Deciphering novel mechanisms of bacterial secondary metabolite biosynthetic pathways

Dissertation
zur Erlangung des Grades
des Doktors der Naturwissenschaften
der Naturwissenschaftlich-Technischen Fakultät III
Chemie, Pharmazie, Bio- und Werkstoffwissenschaften
der Universität des Saarlandes

von
Dominik Pistorius
Saarbrücken
2011

Tag des Kolloquiums: 28. Oktober 2011

Dekan:	Prof. Dr. Wilhelm F. Maier
Berichterstatter:	Prof. Dr. Rolf Müller
	Prof. Dr. Rolf W. Hartmann
	Prof. Dr. Andreas Bechthold
Vorsitz:	Prof. Dr. Alexandra Kiemer
Akad. Mitarbeiter:	Dr. Angelika Ullrich

Diese Arbeit entstand unter der Anleitung von Prof. Dr. Rolf Müller in der Fachrichtung 8.2 Pharmazeutische Biotechnologie der Naturwissenschaftlich-Technischen Fakultät III der Universität des Saarlandes von Oktober 2007 bis August 2011.

Kein Baum fällt auf den ersten Streich.
(Sprichwort)

Danksagung

Mein besonderer Dank gilt Prof. Dr. Rolf Müller für die Überlassung der interessanten und herausfordernden Themen, die Bereitstellung der außergewöhnlichen instrumentellen Ausstattung, sowie seine stete Unterstützung und Diskussionsbereitschaft in wissenschaftlichen Belangen.

Herzlich bedanken möchte ich mich des Weiteren bei:

Prof. Dr. Rolf W. Hartmann für die Übernahme des Koreferates.

Dr. Angelika Ullrich für die Übernahme des Beisitzes in der Prüfungskommission und die „synthetische“ Unterstützung vieler meiner Projekte.

Dr. Yanyan Li für die praktische Einweisung in die Molekularbiologie und Proteinchemie, sowie für ihre Unterstützung beim Verfassen meiner Publikationen.

Dr. Kira J. Weissman für ihre Ratschläge und die ergiebigen wissenschaftlichen Diskussionen.

Kookie Cortina für ihre Expertenmeinung bei der Erstellung von Abbildungen, ihre Hilfe bei unzähligen Softwareproblemen und ihre gute Freundschaft. Dr. Kathrin Buntin für viele gute Tipps und ihre Hilfe beim Start ins postakademische Berufsleben. Janet Lei und Dr. Alberto Plaza für das Korrekturlesen von Teilen dieser Dissertation. Thomas Hoffmann für seine Hilfe und Ratschläge bei analytischen Problemen. Thorsten Klefisch für die gute Zusammenarbeit im Labor und bei der Betreuung der Seminarreihe Klinische Pharmazie. Ole Revermann für seine Hilfe bei der Auswertung von NMR-Daten. Dr. Carsten Volz für die gute Laborbanknachbarschaft und anregende wissenschaftliche und private Gespräche. Jenifer Herrmann für die durchgeführten Aktivitätsassays und MALDI-Messungen. Dr. Daniel Krug für die Hilfe bei IT-Problemen. Eva Luxenburger für ihre Hilfe bei experimentellen Arbeiten.

Der gesamten Myxoarbeitsgruppe für die gute Atmosphäre und Zusammenarbeit im Labor, und im Besonderen bei den guten Freunden, die ich hier gefunden habe, für die lustigen Abende und den guten Zusammenhalt.

Meinen Freunden für ihr Verständnis und ihren Beistand in schwierigen Zeiten, besonders meinen guten Freunden Dr. Alexander Oster und Dr. Christian Steuer, die mit mir gemeinsam den Weg vom Pharmaziestudium bis zur Promotion gegangen sind.

Und nicht zuletzt meinen Eltern für ihre Geduld, ihren Rat und ihre Unterstützung jedweder Art.

Vorveröffentlichungen der Dissertation

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

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

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

Pistorius, D., Ullrich, A., Lucas, S., Hartmann, R.W., Kazmaier, U., Müller, R. (2011) Biosynthesis of 2-alkyl-4(1*H*)-quinolones in *Pseudomonas aeruginosa*: Potential for therapeutic interference with pathogenicity. *ChemBioChem,* **12,** 850-853.

Pistorius, D., Li, Y., Mann, S., Müller, R. (2011) Unprecedented anthranilate priming involving two enzymes of the acyl adenylating superfamily in aurachin biosynthesis. *J. Am. Chem. Soc.,* **133,** 12362-12365.

Pistorius, D., Li, Y., Sandmann, A., Weissman, K.J., Müller, R. (2011) Completing the puzzle of aurachins biosynthesis in *Stigmatella aurantiaca* Sg a15. *Mol. BioSyst.,* submitted.

Pistorius, D., Müller, R. (2011) Discovery of the rhizopodin biosynthetic gene cluster in *Stigmatella aurantiaca* Sg a15 by genome mining. *ChemBioChem,* to be submitted.

Zusammenfassung

Von Bakterien produzierte Sekundärmetabolite weisen enorme strukturelle Diversität und erstaunliche biologische Aktivitäten auf. Diese Arbeit gewährt neue Einblicke in die Biosynthese der myxobakteriellen Sekundärmetabolitfamilie der Tubulysine durch *in vitro* und *in vivo* Experimente zusammen mit der Identifizierung und Charakterisierung neuer struktureller Varianten.

Die komplizierte Biosynthese der Aurachine in *Stigmatella aurantiaca* Sg a15 wurde aufgeklärt. *In vitro* Experimente legten einen einmaligen Mechanismus zur Bereitstellung der Anthranilsäure Startereinheit offen und *in vivo* Inaktivierungsexperimente resultierten in der Identifizierung der modifizierenden Enzyme, die für die schrittweise Umwandlung von Aurachin D zu Aurachin A verantwortlich sind.

Ein weiterer Gegenstand dieser Arbeit war die Untersuchung der Biosynthese von 2-Heptyl-4(1*H*)-chinolin (HHQ), ein wichtiges Quorum sensing Molekül in *Pseudomonas aeruginosa*. Die Rekonstruktion der HHQ-Biosynthese *in vitro* mittels rekombinantem PqsD Protein gab Aufschluss über dieses Enzym und seine Substrate und führte zur Entwicklung eines *in vitro* Testsystems zur Identifizierung von Inhibitoren der Biosynthese von HHQ, ein vielverspechendes alternatives Angriffsziel für die Therapie von Infektionen mit *P. aeruginosa*.

Das kürzlich sequenzierte Genom von *S. aurantiaca* Sg a15 wurde für einen „Genome mining“ Ansatz genutzt. Dieser setzt sich zusammen aus der gezielten Inaktivierung bestimmter Gene und einer statistikgestützten vergleichenden Analyse der Sekundärmetabolitprofile. Dadurch konnte der Biosynthesegencluster von Rhizopodin identifiziert und charakterisiert werden.

Summary

Bacterial secondary metabolites display enormous structural diversity with astonishing biological activities. This thesis confers new insights into the biosynthesis of the myxobacterial secondary metabolites tubulysins by both *in vitro* and *in vivo* experiments, along with the identification and characterization of new structural variants.

The cryptic biosynthetic pathway of the aurachins in *Stigmatella aurantiaca* Sg a15 was elucidated. *In vitro* experiments revealed an unparalleled priming mechanism for the anthranilate starter unit. *In vivo* gene inactivation experiments resulted in the identification of the tailoring functionalities responsible for the successive transformations from aurachin D to aurachin A.

Another topic of this thesis was the investigation of the biosynthesis of 2-heptyl-4(1*H*)-quinolone (HHQ), an important quorum sensing molecule in *Pseudomonas aeruginosa*. *In vitro* reconstitution of HHQ biosynthesis by employing recombinant PqsD protein has shed light on this enzyme and its substrates, leading to the development of an *in vitro* test system for the identification of HHQ biosynthesis inhibitors, a promising alternative target in *P. aeruginosa* therapy.

Exploiting the newly sequenced genome of *S. aurantiaca* Sg a15 a genome mining approach, combining targeted gene inactivation with a statistics based comparative secondary metabolite profile analysis, was adapted and employed. This culminated in the identification and characterization of the rhizopodin biosynthetic gene cluster.

Table of contents

A. Introduction

1. The importance of natural products in drug discovery – From the past to the future

1.1 The past

Natural products have had and still have an invaluable impact on modern-day medicine. Natural products are defined as chemical substances produced by living organisms - plants, microorganisms, or animals. Since this definition holds true for almost every cellular matter, the term natural product usually refers to compounds that exhibit distinct pharmacological or biological activities. A synonymously used term is secondary metabolites, implying that these organic compounds are not directly involved in normal growth, development, or reproduction and are therefore not absolutely essential for the survival of its producer. Mankind's use of plants for the treatment of ailments can be traced back to the prehistoric times. Through the course of time, this knowledge has grown immensely and led to the development of a number of traditional medicines based on the use of indigenous medicinal plants. The identification of the active components of these medicinal plants was very fruitful for the pioneering days of drug discovery. More than 200 years ago, Friedrich Sertürner isolated morphine from *Papaver somniferum*, an event often regarded as the birth of modern natural product research.[1]

An astonishing number of the drugs in clinical use today are either derived from, inspired by, or natural products themselves. A remarkable 52% of all new chemical entities approved between 1981 and 2006 can be attributed to natural products (**Figure A-1**).[2] 5% are natural products themselves, 23% are semisynthetic derivatives of natural products, and another 24% are somehow derived from a natural product pharmacophore. When considering specific indications, the numbers become even more impressive. For drugs with anticancer activity or antibacterial activity, a combined 63% and 69%, respectively, can be traced back to natural products.

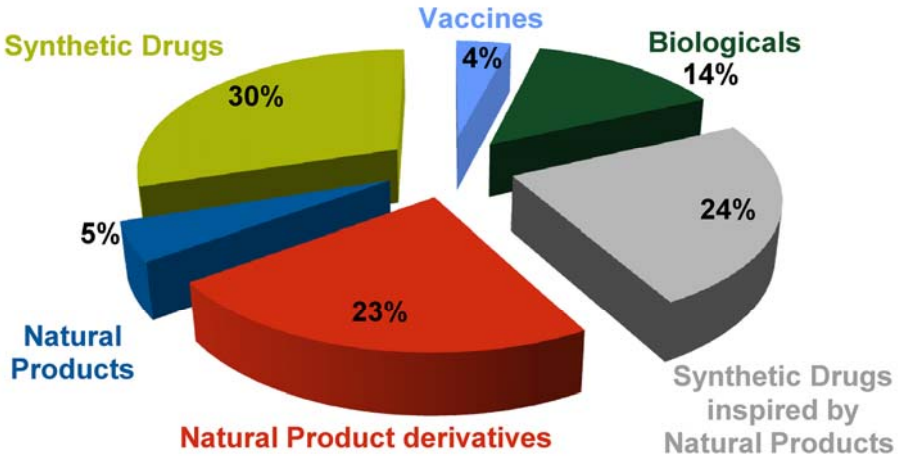


Figure A-1. *Origin of all new chemical entities approved in between 1981 – 2006 (modified from Newman* et al. *[1])*

Nowadays, regulatory agencies request highly detailed knowledge of the chemical structure of the active pharmaceutical ingredient (API), pharmacodynamic and pharmacokinetic parameters, along with proof of efficacy, safety, purity, and content for every single drug submitted for approval. Classic herbal remedies usually present crude mixtures containing, in addition to the active component(s), a jumble of uncharacterized compounds, some exhibiting adverse pharmacological or toxicological properties. A further roadblock to the standardization of these remedies is the questionable content of the active component in the mixture.

Early natural product research focused on the isolation of biologically active compounds from the alkaloid family from plants. It wasn't until the 1930´s that researchers noticed that certain microorganisms, too, were prolific producers of antibacterial and antifungal compounds.These antimicrobials were desperately needed at a time when, despite rapid industrialization, infectious diseases were still the leading cause of death. The most prominent example is certainly the discovery of the β-lactam antibiotic penicillin from *Penicillium notatum* by A. Fleming, H. W. Florey and E. B. Chain.[3] Massive screening programs of microorganisms in the search for new antibiotics and antifungals culminated in a "golden era" in anti-infective research, yielding many antibiotics that are still in use today (**Figure A-2**).

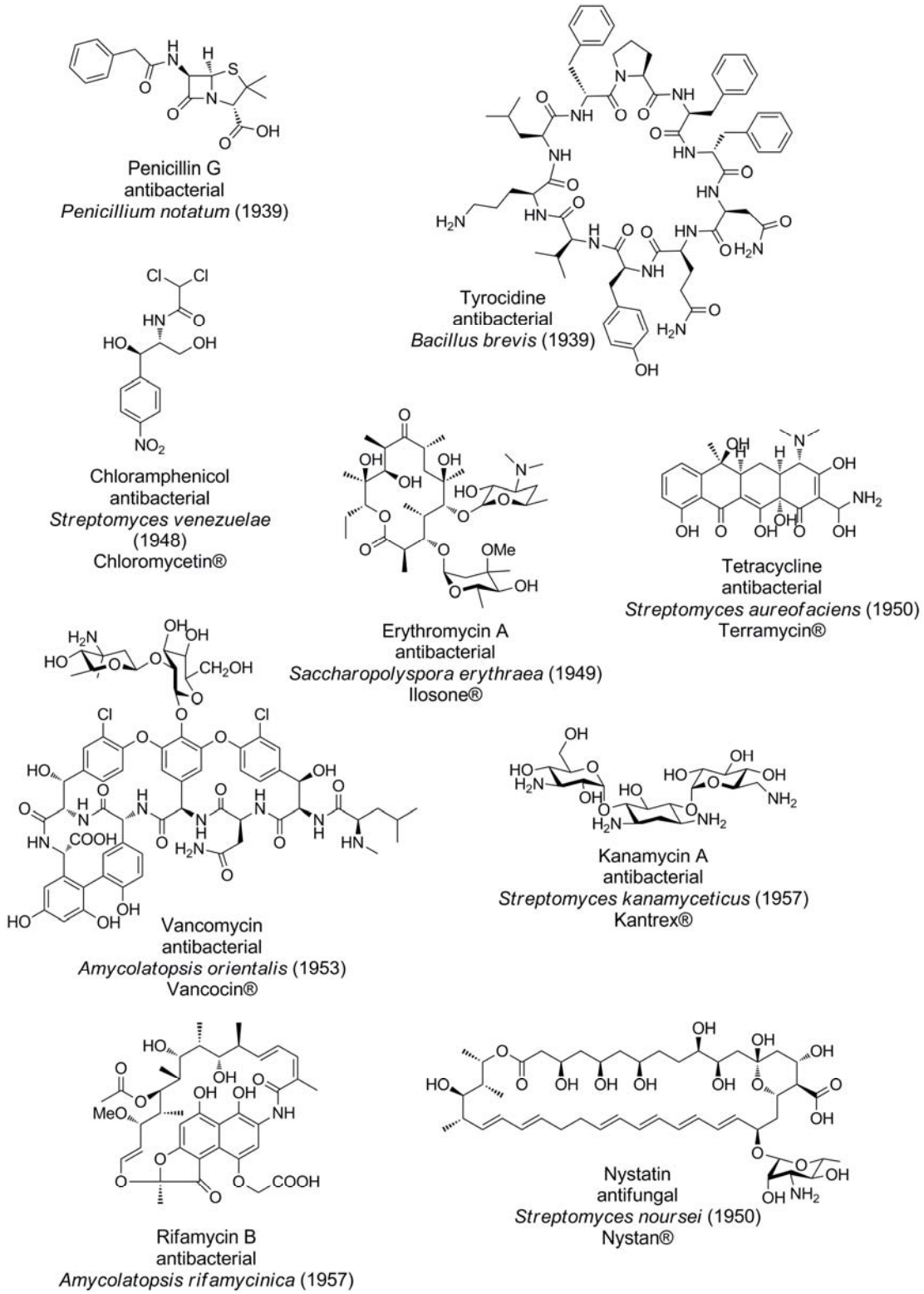


Figure A-2. *Examples of microbial natural products representing important antibiotics in clinical use. Name, activity, native producer, year of first isolation, and trade name (where applicable) are indicated below the respective structure.*

The investigation of their modes of action revealed bacterial cellular functions that could be targeted without having toxic effects in humans, known as the principle of selective toxicity. Among these targets are the bacterial cell wall synthesis (inhibited by β-lactams and

glycopeptides), protein synthesis (inhibited by tetracyclines, chloramphenicol, macrolides, and aminoglycosides), and bacterial RNA polymerase (inhibited by rifamycins). The identification of these targets made them also accessible for further drug development.

