online monitoring of the masses. Using this method, the exact retention times of the compounds could be measured, allowing straightforward collection of the substances. Five fractions were collected and analyzed for the presence of the compounds (representative results are shown in Figure 3.34).
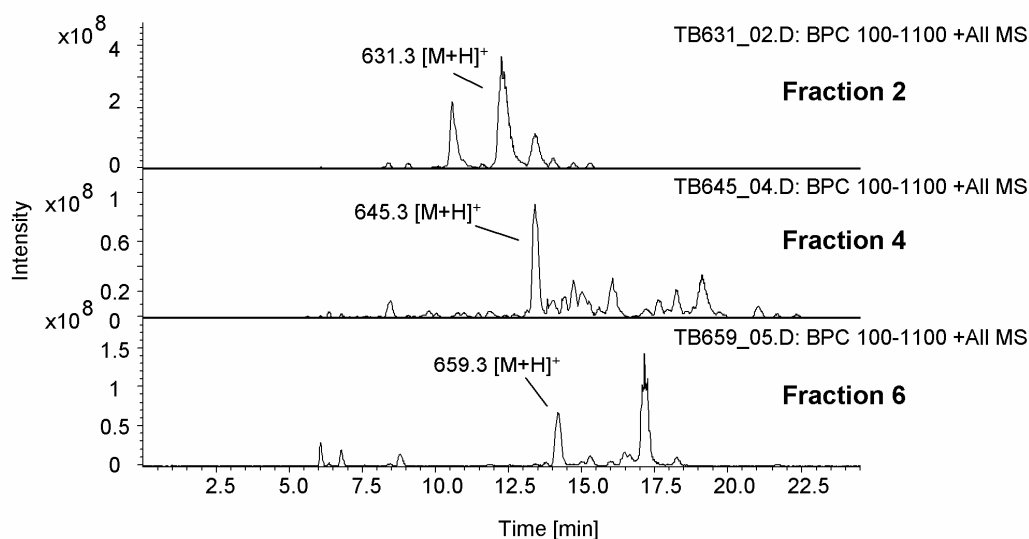
Figure 3.34:   HPLC/MS analysis of different purified fractions after preparative HPLC.

Base peak chromatogram showing molecular ions with the masses m/z [M+H]$^+$ = 631.3, 645.3 and 659.3.

Fractions 2, 4 and 5 contained the compounds 631, 645 and 659, respectively. However, contaminants remained in all of the fractions, so it was necessary to carry out a further purification step. MS analysis revealed that only substances with relatively low masses (< m/z = 300) were present in fraction 2 (containing 631). Fractions 4 (645) and 5 (659) also contained substances with masses similar to the desired compounds (in a range of m/z = 600). These observations prompted us to use size-exclusion chromatography for fraction 2 (631), followed by purification by isocratic preparative HPLC, while fractions 4 (645) and 5 (659) were directly submitted to separation by isocratic preparative HPLC. All 30 samples from each fraction were combined and the solvent evaporated. The extracts of fractions 4 (645) and 5 (659) were resuspended in 500 μl methanol.

The extract of fraction 2 (631) was resuspended in 1 ml of methanol and subsequently separated on a Sephadex LH20 column (28 h), yielding 48 fractions. Every second fraction was analyzed by HPLC/MS. Fractions 9–11 contained the compound and were combined. The solvent was evaporated and the extract was resuspended in 500 μl methanol. All three extracts were then purified by isocratic preparative HPLC (solvent A: H$_2$O containing 0.01 % formic acid; solvent B: acetonitrile containing 0.01% formic acid, 0–10 min 30% B isocratic, 10–20 min 30%–50% B, 20–25 min 50%–95% B, 25–27min 95% B, injection volume 200

µl). Monitoring was again carried out by mass spectrometry. After this separation step, essentially pure compound was obtained in each case (Figure 3.35)
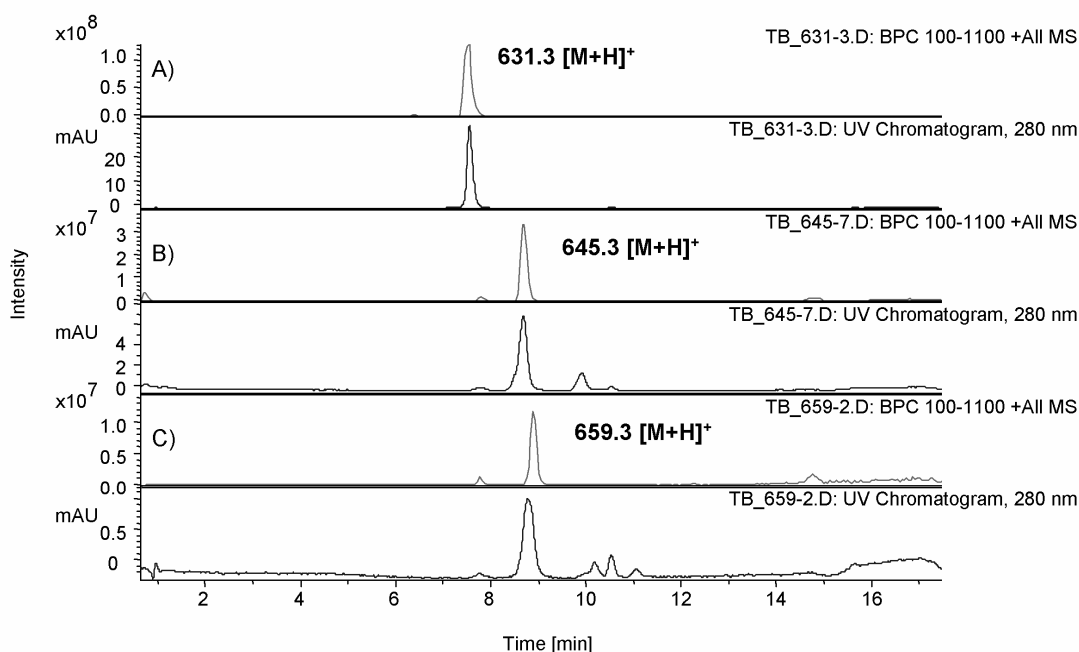
Figure 3.35: **HPLC-MS analysis of different fractions after preparative HPLC.**

Base peak chromatograms and UV chromatograms after preparative HPLC showing molecular ions of A) m/z $[M+H]^+ = 631.3$, B) m/z $[M+H]^+ = 645.3$, and C) m/z $[M+H]^+ = 659.3$.

The extracts were vacuum-dried by lyophilization and weighed. Compound 631 was obtained at a yield of 4.8 mg, compound 645 at 2.3 mg and compound 659, 1.2 mg. Structure elucidation of all three substances was then performed by NMR.

3.2.8 NMR structure elucidation

NMR structure elucidation was started with the compound 631 because purification yielded the highest amount (4.8 mg) of this compound. NMR was carried out using 1D (^{1}H, ^{13}C) and 2D (^{1}H-^{1}H-COSY, HMBC and HSQC) NMR analysis in [D$_6$]DMSO as well as in CD$_3$OD. Although several structural elements have been elucidated to date for compound 631 (Figure 3.36), it has not yet been possible to solve the complete structure.
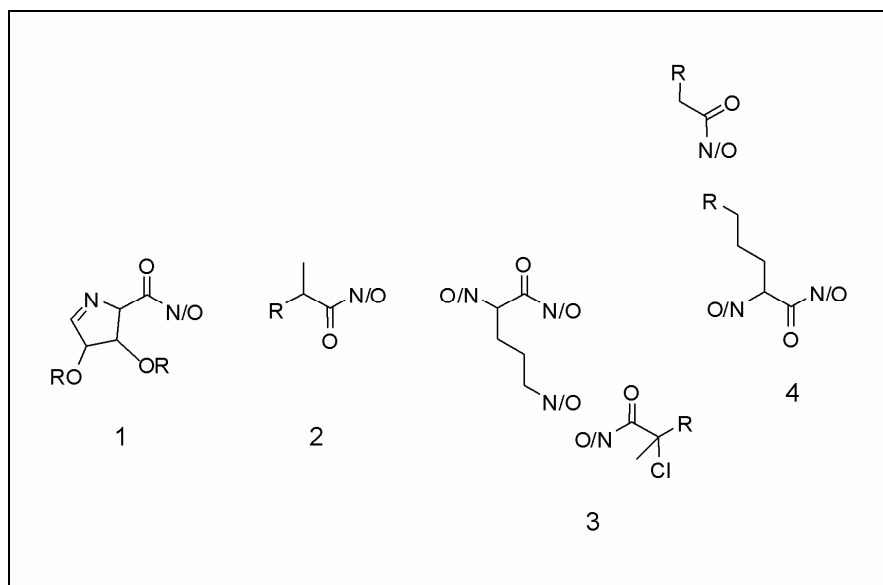
Figure 3.36: **Postulated structural elements of compound 631 identified by NMR spectroscopy.**

N/O means that N or O might be present at that specific position. R indicates that the functionality could not be identified.

(NMR interpretation was done by H. B. Bode) and is not discussed in detail here. The structure is likely to contain four amino acid-derived structural elements (nos. 1–4) including two ornithine moieties (3 and 4) which are probably acylated (H. B. Bode, personal communication). However, these structural elements could not be connected together due to missing correlations in the 2D-NMR experiments (HMBC, COSY). The three compounds 631, 645 and 659 differ in 14 atomic mass units each. To date, it is not clear at which positions these variations occur. Therefore, further experiments will be necessary to complete the structure elucidation.

3.2.9 Identification of the GE81112 biosynthetic gene cluster

3.2.9.1 Identification of a cosmid containing GE81112 biosynthetic genes

As the results described in section 3.2.5 led us to conclude that the identified gene clusters were not involved in the biosynthesis of GE81112, we developed a new strategy to identify the GE81112 gene cluster. The previous cyclodeaminase probe successfully identified a cyclodeaminase, but it was truncated, and therefore presumed to be inactive. However, as such pseudo-genes often arise by gene duplication of an original active copy; we reasoned that there might be a second, complete cyclodeaminase gene in the genome, which should be functional. We therefore decided to generate a new specific cyclodeaminase probe based on the cyclodeaminase sequence found on cosmid FD10. In order to amplify the cyclodeaminase

gene, a PCR reaction was carried out using specific primers (Table 2.12, Cyclo Probe_for and Cyclo Probe_rev). The resulting 400 bp PCR product was labeled with digoxigenin and used to re-probe the cosmid library. To evaluate the likely success of this strategy, we first carried out a Southern Blot to check if the probe hybridized to more than one gene in the genome. Digestion of cosmid FD10 with *Pvu*II produced a 2.5 kb fragment which encompasses the cyclodeaminase sequence. Thus, if the FD10 cyclodeaminase were the only cyclodeaminase in the genome, we would have observed a single 2.5 kb band by Southern Blot.
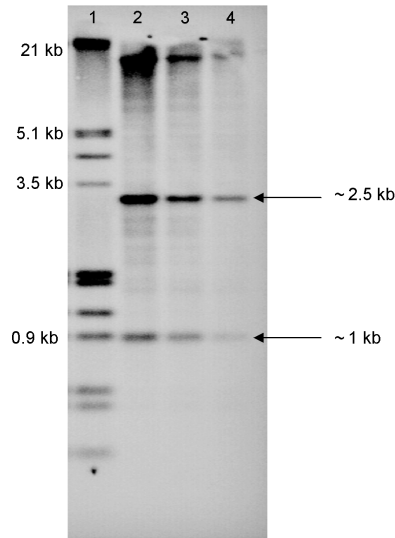
Figure.3.37: **Southern Blot analysis of genomic DNA from *S.* 14386.**

1: Ladder; 2–4: varying concentrations of genomic DNA from *S.* 14386 digested with *Pvu*II.

For the Southern, genomic DNA of *S.* 14386 was isolated and digested with *Pvu*II. Hybridization was carried out at 42 °C under stringent conditions using the new cyclodeaminase probe. Southern Blot analysis showed the expected 2.5 kb band (Figure 3.37). However, two additional signals were detected at approximately 1 kb and 13 kb (Figure 3.37), suggesting that the probe hybridized a total of three times with genes in the genome (1 kb, 2.5 kb and 13 kb band). We assumed that the hybridization had been successful, as non-specific hybridization would likley have resulted in more signals. Thus, we concluded that the genome was likely to contain a second or even a third copy of the cyclodeaminase gene. As the new probe seemed to function well, we aimed to screen the cosmid library again with this probe to identify cosmids that had not hybridized with the old cylodeaminase probe. Hybridization was again carried out at 42 °C (Figure 3.38).
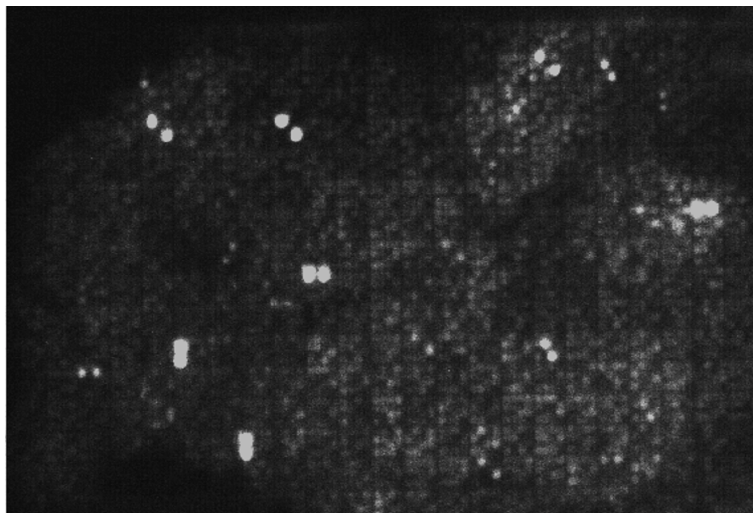
Figure 3.38:   **Screening of *S.* 14386 cosmid library with the specific cyclodeaminase probe.**

The hybridization was successful, as only nine distinct signals were obtained, in addition to that with cosmid FD 10 (positive control). To verify the presence of cyclodeaminase sequence on these nine cosmids, we performed a PCR analysis on the cosmids using the primers employed to generate the probe. PCR products of ca. 400 bp were obtained for all cosmids and the control FD10, confirming in each case that the cyclodeaminase sequence was present. As we expected that NRPS sequence must also be present on the cosmid, we additionally performed a PCR using the degenerate NRPS primers RevA3 and PSLGG [88]. Unfortunately, no PCR products were obtained with these primers. One possible explanation for this result, however, is that the NRPS degenerate primers RevA3 and PSLGG were not optimal. Indeed, attempts by other members of the laboratory to use these primers to amplify A domains from *Streptomcyes* genomes have not been successful, perhaps because they were originally designed for use with myxobacterial DNA.

We therefore designed new primers based on *Streptomyces* A domains using the CODEHOP software [96]; priming was targeted against the core A3 and A6 motifs. The new primers were then used to amplify A domains from the nine cosmids that produced products of the correct size in the cyclodeaminase PCR. This reaction resulted in PCR products from four of the nine cosmids. These fragments were cloned into pCR2.1-TOPO and sequenced. BLAST analysis revealed that all of the sequences showed high homology to A domains. We then analyzed the A domain sequences as described earlier, to predict the substrate specificity of each A domain (Table 3.9).

Table 3.9:   **Putative A domain 'code residues' within A domains from cosmids**

	Position of the amino acid within the A domain								Activated amino acid	Identity to protein
	235	236	239	278	299	301	322	330		
Topo BG23	D	A	S	Q	V	G	E	V	Asp/Asn	75%
Topo DB20	D	A	V	D	F	G	T	I	No hit	
Topo DK18	D	A	V	D	F	G	T	I	No hit	
Topo BI11	D	V	Q	D	I	A	H	M	Pro/Pip	70%

This analysis led to a prediction that the A domain in cosmid BI11 would be specific for proline/pipecolic acid. This result was supported by BLAST analysis of the A domain which predicted homology to a "NRPS for pipecolate incorporation, *Streptomyces* sp. NRRL 30748". We also directly compared the nonribosomal code of this A domain to those of known pipecolic acid-incorporating A domains from the rapamycin and tubulysin systems (Table 3.10). The code residues in each case were similar, supporting the probable specificity for pipecolic acid.

Table 3.10:   **Substrate specificity pocket of pipecolic acid incorporating A domains**

	Position of the amino acid within the A domain								Activated amino acid
	235	236	239	278	299	301	322	330	
A domain RapP	D	Y	Q	Y	C	G	H	L	Pip
A domain BI11	D	V	Q	D	I	A	H	M	
A domain TubB	D	I	Q	Y	I	A	Q	V	Pip

Taken together, these data strongly suggested that cosmid BI11 contained the biosynthetic pathway for GE81112. Furthermore, as we anticipated that the NRPS would contain four modules (30–40 kb), it was possible that entire gene cluster would be located on a single cosmid. To verify this, we end sequenced the cosmid using primers T4 and T7 (Table 2.12.).

Table 3.11: End sequencing results of cosmid BI11 with T7 and T4 primers

Cosmid	Homology	Identity	Score	E-Value
BI11_T4	DNA binding protein, *Streptomyces coelicolor*	36%	143	1e-32
BI11_T7	L-arginine beta-hydroxylase, *Streptomyces vinaceus*	44%	331	4e-89

The T4 end exhibited homology to a DNA binding protein, while sequence at the T7 showed similarity to an arginine β-hydroxylase (Table 3.11). The DNA binding protein was not expected to have any function in the biosynthesis, and thus cosmid BI11 contained at least one end of the gene cluster. However, it was possible that the β-hydroxylase was involved in the biosynthesis, so we could not conclude that the entire cluster was located on the cosmid. Nonetheless, we sequenced BI11, as at least a significant portion of the cluster was likely to be present.

3.2.9.2 Identification of overlapping cosmids

To find a cosmid which overlapped with the 3′ (T7) end of the GE81112 biosynthetic gene cluster, we amplified a 1 kb fragment from the T7 end of cosmid BI11 to serve as a probe. The sequenced PCR product was then used as a template to carry out a PCR reaction using digoxigenin-labeled nucleotides. The cosmid library was then screened with the new probe, with hybridization at 42 °C.
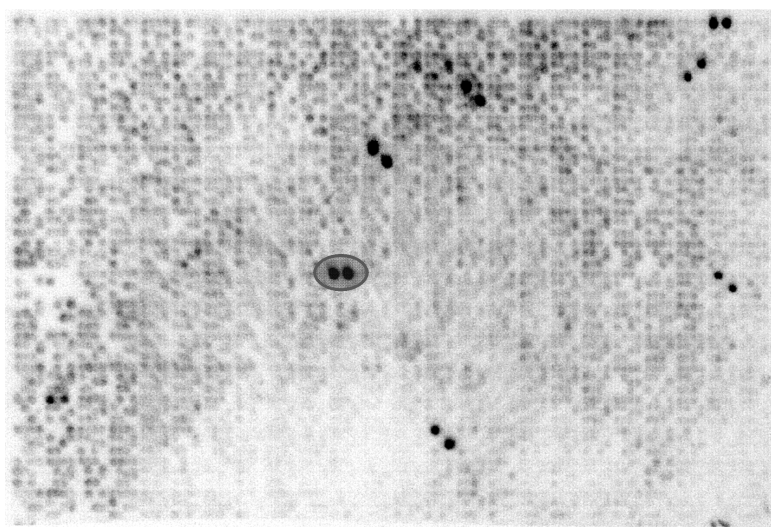
Figure.3.39:   **Screening of the cosmid library to identify cosmids which overlapped the T7 end of cosmid BI11.** Cosmid BI11is labeled.

Seven new cosmids were identified, in addition to cosmid BI11 (marked in Figure 3.39). We initially investigated the cosmids by PCR analysis by amplifying an 800 bp fragment with the primers used to generate the probe (cyclo probe_for and cyclo probe_rev). All of the cosmids produced a positive signal. However, as some of the cosmids gave a similar restriction pattern to that of BI11, we excluded them from further analysis. We chose one cosmid (BA23) which showed the least similar fragmentation pattern and analyzed it further. To ensure that both cosmids (BI11 and BA23) contained similar portions of the cluster, and to determine whether the cloned DNA was co-linear with the *S.* 14386 chromosome, we performed a Southern Blot with both the cosmids and the wild type genomic DNA. In theory, the probe that derives from the T7 end of cosmid BI11 should produce identical signals when hybridized to the cosmids and the wild type DNA. As hybridization was observed to the same fragments in all cases (Figure 3.40), the cosmids were verified to be co-linear with the genome.
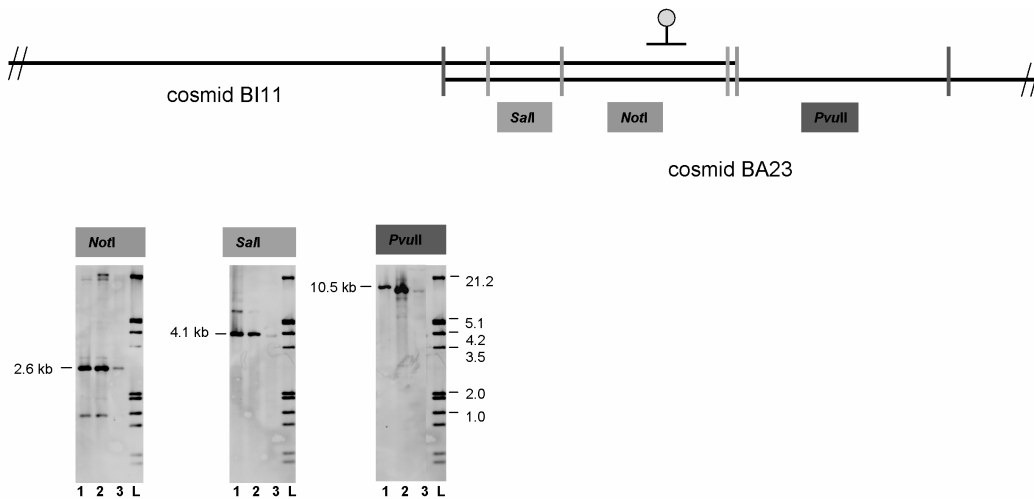
Figure 3.40:   **Southern Blot analysis of cosmid BI11 and the overlapping cosmid.**

Schematic arrangement of cosmids BI11 and BA23. To verify the colinearity between the cosmids and the genome, the cosmids and the genomic DNA were digested with the same enzymes (*Not*I, *Sal*I and *Pvu*II). The probe was designed against the T7 end of cosmid BI11, as shown. 1: Cosmid BI11, 2: cosmid BA23, 3: genomic DNA, L: Ladder

3.2.9.3 Annotation

The two cosmids BI11 and BA23 were sequenced in the HZI. The obtained sequence was analyzed for the presence of putative open reading frames (*orfs*) with FramePlot 2.3.2 [92], and preliminary functional assignments of individual *orfs* were made by comparison of the deduced gene products with proteins of known function in the BLAST database (Table 3.12).

Table 3.12:   **BLAST results for genes on the cosmids BI11 and BA23**

Gene	Homology	Identity	Score	E-Value
orf1	putative partitioning protein ParA, *Nocardia farcinica* IFM 10152	63/222 (29%)	77.0	8e-13
orf2	hypothetical protein SAV_5443 *Streptomyces avermitilis* MA-4680	263/478 (55%)	470	1e-130
orf3	FtsK/SpoIIIE family protein *Streptomyces avermitilis* MA-4680	802/1099 (72%)	1565)	0
orf4	hypothetical protein SAV_6995 *Streptomyces avermitilis* MA-4680	118/142 (83%)	244	2e-63
orf5	serine/threonine protein kinase *Streptomyces avermitilis* MA-4680	249/381 (65%)	395	5e-108
geA	thioesterase type II *Streptomyces atroolivaceus*	89/230 (38%)	142	4e-32
geB	putative ABC-type multidrug transport system ATPase and permease component, *Streptomyces griseus*	269/563 (47%)	449	4e-124
geC	putative ABC transporter ATPase and permease component, *Streptomyces griseus*	291/575 (50%)	533	1e-149
geD	TubZ protein *Angiococcus disciformis*	182/346 (52%)	337	1e-90
geE	syringopeptin synthetase *Pseudomonas syringae*	209/522 (40%)	325	7e-87
geF	putative L-proline 3-hydroxylase protein *Ralstonia solanacearum* GMI1000	81/253 (32%)	123	2e-6
geG	peptide synthetase protein, *Ralstonia solanacearum* GMI1000	765/2401 (31%)	857	0
geH	non-ribosomal peptide synthetase *Myxococcus xanthus* DK 1622	168/530 (31%)	194	2e-47
geI	L-arginine β-hydroxylase *Streptomyces vinaceus*	144/322 (44%)	238	6e-61

Table 3.12:   **Continuation of Table 3.12**

Gene	Homology	Identity	Score	E-Value
geJ	bacitracin synthetase 3; BacC, *Bacillus licheniformis*	223/598 (37%)	368	1e-99
geK	ribosome-associated GTPase *Streptomyces sviceus* ATCC 29083	209/368 (56%)	395	3e-108
geL	halogenase, *Microcystis aeruginosa*	199/500 (39%)	376	2e-102
geM	bacitracin synthetase 3; BacC,*Bacillus licheniformis*	199/500 (39%)	397	2e-108
geN	ChlK, thioesterase type I, *Streptomyces antibioticus*	44/112 (39%),	80.5	4e-14
orf6	3-oxoacid CoA-transferase subunit B *Streptomyces avermitilis*, MA-4680	177/215 (82%)	359	1e-97
orf7	protocatechuate 3,4-dioxygenase α subunit, *Streptomyces avermitilis* MA-4680	185/253 (73%)	375	2e-102
orf8	protocatechuate 3,4-dioxygenase β subunit, *Streptomyces avermitilis* MA-4680	129/182 (70%)	253	5e-66
orf9	3-carboxymuconate cycloisomerase *Streptomyces sviceus* ATCC 29083	319/447 (71%)	570	8e-161
orf10	deoxyribose-phosphate aldolase *Streptomyces avermitilis* MA-4680	150/205 (73%)	285	4e-75
orf11	arthrofactin synthetase/syringopeptin synthetase C-related non-ribosomal peptide synthetase module, *Bradyrhizobium sp*. BTAi1	513/1186 (43%)	777	0.0
orf12	peptide synthetase protein, *Ralstonia solanacearum* GMI1000	221/656 (33%)	244	3e-62
orf13	putative ATP/GTP binding protein *Streptomyces coelicolor* A3(2)	178/193 (92%)	338	3e-91

Table 3.12:   **Continuation of Table 3.12**

Gene	Homology	Identity	Score	E-Value
orf14	integrin-like protein, *Streptomyces avermitilis* MA-4680	200/468 (42%)	299	3e-79
orf15	two-component system sensor kinase, *Streptomyces coelicolor* A3(2)	107/218 (49%)	179	2e-43
orf16	polyketide synthase type I, *Streptomyces aizunensis*	228/435 (52%)	359	4e-97

Based on the chemical structure of GE81112 we expected a NRPS pathway involving a cyclodeaminase and a halogenase. Annotation of the two cosmids revealed 30 *orfs* of which 14, designated *geA–N*, are postulated to be involved in the GE81112 biosynthetic pathway (Figure 3.41). The genes marked in black indicate NRPS encoding enzymes.

