Studies on the biosynthesis and heterologous expression of tubulysin

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

Yi Chai

Saarbrücken

2011

Tag des Kolloquiums: Nov 21, 2011

Dekan: Prof. Dr. William F. Maier

Berichterstatter: Prof. Dr. Rolf Müller

Prof. Dr. Uli Kazmaier

Vorsitz: Prof. Dr. Alexandra K. Kiemer

Akad. Mitarbeiter: Dr. Angelika Ullrich

Dedicated to the wonderful Deutschland and the people inside.

Thank you for five amazing years!

ACKNOWLEDGEMENTS

This thesis would not have been possible without the help and support of many individuals. First of all, I would like to thank my supervisor Prof. Dr. Rolf Müller. I am extremely thankful to Prof. Müller who gave me the opportunity to work in the truly remarkable research environment in the Department of Pharmaceutical Biotechnology at Saarland University, and supported me throughout my thesis work. I learned so much from his research ability and attitude which will benefit my whole research career in my life.

I owe special thanks to my co-advisor Prof. Dr. Kira J. Weissman. She is a wonderful researcher and mentor, who provided me so many advices and guidance for my research. Her support, motivation and kindness greatly helped me go through the tough research process.

I also want to thank Dr. Axel Sandmann, who was my mentor in my first year of Ph.D. He led me into the pharmaceutical biotechnology research world, and also guided me, shared his experiences with me during the beginning of my Ph.D. studies.

I want to especially thank Dr. Youming Zhang from Gene Bridges GmbH. He provided me so many advices and guidance for both my research and career. I appreciate his help a lot. And also gracious thanks to the whole research group of Gene Bridges GmbH at Saarbrücken, especially Shiping Shan who is co-working with me.

At the same time I would like to thank Prof. Dr. Uli Kazmaier, and Dr. Angelika Ullrich from Department of Organic Chemistry. Over the past few years, we cooperated for many projects, and had very pleasant and successful team works. Besides, I also want to thank Prof. Kazmaier for gladly consenting to review my thesis.

Gracious thanks for all my collaborators in Prof. Müller's research group. I am grateful for all those encouragements, discussions, assistance, and support. And I was really fortunate to have so many great times and memorable moments with them. I specially wish to thank Dr. Silke Wenzel, Dr. Daniel Krug, Dr. Shwan Rachid, Dr. Kathrin Buntin, Dr. Dominik Pistorius, Dr. Ronald Garcia, Eva Luxenburger and Ole Revermann for their unremitting assistance during my study, and also valuable answers to all my questions. I also own thanks to Dr. Yanyan Li, Dr. Nora

Luniak and all my other Chinese colleagues for all their support in various ways during these years as both collaborators and friends. I usually have lots of administrative processing stuffs to bother our secretary Ms. Birgitta Lelarge. I own gracious thanks for her kind and generous support for not only my work stuffs but also my living in Germany.

Last but not least my heartfelt thanks to my husband Dr. Zhao Dong and my parents for their support and sacrifices throughout all the years of this thesis. I want to especially mention my 6-month old daughter Emily Y. Dong. She is my angle and her lovely smile is always the best encouragement for me to fulfill my thesis writing.

PUBLICATIONS

Ullrich,A., **Chai,Y.**, Pistorius,D., Elnakady,Y.A., Herrmann,J.E., Weissman,K.J., Kazmaier,U., and Müller,R. (2009). Pretubulysin, a potent and chemically-accessible tubulysin precursor from *Angiococcus disciformis*. Angew. Chem Int. Ed Engl. 48, 4422-4425. (The first two authors contributed equally to this work)

Chai,Y., Pistorius,D., Ullrich,A., Weissman,K.J., Kazmaier,U., and Müller,R. (2010). Discovery of 23 natural tubulysins from Angiococcus disciformis An d48 and Cystobacter SBCb004. Chem. Biol. 17, 296-309.

Chai,Y., Shan,S., Weissman,K.J., Hu,S., Zhang,Y., Müller,R. (2011). Heterologous expression and genetic engineering of the tubulysin biosynthetic gene cluster using Red/ET recombineering and inactivation mutagenesis. Chem. Biol. *submitted*.

CONFERENCE CONTRIBUTIONS

Chai,Y., Pistorius,D., Ullrich,A., Weissman,K.J., Kazmaier,U., Zhang,Y., and Müller,R. (September 2010). More natural tubulysins and heterologous expression of the tubulysin biosynthetic gene cluster. International Workshop: Biology of Bacterial Producers of Natural Compounds, Tübingen, Germany (oral presentation).

Chai,Y., Ullrich,A., Kazmaier,U., Weissman,K.J., and Müller,R. (September 2008). Pretubulysin – a new metabolite from *Angiococcus disciformis* An d48. International Workshop: Biology and Chemistry of Antibiotic-Producing Bacteria, Berlin, Germany (poster).

Chai,Y., Sandmann,A., Ullrich,A., Kazmaier,U., Weissman,K.J., and Müller,R. (October 2007). Mutation and expression analysis of three genes in the tubulysin biosynthetic gene cluster. International workshop: Biology of Bacteria Producing Natural Products, Saarbrücken, Germany (poster).

ABSTRACT

The tubulysins are a family of nine complex peptides with promising cytotoxic activity against multi drug-resistant tumors. This thesis reports a third tubulysin producing strain, *Cystobacter* sp. SBCb004. Although the tubulysin biosynthetic gene clusters appear to be incomplete in both SBCb004 and An d48, comparison of the clusters reveals a conserved 11-gene architecture, which allowed the assignment of cluster boundaries. Knockout mutagenesis and *in vitro* assay of genes located at both ends of the conserved cluster revealed their involvement and possible functions in the tubulysin pathway. Pretubulysin was identified from both natural producer strains, confirming its intermediacy in the pathway. Using a combination of feeding studies and high-resolution tandem mass spectrometry, An d48 and SBCb004 were found to biosynthesize 32 additional tubulysin derivatives. Using Red/ET recombineering, the entire tubulysin core gene cluster from strain SBCb004 was reconstituted and successfully expressed in heterologous host strains *Pseudomonas putida* and *Myxococcus xanthus*. Finally, BLAST analysis of the SBCb004 genome data was carried out to identify candidate monooxygenases, whose involvements in tubulysin assembly were evaluated using a combination of knockout mutagenesis and heterologous expression.

ZUSAMMENFASSUNG

Tubulysine stellen eine Familie von neun komplexe Peptiden dar, welche viel versprechende zytotoxische Eigenschaften gegen multiresistente Tumore aufweisen. In dieser Arbeit wird ein dritter Stamm beschrieben, welcher Tubulysine produziert (*Cystobacter* sp. SBCb004). Obwohl die Tubulysin-Biosynthesegencluster aus SBCb004 und An d48 unvollständig erscheinen, zeigt ein Vergleich der Gencluster eine konservierte Architektur von elf Genen auf, wodurch die Grenzen des Clusters definiert werden konnten. Über Geninaktivierungsstudien und *in vitro* Assays konnte die Beteiligung und mögliche Funktion von Genen, welche an beiden Enden der Gencluster in konservierter Form gefunden wurden, aufgezeigt werden. Prätubulysin wurde in den Extrakten beider natürlichen Produzentenstämme identifiziert und so als Biosyntheseintermediat bestätigt. Zudem wurde basierend auf der Kombination von hochauflösender Massenspektrometrie und Fütterungsexperimenten nachgewiesen, dass An d48 und SBCb004 zusammen 32 zusätzliche Tubulysinderivate herstellen. Der Tubulysin-Biosynthesegencluster aus Stamm SBCb004 ausgehend mittels Red/ET Recombineering rekonstituiert und dann heterolog in *Pseudomonas putida* und *Myxococcus xanthus* exprimiert. BLAST Analysen des Genoms von SBCb004 zeigten die Präsenz von multiplen Monooxygenasen auf, deren Beteiligung an der Tubulysinbiosynthese durch eine Kombination von Inaktivierungsmutagenese und heterologer Expression studiert wurde.

TABLE OF CONTENT

FIGURES

TABELS

ABBREVIATIONS

A domain	Adenylation domain
ADP	Adenosine diphopshate
AMP	Adenosine monophosphate
Amp	Ampicillin
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
Genta	Gentamycin
h	Hour
HPLC	High performance liquid chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Km	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
Pip	Pipecolic acid
PKS	Polyketide synthase
Ppant	Phospopanthetein
PPi	Pyrophosphate
PtK2	Potoroo kidney
RBS	Ribosome binding site
RNA	Ribonucleic acid
Rt	Retention time
RT	Room temperature
S	Sulfur
SAP	Shrimps alkaline phopsphatase

Sec	Seconds
SDS	Sodium dodecyl sulfate
sp.	Species
TE	Thioesterase
Tet	Tetracycline
TetR	Tetracycline repressor protein
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

1 INTRODUCTION

1.1 Natural products in drug discovery

Even thousands of years ago, man was using boiled herb extracts as medicines to treat disease. Insights into the active principles that these extracts contained started to emerge two hundred years ago, when Friedrich Sertürner identified morphine as the pain-relieving principle from the poppy *Papaver somniferum* L [1-3]. Alexander Fleming discovered that the fungus *Penicillium notatum* produced a metabolite capable of killing bacteria, and the active agent pencillin was isolated shortly after that in 1928 [4]. The identification of individual compounds with biological activities represented a major milestone in modern medicine, inaugurating a new era in natural product research. Since then, natural products have been established as invaluable sources for novel drugs and drug-leads. The primary reason for the utility of these compounds is that they are privileged structures selected over millions of years by evolution to interact with a variety of targets in a specific manner [5]. Today, approximately 50% of pharmaceuticals in clinical use are natural products or their derivatives (Figure 1.1).

Natural products are produced by plants, fungi, and both terrestrial and marine microorganisms. These secondary metabolites exhibit a wide variety of biological activities, including anti-bacterial, anti-fungal, antiparasitic and anti-cancer properties [6-9]. In addition to their many applications in human therapy, they are also used in agriculture as plant growth regulators, herbicides and insecticides [10]. A high proportion of these compounds are polyketides, nonribosomal peptides, or hybrids of both classes. In bacteria, the biosynthetic enzymes are usually encoded by genes that are clustered together in the genome, facilitating the identification and analysis of the biosynthetic pathway [11]. The knowledge gained can be used to inform targeted manipulation of the biosynthesis towards the generation of structural derivatives, either through genetic modification of the primary or secondary metabolism of the producing organisms or through the use of individual enzymatic activities *in vitro* [12; 13].

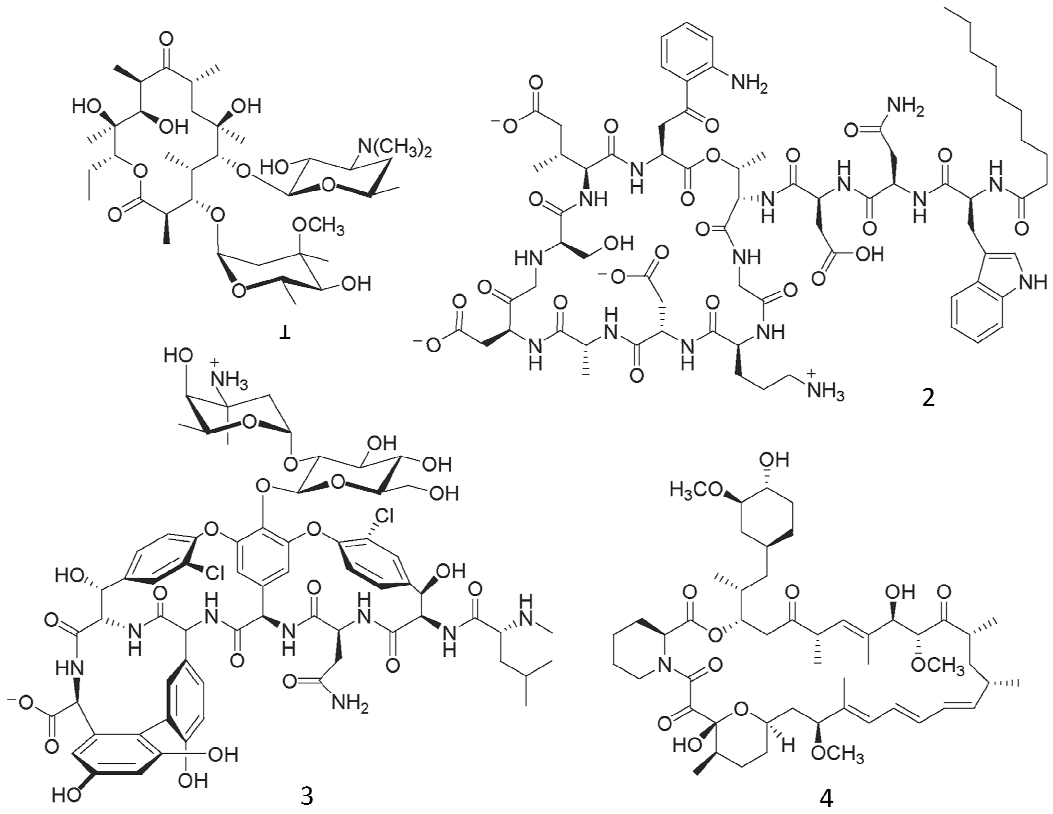


Figure 1.1 Examples of bioactive natural products derived from bacteria.
(1) Erythromycin (polyketide, *Saccharopolyspora erythrea*); (2) Daptomycin (nonribosomal polypeptide, *Streptomyces roseosporus*); (3) Vancomycin (nonribosomal polypeptide, *Amycolatopsis orientalis*); (4) Rapamycin (polyketide-nonribosomal polypeptide hybrid, *Streptomyces hygroscopicus*).

Despite the significant success of natural product in the clinical setting, there remains a pressing need for novel secondary metabolites in particular due to the increase in multi-drug resistance among bacterial pathogens, the lack of highly effective therapeutics for many, and the side effects associated with some current medicines [10; 14]. Efforts have been made to identify new groups of microbial producers, and several underexplored classes of bacteria, such as marine organisms and myxobacteria, have come into focus as promising producers of compounds with both unique structures and bioactivities [14; 15].

1.2 Myxobacteria – a promising source for novel natural products

Myxobacteria are obligate, aerobic, chemotrophic, Gram-negative mesophilic δ-proteobacteria. The cells are typically rod-shaped, which are 4−12 microns in length and 0.7−1.2 microns in width. To date, approximately 7500 different myxobacteria have been isolated, and novel strains, species and even families are continually being discovered [16; 17]. The myxobacteria are united in the order *Myxococcales*, and further divided in three suborders: *Cystobacterineae*, *Sorangiineae* and *Nannocystineae* [18]. They occupy a wide range of habitats, such as soil, the bark of trees, decaying plant materials, herbivore dung, and the marine environment [16; 19; 20].

A special feature of myxobacteria is their ability to move smoothly on solid surfaces by gliding or creeping. During swarming, they leave a thin film of mucus which exhibits a characteristic surface wave pattern [21]. Due to their ability to coordinate their movements, the myxobacteria additionally exhibit an extent of 'social behavior' which is unique among the bacteria. Under starvation conditions, they aggregate to form characteristic multi-cellular structures called fruiting bodies [22; 23]. These structures can exhibit different shapes and colors, depending on the species. Within the fruiting bodies, cells begin as rod-shaped vegetative cells, and develop into rounded myxospores with thick cell walls. The myxospores, which are analogous to spores produced by other organisms, are meant to survive until nutrients are more plentiful. The fruiting process is thought to benefit myxobacteria by ensuring that cell growth is resumed by a group of myxobacteria, rather than by isolated cells. The shape of the fruiting bodies, which range from simple spherical aggregates to palm-shaped structures, is an important means by which the myobacteria can be classified. Another defining feature is that the myxobacteria can access a diversity of biological macromolecules (e.g., cellulose) as food sources via the secretion of extracellular enzymes, and can prey on whole microorganisms such as fungi and bacteria [24; 25]. Taken together, these behavioral features distinguish the myxobacteria from most other bacteria.

Another unique characteristic of the myxobacteria is the size of their genomes. Relative to other bacteria, the myxobacteria have extremely large genomes (about 9−13 Mbp). In fact, the genome of *Sorangium cellulosum* (*S. cellulosum*) at 13.0 million nucleotides, is the largest known bacterial genome (as of 2008) [26-28]. It has recently been noted that there is a direct correlation between the size of bacterial genomes and the ability to produce secondary metabolites [29]. Indeed, the sequenced myxobacterial genomes reveal a high number of gene clusters putatively encoding secondary metabolite biosynthesis, for example strain *S. cellulosum* harbors 17 gene clusters and *Myxococcus xanthus* (*M. xanthus*) harbors 18 gene clusters [26; 30].

Even prior to the advent of whole-genome sequencing, the myxobacteria had been identified as a promising source of natural products (selected examples of myxobacterial secondary metabolites are shown in Figure 1.2). Within the last two decades, at least 100 distinct core structures and approximately 500 derivatives have been characterized from myxobacteria [25]. From analysis of the sequenced genomes, however, it is clear that much of myxobacterial metabolism remains to be characterized, and that genome mining of the strains should provide access to large number of new structures [29; 31; 32]. Of particular interest, myxobacterial natural products exhibit rare or wholly novel modes of action relative to other compounds of bacterial origin, features which make them attractive starting points for drug discovery. For example, argyrin acts as a potent proteasome inhibitor [33], stigmatellin and soraphen both show antifungal bioactivity [34; 35], tubulysin, disorazol, and epothilone all interact with cytoskeleton [36-38]. These findings mark the myxobacteria as a promising source for novel drug leads [25].

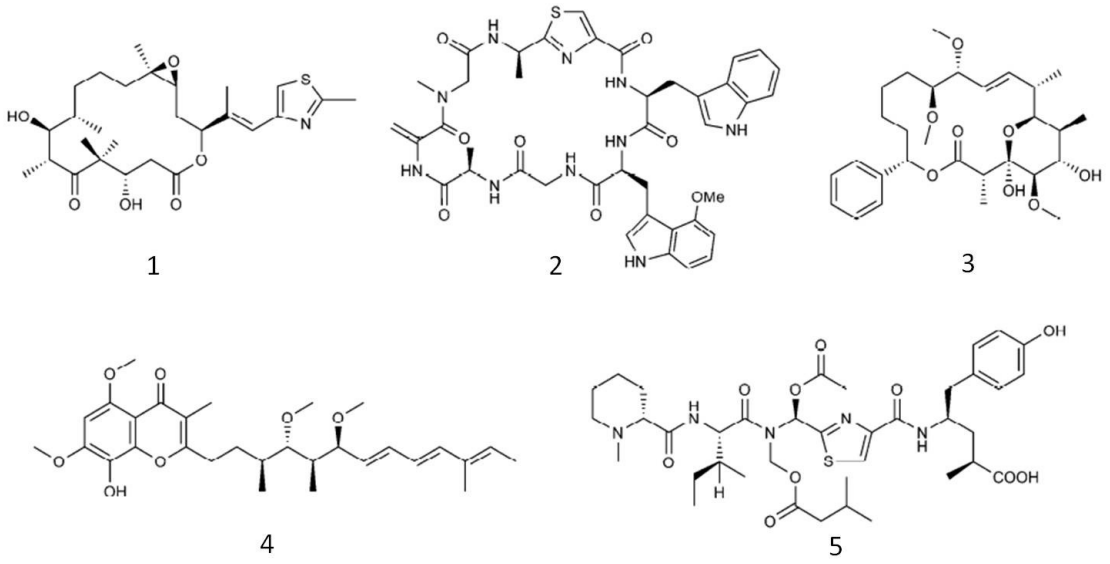


Figure 1.2 Selected examples of myxobacterial secondary metabolites.
(1) Epothilone B (antitumor, *Sorangium cellulosum)*; (2) Argyrin A (antitumor, *Archangium gephyra*); (3) Soraphen A (antifungal, *Sorangium cellulosum*); (4) Stigmatellin A (antifungal, *Stigmatella aurantiaca*); (5) Tubulysin A (cytotoxic, *Archangium gephyra*).

Among the most promising myxobacterial natural products are several compounds which target the microtubule network of eukaryotic cells. Due to their various effects on the network, they can be divided into two groups: either they stabilize microtubules and induce polymerization, or they destabilize them and cause degradation of the microtubule network. Both effects inevitably lead to the collapse of microtubule-dependant cellular processes, and the eventual induction of

apoptosis. The most famous myxobacterial representatives of these anti-mitotic compounds are the epothilones, isolated from the myxobacterium *S. cellulosum* [39]. Indeed, a semi-synthetic epothilone B derivative (Ixabepilone®) was, in 2007, the first myxobacterial natural product to be approved for clinical use by the FDA for the treatment of aggressive breast cancer (Figure 1.2A). Epothilone has been shown to have a similar mechanism of action as that of Taxol, but exhibits much stronger bioactivity, and also induces cell death in Taxol-resistant cancer cell lines. Other myxobacterial compounds which interact with the cytoskeleton include tubulysin and disorazol which destabilize their tubulin target (Figure 1.2) [40], and rhizopodin, which binds to actin preventing its polymerization [37; 38; 41]. Of particular note, the bioactivity of tubulysin is higher than that of both Taxol and epothilone [42].

1.3 Polyketide and nonribosomal peptide biochemistry

Many myxobacterial secondary metabolites are polyketides and nonribosomal polypeptides, which are assembled by gigantic multienzymes called polyketide synthases (PKS) and non-ribosomal peptide synthetase (NRPS), respectively.

1.3.1 PKS biochemistry

The synthesis of polyketides is comparable to that of the fatty acids, as the two pathways have strong homologies both in the nature of the chemistry used in chain extension from a common pool of simple precursors and in the character of the enzymes that are used for chain assembly. Based on the mode of catalysis and the structure of the biosynthetic proteins, PKS are usually classified into type I, II and III, although there are increasing examples of systems which are formally hybrids of several classes [43].

Type I polyketide synthases are multifunctional enzymes that are organized into modules [44]. Each module contains the three core domains required for one round of extension of the growing polyketide chain with a specific carboxylic acid building block. One of the essential domains is the acyl carrier protein (ACP) domain, which serves as a point of covalent attachment of the intermediates to the PKS as well as shepherding the chains to the various catalytic domains within the modules. In order to be functional, the ACP domains must be post-translationally activated by a phosphopantetheinyltransferase (PPTase). These enzymes catalyze the transfer of the 4'-phosphopantetheine (Ppant) moiety of a coenzyme A (CoA) to a highly conserved serine

residue of the ACP, converting the carrier protein domain from its inactive *apo* form to its active *holo* form [45]. During chain assembly, the intermediates are bound to the terminal thiol group of this phosphopantetheine arm as a thioester, and the Ppant acts as swinging arm that is able to shuttle the intermediate to all of the catalytic domains (Figure 1.3).

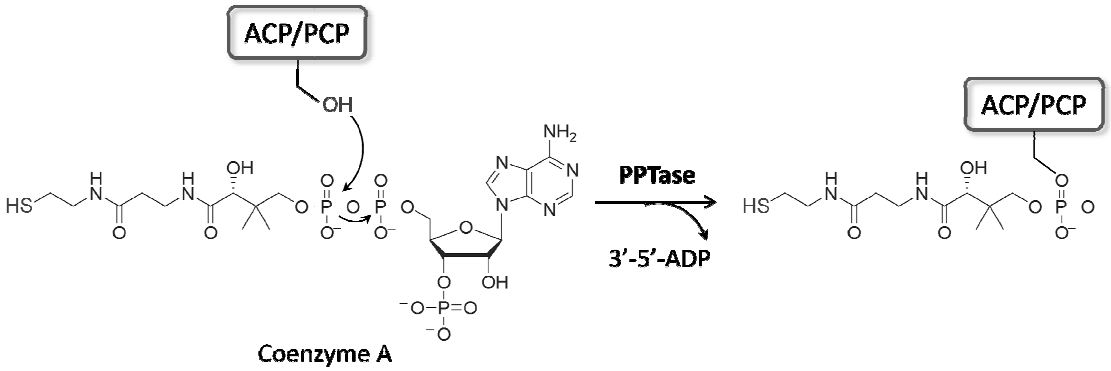


Figure 1.3 Post-translational modification of the ACP/PCP domain by a PPTase.
The PPTase catalyzes the transfer of Ppant from CoA to a conserved serine in the carrier protein domain, converting the inactive *apo* domain to the active *holo* form.

The other two core domains in PKS are the acyltransferase (AT) and the ketosynthase (KS). The role of the AT domain in chain extension modules is to selects short chain dicarboxylic acids (such as malonyl-CoA and methylmalonyl-CoA) as building blocks for the biosynthesis. There is also an AT domain present in the first ('loading') module of the PKS, which is responsible for choosing a starter unit, such as acetyl-CoA or propionyl-CoA. All ATs first catalyze their autoacylation with the building block, and then transfer the substrate to the adjacent ACP. Carbon-carbon bond formation between the newly recruited extender unit and the growing chain is catalyzed by the KS domain. This condensation reaction occurs between the polyketide intermediate tethered to a conserved cysteine of the KS domain, and a carbanion nucleophile generated by KS-catalyzed decarboxylation of the bound extender unit.

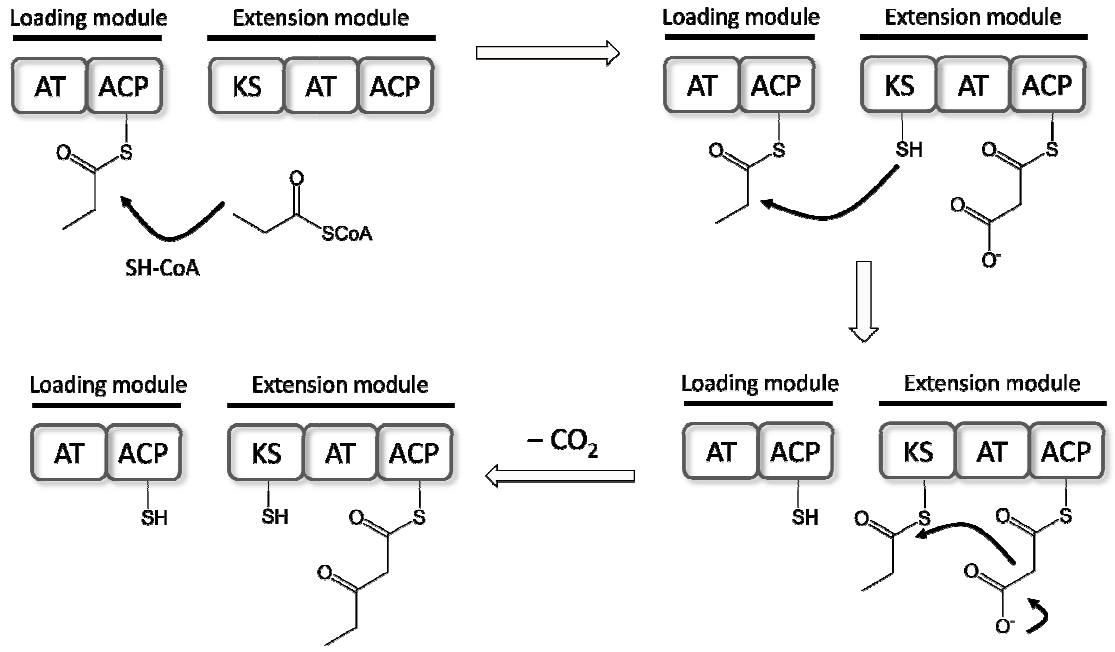


Figure 1.4 Schematic overview of PKS biochemistry.
AT domains within the loading module and the first extention module select their respective acyl-CoA monomers and catalyze the transthiolation of these substrates to the downstream ACP domains. In the next step, the starter unit is transferred from the loading ACP to the KS of the first extension module. Subsequently decarboxylative condensation catalyzed by the KS domain occurs, forming a C–C bond between the starter and extender units, resulting in the elongated ACP-bound intermediate (a diketide in this particular case).

The β-carbonyl groups of the polyketide that result from the condensation can be reduced by several optional domains within the 'reductive loop' of the module [43]. First, the NADPH-dependent ketoreductase (KR) domain can take the carbonyl to the state of an alcohol. The alcohol can then be dehydrated to the double bond by a dehydratase (DH) domain, followed by further NADPH-dependent reduction to the fully-reduced methylene, catalyzed by an enoylreductase (ER) (Figure 1.5). This sequence of reactions is repeated as the growing chain is shuttled from module to module until the biosynthesis is completed, and the intermediate is cleaved from the multienzyme by the terminal thioesterase domain, either by hydrolysis or macrocyclization (Figure 1.4).

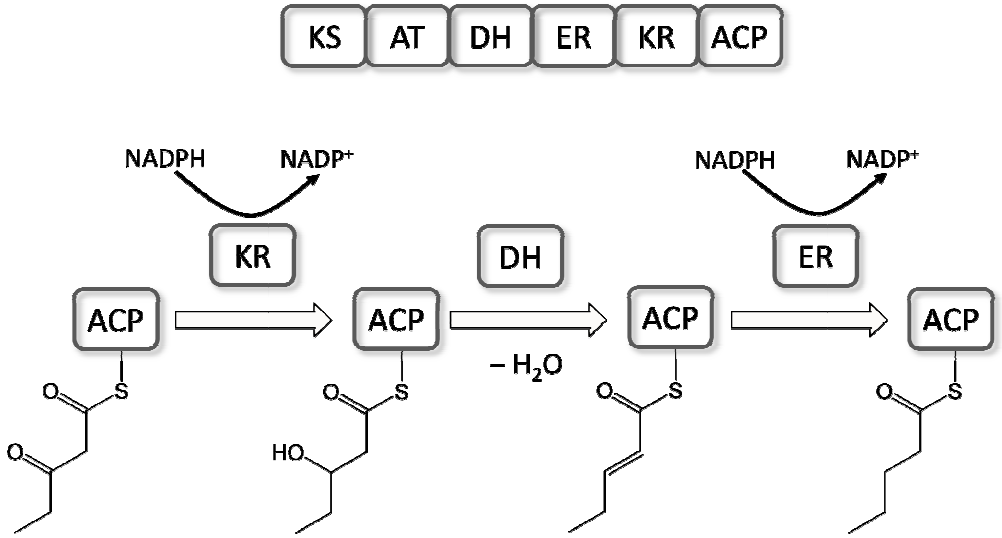


Figure 1.5 Optional β-carbon processing catalyzed by the KR, DH and ER domains.

In type I PKS, although chain extension modules are typically used singly, there are also several exception examples that the PKS modules act iteratively during the biosynthetic process, such as the pathways of stigmatellin in *Stigmatella aurantiaca* [46], and the DKxanthene in *M. xanthus* [47]. The molecular basis for this iterative behavior is presently unknown, but it is reminiscent of the natively iterative type I PKS from fungi [48].

In bacterial type II polyketide synthases, the catalytic domains are present as discrete proteins which must associate into a multienzyme complex. The biosynthetic machinery acts iteratively to produce polycyclic, aromatic metabolites using malonyl-CoA as the substrate for chain extension [44]. In type III PKS systems, the catalytic activity of a single active site accomplishes all reactions necessary for the biosynthesis, including decarboxylation, condensation, cyclization and aromatization reactions. Unlike the other PKS types, type III PKS use CoA-esters instead of ACP- bound intermediates as substrates [49-51].

1.3.2 NRPS biochemistry

Nonribosomal peptide synthetases exhibit a similar organization as the type I polyketide synthases, in that they also comprise modules dedicated to initiation of the biosynthesis, extension of the growing chain and termination of chain assembly [52]. By analogy to the PKS systems, a minimal NRPS module also consists of three core domains: peptidyl carrier protein (PCP), adenylation (A), condensation (C) (Figure 1.6). The A domain is responsible for choosing

the specific amino acid to be added to the chain. It activates the residue to the aminoacyl adenylate in an ATP-dependent reaction, and then transfers it to the adjacent PCP domain within the module, forming the respective aminoacyl-*S*-PCP. The substrate specificity of the A domain is supposed to be determined by eight non-conserved amino acids within the overall hydrophobic binding pocket [53-55]. Notably, the specificity of the A domains is not restricted to the 20 proteinogenic amino acids, as they can also recognize and activate a much wider variety of nonproteinogenic amino and aryl acids as building blocks [56]. This biosynthetic feature contributes to the high structural diversity within this class of compounds. Following substrate selection, the C domain catalyzes the nucleophilic attack of the downstream amino group on the aminoacyl or peptidyl moiety bound to the PCP of the upstream module, resulting in a new peptide bond.

As in PKS systems, optional tailoring domains are also present within NRPS modules, which again enhance the structural diversity of the polypeptide products. These modification domains include the epimerization (E), *N*-methylation (*N*-MT), oxidation (Ox) , heterocyclization (HC), monooxygenase (Mox), reduction (R) and formyltransferase (F) domains [57]. The *N*-MT domain carries out an *S*-adenosyl-methionine (SAM) dependent reaction to transfer a methyl group onto a nitrogen atom of the intermediate, a reaction which occurs prior to the condensation (Figure 1.7A). The function of epimerization domains is to convert the L-amino acids into their D-forms, which occurs after amide bond formation (Figure 1.7C). The HC domain is a variant of a C domain. In addition to peptide bond formation, HC domains catalyze the heterocyclization of the side chains of the amino acids cysteine, serine and threonine onto the peptide backbone: attack of the sulfhydryl group of cysteine or the hydroxyl group of serine and threonine on the adjacent peptide carbonyl results in five-membered hydrolytically labile thiazoline or oxazoline rings. Subsequent oxidization by an FMN-dependent Ox domain results in formation of stable thiazole and oxazole rings (Figure 1.7B). The process of chain extension and modification is repeated until the intermediate reaches the last module where it can be cleaved from the multienzyme by a TE domain as a linear or cyclic product [58], or alternatively released by reduction by an R domain to form the aldehyde or alcohol [59].

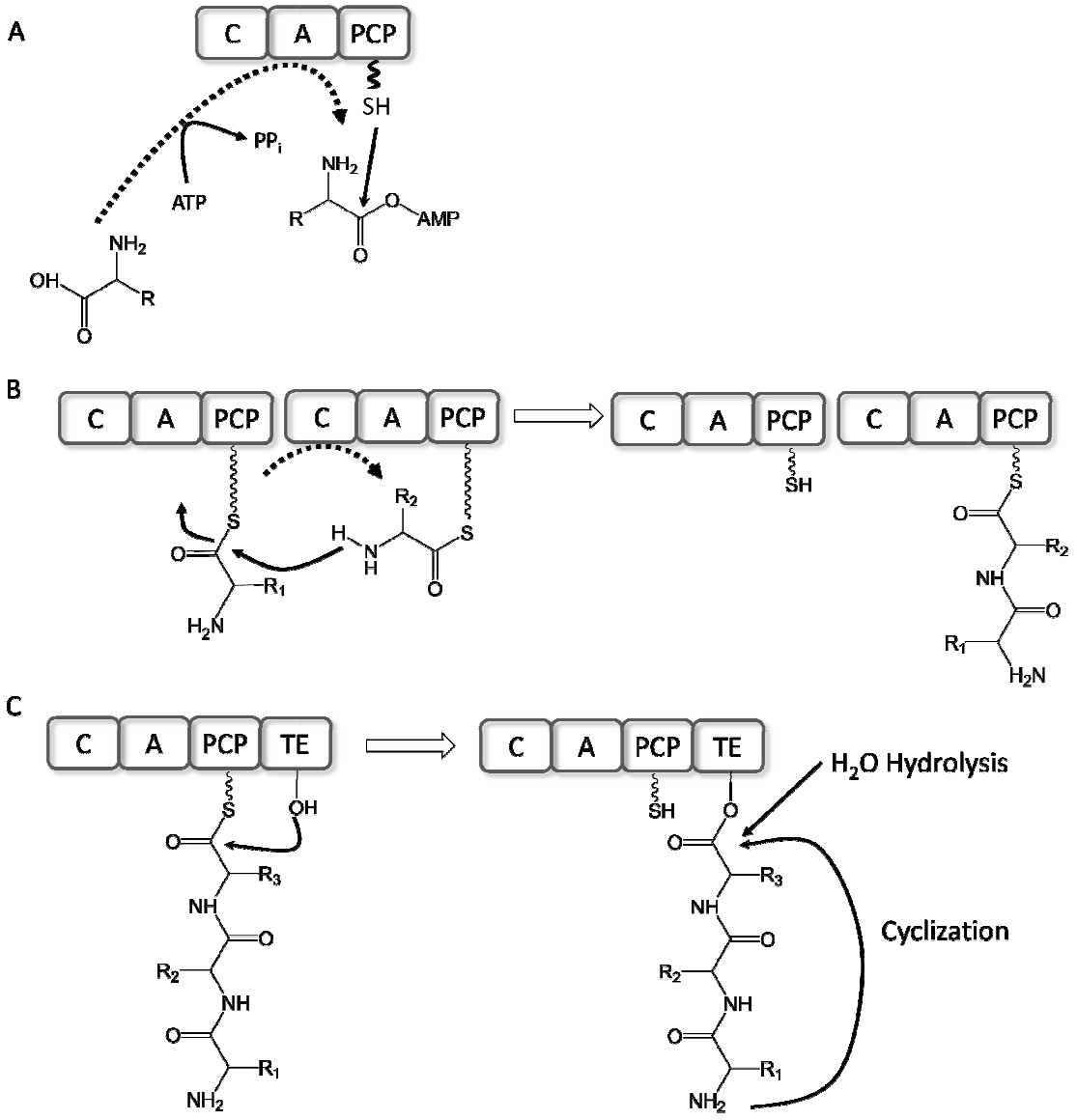


Figure 1.6 Schematic overview of NRPS biochemistry.
(A) Selection and activation of a specific amino acid by the A domain. ATP is consumed in the activation reaction, and the aminoacyl group of the resulting aminoacyl-AMP is transferred to the PCP domain.
(B) The C domain catalyzes a condensation reaction which results in formation of a new peptide bond between two adjacent aminoacyl-*S*-PCP intermediates. Passage from one module to the next results in the elongation of the chain by a single amino acid.
(C) The TE domain releases the peptide from the multienzyme either by hydrolysis or cyclization to terminate the assembly process.

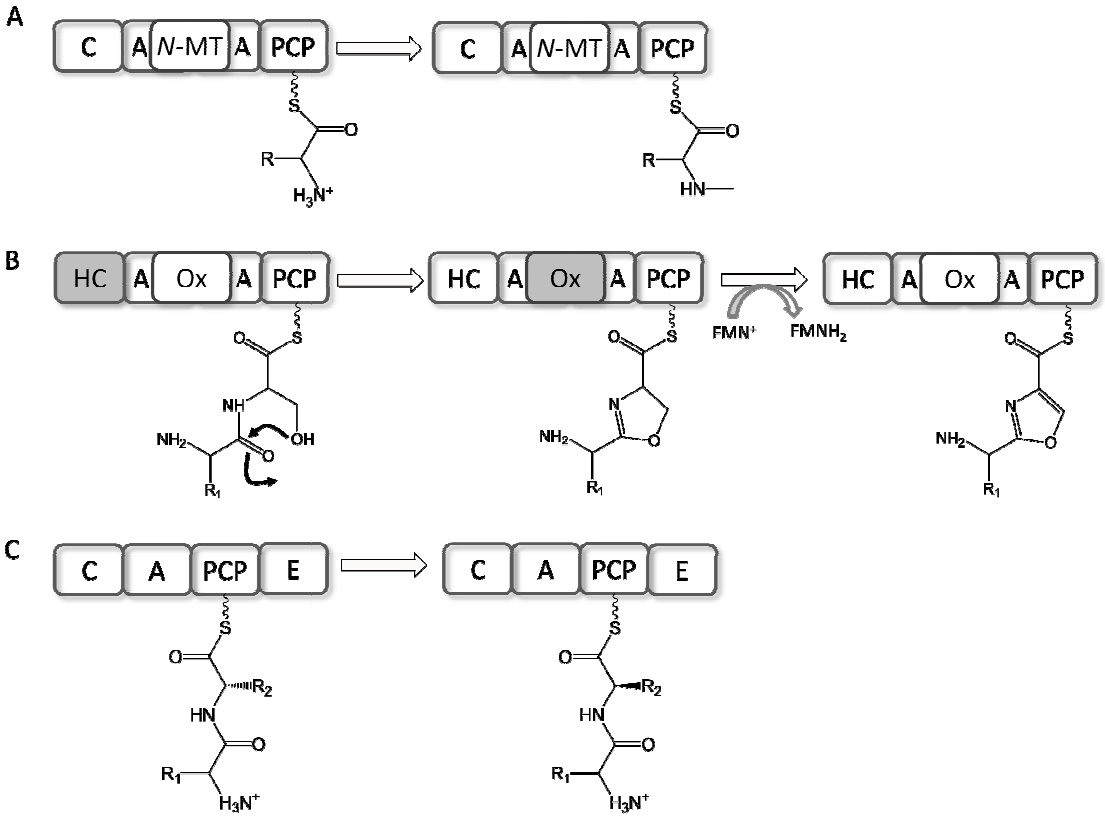


Figure 1.7 Reactions of several NRPS modification domains.
(A) *N*-methylation catalysed by an *N*-MT domain.
(B) Formation of a thiazole ring catalyzed by HC and Ox domains. After carrying out the condensation between two aminoacyl intermediates, the HC domain facilitates the attack of an internal nucleophilic (in this case, a cysteine) on the adjacent carbonyl, followed by dehydration to yield the thiazoline. Subsequently an Ox domain can act on the thiazoline, converting it to the fully-reduced aromatic thiazole using FMN as a hydride acceptor. Domains that are involved in a specific catalytic step are colored in grey.
(C) E-domain catalyzed epimerization.