At the beginning, antibiotic and antifungal agents were the focus of microbial derived natural product research, but this field has also rapidly proven to be a valuable source of secondary metabolites with a vast array of biological activities, finding their way into the clinics as anticancer and antiparasitic agents, as well as immunosuppressives and cholesterol-reducing drugs (**Figure A-3**).

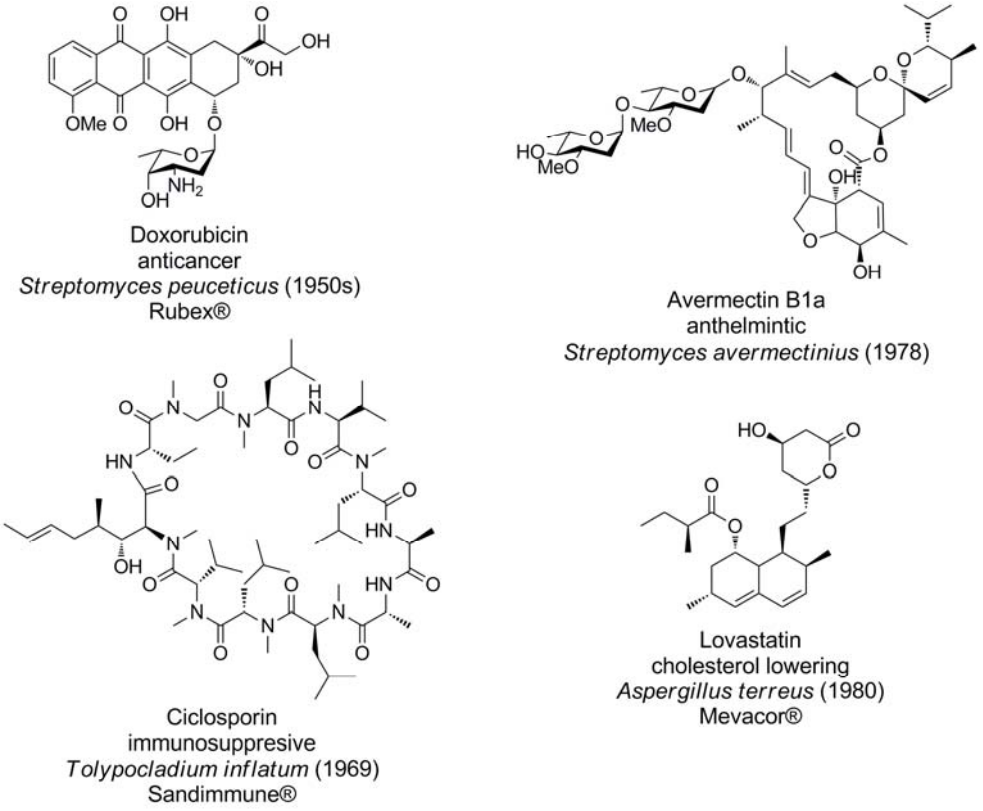


Figure A-3. *Examples of microbial natural products with various biological activities. Name, activity, native producer, year of first isolation, and trade name (where applicable) are indicated below the respective structure.*

1.2 The present and the future

The presented structures provide only a glimpse of the enormous structural diversity covered by microbial secondary metabolites. Although synthetic approaches may provide deeper access to the chemical space, the profit of such an approach might be limited. With an estimated more than 10^{60} different molecules spanning the drug-like chemical space, comprehensive exploitation is nowhere near feasible.[4] The challenge is rather to identify areas within the chemical space containing a high percentage of biologically active

compounds. In this context, it is of particular interest that the vast number of possible primary protein sequences is greatly reduced by the spatial arrangement of secondary structure elements like α-helices and β-sheets. This leads to a predicted 1000 – 8000 different fold types for proteins in nature.[5] Natural products can be considered as privileged structures for specific interactions with diverse biological targets due to their evolutionary selection over a multitude of generations.[6] Statistical evaluation of ~18000 protein cavities, normally representing the most promising targets for small molecule-protein interactions, revealed mainly volumes in the range of 300 to 800 $Å^3$.[7] This value is in very good agreement with the calculated volumes of most of NP scaffolds.

Another major requirement in modern drug development is oral bioavailability. A retrospective evaluation of chemical properties of drugs on the market led to the establishment of Lipinski´s Rule of Five.[8] The bottom line is that most orally active drugs do feature no more than 5 hydrogen bond donors or more than 10 hydrogen bond acceptors in combination with a molecular weight below 500 daltons and an octanol-water partition coefficient (log *P*) below 5. The fact that the majority of natural products disobey this rule of thumb while still exhibiting, in many cases, reasonable bioavailability led to the establishment of Lipinski´s fifth rule which states that the first four rules do not apply to natural products or any molecule recognized by an active transport system.

The high expectations of discovery of drug leads from libraries based on unfocused combinatorial chemistry have not met with success. Today, most industrial drug discovery programs rely on high-throughput screening (HTS) of focused synthetic libraries. The advantages, usually easy access and modifiability of the hits, carry the steep price tag of a perilously low hit rate. Taking the ~7000 known polyketide secondary metabolites into account, natural product library screening has yielded more than 20 approved drugs, corresponding to a hit rate of 0.3% and several orders of magnitude higher than the <0.001% usually obtained from synthetic compound libraries.[9] In return, exploitation of this potential is complicated by the labor intensive identification and purification of microbial secondary metabolites. One approach to bypass this bottleneck is the testing of either fractionated or crude extracts from microbial fermentation in HTS. While a tremendous increase in throughput is achieved, the reliability and robustness of the results might suffer due to a number of problems. The concentration, stability, or solubility of a bioactive compound in the extract may be too low to be detected by assays, leading to missed hits. Furthermore, interpretation of the results may be complicated by synergistic or antagonistic effects of the

complex mixture of compounds. Some antibiotics (*e.g.* members of the aminoglycosides) occur in a high percentage of soil actinomycetes. The activity of these known compounds can mask that of other antimicrobial compounds present in the crude extract. Lastly, every confirmed activity must be correlated to its bioactive compound, requiring labor-intensive processes such as dereplication, upscaling, isolation, and structure elucidation.

As bacterial antibiotic resistance continues to exhaust our supply of effective antibiotics, there is high demand for the discovery of new lead structures in order to prevent the antibiotic pipeline from drying up. The emergence of multiresistant nosocomial pathogens is especially alarming; a group of six bacteria, termed the ESKAPE pathogens, is responsible for two thirds of all health care associated infections. These ESKAPE pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumonia*, *Acinetobacter baumanii*, *Pseudomonas aeruginosa*, and *Enterobacter* species) escape most antibiotic therapies due to their resistance against a large number of antibiotics. Therefore, new drugs are needed to fight them effectively.

Over the last 50 years, streptomycetes have emerged as the major source of antibiotic secondary metabolites. Of the ~12000 antibiotic compounds known by 1995, 55% were derived from streptomycetes and another 11% from other actinomycetes.[10] The genesis for this prolific production of secondary metabolites is the enormous competition in their natural soil habitat. The continuous exploitation of this repository elicits difficulties as there is frequent rediscovery and reisolation of known structures. Within the fungal and bacterial kingdoms, there remain enormous numbers of species that have not been thoroughly investigated for their potential to produce bioactive secondary metabolites. In the barely exploited marine environments alone, around $3x10^{30}$ microorganisms are estimated to be present, giving rise to a gigantic pool of natural products.[11] Unfortunately, there are also estimates that more than 99% of all bacteria cannot be cultured under standard laboratory conditions. It is therefore of utmost importance to exploit the most promising and accessible microbial resources. In this framework, myxobacteria have attracted attention as prolific producers of novel secondary metabolites.

2. Myxobacteria as novel source for bioactive natural products

The bulk of myxobacteria have been isolated from soil, decaying wood, and the dung of herbivores,[12] although recently novel isolates from marine and moderate halophilic environments have also been reported.[13;14] Myxobacteria are Gram-negative, obligate aerobic mesophiles belonging to the δ-group of proteobacteria. The typically rod-shaped vegetative cells exhibit dimensions of 3-12 µm in length and 0.7-1.2 µm in width (**Figure A-4A**).[15] Their lifestyle is rather "micropredatory", preying on other microorganisms than being saprotrophic. Myxobacteria are taxonomically contained in the order *Myxococcales*, which comprises the three suborders *Cystobacterineae*, *Sorangineae*, and *Nannocystineae*.[16] They are characterized by outstanding nutrient exploitation, motility and life cycle behaviours which differ again distinctively between the members of the different suborders or families. *Myxococcus xanthus* (a member of the *Myxococcaceae* family within the suborder *Cystobacterineae*) is, for example, is able to swarm on solid surfaces such as agar by chemotaxis, hunting out nutrients not within its immediate proximity (**Figure A-4B**). These "nutrients" are often other bacteria or yeast digested by *M. xanthus* through secretion of a variety of lytic enzymes that break open the cells of the prey.[17] The high concentration of these extracellular lytic enzymes required for cell lysis is achieved by the coordinated action of many individual cells, leading to the perception of "wolf pack bacteria". The released proteins, lipids and nucleic acids are then further digested and consumed as monomeric building blocks. This lifestyle is also mirrored in liquid culture, where *M. xanthus*, like most myxobacteria, grows preferentially in media containing small peptides from protein digestions in comparison to media containing mainly carbohydrates as carbon and energy sources. *Sorangium cellulosum* deserves special mention as, aside from a characteristic micropredatory lifestyle, they exhibit the unique ability of cellulose degradation, even using it as a sole carbon and energy source.[18] Under low nutrient conditions, myxobacteria can aggregate and undergo differentiation into specialized cell types, a process that finally culminates in characteristic multi-cellular structures called fruiting bodies (**Figure A-4C**). These fruiting bodies house so called myxospores (**Figure A-4D**), which are highly resistant to environmental stress factors as temperature extremes, desiccation, and UV radiation. These myxospores allow the bacteria to outlast periods of starvation and to return to a monocellular vegetative growth lifestyle upon restoration of favorable nutritional and environmental conditions. The shape, size, and

morphology of myxobacterial fruiting bodies vary greatly between members of the different genera and range from spherical cell clusters (e.g. *Stigmatella*) to sophisticated tree-like structures (e.g. *Chondromyces*). Due to these developmental and morphological characteristics of myxobacteria, they have been the subject of microbiological studies for several decades.

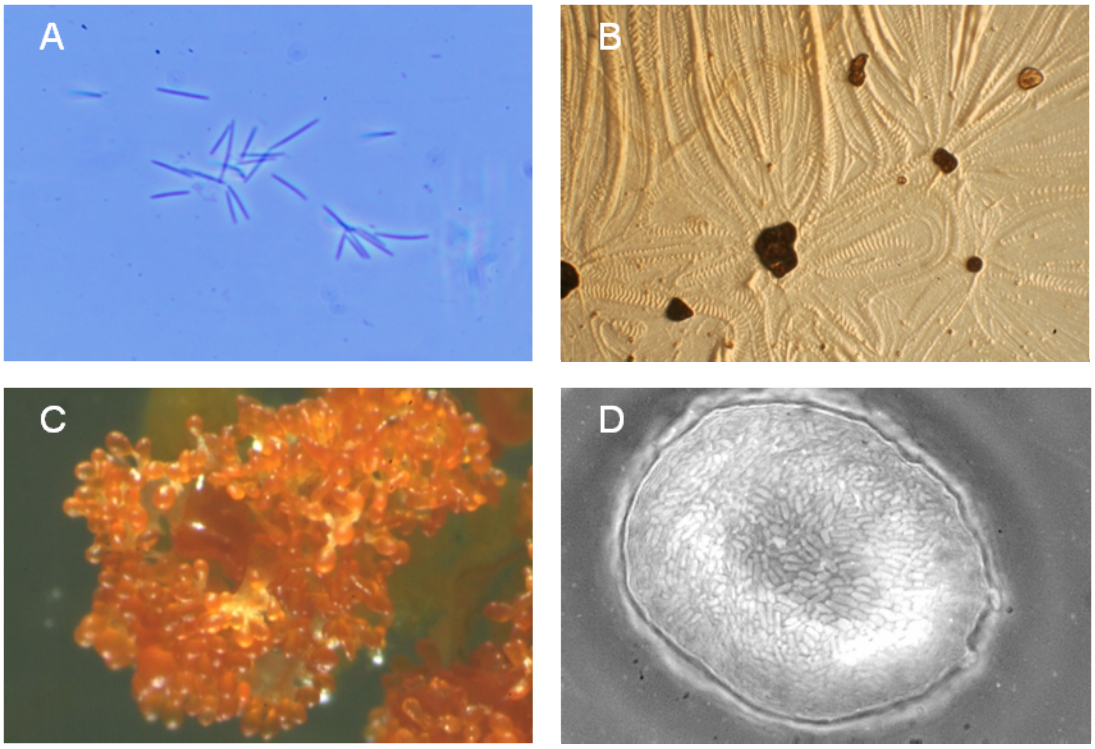


Figure A-4. *Developmental growth stages of* S. aurantiaca *Sg a32.* ***A)*** *Vegetative cells taken from a swarm colony.* ***B)*** *Swarming colony on agar plate showing the rippling and radial vein pattern.*
C) *Cluster of fruiting bodies on agar surface.* ***D)*** *Arrangement of myxospores in a sporangiole. Pictures A) and D) are obtained from phase contrast microscopy and pictures B) and C) from stereoscopic microscopy.*

Another special feature of myxobacteria, revealed recently through enormous progress in sequencing technology, are their very large genomes, comprising 9 – 13 million base pairs (Mb). *Sorangium cellulosum* So ce56, with a size of 13.0 Mb, has the largest genome of all sequenced bacteria[19], although other members of the genus *Sorangium*, such as So ce38, appear to be even larger, with an estimated genome size of 14.5 Mb (unpublished results). The myxobacterial genomes exhibit a high average GC-content of 67 – 72%.[15] Of particular interest is the fact that myxobacteria dedicate a high percentage of their sequence information (usually around 8 – 10%) to genes encoding enzymes involved in secondary metabolism.[19;20] The number of secondary metabolite gene clusters found in one myxobacterial strain can

range from around 10 up to 20. Thereby, their biosynthetic potential clearly outcompetes that of sequenced streptomycetes.[21-23]

Even without the benefit of genetic background knowledge, the potential of myxobacteria as prolific producers of bioactive secondary metabolites was recognized in the early 1980s, leading to the establishment of a program for the systematical exploitation of this yet untapped source at the German Research Center for Biotechnology (GBF). This continuing success story has led to the isolation and structure elucidation of more than 100 new core structures and approximately 500 structural variants.[24] Intriguingly, many of these myxobacterial secondary metabolites exhibit exceptional structural features and modes-of-action not employed by secondary metabolites isolated from other bacterial sources.[25;26]

The most prominent myxobacterial secondary metabolites by far are the epothilones, produced by several *Sorangium* strains.[27] The mode of action of the cytostatic epothilones is identical to that of paclitaxel (Taxol®), approved in 1992 for the treatment of various cancers. Both agents bind β-tubulin and promote tubulin polymerization, disturbing the natural dynamics between polymerization and dissociation of the microtubule network essential for vital cellular functions and finally causing apoptosis. [28] Interestingly, epothilones retain their cytotoxic activity against a number of paclitaxel resistant tumor cell lines and show low susceptibility to multiple tumor resistance mechanisms.[29] Combined with a reasonable side effect profile and pharmacokinetic properties, ixabepilone (Ixempra®), a semisynthetic epothilone B derivative (**Figure A-5**), was approved by the American Food and Drug Administration (FDA) for the treatment of breast cancer in 2007.[30] Because cancer cells divide at an accelerated rate in comparison to normal cells, a process dependent on proper microtubule function, the eukaryotic cytoskeleton is a validated target for anticancer therapy.[31] A cytotoxic effect through interference with the microtubule network can be caused by both the induction and the inhibition of tubulin polymerization. The latter mechanism is employed by the vinca alkaloids vinblastine (Velbe®) and vincristine (Oncovin®), which have been in use in clinical oncology for almost 50 years. Two families of myxobacterial secondary metabolites, the tubulysins and the disorazols, show an identical mode of action as the vinca alkoids.[32;33] The tubulysins inhibit tubulin polymerization and induce microtubule depolymerization and are the most potent tubulin polymerization inhibitors described to date.