The sequenced region starts with the gene *orf1*, which shows homology to a putative partitioning protein ParA from *Nocardia farcinica* IFM 10152. This protein is not expected to be involved in the biosynthesis, as these proteins are known to be involved in bacterial mitosis [97]. Similarly, we excluded *orfs 2* and 3, as they do not show homology to proteins postulated to be involved in the biosynthesis. *orf*2 is a hypothetical protein with no assigned function, and *orf3* encodes for a FtsK/SpoIIIE family protein which belongs to the family of DNA translocases [98]. It is not possible, however, to conclusively identify the boundaries of a gene cluster based solely on the deduced functions of the gene products. Therefore, in future, it will be necessary to carry out gene knockouts to confirm the proposed cluster ends.
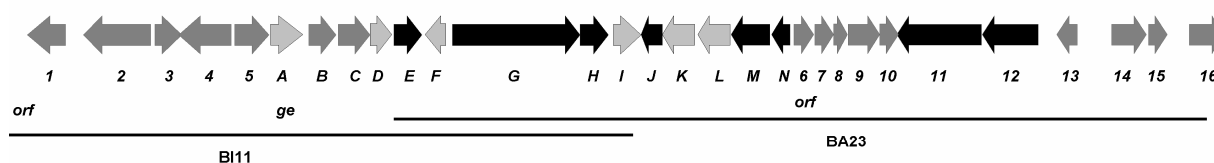

Figure 3.41:   **GE81112 biosynthetic gene cluster on cosmids BI11 and BA23.**

Key: black, NRPS biosynthetic genes; light grey, modification genes; dark grey, *orfs* not involved in the biosynthesis.

The first gene that we predicted to be involved in the biosynthesis is gene *geA*, which encodes a type II thioesterase (TEII). TEIIs are present in many NRPS and PKS systems, where they

perform crucial proof-reading functions by hydrolyzing aberrant substrates from the respective carrier protein domains. In modular PKS, type II TEs are proposed to release from the ACP domains acyl groups that have been produced by aberrant decarboxylation of chain extender units [99]. In NRPS systems, they are proposed to play two functions: removal of acyl groups added to the PCP domains during post-translational priming of the *apo* proteins, and release of amino acids that have been loaded by mistake [100].

geA appears to be in an operon with genes *geB–E*. *geB* and *geC* encode proteins which show homology to ABC transporter systems. Many antibiotic-producing actinomycetes posses at least one ABC (ATP-binding cassette) transporter which forms part of the antibiotic biosynthetic pathway, and in most cases confers resistance to the drug in a heterologous host. In Type I ABC transporters, two genes are involved, one encoding a hydrophilic ATP-binding protein with one nucleotide-binding domain, and the other encoding a hydrophobic membrane protein. The same type of proteins have been found in the GE81112 cluster [101]. *geD* encodes the cyclodeaminase gene that was targeted in our second probing strategy. It shows 52% identity to TubZ, the cyclodeaminase from *Angiococcus disciformis*. This time the protein was not truncated as in cosmid FD10 and was proposed to be functional. The sequence of the truncated cyclodeaminase gene on cosmid FD10 is identical to a region within *geD*, consistent with the idea that the truncated form arose from gene duplication.

geE is the first gene that encodes for a NRPS protein, and most likely starts with a GTG. A putative RBS (GGAG) was found 7 bp upstream of the gene. The next gene *geF* is oriented in the opposite direction, and encodes a protein with homology to a putative L-proline 3-hydroxylase. This enzyme is expected to be responsible for the hydroxylation of the pipecolic acid. The following gene, *geG*, encodes another NRPS protein, and is the likely starting point of a new operon which includes genes *geH* and *geI*. *geG* has an ATG start codon and a putative RBS (GAAGG) is located 19 bp upstream from the start codon. *geH* encodes a third protein with homology to NRPSs, and again starts with a GTG. The last gene in this operon is *geI*, which exhibits homology to an L-arginine β-hydroxylase, which might be responsible for the hydroxylation of one of the incorporated histidines in GE81112. The last operon contains 5 genes *(geJ–geN)*. This operon starts with a change in the transcription direction again. The operon begins with the gene *geJ* which encodes a NRPS enzyme. The next two genes, *geK* and *geL*, encode a GTPase protein and a halogenase, respectively; GeL may be responsible for the chlorination of the GE81112 histidine. The NRPS-enconding gene *geM* starts with a

TTG, and a RBS (AGGG) is located 6 bp upstream of the gene. The last gene of the operon is *geN*, which encodes a protein with homology to a type I thioesterase. A RBS (GAAA) is located 8 bp upstream of the gene which starts with an ATG. The involvement of o*rfs* 6–16 can not yet be excluded.

3.2.9.4 Analysis of NRPS domains

The sequence analysis revealed 14 open reading frames that we predicted to be part of the gene cluster. Five of them showed homology to NRPS genes (Figure 3.41). The constituent domains were assigned using the PKS/NRPS predictor (http://www.tigr.org/jravel/nrps/) and confirmed by manual inspection with BLAST, as the prediction programs can omit some domains.
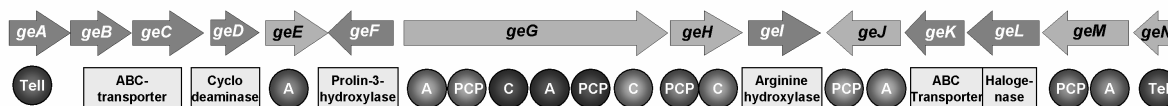

Figure 3.42:   **The GE81112 biosynthetic gene cluster.**

In order to incorporate 4 amino acids into the GE81112 structure, we expected the synthetases to contain 4 modules for the biosynthesis of a tetrapeptide including a loading module followed by three condensation module with the standard C-A-PCP arrangement. However, the sequence analysis revealed that the cluster exhibits a highly nonlinear arrangement that did not fit to what we expected. Although the domain complement for the biosynthesis of a tetrapeptide could be identified one extra module consisting of an A-PCP didomain could be identified. Furthermore the modules showed a highly split arrangement (Figure 3.42). Thus, from the domain assignment alone, the overall order of subunits could not be discerned, and so we carried out further analysis of the domains themselves. We started with the analysis of the A domains, as we hoped that the predicted substrate specificities would give us a first hint about the order of the domains.

3.2.9.4.1   Analysis of the A domains

To determine the order of the A domains in the GE81112 cluster, we analyzed the substrate specificity of the A domains. The eight conserved motifs were identified for each A domain, and the specificity-conferring code (bioinformatics and manual prediction) was then used to

predict the substrate specificity. The first A domain, encoded by gene *geE* (GeEA1) as a discrete protein, showed specificity for the incorporation of proline/pipecolic acid (see section 3.2.9.1). This observation made it a candidate for selection of the GE81112 starter unit. The second A-domain located on gene *geG* (GeGA2) showed homology to ornithine/glutamine/asparagine-incorporating A domains. The third A domain on gene *geG* (GeGA3) was predicted to be specific for the incorporation of tyrosine/tryptophane, and the two latter domains on *geJ* (GeJA4) and on *geM* (GeMA5) for histidine (Table 3.13).

Table 3.13:   **Putative A domain 'code residues' within A domains in the GE81112 cluster**

	Position of the amino acid within the A domain									
	235	**236**	**239**	**278**	**299**	**301**	**322**	**330**	Predicted amino acid	Identity to protein
GeEA1	D	V	Q	Y	I	A	Q	V	Pro/Pip	70%
GeGA2	D	A	Y	N	L	G	L	I	Orn/Gln/Asp	70%
GeGA3	D	A	V	G	V	G	E	V	Tyr/Trp	70%
GeJA4	D	S	A	S	T	A	E	V	His	70%
GeMA5	D	S	A	L	T	A	E	V	His	70%

These results correlated well with our prediction that an ornithine and two histidines are incorporated in the the GE81112 metabolites, in addition to the pipecolic acid. However, it remained unclear why two PCP-A didomains (*geM* and *geJ*) are located within the cluster, although only one is required to give a total of four active modules. To check if any of the domains were inactive in order to account for the prescence of the apparently superfluous module, we continued by analyzing the C and PCP domains.

3.2.9.4.2   Analysis of the C and PCP domains

The GE81112 cluster contains three C domains which were aligned with C domains from the rapamycin [102], the gramicidin [103] and the calcium-dependent antibiotic [104] clusters. The seven conserved regions were identified in all of the C domains. The same analysis was carried out for the PCP domains, revealing the signature sequence and active Ser residue, in all cases. These results demonstrate that in principle, all modules are active, so that it is unlikely that one of the modules is skipped.

3.2.9.4.3   Analysis of the TE domains

TE domains are characterized by a conserved signature sequence GxSxG [91]. There are two classes of TE domains known in NRPS systems. TEI domains are normally integrated into the multienzyme subunits, and located at the C-terminal end of the modules which are involved in adding the last amino acid to the linear peptides [91]. These TEs are responsible for cleavage of the linear peptide products. Type II TEs are normally discrete proteins that perform a crucial proof-reading function by hydrolyzing aberrant substrates from the respective carrier proteins [100;105]. We could identify two discrete TE domains (on genes *geA* and *geN*) in the GE81112 biosynthetic gene cluster which showed the conserved motif, indicating that they are functional enzymes. As they were both discrete proteins, it appeared as if they were both TEII enzymes. However, while BLAST analysis (see Table 3.12) revealed that the protein GeA showed high homology to a type II TE, protein GeN was similar to type I TEs. This finding was surprising, as GeN is a discrete protein like a type II TE, and not integrated into an NRPS as is typically found in bacterial NRPS systems.

3.2.9.4.4   Docking domain analysis

The analysis of all domains with the NRPS subunits revealed that they all contained the respective conserved regions, indicating that all proteins/domains are probably functional. In the vast majority of NRPS complexes, the modules are distributed over two or more NRPSs which have to interact selectively with each other to bring about the synthesis of a defined product. The molecular basis for the selective interaction between NRPSs are the so called 'communication-mediating (COM)' domains [106;107]. A donor COM domain (Com^D) located at the C terminus of an aminoacyl- or peptidyl-donating NRPS and an acceptor COM Domain (Com^A) located at the N terminus of the accepting partner NRPS form a matching set, required for the proper intermolecular interaction between adjacent modules. In contrast, Com^D and Com^A domains of nonpartner NRPSs are considered nonmatching, preventing false contact between enzymes [106]. The C-terminal COM domain, which consists of approximately 20–30 amino acids, most often follows a PCP or E domain, while the N-terminal COM domain, which typically precedes a C or HC domain, comprises 15–25 amino acids. Although the overall sequence similarity between COM domains is low, the amino acid composition is relatively uniform. The C-terminal COM domains incorporate a higher than average proportion of acidic amino acids, while N-terminal COM domains are biased toward

polar residues, suggesting that specificity is mediated largely by polar/or electrostatic interactions between key residues.

To obtain insight into how the different NRPS proteins from the GE81112 cluster might interact with each other to form a defined biosynthetic template, we analyzed the GE81112 NRPS proteins for putative COM domains. In contrast to linear NRPS, the GE81112 system is a nonlinear systems in which many of the modules are split (Figure 3.43). Thus, we devised an NRPS pathway in which individual multienzymes interact to reconstitue standard C-A-PCP modules (for example GeE and GeH, Figure 3.43). This task was aided by the *in silico* prediction of the substrate specificity of the A domains which gave us some hints as to the ordering of the proteins within the pathway.
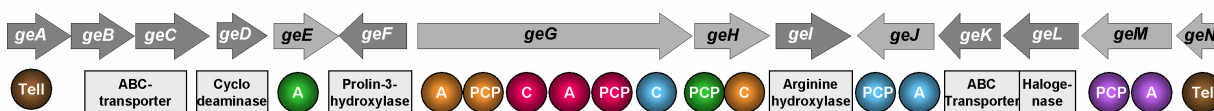

Figure 3.43:   **GE81112 biosynthetic gene cluster.**

Domains that are postulated to interact with each other are highlighted in the same color.

We predict that protein GeE (harboring the single A domain GeEA$_1$ which is predicted to incorporate pipecolic acid) interacts with subunit GeH to form the loading module. Interaction of GeH with GeG would then form the next two modules (A domains GeGA$_2$ (Orn/Gln) and GeGA$_3$ (Trp/Tyr)). Protein GeG would than interact with either GeJ (A domain GeGA$_4$ (His)) or GeM (A domain GeGA$_5$ (His)) to form the last module. The protein GeN has a TEI function and is therefore proposed to hydrolyze the final tetrapeptide from the assembly line.
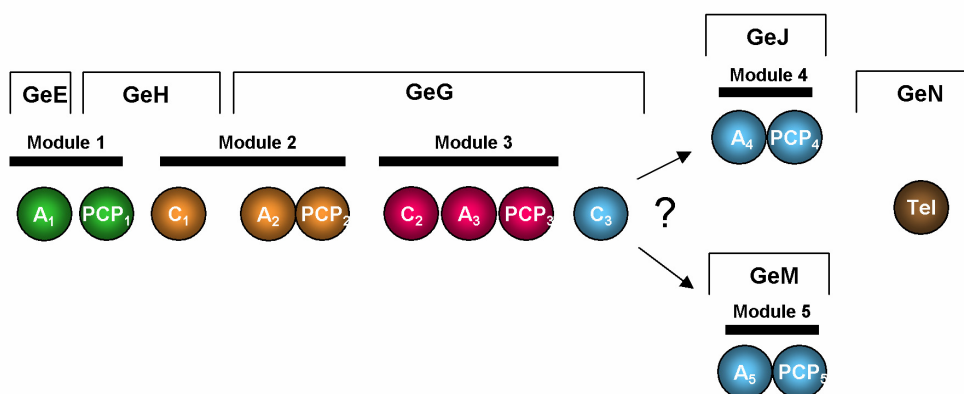

Figure 3.44:   **Predicted interaction of NRPS enzymes from the GE81112 biosynthetic gene cluster.**

It is not clear if proteins GeG and GeJ or GeG and GeM interact with each other to form the last module.

In contrast to standard COM domains which normally facilitate the interaction between PCP and C domains located on separate multienzymes, intersubunit protein-protein interactions in the GE81112 system must reconstitute interfaces between C and A domains (between proteins GeH/GeG, and betwee GeG/GeJ or GeG/GeM, see Figure 3.44) to provide a functional pathway. To determine if we could identify COM domains between C and A domains, GeHC$_1$ and GeGA$_2$ (module 2), and GeGC$_3$ and GeJA$_4$ (module 4) or GeMA (module 5), we constructed a sequence alignment (attached in the appendix, Table 5.1) comprising putative modules 1–5. In this alignment, the sequences of modules 1–5 (Figure 3.44) were compared to standard C-A-PCP modules (modules from NRPS pathway from cosmid FD10). This analysis revealed large putative recognition sequences at the C terminus of the C domains and the N terminus of the A domains in modules 2 and 4 (Figure 3.45).
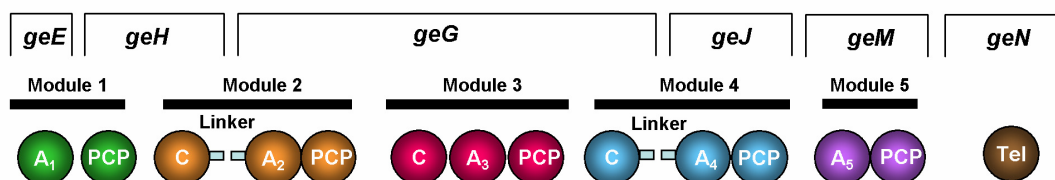

Figure 3.45: **Proposed NRPS interaction and order of subunits in the GE81112 biosynthesis.**

However, these recognition sequences did not match the standard size of COM domains (20–30 amino acid residues). Instead, we found stretches of approximately 100 amino acid residues at each terminus; to reflect this difference, we've designated the sequences as 'recognition elements' instead of COM domains. Nonetheless, the C-terminal recognition elements incorporate a high proportion of acidic amino acids while the N-terminal linker domains are biased towards polar amino acids, providing the basis for electrostatic interactions as in classical COM domains (the recognition elements contained approximately the same proportion of acidic and polar amino acid residues as COM domains). Such recognition sequences are normally not found between C and A domains, indicating that in this special case, these sequences have a docking function, enabling the interaction between NRPS proteins.

At this stage, it remained unclear why the cluster contains an apparently superfluous fifth module. One hypothesis was that both bidomains GeJ and GeM can interact with the C domain of protein GeG to form the fourth module of the NRPS, incorporating histidine in either case. To evaluate this hypothesis, we aligned the sequences of proteins GeJ and GeM,

in order to evaluate the extent of their sequence similarity (Table 5.2, appendix). We reasoned that a high degree of sequence identity would provide evidence for the evolution of the two didomains by gene duplication, supporting the idea that they might both be able to interact with the C terminal C domain from GeG. However, the alignment showed that the proteins exhibit only 25% mutual sequence identity, consistent instead with a lack of shared partner specificity. Indeed, GeM lacks the putative recognition sequence for GeG which is present at the N-terminus of GeJ. Taken together, these data indicated that the didomain protein GeM was not derived from duplication of GeG, and should therefore have a specific function in the pathway (Figure 3.45). To provide further support for the proposals concerning domain order and functions of GeJ and GeM, we investigated the gene cluster using both gene inactivation and expression analysis of recombinant proteins.

3.2.10 Insertional mutagenesis of the GE81112 biosynthetic gene cluster

To first verify the identity of the gene cluster, we aimed to inactive the gene *geE*. For this, a knockout construct was designed by amplifying an internal fragment (660 bp) of the gene, which was then cloned into the knockout vector pKC1132 [10] leading to the construct pKC1132_BI11_PipA. Initial attempts to transform *S.* 14386 with the knockout plasmids pKC1132_FD10 and pKC1132_AI6 however, were unsuccessful, necessitating the development of a revised transformation method. The classical procedures for streptomycetes transformation are protoplast transformation and conjugation [10]. We tried both methods with the knockout vector, using a replicative vector pOJ446 [10] and an integrative vector pSET152 [10], as controls. The protoplast transformation method did not produce any exconjugants using the knockout vector or with the controls. The conjugation was then attempted with varying amounts of *E. coli*. In the standard procedure, a certain amount (normally 500 µl) of *E. coli* culture (OD_{600} = 0.6) is mixed with an equal volume of *Streptomyces* cells. Using this approach, we only obtained a low number of colonies with the replicative and the integrative vector controls, and no colonies with the knockout construct (Table 3.14). We then tried 100-fold and 1000-fold more *E. coli* cells, but obtained the same result.

Table 3.14:   **Transformation efficiency of *S.* 14386 by biparental conjugation**

Transformation Method	*E. coli* amount 0.5×10⁸			*E. coli* amount 50×10⁹		
Biparental conjugation	**pKC1132_BI 11_PipA**	**pOJ446 replicative**	**pSET152 integrative**	**pKC1132_B I11_PipA**	**pOJ446 replicative**	**pSET152 integrative**
	–	20	50	50–100	100–200	500–1000

We next modified the procedure by mixing 5 mL of an overnight culture of *E. coli* which had a very high cell density, with 5 ml of *Streptomyces* cells. This change finally led to several exconjugants containing the knockout vector, indicating that the amount of *E. coli* is crucial for the conjugation efficiency with this strain. The obtained mutants were verified by PCR amplifying the ampramycin resistance gene as well as an internal region from the genomic DNA, proving the correct integration (primers: apra_for and apra_rev, lacZ1 and lacZ2 and BI11A1_pip_for and BI11A1_pip_rev) (section Table 2.12). The mutants were cultivated in production medium, as before. The extracts were then analyzed for the presence of the GE81112 compounds by high-resolution MS (Figure 3.46).
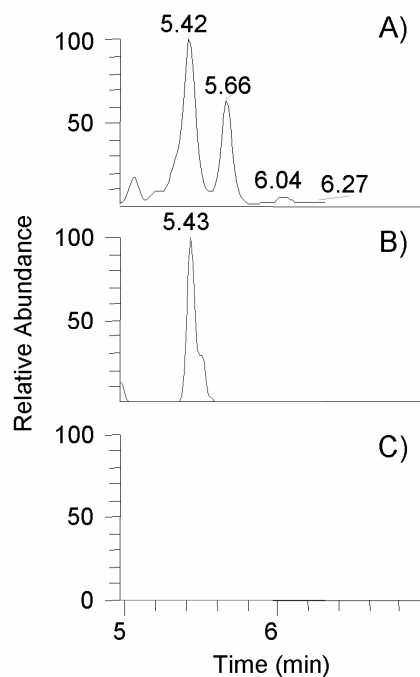
Figure 3.46:   **High resolution MS analysis of extracts of _S._ 14386 wild type and _S._ 14386::pKC1132 _BI11_PipA mutant.**

A) BPC of _S._ 14386 wild type extract. B) EIC of GE81112 factor B showing a molecular ion of the mass _m/z_ = 659.22953 [M+H]$^+$ from _S._ 14386 wild type extract. C) EIC of GE81112 B showing the analysis for a molecular ion of the mass m/z = 659.22953 [M+H]$^+$ of _S._ 14386::pKC1132_BI11_PipA

This analysis clearly showed that production of GE81112 factor B (mass _m/z_ [M+H]$^+$ = 659.22953) was abolished in the mutant, verifying the identity of the gene cluster. As an alternative approach we tried the heterologous expression of the cosmids, as described before 3.2.6.1. Heterologous expression couldn't be observed from the cosmids. This was not surprising as they both didn't contain the whole cluster.

3.2.11 Feeding experiments

Inspection of the GE81112s metabolite structures suggests that they incorporate modified versions of pipecolate, histidine and ornithine/glutamine. Thus, it was possible that the corresponding proteinogenic amino acids were incorporated and then modified, or alternatively that the modifications occurred prior to activation of the building blocks by the A domains. To investigate this question directly, we designed feeding experiments with labelled substrate. As the stereochemistry of GE81112 compounds has not yet been solved, we decided to feed separately both D- and L-pipecolic acid. In an independent experiment, we also fed U^{13} C-labelled histidine. Unfortunately, we did not detect any incorporation of the

labeled amino acids into GE81112, a result which we attribute to the very low production yield of the GE81112 compounds. We also did not observe any incorporation of the labeled amino acids into the newly- discovered substances 631, 645 and 659. As these substances are produced in reasonable amounts and could be clearly detected in HPLC-MS, we could exclude production yield as the underlying problem. Thus, it is likely that none of the amino acids used in this experiment could be incorporated into the substances. These findings are consistent with the structure elucidation results, as to date, no structural elements similar to pipecolic acid or histidine have been found.

3.2.12 Biochemical characterization of the GE81112 biosynthetic gene cluster

An alternative approach to feeding experiments to evaluate the proposed biosynthesis is to express key enzymes involved in substrate biosynthesis or selection in recombinant form, and assay them *in vitro*. As we still had no experimental proof that the expected amino acids are incorporated from the A domains proposed in section 3.2.9.4.4 we tried to express three of the A domains in recombinant form in *E. coli* to assay them *in vitro*. Additionally the cyclodeaminase was chosen for recombinant expression to confirm its proposed function.

3.2.12.1 **Expression and purification of the lysine cyclodeaminase GeD**

Lysine cyclodeaminases have been shown *in vitro* to generate L-pipecolic acid from L-lysine [82;83;108] (and Y. Chai *et al.* unpublished data). The enzymes were obtained in recombinant form from *E. coli* [108] and in some cases it could be determined that the L-stereoisomer of pipecolic acid was formed. Based on this precedent, we intended to express the cyclodeaminase *geD* from the GE81112 cluster in *E. coli* as a C-terminal fusion to GST, and to test its function in an assay *in vitro*. For this, we amplified the open reading frame of *geD* from cosmid BI11 and cloned it as an *Eco*RI-*Not*I fragment into the expression vector pGEX-6P-1 (Figure 3.47).
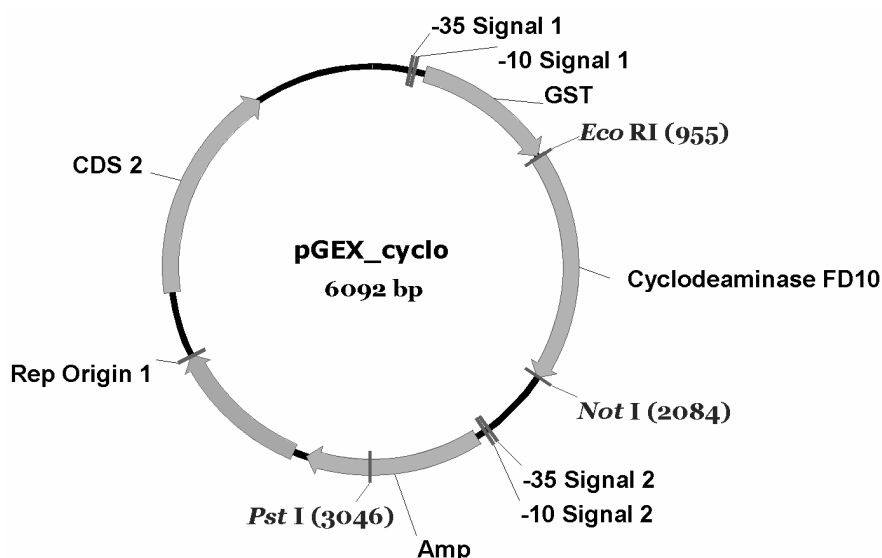
Figure 3.47: **Construct pGEX_cyclo for the expression of geD as a GST fusion protein in _E. coli._**

The expression construct was sequenced, and then transformed into _E. coli_ BL21 by electroporation. Cultivation was then carried out at both 16 °C and 30 °C. Analysis of 10 µl of the supernatant and the pellet using a 12% SDS polyacrylamide gel showed that most of the protein was present in the pellet, following growth at both 16 °C and 30 °C (Figure 3.48).
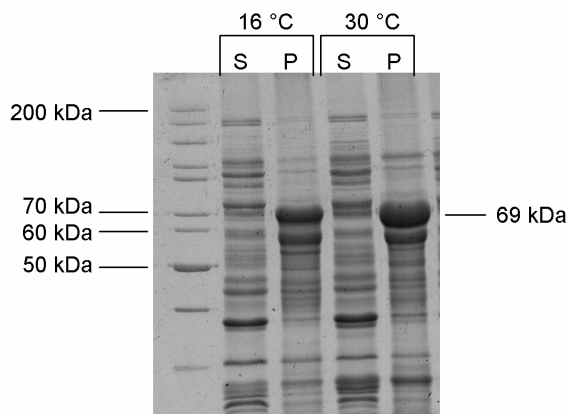

Figure 3.48: **Expression of the construct pGEX_cyclo in _E. coli_ BL21.**

Protein expression was carried out at 16° and 30 °C. The expected GST-fusion protein (69 kDa) was only present in the pellet. Key: P, pellet; S, supernatant.