1.3.3 Post-assembly line modifications

Discrete enzymes, referred to as tailoring enzymes, are able to modify the secondary metabolites after they release from the PKS-NRPS multienzyme complex [60; 61]. Generally the genes encoding these tailoring enzymes are part of the biosynthesis clusters, allowing coupled control of their transcription. These post-assembly line modifications are often responsible for the bioactivity of the natural products.

PKs in actinomycetes are often decorated with additional moieties, such as sugars, hydroxyls and methyl groups [43; 62]. In NRPS system, the typical modifications are carried out by *cis*

acting domains (such as epimerization, methylation and heterocyclization domains that are integrated in the NRPS megasynthetases) during assembly of their scaffold [57], while other post-modifications are mostly catalyzed in *trans* by tailoring enzymes [60]. As a classic example, the biosynthesis pathway of vancomycin incorporates the post-modifications of glycosylation, halogenation, oxidative cross-linking, and hydroxylation [60; 63]. Other examples of glycosylation include erythromycin [64], novobiocin [65], chivosazol [66] and sorangicin [67]. Halogenase exists in the biosynthetic gene cluster of Syringomycin E [60], chondramides [68], and chondrochloren [69]. Furthermore, hydroxylation, methylations and acylations of natural products have also been observed [70].

1.4 Heterologous expression of natural product pathways through Red/ET recombination

The biological activities of natural products are usually structure related. Thus interests are stimulated in using genetic engineering of the biosynthetic pathways to produce analogues for evaluation as drug candidates. But it is always difficult of carrying out such manipulation in the native producing strains, prompting the need for heterologous expression systems in more amendable host strains. Over the last decade, a number of natural product biosynthetic pathways have been heterologously expressed with the aid of Red/ET recombination. In principle, this technique allows the assembly and subsequent transfer of large biosynthetic gene clusters from the natural producer strain into heterologous hosts which are genetically more tractable and easily cultivated. It also provides a platform for the detailed investigation of complex biosynthetic mechanisms, allows identification of the products of silent biosynthetic gene clusters, and the generation of novel analogs through biosynthetic engineering in the genetically tractable *E. coli* [29; 32; 71-75]. The method is thus gaining increasing interest in both biotechnology and drug discovery [76; 77].

Red/ET recombination relies on homologous recombination *in vivo* in a specific strain of *E.coli*. In principle, as the homology regions upon which the technique relies can be chosen freely, any position on a target molecule can be specifically altered in a precise manner, regardless of DNA size. These qualities make the Red/ET recombination technique a revolutionary DNA engineering platform that is capable of engineering large DNA molecules (e.g. cosmids or BACs), and which allows for a wide range of DNA modifications (such as insertions, deletions, substitutions,

fusions, point mutations, direct cloning and subcloning) at any chosen position [78]. In Red/ET recombination, target DNA molecules are precisely altered by homologous recombination in the *E. coli* strains that express the phage-derived protein pairs RecE/RecT from the Rac prophage or Redα/Redβ from λ phage (the pairs are equivalent in terms of function). RecE and Redα are 5'→3' exonucleases, while RecT and Redβ are both DNA annealing proteins. A functional interaction between RecE and RecT, or between Redα and Redβ is necessary to catalyse the homologous recombination reaction. The process is assisted by λ-encoded Gam protein, which inhibits the RecBCD exonuclease activity of *E. coli* (Unlike in yeast, linear dsDNA is unstable in *E.coli* because of the activity of RecBCD) (Figure 1.8) [71; 79]. The homologous recombination occurs through homology regions, termed homology arms (hm). which are regions of DNA sequence that are shared by the two molecules. The modification cassette is typically a small DNA fragment generated by PCR, which harbors the sequence to be introduced and a selection marker, flanked by homology arms. The target DNA can be a gene locus on the *E. coli* chromosome, or any other stretch of DNA in a BAC or plasmid vector (Figure 1.9) [71; 79].

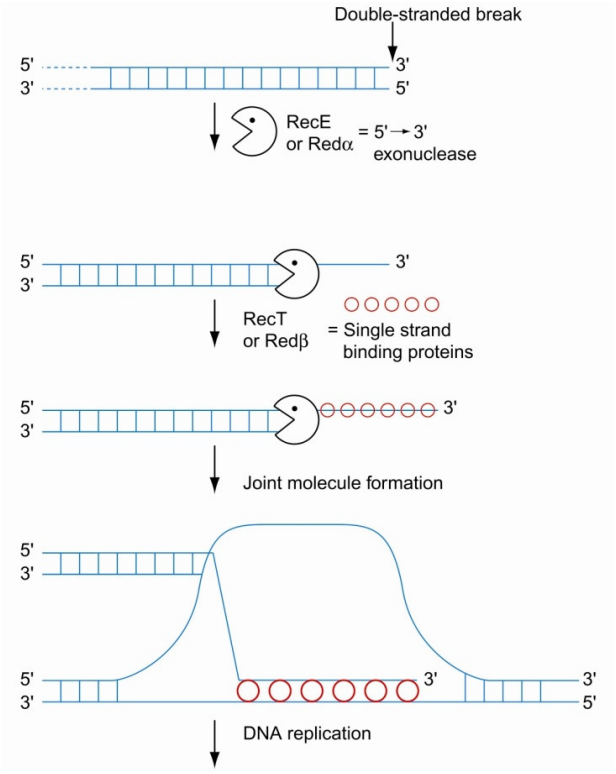


Figure 1.8 Mechanism of Red/ET recombination.
A double-stranded break repair (DSBR) is initiated by the recombinase protein pairs, RecE/RecT or Redα/Redβ. First Redα (or RecE) digests one strand of the DNA from the DSB, leaving the other strand as a 3´ single-stranded DNA overhang. Then Redβ (or RecT) binds to and coats the single strand**.** The protein-

nucleic acid filament then aligns with the homologous DNA. Once aligned, the 3´- end becomes a primer for DNA replication. The diagram was reproduced from www.genebridges.com.

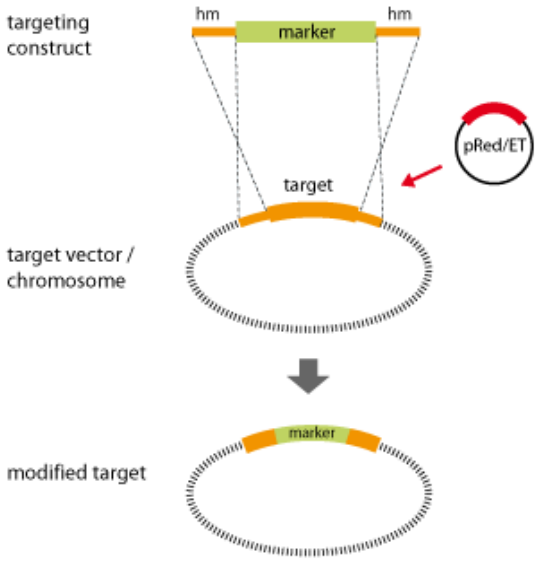


Figure 1.9 Crossover step in Red/ET recombination.
The crossover step occurs between a modification cassette and the target DNA (e.g. a plasmid or the *E. coli* chromosome). The diagram was reproduced from www.genebridges.com.

The general overview of the strategy for cloning and heterologous expression of large natural product assembly lines is shown in Figure 1.10 [71]. First, in order to obtain the natural product biosynthetic gene cluster, a BAC or cosmid library is screened for DNA harboring parts of the gene cluster. As it is often the case that the large biosynthetic clusters are located on multiple BACs/cosmids, Red/ET recombineering in *E. coli* is then used to reconstitute the complete gene sets on one heterologous expression construct. The resulting construct harboring the entire pathway can be further modified in *E. coli* by recombineering, in order for example to add genes that are needed for transformation into the desired heterologous host, promoters and other control elements required for efficient expression, etc. The final construct can then be transferred into suitable heterologous host or hosts, and growth extracts screened for successful production of the desired metabolite [71].

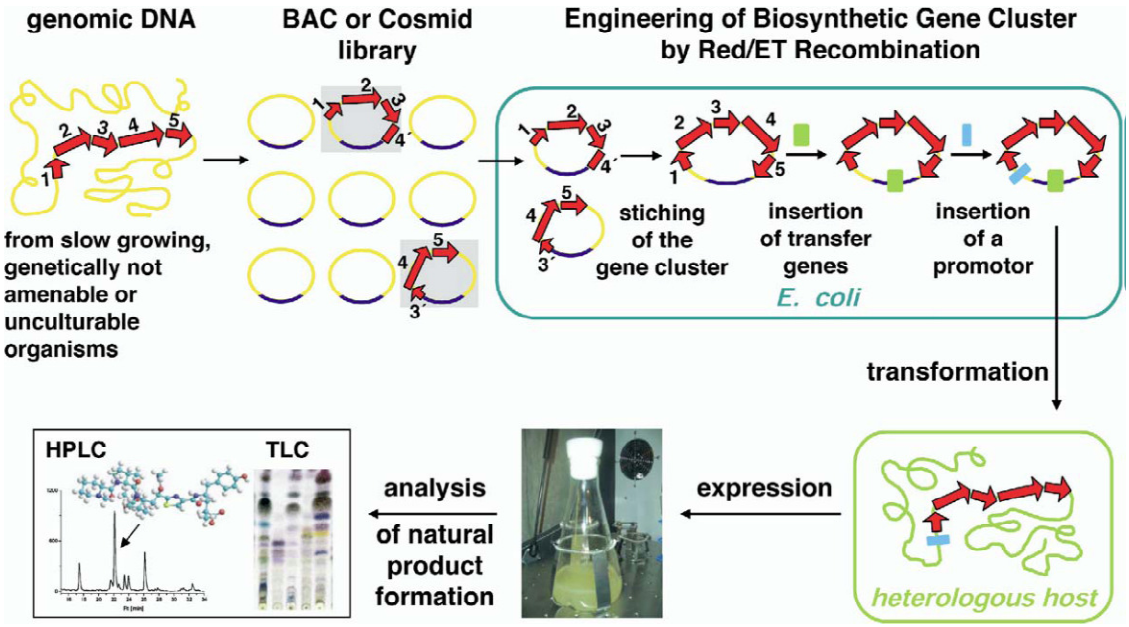


Figure 1.10 General strategy for the heterologous expression of natural product biosynthetic gene clusters.
The figure was reproduced from reference [71].

There are several specific challenges associated with the heterologous expression of polyketide and nonribosomal peptide biosynthetic pathways. For example, the host strain must be able to fold and post-translationally activate the gigantic multienzymes and thus a PPTase of suitably broad specificity must be present. Furthermore, the biosynthesis requires the presence of a specific pool of substrates, potentially including coenzyme A-activated short-chain carboxylic acids, proteinogenic and specific non-proteinogenic amino acids, short-chain fatty acids, and sugars. It is often necessary to adapt fermentation protocols in order to optimize production by ensuring that the required substrates are available in adequate amounts at the correct time. If antimicrobials or toxic agents are to be produced, it is necessary to include inducible promoters so that the effects of the metabolites on the growth of the host can be minimized.

One such inducible promoter system is that is based on tetracycline (pTet promoter). In Gram-negative bacteria, the Tet repressor protein (TetR) regulates the transcription of a family of tetracycline resistance genes [80-82]. In this system, the gene *tetA* encodes the tetracycline resistance protein which is a membrane-spanning H^+–[tet–Mg]$^+$ antiporter. The 81 bp sequence between *tetA* and *tetR* habors the promoter of *tetA*, pTetA, two overlapping promoters of *tetR*, pTetR1 and pTetR2, and two highly homologous control gene sequences $tetO_1$ and $tetO_2$ (Figure

1.11). In the absence of the inducer tetracycline, TetR protein dimmers bind to the two operators *tetO*$_1$ and *tetO*$_2$, shutting down transcription of the downstream resistance gene *tetA*. TetR has high affinity to tetracycline. Therefore, once tetracycline enters into the cell, the TetR protein will complex with the tetracycline in preference to the *tetO* sequence region. This binding lifts the inhibition on the pTetA promoter, allowing for expression of TetA. In summary, the regulatory system based on *tetR* and the 81 bp *tetO* control region can be used to regulate transcription of downstream genes via induction by tetracycline.

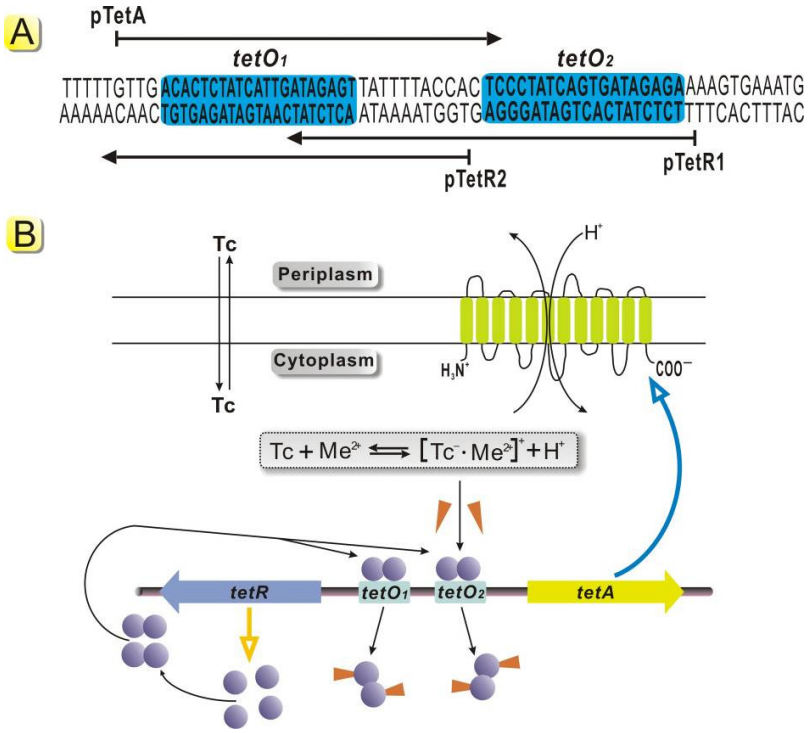


Figure 1.11 pTet promoter regulation system.
(A) The 81 bp gene sequence containing the promoter pTetA and pTetR as well as two operators, *tetO*$_1$ *and tetO*$_2$.
(B) Principle of the regulation of the pTet promoter.
The figure was reproduced from reference [81].

Taking all these challenges into account, it is not surprising that to date, only a few bacterial strains have been demonstrated as suitable heterologous expression hosts, including *E. coli*, *Streptomyces*, Pseudomonads, *Bacilli*, and the myxobacteria [75]. Nonetheless, heterologous expression has been used successfully with a number of natural product assembly lines, including those responsible for polyketides, non-ribosomal polypeptides and their hybrids. Examples

include the expression of some small type II PKS gene clusters in Streptomyces strains (e.g. medermycin, griseorhodin A, aloesaponarin II, precursors of kanamycin) [83-86], landomycin A and prodigiosin in *Streptomyces fradiae* and *Erwinia carotovora* strains [87; 88], daptomycin in *Streptomyces roseosporus* [89], phenalinolactone in *Streptomyces* strains [90], yersiniabactin in *E. coli* [91; 92], myxochromide and myxothiazol in *Pseudomonas putida* (*P. putida*) and the myxobacterium *M. xanthus* [71; 93; 94], and the epothilones in *Streptomyces coelicolor*, *M. xanthus* and *E. coli* [95-98], et al.

1.5 The tubulysins – targets for biosynthetic investigation and heterologous expression

Tubulysins are active against a wide variety of cancer cell lines (ovarian, breast, prostate, colon, lung and leukemia) within the NIC-60 cell line panel, including multi drug-resistant tumors [99] (Figure 1.11). They are among a handful of natural products which interact with the eukaryotic cytoskeleton, inhibiting the polymerization of tubulin at very low concentrations (<50 pg ml^{-1}) [37; 40]. Notably, their growth inhibition potential exceeds that of the anti-cancer drugs epothilone, vinblastine, and taxol, by 20–1000 fold [100]. Mode-of-action studies have demonstrated that the tubulysins disrupt microtubule assembly [37], and that they also show anti-angiogenic effects [101]. Together, these properties make the tubulysins attractive lead structures for development as anti-neoplastic agents by both academic and industrial sectors.

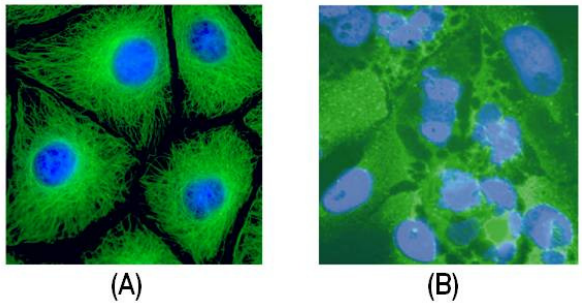
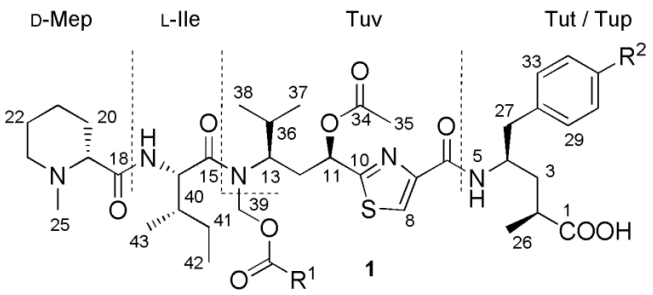


Figure 1.12 Activity of tubulysin against PtK2 **cancer cells.**
(A) Microtubules of control cells. The cells were fixed and immunostained for tubulin. Green: microtubule network; blue: cell nuclei.
(B) Incubation of Potoroo kidney (PtK2) cells with tubulysin A (50 ng/ml) for 24 hr. After 24 hours, the cell nuclei begins to fragment and the microtubule network is completely degraded .
The figures were reproduced from reference [102].

The first nine-members of the tubulysin family (Figure 1.12) were discovered by Höfle, Reichenbach and their co-workers from myxobacterial strains *Angiococcus disciformis* An d48 and *Archangium gephyra* Ar 315, using bioactivity-guided screening with L929 mouse fibroblasts [40; 100]. They were subsequently identified in extracts of additional myxobacteria, including *Stigmatella* and *Cystobacter*. The shared molecular core comprises five amino acids (*N*-methyl pipecolic acid (Mep), isoleucine (Ile), valine, cysteine and phenylalanine or tyrosine), and two units of acetate. In the synthetic literature, valine, cysteine and one unit of acetate have been grouped together into a structural unit referred to as tubuvaline (Tuv), while the acetate chain-extended forms of phenylalanine or tyrosine are designated as tubuphenylalaline (Tup) and tubutyrosine (Tut), respectively. All of the structures include an acetoxy moiety, but exhibit variable functionality (R^1) within the Tuv bis-acyl *N*,*O*-acetal substituent. SAR studies using synthetic analogues of the most potent tubulysin, tubulysin D, have begun to identify the essential structural features underlying its cytotoxicity, as well as suggesting strategies for optimizing the metabolite's pharmacological properties [103-108]. Taken together, these data reveal a surprising tolerance to structural modification, encouraging efforts to synthesize simplified variants of the tubulysins for evaluation as anticancer agents. For example, replacing the chemically labile *N*,*O*-acetal and stereogenic acetate groups of tubuvaline with stable alternatives resulted in only a minor loss of potency [104].



Tubulysin	R^1	R^2
A	$CH_2CH(CH_3)_2$	OH
B	$CH_2CH_2CH_3$	OH
C	CH_2CH_3	OH
D	$CH_2CH(CH_3)_2$	H
E	$CH_2CH_2CH_3$	H
F	CH_2CH_3	H
G	$CH{=}C(CH_3)_2$	OH
H	CH_3	H
I	CH_3	OH

Figure 1.13 The nine tubulysins known when this work was initiated.

Angiococcus disciformis An d48 is known to produce four tubulysins, the tubulysins D, E, F and H [100]. Sequencing of the approximately 40 kb tubulysin gene cluster within the strain revealed that tubulysin assembly is catalyzed, as expected, by a molecular assembly line, consisting of five NRPS modules and two PKS modules [102]. However, the gene set does not

incorporate any obvious functions for carrying out the required post-assembly line oxidation and acyl transfer reactions.

1.6 Outline of this work

The goal of the work described in this thesis was to deepen our understanding of the tubulysin biosynthetic pathway. Studies were carried out with the aim of obtaining further insights into pathway initiation and post-assembly line aspects, as well as identifying novel tubulysin derivatives. A second major goal was to facilitate the eventual engineering of the gene cluster towards the generation of bioactive derivatives.

The identified tubulysin core biosynthetic gene cluster from strain An d48 is incomplete due to the missing post-modification enzymes. As comparative cluster analysis is often informative, the first specific aim of this work was to identify the tubulysin gene cluster in a second myxobacterial strain, *Cystobacter* sp. SBCb004. I wanted to determine the cluster boundaries via comparative analysis of the two clusters, and also wanted to find those missing gene. Detailed comparison of the two clusters allowed determination of the cluster boundaries, supporting a role for several conserved genes (orfs 1, 2, 17 and 18) whose participation in the biosynthesis was not obvious based on automated assignment of gene function. In addition, the more amenable SBCb004 strain was then expected to provide a more convenient platform for further investigation of the tubulysin biosynthetic pathway. As proof-of-principle, an attempt was made to boost tubulysin yields in the strain by manipulation of the promoter region.

Among the auxiliary genes, the first target for investigation was the cyclodeaminase-encoding gene *tubZ*, whose protein product was expected to furnish the unusual amino acid L-pipecolate, the presumed starter unit for tubulysin assembly. This activity was confirmed by analysis *in vitro* of recombinant TubZ produced in *E. coli*. TubZ's role was also explored by inactivation mutagenesis in strains SBCb004 and An d48, which demonstrated that the protein participates in the pathway, but is not essential for tubulysin biosynthesis. A similar inactivation approach was used to investigate the function of genes *orf1*, *orf2*, *orf17* and *orf18*. Although the experiments with orfs 1 and 2 did not clearly reveal their function, they nonetheless showed that the protein products play a role in tubulysin biosynthesis. Inactivation of orfs 17 and 18, on the other hand, demonstrated their explicit involvements, and also provided clues for the function of acyl

hydrolase in the tubulysin assembly. In addition, this work provided support for the role of TubA as one of the missing acyl transferases.

During the course of this work, new metabolites, pretubulysin A or D, were discovered in extracts of wild type SBCb004 and An d48. This result prompted reanalysis of the extracts by HPLC-MS/MS for the presence of additional metabolites which may have escaped detection in earlier experiments. This work resulted in the identification of 30 additional tubulysin derivatives, resulting from skipping of specific catalytic domains or even an entire PKS module, or from the missing of glycerol hydrolysis carried out by the patatin genes, illustrating the inherently diversity-oriented nature of the pathway. As pretubulysin has been demonstrated to retain good bioactivity [109], these additional simplified analogues of tubulysin molecules could find interest as drug leads.

To facilitate the engineering of such tubulysin derivatives, the next goal of this thesis was to reconstitute the entire tubulysin gene cluster using Red/ET recombineering technology, and to heterologously express it in host strains *P. putida* and *M. xanthus*. The established heterologous expression system was then used to confirm the gene inactivation results obtained with the natural hosts strains An d48 and SBCb004.

Furthermore, I aimed to use BLAST analysis of the SBCb004 genome to identify possible candidate monooxygenases. Then coupled with inactivation mutagenesis and heterologous expression, the possible candidate could be studied in detail. Inactivation of identified P450 gene *633P1* in SBCb004 supported its role as one of the missing oxidative functions, although addition of the gene to the heterologous expression construct did not result in oxidized pretubulysins.

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 in this work

Chemical product	Manufacturer
Acetone Acetonitrile Chloroform Ethanol Ethyl acetate Hexane Methanol Magnesium sulfate 2-Propanol Succinic acid	Sigma Aldrich
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	Merck

Potassium sulfhate Sodium chloride Sodium citrate-di-hydrate Sodium hydroxide	
Bacto agar Casitone Casaminoacids Malt extract Tryptone Peptone Yeast extract	Becton Dickinson and Co., USA
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
BSA (Bovine Serum Albumin)	New England Biolabs
XAD 16 adsorber resin	Rohm and Haas

2.2 Enzymes, kits and markers

All the enzymes in this work were used according to the manufacturers' protocols.

Table 2.2 Enzymes, kits and markers

Product	**Manufacturer**

DNA ladder T4 DNA ligase Page RulerTM Prestained Protein Ladder Page RulerTM Protein Ladder Shrimp alkaline phosphatase Restriction endonucleases Plasmid purification kit Nucleotides	MBI Fermentas
Taq Ploymerase	Quiagen
Phusion DNA Polymerase	Finnzymes
KOD DNA Polymerase	Novagen
Topo TA Cloning® Kit Gel-Dry™ Drying Kit	Invitrogen
Proteinase K Lysozyme	Roth
DIG labeleld DNA Molecular Weight Marker III DIG-Labeling and Detection Kit DIG-Blocking solution DIG EasyHyb DIG PCR Labeling Mix	Roche
BugBuster® Protein Extraction Reagent	Novagen
Gigapack® III Gold Packaging Extract	Stratagene

2.3 Buffers and solutions

Table 2.3 Buffers and Solutions

Buffer	**Constituents**	
Molecular biology:		
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
Neutralisation buffer (P 3)	3 M Potassium acetate	60 ml
	Acetic acid	11.5 ml
	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
1 M Tris-HCl	TRIZMA Base	60.5 g
	H_2O	to 500 ml
10% SDS solution	SDS	10 g
	H_2O	to100 ml
5 M Sodium chloride	NaCl	580 g
	H_2O	to 100 ml
Southern Blot and colony hybridization:		
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
20 × SSC	NaCl	155.3 g
	Sodium citrate•2 H_2O	88.2 g
	H_2O	to 1L
Hybridization buffer	DIG Easy Hyb Granules	1 bottle
	H_2O	to 64 ml
2× wash solution	20 × SSC	100 ml
	10% SDS	10 ml

	H_2O	to 1L
1× wash solution	20 × SSC	50 ml
	10% SDS	10 ml
	H_2O	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
Protein expression and *in vitro* assay		
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
Elution buffer	Tris-HCl (pH 7.8)	20 mM
	NaCl	200 mM
	Glycerol	100 ml
	Imidazole	500 mM
	H_2O	to 1 L
Nickel solution	$NiSO_4$	100 mM

Stripping buffer	Sodium phosphate (pH 7.4)	20 mM
	NaCl	500 mM
	EDTA	50 mM
	H_2O	to 1 L
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 (The DTT is added directly before use.)	800 µl
PBS buffer	PBS tablet	1 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

Table 2.4 Mediums used in the cultivation*

Medium	**Constituents**		**pH value**
Media for *E. coli* cultivation			
LB medium	Tryptone	10 g	pH 7.4
	Yeast extract	5 g	
	NaCl	5 g	
	H_2O	to 1 L	
2YT medium	Tryptone	16 g	pH 7.4

	Yeast extract	10 g	
	NaCl	5 g	
	H_2O	to 1L	
Low salt LB medium	Tryptone	10 g	pH 7.4
	Yeast extract	5 g	
	NaCl	2 g	
	H_2O	to 1 L	
Media for myxobacterial cultivation			
Trypton medium	Trypton	10 g	Adjust pH to 7.2
	$MgSO_4 \times 7\ H_2O$	2 g	with KOH (10M)
	H_2O	to 1 L	
M7 medium	Probion	5 g	Adjust pH to 7.2
	$CaCl_2 \times 2\ H_2O$	1 g	with NaOH (5M).
	$MgSO_4 \times 7\ H_2O$	1 g	Glucose and VB_{12}
	Hefe extract	1 g	were added in after
	HEPES	10 g	autoclave.
	Starch	5 g	
	Glucose 20 %	10 ml	
	Vitamin B_{12} Stock solution	1 ml	
	H_2O	to 1 L	
Glucose (20%)	D(+)-glucose-monohydrate	20 g	Filter sterilize.
	H_2O	100 ml	
VB_{12} stock solution	Vitamin B_{12}	100 mg	Filter sterilize.
(1 mg/10 ml)	H_2O	to 1 L	
M-medium	Phytone Peptone	10 g	Adjust pH to 7.2
	Maltose	10 g	with KOH (10M).
	$CaCl_2 \times 2\ H_2O$	1 g	
	$MgSO_4 \times 7\ H_2O$	1 g	
	NaFe-EDTA	8 mg	
	HEPES	11.9 g	
CTT medium	Casiton	10 g	Adjust pH to 7.0
	Tris-HCl, 1M (pH 7.6)	10 ml	with KOH (10M).
	K_2HPO_4, 1M (pH 7.6)	1 ml	

	$MgSO_4$, 0.8 M	10 ml	
	H_2O	to 1 L	
PMM medium	K_2HPO_4	8 g	Adjust pH to 7.0
	KH_2PO_4	5 g	with KOH (10M).
	$(NH_4)_2SO_4$	1 g	
	Succinic acid [Na-salt]	6.6 g	
	$MgSO_4 \times 7\ H_2O$	1.2 mM	
	H_2O	to 1 L	

* For agar plates, 15 g/L agar was added into the medium; for soft agar plates, 7.5 g/L agar was added into the medium.

2.5 Antibiotics

Table 2.5 Antibiotics used in this work

Antibiotic	Stock solution	Working concentration
Kanamycin (Km)	50 mg/ml	30 ~ 100 μg/ml
Ampicillin (Amp)	100 mg/ml	100 μg/ml
Zeocin (Zeo)	100 mg/ml	15 μg/ml
Tetracyline (Tet)	10 mg/ml 75% EtOH	5 μg/ml
Chloramphenicol (Cm)	34 mg/ml	20 μg/ml
Gentamycin (Genta)	20 mg/ml	6 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)
• 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	
Membrane filter 0.22 µm	(Millipore)
Nylon membrane, positively charged	(Roche)
Microspin Columns	(Amersham Pharmacia)

2.7 Bacterial strains

Table 2.6 Bacterial strains used in this work

Strain	Genotype/characteristics	Reference
E.coli DH10B	F-*mcrAΔ*(*mrr-hds*RMS-*mcr*BC) *Ø80dlacZΔM15ΔlacX74 deoR recA1 endA1 araD 139Δ (ara,leu) 7696 galU galK rpsL nupG*	
E. coli HS 996	*F-mcrAΔ(mrr-hdsRMS-mcrBC) Ø80dlacZΔM15ΔlacX74 deoR recA1 endA1 araD 139Δ (ara,leu) 7696 galU galK rpsL nupG fhuA::IS*	Invitrogen
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 hsdSB(r_B^- m_B^-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*	
E. coli Rosetta BL21 (DE3) pLysS/RARE	*F⁻ ompT hsd SB(R_B^- R_B^-) gal dcm λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*	Novagen

	pLysSRARE (Cm^R)	
E.coli SURE	*endA1 glnV44 thi-1 gyrA96 relA1 lac recB recJ sbcC umuC::Tn5 uvrC e14- Δ(mcrCB-hsdSMRmrr)171 F'*[*proAB+ lacIq lacZΔM15* Tn10(tet^R)]	Stratagene
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
YZ2005	HS996 Δ*recT* and *red*α	Genebridges
A. disciform An d48	wild type	
Cystobactor sp. SBCb004	wild type	
M. xanthus DK1622	wild type	
P. putida FG2005	wild type	

2.8 Oligonucleotides and general plasmids

Table 2.7 Oligonucleotides

Primer	Sequence (5´→ 3´)
Primers for knockout mutagenesis	
Kan_for	CAGGGCGCAAGGGCTGC
Kan_rev	GGGAATAAGGGCGACACGG
tubZ_for	CGGGTCAAGCTTCAAAGAGGCT
tubZ_rev	TACGAGCTGGATCCGACTTCGAC
Primer 1	GCTCATGAGCCCGAAGTGG
Primer 2	CGCGCAGACCAAAACGATC
primer A	CCCTTCGGCAGCCCCAACAT
primer B	CAGCGGCAGTGGAGTGAAG
orf1-for	AGATGTAAAGCTTCGGTGGGGTC
orf1-rev	CTGCAAGAGGATCCGACATCCAG
tubA-for	GGCCTTTCAAGCTTGGACCCTGG
tubA-rev	GTCGCGCGTTGGATCCAGGGCTG
Pat1-for	GACCATCGGATCCTATGGGC
Pat1-rev	GATCGCAAGCTTGGGCCAC

Pat2-for	CACTCGGATCCAATACAGGAG
Pat2-rev	CGTTATAAGCTTCCTAGGTAG
patvf_for	GGAGCAGGATTTGGAAGAC
patvf_rev	GACCATATGACGCTGCTG
TubZ-exp-f	GGAGCCGAATTCGTGAATGCAGTG
TubZ-exp-r	GGCGGCCGCTACGGCTGGGGC
Cbt-Z1	GAAGTGACTCATGTTGAGCACGAGGGTG
Cbt-Z2	GAGCGCACCAGCTCGAAGAG
Cbt-Z3	GTCGGCGAGGGATCACTCGCGGGCTGC
Cbt-Z4	TGCCTGAAAGAAGGGTTACCGC
ctubA-a	ACTCTGACCGACGGAGCAGGTTTTCTAC
ctubA-b	CAGCGTTGCATCCACGGTTG
ctubA-c	AATGCGCACTCCCGGTGAAAC
ctubA-d	TCGTCTGCCCGGACTACCTG
corf1-a	TGCCTGAAAGAAGGGTTACCGC
corf1-b	ATGTGGATGCTCAACCCGTAG
corf1-c	GAGCGCACCAGCTCGAAGAG
corf1-d	CATCCGAGAGGATGTGATGG
Cbt-H1	CAACTGATCGCCCTCGAATGGGCACTC
Cbt-H2	TGCTGCCCTGAGGCAGCAAC
Cbt-G1	CAGGTGAAGACCGATGGCCGGCGAAAC
Cbt-G3	CAGTTGAGACAAGTACCGCATCCTCAG
Cbt-G4	AGATCACCTGTCCCTGCTG
Cbt-H3	TAGTGGAGCAGGAGGTGTC
Tn5-a	GGTTTTATGGACAGCAAGCG
Tn5-b	CCCTCAGGCGACATGCGAAACGATCCTC
ctubB-a	GAGGATCGTTTCGCATGTCGCCTGAGGG
ctubB-b	GATGAGTCCCTGCAACTCG
pTOPO-In	CCTCTAGATGCATGCTCGAGC
pTOPO-Out	TTGGTACCGAGCTCGGATCC
24P-a	CGCTTGAAACCCGCACCCGCTCTACCAC
24P-b	TCGGGGAAGAGCGAGGGATC
45M-a	GGACTGAGACAAGGATCCCTACGTCGAG

45M-b	CCGTCGGTATCCACGAGCTG
302P-a	CGAGTGAAGGCAGAACCCCTATCCCTTC
302P-b	AACTGCGCTGGATCGCGGTTC
333P-a	TCACTGACTCAAGGATCCGTTCCCCTTC
333P-b	TGCTCGAAGACGGACTCGTC
399P-a	CATCTGAGTCGTTCGCGACGTCCGGAAG
399P-b	CGCAGACCGTGTCGTCCTCCATC
633P1-a	TCCCTGATACGACCAGCTCCGGAGCTTC
633P1-b	GTCTCGGAACTTCAGCGGATCCC
1042P-a	CGGGTGATTCCCGTACATCCACGGACTC
1042P-b	GTCCGGAAGATGGTGCGACACG
633P2-a	CGCGTGACTGTCGGACCTCTTCTGG
633P2-b	ACGGGCGGTTCGCCGCGAC
Primers for heterologous expression	
15A-Km-tub3	CAGGTGCAACGCAACGCCGCCGAGCGCGCCTCGCGGCAGCAGA AGGCCGCCCAGCAGCGCCGGGCCCATATGAAGCGAGGATCACCC CATGACGCGTAGGCGCTAGCGGAGTGTATACTG
15A-Km-tub5	TGAAGTTGTACGCCTGCTCCAGTCCCCGCTCCCCGGTGAGGGAC GCCTGCAACCGGAGATCCCTCGCGAGGAAGCGCGGCTGTTCGAA GTCAGAAGAACTCGTCAAGAAG
ask-tub5	CGGCCGGCCATTCCCACCACGGCGATGCCCGACAGAGACGCCTC GTCCGCGCTACGCGTCATGGGGTGATCCTCGCTTCATATGCTAGC TAGATTTTTGTCGAACTATTC
ask-tub3	TTCCTCCGAACCACCGGCTCCGCCCGCGCTCGACGTCGTCCTCGC GCCACACGCCGAAGGCGGCGTGGTCGACCGCATATGCTAGCTGA CCTGTGAAGTGAAAAATGG
Zeo-tub5	GCGCTCGACGTCGTCCTCGCGCCACACGCCGAAGGCGGCGTGGT CGACCGCAGCACGTGTTGACAATTAATC
Zeo-tub3	ATGTCGCACAATGTGCGCCATTTTTCACTTCACAGGTCAGCTAGC ATATGTCAGTCCTGCTCCTCGGCCAC
Tps-amp5	GTCGTTCTTCGAGCGGTAGCGTCAAGGTACTCACCGTCCAGCCC GAGAAGGTTCCGGAACATCGAATCAC
Tps-amp3	CCCTCATCAGTGCCAACATAGTAAGCCAGTATACACTCCGCTAG

	CATATGTGCTTCGGGGTCATTATAGC
Genta-tub5	TCACCGAGCAACCGACATCCCCGCAGGCGACCGAGGAGGGCGA GATTTGAAGCTTAGGCACGAACCCAGTTGACATAAG
TetR-tub-ATG	CACAGCCGCACTCCCAGCCCCGCGGCGTGCGCGAGCAATTCTCC CGGGCTCATATAGTGCCTCTTCTCTATCACTG
TetR-tub-TTG	CACAGCCGCACTCCCAGCCCCGCGGCGTGCGCGAGCAATTCTCC CGGGCTCAATTAGTGCCTCTTCTCTATCACTG
H-orf2-for	GTTCCTCTACAAGCGCCAGACCGACACGGCAGCCCGCGAGCGCT CCCTCGGCAGAATAAATAAATCCTGGTGTC
H-orf2-rev	TCGGGCTGGATGTTGTACATCCGCCGCACCTCCTCGAGCGGCAG CTCCATAAATTACGCCCCGCCCTG
H-tubZ-for	CCCACTCGGACCCGAGACACACCAAAGCAAGGGGAAGACACTC ATGTTGAGCAGAATAAATAAATCCTGGTGTC
H-Orf1-for	TTTCAGGCAAGGGCGGTGTTGGTAAGACCACCGTCTCCGCGGCC ACCGCCGCAGAATAAATAAATCCTGGTGTC
H-tubA-for	GTACTCGGACCGACGGAGCAGGTTTTCTACAAGGTACGGCGGCT GCGCCCGCAGAATAAATAAATCCTGGTGTC
H-tubA-rev	GTCTCCGCGAGCACCTCGCAGGTGCGCGAGGCGATGCGCTCGGC CAGCTCAAATTACGCCCCGCCCTG
P450-zeo5	AGGGATTGCGGCGCGTCTCGTCGGAGAGGAAGTCCAGCATTAGT GCCTCTTCTCTATCACTGATAGGGAGTGGTAAAATAACTCTATCA ATGATAGAGTGTCAGTCCTGCTCCTCGGCCA
P450-zeo3	CGACGGCCAGTGAATTGTAATACGACTCACTATAGGGCGAATTC CCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCT TCTGACCTAGCTAGCCTTAAGAGCAC
633P1-1	CCGAGGGCAAAGGAATAGATATGCTGGACTTCCTCTCCGACG
633P1-2	CAGGACTTCAGGAAGGCGAGCAGGTCGGCGTTGACCTGATCCTT GTGCGTGTCATGAGGAGCGTTCCCGCC

Table 2.8 General plasmids

Plasmids	Relevant characteristics	Reference
pCR2.1-TOPO	Cloning vector, *lacZα, t7 , f1ori,*	Invitrogen

	neoR,bla, pUC origin	
pGEX-6-P-1	Expression vector, for expression with Nterminal GST Tag	GE Healthcare
pSC101-BAD-γβαA-Amp	ET cloning vector	Genebridges

2.9 Cultivation of strains

2.9.1 *E. coli* strain

All the *E.coli* strains were cultured in LB medium at 37°C with the appropriate antibiotics if necessary. Generally, 1 ml of sterile LB medium were inoculated with a few cells from a single colony and cultivated at 37 °C at 180 rpm overnight.