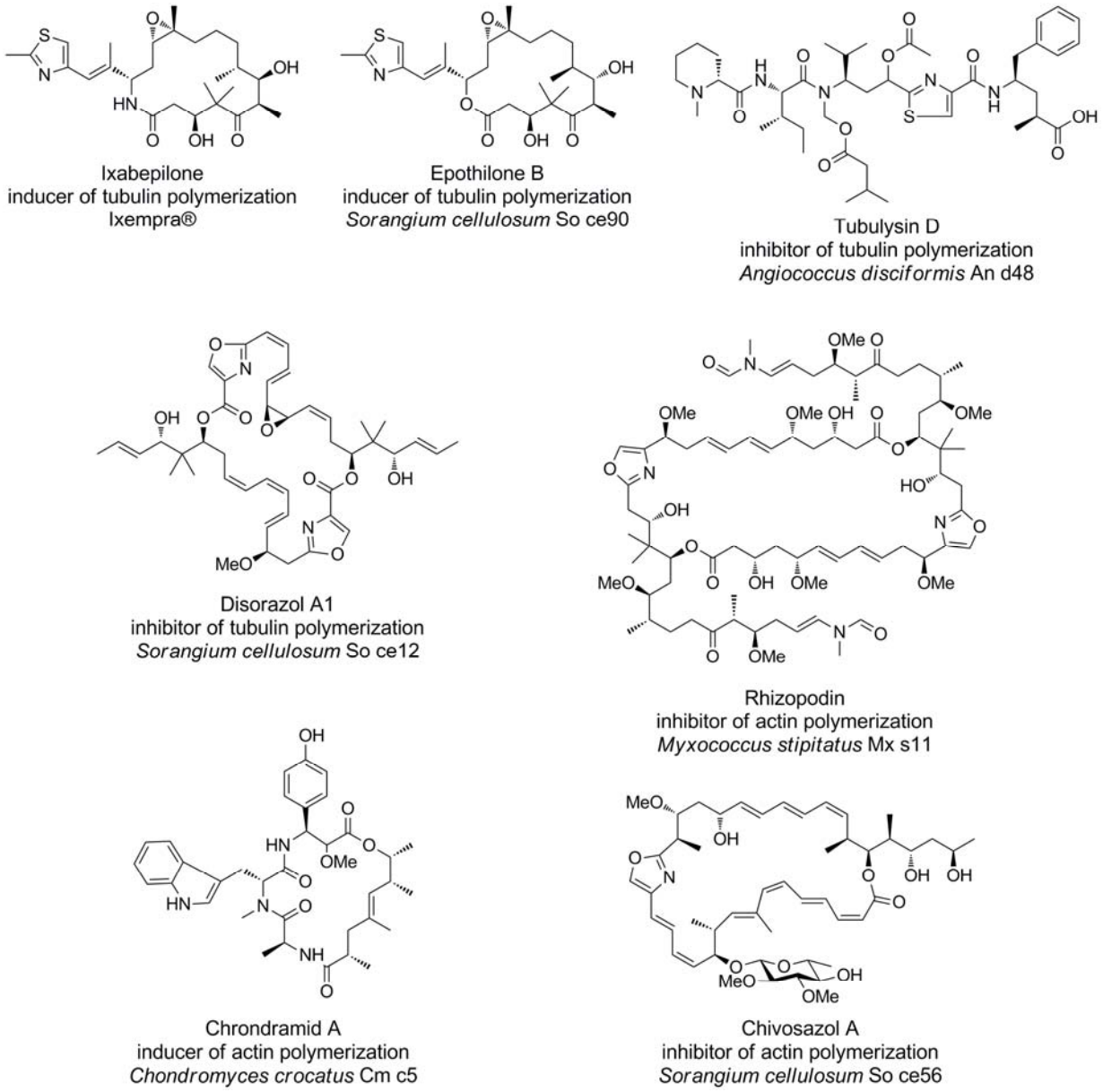


Figure A-5. *Myxobacterial secondary metabolites that target the eukaryotic cytoskeleton. Name, mode of action, and native producer are indicated below the respective structure.*

In general, the eukaryotic cytoskeleton is rarely targeted by natural products of bacterial origin. However, in addition to those acting on microtubules, three other myxobacterial secondary metabolites interact with actin filaments. Once again, both polymerization inducers and inhibitors are represented. The chondramides, produced by several *Chondromyces* strains, enhance actin polymerization.[34] Rhizopodin, a C_2-symmetric dilactone, bears two enamide side chains, each of which binds a single G-actin molecule, resulting in a ternary rhizopodin/G-actin complex and inhibiting actin polymerization.[35] The chivosazoles also inhibit actin polymerization, though incubation with eukaryotic cells suggests a slightly different mode of action from rhizopodin, which itself still has to be elucidated in detail.[36]

There are several other myxobacterial secondary metabolites with rare modes of action (**Figure A-6**). Argyrin A, another compound of potential value in cancer therapy, acts as a proteasome inhibitor.[37] The target of the fungicide soraphen A is the fungal acetyl-CoA carboxylase, blocking *de novo* fatty acid biosynthesis.[38-40] Several myxobacterial secondary metabolites like sorangicin A,[41] corallopyronin A,[42] and ripostatin A[43] inhibit bacterial RNA polymerase. Notably, the latter two retain their activity against rifamycin-resistant *rpoB* mutants of *Staphylococcus aureus*, potentially making them very valuable in therapy.[44] Yet another common target among myxobacterial secondary metabolites is the electron transport chain, which is inhibited by the aurachins,[45] myxalamids,[46] myxothiazoles,[47] and the stigmatellins.[48]

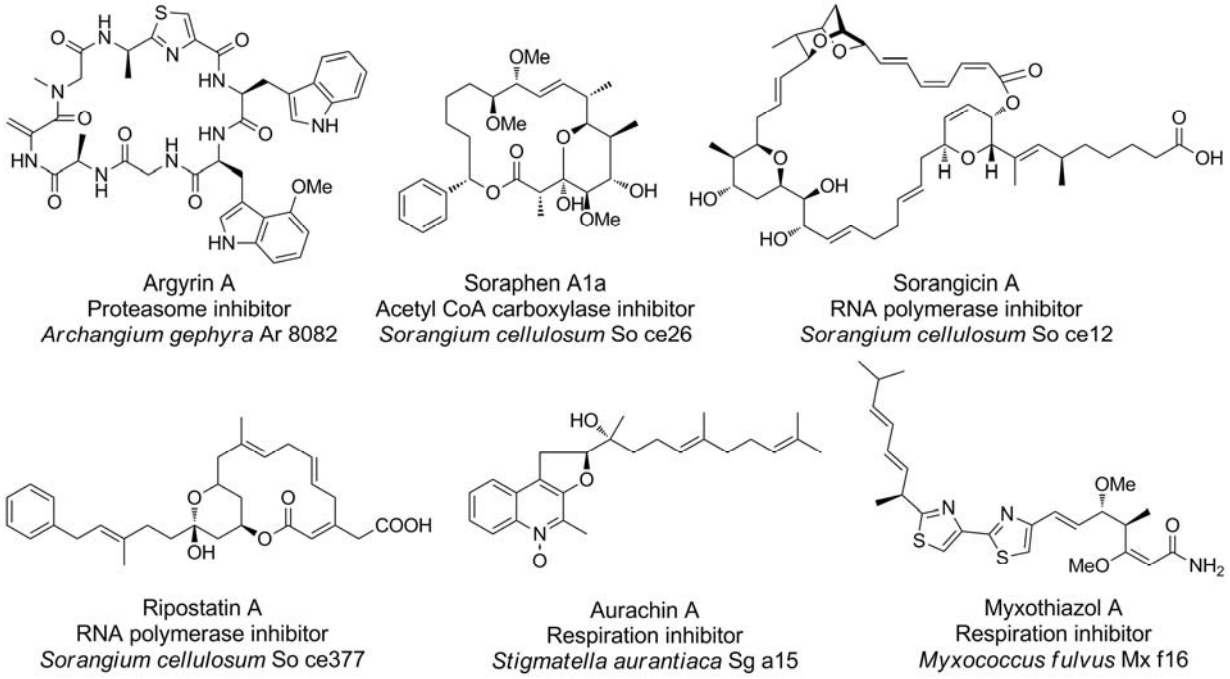


Figure A-6. *Myxobacterial secondary metabolites with rare targets. Name, mode of action, and native producer are indicated below the respective structure.*

3. Biosynthesis of secondary metabolites

3.1 Nonribosomal peptide and polyketide biochemistry

Most of the bacterial secondary metabolites are assembled by nonribosomal peptide synthethases (NRPSs), polyketide synthases (PKSs) or hybrids of the two.[49;50] Both NRPS and modular type I PKS are "biosynthetic assembly lines", consisting of giant multifunctional enzymes containing several repeated functional units known as modules. By definition, each module is responsible for the incorporation of one monomeric building block and thereby for chain extension of the nascent natural product. PKSs are further categorized as iterative or noniterative, depending on whether or not each module catalyzes more than one round of chain extension.[51;52] The modules are further split up into domains that catalyze one of several successive reactions during one extension step. In NRPS, the building blocks are proteinogenic as well as nonproteinogenic amino acids; coenzyme A (CoA) thioesters of short dicarboxylic acids are the building blocks of choice in PKS. Despite differences in building blocks and the chemical reactions catalyzed, there are striking similarities between the NRPS and PKS biosynthetic logics.[53;54] Both systems utilize analogous elongation strategies in which the growing chain is tethered through thioester linkage to the 4´-phosphopantetheine (Ppant) prosthetic group of dedicated carrier proteins (CPs).[54] Post-translational attachment of this non-protein organic molecule to a highly conserved serine residue is catalyzed by dedicated Ppant transferases (PPTases) that convert the catalytically inactive *apo*-CPs to the catalytically active *holo*-CPs.[55] The flexible nature and length (~20 Å) of the Ppant prosthetic group allows for efficient positioning of the tethered intermediates for processing by the spatially distinct domains within a module and transfer of the growing intermediate by interaction with the domain catalyzing the condensation reaction in the downstream module. The covalent linkage to the CP makes these processes far more efficient than simple diffusion of dissolved intermediates between the single domains. Further, the building blocks and the intermediates are activated by the energy-rich thioester linkage which is crucial for the condensation steps.[56] The offloading of the final product from the assembly line NRPS and PKS also employ analogous strategies. Both have evolved dedicated chain-terminating thioesterase (TE) domains responsible for detachment of the fully extended product from the terminal CP modules and off-loading of the final secondary metabolite. In this process, the TE can catalyze either hydrolysis, leading to the release of a linear product, or macrolactonization

by intramolecular nucleophilic attack of a specific amino or hydroxyl moiety on the terminal acyl group of the covalent TE-*O*-intermediate, leading to the formation of lactames or lactones, respectively.[57]

3.2 NRPS biochemistry

NRPS and PKS differ, however, in the core domains liable for building block selection (and activation), building block condensation and downstream processing. The minimal NRPS module required for one extension step consists of an adenylation (A) domain, a peptidyl carrier protein (PCP; also referred to as thiolation (T) domain), and a condensation (C) domain.[58] A typical loading module lacks a C domain.[59] Every cycle of chain extension is initiated by the action of an A domain, which successsively selects, activats and transfers an amino acid to the PCP within the same module. The efficiency with which a specific amino acid building block is selected is crucial for the formation of a natural product with a defined peptide sequence, which is in turn crucial for biological activity. Sequence alignments of A domains with known amino acid specificity, combined with structural data information, led to the identification of 8-10 pivotal residues within the substrate binding pocket that determine the substrate specificity.[60-62] These indicative residues, referred to as the nonribosomal code, present a powerful tool for *in silico* prediction of A domain substrate specificity. The variety of employed building blocks does not only include the pool of proteinogenic amino acids but also nonproteinogenic amino acids and even aryl carboxylic acids.[59] After the specific selection of an amino acid, the A domain catalyzes the formation of the respective aminoacyl-adenylate intermediate with the consumption of ATP. The adenylate, a high-energy acid anhydride, provides the energy for the thioester-forming step, covalently linking the amino acid to the terminal sulfhydryl group of the PCPs Ppant co-factor via its carboxyl function.[63] The C domain catalyzes peptide bond formation between amino acid building blocks tethered to the PCPs of adjacent modules by nucleophilic attack of the amino group of the downstream aminoacyl-*S*-PCP on the acyl group of the upstream peptidyl-*S*-PCP.[64] This condensation reaction yields an extended peptide intermediate covalently linked to the PCP of the downstream module, allowing for further elongation steps (**Figure A-7**).

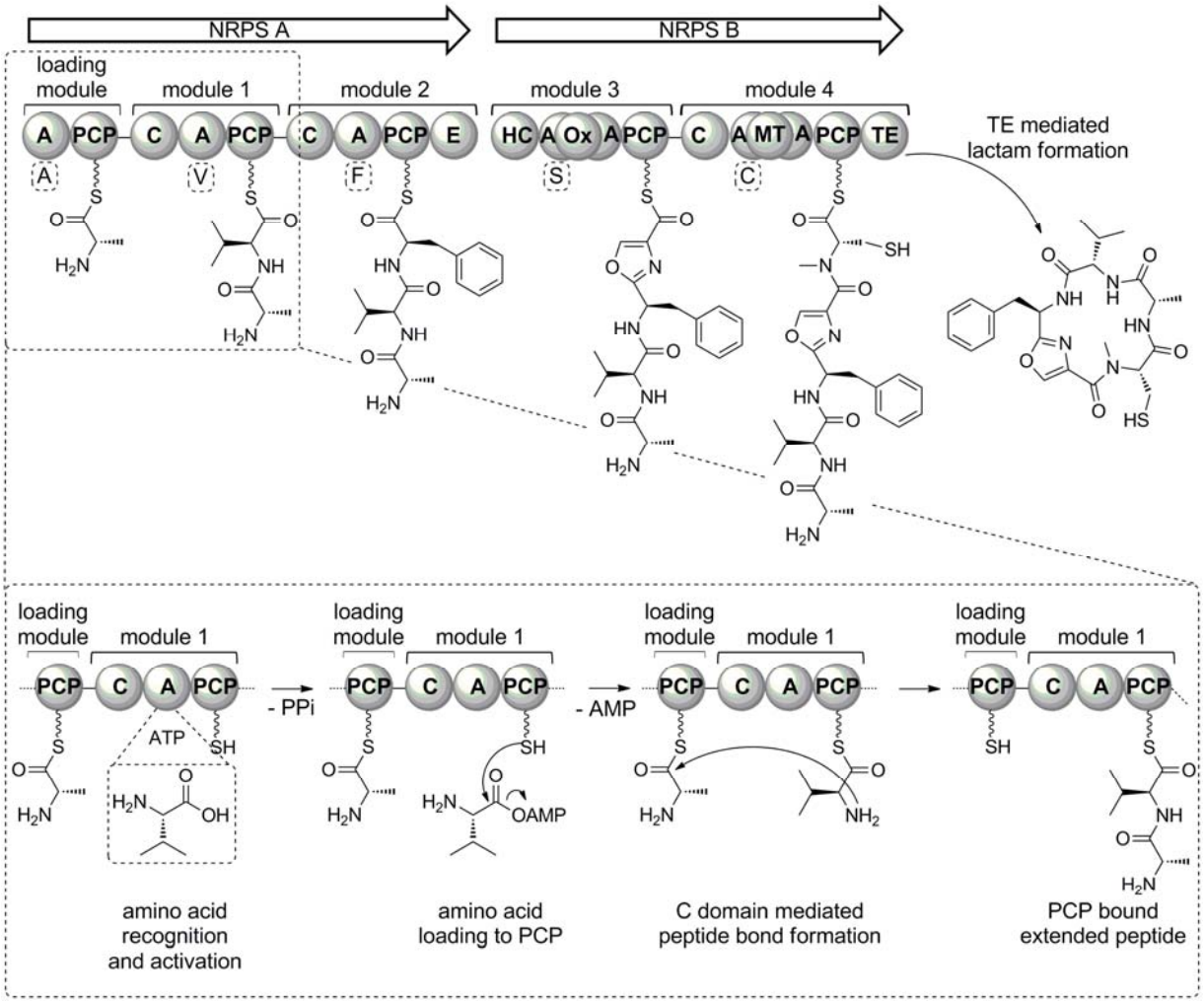


Figure A-7. *Schematic overview of NRPS biosynthesis of a hypothetical nonribosomal polypeptide product and illustration of the discrete reactions within one elongation cycle.*