To attempt to obtain soluble protein, we tried expression in _E. coli Rosetta_ BL21 (DE3) pLysS/RARE. In addition, we evaluated whether two sets of chaperones (Chaperone Plasmid Set, www.takara-bio.eu) would increase the solubility of the protein in _E. coli BL21._ Chaperone set 1 contained chaperones GroE, GroL, trigger factor and chloramphenicol (cm)

resistance [109], while chaperone set 2 contained chaperones GroE, GroL and cm [110] (*E. coli Rosetta* BL21 (DE3) pLysS/RARE was not suitable as an expression host for the chaperones because it already carries a plasmid with chloramphenicol resistance). Both strains were cultivated as described in section 2.18.1.

E. coli Rosetta BL21 (DE3) pLysS/RARE/pGEX_cyclo was induced with 0.1 mM IPTG, and the *E. coli* strains containing the chaperones with 1 mM IPTG (according to the manufacturer's instructions). Cultivations were performed at 16 °C for all three expression cultures. As before, most of the protein was found to be insoluble. However, during expression with chaperone 1 in *E. coli* BL21 and independently in *E. coli Rosetta* BL21 (DE3) pLysS/RARE, some soluble protein was present in the supernatant (Figure 3.49).
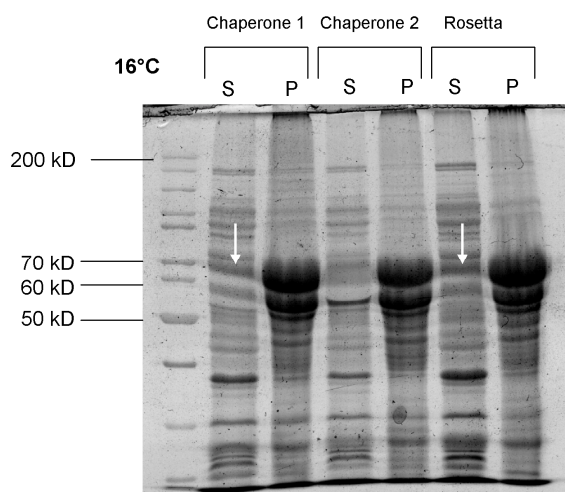
Figure 3.49: **Expression of the cyclodeaminase GeD.**

Expression was performed at 16 °C with chaperone set 1, chaperone set 2 and in *E. coli Rosetta* BL21 (DE3) pLysS/RARE. Soluble protein is indicated by the arrows. Key: P, pellet; S, supernatant.

The GST-fusion protein was purified from the crude cell extract using GST minicolumns, coupled with on-column digest using PreScission Protease. The final yield of purified GeD from expression in *E. coli* Rosetta was approximately 0.8 mg/ml protein. Figure 3.50 shows the GST fusion protein (26 kDa) which was eluted with reduced glutathione, and the cleaved GeD (43 kDa) after on-column cleavage with PresScission Protease.

Figure 3.50:   Purified protein GeD.

GeD expression in *E. coli Rosetta* BL21 (DE3) pLysS/RARE and co-expression with chaperone set 1. The gel shows the protein with the GST tag and following cleavage with PreScission Protease. Cleaved protein is indicated by the arrows.

The identity of the protein (from both samples) was verified by MALDI-MS (data not shown), and frozen at −80 °C until use.

3.2.12.2 Cyclodeaminase assay

On the basis of its homology to TubZ and RapL, GeD was presumed to act as a lysine cyclodeaminase, converting L-lysine to L-pipecolic acid [82]. In order to evaluate whether GeD could catalyze the same reaction, we used a method previously established by A. Sandmann and Y. Chai method to detect L-pipecolic acid by HPLC/MS.

In brief, treatment of the pipecolic acid standard (a racemic mixture of L- and D-) with 9-fluorenylmethyl chloroformate (Fmoc-Cl) converted the amine to the corresponding Fmoc-carbamate. The Fmoc-pipecolic acid was then straightforwardly detected by HPLC/MS ($m/z =$ 571.2 [M-H]$^-$ (Figure 3.51 A)). Additionally pure L- and D-pipecolic acid standards were derivatized and analyzed to establish the retention times of L- and D-pipecolic acid, showing that D-pipecolic acids elutes first. As a positive control, we simultaneously assayed the enzyme TubZ from the tubulysin cluster, which had previously been shown to catalyze the formation of L-pipecolic acid from L-lysine, *in vitro* (Y. Chai, A. Sandmann, unpublished results). As expected, incubation of TubZ with L-lysine resulted in the formation of L-pipecolic acid (Figure 3.51 B)). To verify that exclusively the L-stereoisomer was formed, we

mixed the standard reaction mixture with the product of TubZ. A clear increase of the signal corresponding to the L-stereoisomer was observed (Figure 3.51 C)).
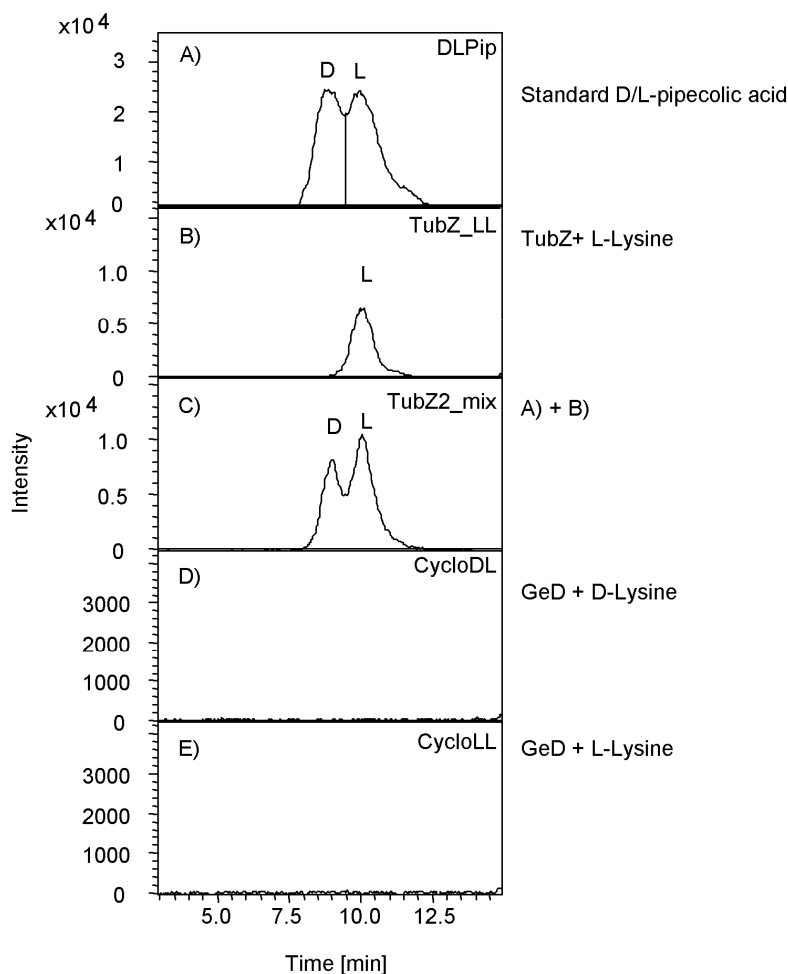

Figure 3.51: **Cyclodeaminase assays.**

A) Standard of D- and L-pipecolic acid. B) L-pipecolic acid formed by TubZ in the presence of L-lysine. C) Mixture of assay reactions A) and B), to verifiy the stereochemistry of the pipecolic acid generated in B). D) Incubation of GeD with D-lysine. E) Incubated of GeD with L-lysine.

Disappointingly, incubation of GeD (both fractions were used from expression in *E. coli* Rosetta and also from coexpression with the chaperones) with both L- and D-lysine under the same conditions failed to produce any pipecolic acid of either stereochemistry. As the assay worked efficiently with the control protein TubZ, GeD is apparently inactive under the assay conditions, or L-lysine (D-lysine) is not the correct substrate.

3.2.12.3 **Expression of A domains**

In order to directly investigate A domain substrate specificity, we aimed to express the first three A domains from cosmid BI11 as N-terminally His$_6$-tagged proteins from pET28b. DNA

fragments coding for the adenylation domains of *geE* (1 domain, $GeEA_1$) and *geG* (2 domains, $GeGA_2$ and $GeGA_3$) were amplified from cosmid BI11, and cloned into pET28b vectors using *Nde*I and *Bam*HI. Cosmid BA23 was not completely sequenced at this timepoint so that we couldn't amplify the other two A domains from *geJ* and *geM* encoding $GeJA_4$ and $GeMA_5$. The constructs were confirmed by sequencing and transformed into *E. coli Rosetta* BL21 (DE3) pLysS/RARE. Cultivation was carried out at 16 °C. Expression was induced with 0.2 mM IPTG, at $OD_{600} = 0.8–1$.
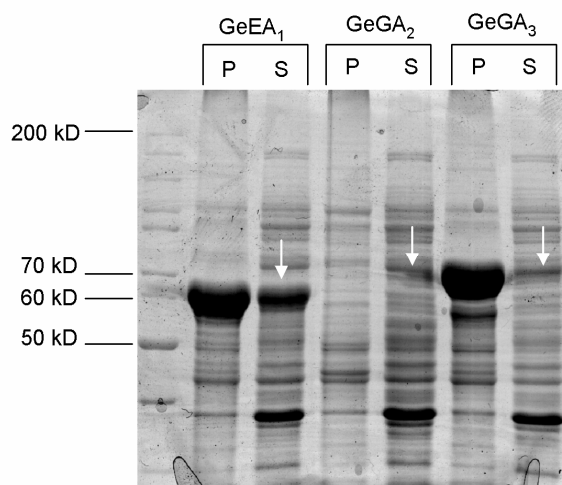
Figure 3.52: **Expression of A domains in *E. coli Rosetta* BL21 (DE3) pLysS/RARE at 16 °C.**
Soluble proteins are indicted by the white arrows. Key: P, pellet; S, supernatant

Protein $GeEA_1$ was found in the soluble fraction after French press lysis of *E. coli* cells, as judged by SDS-PAGE (Figure 3.52). However, only a small amount of both $GeGA_2$ and $GeGA_3$ were present as soluble proteins. Nonetheless, the observed sizes of the proteins were in good agreement with the calculated molecular masses of 56.6, 59.9 and 61.7 kDa, respectively. To try to improve the solubility of $GeGA_2$ and A_3, we co-expressed them with the chaperone set described in section 3.2.12.1. Coexpression with chaperone set 2 led to soluble protein in the case of $GeGA_2$, allowing the enzyme to be purified by Ni^{2+}-affinity chromatography using the Äkta prime system (Figure 3.53 and section 2.18.2.2). $GeGA_3$ could not be expressed as soluble protein until now. In the future further attempts to increase solubility need to be done.
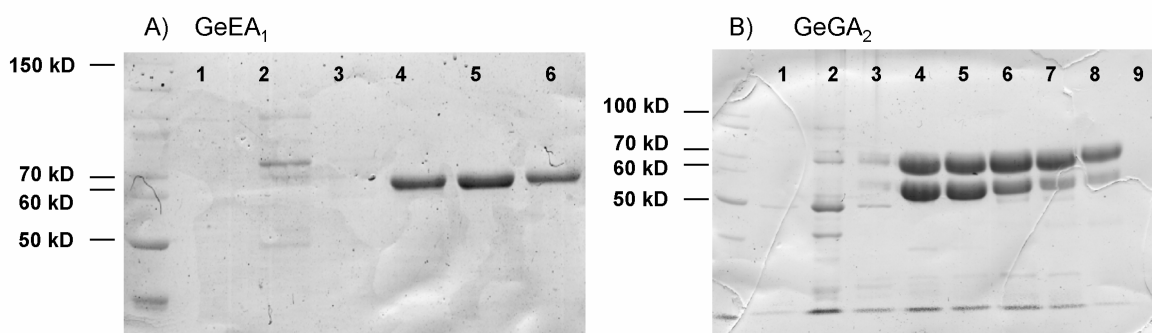
Figure 3.53: **SDS-PAGE of adenylation domains overexpressed in *E. coli*.**

A) Purification of A domain GeEA1. Lanes 1–6: various fractions after Äkta purification; Lanes 4–6 contain the purified A domain used for biochemical analysis. B) Purification of A domain GeGA2. Lane 1–9: various fractions after Äkta purification; Lanes 4–8 contain the purified A domain used for biochemical analysis.

All proteins were analyzed by MALDI-MS, which identified them as the desired proteins. Both the SDS-PAGE (Figure 3.53B) and the MALDI-MS analysis (data not shown) revealed, however, that a proportion of protein GeGA$_2$ was truncated from its C-terminal end. As this was not expected to interfere with the substrate specificity of the intact domain, the mixture was used in subsequent assays.

3.2.12.4 ATP-PP$_i$ exchange assay

The substrate specificity of the two purified adenylation domains was evaluated using the well-established ATP-PP$_i$ exchange assay [58;111]. This assay measures the amino acid-dependent exchange of radiolabel from ^{32}PP$_i$ into ATP. For this, each protein was incubated with a panel of different amino acids, including the anticipated substrate of each domain. As a control, each protein was incubated in the absence of added amino acid.
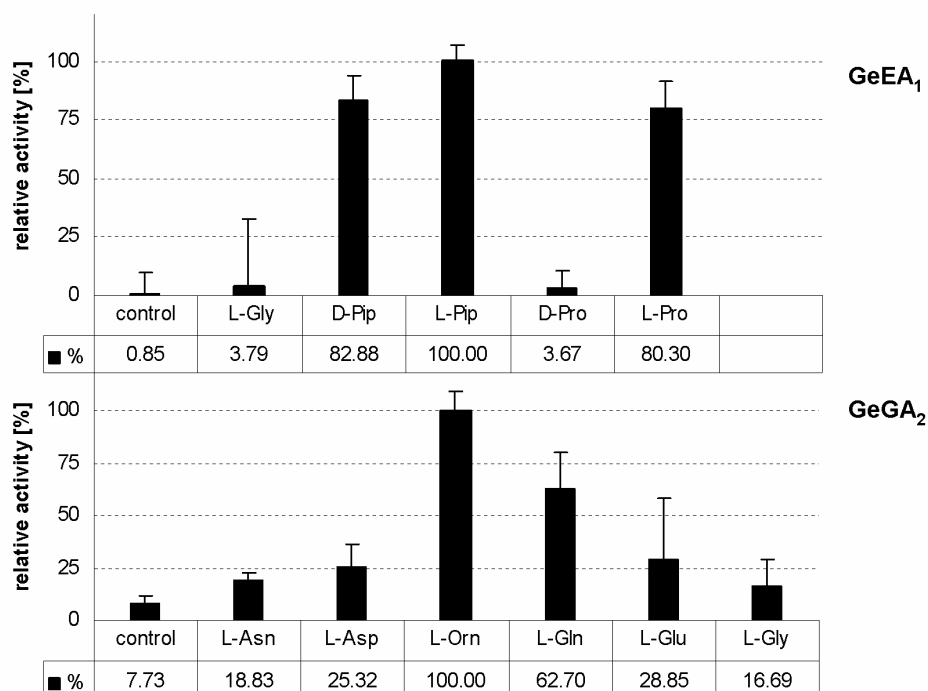
Figure 3.54: Relative substrate specificities of the adenylation domains GeEA$_1$, GeGA$_2$.

The substrate specificity of the A domains was investigated using the ATP-PP$_i$ assay. The highest activities were set to 100%.

GeEA$_1$ activated L-Pip (100%), D-Pip (82.88%) and L-Pro (80.3%). The background and controls were between 0.85% and 3.79%, confirming that the measured activity reflected the true substrate preference of the A domain (Figure 3.54). GeGA$_2$ activated L-Orn (100%) as well as L-Gln (62.7%). L-Glu and L-Asp were also activated to a minor extent. So the preferred activation of ornithine and glutamine was clear. Taken together, our results prove that the two internal enzyme fragments investigated exhibit the enzymatic properties of adenylation domains. The determination of the A domains specificity enabled us to place the first two A domains in the correct order within the synthetase. As predicted previously, the A domain GeEA$_1$ activates pipecolic acid, and thus forms a part of the loading module (module 1). It is followed by module 2, which harbors an A domain (GeGA$_2$) specific for orithine/glutamine. It remains to test the other three A domains GeGA$_3$, GeJA$_4$ and GeMA$_5$ to determine their function.

4 Discussion

4.1 General scope of this work

This thesis deals with the biosynthesis and heterologous production of complex secondary metabolites from streptomycetes. A method for the heterologous expression of these compounds in streptomycetes was developed, using the phenalinolcatone biosynthetic pathway as a model system. The established expression system provided the basis for the modification of the pathway, leading to the production of a new derivative.

Identification and biochemical characterization of the GE81112 biosynthetic gene cluster provided information about the biochemistry of this unique family of secondary metabolites and of nonlinear NRPS megasynthetases more generally. In the course of these studies, genetic manipulation methods for a new streptomycetes strain were developed, leading to the formation of a knockout mutant.

4.2 Phenalinolactone - Heterologous expression of a complex secondary metabolite

4.2.1 Heterologous expression of natural product pathways

Heterologous expression of complex natural product biosynthetic gene clusters in amenable host strains has become an important tool in natural products research and drug discovery. This strategy provides an alternative to (over)producing structurally complex substances that would be difficult or impossible to access by other means, and can enable the generation of novel analogs by combinatorial biosynthesis approaches. However, successful heterologous production of natural products faces significant challenges, deriving both from complicated nature of the secondary metabolites, and the complexity of the biosynthetic enzymes. In addition to laboratory convenience, excellent growth characteristics, and an array of available genetic tools, the heterologous host must be able to express relatively large proteins (e.g. PKS and NRPS proteins) [30;112], post-translationally modify them to their active forms, and produce adequate amounts of activated intracellular building blocks. As the codon usage is usually [113] critical for efficient expression, it is generally important that the GC content of the foreign genes and the host genome are similar. Furthermore, the functionality of regulatory elements and promoter structures as well as mRNA stability need to be considered when choosing the host organism. To date, the most successful heterologous expression

experiments have relied on hosts which are closely-related to the natural producer strain. For example, many clusters derived from *Streptomyces* clusters have been reconstituted in related actinomycete strains [29;33;114-116]. As most of these gene clusters were relatively small (<30 kbp), it was often possible to retrieve the entire gene set from a single cosmid within a genomic library of the natural producer strain [30]. In cases where the libraries were prepared in *E. coli-Streptomyces* shuttle vectors (e.g. the medermycin and griseorhodin A biosynthetic gene clusters [31;117]), transfer of the pathways into heterologous strains was straightforward.

However, many natural product biosynthetic gene clusters are larger than the average insert size of common cosmid vectors. One strategy for overcoming this size limitation is to use BAC shuttle vectors for library construction, as they can accommodate inserts in excess of 100 kbp. This approach enabled the successful expression of the 128 kbp daptomycin biosynthetic gene cluster from *Streptomyces roseosporus* in a related *Streptomyces* strain [118]. Another option is to clone subsets of the biosynthetic gene cluster into compatible expression plasmids, followed by their stepwise introduction and co-expression in suitable host strains [34;119]. As the cloning procedure allows the introduction of artificial promoter regions, heterologous production can be performed even in non-related host organisms. Using this strategy, natural products from *Streptomyces* were produced in *E. coli,* and myxobacterial compounds were obtained from *Streptomyces* host strains [120-122].

An alternative to these classical, time-consuming cloning and mutagenesis approaches is the reassembly of large natural product pathways on a single transferable vector system, using Red/ET technology. This recombination approach was developed by the Stewart group in 1998 [37], and is ideal for manipulating large pieces of DNA as it does not depend on the use of restriction enzymes. Phage-derived protein pairs (RecE/RecT from the Rac prophage or Redα/Redβ from the lambda phage) are employed to precisely alter target DNA molecules by homologous recombination within *E. coli* strains [36]. The first notable proof-of-principle for the cloning of a large PKS/NRPS hybrid pathway using Red/ET recombineering was provided by Wenzel *et al.* [35]. The 29.6 kb pathway directing the biosynthesis of myxochromides S was rebuilt on one molecule and the DNA elements for transfer, stable maintenance, and inducible expression were introduced, enabling the production of the natural product in *Pseudomonas putida.*

In *Streptomyces* strains, Red/ET cloning has been used to modify the backbone of cosmids containing complete biosynthetic pathways, enabling the heterologous expression of the aminocoumarin antibiotics novobiocin and clorobiocin [123]. In these examples, all of the genes were oriented in the same direction, consistent with the presence of a single transcriptional unit. In comparison to the aminocoumarin pathways, the phenalinolactone gene cluster of *Streptomyces* Tü6071 is more complex, as it consists of 35 orfs which appear to be organized into 11 individual transcriptional units [66]. This highly divergent architecture would seem to present a significant challenge to the regulatory apparatus of heterologous host strains.

Thus, in attempting to achieve heterologous expression of the phenalinolactone pathway, we developed a Red/ET-based methodology to reconstitute the entire gene cluster on one construct to ensure that all transcriptional units were intact. Whereas similar approaches had already been used to engineer myxobacterial systems [35;42], the 'stitching together' of complex actinomycete pathways via Red/ET recombineering had not been previously described when the Ph.D. thesis was started which intended us to start this project. Shortly after our work was published another example for the heterologous expression of a gene cluster using Red/ET was published by the Gust group [124]. Using this method they were able to stitch and express the coumermycin A$_1$ biosynthetic gene cluster in *S. coelicolor* M512. Using Red/ET recombineering the entire gene cluster of PL was rebuilt from overlapping cosmids based on the integrative *E. coli-Streptomyces* shuttle vector pOJ436, and transformed into different heterologous hosts.

4.2.2 Heterologous expression of the PL pathway in pseudomonads

Pseudomonads has been shown to be suitable heterologous hosts for the production of natural products [125]. These microbes possess a high GC content, have a relatively short doubling time, and genetic tools are well established. Therefore we aimed to heterologously express the PL gene cluster in *Pseudomonas putida*. As we had integrated an *oriT-tet-trpE* cassette into the expression construct, transformation and integration of the vector into the *Pseudomonas* genome was possible. Integration into the genome was proven by PCR and the obtained mutants were cultivated, extracted and the extracts analyzed by HPLC-MS. Unfortunately we did not observe any heterologous production of PLs in *P. putida*.