2.9.2 Myxobacetria strains

A. discoformis An d48 was cultivated in tryptone medium for homologue growth, while knockout mutants were grown in the presence of kanamycin (50 μg/ml). For analysis of secondary metabolite production, the wild type and mutant strains were all grown in 50 ml M7 medium.

The *Cystobacter* sp. SBCb004 wild type strain was cultivated in M-medium, and all mutants were grown in the presence of kanamycin (100 μg/ml).

Strain *Myxococcus xanthus* DK1622 lived in CTT medium [110]. The tubulysin producing mutants were grown in the presence of kanamycin (50 μg/ml). All the myxobacterial strains were cultivated at 30 °C and 180 rpm.

2.9.3 *Pseudomonas putida*

Heterologous expression strain *Pseudomonads putida* FG2005 was grown in LB medium or in PMM medium at 30 °C (or 16 °C) and 180 rpm [71]. Kanamycin (50 μg/ml) was supplemented for the *P. putida* tubulysin-producing strains.

2.10 Analysis of secondary metabolite production

2.10.1 Analysis of the production from strain An d48

For analysis of secondary metabolite production, the An d48 wild type and mutant were grown in 50 ml M7 medium. 2% absorber resin Amberlite XAD-16 was added to the cultures when the first ball-shaped cell clumps became visible, and cultivation was continued for a further 5–7 days. The absorber resin and cell clumps were separated from the medium by sieving, and then extracted three times with 40 ml of a methanol/acetone mixture (1:1 *v/v*). The extracts were evaporated to dryness and resuspended in 0.5 ml methanol before analysis by high performance liquid chromatography-mass spectrometry (HPLC-MS) or using an LTQ Orbitrap Hybrid FT mass spectrometer.

Feeding studies of An d48 wild type and An d48-*tubZ*$^{-}$ mutant were performed in 50 ml M7 culture, with synthetic D-/L-pipecolic acid (mono-labeled at position C4, and mono- and di-labeled at position C5) (1 mg), d_8 L-valine (2.5 mg) and d_3 L-methionine (25 mg). The compounds were dissolved in water, sterile filtered and administered in three portions to one-day old cultures in three parts (100 µl every 12 h). The cells were cultivated for an additional day. 1% Amberlite XAD absorber resin was then added to the culture, and the cultivation was continued for another 5 days. The resin and cells were processed as described previously.

2.10.2 Analysis of the production from strain SBCb004

The SBCb004 wild type and mutant strains were grown at 30 °C for 5–7 days in 50 ml M-medium containing 2% absorber resin Amberlite XAD-16. The resin was separated from the medium by sieving, and extracted with 40 ml of methanol. The extracts were evaporated to dryness and resuspended in 1 ml methanol before analysis by mass spectrometry.

2.10.3 Analysis of the production from strain *M.xanthus* DK1622

M. xanthus strain was inoculated from an overnight culture (1:20), and cultivated at 30°C for 6 days in 50 ml CTT medium with 50 µg/ml kanamycin, 0.7 µg/ml tetracycline, and 2% absorber resin Amberlite XAD-16 (added 12 hr later after tetracycline induction) [94]. The XAD resin and cell pellets were harvested from the medium by centrifugation, and extracted with 40 ml of an acetone/methanol mixture (1:3 v/v). The extracts were evaporated to dryness in vacuum and the residues were resuspended in 1 ml methanol before analysis by HPLC-MS.

The feeding studies were performed with different amount of commercial pipecolic acid hydrochloride. This compound was first dissolved in 70% ethanol and administered in three portions to cell cultures (adding each portion per 12 hr since tetracycline induction). The cells was then cultivated for another 12 hr before adding the XAD absorber resin. For quantification experiment of strain *M. xanthus*::pTub-ATG, three independent colonies were carried out.

2.10.4 Analysis of the production from strain *P. putida*

For *P. putida*, 50ml culture in PMM medium supplemented with Km (50 µg/ml) was inoculated from an overnight pre-culture in LB medium (1:100). After cultivation at 30°C for 2 hr with shaking, the culture was induced with 0.7 µg/ml tetracyclin (final concentration) for tubulysin production. Then it was continuously grown for 4 days at 30°C (or 1 °C). In between, 12 hr later after tetracycline induction, 2% absorber resin Amberlite XAD-16 was added [71]. The absorber resin and cell pellets were harvested from the medium by centrifugation, and extracted with 40 ml of an acetone/methanol mixture (1:3 v/v). The extracts were evaporated to dryness in vacuum and the residues were resuspended in 0.5 ml methanol before analysis by HPLC-MS.

2.11 Chromatography and mass spectrometry

Standard analysis of crude extracts was performed on an HPLC-DAD system from the Agilent 1100 series, coupled to a Brucker Daltonics HCTultra ESI-MS ion trap instrument operating in positive ionization mode. Compounds were separated on a Luna RP-C_{18} column (125 × 2 mm; 2.5 µm particle diameter; flow rate 0.4 ml/min), with a mobile phase of water/acetonitrile each containing 0.1% formic acid, using a gradient from 5–95% acetonitrile over 20 min. Detection was by both diode array and ESI-MS.

High-resolution mass spectrometry was performed on an Accela UPLC-system coupled to a linear trap-FT-Orbitrap combination (LTQ-Orbitrap), operating in positive ionization mode. Separation was achieved using a BEH RP-C_{18} column (50 × 2 mm; 1.7 µm particle diameter; flow rate 0.6 ml/min), with a mobile phase of water/acetonitrile each containing 0.1% formic acid, using a gradient from 5–95% acetonitrile over 9 min. The UPLC-system was coupled to the LTQ-Orbitrap by a Triversa Nanomate, a chip-based nano-ESI interface. In certain experiments, the extracts were fractionated, and single fractions were analyzed by direct infusion.

The yield of tubulysin D was determined from the peak area (base peak chromatogram) by reference to a standard curve generated with authentic, synthetic material (a kind gift of Prof. H. Steinmetz). Estimates of the yields of all other metabolites were obtained by comparison of the peak areas to those of tubulysin D.

2.12 Gene inactivation

2.12.1 Gene manipulation in strain An d48

A kanamycin resistance gene was amplified from plasmid pCR 2.1-TOPO using the primer pair Km_for and Km_rev (Table 2.7), and the resulting fragment was cloned into vector pSUP102 to generate the plasmid pSUP-Km [40; 100], which was used as the inactivation plasmid in strain An d48.

Here, gene *tubZ* was taken as an example to show the inactivation steps in An d48 wild type strain. Primer pairs, tubZ_for and tubZ_rev (Table 2.7) were used to amplify the knockout region of gene *tubZ* from a cosmid library of *A. disciformis* An d48, introducing restriction sites for both *Bam*HI and *Hind*III. The amplified fragment was cloned into vector pCR 2.1-TOPO, and amplification of the correct product was confirmed by restriction analysis. The fragment was then excised using *Bam*HI and *Hind*III and sub-cloned into vector pSUP-Km, creating the inactivation plasmid pSUP-Km-tubZ. The inactivation vector was then electroporated into the *E. coli* strain SCS110 to enable conjugation with *A. discoformis* An d48.

The *tubZ* inactivation mutant (And48-*tubZ*$^-$) was generated using a modified triparental conjugation method [37]. An d48 wild type was cultivated at 30 ^oC and 170 rpm in tryptone medium to a cell density of 2.0×10^8–1.0×10^9 cells/ml. The cells were harvested by centrifugation at room temperature and resuspended in tryptone medium to a final concentration of 1×10^{10} cell/ml. The *E. coli* strains SCS110/pSUP-Km-tubZ and the helper strain HB101/pRK600, were cultivated at 37 ^oC to an OD_{600} 0.8–1.2. The *E. coli* were then harvested at room temperature, washed 3 times and resuspended in LB medium to yield a cell density of 1×10^{10} cells/ml. 10^9 cells of An d48, SCS110/pSUP-Km-tubZ and HB101/pRK600 were then combined. The mixture was concentrated by centrifugation and the cell pellet was resuspended in 100 µl tryptone medium, followed by spotting onto a tryptone agar plate. The plates were then

incubated at 37 °C overnight. The mixed cells were scraped off the plate and resuspended in 1 ml tryptone medium. The resuspended cells were mixed with 3 ml of tryptone soft agar containing antibiotics (tryptone medium + 0.75% agar + 50 μg/ml kanamycin + 30 μg/ml gentamycin), and immediately plated on tryptone agar plates (tryptone medium + 1.5% agar + 50 μg/ml kanamycin + 30 μg/ml gentamycin). The plates were incubated at 30 °C for 1–2 weeks until colonies became visible.

2.12.2 Gene manipulation in strain SBCb004

The construction of gene knockout in SBCb004 was accomplished by homologous recombination in the encoding regions. Oligos were designed to amplify the knockout region in the middle of target gene and also introduce a stop codon. Primer pairs, Cbt-Z1 and Cbt-Z2, were designed to amplify the knockout region (864 bp) in the middle of target gene *tubZ*, and introduce a stop codon. The amplified DNA fragment was then cloned into pCR 2.1-TOPO vector to obtain the inactivation vector pTOPO-tubZ. Correct integration of plasmid pTOPO–tubZ into the chromosome of SBCb004 was confirmed using three different PCR reactions with primers pTOPO-In, pTOPO-Out, Cbt-Z3 and Cbt-Z4. Primer pair ctubA-a and ctubA-b were used to knockout gene *tubA*, while corf1-a and corf1-b were used to inactivate gene *orf1* (Table 2.7).

The inactivation mutant was generated by electroporation as follows: SBCb004 wild type cells were grown to an OD_{600} 0.8–1.0, and cells were harvested from 2 ml of culture by centrifugation at room ttemperature. The cell pellet was then washed twice with an equal amount of washing buffer (5 mM HEPES (pH 7.2), 0.5 mM $CaCl_2$), and resuspended in 50 μl ddH_2O. Plasmid pTOPO-tubZ(tubA, orf1) (1–3 μg) was then added into the cell suspension, and the mixture was electroporated in a 0.1 cm cuvette (200 Ω, 25 mF, and 0.9 kV/cm). 1 ml M-medium was added to the cells. The suspension was transferred into a 2 ml-centrifuge tube and cultivated for 6 h at 30 °C and 800 rpm on a thermomixer. The cells were then mixed with 1 ml M-medium, and plated on M-agar with 100 μg/ml kanamycin. The plates were incubated at 30 °C for 1–2 weeks until colonies became visible.

The insertion of Tn5 promoter into strain SBCb004 was also carried out similarly. Primer Tn5-a and Tn5-b were used to amplified the Tn5 promoter, while primer ctubB-a and ctubB-b were used to obtained the 1 kb *tubB* encoding region (*tubB'*). An overlap PCR with primer Tn5-a and ctubB-b was then carried out to connect both the Tn5 promoter and the *tubB* encoding region. The resulted DNA fragment *Tn5-tubB'* was electroporated into SBCB004 strain subsequently.

When the homologous recombination happened in the *tubB* encoding region, the Tn5 promoter was directly inserted in front of the core NRPS-PKS genes of tubulysin gene cluster, and resulted in mutant SBCb004-Tn5. The integrations of plasmids into SBCb004 chromosome for all these inactivation mutants were confirmed by PCR reactions as described before, while the Tn5 insertion mutant was verified by PCRs and sequencing.

Those identified monooxygenase genes from SBCb004 were also inactivated by the same method with specially designed primer pairs.

2.13 TubZ from An d48

2.13.1 Cloning, expression and purification of TubZ

The *tubZ* gene of An d48 was amplified from cosmid F5 [102] containing the *Angiococcus disciformis* An d48 tubulysin gene cluster (primer pair TubZ-exp-f and TubZ-exp-r), introducing flanking *Eco*RI and *Not*I sites. The resulting product was then with *Eco*RI and *Not*I, and ligated into pGEX-6P-1, previously digested with both *Eco*RI and *Not*I. The fidelity of the PCR reaction was then confirmed by sequencing. The resulting expression construct pGEX-6P-1-TubZ was transformed into *E. coli* strain BL21(DE3)pLyS. The cells were grown in LB medium (1 L) containing ampicillin (100 mg/ml) at 30 ^{o}C until an OD_{600} of 0.6–0.7 was reached, and then expression was induced by addition of 0.1 mM IPTG. The cultivation was continued at 16 ^{o}C overnight, and then the cells were harvested by centrifugation. The cell pellets were resuspended in ice-cold lysis buffer (25 ml; 1× PBS buffer containing 5 mM DTT and 0.5% Triton X-100) and lysed by French Press. The lysis solution was centrifuged (14,000 *g*) at 4°C for 30 min, and then the clarified lysate was mixed with a 50% slurry of Glutathione Sepharose 4B beads (2 ml), and incubated at room temperature for 30 min, with gentle agitation. The beads were then washed with buffer (1× PBS buffer; 10 column volumes). In the case of the GST-TubZ fusion protein, elution was performed directly with elution buffer (50 mM Tris.HCl (pH 8.0), 10 mM reduced glutathione); the protein was eluted in three fractions (10 min incubation). Discrete TubZ was obtained by on-column digestion of the bound GST-TubZ fusion protein. For this, PreScissionProtease® (5 units) was added to the beads in protease buffer, and incubated at 4 °C overnight. The purified proteins were flash frozen in liquid nitrogen, and stored at −80°C until use.

2.13.2 Characterization of TubZ cyclodeaminase activity

Assays contained TubZ–GST fusion protein (1.5 µM) or discrete TubZ protein (0.6 µM), L- or D-lysine (1 mM), NAD^+ (100 M), and BSA (1 mg/ml) were performed in buffer (5 mM Tris.HCl (pH 8)) in a total volume of 100 µl. Control incubations were performed with L- or D-lysine in the absence of added TubZ. After 24 h of incubation at room temperature, the reactions were quenched by the addition of acetonitrile (200 µl). The assay mixture was chilled at −20 °C for at least 10 min, and then centrifuged (16060 *g*) at 4°C for 10 min. Derivatization was carried out on the assay supernatant (240 µl), using Fmoc-Cl (10 mM; 20 L) in borate buffer (10 mM; 80 L). The reaction mixture was incubated at room temperature for 30 min, and then quenched by addition of 2 volumes of pure acetic acid. The same derivatization procedure was performed on a commercially available racemic mixture of L-/D-pipecolic acid, and enantiomerically pure D-pipecolic acid. Following centrifugation, the supernatants were analyzed using the HPLC-DAD system at 25 °C. A Nucleodex-OH column (200 × 4 mm) was employed for separation with a solvent system consisting of H_2O and acetonitrile, each containing 0.1% formic acid. The following gradient was applied: 30–95% acetonitrile over 35 min, at a flow rate 0.5 ml/min.

2.14 Cloning by Red/ET recombination

In all the Red/ET recombination experiments, linear DNA fragments working as the modification cassette were obtained either by PCR or by restriction. KOD DNA polymerase was used to run the PCR reactions according to the manufacturer's protocol, and the products were concentrated and purified by ethanol precipitation before further applications.

E.coli strain GB05-red and YZ2005 were used in the recombineering experiments [76; 111]. They were cultured in LB medium at 37°C with the appropriate antibiotics (kanamycin, 30 µg/ml; ampicillin, 100 µg/ml; gentamycin, 6 µg/ml; zeocin, 20 µg/ml; chloramphenicol, 25 µg/ml). Low salt LB was used to screen clones harboring Cm or Zeo resistance gene.

Generally, the GB2005-red cells harboring parent vector to be modified were inoculated by an overnight culture (1:40), and cultivated on a thermomixer at 37°C and 1000 rpm. When OD_{600} value reached 0.2, 0.1% L-arabinose was added into the culture to induce the recombinases expression. The cultivation was continued until OD_{600} 0.4. The cells were then harvested by centrifugation, washed twice with ice-cold dH_2O, and resuspended in 40 µl dH_2O. 0.2–0.3 µg PCR product was mixed with the electrocompetent cells and transferred into an ice-cold electroporation cuvette (1

mm). The mixture was shot at 1300 V by an electroporator, followed by adding 1 ml LB medium immediately. It was cultivated at 37°C and 1000 rpm for 75 min before the culture were concentrated and spread on LB agar plates with suitable selection antibiotic. Plates were incubated at 37°C overnight. The selection marker located on the introduced linear DNA fragment can be used to select the correct transformats. Among all these Red/ET recombination reactions, all generated constructs were either verified by restriction analysis or confirmed by sequencing.

2.15 Reconstitution of tubulysin biosynthetic gene cluster

Before stitching the tubulysin gene cluster together, the two cosmids Cos DL18 and Cos AP17 were firstly modified by Red/ET recombination. The 5′- end tub genes on Cos DL18 were cloned into a minimal linear cloning vector *p15A-IR-Km*, flanked by a *NheI* site, and resulted in plasmid pUp. Oligonucleotides used here were 15A-Km-tub5 and 15A-Km-tub3. Two rounds of recombineering were used to engineer the tub genes on Cos AP17. Firstly, a minimal cloning vector *pUC-Amp*, flanked by a *NheI* site was designed to replace the pSuperCos backbone and the overlapped *tubC* portion on Cos AP17 with primers ask-tub5 and ask-tub3. Then, a cassette of zeocin resistance gene and *NheI* site (amplified by oligos Zeo-tub5 and Zeo-tub3) was subcloned onto the above plasmid, resulting in plasmid pDown. After that, the two modified vectors were digested by enzyme *NheI*. Vector pDown was cut into two fragments, and the bigger one (29 kb) was stitched with the linearized pUp by T4 DNA ligase. The ligated DNAs were electroporated into the GB05-red competent cells, and the circularized plasmids harboring the full-length tub gene cluster in correct direction were screened by zeocin and verified by restriction analysis. For expressing in heterologous expression host strain *P. putida* or *M. xanthus*, a conjugation/ transposition cassette *IR-Tps-Amp-oriT* was generated by oligos Tps-amp5 and Tps-amp3, and inserted into plamid p15A-Tub. Ampicillin was used to screen the correct clones. Besides, in order to modify the start codon of gene *tubC*, two parallel cassettes *Genta-TetR-*P_{tet}*-TTG* and *Genta-TetR-*P_{tet}*-ATG* were generated separately by primer pair Genta-tub5 and TetR-tub-TTG (for TTG), and primer pair Genta-tub5 and TetR-tub-ATG (for ATG). They were inserted directly in front of gene *tubC*. The resulted final clones with constructs pTub-TTG or pTub-ATG were selected by both Genta and Amp. Among all these Red/ET recombination reactions, all generated constructs were either verified by restriction analysis or confirmed by sequencing.

2.16 Transformation of tubulysin gene cluster into the heterologous expression host strains

2.16.1 Electroporation of *M. xanthus* DK1622

The transformation of tubulysin gene cluster into the chromosome of *M. xanthus* DK1622 was carried out by electroporation in a 0.1 cm cuvette at 400 Ω, 25 mF, and 0.8 kV/cm. The Km-resistant DK1622 colonies should typically appear after 6 days.

2.16.2 Conjugation of *P. pitida* FG2005

The tubulysin gene cluster were integrated into the genome of *P. putida* FG2005 [93] by triparental conjugation with *E. coli* strain GB2005-red (containing plasmid harboring the entire tub gene cluster), *E. coli* HB101 (containing the helper plasmid pRK2013), and *P. putida* FG2005 [112]. Positive clones were selected on PMM agar plates supplemented with kanamycin at 30^{o}C.

2.16.3 Verification of the plasmid integration

In both *M.xanthus* DK1622 and *P.putida* FG2005 strains, the successful integrations of the tub gene cluster into host strain chromosome were verified by two sets of colony PCRs with specially designed primer pairs: primer pair Cbt-A3 and Cbt-A4 located at 5′- end of tub gene cluster were used to amplify the *tubA-orf2* region (2.4 kb); while primer pair Cbt-H3 and Cbt-G4 located at 3′-end of tub gene cluster were used to amplify the *orf17-orf18* region (2.5 kb).

2.17 Genetic modification of tubulysin heterologous expression system in *E.coli*

A serious of similar cassettes of chloramphenicol resistance gene were generated by primer pairs with varied homology arms. Primer orf2-for and orf2-rev were used to inactive gene *orf2*, primer tubZ-for and orf2-rev were used to delete both *tubZ-orf2* genes, and primer orf1-for and orf2-rev were used to replace the whole operon *orf1-tubZ-orf2*. Primer tubA-for and tubA-rev were used in gene *tubA* deletion. These cassettes were introduced into plasmid pTub-ATG to replace those target genes, and the resulting deletion vectors were screened by both Cm and Amp.

2.18 Co-expression of tubulysin gene cluster and gene *633P1*

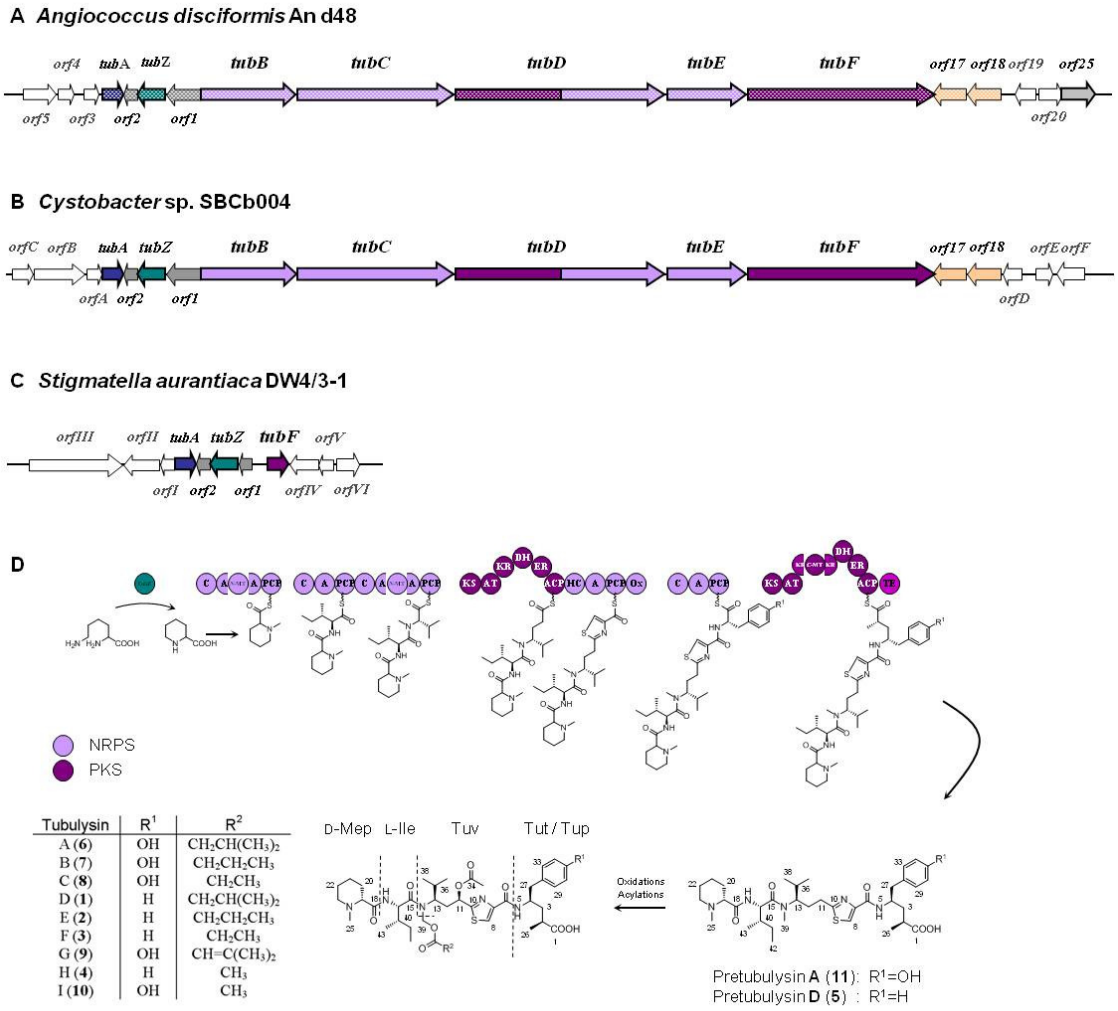
Cassette *Zeo-P_{tet}-633P1* was designed to add gene *633P1* in front of the tub gene cluster on the heterologous expression construct pTub-ATG. Gene *633P1* was obtained by PCR from SBCb004 genome with primer 633P1-1 and 633P1-2, and subsequently cloned into pCR2.1-TOPO vector. The pTOPO-633P1 plasmid was linearized by *EcoRI* digestion. DNA fragment *Zeo-P_{tet}* was amplified by primer P450-zeo5 and P450-zeo3. Both linear DNA fragments *pTOPO-633P1* and *Zeo-P_{tet}* were connected together by the linear to linear recombination in *E.coli* strain YZ2005 [76]. The resulted plasmid pTOPO-Zeo-P_{tet}-633P1 was screened by zeocin. Restriction enzyme *EcoRI* was then used again to release the DNA fragment *Zeo-P_{tet}-633P1*. One step Red/ET recombination was carried out to insert this cassette onto the vector pTub-ATG. The resulted positive clones harboring plasmid pTub-633P1 were selected on low salt LB agar plates containing Amp and Zeo.

3 RESULTS AND DISCUSSION

3.1 Discovery and annotation of the tubulysin biosynthetic gene cluster in *Cystobacter* SBCb004

Sequencing of the ca. 40 kb tubulysin gene cluster within the strain *Angiococcus disciformis* An d48 revealed that tubulysin assembly is catalyzed by a molecular 'assembly line', consisting of PKS and NRPS multienzymes, and a hybrid PKS-NRPS subunit [102]. The gene set does not incorporate any obvious functions for carrying out the required post-assembly line oxidation and acyl transfer reactions. To obtain further insights into tubulysin biosynthesis, I aimed to identify the gene cluster in a second myxobacterial strain, *Cystobacter* sp. SBCb004.

In common with *Archangium gephyra* Ar315, *Cystobacter* SBCb004 generates tubulysins A, B, C, G and I, incorporating tubutyrosine instead of the tubuphenylalanine of tubulysins D, E, F and H (Figure 1.13). To locate the tubulysin gene cluster in *Cystobacter* SBCb004, the SBCb004 genome was shot-gun sequenced using 454 technology, and resulted in 276 contigs [113; 114]. Then, the data were analyzed by BLAST [115] using the *tub* genes from An d48 [102]. Finally, genes within an 82 kb region were functionally annotated. The SBCb004 tubulysin gene cluster closely resembles that in An d48. Comparison of the tubulysin biosynthetic gene clusters in SBCb004 and An d48 reveals a conserved architecture. By identification of the gene set common to both SBCb004 and An d48, the cluster boundaries was able to be defined straightforward in both organisms for the first time. The clusters each incorporate 11 genes, which are present in the same order and orientation (Figures 3.1 and Table 3.1). The core biosynthetic genes (*tubB–tubF*) encode three NRPS subunits, one PKS multienzyme, and a hybrid PKS-NRPS protein. Two conserved orfs (*orf17* and *orf18*) are present downstream of gene *tubF*, while *tubB* is preceded by genes *tubA*, *orf2*, *tubZ* and *orf1*. Like its An d48 counterpart, no genes clearly encoding post-PKS tailoring functions were present in the SBCb004 tubulysin cluster.



Tubulysin	R^1	R^2
A (**6**)	OH	$CH_2CH(CH_3)_2$
B (**7**)	OH	$CH_2CH_2CH_3$
C (**8**)	OH	CH_2CH_3
D (**1**)	H	$CH_2CH(CH_3)_2$
E (**2**)	H	$CH_2CH_2CH_3$
F (**3**)	H	CH_2CH_3
G (**9**)	OH	$CH{=}C(CH_3)_2$
H (**4**)	H	CH_3
I (**10**)	OH	CH_3

Figure 3.1 Comparison of tubulysin biosynthetic gene clusters in An d48 and SBCb004.
(A) Genetic organization of the tubulysin gene cluster in *Angiococcus disciformis* An d48. Based upon the newly-determined cluster boundaries, gene *tubG* [102] was redesignated as *orf25*, reflecting the fact that it is not part of the tubulysin pathway.
(B) Genetic organization of the tubulysin gene cluster in *Cystobacter* sp. SBCb004.
(C) Genetic organization of the tubulysin gene cluster fragment in *Stigmatella aurantiaca* DW 4/3-1.
(D) Model for tubulysin biosynthesis, showing the domain organization of the PKS-NRPS assembly line and the structures of the chain extension intermediates. The structures of the known tubulysins are shown, as well as that of the intermediate structure pretubulysin A and D.

Gene *orf1* is predicted to encode an anion-transporting ATPase, while the *orf2* gene product exhibits similarity to hypothetical proteins. Mutual sequence homology between the SBCb004 and An d48 homologues is high (Orf1: 78% identity/90% similarity; Orf2: 69% identity/82% similarity vs. 68–73% identity/78–81% similarity for the PKS/NRPS proteins) which suggests

evolutionary pressure to maintain their functions. However, neither protein has an obvious role to play in tubulysin biosynthesis.

Table 3.1 Proteins encoded in the sequenced region in the SBCb004 tubulysin biosynthetic gene cluster and their putative functions[a]

NRPS/PKS portion of the gene cluster					
Protein	aa	Proposed function (Protein domains with their position in the sequence)[b]			
TubB	1520	NRPS domains: C (62–493), A (494–1416), *N*-MT (971–1368), PCP (1432–1500)			
TubC	2628	NRPS domains: C (77–515), A (516–1027), PCP (1044–1112), C (1136–1569), A (1570–2499), *N*-MT (2042–2451), PCP (2514–2582)			
TubD	3511	PKS domains: KS (11–441), AT (527–852), KR (939–1370), DH (1371–1652), ER (1678–1980), ACP (2062–2157) NRPS domains: HC (2190–2624), A (2625–3138), PCP (3153–3223), Ox (3231–3509)			
TubE	1161	NRPS domains: C (47–482), A (483–1034), PCP (1051–1121)			
TubF	2843	PKS domains: KS (5–436), AT (522–847), KR (925–1721), *C*-MT (1309–1540), DH (1722–1997), ER (2023–2322), ACP (2404–2506), TE (2541–2815)			
ORFs encoded upstream and downstream of *tubB–tubF*					
Protein	aa	Proposed function of the homologous protein	Source of the homologous protein	Similarity/ identity	GenBank accession no.
OrfC	735	Outer membrane protein, OMP85 family, putative	*Stigmatella aurantiaca*	71%/57%	EAU64371.1
OrfB	1549	Hypothetical protein MXAN_6465	*Myxococcus xanthus*	65%/48%	ABF91418.1
OrfA	273	α/β hydrolase fold	*Geobacter uraniireducens*	91%/81%	ABQ27606.1
TubA	432	TubA	*Stigmatella aurantiaca*	89%/81%	EAU66910.1
Orf2	219	Conserved hypothetical protein	*Stigmatella aurantiaca*	86%/74%	EAU66945.1
TubZ	386	TubZ	*Stigmatella aurantiaca*	91%/86%	EAU66930.1
Orf1	394	Hypothetical protein (Orf1)	*Angiococcus disciformis*	90%/78%	CAF05646.1
Orf17	337	Patatin-like protein (Orf18)	*Angiococcus disciformis*	77%/62%	CAF05653.1
Orf18	368	Patatin-like protein (Orf18)	*Angiococcus disciformis*	82%/68%	CAF05653.1
OrfD	124	Chloroplast heat shock protein 70	*Pennisetum glaucum*	50%/33%	ABP65327.1
OrfE	119	ChpK	*Stigmatella aurantiaca*	93%/89%	EAU64035.1
OrfF	271	Protein kinase	*Stigmatella aurantiaca*	49%/33%	EAU63389.1

[a]For the corresponding proteins in the *A. disciformis* An d48 gene cluster, please see [102].
[b]Domain boundaries were assigned, when possible, based on the solved structures of the individual domains.

The product of gene *tubA* shows similarity to NRPS C domains [102]. Notably, a sequence (DHxxxDx) within core motif C3 is closely similar to a proposed signature sequence (HHxxxDG) for enzymes which effect acyl transfer, such as chloramphenicol acyltransferase and dihyrolipoamide acyltransferase [116]. Thus, TubA may serve as one of the missing acyl transferases in the tubulysin pathway. In support of a shared role for the TubA proteins in the two strains, the SBCb004 and An d48 homologues show a high degree of sequence conservation (56% identity/73% similarity). However, the proposed functions remain to be directly demonstrated.

orf17 and *orf18* both encode patatin-like proteins, which show high mutual sequence homology (60% identity/70% similarity in the case of the SBCb004 proteins), suggesting evolution by gene duplication. Interestingly, the closest homologue to both Orf17 and Orf18 in SBCb004 is Orf18 of An d48, which may imply that *orf17* of An d48 has diverged from the other sequences. Patatin-like proteins exhibit lipid acyl hydrolase activity [117; 118], a function apparently not required in the tubulysin pathway. Intriguingly, however, genes encoding patatin-like proteins are also present in the DKxanthene gene clusters of the myxobacteria *Myxococcus xanthus* DK1622 and *Stigmatella aurantiaca* DW4/3-1 [47], suggesting that the enzymes may play a previously unsuspected role in natural product biosynthesis.

Predictably from the defined boundaries, *orfs* upstream of *tubA* and downstream of *orf18* in both organisms have no obvious function in the tubulysin pathway. Thus, the oxidase activities, and possibly a second acyl transferase enzyme in addition to TubA, are apparently lacking in both clusters, and so are presumed to be located elsewhere in the genome. Such split cluster PKS/NRPS architecture is unusual for *Streptomyces*, but has precedent in the myxobacteria [119-121]. The fact that this split organization is conserved in two different species argues for the presence of the tubulysin genes in an ancestor common to both *Cystobacter* SBCb004 and *Angiococcus* An d48 – that is, that the cluster evolved into its present form prior to the speciation event within the Cystobacterineae suborder. Phylogenetic analysis based on 16S rDNA supports this hypothesis, as An d48 and SBCb004 belong to the same family, sharing membership with another potential tubulysin producing-strain *Archangium gephyra* Ar g1 (Figure 3.2).

Analysis by BLAST also revealed a portion of the tubulysin gene cluster comprising *tubA*, *orf2*, *tubZ*, and fragments of *orf1* and *tubF*, in the publically-available genome of a third myxobacterial strain, *Stigmatella aurantiaca* DW 4/3-1 (Figure 3.1C). This architecture may have arisen from

deletion of the intervening genes (*tubB–tubE*) from an ancestral cluster. Interestingly, the *Cystobacter* TubA, Orf2 and TubZ proteins show higher sequence identity to their *Stigmatella* homologues than to those of *Angiococcus* (Table 3.1), despite the closer evolutionary relationship between *Cystobacter* and *Angiococcus* (as judged on the basis of 16S rDNA) (Figure 3.2). Consistent with the fact that only part of the gene cluster is present in *S. aurantiaca*, the strain is not known to produce any of the tubulysins.

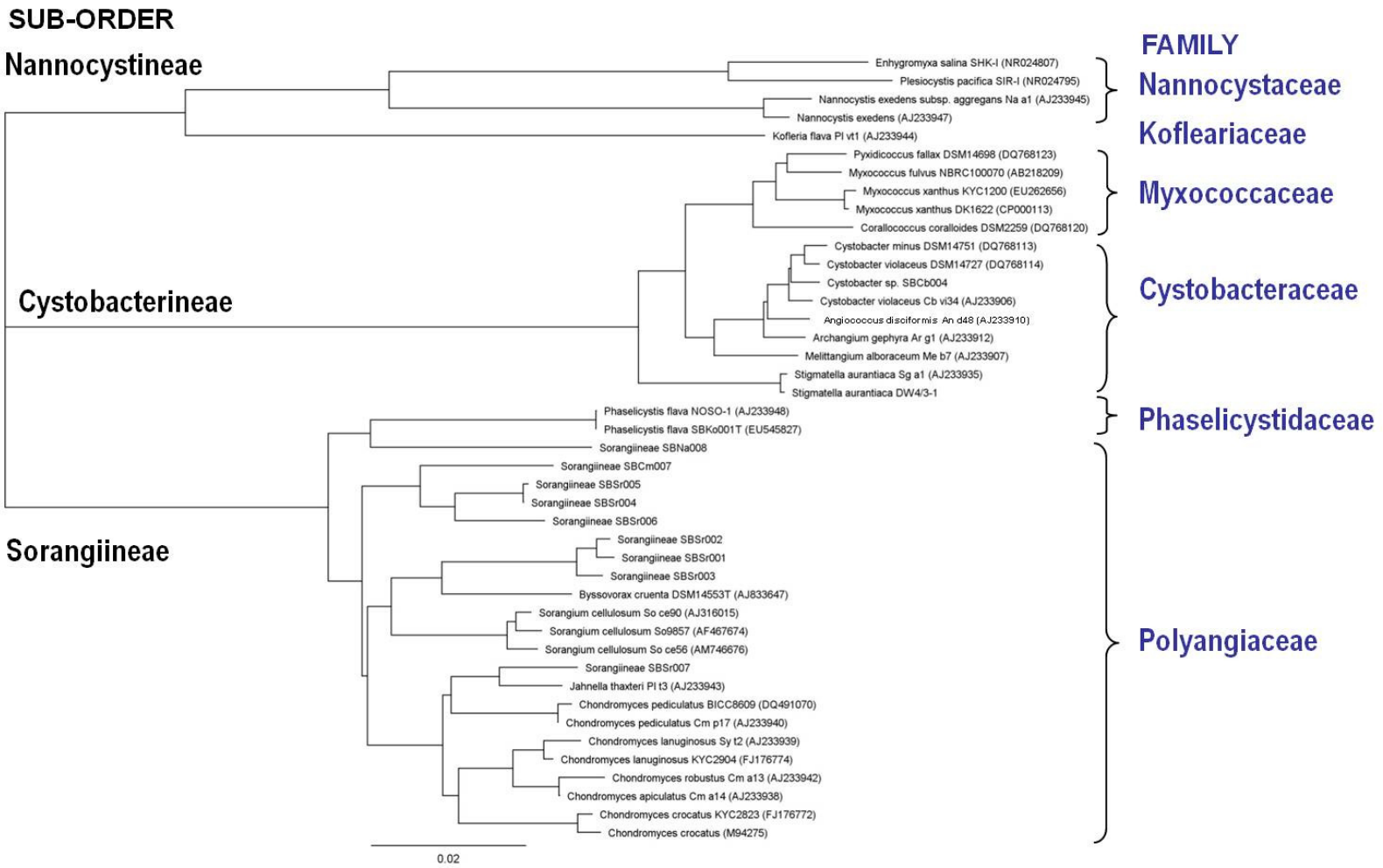


Figure 3.2. Phylogenetic Tree of Selected Myxobacterial Strains Based on 16S rDNA Sequences
The three myxobacterial sub-orders and six families are shown. The known tubulysin producers *Angiococcus disciformis* An d48 and *Cystobacter* sp. SBCb004 belong to the same family, Cystobacteracea. The branch length indicates the number of inferred base pair changes per position.