In addition, optional domains that are responsible for methyl transfer (MT), epimerization (E), heterocyclization (HC) and oxidation (Ox) can increase the structural diversity of the product.[65] These tailoring functionalities are embedded in the modular context and fulfill their actions either on the PCP bound building blocks or intermediates. The MT domains catalyze *N*- or *C*-methylation of the building blocks using *S*-adenosyl-L-methionine (SAM) as cofactor. This modification is performed on the aminoacyl-*S*-PCP intermediate prior to the condensation reaction.[65] *N*-methylation is a structural feature found in many NRP products (*e. g.* ciclosporin) which reduces the proteolytic liability of the compound and thereby enhances its biological activity.[66] A similar effect can be achieved by incorporation of D-amino acids into the product. In most cases, this is realized by incorporation of the respective L-amino acid that is epimerized by an E domain acting on the peptidyl-*S*-PCP intermediate after peptide bond formation.[67] HC domains act on intermediates with

C-terminal cysteine, serine, or threonine residues to yield five-membered thiazoline or oxazoline heterocycles. These HC domains replace the C domain in the respective module and thus catalyze peptide formation and heterocyclization of the nucleophilic amino acid side chains with the *C*-penultimate acyl group.[68] The resulting heterocycles are often further transformed by Ox domains catalyzing flavin mononucleotide (FMN)-dependent oxidation to thiazoles or oxazoles.[69]

3.3 PKS biochemistry

3.3.1 Modular type I PKS

The minimal set of domains within a PKS module consists of an acyltransferase (AT), an acyl carrier protein (ACP), and a β-ketoacylsynthase (KS) domain. The AT is responsible for extender unit selection and transfer to the Ppant co-factor of the ACP. Unlike NRPS, where a low reactivity building block is activated *in situ* by a dedicated domain via consumption of ATP, the AT domain selects pre-activated CoA thioesters of short dicarboxylic acids. Analogous to A domains, AT domains can specifically select particular building blocks and their substrate specificity can be predicted *in silico* based on sequence alignments.[70;71] The commonly used extender units are malonyl-CoA and methylmalonyl-CoA. In addition, the AT can show specificity for one of the more unusual extender units such as ethylmalonyl-CoA, hydroxymalonyl-CoA, methoxymalonyl-CoA, or aminomalonyl-CoA.[72-74] For chain extension, the polyketide intermediate must first be transferred from the ACP of the upstream module to a conserved active site cysteine of the KS by transthioesterification. The KS catalyzes a decarboxylative Claisen thioester condensation between the extender unit and the acyl-*S*-KS intermediate, leading to a C_2 extended β-keto-acyl-*S*-ACP intermediate (**Figure A-8A**).[75] The structural diversity of the product can be increased by processing of the β-keto function by a series of optional reductive domains within the module (**Figure A-8A**). A complete so-called "reductive loop" consists of ketoreductase (KR), dehydratase (DH) and enoylreductase (ER). The KR reduces the β-keto function with NADPH as cofactor, resulting in a β-hydroxy function.[76] This can, in turn, be dehydrated by the DH to yield a α,β-enoyl moiety which can finally be reduced to a methylene function by the ER with the consumption of NADPH. As not all of these domains have to be present within a module, the final product can contain hydroxyl- as well as enoyl- or fully reduced methylene functions.[77] Additional diversity can be introduced by MT domains introducing

either *O*- or *C*-methyl groups. Selection of the starter unit in type I PKS biosynthesis is accomplished by dedicated loading modules. This can be a module of KS-AT-ACP architecture with selection and ACP priming of either malonyl-CoA or methylmalonyl-CoA, followed by KS mediated decarboxylation to yield either acetyl- or propinyl-starter units. In this type of module, the KS is disabled for condensation by the exchange of the active site cysteine against glutamine, and is therefore referred to as a KS^Q domain.[78] Alternatively, the loading module can consist of an AT-ACP didomain in which the AT selects and transfers short straight-chain, branched-chain or aromatic monocarboxylic acids immediately.[79-81]

3.3.2 Special cases of modular type I PKS

In contrast to classical *cis*-AT (or canonical) type I PKSs, there are also *trans*-AT type I PKSs which lack integral AT domains and complement the AT functionality by freestanding enzymes.[82] In addition to the exceptional acyl transfer mechanism, these systems quite often exhibit highly aberrant architectures with unprecedented functionalities, novel domain orders or apparently superfluous functions. Whereas NRPSs and textbook *cis*-AT PKSs show strong correlation between the organization of modules and domains in the megasynthases and the structural elements found in the final product, also known as colinearity rule,[83] this does not hold true for *trans*-AT PKSs. The colinearity rule allows for prediction of metabolite structures from enzyme architectures which can again be predicted from DNA sequences and vice versa, and is an effective tool in the genome mining context. Recently, an innovative approach based on phylogenetic clustering of KS sequences in relationship to their substrate specificity was developed and allowed prediction of core structures from *trans*-AT systems with a significantly improved reliability.[84]

The structural diversity of natural products assembled by NRPS or PKS can further be increased by the cooperation of both systems, resulting in mixed polyketide-polypeptide compounds.[81;85]

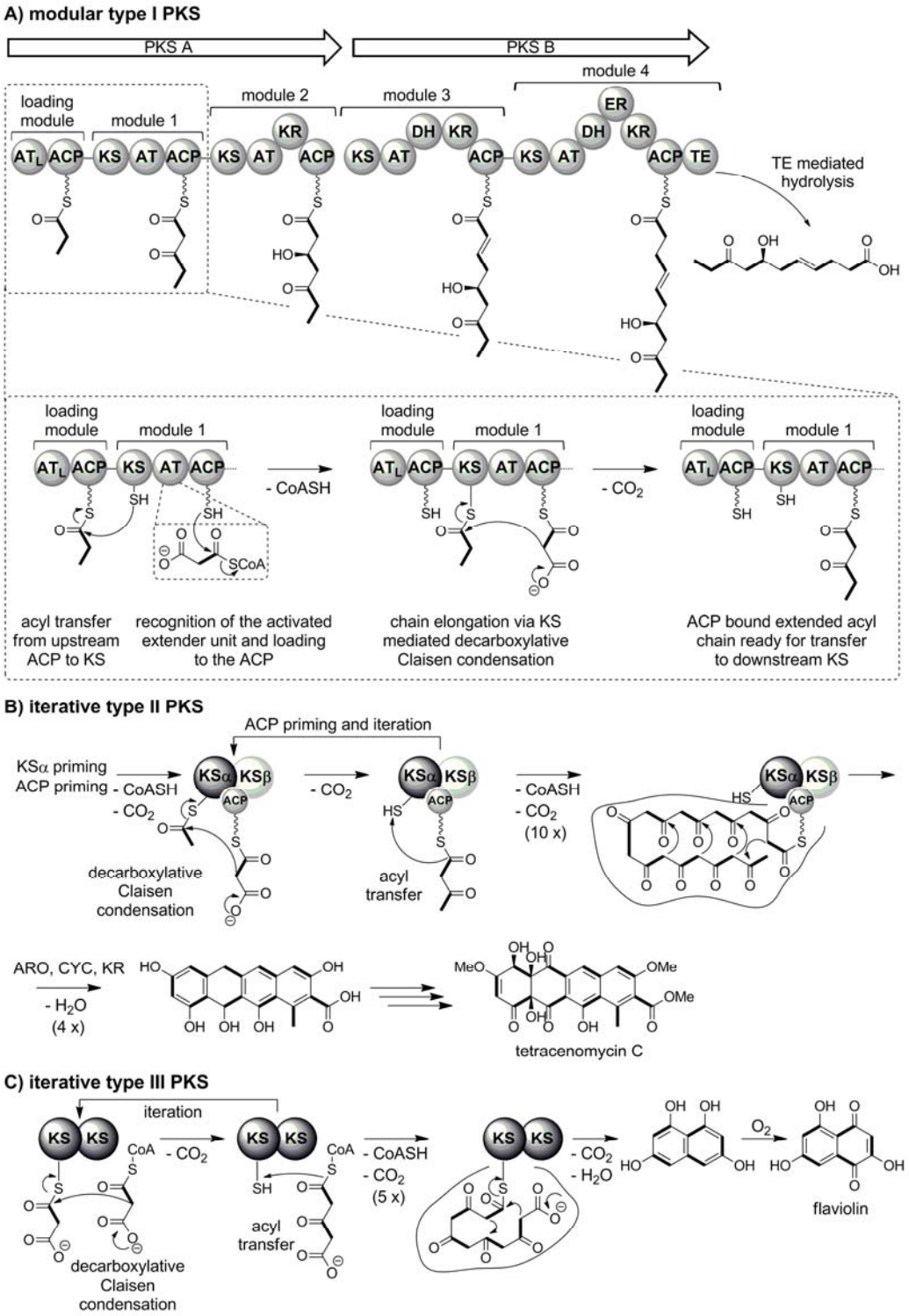


Figure A-8. *Schematic overview of the different types of PKS systems and their respective biochemistry.* ***A)*** *Biosynthesis of a hypothetical polyketide product build up by a modular type I PKS system and illustration of the discrete reactions within one elongation cycle.* ***B)*** *Type II PKS consisting of iteratively acting subunits exemplified for tetracenomycin biosynthesis.* ***C)*** *Type III PKS consisting of one iteratively acting homodimer subunit as exemplified for flaviolin biosynthesis.*

3.3.3 Type II PKS

Aside from these classes of modular and non-iteratively acting multifunctional enzymes, there are two other classes of biosynthetic enzymes that use discrete, iteratively acting monofunctional enzymes for polyketide assembly, the type II PKSs and type III PKSs. Type II PKSs are found almost exclusively in Gram-positive actinomycetes (with only two known examples from Gram-negative bacteria[86;87]) where they are involved in the biosynthesis of aromatic polycyclic polyketides, also called polyphenols.[88] The "minimal type II PKS" is composed of two ketosynthase units, KSα and KSβ (also referred to as chain length factor (CLF)) and an ACP, which serves to tether the growing intermediate and extender unit supplier (**Figure A-8B**).[89] There is ongoing debate in the literature whether or not a malonyl-CoA:ACP transacylase (MCAT), which is usually recruited from fatty acid synthases and serves for rapid malonyl-ACP priming, should be added to the minimal type II PKS functionalities. This set of proteins employs identical overall biochemical logic for the iterative decarboxylative condensation of malonyl-CoA building blocks as type I PKSs. The two KS subunits KSα and KSβ show high sequence homologies to each other, but only KSα possesses the active site cysteine crucial for transient covalent tethering of the intermediate and thereby polyketide assembly.[90] However, both subunits are absolutely required for polyketide biosynthesis. The KSβ subunit is involved in malonyl-CoA loading and generation of acetyl-KS from decarboxylation of malonyl-ACP.[78] Further, the KSβ subunit dictates the carbon chain length of the final product and thereby the number of iterative condensation steps.[91] During polyketide assembly, the growing intermediate is shuttled between the KSα and the ACP. Since there is no intercyclic reductive processing, the nascent chain consists of highly unstable poly-β-keto functions, which should be prone to spontaneous cyclisation. From a structural study, it was deduced that KSα/KSβ form a heterodimer with a cleft that keeping the nascent polyketide chain extended.[92] Spatial separation of the ketide groups is therefore achieved, allowing for stabilization of the intermediate until the final number of elongation steps is completed to yield octaketides (as actinorhodin), decaketides (as tetracenomycin), or dodecaketides (as pradimicin), in the majority of cases. Additional discrete functionalities as ketoreductases, cyclases, and aromatases influence the folding pattern of the final polycyclic product. Further modifications by tailoring enzymes as P450 monooxygenases, methyltransferases, and glycosyltransferases contribute to the huge structural diversity of polyphenolic polyketides found in nature.

3.3.4 Type III PKS

Type III PKS were long thought to derive only from plants, where they are also referred to as chalcone synthases (CHSs) or stilbene synthases (STSs), depending on the generated products. CHSs are found ubiquitously in plants and catalyse the condensation of three malonyl-CoA building blocks with a phenylpropanoid-derived *p*-coumaroyl-CoA starter unit. Subsequent cyclisation leads to the formation of the final product chalcone.[93] It was only in the last decade that a number of homologous enzymes have been discovered in bacteria.[94;95] These bacterial type III PKSs share only weak sequence homology to plant CHSs (typically ~25% identity), and are therefore unsurprisingly involved in the formation of products other than chalcones.

Type III PKSs are structurally and mechanistically quite distinct from type I and type II PKSs. Instead of using CP-Ppant-tethered dicarboxylic acid building blocks, type III PKSs employ free CoA thioesters as a building block supply and anchoring of the polyketide intermediate. They act as homodimeric proteins and utilize a single active site to iteratively perform loading of the acyl-CoA intermediate to the active site cysteine and chain extension via decarboxylative condensation of malonyl-CoA extender units (**Figure A-8C**). In addition, they catalyze intramolecular cyclization to release the final product.

3.4 Further classes of secondary metabolites

Aromatic amino acids, benzoic acids, and cinnamic acids are produced via the shikimate pathway. Further modifications of these starting compounds lead to the formation of lignans and lignin, phenylpropenes, and coumarins.[96] The phenazines are another large class of secondary metabolites whose biosynthetic pathway branches off from the shikimate pathway before anthranilate formation.[97] They share a dibenzo-annulated pyrazine core structure that can be highly diversified by various substituents. Another large and structurally diverse family of natural products is the terpenoids, which are classified according to the number of incorporated isoprenoid units. They are generated by head-to-tail condensation of C_5 dimethylallyl pyrophosphate (DMAPP) and isopentenyl pyrophosphate (IPP) units[98], formed through the mevalonate or deoxyxylulose pathways. “True alkaloids” are defined by a nitrogen-containing heterocycle derived from an amino acid. Although classified mainly based on the nature of their heterocycle, a uniform classification is not feasible due to their enormous structural diversity. Glycosinolates are a group of highly water soluble anions that incorporate sulfur and nitrogen atoms and are derived from glucose and an amino acid. In

contrast to nonribosomal peptides and polyketides (formed by type I PKS) assembled by the previously described multifunctional gigantic enzymes, all secondary metabolites discussed in this section arise from successive reactions catalyzed by mostly monofunctional enzymes.

3.5 Functions of secondary metabolites in the natural context

The exact roles of many if not most secondary metabolites in their original biological context remain unknown. It is likely most confer a selection benefit for the producing organism or serve in their producer's defense. These effects can also be embedded in a symbiotic context. For example, there are reports of *Streptomyces* living in symbiosis with leaf-cutting ants and thus serve to control parasites in the ants' fungus garden. The leaf-cutting ants grow special fungi on harvested leaf material, which serves as their major food source. These fungus gardens can be infected by a specialized pathogenic fungus which is controlled by the ants via symbiosis with antibiotic-producing bacteria. This symbiosis between *Streptomyces* and the ants is highly sophisticated, as the fungicide produced by the bacteria, recently identified as the macrolide candicidin, exhibits strong activity against the pathogenic fungus but does not affect the fungus serving as ants' food source.[99] Secondary metabolites can aid the acquisition of iron from the environment for its producer. These so-called siderophores are small, high-affinity iron chelating compounds secreted into the extracellular compartment and actively taken up again once Fe^{3+}-bound.[100] Others act as virulence factors (rhizoxin,[101] pyocyanin), pigments, or infochemicals. The extent of both inter- and inner-species signaling in bacteria has just recently been revealed. Numerous species of bacteria use elaborate regulatory mechanisms, referred to as quorum sensing (QS), to control the expression of specific genes used in the coordination of activities of a population. QS is based on the production and release of small signaling molecules, autoinducers, that increase in concentration as a function of cell density and activate corresponding transcriptional regulators after a threshold concentration has been reached.[102] Gram-positive bacteria employ mainly small post-translationally modified peptides as autoinducers, whereas Gram-negative bacteria mainly make use of *N*-acyl-L-homoserine lactones (AHLs). These AHL signaling molecules are often produced as mixtures with one species-specific main constituent and variable *N*-acyl chains. By definition, these autoinducers are also regarded as secondary metabolites, although they are not derived from one of the previously discussed biosynthetic machines. AHLs are synthesized by monofunctional enzymes, termed LuxI homologues after the first characterized member of this class. LuxI homologues use

ACP-bound fatty acyl derivatives and SAM as substrates for the production of AHL molecules.[103] Another rather unique QS system, restricted to particular *Pseudomonas* and *Burkholderia* strains, utilises 2-alkyl-4(1H)-quinolone (AQ) autoinducers such as 2-heptyl-3-hydroxy-4(1*H*)-quinolone (the *Pseudomonas* quinolone signal, PQS) and its direct precursor 2-heptyl-4(1*H*)-quinolone (HHQ).[104] The biosynthesis of HHQ is mediated by a four-gene operon (*pqsABCD*) encoding an anthranilate:CoA ligase (*pqsA*) and three β-ketoacyl-acyl carrier protein synthase III (KAS III) homologues.[105] Logically, this biosynthetic operon resembles PKS biosynthesis; however, the molecular details were unknown before this study.