One of the major issues for the heterologous expression of biosynthetic pathways in unrelated bacteria is the supply of adequate precursors. In the case of the polyketide myxothiazol, heterologous expression in *P. putida* initially failed. However, integration into *P. putida* of an operon responsible for production of an essential building block, methylmalonyl-CoA, resulted in myxothiazol production [43]. To investigate whether the absence of a precursor could explain the failure to obtain the PLs from *P. putida*, we considered the biosynthetic model for phenalinoalctone biosynthesis published by Dürr *et al.* [66;80] (Figure 4.1). The pathway can be divided into four major parts: (1) formation of the terpene backbone and its extension with phosphoenolpyruvate; (2) modification and rearrangement of the phosphoenolpyruvate-derived side chain toward the γ-butyrolactone moiety; (3) biosynthesis of L-amicetose and pyrrole-2-carboxylic acid; and (4) group transfer reactions and late-stage tailoring. The precursors for each portion of the biosynthesis are derived from (1) the nonmevalonate pathway (terpene biosynthesis); (2) pyruvate (γ-butyrolactone biosynthesis); (3) glucose (L-amicetose biosynthesis); and (4) L-proline (pyrrole-2-carboxylic acid biosynthesis).
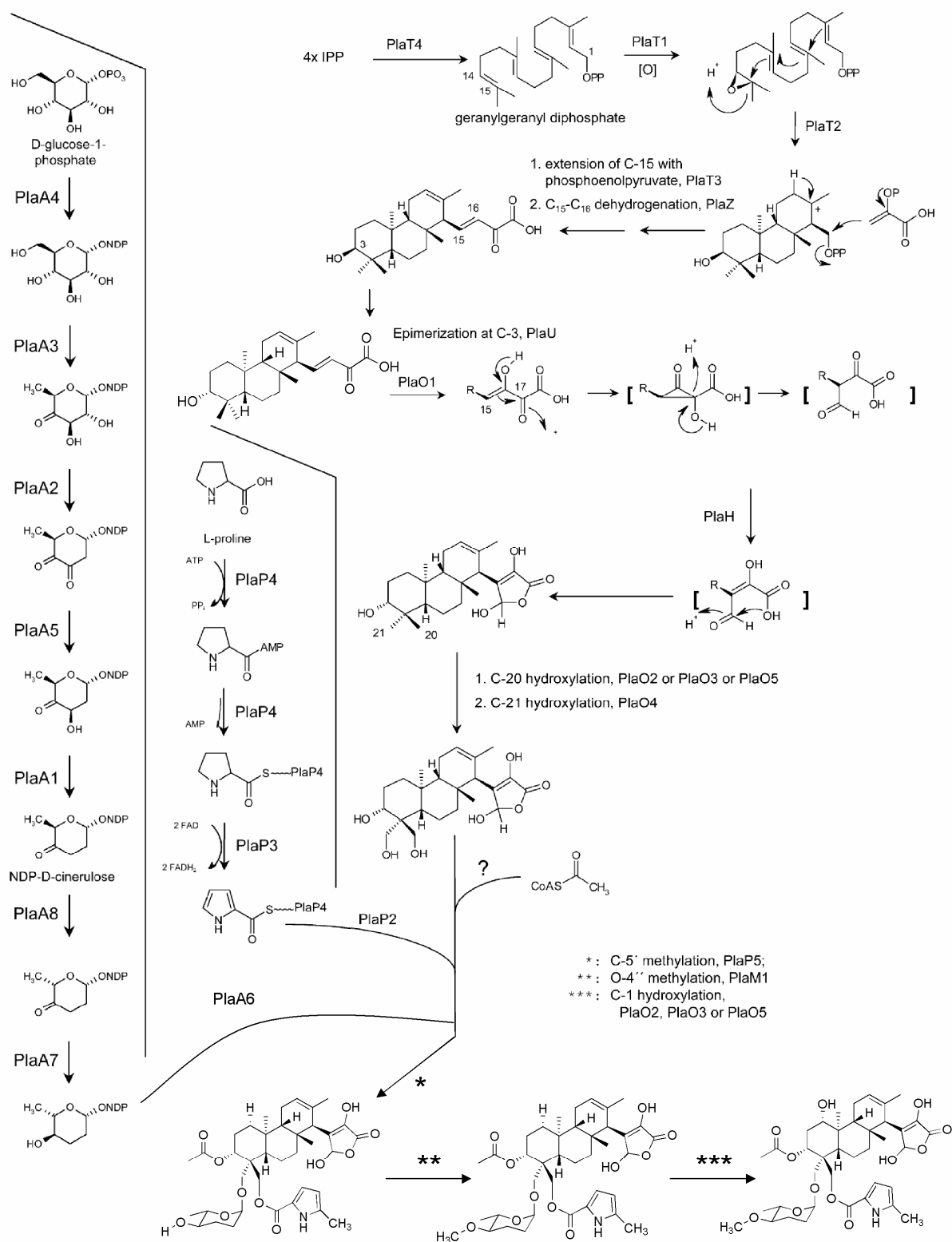
Figure 4.1: **Proposed biosynthetic pathway for phenalinolactone formation.**

Figure adapted by Dürr *et al.* [66].

As the nonmevalonate pathway is present in *P. putida* [126] and the respective genes are also present in the cloned gene cluster, we eliminated this pathway from consideration. The remaining precursors such as pyruvate, glucose and L-proline are derived from the primary metabolism, and were therefore also expected to be present in *P. putida*. Thus, we assumed that *P. putida* is able to produce all precursors which are needed for PL biosynthesis.

As pseudomonads posses a high GC content similar to that of streptomycetes, we also assumed that codon usage was not a problem. This focused attention on the native *Streptomyces* promoters, which may not be recognized in *Pseudomonas putida*. To date, heterologous expression of natural product biosynthetic pathways in pseudomonads has only been succesful when the native promoters were exchanged against the inducible P_m promoter [35;43;125]. For this reason, we also aimed to exchange several promoters in the PL biosynthetic gene cluster to achieve heterologous expression in *P. putida*. As the PL biosynthetic gene cluster contains 35 orfs and 11 transcriptional units, we expected that at least one promoter would be present in each operon, for a total of 11 promoters. As a starting point, we decided to exchange the promoter in front of the operon containing the genes responsible for assembling the core structure of the PLs. The proteins PlaT1–PlaT6 (Figure 4.2) comprise the enzymatic machinery to biosynthesize the phenalionalctone terpenoid scaffold. The isoprene building blocks for phenalinolactone biosynthesis are derived from the nonmevalonate pathway which is also present in *P. putida*. So it can be assumed that the isopentenyl pyrophosphate (IPP) should be produced in primary metabolism of *P. putida*. The genes *plaT5* and *plaT6* in the PL cluster comprise additional copies of the *dxp (*1-deoxy-D-xylulose 5-phosphate) and *hmbpp (*1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate) synthase genes from the nonmevalonate pathway. It is speculated that PlaT5 and PlaT6 probably enhance the IPP production [66]. PlaT4 catalyzes the head to tail condensation of four isoprene monomers toward geranylgeranyl diphosphate (GGDP), the phenalinolactones direct terpene precursor. PlaT1 is suggested to catalyze an epoxidation of the C-14/-15 double bond of GGDP prior to its cyclization. PlaT2 is responsibe for the cyclization of GGDP to PL precursor 1. PlaT3 might be considered for the phosphoenolpyruvate transfer to the terpene backbone. The next step would be the dehydrogenation of the newly established C-C linkage, giving the C-15/-16 double bond. PlaZ would be a candidate for this dehydrogenation reaction (Figure 4.2).
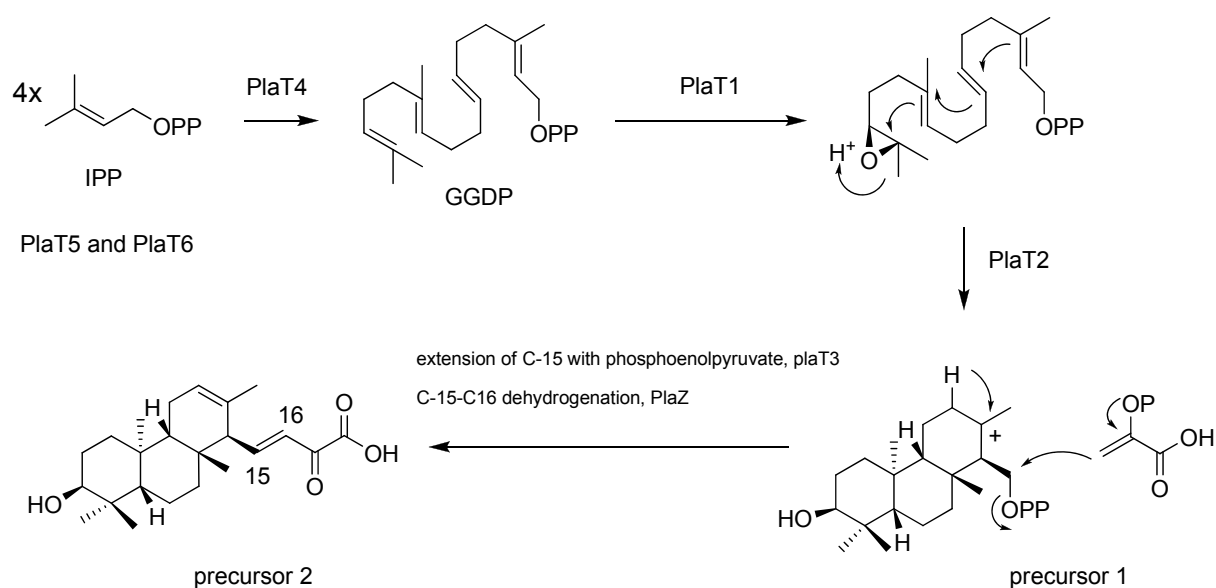
Figure 4.2:   Proposed biosynthesis of possible PL precursors. IPP: isopentenyl pyrophosphate, GGDP: geranylgeranyl diphosphate.

Unfortunately the corresponding genes *plaT1-3* and *plaT4-6* are located in two different operons, making it impossible to attempt to upregulate transcription with a single promoter switch. We started by targeting the operon containing the first three biosynthetic genes, *plaT4-6*. The native promoter was exchanged against a cassette containing the P_m promoter, the *xylS* gene and a resistance gene (chloramphenicol) using Red/ET cloning, and mutants were verified by PCR. We expected at least the direct precursor GGDP to be produced as *plaT4-plaT6* are present in the operon where the promoter was exchanged. But analysis by HPLC-MS analysis failed to reveal the presence of GGDP, PL precursors or PL in the mutants. This result demonstrated that exchange of this single promoter was insufficient to achieve production of any precursors in *P. putida*.

In principle, a follow-up strategy would be to attempt to exchange all of the native promoters against the same P_m promoter cassette or other *Pseudomonas* promoters. However, this approach is likely to cause difficulties, as the introduction of promoters by Red/ET recombination is based on homologous recombination; the presence of the same promoter sequence at multiple locations within the cluster would probably result in undesired recombination events. Alternatively, it might be possible to use a set of different promoters. However, only a few inducible promoters are currently known for pseudomonads. Taking these considerations into account, we concluded that it would not be feasible to exchange all

of the promoters within the PL cluster. We therefore pursued heterologous expression in an alternative host.

4.2.3 Heterologous expression of the PL pathway in streptomycetes

The Red/ET engineering enabled the efficient transfer and integration of the whole PL pathway into several *Streptomyces* strains, as the construct contained a vector backbone with an *attP* site and the integrase gene. The 62 kbp construct was successfully conjugated with high efficiency into the host strains *S. lividans* TK24 and *S. coelicolor* A3(2), and integration into the chromosomes confirmed by PCR. In contrast to the *Pseudomonas* mutants, the resulting mutant strains of *S. lividans* TK24::CPhl8 and *S. coelicolor* A3(2)::CPhl8 were both shown to produce phenalinolactones A and D. This result strongly suggests that the inserted cluster contains all genes necessary for the biosynthesis of phenalinolactone, although participation of enzymes native to the heterologous hosts cannot be ruled out. Evidently, all of the regulatory elements required for transcription of the *pla* genes were recognized in the heterologous hosts, albeit at reduced efficiency relative to the native host strain (Table 3.2).

One potential complication of using *S. lividans* and *S. coelicolor* as heterologous hosts is their endogenous production of the antibiotics prodiginin (Red), actinorhodin (Act) and calcium-dependent antibiotic (CDA), which may hamper the simultaneous biosynthesis (and detection) of a desired new compound. This problem has been circumvented in several cases by using a mutant strain of *S. coelicolor* M512, in which the Act and Red gene clusters have been deleted [123;127]. Therefore, *S. coelicolor* M512 was also used here as an additional heterologous host for phenalinolactone biosynthesis. Indeed, after successful expression of the cluster in the strain, production of phenalinolactone was 5- to 50-fold higher at 28 °C relative to that in *S. lividans* TK24 or *S. coelicolor* A3(2) (Table 3.2). Interestingly, cultivation at 37 °C in *S. lividans* yielded a further enhancement in productivity (up to 100-fold compared to cultivation at 28 °C) although the optimal growth temperature for *S. lividans* is 30 °C. An increase in phenalinolactone biosynthesis was not observed for the *S. coelicolor* strains at 37 °C. In fact, incubation at 37 °C completely abolished PL biosynthesis, consistent with the fact that production by the native host *Streptomyces* Tü6071 is also significantly lower at this temperature (200 μg/L). Thus, under optimized fermentation conditions, it is clear that *S. lividans* is a superior heterologous host for phenalinolactone biosynthesis than the *S.*

coelicolor strains, although the production is still 50-fold lower than from the natural producer.

An attempt was also made to increase phenalinolactone production by feeding of the putative precursor mevalonolactone, but no increase in the yield of PL was observed. Analysis of the *pla* cluster shows that it incorporates genes for producing isoprenoid building blocks via the mevalonate-independent pathway, and indeed this biosynthetic origin has previously been demonstrated [66]. In addition, genes belonging to the mevalonate pathway have not yet been identified in either *S. lividans* and *S. coelicolor* [128;129]. Thus, it seems that the heterologous host strains used here lack the enzymes to metabolize mevalonate, which accounts for our failure to increase phenalinolactone yields by supplementation.

In a further attempt to increase productivity, we exchanged two of the native promoters against the strong constitutive *ermE* promoter using Red/ET recombineering again. The advantage in this strategy was the *ermE* promoters sequence could be introduced with the primers, so that two promoters could be exchanged in a single step. The two promoters were placed in front of the two operons harboring the first biosynthetic genes (*plaT1-3* and *plaT4-6*). It has previously been demonstrated that introduction of the *ermE* promoter can increase the yields of heterologously-expressed natural products, or even lead to initiation of production. For example, cloning of eight validamycin biosynthetic genes under the control of the *ermE* promoter led to heterologous production of this metabolite in *S. lividans* [130]. Another example was the heterologous expression of an intermediate of the kendomycin starter unit. The heterologous expression of a gene set of 7 genes revelaed that 3,5-dihydroxybenzoic acid is an intermediate in the kendomycin starter unit biosynthesis [131]. Successful heterologous production was only achieved in *S. coelicolor* but not in *S. lividans*. Cloning of the *ermE* promoter in front of the gene set initiated the heterologous expression in *S. lividans*. Production under the control of the *ermE* promoter in *S. lividans* was even much more higher (150mg/ml) compared to production in the natural producer or the heterologous production in *S. coelicolor* with and without the *ermE* promoter (S. C. Wenzel unpublished data). Unfortunately, exchange of the native promoters against the *ermE* promoter in the *pla* biosynthetic gene cluster did not yield any increase in productivity. As we introduced two *ermE* promoters in front of both operons containing the first biosynthetic genes we expected at least a production of GGDP or precursor 1 (Figure 4.2). As PlaZ which is probably needed for the formation of precursor 2 was not part of both operons precursor 2 was not directly

expected to be produced. But no GGDP or any other precursors could be detected; instead, PL production was completely abolished. These results led us to conclude that promoter exchange interrupted transcription or led to problems with regulation or both, probably as a result of the complex operon structure of the *pla* cluster. Again, a solution might be to exchange of all the promoters within the cluster, but this approach would raise the same issues as in the pseudomonads. Thus, overall we achieved the greatest success in improving PL production by changing the growth temperature of *S. lividans* showing that standard microbiological optimization methods can still be effective.

4.2.4 Genetic modification

The expression system described here further enabled the generation of an unglycosylated phenalinolactone derivative. For this, Red/ET technology was combined with the FRT/FLP recombinase system to construct a markerless deletion of the glycosyl transferase gene *plaA6*. Heterologous expression of the modified construct resulted in the unglycosylated derivative phenalinolactone E (PL E), demonstrating that the glycosyl transferase PlaA6 catalyzes the attachment of L-amicetose to the hydroxylated terpenoid backbone. Surprisingly, biosynthesis of PL E was only observed in *S. coelicolor* M512, although yields of the parent glycosylated compound from *S. lividans* TK24 were much higher. In this case, the utility of *S. coelicolor* mutant strain M512 as suitable host may stem from its overall lower background of secondary metabolism. The use of the FRT/FLP recombinase system in combination with the Red/ET technology is a powerful tool to create a markerless gene deletion in a single expression construct. This modification is conveniently achieved by introducing the FRT recognition sites into the PCR primers. Recombination with FLP recombinase results in a markerless mutation, in which only a 34 bp scar sequence is left in place of the deleted gene of interest. This scar sequence does not present a problem for further recombination procedures. To construct two mutations simultanesouly, the Cre/loxP system can be employed in parallel.

A similar approach has been used by Heide *et al.* [132]. New aminocoumarin antibiotic derivatives were generated by targeted gene replacement using Red/ET recombineering. However, instead of introducing FRT or loxP recognition sites with the primers, they introduced recognition sites for the restriction enzymes *Spe*I and *Xba*I which are rare cutters in the *Streptomyces* DNA. Digestion of the cosmids by *Spe*I and *Xba*I resulted in compatible overhangs, and religation then resulted in a cosmid in which only an 18 bp scar sequence was

left in place of the deleted gene. One drawback of this approach in comparison to the FRT/FLP or Cre/lox system, however, is that a ligation step is always required. Additionally, no other *Spe*I or *Xba*I sites can be present in the target sequence, whereas the recombinase system is not dependent on restriction sites but on the FRT/loxP recombination site.

To obtain additional PL derivatives, we generated two constructs containing markerless deletions of the genes *plaP2* and *plaP5*. By removal of the acyltransferase- encoding gene *plaP2*, we aimed to generate a PL derivative lacking the 5-methylpyrrole-2-carboxylic acid moiety (Figure 4.3). The gene *plaP5* is proposed to be responsible for the methylation of the pyrrole carboxylic acid [66], and thus deletion of this gene was expected to give an unmethylated PL derivative (Figure 4.3). In this case, we exploited the Cre/lox system instead of the FLP/FRT system. Using both systems in combination, we aimed to obtain a small set of different PL derivatives (Figure 4.3).
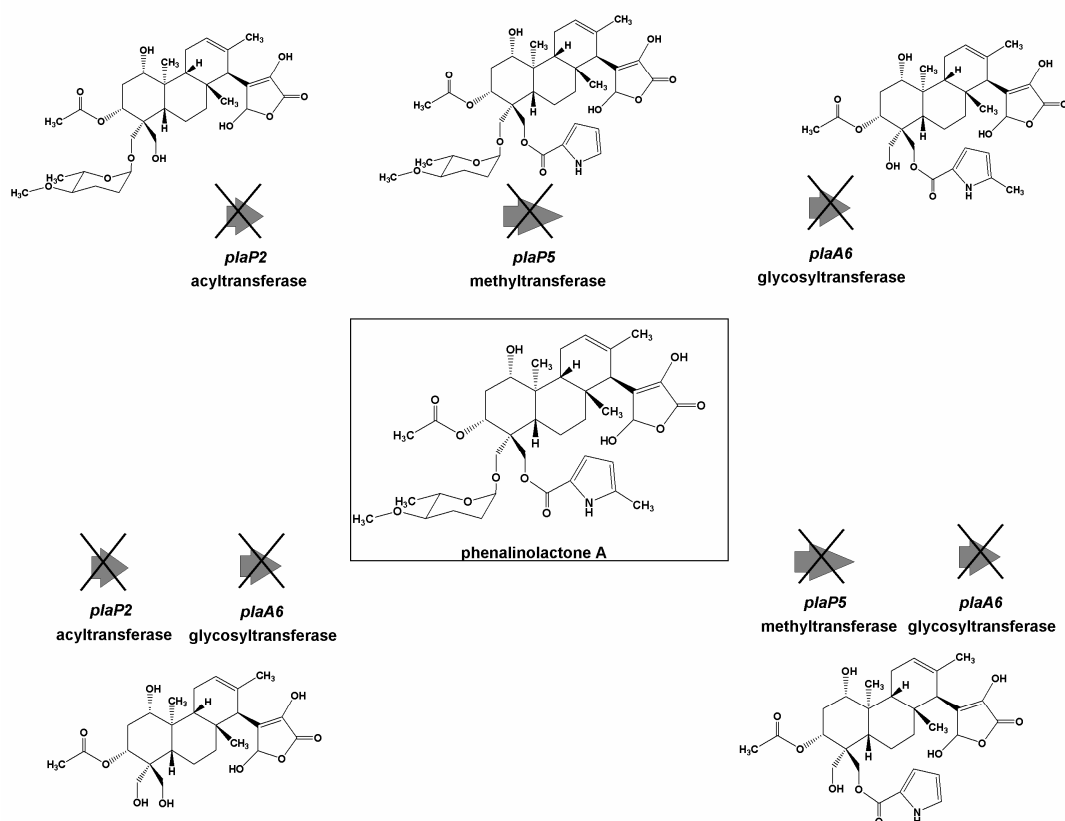
Figure 4.3: PL derivatives expected after deletion of different biosynthetic genes, in heterologous streptomycetes hosts.

Unfortunately, no variant PLs could be detected in extracts of mutants carrying the deletions in *plaP2* and *plaP5*, and instead, PL production in these strains was completely abolished. The failure to detect new derivatives might be explained by very low production levels: as PL

E is already produced in miniscule amounts, it seems reasonable that production of derivatives could fall under the detection limit. However, we cannot rule out the possibility that deletion of the genes interrupted transcription, or that regulatory problems occurred.

4.2.5 Future aspects

Although several *Streptomyces* gene clusters have been relocated to heterologous hosts using Red/ET cloning, transcription in each case was under the control of a single promoter [123]. Thus, the *pla* cluster represents the most complicated gene set yet to be heterologously expressed in any *Streptomyces sp*. In future, efforts to understand the regulatory factors governing phenalinolactone biosynthesis should aid in efforts to optimize heterologous production, for example by overexpressing positive regulators. Also other heterologous actinomycetes hosts could be tried to optimize heterologous production. In addition, the results described here set the stage for further engineering the biosynthetic machinery in *E. coli* towards the generation of additional novel PL analogues.

4.3 GE81112 family compounds - Novel tetrapeptide inhibitors of bacterial protein synthesis

Identification of the biosynthetic gene cluster

The emergence of multi-drug resistant microbial pathogens is driving the search for novel and more effective antibiotics, especially those with new mechanisms of action. The majority of known antibiotics inhibit the fundamental process of protein synthesis. Prominent examples of antibiotics which target this machinery are the macrolides erythromycin and azithromycin, the aminoglycosides kanamycin and tobramycin, and the tetracyclines [133]. Elongation activities such as aminoacyl-tRNA binding and decoding in the ribosomal A-site, transpepdidation, and translocation are frequently found to be affected by antibiotics, albeit by different mechanisms. However, other functions such as translation initiation and termination, are only rarely targeted [133]. Thus, the tetratpeptide GE81112 compounds represent a unique class of novel antibiotics, as they selectively and effectively inhibit the formation of the prokaryotic 30S initiation complex [67]. To obtain a better understanding of the biosynthesis of these unique metabolites, we set out to identify and clone the corresponding biosynthetic gene cluster. We hoped that insights into the biosynthesis would enable the directed engineering of novel GE81112 derivatives.

4.3.1 Attempts to identify the GE81112 biosynthetic gene cluster

4.3.1.1 Generation of a cosmid library and hybridization strategy

At the beginning of the work, we obtained the GE81112 producer strain *S*. 14386 from the company KtedoGen. In order to enable the identification and cloning of the biosynthetic gene cluster, we started by optimizing production conditions for the GE81112 compounds. After establishing cultivation and detection of the GE81112 metabolites, we aimed to locate the corresponding gene cluster by generating a cosmid library, and screening it with suitable probes. Before rapid genome sequencing methods became available, numerous NRPS and/or PKS biosynthetic gene clusters were sucessfully identified using this approach [134-137].

We therefore generated a 2304 clone cosmid library from the genomic DNA of the GE81112 producer strain containing. A suitable probe was obtained by applying a "retrobiosynthetic strategy", which allowed us to predict some aspects of the composition of the GE81112 cluster from the metabolite structures. On this basis [67], we predicted that the biosythesis

would be carried out by a non-ribosomal peptide synthetase, and so designed probes to identify NRPS genes. Furthermore, we were able to predict that the starter unit would probably derive from pipecolic acid, which is known to be formed from lysine via the action of a lysine cyclodeaminase [82]. As cyclodeaminase genes are relatively rare in the genome [86] this gene was used to design a second probe to identify the GE81112 cluster.

4.3.1.2 Identification of an unkown NRPS biosynthetic gene cluster

Screening of the cosmid library with both probes revealed two cosmids apparently harboring overlapping regions of a gene cluster. Furthermore, sequence analysis revealed that both cosmids harbored NRPS biosynthetic genes. As the modules in both clusters contained comparable domains in the order C-A-PCP [56] (Figure 4.4), we expected the two cosmids to contain the same biosynthetic gene cluster. Surprisingly, however, a more detailed sequence analysis revealed that the gene clusters, although highly similar, were not identical. This finding might indicate that they have arisen from a recent gene duplication event, followed by sequence divergence.
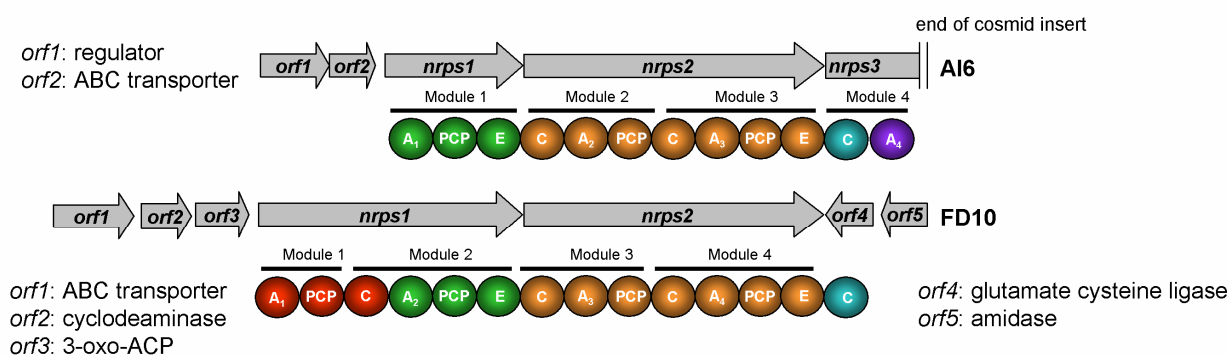
Figure 4.4: **Module and domain arrangement from NRPS biosynthetic gene cluster on cosmid AI6 and FD10.** Identical domains from both clusters are highlighted in the same colors.

The prescence of two highly-similar clusters was unexpected, but might have accounted in principle for the biosynthesis of the GE81112 family of tetrapeptides (e.g. each of the clusters might have given rise to a subset of family members). However, no examples of such 'in parallel' biosynthesis of similar metabolites within the same strain have been described to date, although closely-similar derivatives of a single natural product have been isolated from different strains (e.g. myxochromide A and S) [138].