3.2 Characterization of gene *tubZ* and the stereochemistry of the pipecolic acid starter unit in the tubulysin pathway

Although the tubulysin biosynthetic gene cluster does not incorporate any obvious function for carrying out the required post-assembly line oxidation and acyl transfer reactions, I identified a gene encoding a lysine cyclodeaminase (*tubZ*), whose protein product is likely to be involved in the biosynthesis of the unusual amino acid pipecolate, the presumed starter unit for tubulysin assembly [122; 123]. To demonstrate directly the involvement of *tubZ* in the pathway, gene inactivation by insertional mutagenesis in both An d48 and SBCb004 strain as well as TubZ protein expression and *in vitro* assay were carried out.

3.2.1 Inactivation of *tubZ* in An d48

The gene *tubZ* was knocked out form strain An d48 by homologous recombination. The plasmid pSUP-Km was designed and generated as the inactivation vector, and triparental conjugation was used to transform the inactivation plasmid pSUP-Km-tubZ into strain An d48. Integration of plasmid pSUP-Km-tubZ into An d48 chromosome was demonstrated by three different PCR reactions, using the genomic DNA of An d48 wild type and five exconjugants as templates (Figure 3.3). For these PCR reactions, one set of primers (primers 1 and 2) was designed to bind to the backbone of vector pSUP-Km-tubZ. The other set of primers (A and B) was designed to bind to the genomic DNA adjacent to the expected insertion site (Table 2.7). In the An d48-*tubZ*¯ knockout mutants, PCR using primers 1 and A or primers 2 and B, yielded products of 1.29 kb and 2.0 kb, respectively, consistent with the integration of the knockout plasmid at the expected location (Figure 3.3). In contrast, PCR of the wild type strain using primers 1 or 2 did not yield product, as predicted, due to the absenced of a primer binding site in the genomic DNA. However, a control reaction on the wild type employing primers A and B resulted in the expected band of 2.26 kb, while no product was obtained with this primer combination from the mutant (the product would be 10 kb in length, too long for amplification under the conditions used). Taken together, the PCRs clearly demonstrate that all five analyzed And48-*tubZ*¯ mutants contained the expected knockout in *tubZ*.

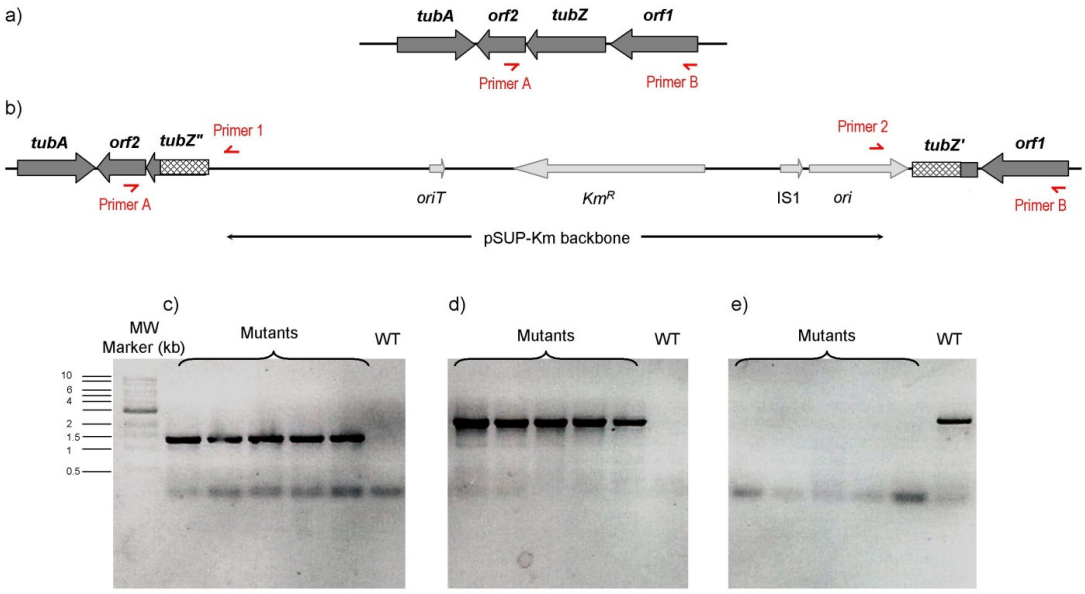


Figure 3.3 Confirmation of the inactivation of *tubZ* in An d48 by PCR.
(A) Genetic organization of chromosomal DNA in *Angiococcus disciformis* wild type.
(B) Genetic organization of chromosomal DNA in mutant An d48-*tubZ*⁻. From right to left: *tubZ* truncated at the 3′-end (*tubZ*′), pSUP-Km vector backbone, *tubZ* truncated at the 5′-end (*tubZ*″).
(C) PCR reaction with primers 1 and A, yielding a product of 1.29 kb from the mutants.
(D) PCR reaction with primers 2 and B, yielding a product of 2.0 kb from the mutants.
(E) PCR reaction with primers A and B. Only the wild type produces a product (2.26 kb). In the figure, the designation 'WT' represents the wild type of An d48, and 'Mutants' represents five distinct And48-*tubZ*⁻ exconjugants.

As this mutation was expected to interrupt the supply of pipecolic acid, it was anticipated that the resulting strain would no longer produce tubulysins. To our surprise, analysis by HPLC-MS revealed tubulysin D in extracts of the mutant An d48-*tubZ*⁻ at approximately 3% of wild type levels, but in addition, substantial amounts of a novel metabolite (*m/z* = 670.4 Da); tubulysins E, F and H were not detected. This result strongly implicates TubZ in provision of the pipecolic acid moiety, but suggests there is a second lysine cyclodeaminase function in *A. disciformis*.

The relationship of metabolite 670.4 to the tubulysins was first confirmed by supplementation of culture broths with commercially-available d_8 L-valine, yielding an incorporation pattern analogous to that of tubulysin D (Figure 3.4). Accurate mass determination on compound 670.4 using a Thermo LTQ Orbitrap Hybrid FT mass spectrometer gave a value of 670.399 Da, consistent with a molecular formula of $C_{36}H_{55}N_5O_5S$ (calculated $[M+H]^+$= 670.400 Da, Δ= –0.939 ppm). As the metabolite appeared to have the structure of the proposed first enzyme-free intermediate in the pathway – that is, the polyketide-nonribosomal polypeptide core minus the

four post-assembly line oxidative and acylation reactions – designating it as 'pretubulysin D'. Furthermore, reanalysis of extracts of wild type *A. disciformis* An d48 revealed low levels of pretublysin D (Figure 3.5A and B), supporting its intermediacy in the pathway. Therefore, it was reasoned that pretubulysin might represent a stable analogue of the more complex tubulysins, and would thus be well suited for evaluation as a drug candidate.

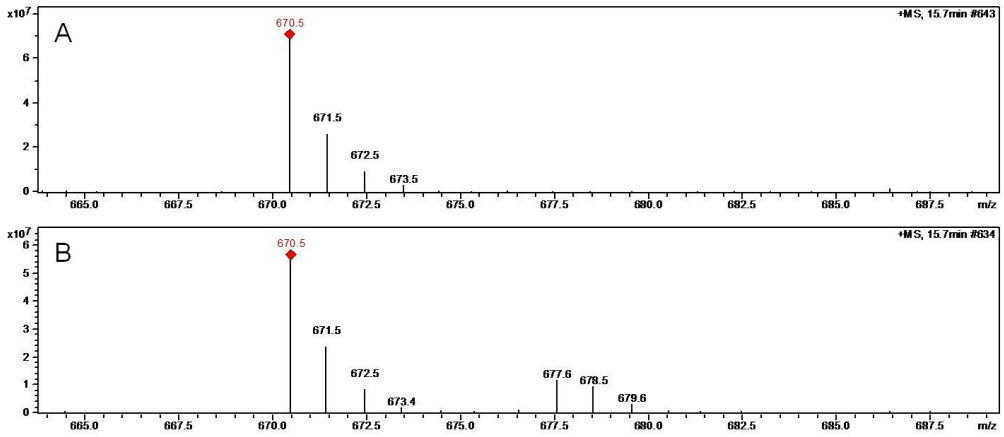


Figure 3.4 Feeding of d_8 L-valine into mutant An d48-*tubZ*$^-$.
(A) HPLC-MS analysis of compound 670.4, none feeding control sample.
(B) HPLC-MS analysis of compound 670.4, feeding with d_8 L-valine.

I aimed to prove the structure of pretublysin D by NMR. However, despite growth of *A. disciformis* on a 20 L scale, it was unable to purify sufficient quantities of of the compound. I therefore used tandem mass spectrometry (MS^2–MS^6) to probe the structure, and compared the fragmentation pattern to that of tubulysin D (the MS^2 data are presented in Figure 3.5C and D). The patterns showed little similarity, suggesting that the presence of the acyl groups in tubulysin D significantly influenced the mode of fragmentation. Nonetheless, accurate mass analysis of pretubulysin D fragments revealed molecular formulae consistent with the proposed structure.

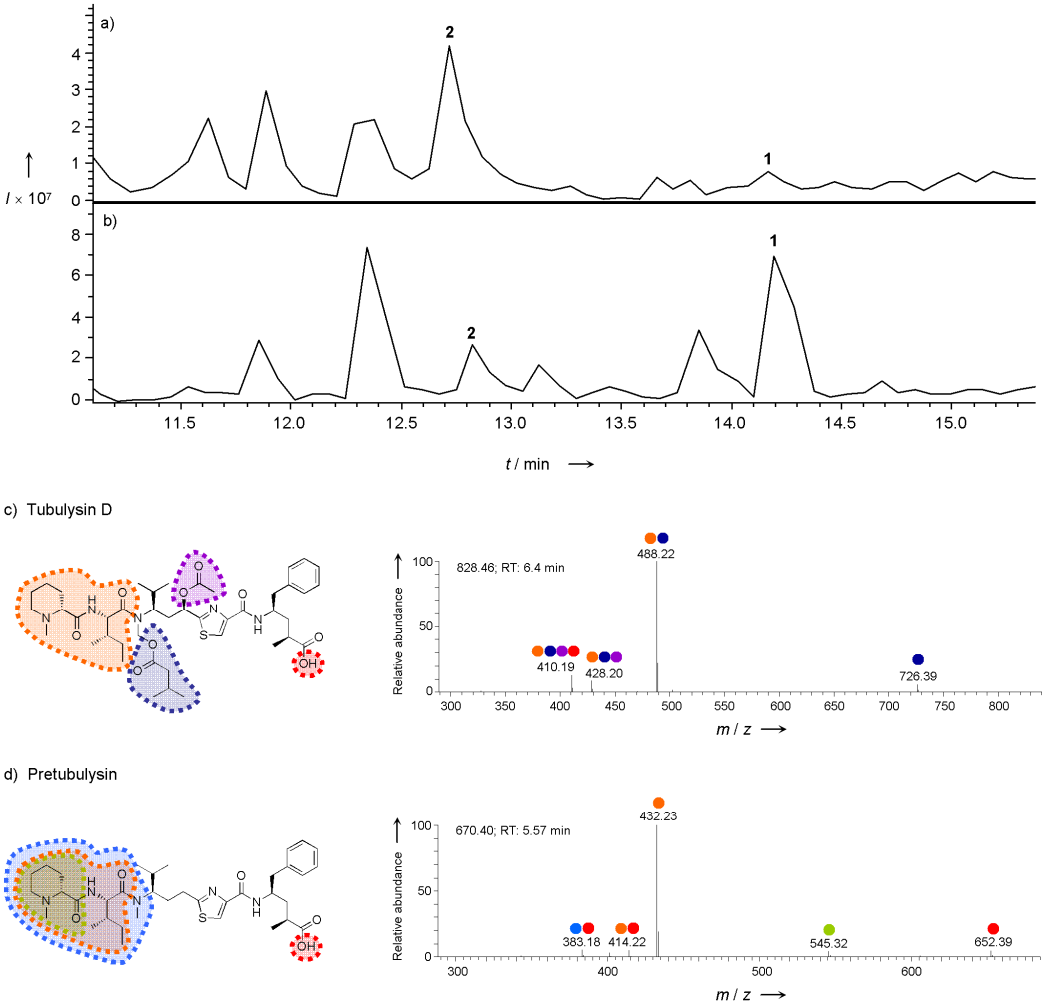


Figure 3.5 Identification of pretubulysin D.
(A) HPLC-MS analysis (base peak chromatogram (BPC)) of mutant An d48-*tubZ*$^{-}$. Peaks corresponding to tubulysin D (**1**) and pretubulysin D (**5**) are indicated.
(B) HPLC-MS analysis (BPC) of extracts of wild type *A. disciformis* An d48, showing peaks corresponding to **1** and **5**.
(C) Comparative analysis of the MS^2 fragmentation patterns of **1.**
(D) Comparative analysis of the MS^2 fragmentation patterns of **5**. Comparable fragments lost from each metabolite are indicated using the same color. The mass spectra include the retention time and molecular mass of the respective parent ions, as well as colored labels to indicate the fragments lost to generate each daughter peak. All data were obtained on a Thermo LTQ Orbitrap Hybrid FT mass spectrometer.

Luckly, our cooperators in Prof. Kazmaier`s group succeeded to synthesize the compound pretubulysin D [124; 125]. To provide final proof for this biosynthetic hypothesis and to investigate the biological activity of pretubulysin in more detail, this synthesized standard compound was used as a comparison template. All mass spectrometric data (MS^2–MS^6) obtained

on the biosynthetic material were essentially identical to those from analysis of authentic, synthetic pretubulysin, confirming the compound's identity (data now shown).

3.2.2 Inactivation of *tubZ* in SBCb004

In order to confirm the function of *tubZ* in strain SBCb004, a similar inactivation mutant was also generated in this strain. A compound with a mass (*m*/*z* 686.4 $[M+H]^+$) consistent with pretubulysin A was identified in extracts of the *Cystobacter* sp. SBCb004 wild type strain, at 42% yield relative to tubulysin A. As the An d48-*tubZ*$^-$ mutant produced increased amounts of pretubulysin D relative to wild type *A. disciformis* [126], I hope to boost the titer of pretubulysin A in SBCb004 by engineering the equivalent *tubZ* mutation. HPLC-MS analysis of the resulting mutant, SBCb004-*tubZ*$^-$, revealed that the strain accumulated pretubulysin A at approximately 5-fold wild type levels, whereas the production of tubulysin A was decreased by 20-fold (Figure 3.6).

The availability of the genome sequence of SBCb004 allowed it possible to probe for the presence of additional lysine cyclodeaminases in the strain which could complement the inactivated TubZ. BLAST analysis revealed two genes with homology to ornithine cyclodeaminases (20% and 25% sequence identity to TubZ) (Table 3.2). As RapL exhibits ornithine cyclodeaminase activity as a side reaction [123], it seems reasonable that the detected ornithine cyclodeaminases may catalyze low levels of lysine cyclodeamination, explaining the continued presence of pipecolic acid in the SBCb004-*tubZ*$^-$ mutant. Given the evolutionary relationship between *Cystobacter* SBCb004 and *Angiococcus disciformis* An d48, it is likely that An d48 harbors homologues of these two genes.

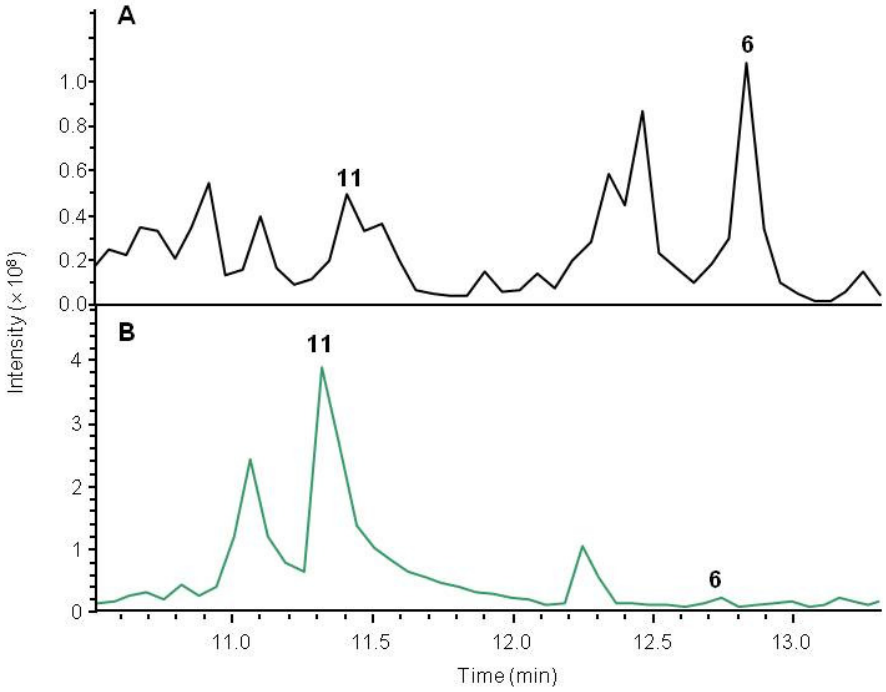


Figure 3.6 Production of tubulysin A (6) and pretubulysin A (11) by SBCb004 wild type and mutant SBCb004-*tubZ*⁻.
(A) HPLC-MS analysis (base peak chromatogram (BPC)) of wild type SBCb004. Peaks corresponding to tubulysin A (**6**) and pretubulysin A (**11**) are indicated.
(B) HPLC-MS analysis (BPC) of extracts of mutant SBCb004-*tubZ*⁻. Peaks corresponding to tubulysin A (**6**) and pretubulysin A (**11**) are indicated.

Table 3.2 The cyclodeaminases annotated in the genome of SBCb004.

Protein	aa	Proposed function of the homologous protein	Source of the homologous protein	Similarity/ identity	GenBank accession no.
TubZ	386	TubZ protein	*Stigmatella aurantiaca* DW4/3-1	91% / 86%	ZP_01462307.1
Contig 124-Orn	320	ornithine cyclodeaminase /mu-crystallin family protein	*Myxococcus xanthus* DK 1622	81% / 70%	YP_635171.1
Contig 256-Orn	346	ornithine cyclodeaminase/ mu-crystallin	*Methylobacterium nodulans* ORS 2060	49% / 33%	YP_002501996.1

In addition, although analysis by mass spectrometry of pretubulysin D from An d48 strongly supported its proposed structure, it is also necessary to obtain sufficient quantities of the pretubulysin compound to allow structural proof by NMR. SBCb004-*tubZ*⁻ mutant was then

grown on large scale in the presence of XAD adsorber resin to generate enough pretubulysin A. Later, it was isolated and analyzed by NMR by Dominik Pistorius [127]. The final structure elucidation result strongly supports the proposed structure for pretubulysin A.

3.2.3 Stereochemistry investigation of the starter unit pipecolic acid

The final configuration of the *N*-methyl pipecolic acid in the tubulysins is D [100], therefore it was hypothesized that free D-pipecolic acid (either generated directly by the cyclodeaminase TubZ, or by epimerization of an initially-formed L-isomer), might be selected as a substrate by the tubulysin PKS-NRPS. Interestingly, the lysine cyclodeaminases RapL from the rapamycin pathway (40% identity/53% similarity to TubZ) and Pip from friulimicin biosynthesis (45% identity/56% similarity to TubZ), have been shown to generate exclusively L-pipecolic acid from L-lysine [122; 123; 128]. To address the stereospecificity of TubZ, the An d48 enzyme was expressed in recombinant form as a C-terminal translational fusion with glutathione-*S*-transferase (GST), and confirmed the identity of the protein by matrix assisted laser desorption ionization-time-of-flight mass spectrometry (MALDI-TOF-MS) analysis.

TubZ was purified as its GST fusion, and also as a discrete protein by on-column cleavage of GST using PreScission Protease® (Figure 3.7A). The recombinant TubZ proteins (GST-TubZ or TubZ) were then incubated with either L- or D-lysine, NAD^+ and BSA at 30 °C for 24 h [123]. The reactions were quenched by the addition of acetonitrile. Prior to analysis by HPLC-MS, the assay mixtures were derivatized with 9-fluorenylmethyl chloroformate (Fmoc-Cl). Crucially, I was able to separate Fmoc-derivatized, commercially-available L- and D-pipecolic acid from each other by chiral chromatography (Figures 3.7B and C). Assays with the lysine substrates and GST-TubZ or discrete TubZ gave identical results. In each case, analysis of the assay containing D-lysine did not reveal any evidence for formation of the product pipecolic acid (Figure 3.7D), relative to controls performed with L-/D-lysine in the absence of TubZ (Figure 3.7G). In contrast, pipecolic acid was detected in the assays with L-lysine (Figure 3.7E), and was shown to be exclusively the L-isomer by spiking the assay mixture with authentic, derivatized D-pipecolic acid (Figure 3.7F). Thus, TubZ shows the same substrate and product stereospecificities as RapL and Pip.

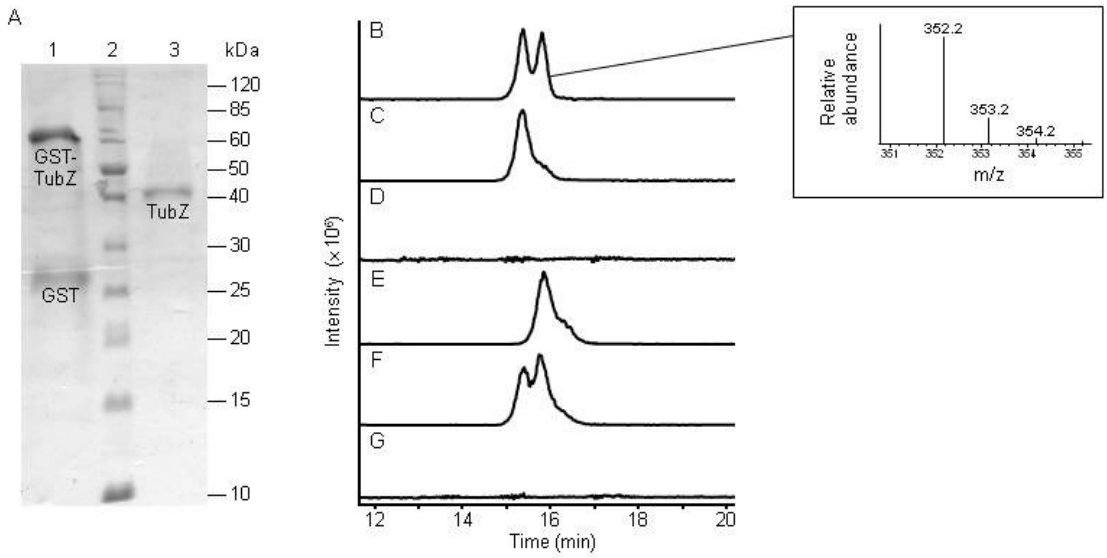


Figure 3.7 Characterization of TubZ cyclodeaminase activity.
(A) SDS-PAGE analysis of GST-TubZ (lane 1) and discrete TubZ (lane 3). Lane 2 contains molecular weight markers, with the indicated masses. The calculated molecular weight of GST-TubZ is 67.698 kDa, while that of TubZ is 40.501 kDa.
(B) Chiral HPLC-MS analysis (extracted ion chromatogram (EIC), *m/z* $[M+H]^+$ = 352.2) of a 1:1 mixture of Fmoc-derivatized, commercially-available L- and D-pipecolic acid. Peak identities were assigned by comparison to analysis of Fmoc-derivatized D-pipecolic acid alone. Inset is the mass spectrum of Fmoc-derivatized L-pipecolic acid (calculated *m/z* $[M+H]^+$ = 352.16).
(C) Chiral HPLC-MS (EIC) analysis of the Fmoc-derivatized commercially-available D-pipecolic acid alone.
(D) Chiral HPLC-MS (EIC) analysis of the TubZ reaction carried out in the presence of D-lysine.
(E) Chiral HPLC-MS (EIC) analysis of the TubZ reaction carried out in the presence of L-lysine.
(F) Mixture of the assay shown in (E) and Fmoc-derivatized D-pipecolic acid (C).
(G) Chiral HPLC-MS (EIC) analysis of a representative negative control reaction carried out with L-lysine in the absence of TubZ.

At this stage, it remained a possibility that the L-pipecolic acid generated by TubZ was epimerized to the D-isomer before activation by TubB by an as yet unidentified epimerase. To probe this question directly, I used the synthesized two enantiomers of pipecolic acid in 4,5-di-deuterium-labeled form (synthesized by Prof. Kazmaier`s laboratory) [126] as feeding substrates. The enantiomerically pure materials were administered to cultures of both wild type An d48 and the An d48-*tubZ*$^-$ mutant [126]. Analysis by HPLC-MS (Figure 3.8) showed that only the L-isomer of pipecolate was incorporated in either case, strongly suggesting that epimerization occurs following attachment of the amino acid to the multienzyme although module TubB lacks a classical epimerization (E) domain [129; 130] to generate the required stereochemistry.

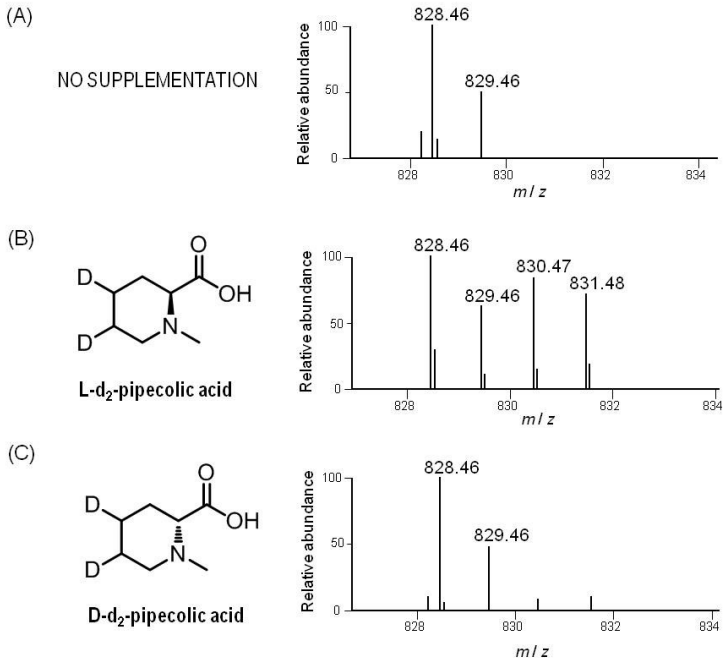


Figure 3.8 Feeding studies with deuterium-labeled L-pipecolic acid.
Peak of tubulysin D.
(A) Absence of supplementation with pipecolic acid.
(B) Supplementation with di-deuterium labeled L-pipecolic acid.
(C) Supplementation with di-deuterium labeled D-pipecolic acid.

Further support for the lack of epimerization by TubB comes from phylogenetic analysis of the TubC condensation domain which accepts the *N*-methyl pipecolic acid as substrate: the C domain clusters together with other C domains with demonstrated specificity for L-amino acids at both their donor and acceptor sites (so-called $^{L}C_{L}$ domains) [131] (Figure 3.9). This analysis also argues against the possibility that the TubC C domain functions as a condensase/epimerase [132], a dual-function domain which has been shown in other NRPS pathways to convert L-amino acids to their D-isomers. Thus, epimerization apparently occurs downstream of the first round of chain extension and is likely to be catalyzed by an external racemase [133]. Intriguingly, only the PCP domain of TubD incorporates a sequence motif at the phosphopantetheine attachment site (GXDSX) which resembles the GGDSI motif previously implicated in recognition between E and PCP domains *in cis* [134], perhaps indicating that the epimerization *in trans* may occur at this stage of chain extension. However, further studies will be required to elucidate the precise timing of this cryptic reaction.

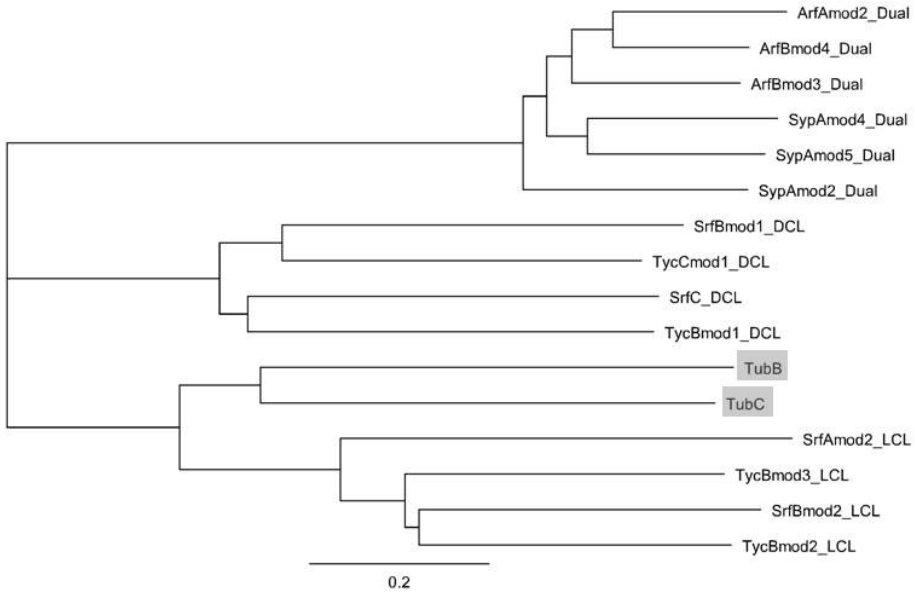


Figure 3.9 Phylogenetic analysis of the C domains of TubB and TubC.
The tree includes representative members of the three known types of C domains [131]: $^{L}C_{L}$, which are specific for L-amino acids at both their donor and acceptor sites; $^{D}C_{L}$, which accept D- amino acids at their donor sites, and L-amino acids at their acceptor sites; and Dual, C domains which show dual condensation and L-to-D epimerization activity. Branch length indicates the number of inferred amino acid changes per position. Abbreviations: Arf, arthrofactin; Syp, syringopeptin; Srf, surfactin; Tyc, tyrocidine. In each case the module from which the sequence derives is indicated.

3.3 Discovery of 22 natural tubulysins from *Angiococcus disciformis* An d48 and *Cystobacter* sp. SBCb004

The presence of the pretubulysins in the wild type strains prompted me to re-analyze the extracts for additional metabolites, which may have escaped detection in earlier experiments. I further found using a combination of feeding studies and high-resolution tandem mass spectrometry, that SBCb004 and An d48 together biosynthesize 22 additional tubulysin derivatives.

I found 11 further candidates for novel compounds in *A. disciformis* An d48, with molecular masses *m/z* $[M+H]^+$ = 628.4, 656.4, 670.4 (different retention time than that of pretubulysin D), 672.4, 686.4 (×2), 700.4, 714.4, 744.4, 772.4, and 786.4 (Figure 3.10). Although most of these metabolites appeared to be minor in comparison to tubulysin D (see Figure 3.10 for yield estimates), compound 670.4 was present in significant amounts (as judged by relative peak area in the HPLC-MS chromatogram). In each case, a likely relationship to the tubulysins was established by feeding deuterium-labeled L-pipecolic acid, valine, or both and by comparison of the incorporation pattern to that of pretubulysin D (**5**). In Figure 3.11, the compound 14 (mass 670.4 but different retention time than that of pretubulysin D) was shown here as an example of the feeding study. Comparison of the fragmentation patterns of metabolites 772.4 and 786.4 to those of the tubulysins D–F and H (Figure 3.13), allowed me to assign the structures to desacetyl tubulysin E (**12**) and desacetyl tubulysin D (**13**), respectively (Figure 3.10). Although biosynthetic considerations (i.e. the defined stereospecificities of adenylation (A) and C-methyltransferase domains) strongly suggest that the stereochemistry of these metabolites should be shared with the parent compounds, I cannot make any definitive statements on this issue in the absence of NMR data.

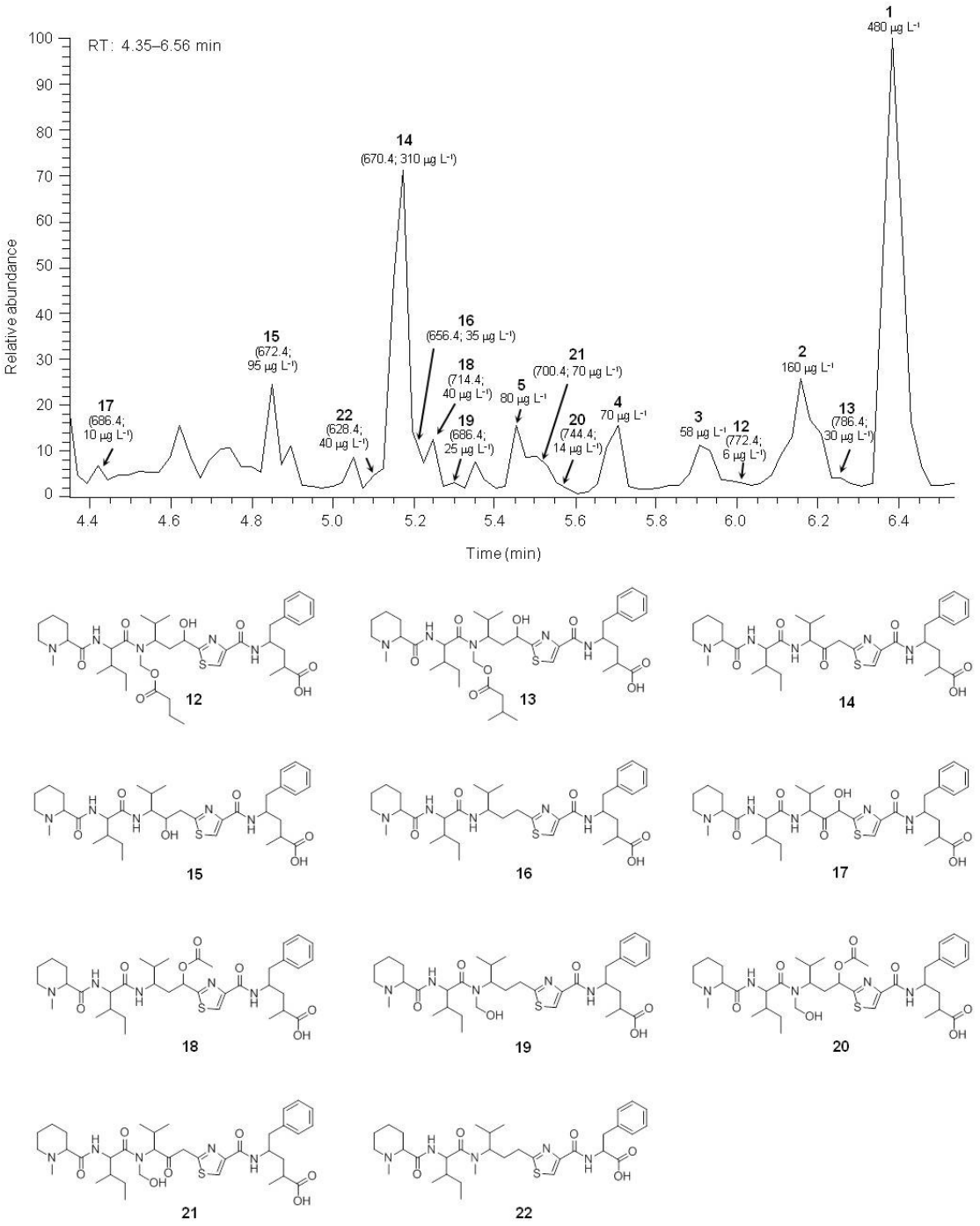


Figure 3.10 Novel tubulysins produced by *A. disciformis* An d48.
HPLC-MS analysis (BPC) of wild type extracts. Peaks corresponding to the five known (**1**–**5**) (Figure 1.13) and eleven novel tubulysins (**12**–**22**) are indicated.

Feeding experiment proved that compound (**14**) belongs to the tubulysin family (Figure 3.11). Based on its accurate mass, the most probable molecular formula of compound 670.4 (**14**) was $C_{35}H_{51}N_5O_6S$. It differs from pretubulysin D by $–CH_4$ and +O, demonstrating that this new compound is not simply a stereoisomer of pretubulysin D. It was hypothesized that the metabolite might have lost a methyl group relative to pretubulysin D (accounting for $–CH_2$), and that it would incorporate an extra oxygen (+O) and a site of unsaturation (–2H) (e.g., a carbonyl functionality) (Figure 3.12). Subsequently, this hypothesis was confirmed by feeding experiment of d_3-methionine and NMR analysis (done by Dominik Pistorius) [127].

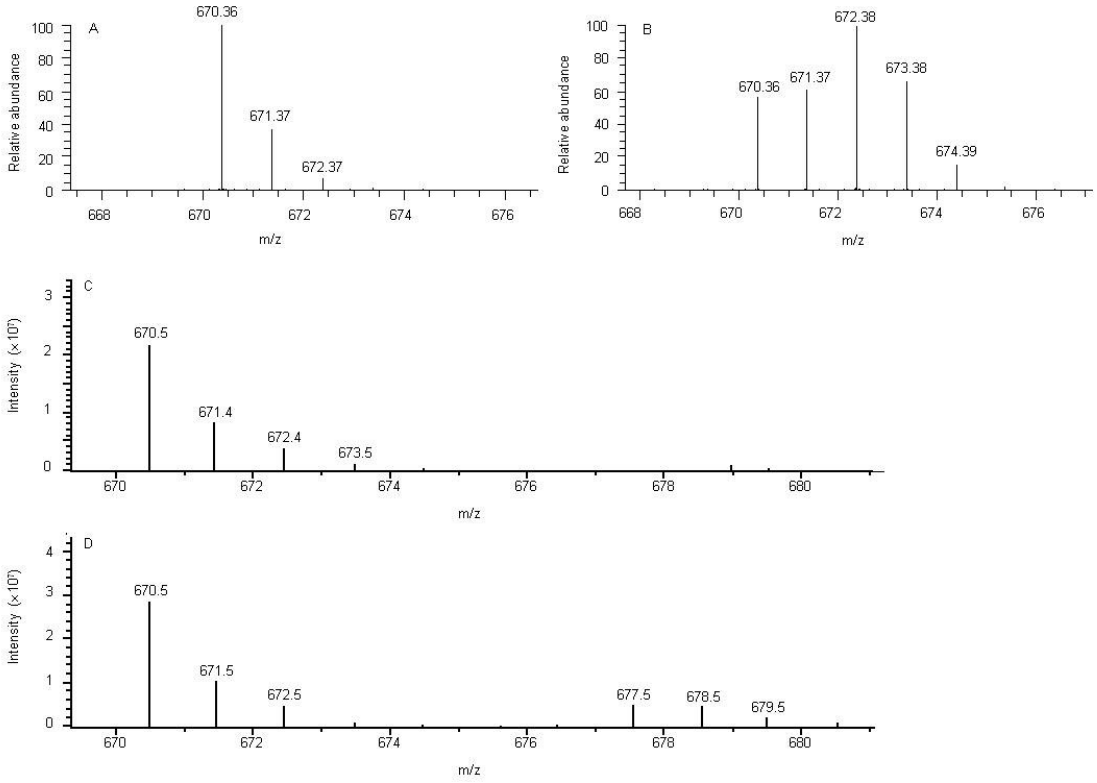


Figure 3.11 Feeding studies of compound (14) (mass 670.4) with deuterium-labeled L-pipecolic acid and d_8 valine. The detailed data for all the other compounds were shown in reference [127].
(A) Absence of supplementation with pipecolic acid.
(B) Supplementation with di-deuterium labeled L-pipecolic acid.
(C) Absence of supplementation with valine.
(D) Supplementation with deuterium-labeled (d_8) valine.
The analyses shown in (A) and (B) were carried out on a Thermo LTQ Orbitrap Hybrid FT mass spectrometer, while analyses (C) and (D) were performed on an Agilent 1100 series HPLC-MS.