4. Outline of the dissertation

The aim of this thesis was to increase the body of knowledge of bacterial secondary metabolite biosynthesis. In doing so, pathways of different biosynthetic classes as well as tailoring modifications were examined and characterised. A detailed biosynthetic background is the basis for tapping the full potential of the microbial secondary metabolome uncovered by the genomic era. The same applies to the discovery of innovative targets for therapeutic intervention that might be represented by key biosynthetic reactions in the pathway of a secondary metabolite crucial to a pathogen´s virulence. In particular, tubulysin biosynthesis in *Angiococcus disciformis* and *Cystobacter* sp. SBCb004, aurachin and rhizopodin biosynthesis in *Stigmatella aurantiaca* Sg a15 and 2-heptyl-4(1*H*)-quinolone (HHQ) biosynthesis in *Pseudomonas aeruginosa* were investigated.

The existence of the previously postulated biosynthetic precursor pretubulysin, lacking the acetoxy and acyl *N,O*-acetal substituents found in mature tubulysin, was confirmed by chemical synthesis and comparison of the reference material with extracts from the natural producer. Furthermore, the cytotoxic activity of this more synthetically accessible, simplified tubulysin derivative was assessed in **Section B Chapter 1**. An d48, one of our novel isolates classified as *Cystobacter* SBCb004, was also shown to produce tubulysins. While An d48 produces D-type tubulysins containing a phenylalanine building block, SBCb004 produces A-type tubulysins containing instead a tyrosine building block.

In depth analysis of extracts from both strains by high performance liquid chromatography coupled to high resolution mass spectrometry (HPLC-HRMS) and

HRMS/MS was performed to search for new tubulysin derivatives. The identification and partial characterization of a multitude of structural variants of A- and D-type tubulysins along with the comparison of the tubulysin biosynthetic gene clusters from An d48 and SBCb004 is reported in **Section B Chapter 2**. Interstrain comparison enabled assignment of cluster boundaries due to conserved architecture. Interestingly, *orf17* and *orf18*, located downstream of the last PKS-gene *tubF*, encode for patatin-like proteins and are present in both clusters. The involvement of these ORFs in tubulysin biosynthesis was investigated by targeted gene inactivation. The mutants accumulated a thus far unknown tubulysin derivative whose structure was studied by stable isotope labeled precursors feeding and HR-MS/MS experiments in combination with compound isolation and NMR-based structure isolation. Further, the biochemical functions of the patatin-like proteins were analyzed by *in vitro* experiments with heterologously expressed proteins, as described in **Section B Chapter 3**.

Details of the biosynthesis of the aurachin quinoline alkaloids from *Stigmatella aurantiaca* Sg a15 have remained elusive so far. The postulated formulation mechanism of the quinoline core comprises the elongation of activated anthranilate with two malonyl-CoA extender units catalyzed by a type II PKS followed by decarboxylation and heterocycle formation.[86] It is of special interest to understand the channeling mechanisms of non-acetate starter units into polyketide biosynthesis due to the associated potential for metabolic engineering. Therefore, the goal was to identify the molecular mechanisms underlying anthranilate priming in aurachin biosynthesis. *In vitro* experiments with recombinant ACP and aryl:CoA ligase proteins were conducted with the previously identified *auaABCDE* operon. These experiments yielded inconclusive results, promting the exploration of genes in the wider genetic neighbourhood of the *auaABCDE* operon for their involvement in anthranilate priming. The discovery of an additional required functionality and the elucidation of an unprecedented mechanism for starter unit priming in polyketide biosynthesis are presented in **Section B Chapter 4**. Further, the functionalities responsible for the successive transformations of the aurachins were identified and characterized to decipher the complete biosynthetic pathway. These results are described in **Section B Chapter 5**.

The aurachins are structurally related to the members of the 4-hydroxy-2-alkylquinoline family (HAQs) produced by *Pseudomonas aeruginosa*. The *Pseudomonas* quinolone signal (PQS) and its direct precursor, 2-heptyl-4(1*H*)-quinolone (HHQ), are important quorum sensing (QS) signalling molecules in *P. aeruginosa*. Due to the structural similarities between the aurachins and the HAQs, possible analogies between their biosynthetic pathways were

investigated. The essential functionalities for the formation of HHQ (*pqsABCD*) are known,[105] but a detailed biosynthesis mechanism has not been established. Recombinant PqsD protein was used to elucidate its function and to identify its substrates. *P. aeruginosa* is an opportunistic pathogen that causes many nosocomial infections, especially in cystic fibrosis (CF) patients.[106] These infections are difficult to treat due to specific resistances and *P. aeruginosa´*s ability to form biofilms that exhibit a low susceptibility to all types of antibiotics. The formation of biofilms and the production of the bulk of virulence factors excreted by *P. aeruginosa* are controlled by QS, which is based on formation and detection of diffusible signal molecules that accumulate as a function of cell density.[107] Interference with QS has been proven to be a promising target in virulence attenuation. The knowledge gained in this study could be employed to investigate the inhibition of HHQ biosynthesis as a potential target for alternative therapeutic approaches against *P. aeruginosa*. The results of this project are summarized in **Section B Chapter 6**.

The field of microbial natural product research is on the verge of a new era as the increasing availability of genome sequences provides enormous potential for the discovery of new natural compounds by so-called genome mining approaches. These approaches may, in the future, replace isolation campaigns based on physico-chemical properties or the tracking of bioactivity, as they show drawbacks in sensitivity and implementation range. The examination of biosynthetic gene clusters that are easily accessible from sequence information can further prevent rediscovery of known compounds. The recently in-house sequenced genome of the myxobacterium *S. aurantiaca* Sg a15 should be screened *in silico* for unknown biosynthetic gene clusters and hits investigated by an approach based on targeted gene inactivation and comparative analysis of the secondary metabolome between mutants and wild type by HPLC-HRMS and data processing with statistical tools. The results of these efforts that led to the identification and characterization of a thus far unknown *trans*-AT PKS/NRPS gene cluster and the correlation of the respective secondary metabolite are illustrated in **Section B Chapter 7**.

B. Publications

Chapter 1

Pretubulysin, a potent and chemically accessible tubulysin precursor from *Angiococcus disciformis*

Ullrich, A., Chai, Y., Pistorius, D., Elnakady, Y.A., Herrmann, J.E.,
Weissman, K.J., Kazmaier, U., Müller, R. (2009)
Angew. Chem. Int. Ed. Engl., **48,** 4422-4425.

This article is available online at:
http://onlinelibrary.wiley.com/doi/10.1002/anie.200900406/pdf

The supporting information is available online at:
http://onlinelibrary.wiley.com/store/10.1002/anie.200900406/asset/supinfo/anie_200900406_sm_miscellaneous_information.pdf?v=1&s=bbd9072566f798f1a15cba06488bed8bfdef6b71

Chapter 2

Discovery of 23 natural tubulysins from *Angiococcus disciformis* An d48 and *Cystobacter* SBCb004

Chai, Y.*, Pistorius, D.*, Ullrich, A., Weissman, K.J., Kazmaier, U., Müller, R. (2010)
Chem Biol., **17,** 296-309. (* Authors contributed equally to this work.)

This article is available online at:
http://www.sciencedirect.com/science/article/pii/S1074552110000487

The supporting information is available online at:
http://www.sciencedirect.com/science/MiamiMultiMediaURL/B6VRP-4YP794P-F/B6VRP-4YP794P-F-1/6240/html/S1074552110000487/cfddc6092639ffbab9fa4cac5420f698/mmc1.pdf

Chapter 3

New insights into tubulysin biosynthesis – Unprecedented involvement of patatin-like proteins in secondary metabolism

In preparation for submission.

Introduction

The tubulysins are a family of nonribosomally synthesized peptides with promising cytotoxic activity,[1] even against multi-drug-resistant tumors. The cytotoxic effect is mediated by inhibition of tubulin polymerization, inducing a depletion of cell microtubules and finally triggering the apoptotic process.[2] The tubulysin core structure consists of the five amino acids, *N*-methyl pipecolic acid (Mep), isoleucine (Ile), valine, cysteine and either phenylalanine or tyrosine, and two acetate units. The fragment consisting of valine and cysteine connected by one acetate unit has been grouped 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 (**Figure 1**).

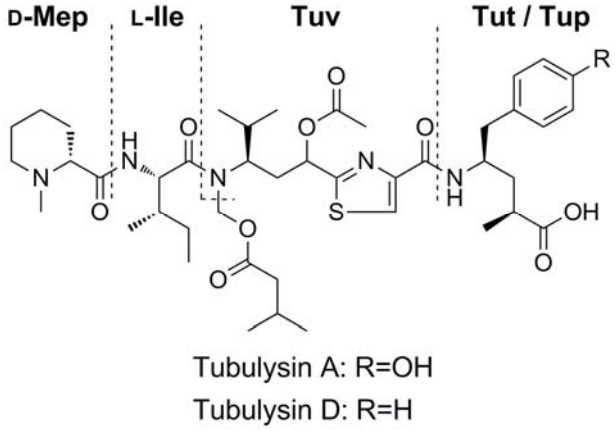


Figure 1. *Chemical structures of tubulysin A and tubulysin D with division into structural subunits*

The core of the corresponding PKS/NRPS hybrid gene cluster for tubulysin biosynthesis was first identified in *Angiococcus disciformis* An d48.[3] Recently, we reported another tubulysin producing strain, *Cystobacter sp.* SBCb004, and compared the tubulysin biosynthetic gene clusters in An d48 and SBCb004. By identification of the gene set common to both strains, it was possible to define the cluster boundaries.[4]

Results and discussion

Conservation of patatin-like proteins

The comparison of the tubulysin biosynthetic gene clusters in An d48 and SBCb004 revealed the presence of two conserved open reading frames (ORFs), *orf17* and *orf18*, downstream of the last PKS gene *tubF* (**Figure 2**). The ORFs further downstream did not show any correlation between the strains. *orf17* and *orf18* both encode patatin-like proteins, which show high mutual sequence homology (in the range of 60% identity and 70% similarity) and suggest evolution by gene duplication.

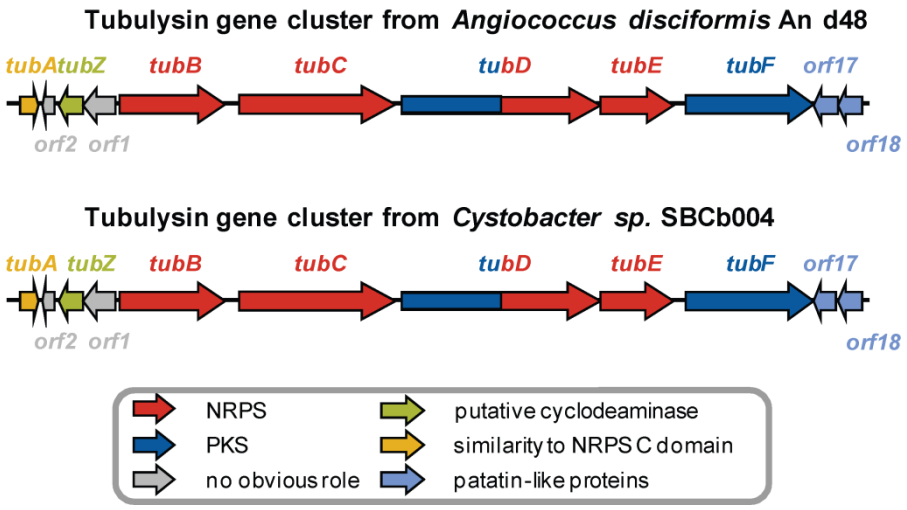


Figure 2. *Comparison of the genetic organizations of the tubulysin biosynthetic gene clusters in* Angiococcus disciformis *and* Cystobacter sp. *SBCb004.*

Patatin-like proteins are known to exhibit lipid acyl hydrolase activity,[5] a function apparently not required in the tubulysin pathway. However, it is worth mentioning that genes encoding patatin-like proteins are also present in the DKxanthene gene clusters of the myxobacteria *Myxococcus xanthus* DK1622 and *Stigmatella aurantiaca* DW 4/3-1,[6] as well as in the leupyrrin gene cluster of *Sorangium cellulosum* So ce690,[7] suggesting that these enzymes may play a unexpected role in natural product biosynthesis.

Characterization of patatin-like protein knockout mutants

To investigate the involvement of the patatin-like proteins in tubulysin biosynthesis, the corresponding genes were inactivated by targeted gene disruption based on homologous recombination. The first mutant obtained and verified was an An d48-*orf18*$^-$ mutant (generated by Yi Chai). HPLC-MS analysis of the extract of this mutant revealed more than a

50-fold decrease in production of tubulysin D, a surprising and extreme result considering patatin-like proteins were not expected to be involved in tubulysin biosynthesis. Detailed analysis of the HPLC-MS data revealed that the An d48-*orf18*$^-$ mutant accumulated an unknown compound not present in wild-type strain extracts. This compound, m/z $[M+H]^+$ = 902.5, was bigger than tubulysin D, with m/z $[M+H]^+$ = 828.5 (**Figure 3**).

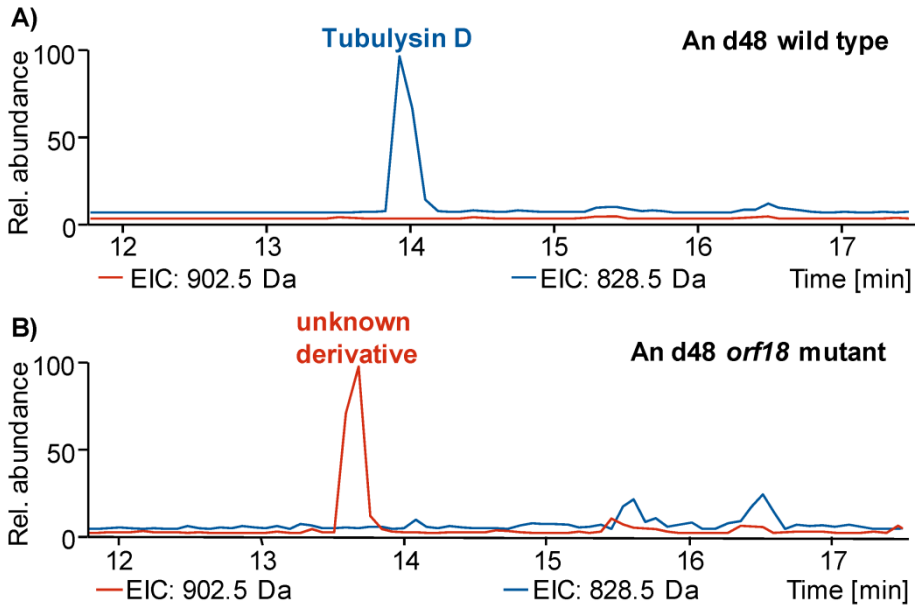


Figure 3. *Extracted ion chromatograms (EICs) for m/z 828.5 (blue) and 902.5 (red) of HPLC-MS analysis of extracts from* ***A****) An d48 wild type and* ***B****) An d48-*orf18$^-$ *mutant.*

Compound characterization by MS experiments

Ensuing HRMS experiments produced a monoisotopic mass of m/z $[M+H]^+$ = 828.4566 for tubulysin D (calculated m/z $[M+H]^+$ = 828.4576). The deduced molecular formula of $C_{43}H_{66}N_5O_9S^+$ matched the reported elemental composition. The monoisotopic mass of the unknown compound was determined to m/z $[M+H]^+$ = 902.4947, leading to the prediction of a molecular formula of $C_{46}H_{72}N_5O_{11}S^+$. Based on similarities between both molecular formulae, it was logical to conclude the unknown compound was a new tubulysin derivative. MS/MS fragmentation based on collision induced dissociation (CID) was employed to investigate this hypothesis. MS/MS data revealed mostly identical fragmentation losses between tubulysin D and the unknown compound (**Figure 4**).