As it was not clear at this stage if the identified clusters were responsible for the biosynthesis of the GE81112 compounds, we analyzed the NRPS domains from each module. For the

incorporation of four amino acids we expected four modules containing the standard C-A-PCP organization and additionally a cyclodeaminase for the generation of the starter unit pipecolic acid. The four module arrangement was present on cosmid FD10 as we predicted, as well as a cyclodeaminase gene. These results supported the hypothesis that the identified assembly line on cosmid FD10 is able to biosynthesize a tetrapeptide or tetrapeptides, as expected in the biosynthesis of the GE81112 family. The gene cluster on cosmid AI6 was not completely present on the cosmid. Three complete modules could be identified on this cosmid indicating that the cluster should at least be able to synthezise a tripeptide. Module four contains a C and an A domain but lacks a PCP where the cluster is interrupted. This could imply that there would still be a PCP in the the genome sequence so that the cluster would also be able to synthetize a tetrapeptide or even a longer peptide, as we don't know what sequence follows. When considering the similarity with FD10 one could also conclude that the cluster on AI6 ends with the C domain on gene *nrps_3* (same domain arrangement than in cosmid FD10) and that the additional A domain would belong to another cluster. To answer these questions the overlapping cosmid providing the whole pathway for the NRPS on AI6 would need to be identified which failed until now.

To test if all domains are likely to be functional, the conserved regions of all C, A, PCP and E domains [91] were analysed and found to be present. Using the specificity-conferring code of adenylation domains, we were able to predict the substrate specificity of some of the A domains (Table 3.8). As both clusters had an overall high content of sequence similarity, it was not surprising to find three A domains from each cluster which showed the same apparent substrate specificity. The identical A domains are found to be in the same order assuming that module 1 has been deleted/added to the FD10 cluster (Figure 4.4: identical domains are highlighted in the same colors). The A domains $FD10\_A_2$ ($AI6\_A_1$) showed substrate specificity for the incorporation of phenylalanine/tryptophan. A domain $FD10\_A_3$ and $AI6\_A_2$ were predicted to show substrate specificity for the activation of threonine. No confident predictions could be made for the remaining A domains, possibily indicating that they active uncommon amino acids.

Overall, the two predicted A domain specificities for both clusters did not fit well with the biosynthetic hypothesis for the GE81112 compounds, as it was expected from the "retrobiosynthetic analysis" that pipecolic acid and two histidines would be incorporated.

In addition to the NRPS genes, we found the expected cyclodeaminase gene, but it was only present on cosmid FD10. Furthermore, sequence analysis revealed that more than half of the protein was missing relative to previously-described cyclodeaminases [82;83;85]. We assumed, therefore, that the truncated protein was no longer functional. This result provided an additional strong hint that the identified gene cluster was probably not responsible for GE81112 biosynthesis. As these results were based solely on *in silico* predictions, however, we aimed to obtain experimental validation for the identity of the gene clusters.

4.3.1.3 **Heterologous expression**

The most straightforward strategy to verify a gene cluster is to knockout one of the NRPS genes which are normally clustered in the genome and essential for the biosynthesis. We tried to do this knockout but, to date, it has not been succesfull. We therefore pursued an alternative strategy to determine the identities of the identified gene clusters, using heterologous expression [30;118;139;140]. Both cosmids were transferred into heterologous hosts *S. lividans* TK24 and *S. coelicolor* M512 and the extracts were checked for GE81112 production. Unfortunately, the desired compounds could not be detected in the heterologous hosts. Instead three new compounds were detected in the heterologous host carrying cosmid FD10, which were not present in the wild type. These substances exhibited exactly the same fragmentation pattern as three compounds that were found during our initial analysis of the natural producer strain *S.* 14386, and which attracted our attention because one of them had the same mass (659) as GE81112 factor B. In addition, their fragmentation patterns suggested that they are chlorinated. Thus, the compounds might be structurally-related to the GE81112 metabolites, but were definately not the desired GE81112 compounds.

Taken together, these results indicated that the biosynthetic gene cluster present on cosmid FD10 is not responsible for biosynthesis of GE81112. The new compounds were not expressed from cosmid AI6, although it contained a highly similar gene cluster. One possible explanation for this result was that the cluster was interrupted in the *nrps_3* gene and was not fully present on the cosmid. When we expect the cluster on AI6 to biosynthesize a tripeptide (as discussed earlier) then this tripeptide should at least incorporate the three identical amino acids than cosmid FD10, which could lead to a similar fragmentation pattern of the compound. But no such tripeptides could be identified. This result rather indicates that the cluster was not complete and was therefore not heterologously expressed as essential genes

are missing. We also concluded that the cluster on AI6 is not responsible for GE81112 biosynthesis, due to its high overall similarity to the gene cluster on FD10 which has been proofen to biosynthesize the new compounds. However, to provide further support for our conclusions, we aimed to purify the novel compounds from the wild type and elucidate their structures. All three compounds could be purified in reasonable amounts for NMR elucidation using standard chromatographic techniques. We focused our initial structure elucidation efforts on compound 631, as it was obtained in the highest yields following purification. The NMR data obtained to date suggest that the structure of compound 631 contains four amino acids, consistent with the four modules found within the NRPS biosynthetic gene cluster on cosmid FD10. However, *in silico* analysis of the A domain specificity predicted that FD10_A_2 would incorporate phenylalanine and FD10_A_3 threonine, while it appears that the compound includes a starter most likely derived from proline and a second amino acid derived from alanine. Analysis of the A domains FD10_A_1 and FD10_A_4 did not yield a clear prediction for their substrate specificity, consistent with the idea that they are uncommon A domains which may activate two unusual, modified ornithines.

The gene cluster found on cosmid FD10 exhibits few additional features that hint as to how to put together the structural elements identified to date by NMR. One element of note is that the cluster ends with a C domain and not with a TE, as is typical for bacterial NRPS. This observation suggests that the C domain might catalyze a head-to-tail condensation to yield a cyclic product, as found for the cyclosporine synthetase [141]. However, evaluating this hypothesis awaits the full structural elucidation of the compounds. Additionally there was also a PKS part found to be present on FD10 but not on AI6. But this gene was also not fully present on the cosmid. So it cannot be excluded that the PKS is somehow involved in the biosynthesis of the new compounds, but this seems rather unlikely as it is not expected to be functional when interrupted. Final evidence would be given by a knockout in the natural producer or even on cosmid FD10 which was heterologously expressed. As the new compounds contained unusual amino acids harboring modifications it is also not clear how they can be biosynthesized from the relatively simple cluster on the cosmid. One explanation would be that the heterologous host provides some of the enzymes which are needed for this most likely complicated biosynthesis. One fact which is supporting this hypothesis is that the new compounds were only expressed in *S. lividans* and not in *S. coelicolor*, indicating that *S. lividans* has probably some advantages (e.g. special enzymes) for the biosynthesis of the new compounds.

4.3.2 New strategy-new probe

Identification of the GE81112 biosynthetic gene cluster

As we had proven by heterologous expression that the gene cluster on cosmid FD10 (and probably also on AI6) were not responsible for the biosynthesis of the GE81112 compounds but for a new class of peptides, we developed an alternative strategy to identify the GE81112 gene cluster. The basis for our approach was to screen with a novel probe. We knew from the earlier experiments that our previous probe had successfully identified the cyclodeaminase gene on cosmid FD10. However, the protein was truncated and therefore presumed to be inactive. As such pseudo-genes typically arise from gene duplication of an original active gene copy, we assumed that there might be a second, complete cyclodeaminase in the genome. To evaluate this hypothesis, we designed a probe specific for additional copies of the cyclodeaminase gene, on the basis of the sequence of the truncated cyclodeaminase. Screening the genome with this probe by Southern Blot revealed two signals in addition to that corresponding to cosmid FD10, indicating that the genome contained a second or even a third copy of the cyclodeaminase gene. Importantly, this gene was likely to be associated with the remainder of the GE81112 gene cluster. We therefore re-probed the cosmid library with the specific cyclodeaminase probe, leading to the identification of new cosmids. We further confirmed that the cosmids contained the cyclodeaminase sequence by PCR. As we strongly expected NRPS sequence to be present on the cosmids, we used degenerate NRPS primers to amplify A domain sequences. Using these primers, we successfully amplified A domains from four of the identified cosmids. Prediction of the A domain substrate specificity revealed one cosmid containing an A domain with specificity for proline/pipecolic acid, exactly as predicted for the GE81112 starter unit. Sequencing of the cosmid revealed a biosynthetic pathway which harbored different genes (e.g. cyclodeaminase) which we expected for GE81112 biosynthesis. However, as the entire gene cluster was not present on the cosmid, we identified an overlapping cosmid. Furthermore, we verified that the genetic organization on both was co-linear with the genome.

To confirm the identity of the cloned biosynthetic locus, we first aimed to establish a method to genetically manipulate the strain. Although genetic manipulaton systems are well established for some streptomycetes strains [10], it turned out that the GE81112 producer was significantly less amenable to genetic manipulation. Nonetheless, using biparental mating, we developed a method for DNA transfer from *E. coli* via conjugation. We started method

development by conjugating both a replicative and an integrative vector as controls to test transformation efficiency. Using the standard amount of *E. coli* as described in Kieser *et al.* [10], we only obtained low numbers of colonies. As it was known from studies with other bacteria that the amount of *E. coli* can be crucial for conjugation efficiency [142], we tested different amounts of *E. coli*. Indeed, we found that transformation efficiency was 10–20-fold higher when large numbers (50×10^9) of *E. coli* were used. With a conjugation method established, we were able to inactivate gene *geE*, which resulted in the complete abolishment of the main GE81112 derivative factor B, which we could detect by high-resolution MS. This result provided clear evidence that the identified cluster was responsible for the production of the GE81112s.

4.3.3 Analysis of the GE81112 biosynthetic gene cluster

Sequence analysis of the identified cluster revealed that it harbored only NRPS components. Many clusters have a co-linear genetic and protein organization, following the so-called co-linearity rule, in which each module is responsible for one discrete chain elongation step and the specific order of the modules defines the sequence of the incorporated amino acids [64]. Contrary to our expectation that we would find such a linear NRPS system providing four modules with the standard C-A-PCP domain order, the cluster exhibited a highly nonlinear genetic arrangement (Figure 4.5).
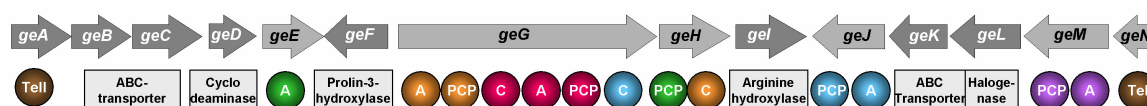
Figure 4.5: **Nonlinear arrangement of the GE81112 biosynthetic gene cluster.**
Domains that are proposed to interact are highlighted in the same colors.

Several examples of such nonlinear NRPS systems have been reported and initially, these NRPS were thought to be rare exceptions to the co-linearity rule [56]. However, during the past few years it has become clear that nonlinear NRPS are not an exception, but constitute a considerable fraction of the NRPS repertoire present in nature. The characteristic feature of these nonlinear NRPSs is at least one unusual arrangement of the core domains C-A-PCP. The GE81112 system also revealed a highly split genetic and module architecture, but when rearranging the domains they provide a linear assembly line as described in the results part. Although four amino acids are incorporated into the structure which would normally require

four modules, five modules harboring five distinct A domains were identified. Further unusual features of the cluster include two freestanding A-PCP didomains encoded by genes *geJ* and *geM*, and a stand-alone A domain, encoded by gene *geE*. Such stand-alone A domains have previously been identified in the nonlinear NRPS pathways to myxochelin [143] and yersiniabactin [144]. Similarly, freestanding A-PCP didomains have been found in the zorbamycin and syringomycin gene clusters [145;146].

The conserved regions of all domains in the GE81112 cluster were analyzed, and predicted to be functional. However, from the domain assignment alone, the order of subunits (modules) could not be predicted, as there were more modules present then expected for the biosynthesis of a tetrapeptide. We therefore began our analysis of the cluster with the prediction of the substrate specificity of the A domains, which we anticipated would give us a hint as to the order of the subunits. Based on analysis of the GE81112 structures, we expected that four amino acids would be the precursors of the biosynthesis incorporated into the compound in the order: pipecolic acid-ornithine/glutamine-histidine-histidine (Figure 4.6).

GE factor A	R=O	R_1=H
GE factor B	R=O	R_1=NH$_2$
GE factor B1	R=NH	R_1=NH$_2$
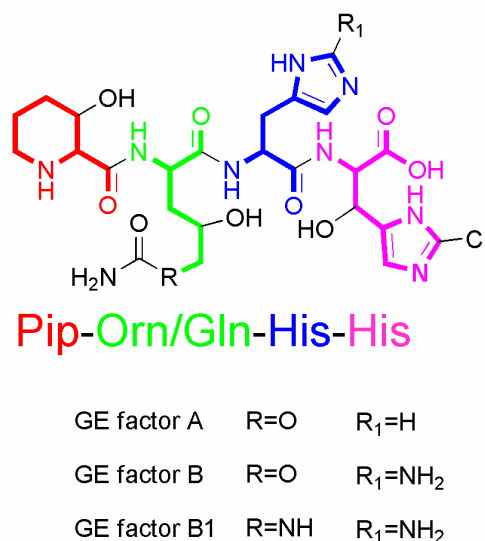
Figure 4.6: **GE81112 tetrapetide structure and corresponding amino acid sequence Pip-Orn/Gln-His-His.**

The predicted specificity for four of the five A domains correlated with our expectations (Table 3.13). Together with an analysis of putative docking regions, we were able to predict how the NRPS proteins interact with each other to biosynthezise the GE81112s (Figure 4.7). Overall, we suggest that the linear assembly line consists of four modules exhibiting the standard C-A-PCP domain arrangement, and that the pathway incorporates one additional A-PCP didomain, where the function was not clear, as well as a stand-alone thioesterase domain

for chain release. On both ends of the cluster additional *orfs* have been (Table 3.12) found where it was not clear if they have a function in the biosynthesis. What we suggest here is a first preliminary model of the biosynthesis where it remains to determine the boundaries of the gene cluster by knockout experiments.
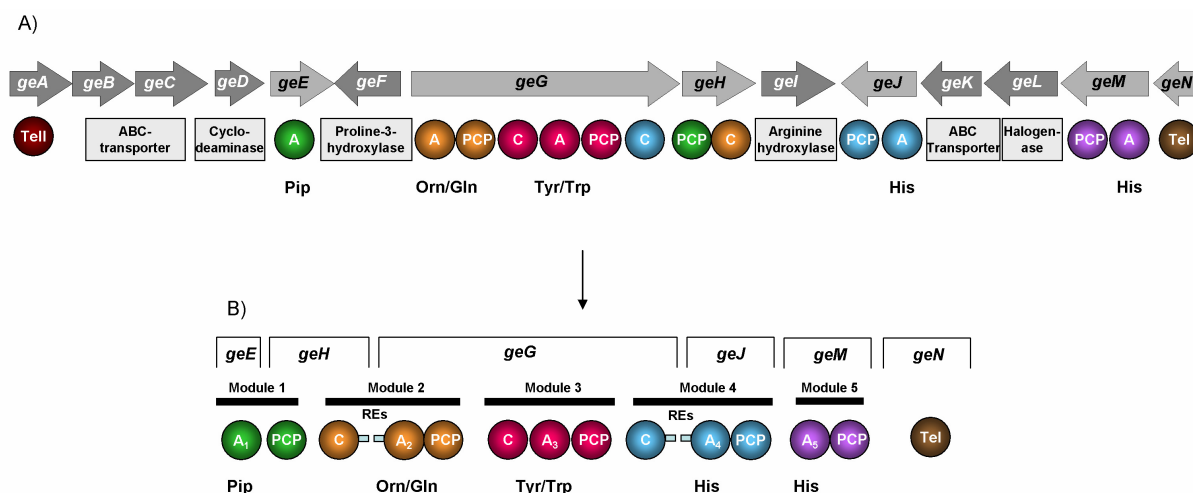
Figure 4.7: **A) Nonlinear genetic arrangement of the GE81112 NRPS and B) proposed interaction of NRPS proteins to yield a linear assembly line.** REs: Recognition elements.

Nonetheless, the function of the apparently superfluous module 5 remained unclear. One hypothesis was that the corresponding gene *geM* arose by gene duplication of *geJ* (or the opposite way around), as both genes showed the same domain arrangement and the encoded A domains were both predicted to activate histidine. If true, the protein GeM would also be expected to be able to interact with GeG to form the last module. However, sequence analysis revealed that the two proteins are not highly homologous, and that in addition, GeM was missing a putative recognition sequence for interaction with GeG as was present in GeJ. We therefore concluded that GeM is unlikely to interact with GeG, and therefore assigned the role of the fourth extension module to protein GeJ. However, as the two domains from GeM were predicted to be functional, we assumed that the didomain must have a specific function in the biosynthesis. A detailed discussion about a possible biosynthesis and the function of the single proteins will be given in the next chapter.

4.3.3.1 **Biosynthetic proposal for the GE81112 compound family**

Our data taken together allowed us to propose a pathway to the GE81112 metabolites, albeit incomplete. GE81112 biosynthesis likely starts with the biosynthesis of L-pipecolic acid from L-lysine via the action of the putative cyclodeaminase (Figures 4.8A and 4.14). Unfortunately we were not able to reconstitute this reaction *in vitro* when we expressed the GE81112

cyclodeaminase GeD, as GeD was inactive under our assay conditions. Pipecolic acid would then be activated by the A domain GeEA$_1$ in module 1 and loaded onto the adjacent PCP (Figures 4.8 B and 4.14). To evaluate the function and substrate specificity of the A domain, we performed an ATP-PP$_i$ exchange assay *in vitro*. This experiment clearly showed that L-pipecolic acid was activated, supporting our hypothesis that this is the starter unit for the biosynthesis. The pipecolic acid moiety in GE81112 is hydroxylated at the β-position; a reaction which we predict is catalyzed by the putative L-proline-3-hydroxylase, GeF. However, the timing of hydroxylation (on the free amino acid [147], while the amino acid is thethered to the PCP [148], or following release of the chain from the assembly line [55;90]), remains unclear and several examples of all three reactions have been reported [55].
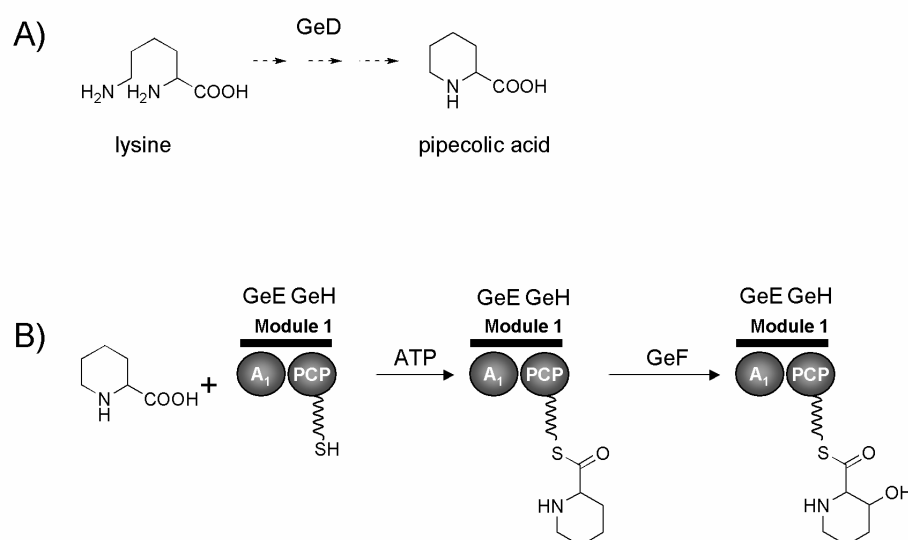
Figure 4.8:   **Proposed biosynthetic mechanisms.**

A) Generation of pipecolic acid from lysine via the action of a cyclodeaminase (GeD). B) Activation and loading of pipecolic acid by module 1 followed by subsequent hydroxylation at position 3 by a putative β-hydroxylase (GeF). Alternative timings of the hydroxylation reaction are possible.

It might have been possible to evaluate whether hydroxylation occurs before or after activation by the module 1 A domain, by assaying for the domain's relative specificity for pipecolic acid and hydroxypipecolic acid. However, as the required hydroxypipecolic acid was not commercially available, it was not possible to carry out this experiment. Nonetheless, we propose, based on literature precedent [148], that GeF most likely hydroxylates the PCP-bound pipecolic acid (Figures 4.8 B and 4.14).

It was not directly obvious from which amino acid the second building block is derived as there are two different structures found in the GE81112 factors A, B and B1 which differ at position 6. In GE81112 factor A and B the second amino acid incorporated into the structure is hydroxypentanoic acid, which is hydroxylated at position 4 and O-carbamoylated at position 6. In factor B1 it is an ornithine which is hydroxylated at position 4 and N-carbamoylated at position 6. To account for the biosynthesis of these two derivatives, we hypothesized that the A domain GeGA$_2$ of module 2 can activate either glutamic acid or glutamine, followed by reduction of the carboxylic groups which would lead to hydroxypentanoic acid (factor A and B) or ornithine (factor B1), respectively. The proposal is shown in Figures 4.9A and B.
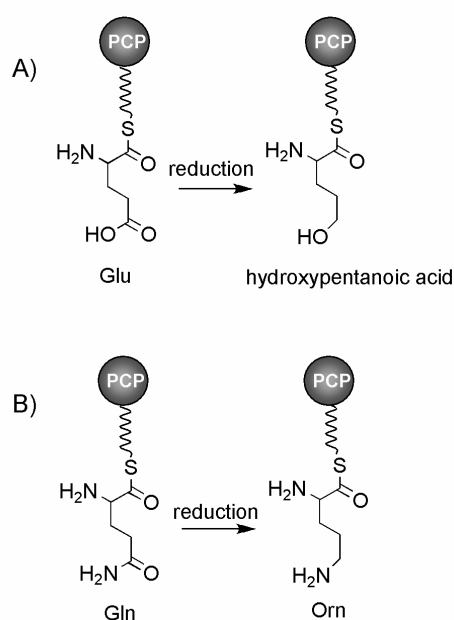
Figure 4.9: **Proposed biosynthetic mechanisms for generation of the second building block.**

A) Reduction of PCP-bound glutamate to hydroxypentanoic acid. B) Reduction of PCP-bound glutamine to ornithine.

The timing of the reduction is not clear, nor is the enzyme which would catalyze this reaction. The next step in either case, however, would be the hydroxylation of the γ-carbon putatively catalyzed by GeI. Finally, the hydroxypentanoic acid/ornithine would be modified by N-/O-carbamoylation. Again, the timing of the carbamoylation is not clear, but precedent examples suggests that it would happen following chain release [149;150]. Carbamoylation is normally catalyzed by a carbamoyltransferase, which we couldn't find in the biosynthetic pathway. So we cannot exclude that the carbamoylation is catalyzed by another mechanism. Additionally it might be possible that the carbamoyltransferase gene is located somewhere else in the

genome. Alternatively, ornithine (factor B1) would be activated directly by the A domain, but this model does not explain how factors A and B are generated. To determine directly which amino acid is activated by this A domain, we tested the function of the enzyme *in vitro* using the ATP-PP$_i$ exchange assay. According to our hypothesis, the A domain must show a broad substrate acceptance, activating at least two different amino acids (glutamate, glutamic acid or ornithine) to generate all three GE81112 derivatives. Indeed, activation of all the three amino acids was observed, although the A domain showed a preference for ornithine. Thus, it appears that ornithine is directly activated by GeGA$_2$, which accounts for the structure of factor B1. The reduced degree of activation of the other two substrates may indicate that modified versions of glutamic acid and glutamine are preferred.

The third amino acid incorporated into the chain is either histidine (in GE81112 factor A) or aminohistidine (as in factors B and B1). In the case of GE81112 factor A, histidine is likely to be incorporated directly, but the origin of the aminohistidine is less obvious. One hypothesis is that the aminohistidine might be derived from cyclization of arginine by nucleophilic attack of the γ-carbon (Figure 4.10). This step could be catalyzed by an arginine cyclase operating on a PCP-bound substrate. This proposal implies that the A domain GeGA$_3$ tolerates different substrates, activating both histidine (factor A) and arginine (factors B and B1). However, an additional possibility is that cyclization of arginine to aminohistidine occurs before the amino acid is thethered to the PCP, making aminohistidine the second substrate of the A domain. Thus, according to this alternative hypothesis, the A domain should be able to activate both histidine and aminohistidine.
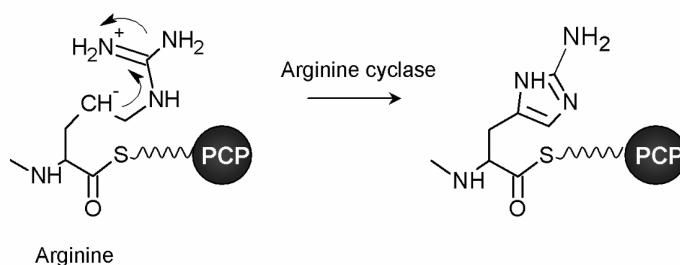
Figure 4.10: Proposed mechanism for the PCP-bound cyclization of arginine by a putative arginine cyclase.

As the cluster lacks an obvious arginine cyclase, it appears likely that aminohistidine is generated by another mechanism. The *in silico* analysis of GeGA$_3$ appears to support this proposal, as it suggests a specificity for Tyr/Trp and not arginine. This prediction may indicate a general preference for aromatic residues, accounting for activation of both histidine

and aminohistidine, but it is not clear how aminohistidine could be generated from tyrosine or tryptophane. Furthermore, amination of the PCP-bound histidine or after release of the chain cannot be excluded. To evaluate one of the hypotheses the suggested putative substrates must be tested in the ATP assay, which remains to be done.