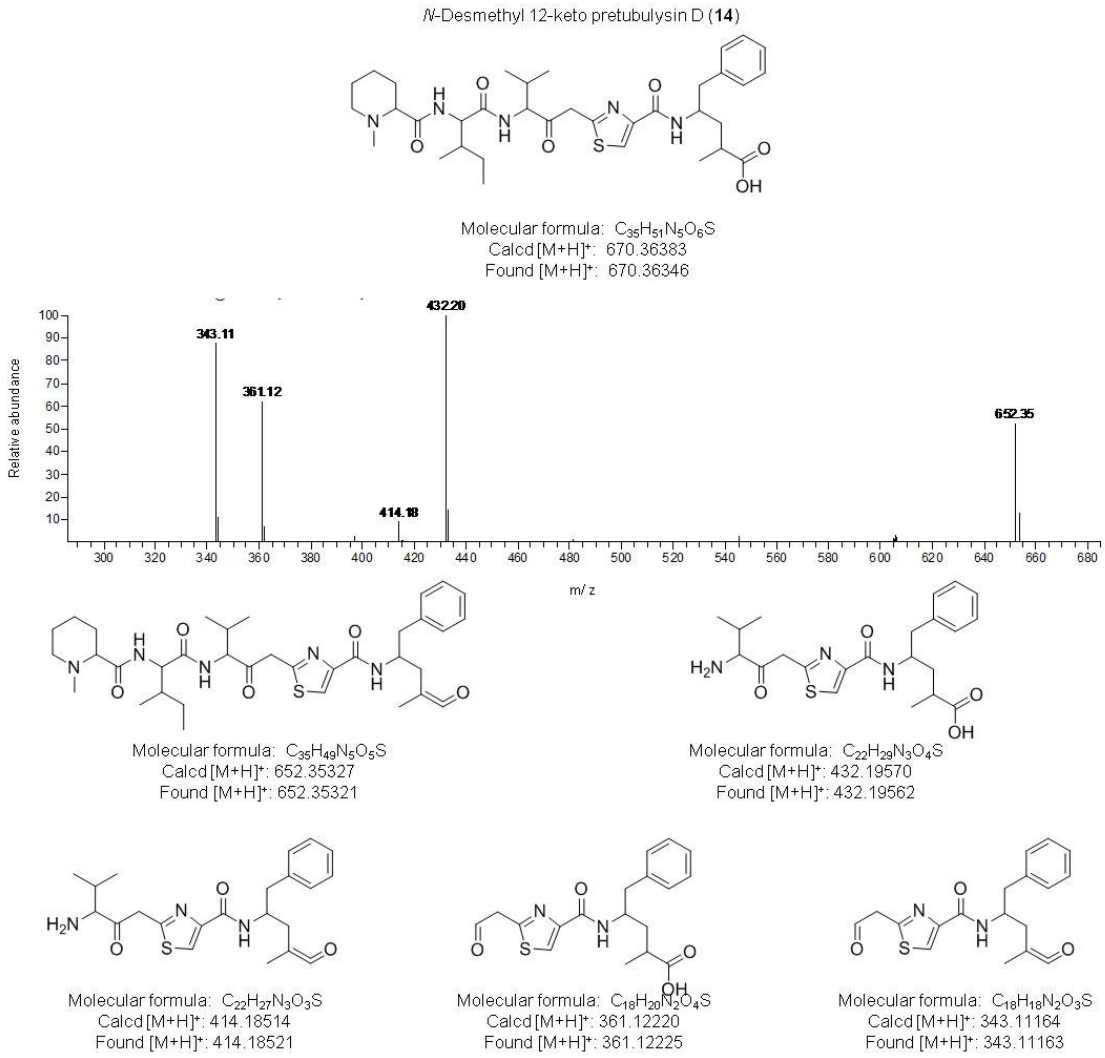


Figure 3.12 Detailed analysis of the MS^2 fragmentation patterns of the novel tubulysins from An d48. Compound (**14**) was taken here as an example. Shown are the proposed structure of the metabolite and all fragments, the molecular formulae and calculated molecular masses of the fragments, and the experimentally-determined molecular mass. Masses shown are typically for the predicted protonated species, but no claim is being made as to the location of the charge on the fragments. Data were obtained on a Thermo LTQ Orbitrap Hybrid FT mass spectrometer. The detailed data for all the other compounds (**10**-**22**) were shown in the reference [127].

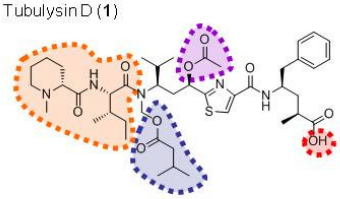
With this structure in hand, it was possible to assign compounds 656.4 and 672.4, to *N*-desmethyl pretubulysin analogues in which the C12 carbonyl group had alternately been reduced to a hydroxyl (672.4; *N*-desmethyl 12-hydroxy pretubulysin D (**15**)), or to a methylene functionality (656.4; *N*-desmethyl pretubulysin D (**16**)); in the case of compound **15**, however, I cannot rule out the alternative that the hydroxyl group is instead located at the adjacent C11

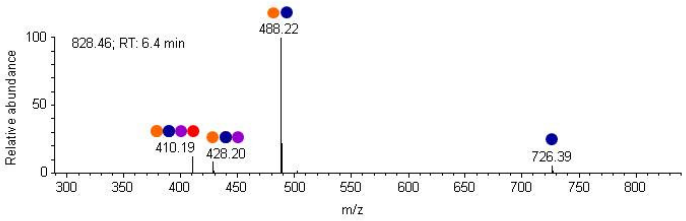
position, consistent with post-assembly line hydroxylation of metabolite **16**. In any case, a compound corresponding to the intermediate dehydrated analogue was not detected. In addition, one of the two compounds with mass *m/z* $[M+H]^+$ = 686.4 (**17**) appears to be the 11-hydroxy version of metabolite **14**. By MS analysis, it was also possible to putatively identify compound 714.4 as *N*-desmethyl 11-acetoxy pretubulysin D (**18**), referred to in earlier synthesis literature as tubulysin U [108; 135; 136], demonstrating that at least two of the normal post-assembly line modifications can occur even in the absence of *N*-methylation. By similar logic, the second peak at 686.4 was attributed to *N*-hydroxymethyl pretubulysin (**19**) and that at 744.4 to its 11-acetoxy analogue (**20**). Analysis of the compound 700.4 revealed fragments consistent with the presence of an *N*-hydroxymethyl functionality and a 12-keto group, leading to its identification as *N*-hydroxymethyl 12-keto pretubulysin D (**21**). Finally, metabolite 628.4 appears to comprise a truncated pretubulysin derivative in which tubuphenylalanine has been replaced by phenylalanine (**22**).

I next analyzed extracts of wild type SBCb004 for the corresponding novel tubulysins, using the same analytical methodology. This comparison revealed a tubutyrosine analogue for each of the new compounds (compounds **23**–**33**) (Figure 3.14, Table 3.3). As with the tubulysin D series, the amounts of these compounds relative to tubulysin A were uniformly low; however the specific yield profiles for the various analogs differ between the strains.

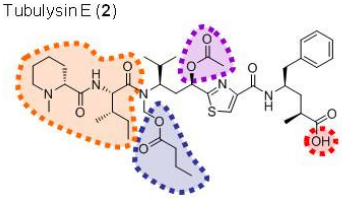
Although the amounts of the new metabolites (Table 3.3) enabled "structure elucidation" by mass spectrometry, the yields were insufficient to allow full structure proof by NMR, or evaluation of cytotoxicity relative to the parent compounds. Nonetheless, the high resolution MS analysis of the parent ions and all fragments is fully consistent with the suggested structures, there is strong internal concordance among the proposed fragmentation patterns (Figure 3.13), and the structural assignments are in complete agreement with the feeding studies. In addition, the mode of biosynthesis significantly restricts the number of possible structures (including stereochemistry) that can result from the tubulysin pathway. Importantly, in every case, the proposed structures can be rationalized in terms of reasonable variations in the action of a specific domain or domains within the assembly lines.

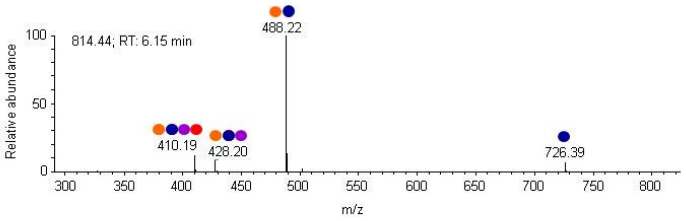
Tubulysin D (**1**)



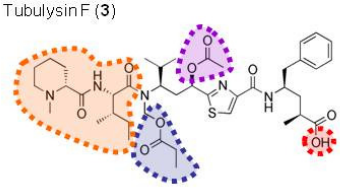


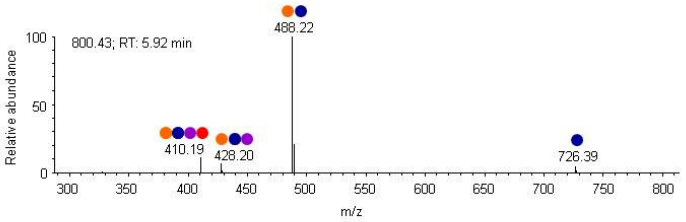
Tubulysin E (**2**)



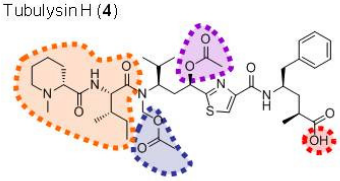


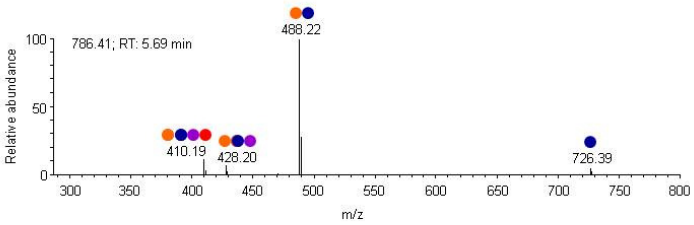
Tubulysin F (**3**)



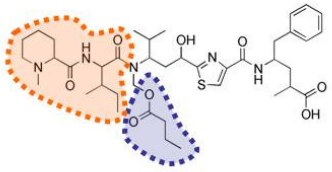


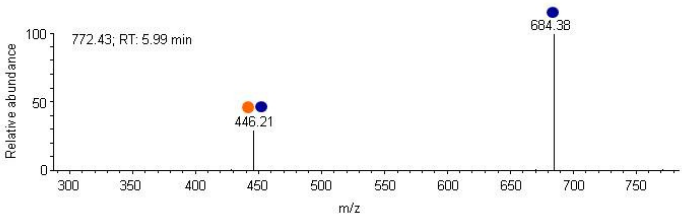
Tubulysin H (**4**)



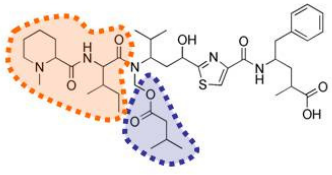


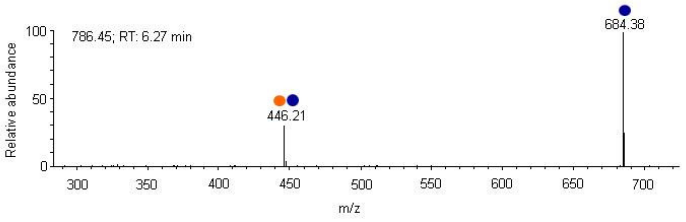
Desacetyl tubulysin E (**12**)



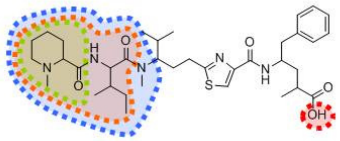


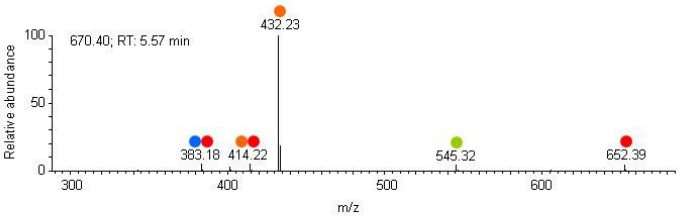
Desacetyl tubulysin D (**13**)



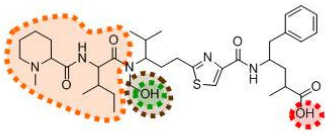


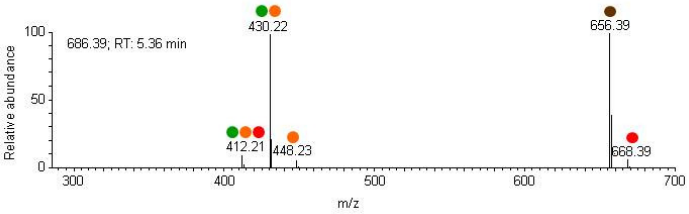
Pretubulysin D (**5**)



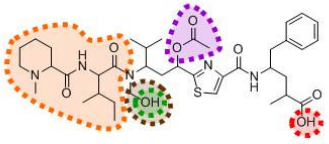


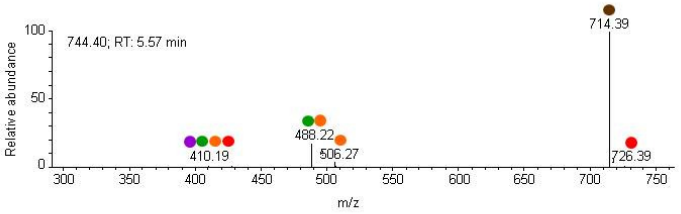
N-Hydroxymethyl pretubulysin D (**19**)



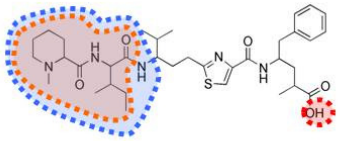


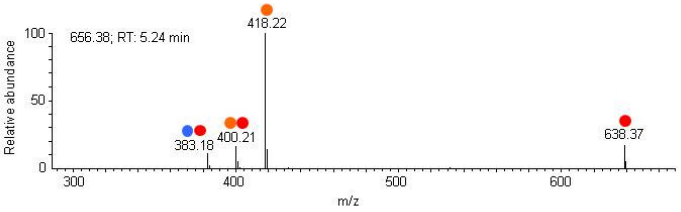
N-Hydroxymethyl 11-acetoxy pretubulysin D (**20**)



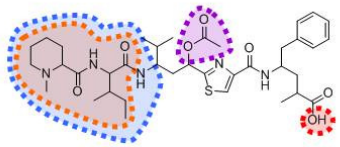


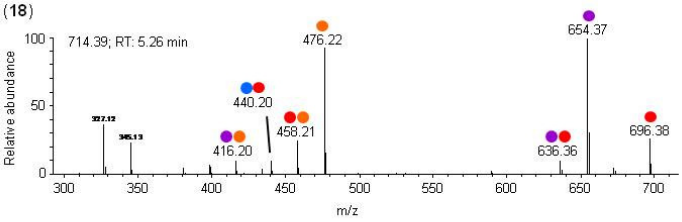
N-Desmethyl pretubulysin D (**16**)



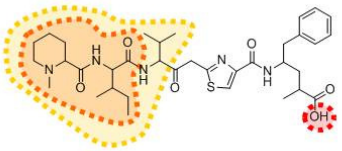


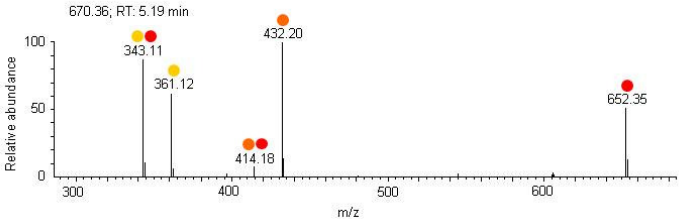
N-Desmethyl 11-acetoxy pretubuylysin D (tubulysin U) (**18**)





N-Desmethyl 12-keto pretubulysin D (**14**)





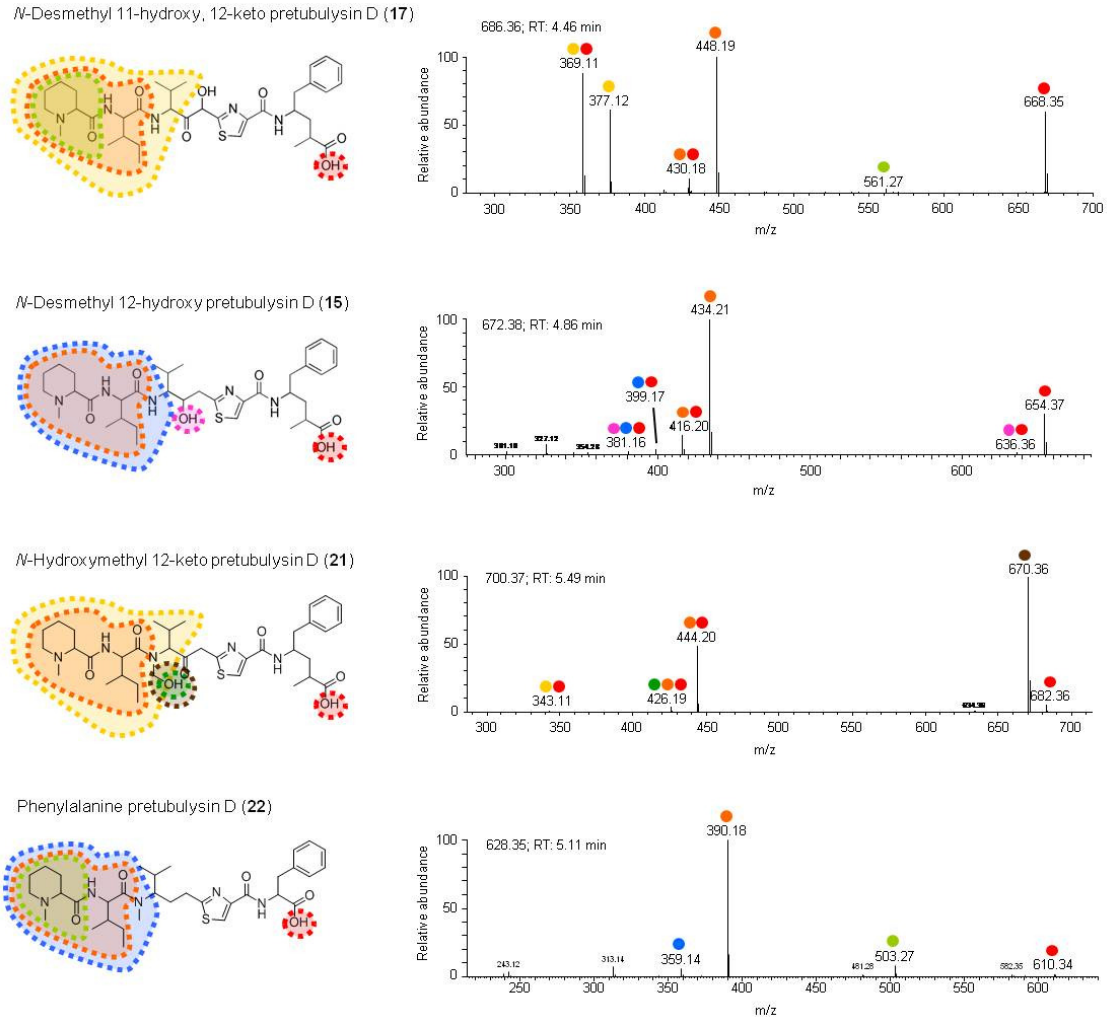


Figure 3.13 Comparative analysis of the MS^2 fragmentation patterns of the tubulysins from *A. disciformis* An d48.
Comparable fragments lost from each metabolite are indicated using the same color. The mass spectra include the retention time and molecular mass of the respective parent ions, as well as colored labels to indicate the fragments lost to generate each daughter peak. All data were obtained on a Thermo LTQ Orbitrap Hybrid FT mass spectrometer.

Table 3.3 Novel metabolites identified in strains An d48 and SBCb004.

Metabolites in strain An d48			Metabolites in strain SBCb004		
No.	Compound name	Mass $[M+H]^+$	No.	Compound name	Mass $[M+H]^+$
5	Pretubulysin D	670.40	11	Pretubulysin A	686.39
12	Desacetyl tubulysin E	772.43	23	Desacetyl tubulysin B (Tubulysin W)	788.43
13	Desacetyl tubulysin D	786.45	24	Desacetyl tubulysin A	802.44
14	*N*-desmethyl 12-keto pretubulysin D	670.36	25	*N*-desmethyl 12-keto pretubulysin A	686.36
15	*N*-desmethyl 12-hydroxy pretubulysin D	672.38	26	*N*-desmethyl 12-hydroxy pretubulysin A	688.37

16	*N*-desmethyl pretubulysin D	656.38	27	*N*-desmethyl pretubulysin A	672.38
17	*N*-desmethyl 11-hydroxy, 12-keto pretubulysin D	686.36	28	*N*-desmethyl 11-hydroxy, 12-keto pretubulysin A	702.35
18	*N*-desmethyl 11-acetoxy pretubuylysin D (Tubulysin U)	714.39	29	*N*-desmethyl 11-acetoxy pretubuylysin A (Tubulysin X)	730.39
19	*N*-hydroxymethyl pretubulysin D	686.39	30	*N*-hydroxymethyl pretubulysin A	702.39
20	*N*-hydroxymethyl 11-acetoxy pretubulysin D	744.40	31	*N*-hydroxymethyl 11-acetoxy pretubulysin A	760.39
21	*N*-hydroxymethyl 12-keto pretubulysin D	700.37	32	*N*-hydroxymethyl 12-keto pretubulysin A	716.37
22	Phenylalanine pretubulysin D	628.35	33	Tyrosine pretubulysin A	644.35

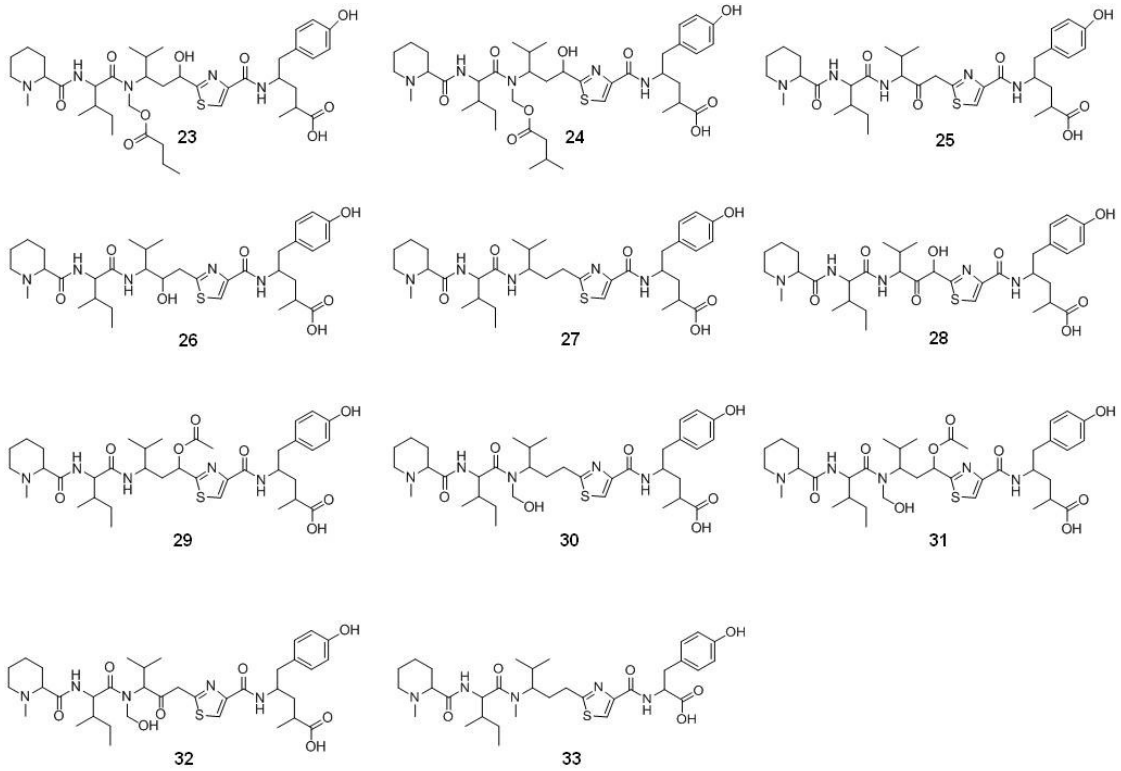


Figure 3.14 Proposed structures of the novel tubulysins identified in *Cystobacter* sp. CBSb004.

In summary, the analysis of the molecular basis for this diversity-oriented biosynthesis provides new insights into the function of PKS and NRPS natural product 'assembly lines'. It was also showedn that pretubulysin D retains much of the potency of its more structurally elaborate relatives (see paper). Thus, the 22 additional simplified analogues are attractive targets for chemical synthesis, with the aim of deepening our understanding of structure-activity relationships within the tubulysin family of compounds.

3.4 Genetic modification of tubulysin biosynthetic genes in natural producer strains

As noted earlier, the conserved tubulysin biosynthetic gene cluster in both An d48 and SBCb004 strains contains several open reading frames whose functions have been unclear to date. These include the genes *tubA*, *orf2*, and *orf1* at the 5′ end of the gene set. The observation that they share a high degree of mutual sequence homology (56–78% identity/73–90% similarity) [127] suggests that they play a role in the biosynthesis. Based on the sequence analysis, TubA was early supposed to serve as one of the missing acyl tranferases in the tubuysin biosynthesis process [102]. However, neither *orf1* (predicted to encode an anion transporting ATPase) nor *orf2* (similar to hypothetical proteins only) has an obvious role to play in tubulysin biosynthesis.

Both orf17 and *orf18* encode patatin-like proteins, which show high mutual sequence homology (60% identity/70% similarity in the case of the SBCb004 proteins). Patatin-like proteins exhibit lipid acyl hydrolase activity, a function apparently not required in the tubulysin pathway. Therefore, to obtain further insights into the potential functions of these genes in the tubulysin pathway, we aimed to inactivate these genes in the both the An d48 and the SBCb004 wild type strain.

3.4.1 Knockout mutagenesis of *tubA*

I was able to generate the insertion knockout mutants of genes *tubA* in both natural producer strains. The construction of gene inactivation was accomplished by homologous recombination in the encoding regions. Oligos were designed to amplify the knockout region in the middle of target gene and also to introduce a stop codon inside. In strain SBCb004, primer pair ctubA-a and ctubA-b were used to knockout gene *tubA* (Figure 3.15, Table 2.7). The inactivation mutant of *tubA* was also obtained in strain An d48.

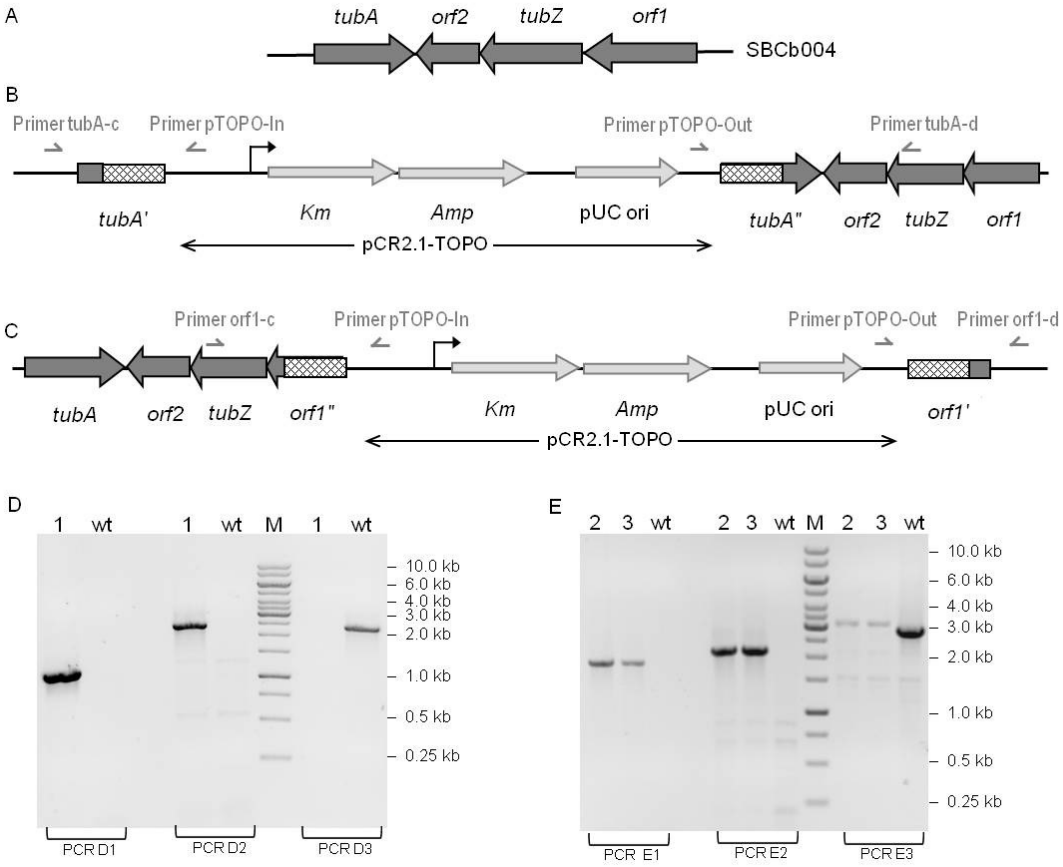


Figure 3.15 Inactivation of *tubA* and *orf1* in SBCb004.
(A) Genetic organization of chromosomal DNA in SBCb004 wild type.
(B) Genetic organization of chromosomal DNA in mutant SBCb004-*tubA*$^{-}$. From left to right: *tubA* truncated at the 3′ end (*tubA′*), pCR2.1-TOPO vector backbone, *tubA* truncated at the 5′ end (*tubA″*).
(C) Genetic organization of chromosomal DNA in mutant SBCb004-*orf1*$^{-}$. From right to left: *orf1* truncated at the 3′ end (*orf1′*), pCR2.1-TOPO vector backbone, *orf1* truncated at the 5′ end (*orf1″*).
(D) PCR reactions for verification of SBCb004-tubA mutant: PCR D1 yield a product of 1.1 kb from the mutant with primers tubA-c and pTOPO-in; PCR D2 yield a product of 2.3 kb from the mutant with primers pTOPO-out and tubA-d; PCR D3 amplify a product (2.3 kb) only from the wild-type with primers tubA-c and tubA-d.
(E) PCR reactions for verification of SBCb004-orf1 mutant: PCR E1 yield a product of 1.8 kb from the mutant with primers orf1-d and pTOPO-out; PCR D2 yield a product of 2.1 kb from the mutant with primers pTOPO-in and orf1-c; PCR D3 amplify a product (2.8 kb) only from the wild-type with primers orf1-c and orf1-d.
In the figure, the designation 'wt' represents the wild type of SBCb004, '1' represents SBCb004-*tubA*$^{-}$ exconjugant, and '2' and '3' represents two distinct clones of SBCb004-*orf1*$^{-}$ exconjugant. All primers listed are shown in Table 2.7.

By HPLC-MS analysis, mutant SBCb004-*tubA*$^{-}$ was found to produce similar amount of pretubulsyin A with the wild type SBCb004 strain, and in no case tubulysin A presents in the culture extracts. Only small amount of additional pretubulysin derivatives were also detected,

including the tyrosine pretubulysin A (**33**), *N*-desmethyl 12-keto pretubulysin A (**25**), *N*-desmethyl 12-hydroxy pretubulysin A (**26**), as well as the *N*-hydroxymethyl 12-keto pretubulysin A (**32**) (Table 3.3). Except the only hydroxyl pretubulysin *N*-hydroxymethyl 12-keto pretubulysin A (**32**) produced in tiny amount, no other mature tubulysins that incorporate acyl group were observed from the mutant SBCb004-*tubA*$^{-}$. Similar result was obtained from the *tubA* knockout mutant in strain An d48 (Figure 3.16). It proved that no function of acyl transferase was applied in the *tubA* inactivation mutants. This fact suggests that TubA may serve as one (or both) of the missing acyltransferases in tubulysin assembly. Till now, no more proof was draw for the acylation function from the discovery of natural strain. To further confirm this hypothesis, *in vitro* assay of expressed TubA protein is necessary to carry out in the future.

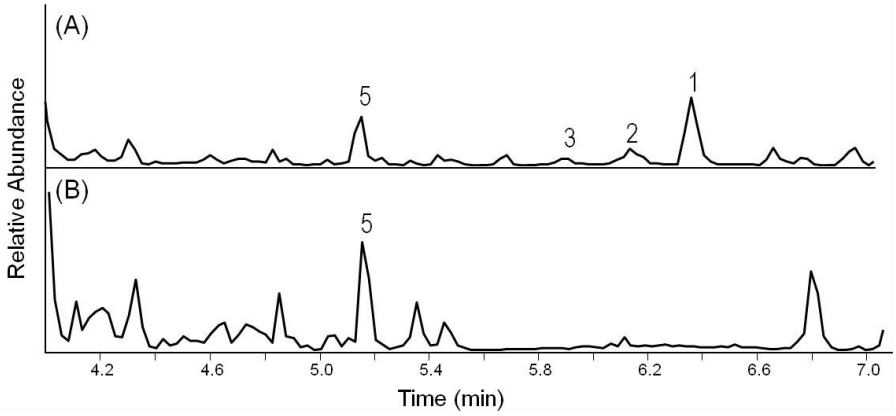


Figure 3.16 HPLC-MS analysis (BPC) of the An d48 wild type and mutant An d48-*tubA*$^{-}$.
Peaks corresponding to tubulysin D (**1**), E (**2**), F (**3**) and pretubulysin D (**5**) are indicated.
(A) An d48 wild type.
(B) Mutant An d48-*tubA*$^{-}$.

3.4.2 Knockout mutagenesis of *orf1*

Given the operon structure of genes *orf1-tubZ-orf2,* I anticipated polar effects on *tubZ* and *orf2* from disruption of *orf1*, and thus inactivation of *orf1* was assumed to knockout the entire operon. I have shown previously that insertion mutants of *tubZ* in both the An d48 and SBCb004 produce more pretubulysin D or A and substantially less tubulysin D or A than the respective wild type strains [109; 127]. The production profile from the *orf1* mutant in both strains (SBCb004-*orf1*$^{-}$ and An d48-*orf1*$^{-}$) closely resembled that of the *tubZ* knockout strain (Figure 3.17). This result confirms that those strains must contain an additional cyclodeaminase function, capable of supplying L-pipecolate to the assembly line in the absence of a functional TubZ. The *orf1* and *tubZ* inactivation mutants in both strains were also found to produce small amounts of tubulysin

D or A. It might because *orf1* serve to recruit the oxidases and/or acyl transferases to the pathway via specific protein-protein interactions. Unfortunately, I was unable to obtain any *orf2* knockout mutant in both wild type strains due to the small size of this gene (around 660 bp), but a *orf2* mutant was generated in the heterologous expression system (This will be addressed in the chapter 3.5.4). Of course, the proposed functions of gene *orf1* and *orf2* still remain to be directly demonstrated in future works, for example *in vitro* with recombinant protein.

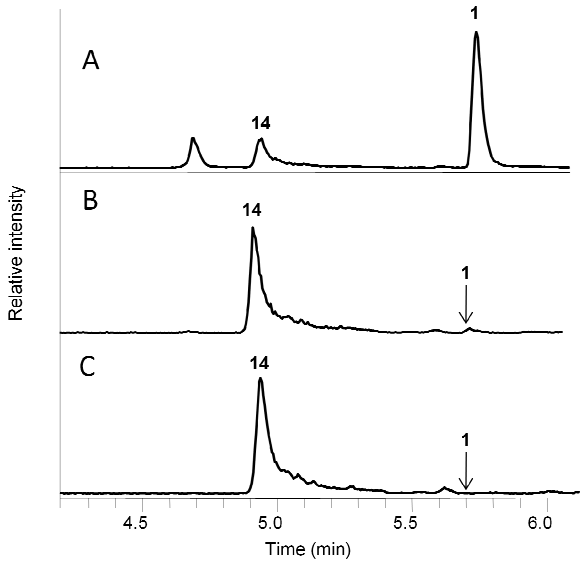


Figure 3.17 High-resolution MS analysis (EIC) at *m/z* 686.4 [*M*+H]$^+$ (pretubulysin A [11]) and *m/z* 844.4 [*M*+H]$^+$ (tubulysin A [6]).
(A) SBCb004 wild type.
(B) Mutant SBCb004-*tubZ*$^-$.
(C) Mutant SBCb004-*orf1*$^-$.

3.4.3 Knockout mutagenesis of patatin genes

I got one patatin mutant An d48-*orf18*$^-$ in strain An d48 as well as two patatin mutants SBCb004-*orf17*$^-$ and SBCb004-*orf18*$^-$ in strain SBCb004. Similar production profiles of *orf18* mutants in two strains were observed, and the major tubulysin productions in two mutant strains (tubulysin A, B, C and D, E, F) declined dramatically to very tiny amount. On the contrary, a series of new peaks showed up in the chromatogram (Figure 3.18). These new group of compounds have a common characteristic that their mass all equal to 74 plus the mass of known tubulysins. For example, the mass of compound **34** in strain An d48-*orf18*$^-$ is 744.4 [*M*+H]$^+$, which is 74 more than the mass of pretubulysin D (670.4 [*M*+H]$^+$); the mass of compound **37** in

strain SBCb004-*orf18*$^{-}$ is 918.4 $[M+H]^+$, which is also 74 more than the mass of tubulsyin A (844.4 $[M+H]^+$) (Figure 3.18). Feeding experiments showed clear incorporations of deuterium-labeled pipecolic acid and d_8-labeled valine into these compounds (Figure 3.19). This fact demonstrates that they harbor at least these two core building blocks of the tubulysin structure. This phenomenon encouraged me to hypothesize this new group of compounds also as tubulysin derivatives.

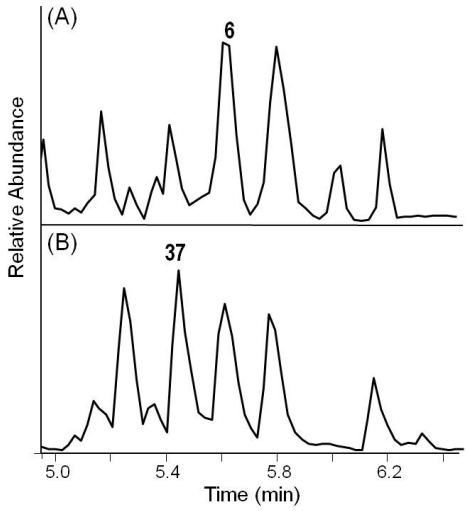


Figure 3.18 High-resolution MS analysis (BPC) of the SBCb004 wild type and mutant SBCb004-*orf18*$^{-}$.
(A) SBCb004 wild type. Peaks corresponding to tubulysin A (**6**) is indicated.
(B) Mutant SBCb004-*orf18*$^{-}$. Peak (**37**) represents a new found compound at *m/z* 918.5 $[M+H]^+$.