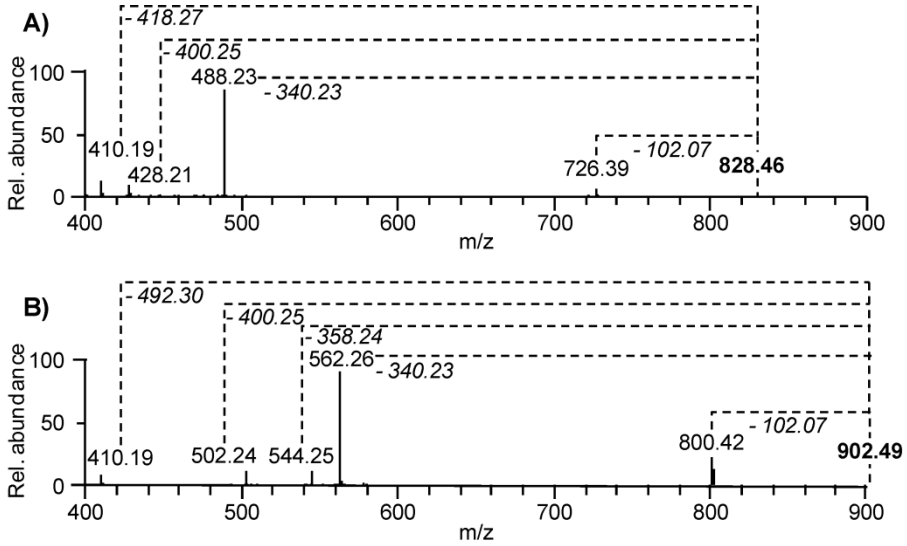


Figure 4. *HRMS/MS spectra of* ***A****) tubulysin D and* ***B****) the unknown compound accumulating in the An d48*-orf18^{-} *mutant. The fragments are connected to the mother ion by dashed lines and the italic numbers in the upper corners of the respective connection indicate the loss in m/z upon fragmentation. For clarity, only the first two decimal places are given.*

Both compounds exhibit common fragmentation losses of 102.07, 340.23 and 400.25 Da. Furthermore, they share a fragment of m/z = 410.19. For the unknown compound, there is another fragment of m/z = 544.25 which, neither itself or the corresponding fragmentation loss of 358.24 Da, is found in the MS/MS-spectra of tubulysin D. Mapping of the most likely fragmentation breaking points of tubulysin D allowed assignment of the four main fragments. These four fragments show a sequential relationship to each other. The fragment of m/z = 726.3891 arises from tubulysin D by elimination of 3-methylbutanoic acid from the bis-acyl *N*,*O*-acetal substituent attached to the tubuvaline moiety. The fragment of m/z = 488.2208 has, in addition, eliminated the *N*-terminal dipeptide consisting of *N*-methyl pipecolic acid (Mep) and isoleucine (Ile). This fragment can further eliminate acetic acid from the acetoxy substituent of tubuvaline to yield the fragment of m/z = 428.2001. The smallest fragment of m/z = 410.1895 arises from additional elimination of water from the *C*-terminal carboxyl group of the tubuphenylalanine moiety (**Figure 5**). Since this fragment was also found in the unknown compound it could be concluded that an additional substituent is linked to the carboxyl group of tubulysin D and, upon elimination of the entire moiety, the common fragment of m/z = 410.1895 is formed.

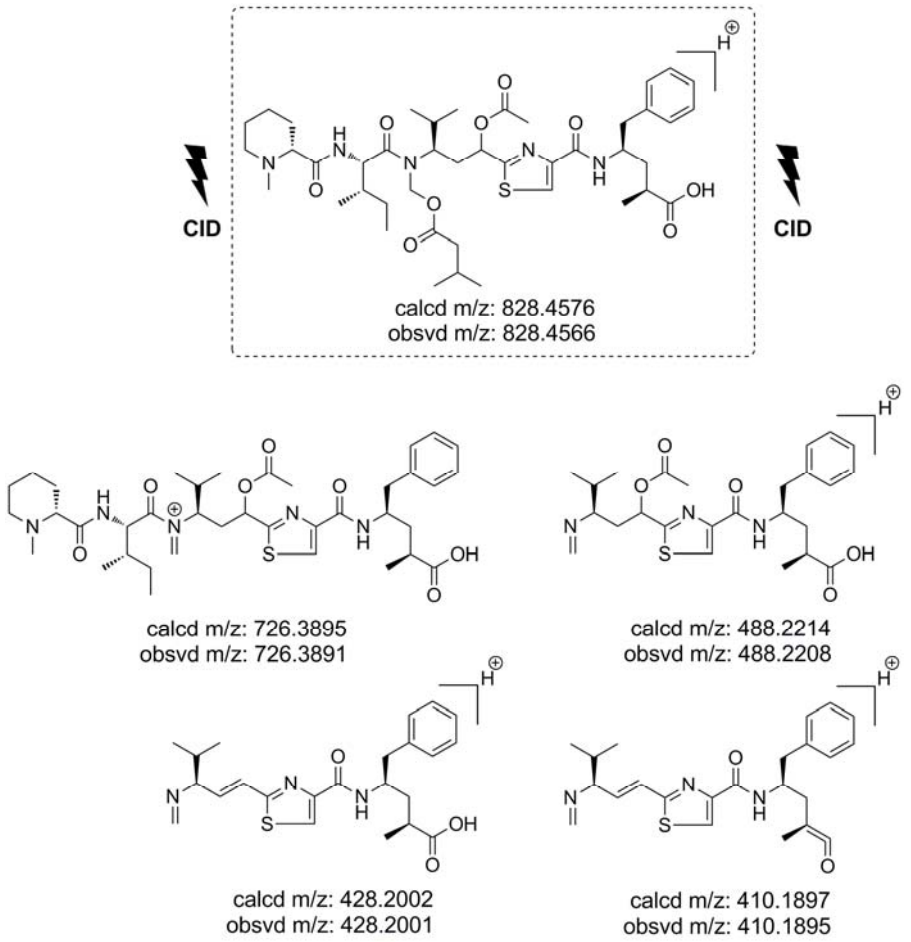


Figure 5. *Postulated structural assignments for CID fragments obtained for tubulysin D. Calculated (calcd) and the observed (obsvd) m/z-values for the charged species are given below the structures.*

Based on a difference to the accurate mass of 74.0371 Da as well as a difference in predicted molecular formula to tubulysin D, the unknown compound was predicted to contain an additional $C_3H_6O_2$. Taking into account the high relative amount of oxygen in $C_3H_6O_2$ and that this moiety would have to be attached in a biosynthetically comprehensible way to the *C*-terminal carboxyl group of tubulysin D greatly reduced the pool of possible structures. The most likely candidate was a glycerol moiety hypothesized to be attached to tubulysin D via ester bond formation. In order to support this hypothesis, the An d48-*orf18*$^-$ mutant was fed with labeled $^{13}C_3$-glycerol (experiment performed by Yi Chai). HPLC-HRMS analysis revealed the incorporation of one glycerol molecule by peak intensities at +3 Da and the successive isotope signals (**Figure 6**).

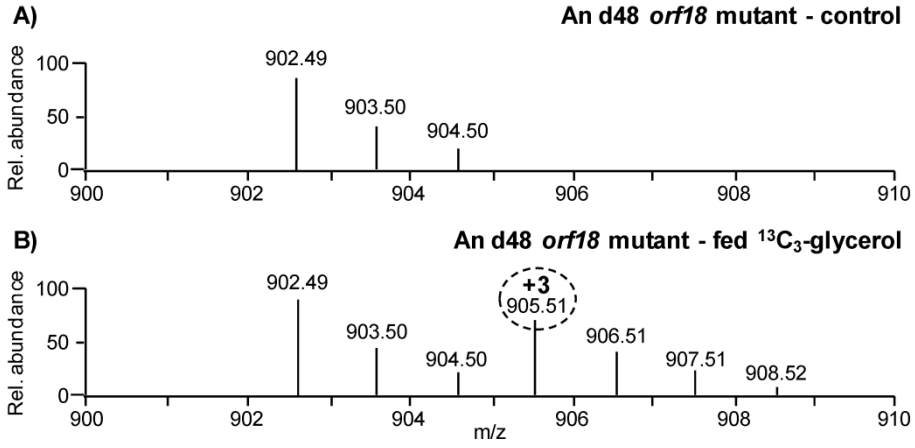


Figure 6. *Isotope patterns of the unknown compound accumulating in An d48-*orf18$^-$ *mutant. Patterns were obtained by HPLC-HRMS measurements of extracts from* ***A****) An d48-*orf18$^-$ *mutant control and* ***B****) An d48-*orf18$^-$ *mutant fed with stable isotope-labeled* $^{13}C_3$*-glycerol. For clarity, only the first two decimal places are given.*

Based on these results, we correlated the MS/MS fragments obtained for the unknown compound to the structural proposal of tubulysin D glycerol ester and were able to rationalize all differences between masses of fragments of tubulysin D and the new compound. The common fragment of m/z = 410.1895 arises, in this case, from the elimination of glycerol instead of water. The structural proposal also explains the presence of a fragment of m/z = 544.2474, found in the new compound as a result of a loss of 3-methylbutanoic acid, the *N*-terminal Ile-Mep dipeptide and one molecule of water from the *C*-terminal glycerol moiety (**Figure 7**).

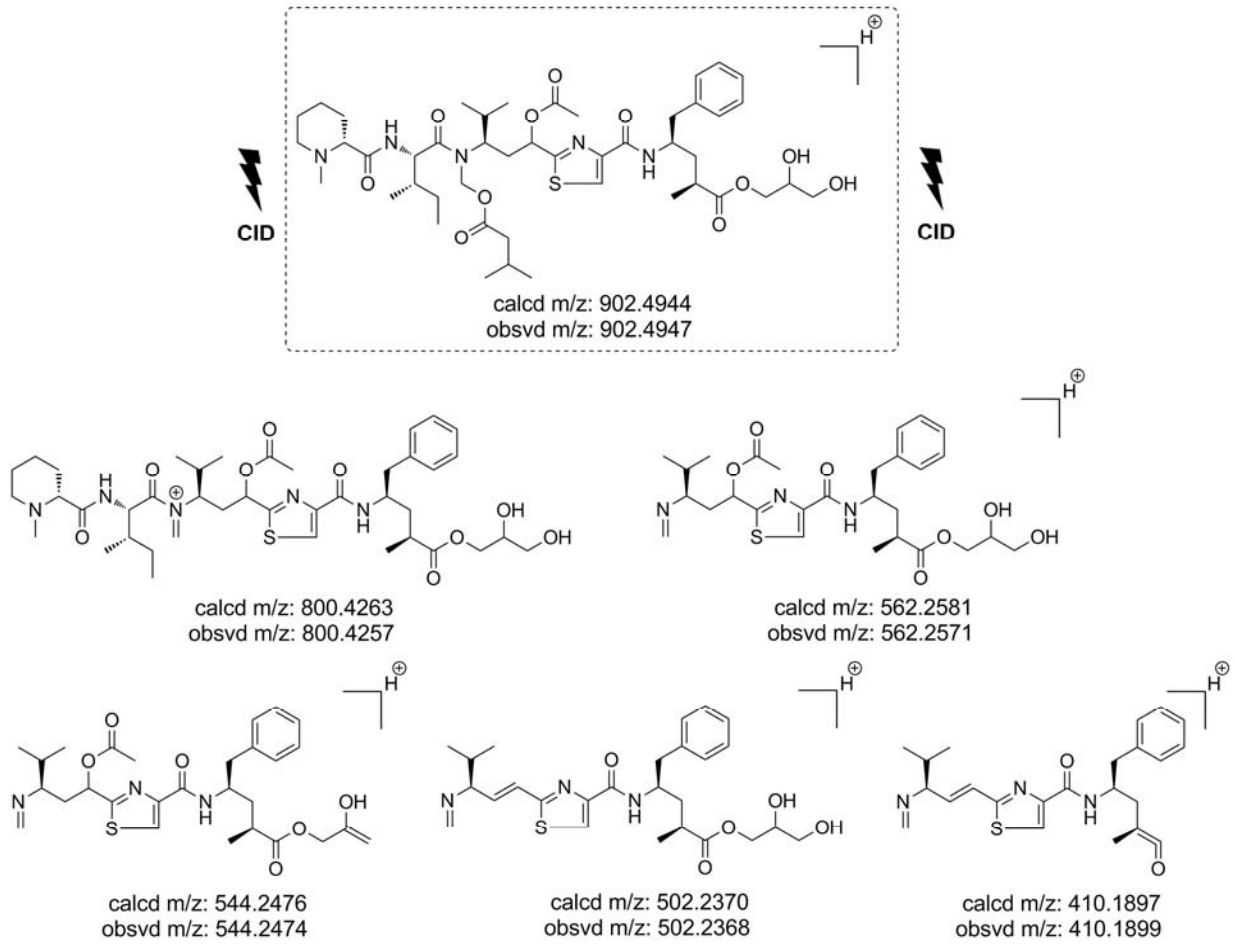


Figure 7. *Postulated structural assignments for the CID fragments obtained for the unknown compound accumulating in the An d48-*orf18^{-} *mutant based on the structural proposal of tubulysin D glycerol ester. Calculated (calcd) and the observed (obsvd) m/z-values for the charged species are given below the structures.*

Compound isolation and structure elucidation by NMR

To confirm the MS-based proposed structure of the tubulysin D glycerol ester, the compound was isolated for more conclusive structural elucidation by nuclear magnetic resonance (NMR) spectroscopy. The purification procedure is described in detail in the experimental section. 0.6 mg of a slightly beige amorphous solid were obtained after final purification and analysed by one-dimensional and two-dimensional NMR experiments. **Table 1** lists the observed ^{1}H NMR signals and their assignments to the respective positions.

Table 1. *^{1}H NMR spectroscopic data from tubulysin D glycerol ester*

Position (H)	δ_H [ppm]	Multiplicity	J_H [Hz]
2	2.48	under DMSO signal	
3	1.86	m	
4	4.21	m	
5	7.97	d	9.1
8	8.18	s	
11	5.74	dd	11.3; 1.6
12	2.41	m	
13	4.20	m	
16	4.40	t	9.1
17	7.93	d	8.8
19	2.47	dd	10.7; 3.2
20a	1.37	m	
20b	1.56	m	
21a	1.15	m	
21b	1.63	m	
22a	1.42	m	
22b	1.51	m	
23a	1.92	m	
23b	2.82	m	
25	2.04	s	
26	1.07	d	7.3
27a	2.80	dd	13.6; 6.3
27b	2.87	dd	13.6; 7.6
29	7.19	AA´	
30	7.24	BB´	
31	7.16	C	
32	7.24	BB´	
33	7.19	AA´	
35	2.10	s	
36	1.81	m	
37	0.67	d	6.6
38	0.98	d	6.3
39a	5.25	d	12.0
39b	6.19	brd	11.7
40	1.91	m	
41	1.47	m	
42	0.81	m	
43	0.81	m	
45a	2.08	m	
45b	2.13	m	
46	1.92	m	
47	0.81	m	
48	0.81	m	
49a	3.87	dd	11.0; 6.3
49b	4.01	dd	11.0; 4.7
50	3.62	m	
51a	3.33	under water signal	
51b	3.67	m	

The observed ^{1}H NMR signals further support the proposed structure and correlate quite well with those reported for tubulysin A.[1] Evaluation of the [^{1}H, ^{1}H]-COSY couplings resulted in the successful assignment of all alkyl- and aryl-moieties present in tubulysin D glycerol ester (**Figure 8**).

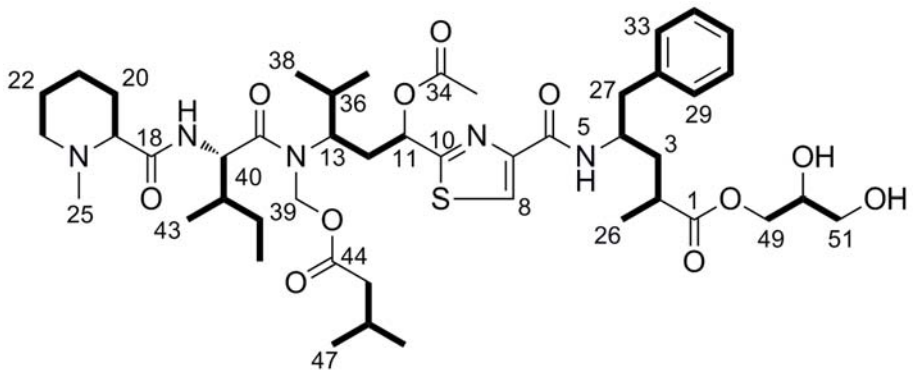


Figure 8.*Chemical structure of tubulysin D glycerol ester with numbering used for ^{1}H NMR assignments and observed [^{1}H, ^{1}H]-COSY couplings indicated by bold bonds.*

Combination of the ^{1}H NMR and [^{1}H, ^{1}H]-COSY data prove the chemical structure of tubulysin D glycerol ester and further suggest linkage of the glycerol moiety to the carboxylic acid of tubulysin D via the primary and not the secondary alcohol function. In the case of ester bond formation with the secondary alcohol, the signals of the protons at the two neighboring primary alcohol functions would not be split due to the fact that they are chemically and magnetically identical; in this case, one duplet with an intensity equaling four protons would be expected. The proton of the carbon carrying the secondary alcohol function should then give a triplet with an intensity of one. However, this situation is not observed in the spectrum. Instead, the signals for the primary alcohol functions were split up. The observation of distinct chemical shifts for the protons of one of the primary alcohol functions (assigned as 49a and 49b) and the most downfield shifted signals for the glycerol moiety indicates its participation in the ester bond. This is further confirmed by reciprocal couplings of the protons 49a and 49b in the [^{1}H, ^{1}H]-COSY experiment. Due to the separated proton pairs 49a and 49b, and 51a and 51b, the signal for the proton at position 50 should theoretically exhibit a multiplicity of dddd with an intensity of one. Although the multiplicity of the signal is not completely resolved, proton 50 is most likely represented by the multiplet at 3.62 ppm. Thereby, the chemical nature of the glycerol ester is resolved.