The last amino acid to be incorporated in GE81112 is a 6-chlorinated histidine. The chlorination reaction is expected to be catalyzed by a halogenase. Halogenation is a common modification of natural products and many halogenated compounds are known [151-153]. The timing of halogenation reactions during natural product biosynthesis has been shown to vary. For example, pre-assembly line chlorination of tryptophane occurs in rebbecamycin assembly [154], a PCP-bound threonine is modified during the biosynthesis of syringomycin [155], while β-tyrosine is chlorinated following release of the peptide chain, in the assembly of chloroereomycin [156]. A candidate for catalyzing the chlorination of the GE81112 compounds is the halogenase GeF. A number of different scenarios can be proposed for how the histidine is chlorinated and incorporated into the metabolites (Figures 4.11 A, B and C). Common to all three is the role of module 4 in incorporating the (chloro)histidine, which is also hydroxylated at the β-position.
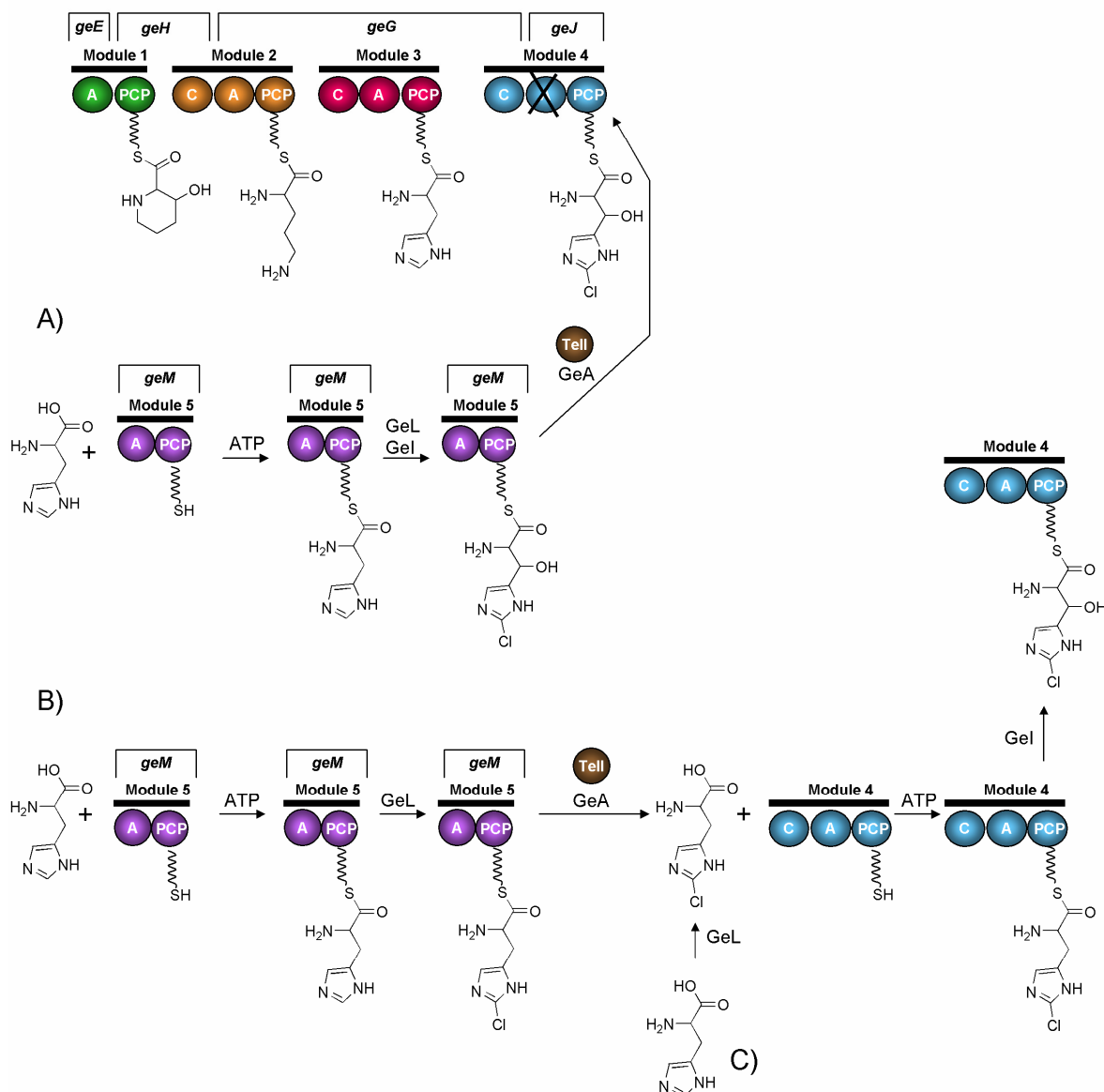
Figure 4.11: **A linear model for the GE81112 NRPS templated assembly from four amino acids.**

In the first proposal (A), biosynthesis starts with the activation of histidine by the A domain of GeM (which is predicted to show specificity for this amino acid) which is then bound to the PCP of GeM (didomain protein). The putative halogenase GeL would then chlorinate the PCP-bound histidine to afford chlorohistidine, as has been proposed for the chlorination of L-threonine in the syringomycin biosynthesis [146]. The hydroxylation could then occur on the Cl-His-*S*-GeM intermediate; alternatively hydroxylation would occur on the free histidine (prior to PCP attachment), or following release of the tetrapeptide from the synthetase. However, as the majority of hydroxylations have been shown to occur on PCP-bound amino acids [148], we believe that this mechanism is most likely. Nonetheless, it is unclear whether hydroxylation proceeds or follows the chlorination. The 3-hydroxy-6-chlorohistidine would

then be transferred onto the PCP of GeJ. There are several examples of such PCP-to-PCP shuttling in the literature. In the syringomycin pathway, for example, the freestanding A-PCP protein SyrB1 activates and loads threonine. The halogenase SyrB2 then acts on the Thr-*S*-SyrB1 intermediate to produce 4-Cl-Thr-*S*-SyrB1. The Cl-Thr is then shuttled *in trans* to the PCP of a C-PCP didomain of the NRPS synthetase SyrE. This reaction is catalyzed by the acyltransferase SyrC, a protein belonging to the hydrolase superfamily [157]. Postulated functions for members of this protein family also include thioesterase and haloperoxidase [158]. A similar mechanism has been elucidated recently for zorbamycin biosynthesis [145], in which L-valine is loaded onto the PCP domain of ZmbVIIb, hydoxylated by ZmbVIIc, and subsequently transferred onto the PCP domain of ZmbVIIa by the predicted type II thioesterase ZmbVIId (Figure 4.12).
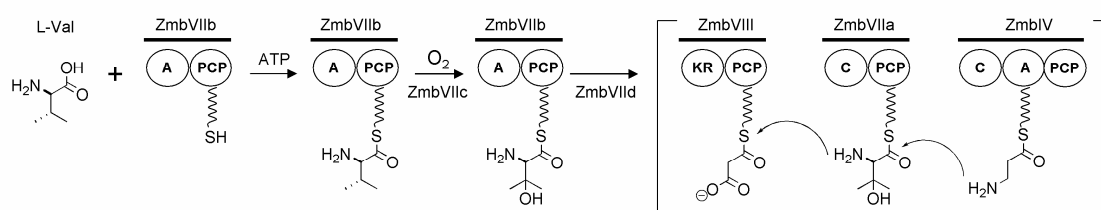
Figure 4.12: **Proposed mechanism for the biosynthesis and incorporation of L-hydroxyvaline in the zorbamycin biosynthetic pathway.**

The coronamic acid biosynthetic pathway incorporates a further example of the shuttling of aminoacyl moieties [159]. In this example, the aminoacyltransferase CmaE shuttles an *allo*-Ile moiety between the pantetheinyl arm of the PCP domain of CmaA (A-PCP didomain) and the freestanding PCP domain of CmaD. By mechanistic analogy, the histidine in GE81112 biosynthesis may be loaded onto the PCP domain of GeM, chlorinated by GeL and/or hydroxylated by GeI, and subsequently transferred onto the GeJ PCP domain by the predicted type II thioesterase GeA, acting as acyltransferase (Figure 4.11A). In this model, the A domain in module 4 would not be necessary.

Alternatively (pathway B) histidine is loaded onto the PCP of GeM and chlorinated by GeL leading to chlorohistidine. The chlorohistidine would then be released by GeA acting as a thioesterase, as described for BarC from the barbamide biosynthetic pathway [160]. The free chlorohistidine would then be activated by the A domain GeJA$_4$ and loaded onto the PCP of module 4. In the next step, β-hydroxylation catalyzed by GeI would take place on the PCP-bound species.

In these proposed pathways, GeA is predicted to function either as an acyltransferase or as a type II thioesterase, both of which are known to release amino acids from PCP domains [100]. To attempt to provide support for one or the other of the pathways, we aligned GeA against other known acyltransferases/thioesterases. The aminoacyltransferases identified to date have been assigned into two groups [145]. The first group, comprising SyrC, CmaE, ZmbVIId, contains a GxCxG motif harboring an active site cysteine; enzymes in this group are predicted to act as acyltransferases. Indeed, the active-site cysteine of these enzymes has been shown to be directly involved in aminoacyltransfer [159]. The second group contains acyltransferases BarC [160] and CouN7 [161], which both contain an active site serine (GxSxG) and are predicted to function as ordinary thioesterases. For example, in the barbamide biosynthetic pathway, BarC is proposed to release trichloroleucine by hydrolysis [160]. When we aligned the two thioesterase-like proteins GeA and GeN from the GE81112 biosynthetic gene cluster, the presence of an active site serine was clear, suggesting that the proteins both function as normal thioesterases (Figure 4.13). Thus, it seems likely that GeA is acting as a type II TE catalyzing the release of chlorohistidine (pathway B). The second thioesterase, GeN, is expected to catalyze the realease of the peptide from the assembly line (Figure 4.14), even though it is a discrete protein (i.e. type II architecture).

```
Cons.                                                  C
                                                    GxSxG

GeA    LLAVQLPGRE--NRWREEPATGMAQVLDELVPALASRLDLPY---AVFGHSMGALVGYEL
BarC   LVAIQLPGRE--KRCYEPLMSNIEDINLVLAQELSPFLSKPF---AFMGHSMGALIAFDL
GeN    PVLLQTPGRE--NRLAEPAVDDIDALVTQLATALRGHLDAPF---ALFGHSMGSLVAFEL
Syr    VVTWETRGLFGACEAFDQIAVDTDAQVADMISVMNHFGLSTA---HLMGICAGAVIALSA
CmaE   LIILESDTWL---AYANEAGVNPEEGVADFIRQFNAALPEPVRVDALVGYCSSAPLALLA
```

Figure 4.13: Sequence alignment of acyltransferase/thioeseterases found in NRPS clusters. TheGxC/SxG is highlighted.

The final alternative (pathway C) is that free histidine is chlorinated, and then activated and loaded onto the PCP of module 4. The remainder of the pathway would then follow that of proposal B. Thus in pathway C, module 5 does not play a role, but according to sequence analysis it is probably active.
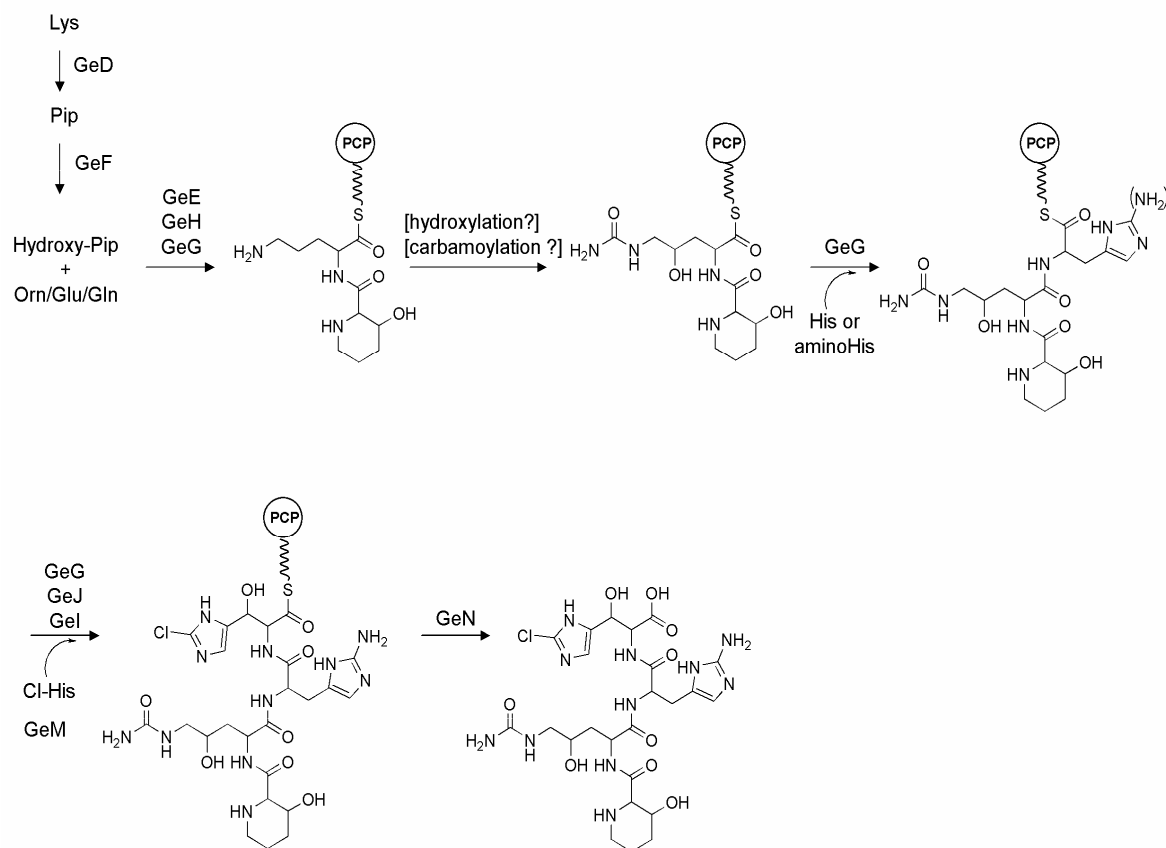
Figure 4.14:   Proposed GE81112 biosynthetic pathway of factor B1.

Overall, pathway B appears to be the most reasonable, as the TE is clearly predicted to act as a thioesterase catalyzing the release of the chlorohistidine, rather than its transfer to GeJ. Further supporting pathway B is the fact that in pathways A and C, no function is required of a single domain or a didomain, respectively, when there is no evidence from sequence analysis that any of these domains should be inactive. Indeed, there are a number of examples in the literature in which apparently superfluous modules have been shown to be essential for natural product biosynthesis [157;159-161].

4.3.4 Summary and Future aspects

The cloned GE81112 biosynthetic pathway could be identified and cloned and comprises a complex NRPS biosynthetic machinery. Many unusual features were found in its modular architecture, including A domains which likely incorporate unusual amino acids, such as pipecolic acid and chlorohistidine. Although many questions about the biosynthesis remain, the availability of the GE81112 cluster as well as tools for genetic manipulation now provide a platform for attempts to decipher these issues, and to generate new GE81112 derivatives by genetic engineering. A future goal will be to define the cluster boundaries by further gene inactivation experiments. To obtain further experimental evidence for the proposed biosynthetic pathway selected proteins can be studied *in vitro*. At first the three remaining A domains should be expressed and tested in the ATP assay for the activation of the suggested substrates. Furthermore it would be interesting to characterize the PCP of module 5 to demonstrate that this module is really needed for the biosynthesis. Another possibility to further investigate the gene cluster and to increase the production of the GE81112 compounds would be the heterologous expression. Therefore the cluster would need to be stitched from the two cosmids as it was already described for the phenalinolactone cluster and transferred into heterologous hosts, such as *S. lividans*.

Additionally to the GE81112 biosynthetic gene cluster another novel NRPS cluster could be identified from *Streptomyces* 14386. Successful heterologous expression of this gene cluster revealed that it was responsible for the biosynthesis of three new peptides, which we speculated to be related to the GE81112 family as they seemed to be also chlorinated tetrapeptides. To get an idea about the structure of this compound family we purified all three compounds and tried to elucidate their structure by NMR. Although structure elucidation was not completely finished, it turned out that theses substances incorporate unusual amino acids. To finally determine if these substances are related to the GE81112 family, full structure elucidation needs to be done. Once the structure elucidation is finished investigations on the mode of action of these substances can also be done. However, with the heterologous expression of these substances and the genetic manipulation methods developed for *S.* 14386, we now have the basis for gene deletion experiments to investigate the biosynthesis of this new class of peptides.

5   Appendix

Table 5.1:   **Sequence alignment of putative modules 1-5 from the GE81112 biosynthteic pathway with two modules of the NRPS pathway from cosmid FD10.**

```
C-A-PCP_FD10_2      1    --------------------------MLFLALYDEQAVDTYTFQLAFELEGDVDAAGLRAAGDALLRRY
C-A-PCP-FD10_3      1    -----------PYARPDVIPLSVGQRGQWFLNRFDD-GSGAYNILHTLRLTGALDREALRLAIDDVVARH
C-A-PCP_4           1    -------------ADDGFLPLSRAQRRVWLTSRRAG--SETFVISDLVRLGRRIDADALRRALGAVVERQ
C-A-PCP_5           1    -------------ADDGFLPLSRAQRRVWLTSRRAG--SETFVISDLVRLGRRIDADALRRALGAVVERQ
C-A-PCP-3           1    ---------VLWPETPGPVPASNAQRRMWFLEQYEGGELRPYNMVEAFRLPGRHEAADVAVALERTVRRH
C-A-PCP-2           1    EPAGPAVPRQRPADPLPGTEVSAATRWLWLERQRADGDFGAYTVPCLLEGDPLDTDRLAAALTVVARRH
A-PCP_Loading1      1    -------------------------------------------------------------------

C-A-PCP_FD10_2      44   PNLRAGFRHEKLSRPVQVIPHQ-AEVPWREVDLRAESADGRAAALARLADEDRTHRFDLSRPPLLRFTLV
C-A-PCP-FD10_3      59   ESLRT-IYPEIDGISRQVILEG-AEVDWTVREVPEEQLAG--ALAAEAAEG----FDLSLDIPVRAALF
C-A-PCP_4           56   DMLRAQALPRGTEAELTVLDRLPDGVPLVVVELPCADPSGPEAAAALRDARRTG--FALDRAPLFELRLL
C-A-PCP_5           56   DMLRAQALPRGTEAELTVLDRLPDGVPLVVVELPCADPSGPEAAAALRDARRTG--FALDRAPLFELRLL
C-A-PCP-3           62   QALRTVFRTEPEGLRQIVLTPGEARPELRVHRARGTDMAG---LLRRVADAEQRHVFDPATGPLLRAHWL
C-A-PCP-2           71   RALRTVFTEAHGEPRAVVTDDVPALEVADLRGQDAG-----DARFAACVDERVAVPFDPATGPLVRATVF
A-PCP_Loading1      1    -------------------------------------------------------------------

C-A-PCP_FD10_2     113   ----RLGD--------------------EQYRLMIVLHHILLDGWSFPLLVNDLFELYGRRGDETGLPR
C-A-PCP-FD10_3     120   ----VLGD--------------------TECVLLLSMHHIASDGWSLAPLSRDLSVAYRARLAGAAPSW
C-A-PCP_4          124   ----TGVV--------------------GGDLLTVTAHHLIYDGASVDVLLRDLFTAYEQAVAGEPPRL
C-A-PCP_5          124   ----TGVV--------------------GGDLLTVTAHHLIYDGASVDVLLRDLFTAYEQAVAGEPPRL
C-A-PCP-3          129   ----VEGANAGGDEGGDKSGDESGDSGGEAGGTLLLSVHHSVCDGWSFAVILRDLLRFLDGTEELPAPTQ
C-A-PCP-2          136   LLPG----------------------DRWSLLLLADHLVCDGRSIELLADQLVAAYADPASDVGRPP
A-PCP_Loading1      1    -------------------------------------------------------------------

C-A-PCP_FD10_2     158   VTPYREYLAWLSRQ-----DKAGAEA-------AWRGALDGVTEPTLLAPAAAANAVRVAPEEFTVDLSQ
C-A-PCP-FD10_3     165   AGLPVQYADEALWQREVLGDEADPESELCRQLSFWTGRLDGLPQ-ELTLPADRQRPTESSGRGGLVDFTL
C-A-PCP_4          169   PPLTYTYQEWVREERE---WLAGPEAER--EVAFWRERLRGLAE--SPDPVDPARRRSRRGRTGQVHRTV
C-A-PCP_5          169   PPLTYTYQEWVREERE---WLAGPEAER--EVAFWRERLRGLAE--SPDPVDPARRRSRRGRTGQVHRTV
C-A-PCP-3          195   YGAYALGQHTRRDDMEG-------------LEFWRRTLDGAPA--VDLPLDRRRPRERGSAAGTVRLAV
C-A-PCP-2          181   GGEPAEETPAARPADPR-----------AALDHWRSVLEPPPA-PLPLPVRGPRTEPAGSTTGIAAREI
A-PCP_Loading1      1    -------------------------------------------------------------------

C-A-PCP_FD10_2     216   ET--TAALDSLARARGLTMNTLVQGAWGLLLGSATGRSDVLFGATVSGRPAELPGVQSMVGLFINTIPVR
C-A-PCP-FD10_3     234   EPGLHQGLVELARSGRASVFMVLQACVAALLSRLGAGGSDIPIGTPVAGRTDEA--LDDLVGEFLNTLVLR
C-A-PCP_4          232   PA-------VLLRATAATPFAQVVTAFALTVRHHTGATDLVLGFAAGLRDRPE--ADQLIGYLVNAVPLR
C-A-PCP_5          232   PA-------VLLRATAATPFAQVVTAFALTVRHHTGATDLVLGFAAGLRDRPE--ADQLIGYLVNAVPLR
C-A-PCP-3          249   GEDTTRAVQGLCRELGTTSFTAMVAAVRVLLLRWSGADDIVLGTVVSGRDQPE--LADSVGLFVNTVALR
C-A-PCP-2          238   GPDVLRRLTDLGRHHHAGPFVPLAAAVARTLAGLTCSADICLGTAVDRRARAG--AKDAVGFHVATVPLR
A-PCP_Loading1      1    -------------------------------------------------------------------

C-A-PCP_FD10_2     284   VGLDPAESVAGLFGRIQDQQTELMDFHHVGLT------EIQRLTGQSVLFDTLAVFEN---YPLDSGELE
C-A-PCP-FD10_3     302   TDTSGDPAFDELVGRVREVNLAAYAEQDVPFERLVEALKPERSTARHPLQVMALQNQAEARLELPGLH
C-A-PCP_4          293   VTLDDTVTGGELLPRVQHGIVEAYEHARLPFDVLAERLALRAGPGRSLLLDLGVSWEN---------AAL
C-A-PCP_5          293   VTLDDTVTGGELLPRVQHGIVEAYEHARLPFDVLAERLALRAGPGRSLLLDLGVSWEN---------AAL
C-A-PCP-3          317   TPVEPERGFAELVAAVAAHAREAAHEEYPFEHLVEALGAERVMGRNPLFDVLVEASVSG---TDPLGSG
C-A-PCP-2          306   LDVRDEMSADDLVRHVARRTMDAVDFSDVTFDGLVAALDPPRSPGRMPFFDVWVALYP-----RIDTGPR
A-PCP_Loading1      1    -------------------------------------------------------------------

C-A-PCP_FD10_2     345   KSVDGVRVRGLTARDATHYPLSLIAYPG-EHMVLRVGHRPDLVDRETAETLVGRLVRILEAVAADPDRSV
C-A-PCP-FD10_3     372   VVPEPVDIATSKFDLALDFTERFAADGTPDGVDGFLQYSADLFDRETAETLVARLSRFLGSVAAEPGQRI
C-A-PCP_4          354   RPESQVVEDLLPDDLPATSDLWLYARVKGDELHLDLTYDDNLLDEAEAGAYADDVAAVLRCLAAGATAPP
C-A-PCP_5          354   RPESQVVEDLLPDDLPATSDLWLYARVKGDELHLDLTYDDNLLDEAEAGAYADDVAAVLRCLAAGATAPP
C-A-PCP-3          384   DGPAAEHIRLDSGAEGFDLAFSFTEPAPGRVALTYREDVLDAAIVHRAAGQLRHLLTLAVADPSVPV
C-A-PCP-2          371   PPDGIALRGGPTPLRVGMFELSFQCVEHAHGMRLTLQYDALRYAADTAERMVQRLGDEVARLLGDPTEPA
A-PCP_Loading1      1    -------------------------------------------------------------------

C-A-PCP_FD10_2     414   ASLDMLSVG----------ERRLVLGEWAGVGGVVSED-----------------------------
C-A-PCP-FD10_3     442   CAVEVLAPA----------ERERVLVEWNDSAHELAP-----------------------------
C-A-PCP_4          424   AQPTVPPPCPPPPAAPPAPNPSAPGPVWSDTRYDFLIPEQGKVRMTSDSAVPATTVRPRRLGPAHSAPAR
C-A-PCP_5          424   AQPTVPPPCPPPPAAPPAPNPSAPGPVWSDTRYDFLMTPENELKAP----------------------
C-A-PCP-3          454   CSAPLLPPD----------QRDALLAAGTGPRGPVPRTAG-------------------------
C-A-PCP-2          441   PETAPAHRSFGGFQFDALAAGHQVPFLDVVSSAGGTPECALDPSATERLDALTGGDPTG----------L
A-PCP_Loading1      1    -------------------------------------------------------------------
```

```
C-A-PCP_FD10_2    442  ------------------------------------------------------------
C-A-PCP-FD10_3    469  ------------------------------------------------------------
C-A-PCP_4         494  ARLPLTGAAADGLAKVSGGRPLEELVGLTAAATVVVGSAEHTDEPVVAVSGPSGPVLCGIGAAAPATVAD
C-A-PCP_5         470  -----------------------------------------------------------L
C-A-PCP-3         484  ------------------------------------------------------------
C-A-PCP-2         501  LLLTVAAARLALAALDEESRHVVLVPAPTNASRADPAADADTAPRELALCAPLDRSAPASVFLTSLHAEL
A-PCP_Loading1      1  ------------------------------------------------------------


C-A-PCP_FD10_2    442  ------------------------------------------------------------
C-A-PCP-FD10_3    469  ------------------------------------------------------------
C-A-PCP_4         564  LVRAVDAALRSAPAADVRDAALAGAVLVSSARVGTAGAPAVRRDAIELTLSEPGADGGRDLVAEAAPERA
C-A-PCP_5         471  ITRSVGARLRESRIHDN----------------------------------SAGGRLSTNRDDPK--
C-A-PCP-3         484  ------------------------------------------------------------
C-A-PCP-2         571  ESALPLAWHDREDVASRLRLAGVPAAHALARLGVVCEEAGGKLLQPVGLCLTLRRSAGRLVLTAEAGGEL
A-PCP_Loading1      1  ------------------------------------------------------------


C-A-PCP_FD10_2    442  ---------------------------------------------------VTLPGLFEAQVVRAPG
C-A-PCP-FD10_3    469  ---------------------------------------------------ATLADLFEAQAARTPA
C-A-PCP_4         634  EAWFLEVLLRSVARVLAGFTDPHDAVADLPRAGADDVAAATDFGRRGFEPDGVTTTLIAPIEAAVHAYPD
C-A-PCP_5         502  -----------------------------------------------------TLTFLFEECARNVPG
C-A-PCP-3         484  -------------------------------------------------------TLLDLVLAQVRRAPE
C-A-PCP-2         641  PVAATALLRCAAAALEALGRSPRHPVAALDVLGPAQRAELERWSGLPVVAEFADATLTGVLDDAAARFPG
A-PCP_Loading1      1  ----------------------------APSPPGHAGPVVDTGVSTVLELIDREVAAHPGDPAVREPG


C-A-PCP_FD10_2    458  STALVFGGESVSYGELNARANRLARFLIARG-VGPERFVAVSLPRSVELVVAVLAVLKAGGAYVPVDPDY
C-A-PCP-FD10_3    485  GTALVHGEESLSYGELNARANRLARLLTGRG-VGPEQVVALALPRSPDLVVAVLLAVVKAGAAYLPVDPEY
C-A-PCP_4         704  RTAVVAGARTLTYREFWRATEAVADRLRASG-IGRGQRVGVLTGKSVHAVPAITGTLLAGAAYVPLDPRS
C-A-PCP_5         517  NIALEFEGEFLSYGELNRRANKLAHHLRETGGVRPDDRVALLVRHPPDVIIAELAVLKAGGAYVPLDPDN
C-A-PCP-3         499  GRAVVCGDRVLTYGDLDRRADALAARIAAVAPTGPDRVVAVVCDRSEWPAALLAVLKTGSAFLPLDPEQ
C-A-PCP-2         711  REAVSDGTVVVWSHRDLAERSSRAARTLALDHGLKPGDRVALLPKSPELVVAVLAVLRAGASFVPLDPSH
A-PCP_Loading1     41  REVT-------YAELDRLADTVAGRLVGEFGIGRGDVVLLAARAGADFTAAVLGTLRAGAAYLPVDTAY


C-A-PCP_FD10_2    527  PAERVAYLLEDSAPSFVLD---------------EAAFAGADVSGFSDVDVRDGERLGVLSGLS------
C-A-PCP-FD10_3    554  PAERIAFMLRDAAPALVLTTSGTGIGTTAPALPLDAAETVAALADHSDADLTDADRTTPLTDRH-----
C-A-PCP_4         773  PAARLAEEILDDAACGAVLVAPDMLDALPEGLAAPVIDLAAVAGGDAGAPAPGEQSAPAAPTAPTG-----
C-A-PCP_5         587  PPCVTQRRIIEDVTPRAMVIES------SAAAGAVFFDGELFVIDVMSDALETPESDPEPEVTPS------
C-A-PCP-3         569  STARLVGVLRDSCAVAAVASAQFAGVTAGAGLPTVTVGAPGAAAGQDAPVRGARDQDAPVLDEPREARAA
C-A-PCP-2         781  PPARVARQLRLSCAVCVISARDELPGDLAVPVVSP---RALSAAPGPCATADACTPSGPAPIDE------
A-PCP_Loading1    103  PPERIAQILGASSAALLVLADPGSLDTGAAGTPATGLACLVAPEPSGCPVPAAAVRRGGPAPEDP------


C-A-PCP_FD10_2    576  -----AAYVIYTSGSTGRPKGVVVPHGNVVRLFSATDRWFDFGPDDVWTLFHSFAFDFSVWEIWGPLLHG
C-A-PCP-FD10_3    618  -----PVYVIYTSGSTGTPKGVVVTGALTGQIRWLAAETELTPRDVVLARTPVSFDAAGAELWPPLVSG
C-A-PCP_4         838  PLPDDLAYVVYTSGSTGRPKGVETRHRAIASYVHWKLRNHGLDHDTRVLQLPSLAFDSSVADLFPVLASG
C-A-PCP_5         645  ----DLAYVIYTSGTTGTPKGIAVEHQAIVNTLTWRNAYYGEGTDAVNLAIPRPSFDSSVADTFCSLTTG
C-A-PCP-3         639  AGHADLAYVVYTSGSTGAPKGVMVEHGGIVNSVRERVDHYGLGADGAVLQVDPIHADAGIVDVFSALASG
C-A-PCP-2         842  ------AILFFTSGSTGLPRPVALTERQLAHKVLASGRLVGFDEEIRCALLSAVTSDALTYQIFTTLAAG
A-PCP_Loading1    167  ------AYVIFSSGSTGAPKGIVQTERCLANFISWQVEGSGLGRGRRVLQVAPLTFDVSVQEMFYTLASG


C-A-PCP_FD10_2    641  GRLVVVPFAVSRSPREFRALLVREGVTVLSQTPSAFYQLMAADAECGAGDCELVLRRVVFGGEALDLWRL
C-A-PCP-FD10_3    683  AAVALASAEVTRDPEQLLAFIGHHGVTVAQFVPS----LLAATPLDERGRG---IRVLLMGGEALPAALA
C-A-PCP_4         908  GLLVLADTHKLVPR-QLAELAEQHRVTHMTTVPSLYRVLLDELP-----RAADSLRAVTVAGEATTPDLV
C-A-PCP_5         711  TRLVLPRRDRITDRRYLTGLMESAGVTHFLITPVLYKRLLGAMDA----QRLKTLRCVTVAGEWFTSSLT
C-A-PCP-3         709  APMVIITRDQLLTPEEVAAVVRQHPIRHVLLVPSLYQVLLDEVGP-----AFRGVREIVLGGEVVTRALA
C-A-PCP-2         906  GCVVPMDSPQTLAPQEFWTTTRRLRVNVINCVPSLLAVMAEG----LPPGAAEDVRICLLGGDEIPAGFL
A-PCP_Loading1    231  GCLHVPEAHVRRDPRDLIDFVIDERIEVVDFPQSLIDVVMTLPT--TFEHAADLRHIISAGETVRVNEAL


C-A-PCP_FD10_2    711  GDWYARHGDAAPVLVNMYGITETTVHVSYVALDAG-------------RVACNAGSVVGRGLADLRVYVLD
C-A-PCP-FD10_3    746  GQAAAAWG---AQVINVYGPTEATICATSGRPDGS-----------GDDATT-VPIGRPAWNTRVYVLD
C-A-PCP_4         972  RRHHELLP--GVRLTNEYGPTECSVGATAFDHG--------------TDAGPGVPIGWPISNTVATVTG
C-A-PCP_5         777  KEHYRRLP--EVGLTNEYGPSENAVCSTVHPLA-----------ATDES-VLIGKPVDHTEAFVLD
C-A-PCP-3         774  ARHAELLP--GVRLTNEYGPAEDSVLTTVELVEPAGTDGGSGAPDTRSPTASGDASIGRPLPGKWVDLLD
C-A-PCP-2         972  PRLADRLR--VGTFANLYGPTEATTEATTFTCPGSALP-----------ALTTVPVGREGFGVVVLT
A-PCP_Loading1    299  EALLTRRP--EITLHNHYGPAENHMVAAHSMS-GAAGN-----------LEPGPPVGSLVWNTYIYLLD
```