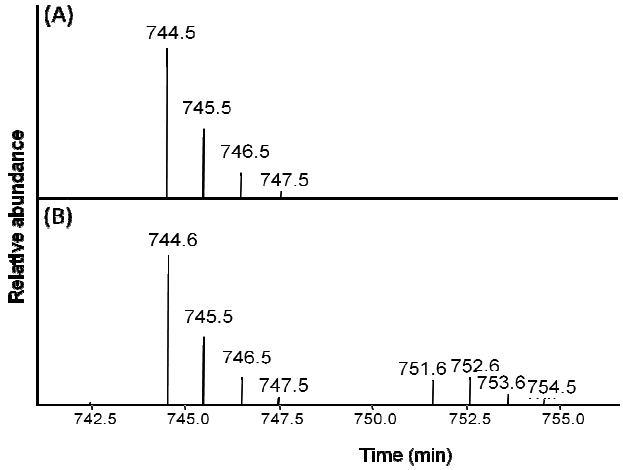


Figure 3.19 Feeding study with d_8-labeled valine in mutant An d48-*orf18*$^{-}$.
Shown is the peak of the new found compound (**34**).
(A) Absence of supplementation.
(B) Supplementation with d_8-labeled-valine.

Tandem high-resolution mass spectrometry, as described previously, was applied to estimate the structures of this new group of compounds. On the basis of relative formula estimation and the comparative analysis of MS^n fragmentation pattern between the unknown compounds (mass + 74) and corresponding known tubulysins, this new group of compounds were assumed to share a similar structure: tubulysin binding with a molecular of glycerol at the C-terminal. Detailed MS^2 fragmentation pattern analysis of compound **35** from An d48-*orf18*$^-$ mutant is shown in Figure 3.20. The proposed structures of the metabolites and all the fragments, the molecular formulae and calculated molecular masses of the fragments, and the experimentally-determined molecular masses all fit well to the estimated structure: tubulysin D–glycerol ester (Figure 3.22).

In order to confirm the prediction, two kinds of ^{13}C labeled glycerols were used as feeding substracts to carry out the feeding experiment in mutant An d48-*orf18*. As expected, the result perfectly showed the incorporation of glycerols in compound **35** with either ^{13}C labeled glycerol substrracts (Figure 3.21). All these results proved that the structure of compound **35** should be the tubulysin D–glycerol. However, neither the tandem high-resolution mass spectrometry mass analysis nor the feeding experiments with ^{13}C labeled glycerols can elucidate the detailed binding carbon of glycerol to tubulysin. Therefore, based on these foundings, my colleague Dominik Pistorius isolated and purified compound 35, and the structure was finally identified via NMR that the C_1 of glycerol molecular binds to the C-terminal of tubulysin (Figure 3.22).

By analyzing the *orf18* knockout mutants in two strains, a series of tubulysin-glycerol metabolites were identified, including the tubulysin D–glycerol (**34**), tubulysin E–glycerol (**35**), pretubulysin D–glycerol (**36**) and *N*-Desmethyl 12-keto pretubulysin D–glycerol (**37**) from strain An d48-*orf18*$^-$, as well as tubulysin A–glycerol (**38**), tubulysin B–glycerol (**39**), pretubulysin A–glycerol (**40**) and *N*-Desmethyl 12-keto pretubulysin A–glycerol (**41**) from strain SBCb004-*orf18*$^-$ (Table 3.4, Figure 3.22). The production profiles of *orf18* mutants in both strains are the same: the major tubulysin products change from tubulysin D and E or A and B to their corresponding glycerol ester compounds (Table 3.5). In addition, these glycerol ester compounds also appeared in the wild type strain when I re-analyzed the production of SBCb004 (Table 3.5).

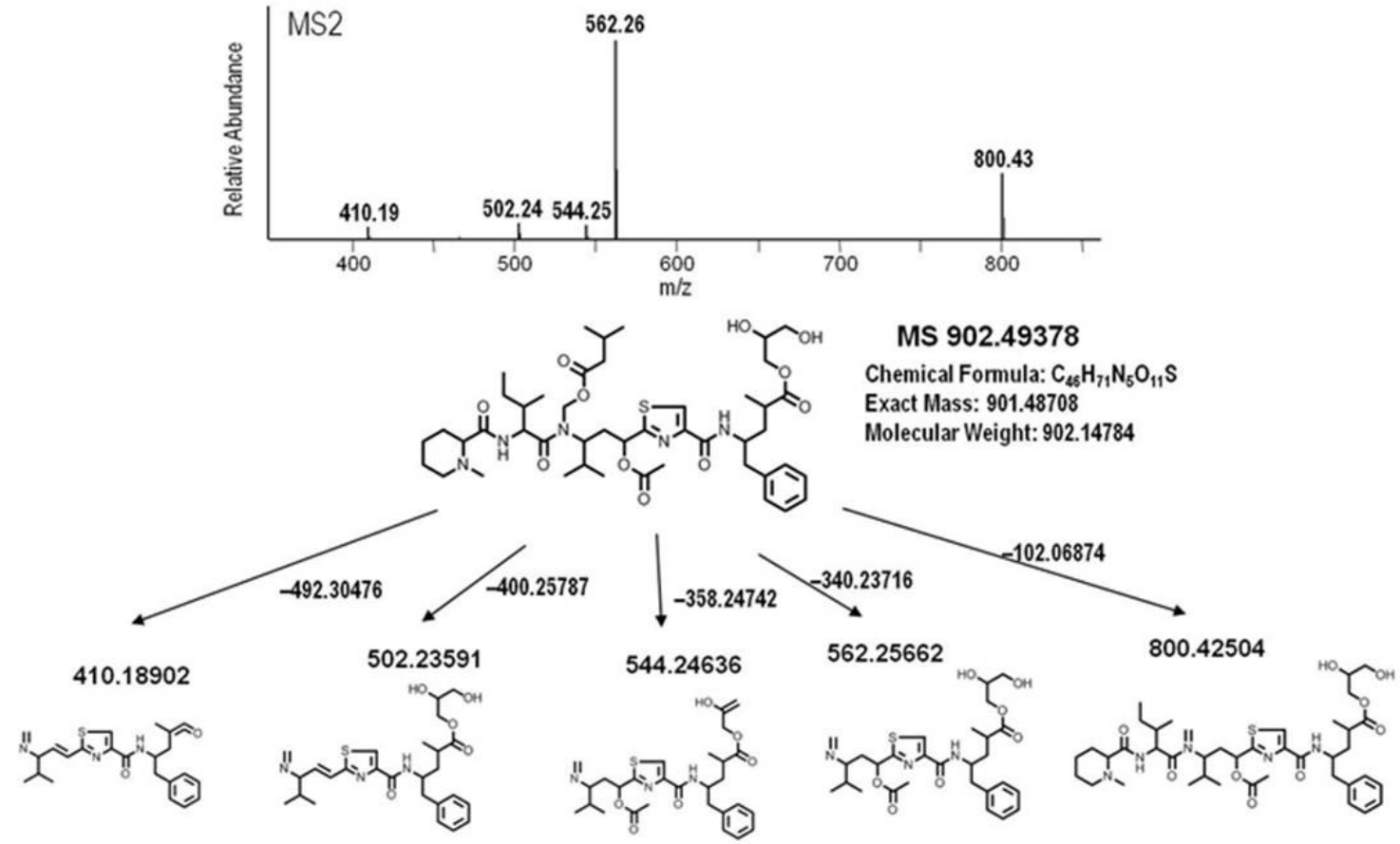


Figure 3.20 Detailed analysis of the MS^2 fragmentation patterns of the new found compound (35) from mutant An d48-*orf18*$^-$.

Shown are the proposed structures of the metabolites and all fragments, the molecular formulae and calculated molecular masses of the fragments, and the experimentally-determined molecular masses. All data were obtained on a Thermo LTQ Orbitrap Hybrid FT mass spectrometer.

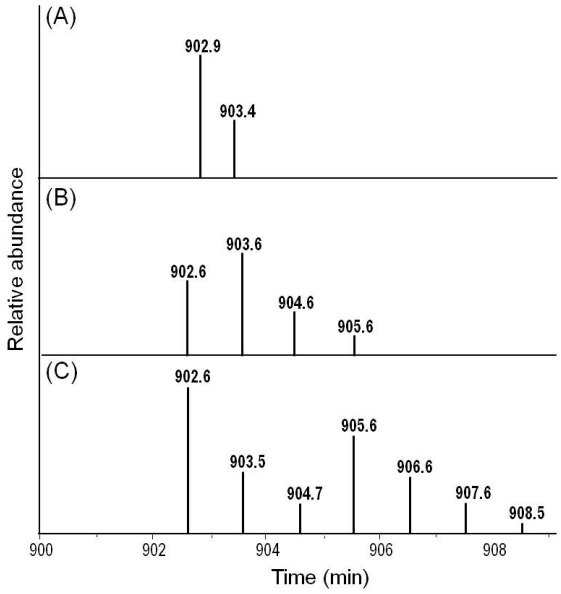


Figure 3.21 Feeding study with ^{13}C labeled glycerol in mutant An d48-*orf18*$^{-}$.
Shown is the peak of the new found compound (**35**).
(A) Absence of supplementation.
(B) Supplementation with 2-^{13}C-labeled-glycerol.
(C) Supplementation with globle-^{13}C-labeled-glycerol.

Table 3.4 More novel metabolites identified in An d48 and SBCb004 mutants (Continue to Table 3.3).

Metabolites in strain An d48			Metabolites in strain SBCb004		
No.	Compound name	Mass $[M+H]^+$	No.	Compound name	Mass $[M+H]^+$
34	Tubulysin D-glycerol	902.5	38	Tubulysin A-glycerol	918.5
35	Tubulysin E-glycerol	888.5	39	Tubulysin B-glycerol	904.5
36	Pretubulysin D-glycerol	744.4	40	Pretubulysin A-glycerol	686.39
37	*N*-desmethyl 12-keto pretubulysin D-glycerol	744.4	41	*N*-desmethyl 12-keto pretubulysin A-glycerol	686.39

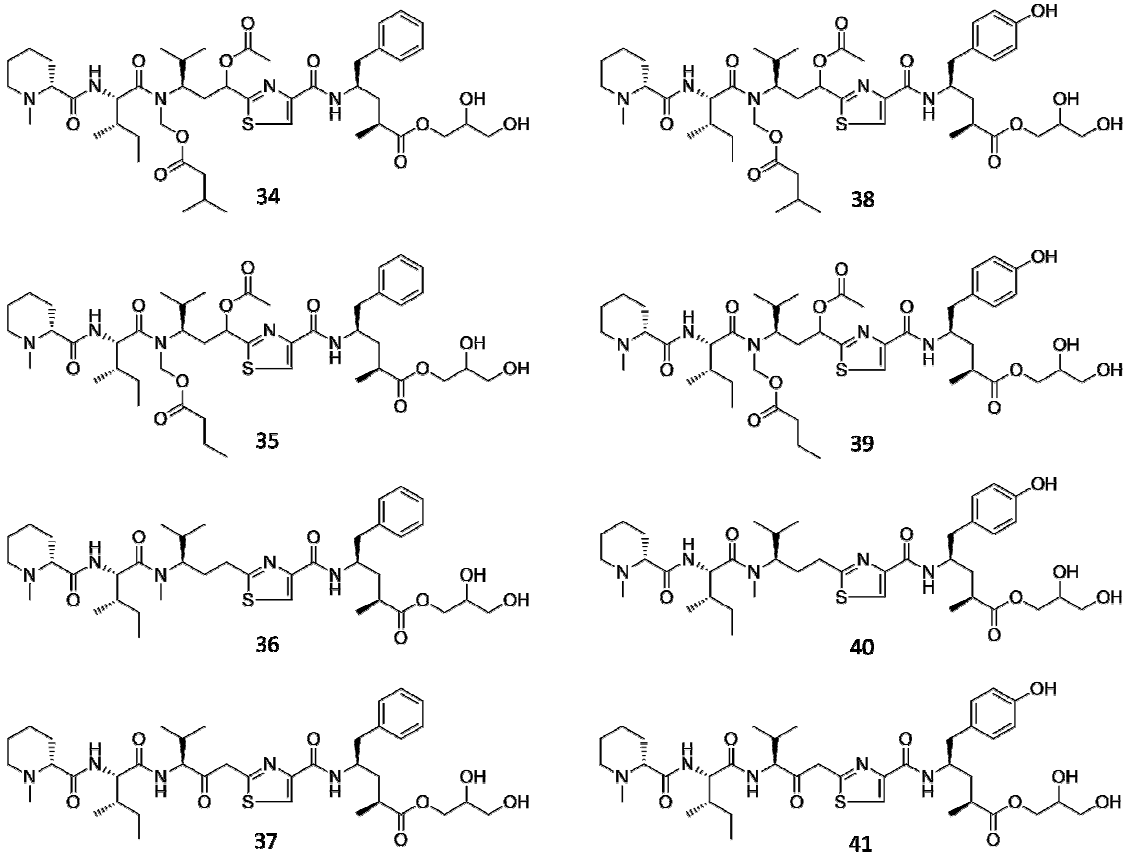


Figure 3.22 Proposed structures of tubulysin–glycerol esters.

Table 3.5 Major production profile of SBCb004 wild type and the patatin gene knockout mutants*.

	SBCB004 WT	SBCb004-*orf17*⁻ mutant	SBCb004-*orf18*⁻ mutant
N-desmethyl 12-keto pretubulysin A	5.4%	0.5%	1.2%
N-desmethyl 12-keto pretubulysin A-glycerol	3.4%	0.5%	12.6%
Pretubulysin A	8.6%	2.0%	6.3%
Pretubulysin A-glycerol	4.8%	4.6%	16.5%
Tubulysin A	20.3%	1.0%	4.0%
Tubulysin A-glycerol	0.7%	0.3%	50.5%
Tubulysin B	20.0%	0.6%	3.0%
Tubulysin B-glycerol	0.5%	0.4%	43.7%

*The peak area integration from the HPLC-MS spectrometry represents the yield of each compound, and the percentage was calculated relative to an internal standard Stigmatallin.

Based on the high homology of the two patatin genes *orf17* and *orf18* (similarity 80%), they were first supposed to carry out the same catalytic function in the tubulysin biosynthesis.

However, SBCb004-*orf17*⁻ mutant showed a different production profile comparing to that of the *orf18* inactivation mutant. Although the tubulyisn–glycerol esters were observed from *orf17* mutant, not like its partner *orf18* their yields were lower than that of the wild type strain. However, they were still produced in comparative amounts to the other none glycerol tubulysins since the yields of all tubulysin derivatives produced in SBCb004-*orf17*⁻ were very low (Table 3.4). One possible explanation for this might be that the inactivation of *orf17* affects the NRPS-PKS proteins' transcription in the tubulysin biosynthesis. In the tubulysin biosynthetic gene cluster, gene *orf17* overlapped with *tubF* for 83 bp. Therefore when gene *orf17* is knocked out by insertion mutagenesis, the truncated *orf17'* and the genes on the pTOPO vector backbone may affect, and further interrupt the transcription of TubF, resulting in a dramatic decrease of tubulysin production. Of course, more evidence will be needed to prove my assumption.

The above results drew me to consider the roles of patatin genes in the tubulysin biosynthesis. I supposed the tubulysin-glycerol ester to be the originally biosynthesized tubulysin products in the wild type strains. The glycerol might be attached to the C-terminal of tubulysin structure by the TE domain when the intermediate pretubulysin is released from the assembly multienzymes. As patatin-like proteins, both Orf17 and Orf18 are assigned the function of acyl hydrolase activity. Thus they are capable to cut the bound glycerol from the tubulysin ester molecular, resulting in the common known tubulysin structures. When patatin gene is inactivated, the hydrolysis step is jumped over so that the tubulysin–glycerol would be kept as major products instead. In the SBCb004 wild type, the pretubulysin–glycerol family compounds (**40** and **41**) are produced in similar level to their corresponding pretubulysins, but the tubulysin–glycerol family compounds (**38** and **39**) are yielded much less. This fact indicates that the patatin protein(s) prefer to use the matural tubulysins as their substracts and reveal better catalytic ability to them than the pretubulysin family compounds.

3.4.4 Promoter insertion into strain SBCb004

Developing natural products as drug leads critically depends on obtaining sufficient amounts of the target compounds. The tubulysins are produced at relatively low yields by the natural producer strains An d48 (0.48 mg/ml tubulysin D) and SBCb004 (0.35 mg/ml tubulysin A). Furthermore, the titers of the tubulysin variants are even lower. These observations prompted us to look for molecular methods for increasing the tubulysin yields. It has been previously demonstrated that overexpression of single genes or multigene transcriptional units by promoter

exchange in other myxobacterial strains can lead to increased secondary metabolite formation [98; 137-139]. Thus, I aimed to insert a strong constitutive Tn5-derived *npt* promoter [139] directly in front of the tubulysin core gene cluster *tubB–tubF* in strain SBCb004. This strategy had been used successfully for myxochromide and ajudazol overproduction in *M. xanthus* and *Chondromyces crocatus*, respectively [139]

Tn5 promoter insertion was also achieved by homologous recombination (experiment was carried out similarly as the other gene knockout) (Figure 3.23). And the generated clones bearing the inserted promoter (SBCb004-Tn5) were confirmed by PCRs and sequencing (data not shown). Cultures of SBCb004-Tn5 were grown in triplicate alongside the SBCb004 wild type strain. Yields of the tubulysins as judged by integration of peak areas in the base peak chromatagram were calibrated against a second set of metabolites produced by the SBCb004 strain, the stigmatellins A, B and C [140]. The amounts of three major products tubulysin A, pretubulysin A, and *N*-desmethyl 12-keto pretubublysin A were calculated as a percentage of the yields of the internal standard (Figure 3.23). Unfortunately, the inserted Tn5 promoter did not upregulate the production, but instead decreased the titers of all the major metabolites by one-third relative to the wild type strain. A possible explanation for this finding is that the promoter insertion method somehow affected normal expression of the tubulysin assembly line, but determining the exact cause of the reduction in yields will require further investigation.

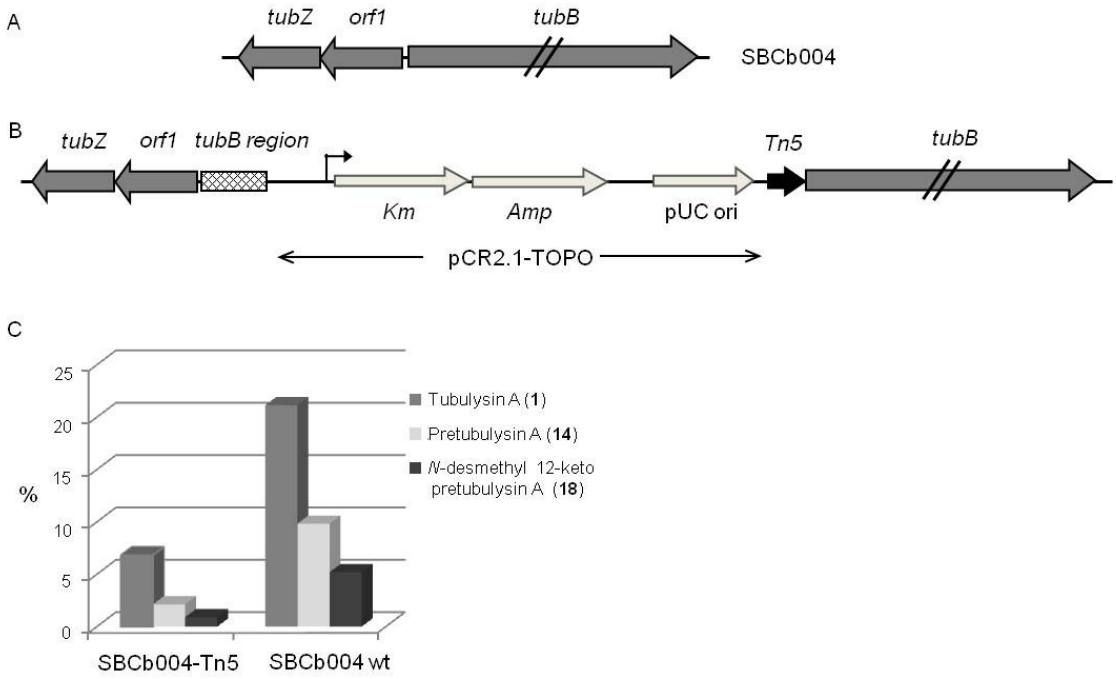


Figure 3.23 Tn5 promoter inserted mutant in SBCb004 strain: SBCb004-Tn5.
(A) Genetic organization of chromosomal DNA in SBCb004 wild type.
(B) Genetic organization of chromosomal DNA in mutant SBCb004-Tn5. From left to right: *tubZ*, *orf1*, *tubB*'s first region (1 kbp), pCR2.1-TOPO vector backbone, Tn5 promoter, complete gene *tubB*.
(C) Comparison of three major tubulysin products (tubulysin A [**1**], pretubulysin A [**14**], and *N*-desmethyl 12-keto pretubulysin A [**18**]) between SBCb004-Tn5 mutant and SBCb004 wild type. The second product of SBCb004 strain stigmatellins were used as the internal standard, and the compounds yields (%) in two strains were calculated by the percentage to the stigmatellins' production.

3.5 Heterologous expression and genetic engineering of the tubulysin biosynthetic gene cluster using Red/ET recombineering

In order to deepen the understanding of the post-assembly line aspects of tubulysin biosynthesis, and to facilitate the engineering of further derivatives for evaluation as drug leads, I endeavored to reconstitute the entire gene cluster using Red/ET recombineering technology.

3.5.1 Reconstitution of the tubulysin biosynthetic gene cluster of *Cystobactor* sp. SBCb004

The tubulysin biosynthetic gene cluster from strain SBCb004 was found to be located on two cosmids (Cos DL18 and AP17) within a cosmid genomic library (data not shown). Cos DL18 contained the 5′ end of the gene cluster (ending in the middle of *tubD*), while the 3′ end was located on Cos AP17 (starting in the middle of *tubC*) (Figure 3.22A and B); the region of overlap which encompassed the *tubC* and *tubD* genes, was ca. 14 kbp. As a first step in reconstituting the cluster, I replaced the pSuperCos cosmid backbone and the imcompleted *tubD* gene on Cos DL18 with a cassette containing the p15A replication origin (p15A ori), an inverted repeat (IR), and a kanamycin selection marker flanked by a *Nhe*I site, resulting in plasmid pUp (Figure 3.22C). p15A is a low copy origin, which has been found to aid efficient further Red/ET modification steps, the *Nhe*I site facilitated subsequent cluster assembly, while the inserted IR was part of the designed transposition system which enables efficient integration of the cluster into the host genome. Cos AP17 was also modified by replacement of the pSuperCos backbone as well as the *tubC* fragment with a gene cassette containing the pUC origin of replication and an ampicillin resistance gene flanked by a *Nhe*I site (Figure 3.22C). To enable efficient selection in the subsequent 'stiching' step, a zeocin resistance gene with a flanking *Nhe*I site was introduced at the opposite end of the cosmid to yield plasmid pDown (Figure 3.22C). Both pUp and pDown were then digested with *Nhe*I (Figure 3.22D). The resulting 29 kbp fragment from pDown was ligated with the linearized pUp, and the ligation mixture was transformed into *E. coli* GB05-red cells. Screening for zeocin resistance yielded the desired construct p15A-Tub which harbors the complete 11 tubulysin biosynthetic genes in the correct orientation (Figure 3.22E). The reconstituted tubulysin gene cluster on plasmid p15A-Tub is nearly identical to the parent in SBCb004. The one difference is that the wild type genes *tubC* and *tubD* overlap by 11 bp, while

on plasmid p15A-Tub, they are separated by 29 bp (which includes the introduced *Nhe*I site) (Figure 1E).

A transposition cassette comprised of two IRs, the corresponding MycoMar transposase (Tps), the origin of transfer oriT, and an Amp resistance gene, was designed to enable transfer and integration of the gene cluster into the chromosome of heterologous host strains. The mariner transposon MycoMar is frequently used in Gram-negative hosts for genetic modification. In the heterologous expression of epothilone and myxochromide S [98], the MycoMar transposon-mediated transformation of large gene was more efficient than that of homologous recombination using a smaller construct. Thus, as a powerful delivery tool for large transgenes, it was employed in the tubulysin heterologous expression system to facilitate the transformation and integration of the gene cluster into the genome of host strains. The incorporation of *oriT* was previously shown to be necessary for transposon-mediated integration of plasmids into the chromosome of the *P. putida* strain [98]. Introduction of this cassette *IR-Tps-Amp-oriT* (the other IR is already on the plasmid) into construct p15A-Tub was also accomplished enabling the deletion of zeocin and several open reading frames downstream of the cluster (*orf21*-orf22-*orf23*), resulting in plasmid pTps-Tub.

The natural start codon of *tubC* is the rare TTG. As I worried that this would limit the efficiency of expression in the target heterologous hosts [141], I also aimed to alter the codon using Red/ET recombineering. First, I tried to insert a gentamycin resistance gene in front of *tubC* as well as to change the start codon to ATG, but no correct clones were obtained: all of the resulting recombinants incorporated a mutation at the 5' end of *tubC*. This problem had been encountered previously during attempts to insert a constitutive promoter in front of the myxochromide gene cluster (Y. Zhang, unpublished result). It may be the case that the expression of the downstream tubulysin gene cluster (*tubC–tubF*) when driven by the gentamycin constitutive promoter, is toxic to the *E. coli* host. Therefore, to stable the tubulysin heterologous expression plasmid in *E.coli* strain, I aimed to insert an inducible promoter 5' of *tubC* to regulate the expression of following genes. For this, I generated a gene cassette containing a Genta resistance gene and an inducible promoter tetR-P_{tet}. The tetR-P_{tet} is a tetracycline (Tet) based regulatory system which is generally used to selectively control downstream gene expression. It includes a tetracycline resistance gene *tetR* (encoding the tetracycline repressor protein TetR) and an 81 bp gene sequence which harbors bidirectional promoters which overlap with two highly homologous control sequences. In the absence of the inducer tetracycline, TetR protein dimers inhibit the bidirectional promoters by

binding to the two control sequences, shutting down the transcription of following gene/operon under P_{tet} which is opposite to *tetR*. When tetracycline is present, however, the TetR protein complexes with it in preference to the control region, resulting in expression of the target genes. The cassette was amplified by PCR using primers with short homology arms to enable integration between the *tubB* and *tubC* genes, as well as to modify the start codon; one version contained the original TTG start codon of *tubC*, and the second, a mutant ATG. The resulting fragments were then inserted into pTps-Tub. This strategy yielded the desired plasmids pTub-TTG and pTub-ATG (Figure 1F) (The construct pTuy-TTG was generated by Shiping Shan). In these plasmids, the gentamycin resistance gene and the tetracycline inducible promoter tetR-P_{tet} are located in front of *tubC*, being a difference to the wild type cluster. The integrity of all obtained recombination constructs was verified by restriction analysis using a diagnostic set of enzymes, or by sequencing.

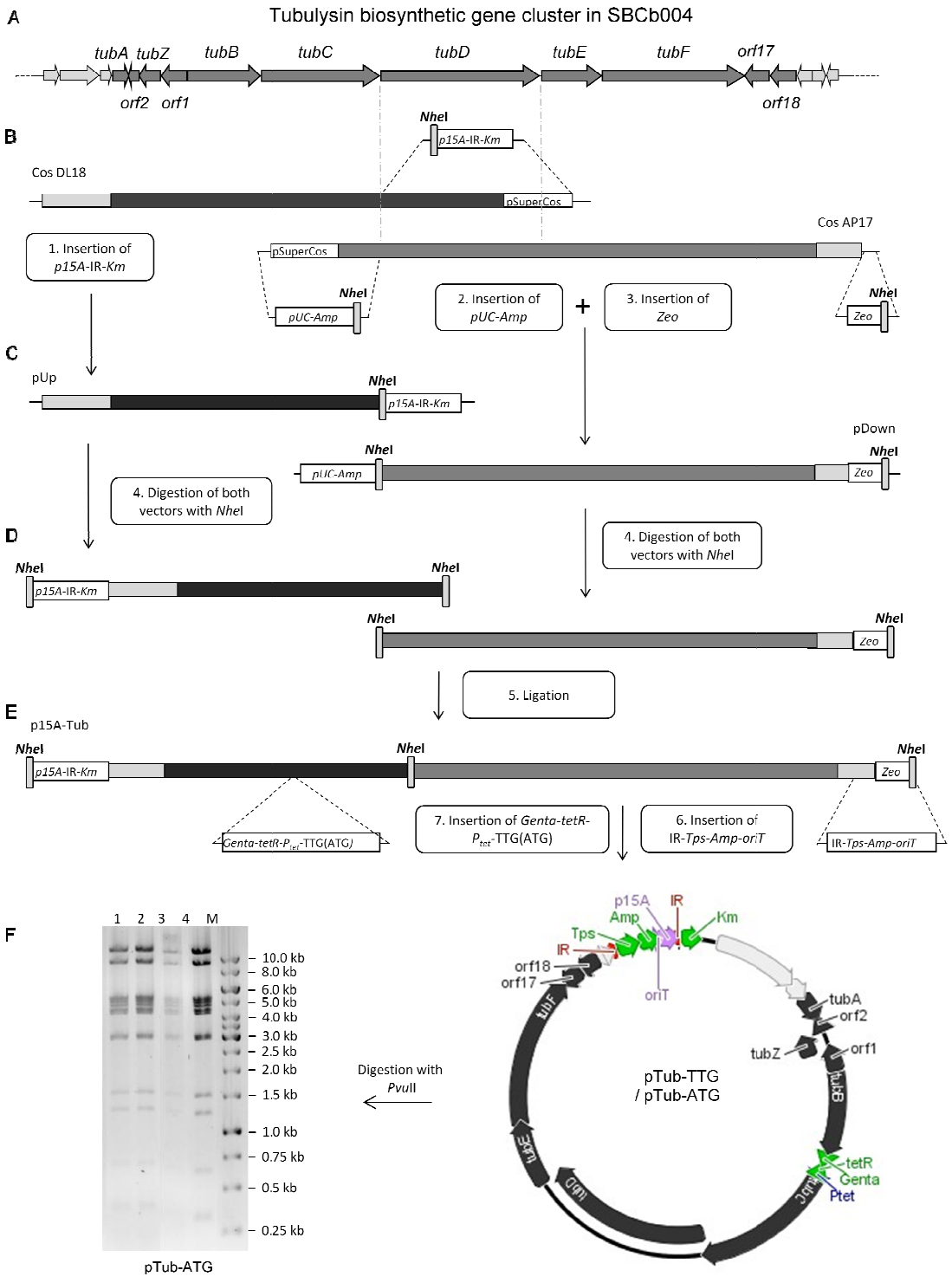


Figure 3.24 Description of the cloning strategy of tubulysin biosynthetic gene cluster reconstitution.
(A) The indentified tubulysin biosynthetic gene cluster from strain SBCb004.
(B) Two overlapping cosmids Cos DL18 (harboring 5′- end tub gene cluster) and Cos AP17 (harboring 3′- end tub gene cluster).

(C) A *p15A-IR-Km* cassette was used to modify Cos DL18 via Red/ET recombineering to replace the pSuperCos backbone as well as the *tubD* portion, resulting in plasmid pUp. Two rounds of Red/ET recombinations were carried out on Cos AP17, leading to the plasmid pDown. Cassette *pUC-Amp* was used to exchange the pSuperCos backbone and the *tubC* portion, while the inserted zeocin restriction gene was used as the selection marker in the following stitching step.
(D) Digest both plasmids with *Nhe*I. Plasmid pUp was linearized, while plasmid pDown was cut into two pieces: a 29 kb fragment (with 3′ end tub gene cluster) and a 4 kb fragment.
(E) Ligate the linearize pUp and 29 kb digested DNA fragment from pDown together to stitch the entire tub gene cluster together into one plasmid p15A-Tub.
(F) Insert *IR-Tps-Amp-oriT* cassette onto the construct to generate a transpositional system. Then, two parallel modification cassettes *Genta-TetR-*P_{tet}-TTG *and Genta-TetR-*P_{tet}-ATG were inserted in front of gene *tubC*, resulting in two heterologous expression construct pTub-TTG and pTub-ATG, respectively. The constituted plasmids were confirmed by *Pvu*II restriction analysis and sequencing. In the DNA gel figure, line "1–4" represented 4 distinct clones of plasmid pTub-ATG, whild "M" represented the DNA marker.

3.5.2 Heterologous expression of the tubulysin gene cluster in *Pseudomonads putida*

Both constructs pTub-TTG and pTub-ATG were introduced into the host organism *P. putida* by triparental conjugation, with selection for kanamycin resistance on PMM medium (The experiments with strain *P. putidas* were carried out by Shiping Shan, including the transformation, fermentation, and product extraction.) For each construct, 5–10 randomly chosen clones were cultivated and the integration of the constructs into the genome was verified by colony PCR. Positive clones were then cultivated in PMM medium in the presence of tetracycline to activate the TetR-P_{tet} promoter, and the culture extracts were analyzed by HPLC-MS.

Two members of the pretubulysin subfamily, pretubulysin A and tyrosine pretubulysin A, were observed in the extracts of both *P. putida*::pTub-TTG and *P. putida*::pTub-ATG strains, although they were both produced at disappointingly low levels (the yield of pretubulysin A (**14**) from *P. putida*::pTub-TTG was approximately 0.04 μg/L, and that from *P. putida*::pTub-ATG was 0.2 μg/L.) (Figure 3.25). This result demonstrates that, as predicted, the core gene cluster contains all genes required to produce the polyketide-nonribosomal peptide skeleton of the tubulysins [102; 127]. However, the fact that neither of the products incorporated the acyl groups of the mature tubulysins, demonstrated that at least some of the post-assembly line genes are absent from the heterologous expression construct (i.e. the acyl transferase might have been present, but unable to act due to the absence of a required oxidase). Alternatively, but much less likely some genes may not be expressed because the promoter(s) are not recognized in the heterologous host. Unexpectedly, I also detected trace amounts of tubulysin A in *P. putida*::pTub-TTG(ATG). A possible explanation for this finding is that another oxidase and/or acyltransferase encoded in the *P. putida* chromosome was able to complement the missing post-modification function(s) in

tubulysin A biosynthesis. I have previously observed a similar complementation event in strains An d48 and SBCb004: inactivation of the cyclodeaminase gene *tubZ* did not abolish biosynthesis of the tubulysins, and instead the mutant strains generated substantial amounts of pretubulysin D/A as well as small quantities of tubulysin D/A. This result suggested that the genomes harbored additional cyclodeaminase-encoding genes, and indeed BLAST analysis of the SBCb004 genome revealed two genes with homology to ornithine cyclodeaminasese [109; 127]. I assume that the *P. putida* genome encodes a number of oxidases and acyltransferases, and their efficiency with the tubulysins as substrates is low, explaining the small yield of tubulysin A from the heterologous expression strain.

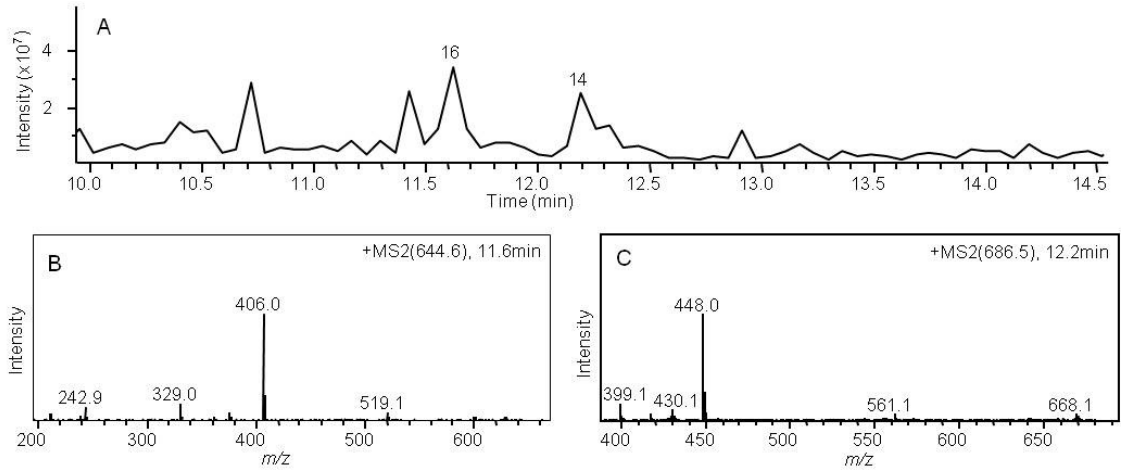


Figure 3.25 Main produced tubulysins by *P.putida*::pTub-ATG.
0.1 mg/ml pipecolic acid was supplemented, and cultivation is at 30 °C.
(A) HPLC-MS analysis (base peak chromatogram [BPC]) of *P. putida*::pTub-ATG.
(B) The MS_2 fragment pattern of tyrosine pretubulysin A (**33**).
(C) The MS_2 fragment pattern of pretubulysin A (**11**).

L-pipecolic acid serves as the starter unit for tubulysin biosynthesis. Reasoning that the Pip supply might limit the overall yields, attempt was carried out to increase tubulysin production by feeding Pip to cultures of *P. putida*::pTub-TTG(ATG). Although racemic D, L-pipecolic acid was used in all the experiments, only the L-isomer is incorporated into the tubulysins (Feeding experiment was carried out in M. xanthus::pTub-ATG. Data was shown in the following chapter, in Figure 3.27) [127]. As hoped, a substantial increase in tubulysin titers was obtained. For example, the production of pretubulysin A (**14**) by *P. putida*::pTub-ATG increased from 0.2 µg/L to 1.76 µg/L in the presence of 0.1 mg/ml supplemental Pip. Thus, it appears that the activity of the heterologous TubZ in *P. putida* limits the overall yields from the biosynthetic pathway.

In order to explore the effect of the start codon on tubulysin biosynthesis in the heterologous expression strain, I compared the yields of pretubulysin A (**14**) from *P. putida*::pTub-TTG and *P. putida*::pTub-ATG. Production by strain *P. putida*::pTub-ATG containing the modified ATG start codon was higher than that incorporating the wild type TTG (0.2 μg/L vs. 0.04 μg/L). Similarly, in the presence of 0.1 mg/ml supplemental Pip, production of pretubulysin A (**14**) by the ATG-modified strain was 1.76 μg/L, approximately three-fold higher than that by the TTG strain (0.66 μg/L). This result supports the idea that the rare TTG codon can inhibit translation of TubC and ultimately lower the overall yields of the tubulysins. Thus, optimization of start codons may be a useful general strategy for improving titers from heterologous expression constructs in the future.

During heterologous expression of myxochromide S in *P. putida*, it was noted that the cultivation temperature dramatically affected the product yield: cultivation at 16 °C resulted in 1000-fold more myxochromide S than incubation at 30 °C [71]. In order to test the effect of temperature on tubulysin heterologous expression in *P. putida*, *P. putida*::pTub-ATG was cultivated at both 30 and 16 °C. Surprisingly, an opposite result was obtained: under supplementation with 0.1 mg/ml Pip, the yield of pretubulysin A (**14**) at 30 °C (1.76 μg/L) was 2-fold higher than that at 16 °C (0.84 μg/L). In the case of myxochromide S, a P_m promoter was introduced in front of the gene cluster in order to boost metabolite production by up-regulating expression of the biosynthetic genes. Thus, high expression at 30 °C may have resulted in improper folding and/or aggregation of the large PKS and NRPS proteins, so that the overall yield of myxochromides was reduced. In the case of tubulysin, however, expression is under the presumably weak, native promoter(s) from the SBCb004 genome. Thus, the influence of temperature on the *P. putida*::pTub-ATG strain is relatively small, and consistent with the general preference of *P. putida* for growth at 30 °C.