A compound with a mass (m/z $[M+H]^{+}$ = 760.4) consistent with pretubulysin A glycerol ester was identified in the crude extract obtained from fermentation of the SBCb004 *tubZ*⁻ mutant from which pretubulysin A was purified (described in **Chapter 2**). Analogous HRMS and HRMS/MS experiments as those performed for tubulysin D glycerol ester substantiated this assumption. Since the compound was present in reasonable amounts, it was also purified to provide structural proof by NMR. Successive chromatographic purification steps yielded

7.5 mg pure compound to be analysed by NMR. **Table 2** lists the obtained ^{1}H NMR signals and their assignments to the respective atoms.

Table 2. *^{1}H NMR signals for pretubulysin A glycerol ester*

Position (H)	δ_H [ppm]	Multiplicity	J_H [Hz]
2	2.63	m	
3a	1.73	m	
3b	1.99	m	
4	4.30	m	
8	7.97	s	
11a	2.86	m	
11b	2.96	m	
12a	1.99	m	
12b	2.17	m	
13	-		
16	4.73	d	8.5
19	2.57	m	
20a	1.81	m	
20b	1.91	m	
21	1.52	m	
22a	1.68	m	
22b	1.72	m	
23a	2.62	m	
23b	3.25	m	
25	2.49	s	
26	1.17	d	8.7
27	2.80	m	
29	7.03	d	8.2
30	6.67	d	8.2
32	6.67	d	8.2
33	7.03	d	8.2
36	1.80	m	
373	0.79	d	6.5
38	0.97	d	6.5
39	3.09	s	
40	1.93	m	
41a	1.22	m	
41b	1.54	m	
42	0.91	t	7.4
43	1.00	d	6.8
49a	3.98	dd	6.3
49b	4.19	dd	4.4
50	3.82	m	
51a	3.54	m	
51b	3.88	m	

Evaluation of the ^{1}H NMR data strengthens the structural proposal of a pretubulysin A glycerol ester. The quality of the data is, likely due to residual impurities, inferior to that of pretubulysin A, making assignment of all proton signals very difficult. Most of the signals are in very good accordance with those for pretubulysin A (reported in **Chapter 2**). However, for some parts of the molecule, especially for the piperazin heterocycle, some signals show a strong shift compared to the spectrum of pretubulysin A. Strong overlap with other signals

compounded our elucidation efforts. However, the signals for the glycerol moiety were clearly identified and confirmed the structural proposal of the ester bond between the carboxyl group and one of the primary alcohol functions of glycerol.

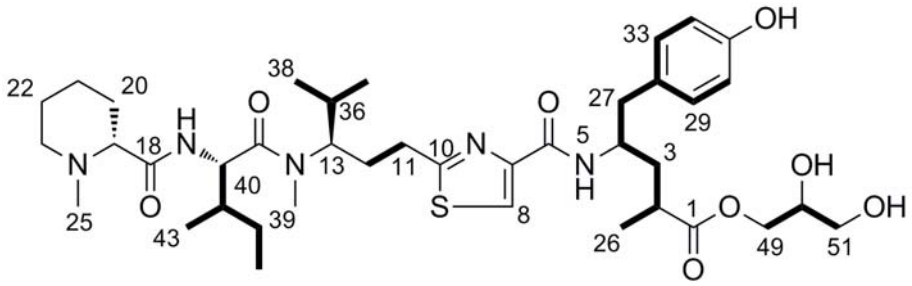


Figure 9. *Chemical structure of pretubulysin A glycerol ester with numbering used for ^{1}H NMR assignments and observed [^{1}H, ^{1}H]-COSY couplings indicated by bold bonds.*

Evaluation of the [^{1}H, ^{1}H]-COSY couplings resulted in the assignment of most alkyl- and aryl-moieties present in pretubulysin A glycerol ester (**Figure 9**).

Characterization of recombinant patatin-like proteins

The fact that tubulysin D glycerol ester accumulates in the An d48-*orf18*$^{-}$ mutant implicates that ORF18 hydrolyzes the ester bond between the *C*-terminal tubulysin carboxyl group and the glycerol moiety. Inactivation of *orf17* in An d48 resulted in a mutant which also accumulated tubulysin D glycerol ester. Inactivations of both genes in SBCb004 yielded mutants with analogous tubulysin A glycerol ester-accumulating phenotypes (experiments performed by Yi Chai, data not shown). To support and further investigate this hypothesis, ORF17 and ORF18 from both An d48 and SBCb004 were heterologously expressed and employed in *in vitro* assays.

Initial expression constructs were designed using a pET-29 b (+) backbone, resulting in heterologous proteins with a *C*-terminal fused hexahistidine tag for convenient purification by Ni^{2+} immobilized metal ion affinity chromatography (IMAC). Expression and purification of ORF18 from both An d48 and SBCb004 was straightforward and yielded reasonably pure preparations (**Figure 9**). However, expression and purification of ORF17 from both strains proved to be very challenging. Even though ORF17 and ORF18 show high mutual homology in both strains, heterologous expression of ORF17 yielded only insoluble protein. To overcome this problem, a number of different affinity tags as glutathione *S*-transferase (GST),[8] maltose-binding protein (MBP),[9] and small ubquitin-like modifier (SUMO)[10] protein were screened in order to improve the yield of soluble protein. Unfortunately, none of

these tags conferred satisfactory improvements concerning the solubility. Therefore, the *C*-terminal hexahistidine tagged version of ORF17 was co-expressed with several chaperone systems.[11-13] Plasmid pGro7 (encoding for *E. coli* GroEs-GroEL) proved to have the best effect on the amount of soluble heterologous protein and was therefore chosen for co-expression. Preparations of ORF17 obtained from co-expression and one-step Ni^{2+}-IMAC purification were still quite heterogeneous but bands corresponding to the respective heterologous proteins could be clearly identified (**Figure 10**) and verified by matrix assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-ToF-MS) of tryptic in-gel digests.

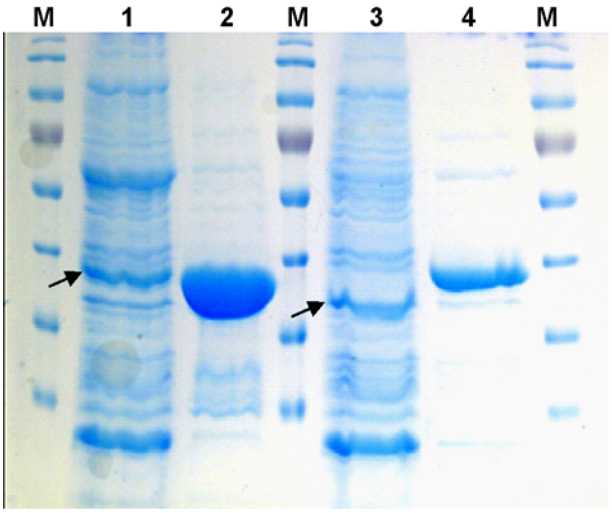


Figure 10. *SDS-PAGE of heterologously expressed patatin-like proteins.* ***1****: An d48 ORF17, indicated by arrow (calc. size 41.5 kDa),* ***2****: An d48 ORF18 (calc. size 40.0 kDa),* ***3****: SBCb004 ORF17, indicated by arrow (calc. size 39.3 kDa),* ***4****: SBCb004 ORF18 (calc. size 41.6 kDa),* ***M****: Marker (sizes descending in kDa: 160 – 110 – 90 – 70 (red) – 55 – 45 – 35 – 25)*

The availability of the recombinant patatin-like proteins as well as their putative substrates tubulysin D glycerol ester and pretubulysin D glycerol ester (synthesized by Angelika Ullrich from AK Kazmaier) allowed for investigation of the proteins' capacities *in vitro*. The patatin-like proteins were incubated with tubulysin glycerol esters in 20 mM Tris-HCl pH 7.5 at 30°C for 1 h and the assays analysed by HPLC-MS. Evaluation of these experiments revealed that, in presence of patatin-like proteins, tubulysin glycerol esters were turned over to "free" tubulysin, as exemplified in tubulysin D glycerol ester incubated with recombinant ORF18 protein from An d48 (**Figure 11**).

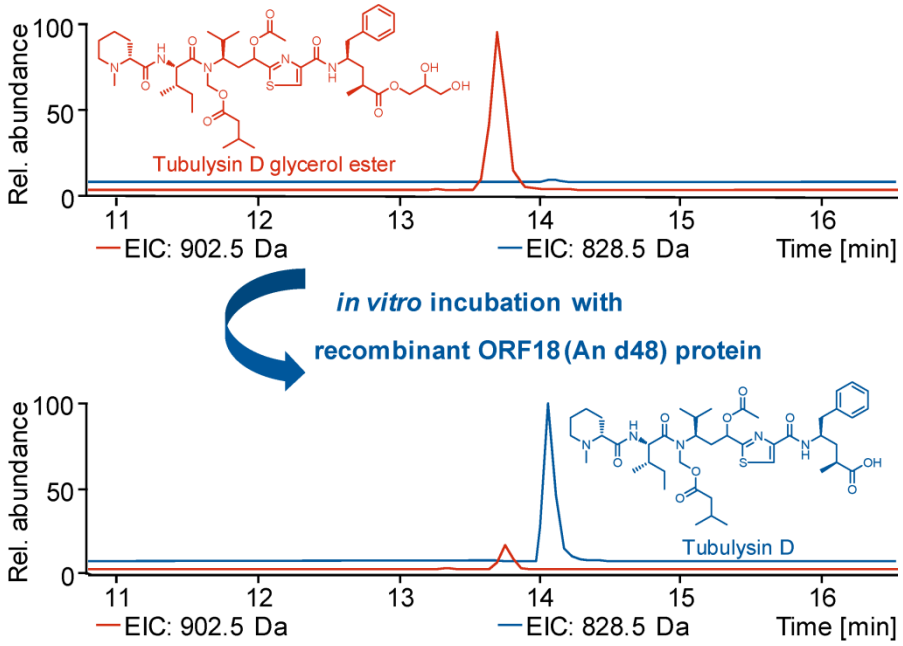


Figure 11. *Representative HPLC-MS analysis (EICs for m/z 828.5 [blue] and 902.5 [red]) of purified tubulysin D glycerol ester before (upper chromatogram) and after (lower chromatogram) incubation with recombinant patatin-like proteins (in this case, ORF18 from An d48).*

These experiments therefore confirmed the postulated function of the patatin-like proteins ORF17 and ORF18, hydrolytic cleavage of the ester bond between the carboxyl group of tubuphenylalanine (or tubutyrosine) and glycerol. This reaction was strictly enzyme-dependent, as negative controls containing either no or heat-inactivated protein did not exhibit any turnover of tubulysin D glycerol ester to tubulysin D, nor did incubation with crude cell extract of the expression strain harbouring the empty expression plasmid. To investigate differences between the patatin-like proteins and their tentative preferences for mature tubulysin D glycerol ester versus pretubulysin D glycerol ester, all possible combinations of recombinant proteins and substrates were tested. The accomplishment of well-executed kinetic studies was hampered by the lack of homogeneous fractions of recombinant ORF17 from both An d48 and SBCb004, hindering the precise measurement of enzyme concentrations, and the limited substrate availability. Therefore equimolar concentrations of recombinant patatin-like proteins (assessed as precisely as possible) were incubated for a defined period of time with one defined concentration of either tubulysin D glycerol ester or pretubulysin D glycerol ester. From these experiments, conversion rates based on the integrals of "free" tubulysins and tubulysin glycerol esters were calculated (listed in **Table 3**).

Table 3. *Conversion rates of substrates by recombinant patatin-like proteins*

Protein	Conversion of pretubulysin glycerol ester	Conversion of tubulysin D glycerol ester
ORF17 (An d48)	54.2 %	96.4 %
ORF18 (An d48)	90.0 %	86.3 %
ORF17 (SBCb004)	57.0 %	89.0 %
ORF18 (SBCb004)	1.1 %	1.7 %

Incubation of pretubulysin D glycerol ester with ORF17 from An d48 resulted in 54.2% cleavage whereas 96.4% tubulysin D glycerol ester was converted. Incubation with ORF18 from An d48 revealed 90.0% pretubulysin D glycerol ester cleavage and 86.3% tubulysin D glycerol ester cleavage. These results suggest that ORF17 from An d48 is acting preferentially on mature tubulysin D glycerol ester whereas ORF18 from An d48 does not appear to discriminate between the two substrates. Preferences for mature tubulysins may be reasonable, assuming the glycerol moiety might fulfil some purpose in post-NRPS/PKS tailoring reactions. To confirm these presumed functions, detailed kinetic studies with pure protein fractions are required. From the assay performed with ORF17 from SBCb004, conversion rates of 57.0% for pretubulysin D glycerol ester and 89.0% for tubulysin D glycerol ester were obtained. These results are in good agreement with those of ORF17 from An d48 and reinforces the hypothesis that tubulysin D glycerol ester is the favoured substrate for ORF17 from both organisms. Surprisingly, incubation of ORF18 from SBCb004 revealed neither turnover for pretubulysin D glycerol ester nor for tubulysin D glycerol ester.

Sequence analysis of the patatin-like proteins

To explain the inactivity of ORF18 from SBCb004, the corresponding patatin-like proteins from both strains were compared and a detailed sequence analysis was performed (**Table 4**).

Table 4. *Results from sequence comparisons among the patatin-like proteins from both strains*

Protein sequences to be compared	Identity [%]	Similarity [%]
ORF17 (An d48) vs. ORF18 (An d48)	78	86
ORF17 (SBCb004) vs. ORF18 (SBCb004)	67	78
ORF17 (An d48) vs. ORF17 (SBCb004)	61	74
ORF18 (An d48) vs. ORF18 (SBCb004)	70	84
ORF17 (An d48) vs. ORF18 (SBCb004)	62	77
ORF18 (An d48) vs. ORF17 (SBCb004)	62	77

Both ORF17 and ORF18 from An d48 have a size of 351 amino acids and show a high mutual sequence homology (78% identity, 86% similarity), suggesting evolution by gene duplication. ORF17 from SBCb004 has a size of 337 amino acids, whereas ORF18, with 368 amino acids, is a little bigger. Their mutual sequence homology among each other (67% identity/78% similarity) is lower than that observed for the patatin-like proteins from An d48. Comparison of the patatin-like proteins from both strains revealed ORF18 from An d48 as the closest homolog to both ORF17 and ORF18 in SBCb004, perhaps implying that *orf17* of An d48 has diverged from the other sequences.

The term "lipolytic enzymes" characterizes enzymes that hydrolyze hydrophobic carboxylic acid esters to release fatty acids and glycerol and these enzymes comprise the families of lipases and esterases. The addition of the family of patatin-like proteins to the lipolytic enzymes has recently been proposed. The potato patatin B_2, after which this family has been named, is the major soluble storage protein in potato tubers and has been shown to exhibit lipid acyl hydrolase activity.[14] Whereas lipases and esterases usually contain a Ser-His-Asp catalytic triad, the active site of patatin B_2 consists of a Ser-Asp catalytic dyad.[15] ORFs coding for patatin-like proteins have been identified in almost half of all sequenced bacterial genomes sequenced and are highly represented in the genomes of pathogens/symbionts, suggesting that they confer an advantage, although their exact function is not known for the majority of cases. Bacterial patatin-like proteins have been shown to contain four conserved domains (blocks I-IV)[5;16]. The first block consists of a Gly rich region with a conserved Arg or Lys, which is assumed to function as an oxyanion hole. The second block neighbours block I (usually 10 to 20 amino acids distance) and includes the hydrolase motif G-X-S-X-G with the putative active-site Ser. Block III features a conserved

Ser with vaguely described function (both roles as hydrogen bond donor and as potential phosphorylation site are conceivable). The highly conserved Pro residues in blocks III and IV may be important for proper conformation of the protein. Block IV contains, in addition, the second member of the catalytic dyad, an Asp residue.