Table 5.2: **Sequence alignment of proteins GeJ and GeM from the GE81112 biosynthetic pathway**
Sequence homology was 29%. Protein GeJ shows an N-terminal linker sequence which is not present in GeM.

```
────────────────────────────────────────────────────── Section 1
        (1) 1        ,10        20         ,30        ,40          57
     geJ (1) LIPEQGKVRMTSDSAVPATTVRPRRLGPAHSAPARARLPLTGAAADGLAKVSGGRPL
     geM (1) --------------------------------------------------------
Consensus (1)

────────────────────────────────────────────────────── Section 2
       (58) 58         ,70        ,80        ,90        ,100        114
     geJ (58) EELVGLTAAATVVVGSAEHTDEPVVAVSGPSGPVLCGIGAAAPATVADLVRAVDAAL
     geM (1)  --------------------------------------------------------
Consensus (58)

────────────────────────────────────────────────────── Section 3
      (115) 115  ,120        ,130        ,140        ,150        ,160      171
     geJ (115) RSAPAADVRDAALAGAVLVSSARVGTAGAPAVRRDAIELTLSEPGADGGRDLVAEAA
     geM (1)   --------------------------------------------------------
Consensus (115)

────────────────────────────────────────────────────── Section 4
      (172) 172     ,180        ,190       200        210           228
     geJ (172) PERAEAWFLEVLLRSVARVLAGFTDPHDAVADLPRAGADDVAAATDFGRRGFEPDGV
     geM (1)   -------------MTPENELKAPLITRSVGARLRESRIHDNSAGGRLSTNRDDPK--
Consensus (172)            S      L A       A L   A   D AAA      DP

────────────────────────────────────────────────────── Section 5
      (229) 229        240        ,250        260        270          285
     geJ (229) TTTLIAPIEAAVHAYPDRTAVVAGARTLTYREFWRATEAVADRLRASGIGR-GQRVG
     geM (43)  --TLTFLFEECARNVPGNIALEFEGEFLSYGELNRRANKLAHHLRETGGVRPDDRVA
Consensus (229)   TL     E        P   AL   A   LSY E  R     LA  LR SG  R    RVA

────────────────────────────────────────────────────── Section 6
      (286) 286        ,300        ,310       320        ,330        342
     geJ (285) VLTGKSVHAVPAITGTLLAGAAYVPLDPRSPAARLAEILDDAACGAVLVAPDMLDAL
     geM (98)  LLVRHPPDVIIAELAVLKAGGAYVPLDPDNPPGVTQRIIEDVTPRAMVIESSAAAGA
Consensus (286) LL      I A   A L AGAAYVPLDP  P A    IIDD     AMLI       A

────────────────────────────────────────────────────── Section 7
      (343) 343    ,350        ,360        370        ,380          399
     geJ (342) PEGLAAPVIDLAAVAGGDAGAPAPGEQSAPAAPTAPTGPLPDDLAYVVYTSGSTGRP
     geM (155) VFFDGELFVIDVMSDALETPESDPEPE-----------VTPSDLAYVIYTSGTTGTP
Consensus (343)     A   I     A D      P                 P DLAYVIYTSGSTG P

────────────────────────────────────────────────────── Section 8
      (400) 400     ,410        420        ,430        ,440          456
     geJ (399) KGVEIRHRAIASYVHWKLRNHGLDHDTRVLQLPSLAFDSSVADLFPVLASGGLLVLA
     geM (201) KGIAVEHQAIVNTLTWRNAYYGFGTDAVNLAIPRPSFDSSVADTFCSLTTGTRLVLP
Consensus (400) KGI  I H AI    L WK   HG   D   L IP   AFDSSVAD F  L SG   LVL
```

```
                                                                    Section 9
         (457) 457         470        480         490        500          513
    geJ (456) DTHKLVPR-QLAELAEQHRVTHMTTVPSLYRVLLDELP-RAADSLRAVTVAGEATTP
    geM (258) RRDRITDRRYLTGLMESAGVTHFLITPVLYKRLLGAMDAQRLKTLRCVTVAGEWFTS
Consensus (457)    KI  R   L   L E    VTH    P LYK LL   L      SLR VTVAGE   T
                                                                    Section 10
         (514) 514    520         530         540        550         560      570
    geJ (511) DLVRRHHELLPGVRLINEYGPTECSVGATAFDHGTDAGPGVPIGWPISNTVATVTGA
    geM (315) SLTKEHYRRLPEVGLFNEYGPSENAVCSTVHPLAATDES-VLIGKPVDHTEAFVLDE
Consensus (514)  L K HH   LP V L NEYGPSE AV AT      A       V IG PI   T A V
                                                                    Section 11
         (571) 571       580        590        600        610           627
    geJ (568) DGRALPVGFVGELRLAGHGLANGYRNRPELTARAFVAD--PEAPGGRGYRTGDLVWW
    geM (371) DGAPVEPGGTGELHLAGARLARGYLGDPELTAERFFTPSGPPAAGKRVYRTGDIVRL
Consensus (571) DG   L   G  GEL LAG  LA GY    PELTA  F       P A G R YRTGDIV
                                                                    Section 12
         (628) 628        640        650        660        670           684
    geJ (623) RPDGMLEFAGRADDQVKIRGHRVEQGEVESVLGVLPGVLQAAVAVVSGPDGAPALAA
    geM (428) LEGGDLQFIERRDGQVKIRGRRVELGHVATVLSKDPSVGHVHLMRHGTESGTPLLVA
Consensus (628)     G L F   R D QVKIRG RVE G V SVL     P V     L       G P L A
                                                                    Section 13
         (685) 685    690         700        710         720          730      741
    geJ (680) WLKAVDT-TVEEIRARAGAALPPAMVPTSITLVDELPRMVSGKIDRGALIRLAETSP
    geM (485) FVQGPTARDLDRLTALAQDDLPEYMVPSAFIPVDTVPLTGHGKVDEAALAALYEEAT
Consensus (685) FL A      LD I A A   LP  MVPSA   VD LP    GKID AAL   L E A
                                                                    Section 14
         (742) 742    750        760        770        780            798
    geJ (736) APAPVSAQAPAPAPEHAASGGEGLEATVASIFEEVLASGPIGPDADFFDEGGHSLLA
    geM (542) G-----RRGAGPEPATET------EAVLLEIWQKLFAPLRIGLDDDFFDLGGDSLSV
Consensus (742) A       A AP P       EA L  IF  L A    IG D DFFD GG SL
                                                                    Section 15
         (799) 799        810        820        830        842
    geJ (793) ISVLDAIEKRLGVSVDLGDFFLTPTVGEVAELIRKADP-EPGR-
    geM (588) MDLVSGVEERLGVQLDNSAVYTDRTIRSLALTIQNRRNNEAGTS
Consensus (799) I LL AIE RLGV LD    F   TI  LA  I        E G
```

6 References

1. **Drews J**: Drug discovery: a historical perspective. *Science* 2000, 287:1960-1964.

2. **Newman DJ, Cragg GM**: Natural products as sources of new drugs over the last 25 years. *J.Nat.Prod.* 2007, 70:461-477.

3. **Clardy J, Fischbach MA, Walsh CT**: New antibiotics from bacterial natural products. *Nat.Biotechnol.* 2006, 24:1541-1550.

4. **Nathan C**: Antibiotics at the crossroads. *Nature* 2004, 431:899-902.

5. **Hopwood D, Levy S, Wenzel RP, Georgopapadakou N, Baltz RH, Bhavnani S, Cox E**: A call to arms. *Nat.Rev.Drug Discovery* 2007, 6:8-12.

6. **Baker DD, Chu M, Oza U, Rajgarhia V**: The value of natural products to future pharmaceutical discovery. *Nat.Prod.Rep.* 2007, 24:1225-1244.

7. **Demain AL**: Pharmaceutically active secondary metabolites of microorganisms. *Appl Microbiol Biotechnol* 1999, 52:455-463.

8. **Baltz RH**: Genetic manipulation of antibiotic-producing Streptomyces. *Trends Microbiol* 1998, 6:76-83.

9. **Thompson CJ, Fink D, Nguyen LD**: Principles of microbial alchemy: insights from the *Streptomyces coelicolor* genome sequence. *Genome Biol.* 2002, 3:REVIEWS1020.

10. **Kieser T, Bibb M, Buttner MJ, Chater KF, Hopwood DA**: *Practical Streptomyces Genetics*. Norwich, England: The John Innes Foundation; 2000.

11. **Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA**: Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 2002, 417:141-147.

12. **Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, Sakaki Y, Hattori M, Omura S**: Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat.Biotechnol.* 2003, 21:526-531.

13. **Smith T**: Antibiotics from soil bacteria. *Nat.Struct.Biol.* 2000, 7:189-190.

14. Pharmaceutical Biotechnology. Drug Discovery and Clinical Applications. Edited by **Kayser O, Müller RH.** Weinheim: WILEY-VCH; 2008.

15. **Washington JA, Wilson WR**: Erythromycin: a microbial and clinical perspective after 30 years of clinical use (1). *Mayo Clin.Proc.* 1985, 60:189-203.

16. **Nakano T, Miyake K, Endo H, Dairi T, Mizukami T, Katsumata R**: Identification and cloning of the gene involved in the final step of chlortetracycline biosynthesis in *Streptomyces aureofaciens*. *Biosci Biotech Bioch* 2004, 68:1345-1352.

17. **Walczak RJ, Dickens ML, Priestley ND, Strohl WR**: Purification, properties, and characterization of recombinant Streptomyces sp. strain C5 DoxA, a cytochrome P-450 catalyzing multiple steps in doxorubicin biosynthesis. *J.Bacteriol.* 1999, 181:298-304.

18. **Harada K**: Production of secondary metabolites by freshwater cyanobacteria. *Chem Pharm Bull (Tokyo)* 2004, 52:889-899.

19. **Behal V**: Alternative sources of biologically active substances. *Folia Microbiol.* 2003, 48:563-571.

20. **Haefner B**: Drugs from the deep: marine natural products as drug candidates. *Drug Discov Today* 2003, 8:536-544.

21. **Reichenbach H, Höfle G**: Myxobacteria as producers of secondary metabolites. In *Drug Discovery from Nature*. Edited by Grabley S, Thiericke R. Berlin: Springer; 1999:149-179.

22. **Malpartida F, Hopwood DA**: Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host. *Nature* 1984, 309:462.

23. **Weber I, Welzel K, Pelzer S, Vente A, Wohlleben W**: Exploiting the genetic potential of polyketide producing streptomycetes. *J.Biotechnol.* 2003, 106:221-232.

24. **Watve MG, Tickoo R, Jog MM, Bhole BD**: How many antibiotics are produced by the genus *Streptomyces*? *Arch.Microbiol.* 2001, 176:386-390.

25. **Lorenz P, Eck J**: Metagenomics and industrial applications. *Nat.Rev.Microbiol.* 2005, 3:510-516.

26. **Streit WR, Schmitz RA**: Metagenomics - the key to the uncultured microbes. *Curr.Opin.Microbiol.* 2004, 7:492-498.

27. **Bormann C, Mohrle V, Bruntner C**: Cloning and heterologous expression of the entire set of structural genes for nikkomycin synthesis from *Streptomyces tendae* Tu901 in *Streptomyces lividans*. *J.Bacteriol.* 1996, 178:1216-1218.

28. **Jung WS, Lee SK, Hong JSJ, Park SR, Jeong SJ, Han AR, Sohng JK, Kim BG, Choi CY, Sherman DH, Yoon YJ**: Heterologous expression of tylosin polyketide synthase and production of a hybrid bioactive macrolide in *Streptomyces venezuelae*. *Appl.Microbiol.Biotechnol.* 2006, 72:763-769.

29. T**hapa LP, Oh TJ, Lee HC, Liou K, Park JW, Yoon YJ, Sohng JK**: Heterologous expression of the kanamycin biosynthetic gene cluster (pSKC2) in *Streptomyces venezuelae* YJ003. *Appl.Microbiol.Biotechnol.* 2007,DOI 10.1007/s00253-007-1096-4.

30. **Wenzel SC, Müller R:** Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Curr.Opin.Biotechnol.* 2005, 16:594-606.

31. I**chinose K, Ozawa M, Itou K, Kunieda K, Ebizuka Y**: Cloning, sequencing and heterologous expression of the medermycin biosynthetic gene cluster of Streptomyces sp. AM-7161: Towards comparative analysis of the benzoisochromanequinone gene clusters. *Microbiology* 2003, 149:1633-1645.

32. **Bai L, Li L, Xu H, Minagawa K, Yu Y, Zhang Y, Zhou X, Floss HG, Mahmud T, Deng Z**: Functional analysis of the validamycin biosynthetic gene cluster and engineered production of validoxylamine A. *Chem.Biol.* 2006, 13:387-397.

33. **Xu Z, Jakobi K, Welzel K, Hertweck C**: Biosynthesis of the antitumor agent chartreusin involves the oxidative rearrangement of an anthracyclic polyketide. *Chem.Biol.* 2005, 12:579-588.

34. **Tang L, Shah S, Chung L, Carney J, Katz L, Khosla C, Julien B**: Cloning and heterologous expression of the epothilone gene cluster. *Science* 2000, 287:640-642.

35. **Wenzel SC, Gross F, Zhang Y, Fu J, Stewart FA, Müller R**: Heterologous expression of a myxobacterial natural products assembly line in pseudomonads via Red/ET recombineering. *Chem.Biol.* 2005, 12:349-356.

36. **Muyrers JPP, Zhang Y, Stewart AF**: Techniques: Recombinogenic engineering - new options for cloning and manipulating DNA. *Trends Biochem.Sci.* 2001, 26:325-331.

37. **Zhang Y, Buchholz F, Muyrers JPP, Stewart FA**: A new logic for DNA engineering using recombination in *Escherichia coli. Nat.Genet.* 1998, 20:123-128.

38. **Zhang Y, Muyrers JPP, Testa G, Stewart AF**: DNA cloning by homologous recombination in *Escherichia coli. Nat.Biotechnol.* 2000, 18:1314-1317.

39. **Wenzel SC, Müller R**: Heterologe Expression von komplexen myxobakteriellen Naturstoff-Biosynthesewegen in Pseudomonaden. *Biospektrum* 2005, 11:628-631.

40. **Wenzel SC**: Biosynthesis and heterologous production of myxobacterial secondary metabolites. 48. 2005. Braunschweig, Technische Carolo-Wilhemina-Universität, Fachbereich Biowissenschaftern und Psychologie.

41. **Muyrers JPP, Zhang Y, Benes V, Testa G, Ansorge W, Stewart AF**: Point mutation of bacterial artificial chromosomes by ET recombination. *EMBO Rep.* 2000, 1:239-243.

42. **Perlova O, Fu J, Kuhlmann S, Krug D, Stewart F, Zhang Y, Müller R**: Reconstitution of myxothiazol biosynthetic gene luster by Red/ET recombination and heterologous expression in *Myxococcus xanthus. Appl.Environ.Microbiol.* 2006, 72:7485-7494.

43. **Gross F, Ring MW, Perlova O, Fu J, Schneider S, Gerth K, Kuhlmann S, Stewart AF, Zhang Y, Müller R**: Metabolic engineering of *Pseudomonas putida* for methylmalonyl-CoA biosynthesis to enable complex heterologous secondary metabolite formation. *Chem Biol* 2006, 13:1253-1264.

44. **Hu YF, Phelan VV, Farnet CM, Zazopoulos E, Bachmann BO**: Reassembly of anthramycin biosynthetic gene cluster by using recombinogenic cassettes. *ChemBioChem* 2008, 9:1603-1608.

45. **Fu J, Wenzel SC, Perlova O, Gross F, Tang Z, Yin Y, Stewart AF, Müller R, Zhang Y**: Efficient transfer of two large secondary metabolite pathways into heterologous hosts by transposition. *Nucleic Acids Res.* 2008, 36:e113.

46. **Staunton J, Weissman KJ**: Polyketide biosynthesis: a millennium review. *Nat.Prod.Rep.* 2001, 18:380-416.

47. **Finking R, Marahiel MA**: Biosynthesis of nonribosomal peptides. *Annu Rev Microbiol* 2004, 58:453-488.

48. **Walsh CT, Gehring AM, Weinreb PH, Quadri LE, Flugel RS**: Post-translational modification of polyketide and nonribosomal peptide synthases. *Curr.Opin.Chem.Biol.* 1997, 1:309-315.

49. **Hong H, Demangel C, Pidot SJ, Leadlay PF, Stinear T**: Mycolactones: immunosuppressive and cytotoxic polyketides produced by aquatic mycobacteria. *Nat.Prod.Rep.* 2008, 25:447-454.

50. **Schwarzer D, Marahiel MA**: Multimodular biocatalysts for natural product assembly. *Naturwissenschaften* 2001, 88:93-101.

51. **Mootz HD, Finking R, Marahiel MA**: 4 '-phosphopantetheine transfer in primary and secondary metabolism of *Bacillus subtilis. J.Biol.Chem.* 2001, 276:37289-37298.

52. **Frueh DP, Arthanari H, Koglin A, Vosburg DA, Bennett AE, Walsh CT, Wagner G**: Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature* 2008, 454:903-906.

53. **Koglin A, Lohr F, Bernhard F, Rogov VV, Frueh DP, Strieter ER, Mofid MR, Guntert P, Wagner G, Walsh CT, Marahiel MA, Dotsch V**: Structural basis for the selectivity of the external thioesterase of the surfactin synthetase. *Nature* 2008, 454:907-U68.

54. **Walsh CT, Chen HW, Keating TA, Hubbard BK, Losey HC, Luo LS, Marshall CG, Miller DA, Patel HM**: Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr.Opin.Chem.Biol.* 2001, 5:525-534.

55. **Samel SA, Marahiel MA, Essen LO**: How to tailor non-ribosomal peptide products - new clues about the structures and mechanisms of modifying enzymes. *Molecular Biosystems* 2008, 4:387-393.

56. **Mootz HD, Schwarzer D, Marahiel MA**: Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *ChemBioChem* 2002, 3:490-504.

57. **Mootz HD, Marahiel MA**: Design and application of multimodular peptide synthetases. *Curr.Opin.Biotechnol.* 1999, 10:341-348.