3.5.3 Heterologous expression of the tubulysin gene cluster in *Myxococcus xanthus*

Given the closer relationship between *Myxococcus xanthus* and the native tubulysin producer SBCb004, I anticipated that *M. xanthus* might – despite its inferior growth characteristics – be a better heterologous expression host strain for tubulysin biosynthesis. To test this hypothesis, construct pTub-ATG was transformed into *M. xanthus* DK1622 by electroporation followed by selection for kanamycin resistance. Randomly selected clones were then cultivated and the integration of gene cluster was verified by colony PCR (Figure 3.26). Correct *M. xanthus*::pTub-

ATG clones were then fermented in CTT liquid medium under tetracycline induction, and the culture extracts were analyzed by HPLC-MS. Expression of the gene cluster in DK1622 produced a similar metabolite profile as *P. putida*::pTub-ATG, but the tubulysins were obtained at higher yield (Figure 3.27). For example, in the presence of 0.1 mg/ml supplemental Pip, the yield of pretubulysin A (**14**) from *M. xanthus*::pTub-ATG (0.19 mg/L) was 100-fold higher than that from *P. putida*::pTub-ATG (1.76 μg/L). In addition, due to the higher titres, I was able to detect more tubulysin variants from the *M. xanthus*::pTub-ATG strain via their MS^2 fragmentation pattern during HPLC-MS analysis. These variants included tyrosine pretubulysin A (**33**), *N*-desmethyl 12-keto pretubulysin A (**25**), *N*-desmethyl 12-hydroxy pretubulysin A (**26**), *N*-desmethyl pretubulysin A (**27**), and also trace amounts of tubulysin A (**6**) (Figure 3.1 and Table 3.3). Again, biosynthesis of tubulysin A is likely due to adventitious action of oxidase and/or acyltransferase enzymes encoded in the *M. xanthus* DK1622 genome.

Supplementation with Pip also increased production of *M. xanthus*::pTub-ATG. In order to quantify the relationship between the Pip supply and pretubulysin production, I fed various amounts of Pip to *M. xanthus*::pTub-ATG. In order to estimate the production yield, a reference curve was generated using synthetic pretubulysin D (**5**) (kind gift of Prof. U. Kazmaier) [109]. Assuming that the ionization efficiency of the derivatives was the same as that for pretubulysin D, the yields of all tubulysins were judged based on their relative peak areas in the HPLC-MS chromatogram in comparison with the standard curve. Four individual clones of *M. xanthus*::pTub-ATG were grown in parallel, and the quantities of the two major products pretubulysin A (**11**) and tyrosine pretubulysin A (**33**) were determined from the average of the three values. Yields from the SBCb004 wild type strain were used as a benchmark (Figure 3.28). As the integration of the cluster into the *M. xanthus* genome was transposon-mediated, the site of integration is likely to vary between randomly-selected clones. In the case of *Streptomyces*, it has been observed that integration into different loci can affect the protein expression and production yield [142]. However, the four randomly selected clones, which likely represent different integration sites (though this was not directly determined), produced similar amounts of the tubulysins. For example, in the presence of 0.08 mg/ml supplementary Pip, the yields of pretubulysin A (**11**) from the four clones were 0.176, 0.154, 0.165, and 0.173 mg/L, respectively (average yield 0.167 ± 0.008 mg/L). A similar result was also observed when this marine transposon strategy was used to integrate epothilone and myxochromide gene clusters into *M. xanthus* [98].

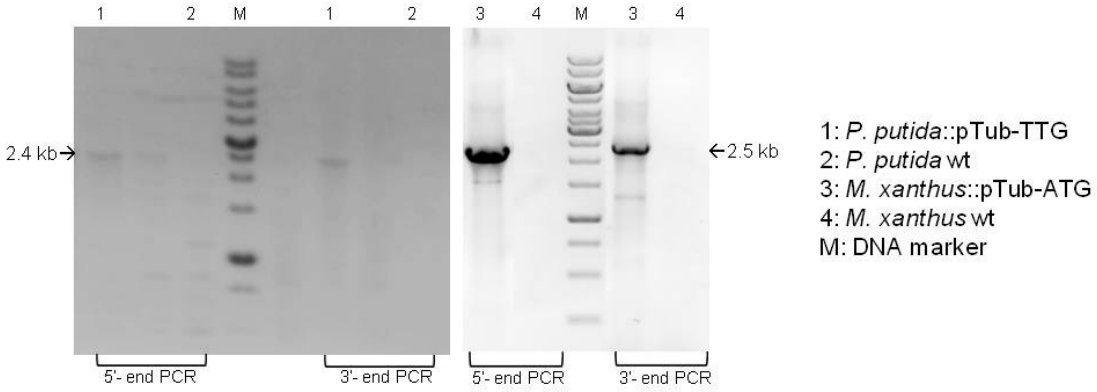


Figure 3.26 Clony PCR of tubulysin heterologous expression construct in *P. putida* and *M. xanthus*. 5′ end PCR amplified the *tubA-orf2* region (2.4 kb) of the tub gene cluster with primer Cbt-A3 and Cbt-A4, and the 3′ end PCR amplified the *orf17-orf18* region (2.5 kb) with primer Cbt-H3 and Cbt-G4 (Table 2.7).

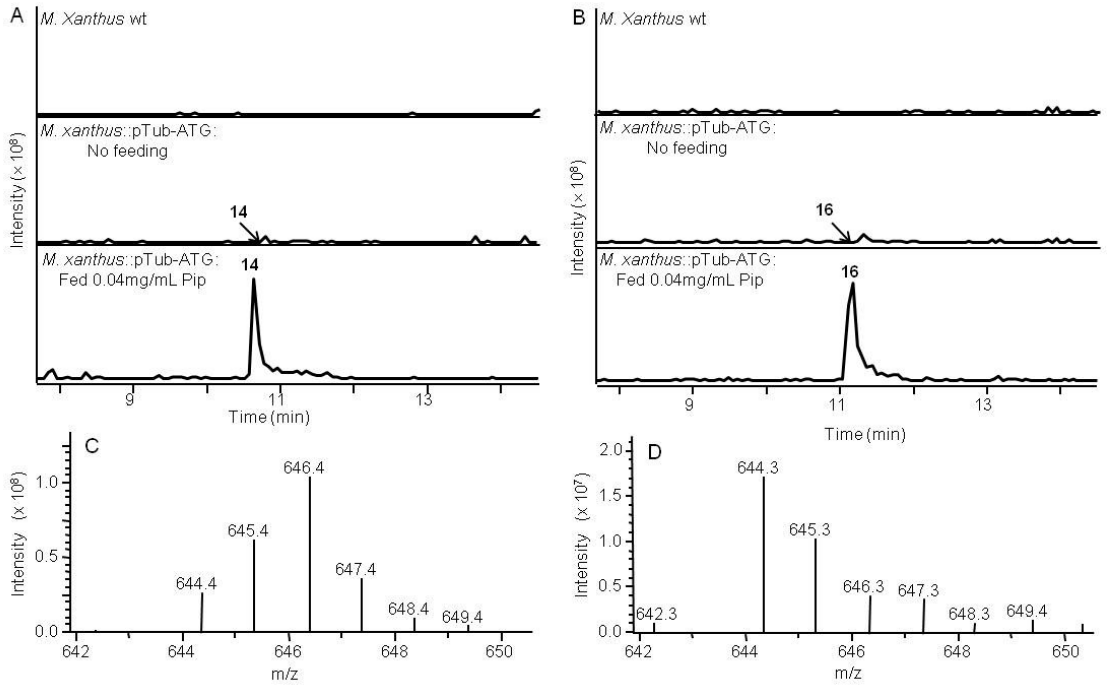


Figure 3.27 Feeding studies with pipecolic acid in strain *M. xanthus*::pTub-ATG.
(A) HPLC-MS Analysis (Extracted Ion Chromatograph [EIC]) at *m/z* 686.5 $[M+H]^+$ (pretubulysin A [**14**]). Fed pipecolic acid is a L-/D- mixture.
(B) HPLC-MS Analysis (Extracted Ion Chromatograph [EIC]) at *m/z* 644.4 $[M+H]^+$ (tyrosine pretubulysin A). Fed pipecolic acid is a L-/D- mixture.
(C) HPLC-MS Analysis: supplementation with di-deuterium labeled L-pipecolic acid.
(D) HPLC-MS Analysis: supplementation with di-deuterium labeled D-pipecolic acid.

In the case of SBCb004, production of both compounds initially increased during supplementation with Pip (to 0.14 mg/ml), but declined thereafter, suggesting an inhibitory effect on the cells. The yields of both metabolites from *M. xanthus*::pTub-ATG also initially increased with the concentration of Pip, but no further augmentation was seen at concentrations above 1 mg/ml (Figure 3.28). In addition, in the case of pretubulysin A, the overall yields remained approximately 50% lower than that of supplemented SBCb004, although they were comparable to the non-supplemented wild type. These results confirm that the supply of L-pipecolate is limiting to tubulysin biosynthesis in both strains, but equally that supplementation can only boost metabolite quantities to a modest extent. It is worth mentioning that the yield of tubulysin is currently limiting further development of the compounds as neither synthetic nor fermentative approaches have led to sufficient production strategies. Only the less potent pretubulysin offers access in reasonable yields via chemical synthesis [109].

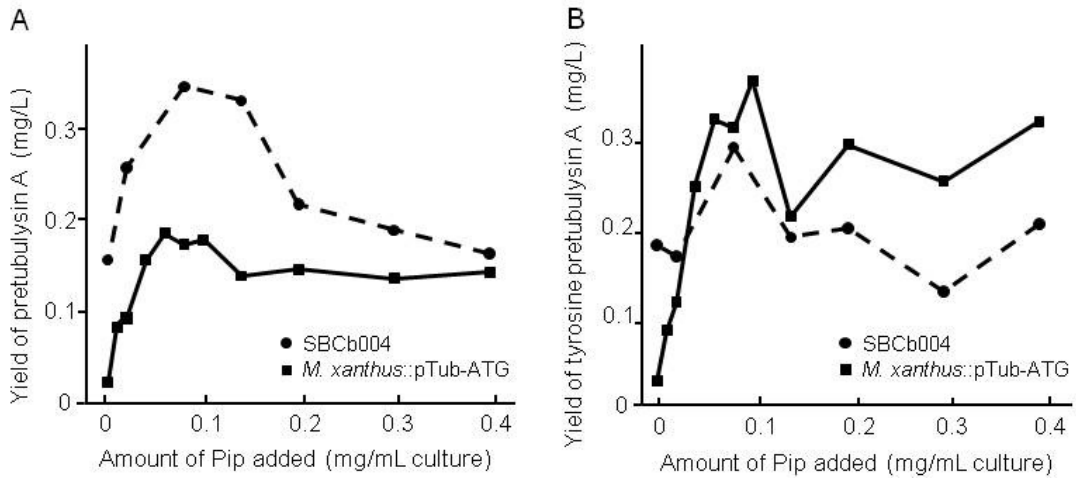


Figure 3.28 Feeding pipecolic acid into strain SBCb004 (dashed line) and *M. xanthus*::pTub-ATG (solid line).
(A) Yield of pretubulysin A.
(B) Yield of tyrosine pretubulysin A.

3.5.4 Genetic modification of the tubulysin heterologous expression system

As noted earlier, the conserved tubulysin biosynthetic gene cluster in both An d48 and SBCb004 strains contains several open reading frames whose functions have been unclear to date. These include the genes *tubA*, *orf2*, and *orf1* at the 5′ end of the gene set. The observation that they share a high degree of mutual sequence homology (56–78% identity/73–90% similarity) [127] suggests that they play a role in the biosynthesis. To obtain further insights into the

potential functions of these genes in the tubulysin pathway, I aimed to inactivate these genes on the pTub-ATG expression construct.

For this, I replaced the target genes by a chloramphenicol (Cm) resistance gene, using Red/ET recombineering. As genes *orf1, tubZ,* and *orf2* are likely to form an operon, deletion of any gene would likely affect the function of all downstream gene(s) due to polar effects. Therefore, in order to obtain the clearest picture from the results, in each case, I deleted the target gene together with any downstream gene(s) (Figure 3.30). The deletion constructs pTubA, pOrf2, pTubZ (deletion of both *tubZ* and *orf2*), and pOrf1 (deletion of the entire *orf1-tubZ-orf2* operon) were screened using low salt LB plates containing Amp and Cm, and then verified by restriction analysis (Figure 3.29). The modified constructs were transformed into *M. xanthus* DK1622, and the production profile of positive clones was analyzed by HPLC–MS.

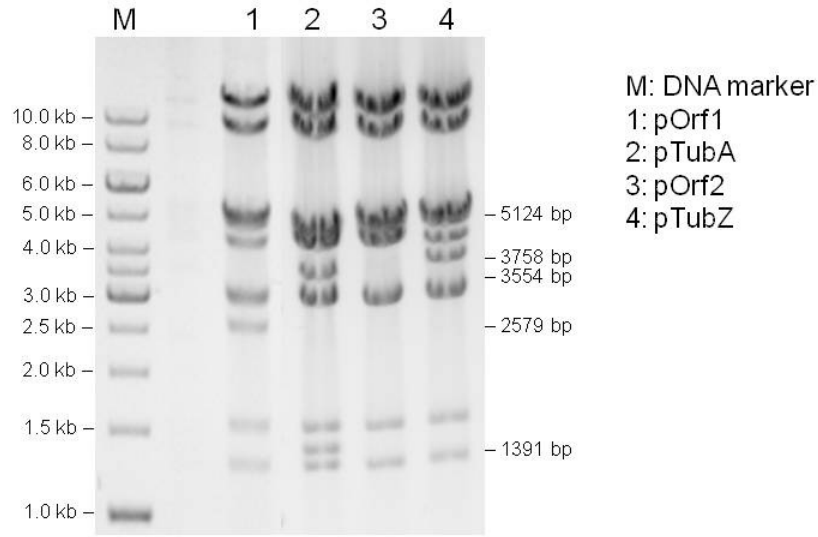


Figure 3.29 Restriction analysis of the gene deletion constructs with enzyme *Pvu*II.
The size of restricted bands that changed in the deletion constructs are shown as following. The changes of band are based on the restriction of construct pTub-ATG. And only the bands bigger than 1 kb are listed.
pOrf1: Arised band: 5124 bp, 2579 bp; Disappeared band: 5285 bp, 4447 bp.
pTubA: Arised band: 3554 bp, 1391 bp; Disappeared band: 5285 bp.
pOrf2: Arised band: 5124 bp; Disappeared band: 5285 bp.
pTubZ: Arised band: 5124 bp, 3758 bp; Disappeared band: 5285 bp, 4447 bp.

The resulting strains still produced both pretubulysin A (**11**) and tyrosine pretubulysin A (**33**), albeit in reduced yields. For example, when fed with saturating Pip (0.08 mg/ml), the titers of pretubulysin A from all mutants was lower than that from *M. xanthus*::pTub-ATG strain (pTubA

24% relative yield, pOrf2 45%, pTubZ 51%, and pOrf1 18%) (Figure 3.30). In addition, even with Pip supplementation (0.08 mg/ml), no tubulysin A was detected from these strains. The production of the pretubulysin subfamily compounds in the absence of *orfs 1*, *2* and *tubA* indicates that these genes are not essential for assembling the tubulysin core structure. However, the reduction in yield relative to that from the pTub-ATG strain shows that they nonetheless contribute to pretubulysin biosynthesis in some manner.

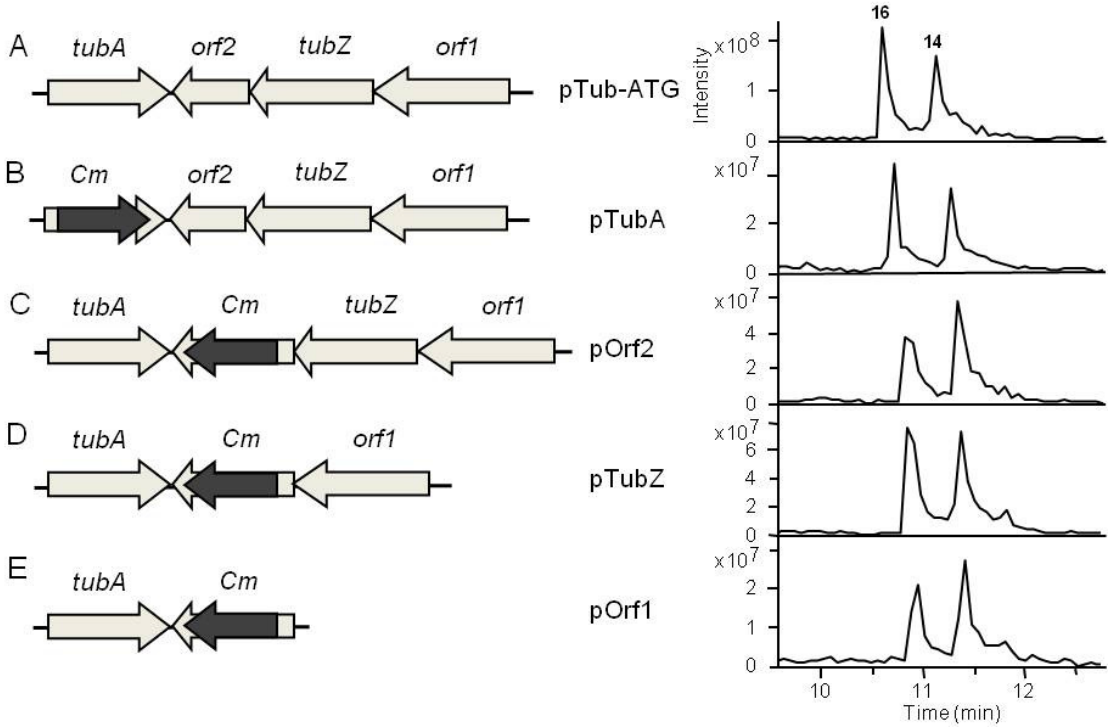


Figure 3.30 Genetic modification of tubulysin heterologous expression system.
Left: Tubulysin 5′ end gene deletion in the deletion constructs. Right: The HPLC-MS analysis (Extracted Ion Chromatograph [EIC]) at *m/z* 686.5 $[M+H]^+$ (pretubulysin A) and *m/z* 644.4 $[M+H]^+$ (tyrosine pretubulysin A) of the corresponding deletion constructs in *M. xanthus* strain. 0.08 mg/ml pipecolic acid was supplemented.
(A) *M. xanthus*::pTub-ATG strain.
(B) *M. xanthus*::pTubA deletion strain
(C) *M. xanthus*::pOrf2 deletion strain.
(D) *M. xanthus*::pTubZ deletion strain.
(E) *M. xanthus*::pOrf1 deletion strain.

3.6 Discovery of the missing monooxygenase in the tubulysin biosynthetic gene cluster

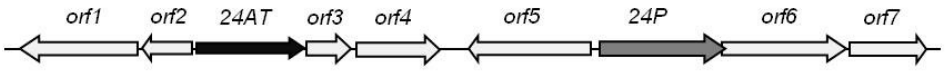
All evidence suggests that the tubulysin oxidase(s) and possibly a second acyl transferase are absent from both clusters in strains An d48 and SBCb004, and so are presumed to be located elsewhere in the genomes. Such split cluster PKS/NRPS architecture is unusual for *Streptomyces*, but has increasing precedent in the myxobacteria [119-121]. Therefore, the BLAST analysis of the SBCb004 genome data was used to identify multiple candidate monooxygenases, whose involvement in tubulysin assembly was evaluated using a combination of knockout mutagenesis and heterologous expression.

3.6.1 Identification of candidate genes for the missing monooxygenase in the SBCb004 genome

In order to locate the missing oxygenase(s), a P450-encoding gene *ajuJ* from the ajudazol biosynthetic gene cluster of the myxobacterium *Chondromyces crocatus* Cm c5 [143] was used as a probe in a BLAST analysis [115] of whole genome sequencing data from strain SBCb004, obtained by shot-gun sequencing using 454 technology [113; 114]. This analysis identified an unprecedented number of 49 candidate oxygenases. Annotation of the flanking regions in each case identified six P450 genes and one monooxygenase gene that were co-located with at least one acyl transferase, indicating their possible involvement in the tubulysin pathway (Figure 3.31).

A

Portion of Contig 24 of SBCb004 genome



Proteins Encoded in the Portion of Contig 24 of SBCb004 Genome and their Putative Functions.

Protein	aa	Proposed function of the homologous protein	Source of the homologous protein	Similarity/ identity
Orf1	374	Cytochrome c peroxidase	*Myxococcus xanthus*	75%/65%
Orf2	172	Hypothetical protein MXAN_5564	*Myxococcus xanthus*	64%/52%
24AT	415	Acyltransferase family protein	*Myxococcus xanthus*	69%/56%
Orf3	138	Ypothetical protein Noc_2710	*Nitrosococcus oceani*	46%/35%
Orf4	361	Cytochrome c551 peroxidase precursor	*Sorangium cellulosum*	72%/60%
Orf5	391	DTW domain protein	*Myxococcus xanthus*	80%/70%
24P	435	Probable p450	*Kitasatospora griseola*	62%/45%
Orf6	428	Hypothetical protein MXAN_6922	*Myxococcus xanthus*	42%/30%
Orf7	249	Conserved hypothetical protein TIGR00046	*Myxococcus xanthus*	89%/84%

B

Portion of Contig 45 of SBCb004 genome

Proteins Encoded in the Portion of Contig 45 of SBCb004 Genome and their Putative Functions.

Protein	aa	Proposed function of the homologous protein	Source of the homologous protein	Similarity/ identity
Orf1	322	Transcriptional Regulator, AraC family protein	*Plesiocystis pacifica*	51%/35%
45AT	191	GCN5-related N-acetyltransferase	*Mycobacterium sp.*	42%/29%
Orf2	157	Hypothetical protein PPSIR1_16180	*Plesiocystis pacifica*	63%/47%
Orf3	328	Putative AraC-family transcriptional regulator	*Plesiocystis pacifica*	68%/57%
45M	596	Monooxygenase, flavin-binding family	*Myxococcus xanthus*	86%/75%
Orf4	199	Transcriptional regulator, TetR family	*Myxococcus xanthus*	82%/66%
Orf5	281	Short-chain dehydrogenase/reductas	*Mycobacterium abscessus*	75%/64%
Orf6	326	Alpha/beta hydrolase fold-3	*Mycobacterium sp.*	66%/57%
Orf7	424	Putative Beta-lactamase	*Bradyrhizobium sp.*	52%/35%
Orf8	557	Putative oxidoreductase	*Janibacter sp.*	68%/52%
Orf9	246	Hypothetical protein MED297_04739	*Reinekea sp.*	76%/58%
45P	1063	Cytochrome P450	*Herpetosiphon aurantiacus*	77%/59%
Orf10	218	Transcriptional regulatory protein	*Bradyrhizobium japonicum*	57%/34%

C

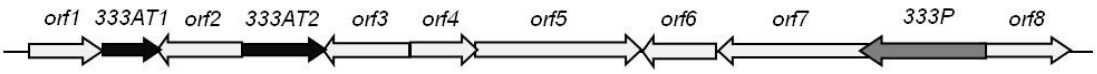
Portion of Contig 302 of SBCb004 genome

Proteins Encoded in the Portion of Contig 302 of SBCb004 Genome and their Putative Functions.

Protein	aa	Proposed function of the homologous protein	Source of the homologous protein	Similarity/ identity
Orf1	407	Hypothetical protein Acid_1970	*Solibacter usitatus*	65%/51%
302P	412	Cytochrome P450 CYP1	*Sorangium cellulosum*	73%/58%
Orf2	965	Probable oxidoreductase	*Roseovarius sp.*	53%/39%
Orf3	175	Ferredoxin	*Roseovarius sp.*	70%/52%
Orf4	333	Conserved hypothetical protein	*Burkholderia ambifaria*	58%/46%
Orf5	500	Endo-beta-1,6-galactanase	*Stigmatella aurantiaca*	97%/95%
Orf6	383	Oxidoreductase domain protein	*Sphingomonas wittichii*	52%/36%
Orf7	186	Isochorismatase hydrolase	*Acidobacteria bacterium*	68%/54%
Orf8	1333	Protein kinase	*Sorangium cellulosum*	58%/39%
Orf9	264	MxcB	*Stigmatella aurantiaca*	77%/64%
302AT	81	GCN5-related N-acetyltransferase	*Rhizobium leguminosarum*	44%/30%
Orf10	268	Autotransporter-associated beta strand repeat protein	*Opitutus terrae*	69%/51%

D

Portion of Contig 333 of SBCb004 genome

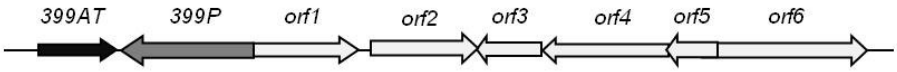


Proteins Encoded in the Portion of Contig 333 of SBCb004 Genome and their Putative Functions.

Protein	aa	Proposed function of the homologous protein	Source of the homologous protein	Similarity/ identity
Orf1	332	Transcriptional regulator, AraC family	*Methylococcus capsulatus*	82%/65%
333AT1	186	Maltose o-acetyltransferase	*methanogenic archaeon*	72%/57%
Orf2	343	Hypothetical protein OsJ_028952	*Oryza sativa*	48%/31%
333AT2	278	Sialate O-acetylesterase	*Opitutaceae bacterium*	47%/31%
Orf3	267	Hypothetical protein STIAU_1365	*Stigmatella aurantiaca*	47%/34%
Orf4	576	Hypothetical protein STIAU_2917	*Stigmatella aurantiaca*	57%/47%
Orf5	258	Putative serine/threonine protein kinase	*Stigmatella aurantiaca*	55%/44%
Orf6	660	Predicted: similar to differentially expressed in FDCP 8	*Pan troglodytes*	60%/35%
Orf7	411	Hypothetical protein Adeh_3596	*Anaeromyxobacter dehalogenans*	41%/28%
333P	411	Cytochrome P450 family protein	*Myxococcus xanthus*	60%/41%
Orf8	350	UbiA prenyltransferase	*Chloroflexus aggregans*	61%/47%

E

Portion of Contig 399 of SBCb004 genome

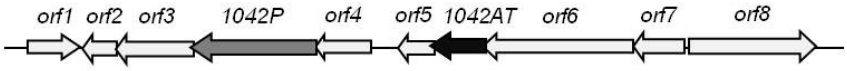


Proteins Encoded in the Portion of Contig 45 of SBCb004 Genome and their Putative Functions.

Protein	aa	Proposed function of the homologous protein	Source of the homologous protein	Similarity/ identity
399AT	293	Acetyltransferase	*Myxococcus xanthus*	87%/76%
399P	459	Cytochrome P450 family protein	*Myxococcus xanthus*	61%/47%
Orf1	349	Methyltransferase type 11	*Roseiflexus castenholzii*	45%/30%
Orf2	358	L-asparaginase	*Stigmatella aurantiaca*	81%/70%
Orf3	218	Phosphoribosyltransferase	*Rubrobacter xylanophilus*	67%/51%
Orf4	447	Nicotinate phosphoribosyltransferase	*Myxococcus xanthus*	79%/68%
Orf5	153	Thioredoxin 2	*Stigmatella aurantiaca*	81%/66%
Orf6	530	Universal stress protein family	*Stigmatella aurantiaca*	77%/61%

F

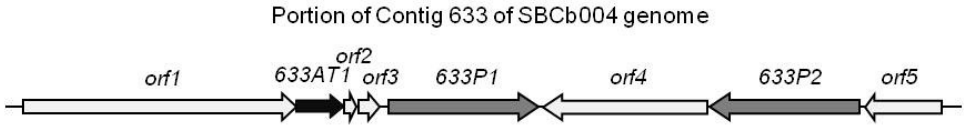
Portion of Contig 1042 of SBCb004 genome



Proteins Encoded in the Portion of Contig 1042 of SBCb004 Genome and their Putative Functions.

Protein	aa	Proposed function of the homologous protein	Source of the homologous protein	Similarity/ identity
Orf1	139	Transcriptional regulator, MarR family	*Frankia sp.*	68%/47%
Orf2	118	Ripartite motif-containing 11	*Homo sapiens*	41%/29%
Orf3	308	Ribonuclease Z	*Plesiocystis pacifica*	69%/53%
1042P	440	COG2124: Cytochrome P450	*Nostoc punctiforme*	53%/33%
Orf4	168	Hypothetical protein Mmar10_0508	*Maricaulis maris*	59%/45%
Orf5	113	Hypothetical protein STIAU_3562	*Stigmatella aurantiaca*	53%/46%
1042AT	144	Acetyltransferase, GNAT family family	*Bacillus cereus*	67%/45%
Orf6	545	Phosphoglucomutase, alpha-D-glucose phosphate-specific	*Stigmatella aurantiaca*	92%/85%
Orf7	157	Putative lipoprotein	*Myxococcus xanthus*	48%/36%
Orf8	417	Glucose-1-phosphate adenylyltransferase	*Stigmatella aurantiaca*	90%/82%

G



Proteins Encoded in the Portion of Contig 633 of SBCb004 Genome and their Putative Functions.

Protein	aa	Proposed function of the homologous protein	Source of the homologous protein	Similarity/ identity
Orf1	582	Hypothetical protein MXAN_6283	*Myxococcus xanthus*	79%/71%
633AT1	164	Acetyltransferase, gnat family	*Stigmatella aurantiaca*	54%/37%
Orf2	65	TonB-system energizer ExbB	*Physcomitrella patens subsp.*	55%/46%
Orf3	79	Predicted protein	*Methylobacterium sp.*	56%/31%
633P1	395	Cytochrome P450 CYP109C2	*Sorangium cellulosum*	86%/75%
Orf4	456	Putative Na/H(+) antiporter	*Azoarcus sp.*	60%/43%
633P2	447	Cytochrome P450	*Aedes aegypti*	51%/32%
Orf5	314	FMN-dependent alpha-hydroxy acid dehydrogenase	*Bacterium Ellin*	70%/53%

Figure 3.31 Annotation of the flanking regions of these identified monooxygenase genes from seven contigs of the SBCb004 genome.
(A) Gene annotation from contig 24 in SBCb004 Genome.
(B) Gene annotation from contig 45 in SBCb004 Genome.
(C) Gene annotation from contig 302 in SBCb004 Genome.
(D) Gene annotation from contig 333 in SBCb004 Genome.
(E) Gene annotation from contig 399 in SBCb004 Genome.
(F) Gene annotation from contig 1042 in SBCb004 Genome.
(G) Gene annotation from contig 633 in SBCb004 Genome.

3.6.2 Knockout mutagenesis of the monooxygenase candidates in the SBCb004 genome

In order to study their involvements in the tubulysin pathway, these seven genes were targeted by knockout mutagenesis in SBCb004, using the method described earlier. The tubulysin production profiles of six mutants (SBCb004-*24P*$^-$, SBCb004-*45M*$^-$, SBCb004-*302P*$^-$, SBCb004-*333P*$^-$, SBCb004-*399P*$^-$, and SBCb004-*1042P*$^-$) were essentially identical with that of the SBCb004 wild type, indicating that the mutated enzymes are not involved in tubulysin assembly. However, culture extracts of a strain containing a mutated P450 gene from contig 633 (*633P1*) produced only several pretubulysin subfamily compounds pretubulysin A (**11**), *N*-desmethyl 12-keto pretubulysin A (**25**), *N*-desmethyl 12-hydroxy pretubulysin A (**26**), and *N*-desmethyl pretubulysin A (**27**) (Figure 3.1 and Table 3.3), while tubulysin variants incorporating post-PKS hydroxylation were completely absent (Figure 3.32). In addition, the production of the pretubulysins decreased dramatically. This result identified gene *633P1* as a promising candidate for one of the missing oxygenases. However, it is unclear why the yields of the pretubulysin should have been affected by its inactivation.

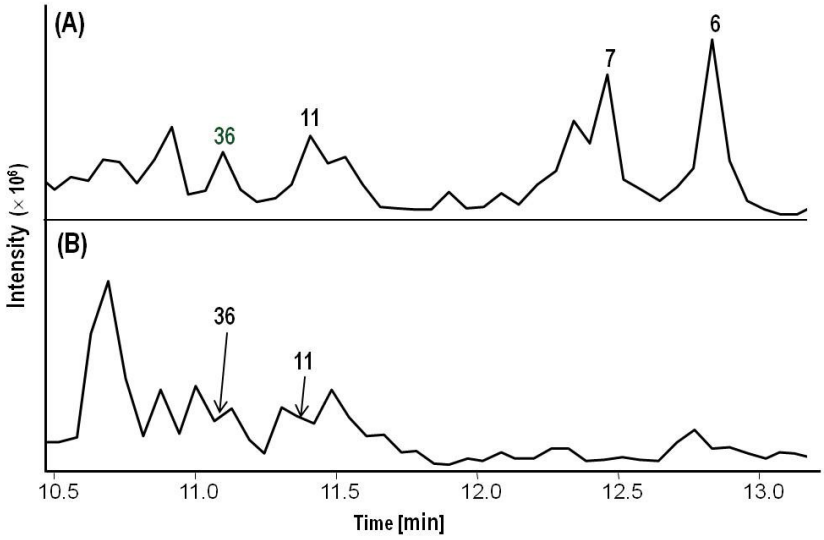


Figure 3.32 HPLC-MS analysis (BPC) of the SBCb004 wild type and mutant SBCb004-*633P1*$^{-}$.
Peaks corresponding to tubulysin A (**6**), tubulysin B (**7**), pretubulysin A (**11**) and pretubulysin A-glycerol (**36**) are indicated.
(A) SBCb004 wild type.
(B) Mutant SBCb004-*633P1*$^{-}$.

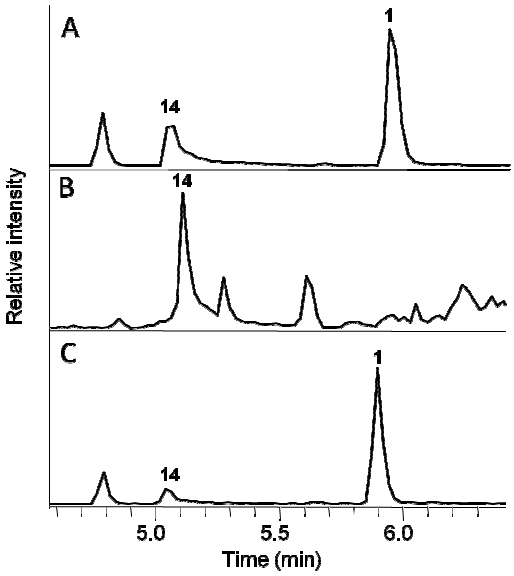


Figure 3.33 High-resolution MS analysis of *633P1* and *633P2* knockout mutant.
EIC at *m/z* 686.5 $[M+H]^+$ (pretubulysin A [11]) and *m/z* 844.4 $[M+H]^+$ (tubulysin A [6]).
(A) SBCb004 wild type.
(B) Mutant SBCb004-*633P1*$^{-}$.
(C) Mutant SBCb004-*633P2*$^{-}$.

Interestingly, gene *633P1* is located in close proximity to a second P450 gene *633P2* as well as an acyltransferase, 633AT (Figure 3.31G), identifying these additional genes as candidates for participation in tubulysin biosynthesis. However, inactivation of gene *633P2* in SBCb004 strain yielded a metabolite profile identical to that of the wild type strain, likely excluding its involvement in the pathway (Figure 3.33). Thus, the P450 gene *633P1* may carry out both of the post-assembly line oxidations in tubulysin assembly.

3.6.3 Co-expression of the tubulysin gene cluster and gene *633P1* in *P. putida*

To attempt to confirm the involvement of gene *633P1* in tubulysin biosynthesis, the gene was introduced into construct pTub-ATG as a cassette along with a zeocin resistance gene and a P_{tet} promoter, via Red/ET recombination (Experiment was did by Shiping Shan) (Figure 3.34). The resulting pTub-633P1 construct was conjugated into *P. putida*, and expression induced with tetracycline (Experiment was done by Shiping Shan). However, analysis by HPLC-MS showed that strain *P. putida*::pTub-633P1 generated pretubulysin A (**11**) and tyrosine pretubulysin A (**33**), but that no metabolites incorporating post-assembly line hydroxylation and/or acylation were present. These data were surprising given the clear results of inactivation of *633P1* in the wild type strain, but may indicate that the P450 requires additional gene products to be present in order to be active, which are currently absent from the heterologous expression construct (e.g. ferredoxins and ferredoxin reductases). In any case, further work will be required to confirm the function of 633P1 as the tubulysin monooxygenase.

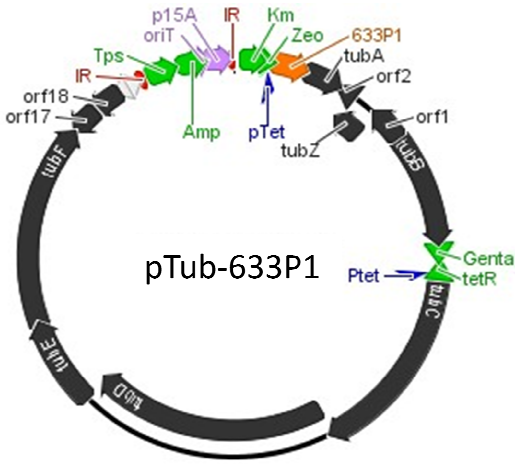


Figure 3.34 The generated heterologous expression construct pTub-633P1.

4 SUMMARY AND CONCLUSION

The present thesis deals with the identification and characterization of tubulysin biosynthetic pathways. I specially focused on the functional analysis of protein TubZ, knockout mutagenesis of *tubA*, as well as orfs 1, 17 and 18, discovery of novel tubulysin derivatives, heterologous expression of the tubulysin gene cluster in other host strains, and identification of the post-assembly line monooxygenase.

4.1 Identification of tubulysin biosynthetic gene cluster

Recently, *Cystobacter* sp. SBCb004 was found out to be a third tubulysin producer strain. In common with *Archangium gephyra* Ar315, SBCb004 generates tubulysins A, B, C, G and I, which incorporate tubutyrosine instead of the tubuphenylalanine of tubulysins D, E, F and H. Based on a cosmid library coupled with complete genome sequence of strain SBCb004, I could identify the tubulysin biosynthetic gene cluster in the SBCb004 strain. Comparison analysis of the cluster in both SBCb004 and An d48 strains reveals the anticipated hybrid NRPS-PKS core genes and six additional small genes. The gene sets show a conserved architecture, allowing the straightforward determination of the cluster boundaries in two myxobacteria strains. Unfortunately, like its An d48 counterpart, the SBCb004 tubulysin gene cluster does not contain any gene encoding post-modification tailoring function.

4.2 Starter unit supplement of the tubulysin assembly

A gene encoding a lysine cyclodeaminase (*tubZ*), identified at 5′ end of the tubulysin gene cluster, was assumed to participate in provision of the starter unit pipecolic acid in the tubulysin pathway. To directly probe the function of *tubZ*, I inactivated the gene by insertional mutagenesis in both An d48 and SBCb004 strains. The resulting mutants An d48-*tubZ*$^-$ and SBCb004-*tubZ*$^-$ were first expected to be deficient in tubulysin biosynthesis, but actually they produced tubulysin D or A at approximately 3%-5% of wild type levels instead. The continued availability of

pipecolic acid to the pathway was attributed to the presence of additional lysine cyclodeaminase function(s) in both strains.

The TubZ protein was also expressed in *E.coli* and analyzed *in vitro*. As expected, the assay shows that TubZ is able to form the starter unit pipecolic acid. In addition, it is also proved that TubZ catalyze the reaction only from L-lysine to L-pipecolic acid. Feeding studies in strain An d48 further confirmed this result that TubZ specially took the L-lysine as its catalysis substrate since only L-lysine was incorporated into the final tubulysin products.

4.3 Discovery of novel tubulysin derivatives

Prior to this study, only nine tubulysins have been described from the myxobacterial strains [40]. Here I found out the presumed first enzyme-free intermediate, pretubulysin D (in strain An d48) and pretubulysin A (in strain SBCb004), in the tubulysin pathway from *tubZ* knockout mutants. The metabolites were subsequently detected in the corresponding wild type strains, supporting their intermediacy in the biosynthesis. The "pretubulysin" is only known as a theoretical structure before my works. Now, for the first time, the intermediates were found out to be produced naturally.

The founding of pretubulysins encouraged me to detailedly investigate the productions of tubulysin natural producer strains An d48 and SBCB004 again. Finally, through the combination of high-resolution tandam mass spectrometry and feeding experiments, I was able to found out the natural biosynthesises of additional 22 tubulysin derivatives. Their assemblies all rely on the same core NRPS-PKS modules. In the patation knockout mutants, I could also identify another 8 novel derivatives, the tubulysin–glycerol ester compounds (This will be addressed in the chapter 4.4.3).

The identification of novel tubulysins in *A. disciformis* and *Cystobacter* reveals new aspects of the biosynthesis, as well as supporting earlier proposals. For example, my results strongly suggest that pretubulysin D/A are the products of the hybrid PKS-NRPS in each strain. More generally, the data show that several steps in the pathway do not go to completion (Figure 4.1). The presence in the extracts of pretubulysins, and compounds **12**, **13**, **19**, **20** and **23**, **24**, **30**, **31** missing one or more of the oxidation and acylation reactions, demonstrates that post-assembly

line modification is relatively inefficient, even in the wild type strains. This observation is consistent with the absence of obvious tailoring genes (oxidases and perhaps an acyltransferase) within the clusters, arguing for a lack of co-evolution of this portion of the pathways. The *N*-methyltransferase domain of subunit TubC also operates only imperfectly – in fact, the *N*-desmethyl compounds (**14**–**18**) and (**25**–**29**) constitute major metabolites of the strain. Apparently, the lack of *N*-methylation somehow induces skipping of some, if not all of the reductive steps carried out by the adjacent PKS module. However, these activities do occasionally function, as the corresponding hydroxyl (**15**, **26**) and methylene (**16**, **27**) analogues are produced at low levels. Processing by post-assembly line enzymes at C11 can also occur, and is more efficient when full reduction has taken place at the adjacent C12 position. This mechanism is inconsistent, however, with the discovery in extracts of *N*-hydroxymethyl 12-keto pretubulysins (**21**, **32**), as the reductive steps have been bypassed even though the *N*-methyl group was added. One possible explanation is that, in a small number of cases, the *N*-methyl is hydroxylated while the intermediate is attached to the assembly line, and this modification also disrupts reductive processing by the PKS module. Finally, the truncated phenylalanine analogues (**22**, **33**) appear to arise from failure of the final chain extension step with malonate, catalyzed by TubF. At this stage it is not clear whether premature release from the assembly line is spontaneous due to stalling of the intermediate at the TubE/TubF interface, or alternatively enzyme-catalyzed, for example by the TE domain, as demonstrated in pikromycin biosynthesis [144].

It is intriguing to consider whether the apparent 'failure' of various enzymes in the tubulysin pathway is in fact favored by evolution to increase the diversity of structures generated by the strain, a proposal known as the 'screening hypothesis' [145]. Indeed, the production of sometimes large families of metabolites which differ in their core structures appears to be a general feature of myxobacterial secondary metabolism, in which post-assembly line processing is relatively minimal [31]. However, relevant information on the biological roles of the various tubulysins for the producing organisms is lacking at present. Clearly, though, the PKS-NRPS can process a range of structurally-divergent intermediates, encouraging the view that the assembly line could be tailored by genetic engineering to produce further variants. Similar observations have been made for the epothilone assembly line, from which multiple epothilone derivatives have been generated by directed modification. In the meantime, the novel analogues can be targeted for total synthesis, in order to obtain sufficient materials to extend the ongoing structure-activity studies of the tubulysin metabolites.

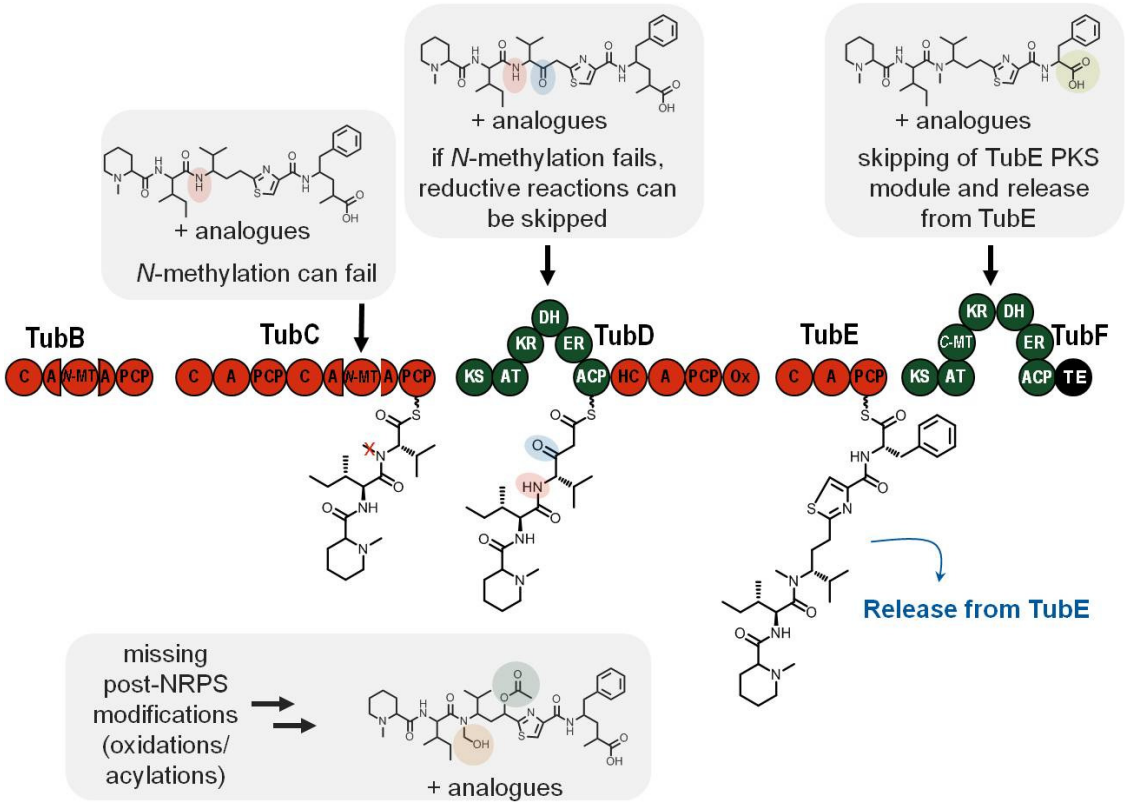


Figure 4.1 Causes for the biosynthetic diversity in tubulysin assembly.

4.4 Inactivation mutagenesis of tubulysin pathway

In order to investigate the roles of the accessory genes in the tubulysin biosynthesis, I carried out the insertional inactivation mutagenesis in both wild type strains An d 48 and SBCb004.

4.4.1 TubA – the possible acyl transferase in tubulysin biosynthesis

The inactivation of gene tubA in two natural producer strains An d48 and SBCb004 resulted in none-acyl tubulysin productions. Further proof was also draw from a *tubA* knockout in the heterologous expression system in strain *M. xanthus* DK1622. Such fact strongly suggests that TubA may serve as at least one of the post-modification acyltransferase in the tubulysin biosynthesis.

4.4.2 *orf1* and *orf2* may be involved in tubulysin biosynthesis

Inactivation of the gene *orf1* or *orf2* (*orf2* only succeeded in the heterologous expression system.) resulted in the accumulation of the pretublysin, suggesting their effects on the extent of post-assembly line modification. The results could not prove their functions clearly, but the pretubulysin subfamily compounds production in the absence of *orf1* and *orf2* indicates that they still contribute to pretubulysin biosynthesis in some manner even though they are not essential for assembling the tubulysin core structure.

4.4.3 Function of patatin genes *orf17* and *orf18*

Inactivation of the patatin gene *orf18* in both An d48 and SBCb004 strains opened a new window for the tubulysin biosynthesis. The novel tubulysin–glycerol or pretubulysin–glycerol compounds in the *orf18* inactivation mutants clearly indicates the function of Orf18: it cuts off the bound glycerol from biosynthesized tubulysin–glycerol and results in the general matural tubulysins, although currently we have no idea how the glycerol is attached on. Gene *orf17* shows high similarity to *orf18* and is supposed to have the same function. However, inactivation of *orf17* in strain SBCb004 showed a different production profile: instead of glycerol ester compounds boosting, the productions of all tubulysins decreased. The explanation might be the *orf17* knockout affects the normal transcription of its overlapped neighbor gene *tubF*, resulting in the decline of tubulysin productions.

4.5 Heterologous expression of tubulysin pathway

I have applied Red/ET recombineering to reconstitute the 11 gene tubulysin cluster from two cosmids. Addition of a transposon was carried out to facilitate the transformation in other host strains. The stitched cluster was then transferred into two heterologous hosts, *Pseudomonas putida* and the myxobacterium *Myxococcus xanthus*. They majorly produce pretubulysin A and tyrosine pretubulysin A, the predicted products of the gene cluster. The yields of tubulysins in both heterologous expression strains are lower than those of the natural producer strains even they were supplied with sufficient starter unit pipecolic acid.

The established heterologous expression system provides a good platform that can be used to manipulate the tubulysin gene cluster in *E.coli* strain, which makes the investigation of tubulysin pathway much easier than before. Generating the equivalent mutations (such as *tubA*, *orf1*, *tubZ* and *orf2*) in the heterologous expression system in host strain *M. xanthus* DK1622 was a good

application, particularly for the small gene *orf2*. In the natural producer myxobacterial strains, it was difficult to knock out such a small gene. But in the heterologous expression system, the gene engineering only happens in *E. coli* strain, making it much easier.

4.6 Searching the missing monooxygenase in the SBCb004 genome

Although the identified tubulysin cluster gives rise to bioactive pathway intermediates, it lacks the post-modification oxidase and/or acyl transferase functions responsible for generating mature tubulysins. By BLAST analysis of the SBCb004 genome, multiple candidate monooxygenases were identified by annotation of their flanking regions. The involvements of selected seven monooxygenases in tubulysin assembly were investigated using knockout mutagenesis.

A mutant of a P450 gene from contig 633 (*633P1*) was found out to produce only the pretubulysin subfamily compounds **14**, **18**, **20**, and **28** at dramatically decreased amount, and the tubulysin variants incorporating post-PKS hydroxylation were completely absent. This result makes gene *633P1* the most promising candidate for one of the missing oxygenases, even though it is still unclear why the pretubulysins' yield is also affected by *633P1* inactivation.

To attempt to confirm the involvement of gene *633P1* in tubulysin biosynthesis, the heterologeous expression system was employed too. *633P1* was introduced into the heterologous expression construct and co-expressed with the tubulysin gene cluster in *P. putida*. Unfortunately, the tubulysin production profile with *633P1* was the same as before. It is possible that the P450 enzyme requires additional gene products to be present in order to be active during the tubulysin assembly, which are currently absent from the heterologous expression construct (e.g. ferredoxins and ferredoxin reductases).

5 REFERENCES

1. Newman D.J., Cragg G.M., Snader K.M., Natural products as sources of new drugs over the period 1981-2002, J.Nat.Prod. 66 (2003) 1022-1037.
2. Newman D.J., Cragg G.M., Natural products as sources of new drugs over the last 25 years, J.Nat.Prod. 70 (2007) 461-477.
3. Norn S., Kruse P.R., Kruse E., History of opium poppy and morphine, Dan.Medicinhist.Arbog. 33 (2005) 171-184.
4. Lerner P.I., Producing penicillin, N Engl J Med 351 (2004) 524.
5. Thompson C.J., Fink D., Nguyen L.D., Principles of microbial alchemy: insights from the Streptomyces coelicolor genome sequence, Genome Biol. 3 (2002) Reviews 1020.
6. Harada K., Production of secondary metabolites by freshwater cyanobacteria, Chem Pharm Bull (Tokyo) 52 (2004) 889-899.
7. Behal V., Alternative sources of biologically active substances, Folia Microbiol. 48 (2003) 563-571.
8. Haefner B., Drugs from the deep: marine natural products as drug candidates, Drug Discov Today 8 (2003) 536-544.
9. Reichenbach H., Höfle G., Myxobacteria as producers of secondary metabolites, in: Grabley S., Thiericke R. (Eds.), Drug Discovery from Nature, Springer, Berlin, 1999, pp. 149-179.
10. Demain A.L., From natural products discovery to commercialization: a success story, J.Ind.Microbiol.Biotechnol. 33 (2006) 486-495.
11. Washington J.A., Wilson W.R., Erythromycin: a microbial and clinical perspective after 30 years of clinical use, Mayo Clin.Proc. 60 (1985) 189-203.
12. 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 68 (2004) 1345-1352.
13. Walczak R.J., Dickens M.L., Priestley N.D., Strohl W.R., Purification, properties, and characterization of recombinant Streptomyces sp. strain C5 DoxA, a cytochrome P-450 catalyzing multiple steps in doxorubicin biosynthesis, J.Bacteriol. 181 (1999) 298-304.
14. Fischbach M.A., Walsh C.T., Antibiotics for emerging pathogens, Science 325 (2009) 1089-1093.
15. Gulder T.A.M., Moore B.S., Chasing the treasures of the sea - bacterial marine natural products, Curr.Opin.Microbiol. 12 (2009) 252-260.
16. Garcia R.O., Krug D., Müller R., Discovering natural products from myxobacteria with emphasis on rare producer strains in combination with improved analytical methods, Methods Enzymol. 458 (2009) 59-91.
17. Garcia R.O., Reichenbach H., Ring M.W., Müller R., Phaselicystis flava gen. nov., sp. nov., an arachidonic acid-containing soil myxobacterium, and the description of Phaselicystidaceae fam. nov., International Journal of Systematic and Evolutionary Microbiology 59 (2009) 1524-1530.

18. Reichenbach H., Order VIII. Myxococcales. Tchan, Pochon and Pre'vot 1948, 398AL, in: Brenner D.J., Krieg N.R., Staley J.T. (Eds.), Bergey's manual of systematic bacteriology, Springer, 2005, pp. 1059-1144.
19. Ravenschlag K., Sahm K., Pernthaler J., Amann R., High bacterial diversity in permanently cold marine sediments, Appl.Environ.Microbiol. 65 (1999) 3982-3989.
20. Reichenbach H., The ecology of the myxobacteria, Environ.Microbiol. 1 (1999) 15-21.
21. Dawid W., Occurrence and distribution of fruiting body-forming myxobacteria in Siebengebirge. Comparative studies with special reference to characteristic biotypes, Z Allg Mikrobiol 19 (1979) 705-719.
22. Shimkets L.J., Social and developmental biology of the myxobacteria, Microbiol.Rev. 54 (1990) 473-501.
23. Reichenbach H., The myxobacteria: common organisms with uncommon behaviour, Microbiol Sci 3 (1986) 268-274.
24. Kaiser D., Coupling cell movement to multicellular development in myxobacteria, Nat.Rev.Microbiol. 1 (2003) 45-54.
25. Reichenbach H., Myxobacteria, producers of novel bioactive substances, J.Ind.Microbiol.Biotechnol. 27 (2001) 149-156.
26. Schneiker S., Perlova O., Kaiser O., Gerth K., Alici A., Altmeyer M.O., et. al., Complete genome sequence of the myxobacterium Sorangium cellulosum, Nat.Biotechnol. 25 (2007) 1281-1289.
27. Pradella S., Hans A., Sproer C., Reichenbach H., Gerth K., Beyer S., Characterisation, genome size and genetic manipulation of the myxobacterium Sorangium cellulosum So ce56, Arch.Microbiol. 178 (2002) 484-492.
28. Chen H., Keseler I.M., Shimkets L.J., Genome size of Myxococcus xanthus determined by pulsed-field gel electrophoresis, J.Bacteriol. 172 (1990) 4206-4213.
29. Wenzel S.C., Müller R., Myxobacterial natural product assembly lines: fascinating examples of curious biochemistry, Nat.Prod.Rep. 24 (2007) 1211-1224.
30. Krug D., Zurek G., Revermann O., Vos M., Velicer G.J., Müller R., Discovering the hidden secondary metabolome of Myxococcus xanthus: a study of intraspecific diversity, Appl.Environ.Microbiol. 74 (2008) 3058-3068.
31. Weissman K.J., Müller R., A brief tour of myxobacterial secondary metabolism, Bioorg.Med.Chem. 17 (2009) 2121-2136.
32. Bode H.B., Müller R., Analysis of myxobacterial secondary metabolism goes molecular, J.Ind.Microbiol.Biotechnol. 33 (2006) 577-588.
33. Nickeleit I., Zender S., Sasse F., Geffers R., Brandes G., Sörensen I., Steinmetz H., Kubicka S., Carlomagno T., Menche D., Gütgemann I., Buer J., Gossler A., Manns M.P., Kalesse M., Frank R., Malek N.P., Argyrin A reveals a critical role for the tumor suppressor protein p27kip1 in mediat$_{ing}$ antitumor activities in response to proteasome inhibition, Cancer Res. 14 (2008) 23-35.
34. Thierbach G., Kunze B., Reichenbach H., Höfle G., The mode of action of stigmatellin, a new inhibitor of the cytochorme b-c1 segment of the respiratory chain, Biochim.Biophys.Acta 765 (1984) 227-235.
35. Vahlensieck H.F., Pridzun L., Reichenbach H., Hinnen A., Identification of the yeast ACC1 gene product (acetyl-CoA carboxylase) as the target of the polyketide fungicide soraphen A, Curr.Genet. 25 (1994) 95-100.

36. Goodin S., Kane M.P., Rubin E.H., Epothilones: Mechanism of action and biologic activity, J Clin Oncol 22 (2004) 2015-2025.
37. Khalil M.W., Sasse F., Lünsdorf H., Elnakady Y.A., Reichenbach H., Mechanism of action of tubulysin, an antimitotic peptide from myxobacteria, ChemBioChem 7 (2006) 678-683.
38. Elnakady Y.A., Sasse F., Lünsdorf H., Reichenbach H., Disorazol A1, a highly effective antimitotic agent acting on tubulin polymerization and inducing apoptosis in mammalian cells, Biochem.Pharmacol. 67 (2004) 927-935.
39. Bollag D.M., McQueney P.A., Zhu J., Hensens O., Koupal L., Liesch J., Goetz M., Lazarides E., Woods M., Epothilones, a new class of microtubule-stabilizing agents with Taxol-like mechanism of action, Cancer Res. 55 (1995) 2325-2333.
40. Sasse F., Steinmetz H., Heil J., Höfle G., Reichenbach H., Tubulysins, new cytostatic peptides from myxobacteria acting on microtubuli: Production, isolation, physico-chemical and biological properties, J.Antibiot. 53 (2000) 879-885.
41. Hagelueken G., Albrecht S.C., Steinmetz H., Jansen R., Heinz D.W., Kalesse M., Schubert W.D., The absolute configuration of rhizopodin and its inhibition of actin polymerization by dimerization, Angew Chem Int Ed Engl 48 (2009) 595-598.
42. Khalil M.W.M., Tubulysin aus Mmyxobakterien: Untersuchungen zum Wirkungsmechanismus, Dissertationsschrift: TU Carolo-Wilhelmina, Braunschweig, 1999, pp. 122.
43. Staunton J., Weissman K.J., Polyketide biosynthesis: a millennium review, Nat.Prod.Rep. 18 (2001) 380-416.
44. Hopwood D.A., Genetic contributions to understanding polyketide synthases, Chem Rev 97 (1997) 2465-2497.
45. Lambalot R.H., Gehring A.M., Flugel R.S., Zuber P., LaCelle M., Marahiel M.A., Reid R., Khosla C., Walsh C.T., A new enzyme superfamily - the phosphopantetheinyl transferases, Chem.Biol. 3 (1996) 923-936.
46. 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, in: Anonymous, 2002, pp. 13082-13090.
47. Meiser P., Weissman K.J., Bode H.B., Krug D., Dickschat J.S., Sandmann A., Müller R., DKxanthene biosynthesis -- understanding the basis for diversity-oriented synthesis in myxobacterial secondary metabolism, Chem.Biol. 15 (2008) 771-781.
48. Hendrickson L., Davis C.R., Roach C., Nguyen D.K., Aldrich T., McAda P.C., Reeves C.D., Lovastatin biosynthesis in Aspergillus terreus: characterization of blocked mutants, enzyme activities and a multifunctional polyketide synthase gene, Chem.Biol. 6 (1999) 429-439.
49. Abe I., Morita H., Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases, Nat.Prod.Rep. 27 (2010) 809-838.
50. Austin M.B., Noel J.P., The chalcone synthase superfamily of type III polyketide synthases, Nat.Prod.Rep. 20 (2003) 79-110.
51. Schröder J., A family of plant-specific polyketide synthases: facts and predictions, Trends Plant Sci. 2 (1997) 373-378.
52. Cane D.E., Walsh C.T., The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases, Chem.Biol. 6 (1999) 319-325.

53. Conti E., Stachelhaus T., Marahiel M.A., Brick P., Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S, EMBO J. 16 (1997) 4174-4183.
54. Challis G.L., Ravel J., Townsend C.A., Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains, Chem.Biol. 7 (2000) 211-224.
55. Stachelhaus T., Mootz H.D., Bergendahl V., Marahiel M.A., Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain, J.Biol.Chem. 273 (1998) 22773-22781.
56. Finking R., Marahiel M.A., Biosynthesis of nonribosomal peptides, Annual Review of Microbiology 58 (2004) 453-488.
57. Sieber S.A., Marahiel M.A., Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics, Chem Rev 105 (2005) 715-738.
58. Marahiel M.A., Stachelhaus T., Mootz H.D., Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis, Chem Rev 97 (1997) 2651-2674.
59. Schwarzer D., Finking R., Marahiel M.A., Nonribosomal peptides: From genes to products, Nat.Prod.Rep. 20 (2003) 275-287.
60. Samel S.A., Marahiel M.A., Essen L.O., How to tailor non-ribosomal peptide products - new clues about the structures and mechanisms of modifying enzymes, Molecular Biosystems 4 (2008) 387-393.
61. Walsh C.T., Chen H.W., Keating T.A., Hubbard B.K., Losey H.C., Luo L.S., Marshall C.G., Miller D.A., Patel H.M., Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines, Curr.Opin.Chem.Biol. 5 (2001) 525-534.
62. Rix U., Fischer C., Remsing L.L., Rohr J., Modification of post-PKS tailoring steps through combinatorial biosynthesis, Nat.Prod.Rep. 19 (2002) 542-580.
63. van Wageningen A.M.A., Kirkpatrick P.N., Williams D.H., Harris B.R., Kershaw J.K., Lennard N.J., Jones M., Jones S.J.M., Solenberg P.J., Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic, Chem.Biol. 5 (1998) 155-162.
64. Staunton J., Wilkinson B., Biosynthesis of erythromycin and rapamycin, Chem Rev 97 (1997) 2611-2630.
65. Steffensky M., Muhlenweg A., Wang Z.X., Li S.M., Heide L., Identification of the novobiocin biosynthetic gene cluster of Streptomyces spheroides NCIB 11891, Antimicrob.Agents Chemother. 44 (2000) 1214-1222.
66. Irschik H., Jansen R., Gerth K., Höfle G., Reichenbach H., The sorangicins, novel and powerful inhibitors of eubacterial RNA polymerase isolated from myxobacteria, J.Antibiot. 40 (1987) 7-13.
67. Jansen R., Irschik H., Reichenbach H., Höfle G., Antibiotics from gliding bacteria, LXXX. Chivosazoles A-F: Novel antifungal and cytotoxic macrolides from Sorangium cellulosum (Myxobacteria), Liebigs Ann Chem (1997) 1725-1732.
68. 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. 14 (2006) 667-681.
69. Buedenbender S., Rachid S., Müller R., Schulz G.E., Structure and action of the myxobacterial chondrochloren halogenase CndH: a new variant of FAD-dependent halogenases, J.Mol.Biol. 385 (2009) 520-530.

70. Wenzel S.C., Müller R., The biosynthetic potential of myxobacteria and their impact on drug discovery, Curr Opin Drug Discov Devel 12 (2009) 220-230.
71. Wenzel S.C., Gross F., Zhang Y., Fu J., Stewart F.A., Müller R., Heterologous expression of a myxobacterial natural products assembly line in pseudomonads via Red/ET recombineering, Chem.Biol. 12 (2005) 349-356.
72. 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. 185 (2006) 28-38.
73. Schmidt E.W., Nelson J.T., Rasko D.A., Sudek S., Eisen J.A., Haygood M.G., Ravel J., Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella, P.Natl.Acad.Sci.USA 102 (2005) 7315-7320.
74. Kodumal S.J., Patel K.G., Reid R., Menzella H.G., Welch M., Santi D.V., Total synthesis of long DNA sequences: Synthesis of a contiguous 32-kb polyketide synthase gene cluster, P.Natl.Acad.Sci.USA 101 (2004) 15573-15578.
75. Wenzel S.C., Müller R., Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways, Curr.Opin.Biotechnol. 16 (2005) 594-606.
76. Zhang Y., Muyrers J.P.P., Testa G., Stewart A.F., DNA cloning by homologous recombination in Escherichia coli, Nat.Biotechnol. 18 (2000) 1314-1317.
77. Zhang Y., Buchholz F., Muyrers J.P., Stewart F.A., A new logic for DNA engineering using recombination in Escherichia coli, Nat.Genet. 20 (1998) 123-128.
78. Muyrers J.P.P., Zhang Y., Benes V., Testa G., Ansorge W., Stewart A.F., Point mutation of bacterial artificial chromosomes by ET recombination, EMBO Rep. 1 (2000) 239-243.
79. Wenzel S.C., Müller R., Heterologe Expression von komplexen myxobakteriellen Naturstoff-Biosynthesewegen in Pseudomonaden, Biospektrum 11 (2005) 628-631.
80. Takahashi M., Altschmied L., Hillen W., Kinetic and equilibrium characterization of the Tet repressor-tetracycline complex by fluorescence measurements. Evidence for divalent metal ion requirement and energy transfer, J.Mol.Biol. 187 (1986) 341-348.
81. Hillen W., Berens C., Mechanisms underlying expression of Tn10 encoded tetracycline resistance, Annual Review of Microbiology 48 (1994) 345-369.
82. Berens C., Hillen W., Gene regulation by tetracyclines. Constraints of resistance regulation in bacteria shape TetR for application in eukaryotes, Eur.J.Biochem. 270 (2003) 3109-3121.
83. Ichinose 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 149 (2003) 1633-1645.
84. Kim C.Y., Park H.J., Kim E.S., Heterologous expression of hybrid type II polyketide synthase system in Streptomyces species, J.Microbiol.Biotechnol. 13 (2003) 819-822.
85. Li A., Piel J., A gene cluster from a marine Streptomyces encoding the biosynthesis of the aromatic spiroketal polyketide griseorhodin A, Chem.Biol. 9 (2002) 1017-1026.
86. Gould S.J., Hong S.T., Carney J.R., Cloning and heterologous expression of genes from the kinamycin biosynthetic pathway of Streptomyces murayamaensis, J.Antibiot. 51 (1998) 52-57.

87. 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. 230 (2004) 91-97.
88. Thomson N.R., Crow M.A., McGowan S.J., Cox A., Salmond G.P., Biosynthesis of carbapenem antibiotic and prodigiosin pigment in Serratia is under quorum sensing control, Mol.Microbiol. 36 (2000) 539-556.
89. Miao V., Coeffet-LeGal M.F., Brian P., Brost R., Penn J., Whiting A., Martin S., Ford R., Parr I., Bouchard M., Silva C.J., Wrigley S.K., Baltz R.H., Daptomycin biosynthesis in Streptomyces roseosporus: Cloning and analysis of the gene cluster and revision of peptide stereochemistry, Microbiology 151 (2005) 1507-1523.
90. Binz T.M., Wenzel S.C., Schnell H.J., Bechthold A., Müller R., Heterologous expression and genetic engineering of the phenalinolactone biosynthetic gene cluster by using Red/ET recombineering, ChemBioChem 9 (2008) 447-454.
91. Pfeifer B.A., Admiraal S.J., Gramajo H., Cane D.E., Khosla C., Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli, Science 291 (2001) 1790-1792.
92. Pfeifer B.A., Wang C.C.C., Walsh C.T., Khosla C., Biosynthesis of yersiniabactin, a complex polyketide-nonribosomal peptide, using Escherichia coli as a heterologous host, Appl.Environ.Microbiol. 69 (2003) 6698-6702.
93. Gross F., Ring M.W., Perlova O., Fu J., Schneider S., Gerth K., Kuhlmann S., Stewart A.F., Zhang Y., Müller R., Metabolic engineering of Pseudomonas putida for methylmalonyl-CoA biosynthesis to enable complex heterologous secondary metabolite formation, Chem.Biol. 13 (2006) 1253-1264.
94. Perlova O., Fu J., Kuhlmann S., Krug D., Stewart F., Zhang Y., Müller R., Reconstitution of myxothiazol biosynthetic gene cluster by Red/ET recombination and heterologous expression in Myxococcus xanthus, Appl.Environ.Microbiol. 72 (2006) 7485-7494.
95. Tang L., Shah S., Chung L., Carney J., Katz L., Khosla C., Julien B., Cloning and heterologous expression of the epothilone gene cluster, Science 287 (2000) 640-642.
96. Julien B., Shah S., Heterologous expression of epothilone biosynthetic genes in Myxococcus xanthus, Antimicrob.Agents Chemother. 46 (2002) 2772-2778.
97. Mutka S.C., Carney J.R., Liu Y., Kennedy J., Heterologous production of epothilone C and D in Escherichia coli, Biochemistry-US 45 (2006) 1321-1330.
98. Fu J., Wenzel S.C., Perlova O., Wang J., Gross F., Tang Z., Yin Y., Stewart A.F., Müller R., Zhang Y., Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition, Nucleic Acids Res. 36 (2008) e113.
99. Dömling A., Richter W., Myxobacterial epothilones and tubulysins as promising anticancer agents, Mol.Divers. 9 (2005) 141-147.
100. Steinmetz H., Glaser N., Herdtweck E., Sasse F., Reichenbach H., Höfle G., Isolation, crystal and solution structure determination, and biosynthesis of tubulysins-powerful inhibitors of tubulin polymerization from myxobacteria, Angew.Chem.Int.Ed. 43 (2004) 4888-4892.
101. Kaur G., Hollingshead M., Holbeck S., Schauer-Vukasinovic V., Camalier R.F., Domling A., Agarwal S., Biological evaluation of tubulysin A: a potential anticancer and antiangiogenic natural product, Biochem.J. 396 (2006) 235-242.

102. 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. 11 (2004) 1071-1079.

103. Balasubramanian R., Raghavan B., Steele J.C., Sackett D.L., Fecik R.A., Tubulysin analogs incorporating desmethyl and dimethyl tubuphenylalanine derivatives, Bioorg Med Chem Lett Bioorg Med Chem Lett 18 (2008) 2996-2999.

104. Raghavan B., Balasubramanian R., Steele J.C., Sackett D.L., Fecik R.A., Cytotoxic simplified tubulysin analogues, J.Med.Chem. 51 (2008) 1530-1533.

105. Patterson A.W., Peltier H.M., Ellman J.A., Expedient synthesis of N-methyl tubulysin analogues with high cytotoxicity, J.Org.Chem. 73 (2008) 4362-4369.

106. Patterson A.W., Peltier H.M., Sasse F., Ellman J.A., Design, synthesis, and biological properties of highly potent tubulysin D analogues, Chemistry 13 (2007) 9534-9541.

107. Wang Z.Y., McPherson P.A., Raccor B.S., Balachandran R., Zhu G.Y., Day B.W., Vogt A., Wipf P., Structure-activity and high-content imaging analyses of novel tubulysins, Chemical Biology & Drug Design 70 (2007) 75-86.

108. Dömling A., Beck B., Eichelberger U., Sakamuri S., Menon S., Chen Q.Z., Lu Y., Wessjohann L.A., Total synthesis of tubulysin U and V, Angew.Chem.Int.Ed. 45 (2006) 7235-7239.

109. Ullrich A., Chai Y., Pistorius D., Elnakady Y.A., Herrmann J.E., Weissman K.J., Kazmaier U., Müller R., Pretubulysin, a potent and chemically-accessible tubulysin precursor from Angiococcus disciformis, Angew.Chem Int.Ed Engl. 48 (2009) 4422-4425.

110. Hodgkin J., Kaiser D., Cell-to-cell stimulation of movement in non-motile mutant of Myxococcus, P.Natl.Acad.Sci.USA 74 (1977) 2932-2942.

111. Chen Y.H., Wang C.C., Greenwell L., Rix U., Hoffmeister D., Vining L.C., Rohr J.R., Yang K.Q., Functional analyses of oxygenases in jadomycin biosynthesis and identification of JadH as a bifunctional oxygenase/dehydrase, J.Biol.Chem. 280 (2005) 22508-22514.

112. Hill D.S., Stein J.I., Torkewitz N.R., Morse A.M., Howell C.R., Pachlatko J.P., Becker J.O., Ligon J.M., 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. 60 (1994) 78-85.

113. Margulies M., Egholm M., Altman W.E., Attiya S., Bader J.S., Bemben L.A., Berka J., et. al., Genome sequencing in microfabricated high-density picolitre reactors, Nature 437 (2005) 376-380.

114. Rothberg J.M., Leamon J.H., The development and impact of 454 sequencing, Nat.Biotechnol. 26 (2008) 1117-1124.

115. Altschul S.F., Madden T.L., Schaffer A.A., Zhang J.H., Zhang Z., Miller W., Lipman D.J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389-3402.

116. De Crecy-Lagard V., Marliere P., Saurin W., Multienzymatic non ribosomal peptide biosynthesis: identification of the functional domains catalysing peptide elongation and epimerisation, C.R.l'Academie Sci., Ser.III 318 (1995) 927-936.

117. Moraleda-Munoz A., Shimkets L.J., Lipolytic enzymes in Myxococcus xanthus, J.Bacteriol. 189 (2007) 3072-3080.

118. Banerji S., Flieger A., Patatin-like proteins: a new family of lipolytic enzymes present in bacteria?, Microbiology 150 (2004) 522-525.

119. Carvalho R., Reid R., Viswanathan N., Gramajo H., Julien B., The biosynthetic genes for disorazoles, potent cytotoxic compounds that disrupt microtubule formation, Gene 359 (2005) 91-98.
120. Kopp M., Irschik H., Pradella S., Müller R., Production of the tubulin destabilizer disorazol in Sorangium cellulosum: biosynthetic machinery and regulatory genes, ChemBioChem 6 (2005) 1277-1286.
121. Perlova O., Gerth K., Hans A., Kaiser O., Müller R., Identification and analysis of the chivosazol biosynthetic gene cluster from the myxobacterial model strain Sorangium cellulosum So ce56, J.Biotechnol. 121 (2006) 174-191.
122. Khaw L.E., Bohm G.A., Metcalfe S., Staunton J., Leadlay P.F., Mutational biosynthesis of novel rapamycins by a strain of Streptomyces hygroscopicus NRRL 5491 disrupted in rapL, encoding a putative lysine cyclodeaminase, J.Bacteriol. 180 (1998) 809-814.
123. Gatto G.J., Boyne M.T., Kelleher N.L., Walsh C.T., Biosynthesis of pipecolic acid by RapL, a lysine cyclodeaminase encoded in the rapamycin gene cluster, J.Am.Chem.Soc. 128 (2006) 3838-3847.
124. Peltier H.M., McMahon J.P., Patterson A.W., Ellman J.A., The total synthesis of tubulysin D, J Am Chem Soc 128 (2006) 16018-16019.
125. Neri D., Fossati G., Zanda M., Efforts toward the total synthesis of tubulysins: new hopes for a more effective targeted drug delivery to tumors, ChemMedChem. 1 (2006) 175-180.
126. Ullrich A., Herrmann J., Müller R., Kazmaier U., Synthesis and Biological Evaluation of Pretubulysin and Derivatives, Eur.J.Org.Chem. 2009 (2009) 6367-6378.
127. Chai Y., Pistorius D., Ullrich A., Weissman K.J., Kazmaier U., Müller R., Discovery of 23 natural tubulysins from Angiococcus disciformis An d48 and Cystobacter SBCb004, Chem.Biol. 17 (2010) 296-309.
128. 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. 51 (2007) 1028-1037.
129. Luo L., Kohli R.M., Onishi M., Linne U., Marahiel M.A., Walsh C.T., Timing of epimerization and condensation reactions in nonribosomal peptide assembly lines: kinetic analysis of phenylalanine activating elongation modules of tyrocidine synthetase B, Biochemistry-US 41 (2002) 9184-9196.
130. Stachelhaus T., Walsh C.T., Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase, Biochemistry-US 39 (2000) 5775-5787.
131. Rausch C., Hoof I., Weber T., Wohlleben W., Huson D.H., Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution, BMC Evolutionary Biology 7 (2007) 78-92.
132. Balibar C.J., Vaillancourt F.H., Walsh C.T., Generation of D-amino Acid residues in assembly of arthrofactin by dual condensation/epimerization domains, Chem.Biol. 12 (2005) 1189-1200.
133. Dubern J.F., Coppoolse E.R., Stiekema W.J., Bloemberg G.V., Genetic and functional characterization of the gene cluster directing the biosynthesis of putisolvin I and II in Pseudomonas putida strain PCL1445, Microbiology 154 (2008) 2070-2083.

134. Linne U., Doekel S., Marahiel M.A., Portability of epimerization domain and role of peptidyl carrier protein on epimerization activity in nonribosomal peptide synthetases, Biochemistry-US 40 (2001) 15824-15834.
135. Balasubramanian R., Raghavan B., Begaye A., Sackett D.L., Fecik R.A., Total synthesis and biological evaluation of tubulysin U, tubulysin V, and their analogues, J.Med.Chem 52 (2009) 238-240.
136. Sani M., Fossati G., Huguenot F., Zanda M., Total synthesis of tubulysins U and V, Angew.Chem Int.Ed Engl. 46 (2007) 3526-3529.
137. Richter C.D., Nietlispach D., Broadhurst R.W., Weissman K.J., Multienzyme docking in hybrid megasynthetases, Nat Chem Biol Nat Chem Biol 4 (2007) 75-81.
138. Meiser P., Müller R., Two functionally redundant Sfp-type 4'-phosphopantetheinyl transferases differentially activate biosynthetic pathways in Myxococcus xanthus, ChemBioChem 9 (2008) 1549-1553.
139. Buntin K., Irschik H., Weissman K.J., Luxenburger E., Blöcker H., Müller R., Biosynthesis of thuggacins in myxobacteria: comparative cluster analysis reveals basis for natural product structural diversity, Chem.Biol. 17 (2010) 342-356.
140. Kunze B., Kemmer T., Höfle G., Reichenbach H., Stigmatellin, a new antibiotic from Stigmatella aurantiaca (Myxobacterales). I. Production, physico-chemical and biological properties, J.Antibiot. 37 (1984) 454-461.
141. Vervoort E.B., van R.A., van Peij N.N., Heikoop J.C., van Haastert P.J., Verheijden G.F., Linskens M.H., Optimizing heterologous expression in dictyostelium: importance of 5' codon adaptation, Nucleic Acids Res. 28 (2000) 2069-2074.
142. Kuhstoss S., Richardson M.A., Rao R.N., Plasmid cloning vectors that integrate site-specifically in Streptomyces spp, Gene 97 (1991) 143-146.
143. Buntin K., Rachid S., Scharfe M., Blöcker H., Weissman K.J., Müller R., Production of the antifungal isochromanone ajudazols A and B in Chondromyces crocatus Cm c5: biosynthetic machinery and cytochrome P450 modifications, Angew.Chem.Int.Ed. 47 (2008) 4595-4599.
144. Kittendorf J.D., Beck B.J., Buchholz T.J., Seufert W., Sherman D.H., Interrogating the molecular basis for multiple macrolactone by the pikromycin polyketide ring formation synthase, Chemistry & Biology 14 (2007) 944-954.
145. Firn R.D., Jones C.G., Natural products--a simple model to explain chemical diversity, Nat.Prod.Rep. 20 (2003) 382-391.