58. **Mootz HD, Marahiel MA**: The tyrocidine biosynthesis operon of Bacillus brevis: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J.Bacteriol.* 1997, 179:6843-6850.

59. **Lee YK, Yoon BD, Yoon JH, Lee SG, Song JJ, Kim JG, Oh HM, Kim HS**: Cloning of srfA operon from Bacillus subtilis C9 and its expression in E-coli. *Appl.Microbiol.Biotechnol.* 2007, 75:567-572.

60. **De Crecy-Lagard V, Blanc V, Gil P, Naudin L, Lorenzon S, Famechon A, Bamas-Jacques N, Crouzet J, Thibaut D**: Pristinamycin I biosynthesis in Streptomyces pristinaespiralis: molecular characterization of the first two structural peptide synthetase genes. *J.Bacteriol.* 1997, 179:705-713.

61. **Zhou Z, Lai JR, Walsh CT**: Interdomain communication between the thiolation and thioesterase domains of EntF explored by combinatorial mutagenesis and selection. *Chem.Biol.* 2006, 13:869-879.

62. **May JJ, Wendrich TM, Marahiel MA**: The dhb operon of Bacillus subtilis encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin. *J.Biol.Chem.* 2001, 276:7209-7217.

63. **Kohli RM, Trauger JW, Schwarzer D, Marahiel MA, Walsh CT**: Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases. *Biochemistry-US* 2001, 40:7099-7108.

64. **Wenzel SC, Müller R**: Formation of novel secondary metabolites by bacterial multimodular assembly lines: deviations from text book biosynthetic logic. *Curr.Opin.Chem.Biol.* 2005, 9:447-458.

65. **Keating TA, Suo Z, Ehmann DE, Walsh CT**: Selectivity of the yersiniabactin synthetase adenylation domain in the two-step process of amino acid activation and transfer to a holo-carrier protein domain. *Biochemistry-US* 2000, 39:2297-2306.

66. **Dürr C, Schnell HJ, Luzhetskyy A, Murillo R, Weber M, Welzel K, Vente A, Bechthold A**: Biosynthesis of the terpene phenalinolactone in Streptomyces sp. Tu6071: Analysis of the gene cluster and generation of derivatives. *Chem.Biol.* 2006, 13:365-377.

67. **Brandi L, Lazzarini A, Cavaletti L, Abbondi M, Corti E, Ciciliato I, Gastaldo L, Marazzi A, Feroggio M, Fabbretti A, Maio A, Colombo L, Donadio S, Marinelli F, Losi D, Gualerzi CO, Selva E**: Novel tetrapeptide inhibitors of bacterial protein synthesis produced by a Streptomyces sp. *Biochemistry-US* 2006, 45:3692-3702.

68. **Grant SG, Jessee J, Bloom FR, Hanahan D**: Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *P.Natl.Acad.Sci.USA* 1990, 87:4645-4649.

69. **Studier FW**: Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J.Mol.Biol.* 1991, 219:37-44.

70. **Jerpseth B, Greener A, Short JM, Viola J, Kretz PL**: XL1-Blue MRF′ *E.coli* cells: MrcA-, MrcB-,Mrr-,HsdR-derivative of XL1-Blue cells. *Strategies* 1992, 5:81-83.

71. **Ramos-Gonzalez MI, Molin S**: Cloning, sequencing, and phenotypic characterization of the rpoS gene from Pseudomonas putida KT2440. *J.Bacteriol.* 1998, 180:3421-3431.

72. **Yin XH, Gerbaud C, Francou FX, Guerineau M, Virolle MJ:** *amlC*: Another amylolytic gene maps close to the *amlB* locus in *Streptomyces lividans* TK24. *Gene* 1998, 215:171-180.

73. **Floriano B, Bibb M**: afsR is a pleiotropic but conditionally required regulatory gene for antibiotic production in Streptomyces coelicolor A3(2). *Mol.Microbiol.* 1996, 21:385-396.

74. **Kawasaki T, Kuzuyama T, Furihata K, Itoh N, Seto H, Dairi T**: A relationship between the mevalonate pathway and isoprenoid production in actinomycetes. *J.Antibiot.* 2003, 56:957-966.

75. **Brandi L, Fabbretti A, La Teana A, Abbondi M, Losi D, Donadio S, Gualerzi CO**: Specific, efficient, and selective inhibition of prokaryotic translation initiation by a novel peptide antibiotic. *P.Natl.Acad.Sci.USA* 2006, 103:39-44.

76. **Hill DS, Stein JI, Torkewitz NR, Morse AM, Howell CR, Pachlatko JP, Becker JO, Ligon JM**: Cloning of genes involved in the synthesis of pyrrolnitrin from *Pseudomonas fluorescens* and role of pyrrolnitrin synthesis in biological control of plant disease. *Appl.Environ.Microbiol.* 1994, 60:78-85.

77. **Sambrook J, Fritsch EF, Maniatis T**: *Molecular cloning: A laboratory manual*, edn 2. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1989.

78. **Laemmli UK**: Cleavage of structural proteins during the assembly of the head of bacteriophage T7. *Nature* 1970, 227:680-685.

79. **Martinez A, Kolvek SJ, Yip CLT, Hopke J, Brown KA, MacNeil IA, Osburne MS**: Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Appl.Environ.Microbiol.* 2004, 70:2452-2463.

80. **Dürr C**: Untersuchungen zur Phenalinolactonbiosynthese sowie Charakterisierung der ungewöhnlichen Glycosyltransferasen UrdGT2 und AviGT4. 1-181. 2005. Fakultät für Chemie, Pharmazie und Geowissenschaften der Albert-Ludwigs-Universität Freiburg im Breisgau

81. **Singh D, Seo MJ, Kwon HJ, Rajkarnikar A, Kim KR, Kim SO, Suh JW**: Genetic localization and heterologous expression of validamycin biosynthetic gene cluster isolated from *Streptomyces hygroscopicus* var. *limoneus* KCCM 11405 (IFO 12704). *Gene* 2006, 376:13-23.

82. **Gatto GJ, Boyne MT, Kelleher NL, Walsh CT**: Biosynthesis of pipecolic acid by RapL, a lysine cyclodeaminase encoded in the rapamycin gene cluster. *J.Am.Chem.Soc.* 2006, 128:3838-3847.

83. **Müller C, Nolden S, Gebhardt P, Heinzelmann E, Lange C, Puk O, Welzel K, Wohlleben W, Schwartz D**: Sequencing and analysis of the biosynthetic gene cluster of the lipopeptide antibiotic Friulimicin in *Actinoplanes friuliensis. Antimicrob.Agents Chemother.* 2007, 51:1028-1037.

84. **Pulsawat N, Kitani S, Nihira T**: Characterization of biosynthetic gene cluster for the production of virginiamycin M, a streptogramin type A antibiotic, in *Streptomyces virginiae. Gene* 2007, 393:31-42.

85. **Sandmann A, Sasse F, Müller R**: Identification and analysis of the core biosynthetic machinery of tubulysin, a potent cytotoxin with potential anticancer activity. *Chem.Biol.* 2004, 11:1071-1079.

86. **He M**: Pipecolic acid in microbes: Biosynthetic routes and enzymes. *J.Ind.Microbiol.Biotechnol.* 2006,1-7.

87. **Khaw LE, Bohm GA, Metcalfe S, Staunton J, Leadlay PF**: Mutational Biosynthesis of Novel Rapamycins by a Strain of Streptomyces hygroscopicus NRRL 5491 Disrupted in rapL, Encoding a Putative Lysine Cyclodeaminase. *J.Bacteriol.* 1998, 180:809-814.

88. **Rachid S, Krug D, Kochems I, Kunze B, Scharfe M, Blöcker H, Zabriski M, Müller R**: Molecular and biochemical studies of chondramide formation - highly cytotoxic natural products from *Chondromyces crocatus* Cm c5. *Chem.Biol.* 2006, 14:667-681.

89. **Kopp M**: Identifizierung und Charakterisierung von Biosynthesegenclustern aus Sorangium cellulosum. 227. 2005. Saarbrücken, Universität des Saarlandes, Naturwissenschaftlich-Technische Fakultät.

90. **Buntin K, Rachid S, Scharfe M, Blöcker H, Weissman KJ, Müller R**: Production of the highly antifungal isochromanone ajudazols in *Chondromyces crocatus* Cm c5: biosynthetic machinery and cytochrome P450 tailoring modifications. *Angew.Chem.Int.Ed.* 2008, 47:4595-4599.

91. **Marahiel MA, Stachelhaus T, Mootz HD**: Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem Rev* 1997, 97:2651-2674.

92. **Ishikawa J, Hotta K**: FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G + C content. *FEMS Microbiol.Lett.* 1999, 174:251-253.

93. **Stachelhaus T, Mootz HD, Marahiel MA**: The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem.Biol.* 1999, 6:493-505.

94. **Challis GL, Ravel J, Townsend CA**: Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem.Biol.* 2000, 7:211-224.

95. **Stachelhaus T, Huser A, Marahiel MA**: Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. *Chem.Biol.* 1996, 3:913-921.

96. **Rose TM, Schultz ER, Henikoff JG, Pietrokovski S, McCallum CM, Henikoff S**: Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res.* 1998, 26:1628-1635.

97. **Ebersbach G, Gerdes K**: Bacterial mitosis: partitioning protein ParA oscillates in spiral-shaped structures and positions plasmids at mid-cell. *Mol.Microbiol.* 2004, 52:385-398.

98. **Parsons JA, Bannam TL, Devenish RJ, Rood JI**: TcpA, an FtsK/SpoIIIE homolog, is essential for transfer of the conjugative plasmid pCW3 in *Clostridium petfringens*. *J.Bacteriol.* 2007, 189:7782-7790.

99. **Schwarzer D, Mootz HD, Linne U, Marahiel MA**: Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *P.Natl.Acad.Sci.USA* 2002, 99:14083-14088.

100. **Weissman KJ, Müller R**: Protein-protein interactions in multienzyme megasynthetases. *ChemBioChem* 2008, 9:826-848.

101. **Mendez C, Salas JA**: ABC transporters in antibiotic-producing actinomycetes. *FEMS Microbiol.Lett.* 1998, 158:1-8.

102. **Aparicio JF, Molnar I, Schwecke T, Konig A, Haydock SF, Khaw LE, Staunton J, Leadlay PF**: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hygroscopicus: analysis of the enzymatic domains in the modular polyketide synthase. *Gene* 1996, 169:9-16.

103. **Turgay K, Krause M, Marahiel MA**: Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Mol.Microbiol.* 1992, 6:529-546.

104. **H**ojati Z, Milne C, Harvey B, Gordon L, Borg M, Flett F, Wilkinson B, Sidebottom P, Rudd B, Hayes M, Smith C, Micklefield J**: Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from Streptomyces coelicolor. *Chem.Biol.* 2002, 9:1175-1187.

105. **Yeh E, Kohli RM, Bruner SD, Walsh CT**: Type II Thioesterase Restores Activity of a NRPS Module Stalled with an Aminoacyl-S-enzyme that Cannot Be Elongated. *ChemBioChem* 2004, 5:1290-1293.

106. **Hahn M, Stachelhaus T**: Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. *P.Natl.Acad.Sci.USA* 2006, 103:275-280.

107. **Hahn M, Stachelhaus T**: Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. *P.Natl.Acad.Sci.USA* 2004, 101:15585-15590.

108. **Tsotsou GE, Barbirato F**: Biochemical characterisation of recombinant *Streptomyces pristinaespiralis* L-lysine cyclodeaminase. *Biochimie* 2007, 89:591-604.

109. **Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T**: Chaperone coexpression plasmids: Differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2 in *Escherichia coli. Appl.Environ.Microbiol.* 1998, 64:1694-1699.

110. **Nishihara K, Kanemori M, Yanagi H, Yura T**: Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli. Appl.Environ.Microbiol.* 2000, 66:884-889.

111. **Thomas MG, Burkart MD, Walsh CT**: Conversion of L-proline to pyrrolyl-2-carboxyl-S-PCP during undecylprodigiosin and pyoluteorin biosynthesis. *Chem.Biol.* 2002, 9:171-184.

112. **Rude MA, Khosla C**: Engineered biosynthesis of polyketides in heterologous hosts. *Chem Eng Sci* 2004, 59:4693-4701.

113. **Gustafsson C, Govindarajan S, Minshull J**: Codon bias and heterologous protein expression. *Trends Biotechnol.* 2004, 22:346-353.

114. **Kim CY, Park HJ, Kim ES**: Heterologous expression of hybrid type II polyketide synthase system in Streptomyces species. *J.Microbiol.Biotechnol.* 2003, 13:819-822.

115. **von Mulert U, Luzhetskyy A, Hofmann C, Mayer A, Bechthold A**: Expression of the landomycin biosynthetic gene cluster in a PKS mutant of Streptomyces fradiae is dependent on the coexpression of a putative transcriptional activator gene. *FEMS Microbiol.Lett.* 2004, 230:91-97.

116. **Luzhetskyy A, Mayer A, Hoffmann J, Pelzer S, Holzenkamper M, Schmitt B, Wohlert SE, Vente A, Bechthold A**: Cloning and heterologous expression of the aranciamycin biosynthetic gene cluster revealed a new flexible glycosyltransferase. *ChemBioChem* 2007, 8:599-602.

117. **Li A, Piel J**: A gene cluster from a marine Streptomyces encoding the biosynthesis of the aromatic spiroketal polyketide griseorhodin A. *Chem.Biol.* 2002, 9:1017-1026.

118. **Miao V, Coeffet-LeGal MF, Brian P, Brost R, Penn J, Whiting A, Martin S, Ford R, Parr I, Bouchard M, Silva CJ, Wrigley SK, Baltz RH**: Daptomycin biosynthesis in Streptomyces roseosporus: Cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology* 2005, 151:1507-1523.

119. **Xue Q, Ashley G, Hutchinson CR, Santi DV:** A multiplasmid approach to preparing large libraries of polyketides. *P.Natl.Acad.Sci.USA* 1999, 96:11740-11745.

120. **Zirkle R, Ligon JM, Molnar I**: Heterologous production of the antifungal polyketide antibiotic soraphen A of Sorangium cellulosum So ce26 in Streptomyces lividans. *Microbiology* 2004, 150:2761-2774.

121. **Pfeifer BA, Wang CCC, Walsh CT, Khosla C**: Biosynthesis of yersiniabactin, a complex polyketide-nonribosomal peptide, using Escherichia coli as a heterologous host. *Appl.Environ.Microbiol.* 2003, 69:6698-6702.

122. **Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C**: Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. *Science* 2001, 291:1790-1792.

123. **Eustaquio AS, Gust B, Galm U, Li SM, Chater KF, Heide L**: Heterologous expression of novobiocin and clorobiocin biosynthetic gene clusters. *Appl.Environ.Microbiol.* 2005, 71:2452-2459.

124. **Wolpert M, Heide L, Kammerer B, Gust B:** Assembly and heterologous expression of the coumermycin A(1) gene cluster and production of new derivatives by genetic engineering. *ChemBioChem* 2008, 9:603-612.

125. **Gross F, Luniak N, Perlova O, Gaitatzis N, Jenke-Kodama H, Gerth K, Gottschalk D, Dittmann E, Müller R**: Bacterial type III polyketide synthases: Phylogenetic analysis and potential for the production of novel secondary metabolites by heterologous expession in pseudomonads. *Arch.Microbiol.* 2006, 185:28-38.

126. **Kuzuyama T**: Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units. *Biosci Biotech Bioch* 2002, 66:1619-1627.

127. **Freitag A, Rapp H, Heide L, Li SM**: Metabolic engineering of aminocoumarins: Inactivation of the methyltransferase gene cloP and generation of new clorobiocin derivatives in a heterologous host. *ChemBioChem* 2005, 6:1411-1418.

128. **Kuzuyama T, Dairi T, Yamashita H, Shoji Y, Seto H**: Heterologous mevalonate production in *Streptomyces lividans* TK23. *Biosci Biotech Bioch* 2004, 68:931-934.

129. **Cane DE, Chow C, Lillo A, Kang I**: Molecular cloning, expression and characterization of the first three genes in the mevalonate-independent isoprenoid pathway in *Streptomyces coelicolor. Bioorganic & Medicinal Chemistry* 2001, 9:1467-1477.

130. **Singh D, Kwon HJ, Rajkarnikar A, Suh JW**: Glucoamylase gene, vldI, is linked to validamycin biosynthesis in *Streptomyces hygroscopicus* var. *limoneus*, and *vldADEFG* confers validamycin production in *Streptomyces lividans*, revealing the role of VldE in glucose attachment. *Gene* 2007, 395:151-159.

131. **Wenzel SC, Bode HB, Kochems I, Müller R**: A type I/type III polyketide synthase hybrid biosynthetic pathway for the structurally unique *ansa* compound kendomycin. *ChemBioChem* 2008, 9:2711-2721.

132. **Heide L, Gust B, Anderle C, Li SM**: Combinatorial biosynthesis, metabolic engineering and mutasynthesis for the generation of new aminocoumarin antibiotics. *Current Topics in Medicinal Chemistry* 2008, 8:667-679.

133. **Walsh C:** *Antibiotics: Actions, Origins, Resistance*. Washington, D.C.: ASM Press; 2003.

134. **Shen B, Du LC, Sanchez C, Edwards DJ, Chen M, Murrell JM**: Cloning and characterization of the bleomycin biosynthetic gene cluster from Streptomyces verticillus ATCC15003. *J.Nat.Prod.* 2002, 65:422-431.

135. **Wenzel SC, Kunze B, Höfle G, Silakowski B, Scharfe M, Blöcker H, Müller R**: Structure and biosynthesis of myxochromides S1-3 in *Stigmatella aurantiaca*: Evidence for an iterative bacterial type I polyketide synthase and for module skipping in nonribosomal peptide biosynthesis. *ChemBioChem* 2005, 6:375-385.

136. **Gaitatzis N, Silakowski B, Kunze B, Nordsiek G, Blöcker H, Höfle G, Müller R**: The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. *J.Biol.Chem.* 2002, 277:13082-13090.

137. **Silakowski B, Kunze B, Müller R**: Multiple hybrid polyketide synthase/non-ribosomal peptide synthetase gene clusters in the myxobacterium *Stigmatella aurantiaca. Gene* 2001, 275:233-240.

138. **Wenzel SC, Meiser P, Binz T, Mahmud T, Müller R**: Nonribosomal peptide biosynthesis: Point mutations and module skipping lead to chemical diversity. *Angew.Chem.Int.Ed.* 2006, 45:2296-2301.

139. **Oja T, Palmu K, Lehmussola H, Lepparanta O, Hannikainen K, Niemi J, Mantsala P, Metsa-Ketela M:** Characterization of the alnumycin gene cluster reveals unusual gene products for pyran ring formation and dioxan biosynthesis. *Chem.Biol.* 2008, 15:1046-1057.

140. **Binz TM, Wenzel SC, Schnell HJ, Bechthold A, Müller R**: Heterologous expression and genetic engineering of the phenalinolactone biosynthetic gene cluster by using Red/ET recombineering. *ChemBioChem* 2008, 9:447-454.

141. **Sieber SA, Marahiel MA**: Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem Rev* 2005, 105:715-738.

142. **Kopp M, Irschik H, Gross F, Perlova O, Sandmann A, Gerth K, Müller R**: Critical variations of conjugational DNA transfer into secondary metabolite multiproducing *Sorangium cellulosum* strains So ce12 and So ce56: development of a mariner-based transposon mutagenesis system. *J.Biotechnol.* 2004, 107:29-40.

143. **Gaitatzis N, Kunze B, Müller R**: In vitro reconstitution of the myxochelin biosynthetic machinery of *Stigmatella aurantiaca* Sg a15: Biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases. *P.Natl.Acad.Sci.USA* 2001, 98:11136-11141.

144. **Miller DA, Luo LS, Hillson N, Keating TA, Walsh CT**: Yersiniabactin synthetase: A four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of Yersinia pestis. *Chem.Biol.* 2002, 9:333-344.

145. **Galm U, Wendt-Pienkowski E, Wang L, George NP, Oh TJ, Yi F, Tao M, Coughlin JM, Shen B**: The biosynthetic gene cluster of zorbamycin, a member of the bleomycin family of antitumor antibiotics, from *Streptomyces flavoviridis* ATCC 21892. *Mol.Biosyst.* 2009, 5:77-90.

146. **Vaillancourt FH, Vosburg DA, Walsh CT**: Dichlorination and bromination of a threonyl-S-carrier protein by the non-heme Fe(II) halogenase SyrB2. *ChemBioChem* 2006, 7:748-752.

147. **Yin X, Zabriskie TM**: VioC is a Non-Heme Iron, alpha-Ketoglutarate-Dependent Oxygenase that Catalyzes the Formation of 3S-Hydroxy-L-Arginine during Viomycin Biosynthesis. *ChemBioChem* 2004, 5:1274.

148. **Chen HW, Thomas MG, O'Connor SE, Hubbard BK, Burkart MD, Walsh CT**: Aminoacyl-S-enzyme intermediates in beta-hydroxylations and alpha,beta-desaturations of amino acids in peptide antibiotics. *Biochemistry-US* 2001, 40:11651-11659.

149. **Steffensky M, Muhlenweg A, Wang ZX, Li SM, Heide L**: Identification of the novobiocin biosynthetic gene cluster of Streptomyces spheroides NCIB 11891. *Antimicrob.Agents Chemother.* 2000, 44:1214-1222.

150. **Yu TW, Bai L, Clade D, Hoffmann D, Toelzer S, Trinh KQ, Xu J, Moss SJ, Leistner E, Floss HG**: The biosynthetic gene cluster of the maytansinoid antitumor agent ansamitocin from Actinosynnema pretiosum. *P.Natl.Acad.Sci.USA* 2002, 99:7968-7973.

151. **Neumann CS, Fujimori DG, Walsh CT**: Halogenation strategies in natural product biosynthesis. *Chemistry & Biology* 2008, 15:99-109.

152. **Eustaquio AS, Gust B, Luft T, Li SM, Chater KF, Heide L**: Clorobiocin Biosynthesis in Streptomyces: Identification of the Halogenase and Generation of Structural Analogs. *Chem.Biol.* 2003, 10:279-288.

153. **Yeh E, Blasiak LC, Koglin A, Drennan CL, Walsh CT**: Chlorination by a long-lived intermediate in the mechanism of flavin-dependent halogenases. *Biochemistry-US* 2007, 46:1284-1292.

154. **Yeh E, Garneau S, Walsh CT**: Robust in vitro activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. *P.Natl.Acad.Sci.USA* 2005, 102:3960-3965.

155. **Vaillancourt FH, Yin J, Walsh CT**: SyrB2 in syringomycin E biosynthesis is a nonheme FeII {alpha}-ketoglutarate- and O2-dependent halogenase. *P.Natl.Acad.Sci.USA* 2005, 102:10111-10116.

156. **Walsh CT**: Natural insights for chemical biologists. *Nat Chem Biol Nat Chem Biol* 2005, 1:122-124.

157. **Singh GM, Vaillancourt FH, Yin J, Walsh CT**: Characterization of SyrC, an aminoacyltransferase shuttling threonyl and chlorothreonyl residues in the syringomycin biosynthetic assembly line. *Chem.Biol.* 2007, 14:31-40.

158. **Guenzi E, Galli G, Grgurina I, Pace E, Ferranti P, Grandi G**: Coordinate transcription and physical linkage of domains in surfactin synthetase are not essential for proper assembly and activity of the multienzyme complex. *J.Biol.Chem.* 1998, 273:14403-14410.

159. **Strieter ER, Vaillancourt FH, Walsh CT**: CmaE: A transferase shuttling aminoacyl groups between carrier protein domains in the coronamic acid biosynthetic pathway. *Biochemistry-US* 2007, 46:7549-7557.

160. **Chang ZX, Flatt P, Gerwick WH, Nguyen VA, Willis CL, Sherman DH**: The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit. *Gene* 2002, 296:235-247.

161. **Garneau-Tsodikova S, Stapon A, Kahne D, Walsh CT**: Installation of the pyrrolyl-2-carboxyl pharmacophore by CouN1 and CouN7 in the late biosynthetic steps of the aminocoumarin antibiotics clorobiocin and coumermycin A(1). *Biochemistry-US* 2006, 45:8568-8578.

Lebenslauf/Curriculum vitae

Tina Maria Binz
Diplom-Biologin
mit Schwerpunkt Human-und Molekularbiologie
* 27.09.1979 in Saarlouis, Deutschland

Am Homburg 59
66123 Saarbrücken

Schulausbildung:

1986-1990	Grundschule Bachem
1990-1999	Gymnasium am Stefansberg, Merzig, Abitur

Hochschulstudium:

10/1999-12/2004	Studium der Human-und Molekularbiologie an der Universität des Saarlandes, Saarbrücken
10/2001	Diplomvorprüfung
09/2002-10/2002	Forschungspraktikum im Rahmen des Studiums, am Institut für Technische Biochemie unter der Leitung von Prof. Dr. Ing. E. Heinzle, Universität des Saarlandes Titel: „Analytische Methoden zur Quantifizierung intrazellulärer Metabolite in Bakterien"
08/2003-11/2003	Forschungspraktikum im Rahmen des ERASMUS Programms an der „University of Edinburgh", Schottland, am Institut für Reproduktive Biologie unter der Leitung von Prof. Dr. Ian Mason Titel: „Construction and characterization of fusion proteins containing human type2 3β-hydroxysteroid dehydrogenase and human type1 11β-hydroxysteroid dehydrogenase"
02/2002	Diplom Prüfung
04/2004-12/2004	Diplomarbeit unter der Leitung von Prof. Dr. Ing. E. Heinzle am Institut für Technische Biochemie, Universität des Saarlandes Titel: „Physiologie von *Corynebacterium glutamicum*- Einfluss des Nährmediums auf die Lysin Produktion"
01/2005	Praktikum unter der Leitung von Prof. Dr. Rolf Müller am Institut für Pharmazeutische Biotechnologie, Universität des Saarlandes
02/2005	Beginn der Promotion unter der Leitung von Prof. Dr. Rolf Müller, Titel: "Studies on the biosynthesis and heterologous expression of complex secondary metabolites from streptomycetes"