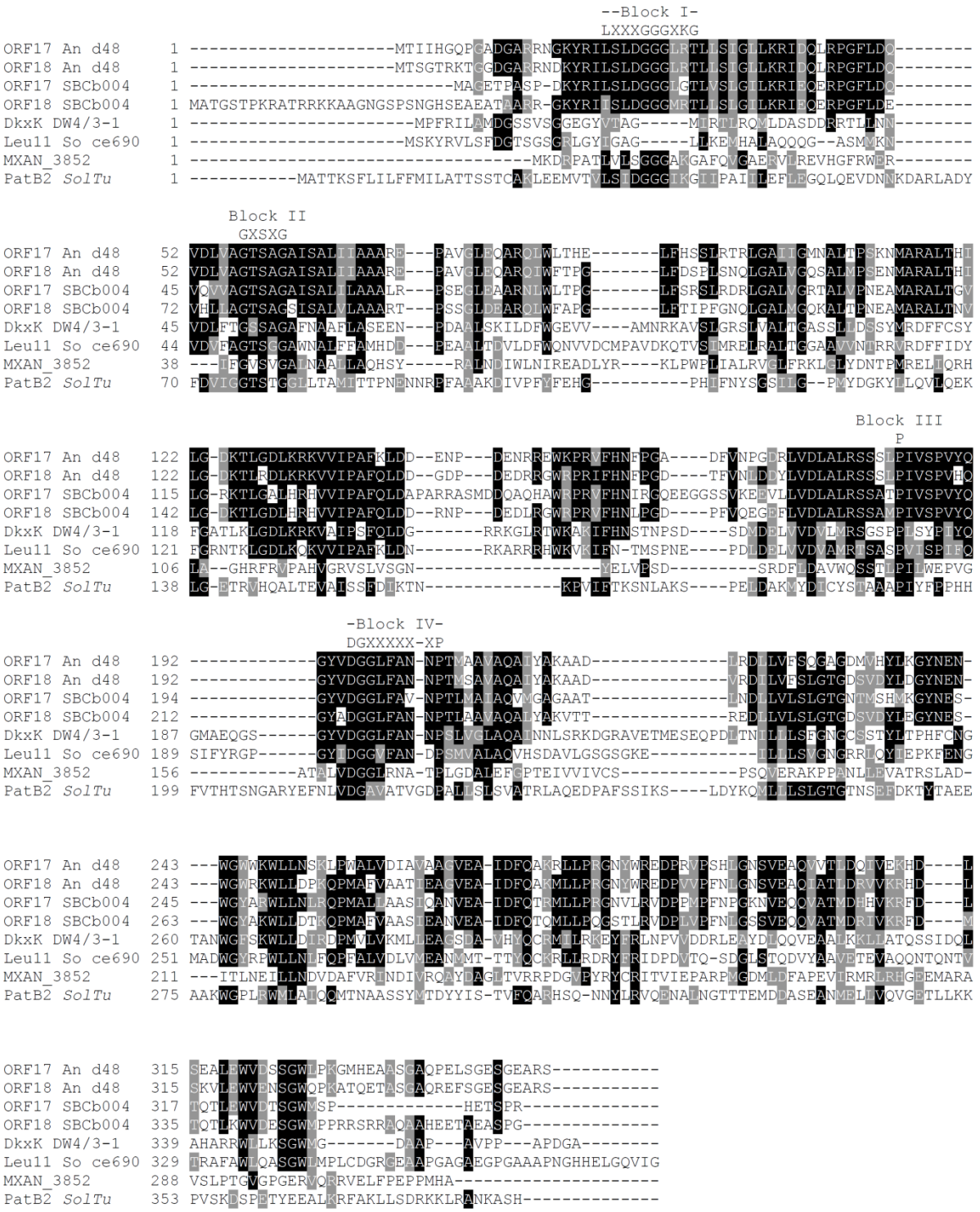


Figure 12. *Sequence alignment of the patatin-like proteins from the tubulysin biosynthetic gene clusters with other myxobacterial patatin-like proteins and patatin B_2 from* Solanum tuberosum*. The alignment*

was performed with ClustalW2 and formatted with Boxshade 3.21. The consensus sequences of the conserved blocks are given on top of the alignment. Abbreviations: DkxK: patatin-like protein associated with DKxanthene biosynthetic gene cluster in Stigmatella aurantiaca *DW4/3-1, accession no. BN001209; Leu 11: patatin-like protein associated with leupyrrin biosynthetic gene cluster in* Sorangium cellulosum *So ce690, accession no. HM639990; MXAN_3852: patatin-like protein from* Myxococcus xanthus *DK1622 involved in development, accession no. ABF93087; PatB2: patatin B2 from* Solanum tuberosum*, accession no. P15477.*

These four blocks are generally well-conserved for the patatin-like proteins ORF17 and ORF18 from both strains (**Figure 12**). Detailed analysis of the most crucial residues within the single blocks revealed perfect conservation of all residues for ORF17 (An d48), ORF18 (An d48) and ORF18 (SBCb004). The conserved Lys residue in block I is uniformly exchanged for an Arg, thereby retaining the positive net charge required for formation of the oxyanion hole. Only ORF17 (SBCb004) is missing this residue (replaced by Gly). Since ORF17 (SBCb004) retains its activity, this exchange is probably compensated by other residues. The region of +12 to -12 aa around the conserved Arg (Lys in patatin B_2 from *Solanum tuberosum*) contains four more basic residues conserved in both ORF17 and ORF18 (from both strains; two Arg and two Lys) that are most likely also involved in the formation of the oxyanion hole and can seemingly retain the functionality of the enzyme. Based on sequence analysis, there are no obvious reasons for the inactivity of ORF18 (SBCb004). All active-site residues are conserved and only two minor sequence variations compared to highly conserved regions in ORF18 (An d48) and both ORF17s (An d48 and SBCb004) are found. At position 82, directly after the G-X-S-X-G motif, the Ala residue common to all the three other proteins is exchanged for Ser. In block IV at position 214, directly preceding the active-site Asp, the consensus Val is substituted by an Ala. However, there is no evident rationale why these substitutions should lead to an inactive enzyme.

The sequences of the proteins DkxK and Leu11, associated with the biosynthetic gene clusters for DKxanthene and leupyrrin respectively, also show good conservation of the four blocks. However, the Gly rich region within block I (a GGG-motif is conserved for all other aligned patatin-like proteins) is not particularly well conserved in both DkxK and Leu11. Their involvement or putative functions in the biosynthesis of DKxanthene or leupyrrin are unknown. A *dkxK* knockout mutant of *Myxococcos xanthus* DK1622 did not reveal any differences regarding DKxanthene production in comparison to the wild type strain nor did this mutant accumulate any compounds with masses corresponding to glycerol esters of DKxanthenes (Thorsten Klefisch, unpublished results).

Sequence analysis of the thioesterase domains

The biochemistry underlying the formation of tubulysin D glycerol ester is still unknown. There are no obvious functionalities encoded in the core gene cluster predicted to catalyze this reaction. Therefore, the thioesterase (TE) domain situated in the last module of TubF was postulated as most likely responsible candidate. TE domains belong to the α/β hydrolase family and contain an active site triad comprised of Ser, Asp, and His. Thioesterases can be divided into two groups: type I TEs exist as integral domains within a PKS or NRPS subunit and catalyze the controlled off-loading of the final product from the assembly line[17], whereas type II TEs are discrete proteins that hydrolyze the acyl thioesters of incorrectly acylated carrier proteins to maintain the functionality of the assembly line;[18] the following discussion focuses on type I TEs. While some of these TEs act as true hydrolases, e. g., those for vancomycin or daunomycin, others catalyze regio- and stereospecific macrolactonization or macrolactamization.[19] In case of tubulysin release from the assembly line, the TE should, rather than employ one molecule of water as nucleophile for hydrolysis of the TE-*O*-acyl intermediate, select one molecule of glycerol. Phylogenetic analyses of thioesterase domains revealed a clustering of the sequences according to the class of secondary metabolites released (either polyketides, nonribosomal polypeptides or polypeptide/polyketide hybrids).[20] To check for obvious special features in the sequence of the tubulysin TEs that might explain for the special postulated reactions catalysed, the sequences of the tubulysin TE domains from An d48 and SBCb004 were aligned with TE domains from other myxobacterial polypeptide/polyketide hybrids (i. e. epothilone, chondramide, DKxanthene, leupyrrin, and myxothiazol; **Figure 13**)

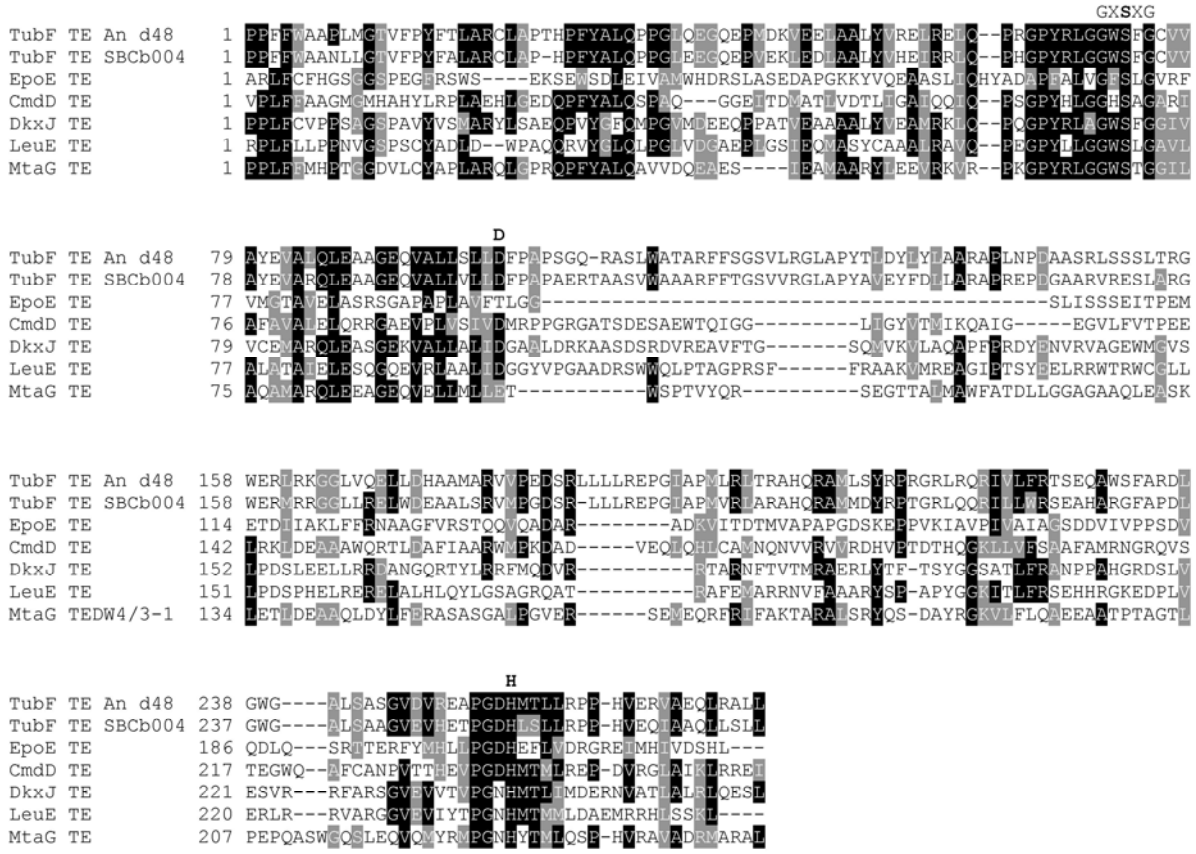


Figure 13. *Sequence alignment of the thioesterase domains from the tubulysin biosynthetic gene clusters and selected thioesterase domains from myxobacterial biosynthetic gene clusters of mixed polypeptide-polyketide secondary metabolites. Domain boundaries were predicted by using the TIGR PKS/NRPS prediction tool (http://nrps.igs.umaryland.edu/nrps/). Alignment was performed with ClustalW2 and formatted with Boxshade 3.21. Active site residues are indicated in bold letters on top of the alignment. Abbreviations: EpoE TE: thioesterase domain from epothilone biosynthetic gene cluster in* Sorangium cellulosum *So ce90, accession no. AAF26925; CmdD TE: thioesterase domain from chondramide biosynthetic gene cluster in* Chondromyces crocatus *Cm c5, accession no. CAJ46692; DkxJ TE: thioesterase domain from DKxanthene biosynthetic gene cluster in* Stigmatella aurantiaca *DW4/3-1, accession no. BN001209; LeuE TE: thioesterase domain fromh leupyrrin biosynthetic gene cluster in* Sorangium cellulosum *So ce690, accession no. HM639990; MtaG TE: thioesterase domain from myxothiazol biosynthetic gene cluster in* Stigmatella aurantiaca *DW4/3-1, accession no. AAF19815.*

The first distinctive feature is the difference in sequence length of the particular TE domains. The epothilone TE exhibits the shortest sequence with 223 aa, while the myxothiazol TE (249 aa), leupyrrin TE (256 aa), chondramide TE (257 aa) and DKxanthene TE (261 aa) are all in the range of 255 aa; the tubulysin TE (An d48) with 275 aa and the tubulysin TE (SBCb004) with 276 aa are around 20 aa residues longer. The *N*-terminal region leading up to the Asp active site residue (~ position 100) shows good mutual agreement for all aligned sequences. The difference in length between the tubulysin and the remaining TE domains is due mainly to aa insertions in the regions directly after the Asp active site residue and halfway between the Asp and the His active site residues. This region between the Asp and the His

active site residues shows lower mutual agreement between all TE domains, whereas the C-terminal block around the His active site residue is more highly conserved. Examination of the crystal structure of the DEBS TE (catalyzing macrolactonization in 6-deoxyerythronolide B biosynthesis) revealed that several residues downstream of the Asp active site residue are involved in the lining of the substrate channel.[21] The additional amino acids inserted in the tubulysin TEs in this region might therefore be involved in formation of a pocket to accommodate the glycerol moiety suspected to replace water as the attacking nucleophile. However, this speculation still requires experimental verification. Initial attempts to heterologously express the tubulysin TE domain for investigation of the postulated capacity of tubulysin D glycerol ester formation *in vitro* have not been successful and require additional work.

Biochemical purpose of the glycerol moiety

The purpose of the attachment of glycerol to tubulysin D and the existence of special enzymes dedicated to the cleavage of this moiety is unclear. One idea was that the tubulysin D glycerol ester might be further modified by esterification of the glycerol moiety with two fatty acid molecules. This tubulysin diacyl glycerol might then be situated in the cell membrane thereby serving as storage for tubulysin. If this was true, this fraction of tubulysin would not be extracted by the standard procedure of stirring with methanol. Therefore, cell mass of An d48-*orf18*$^-$ mutant was first extracted according to the normal procedure (to remove tubulysins present in the cells) and the remaining cell pellet was hydrolyzed, extracted and analyzed. In order to investigate the impact of the hydrolysis on tubulysin D glycerol ester and other tubulysin derivatives (possible breakdown of the backbone), the extract obtained from initial extraction of the An d48-*orf18*$^-$ mutant was treated the same way. Comparison of the hydrolyzed extract with the original extract revealed the disappearance of all glycerol ester derivatives of tubulysins as well as all acylated and *N*-formylated tubulysin derivatives. 12-keto *N*-desmethyl pretubulysin and pretubulysin levels remained unchanged. Instead, a peak with m/z = 672.4, present only at very low levels in the original extract, increased in the hydrolyzed extract. This compound was identified as 11-hydroxy *N*-desmethyl pretubulysin (reported in **Chapter 2**) on the basis of retention time and fragmentation pattern comparisons. It was therefore confirmed that the conditions used for hydrolysis do not cause a breakdown of the tubulysin backbone. Based on these results, there should be an increase of 11-hydroxy *N*-desmethyl pretubulysin in the hydrolysed cell mass sample if the storage theory was

correct. However analysis of the hydrolyzed cell mass did not reveal the presence of 11-hydroxy *N*-desmethyl pretubulysin and therefore contradicts the storage theory, leaving the actual role of the glycerol moiety unknown.

Experimental procedures

Culture conditions

For analysis of secondary metabolite production, *A. discoformis* An d48, An d48-*orf17*$^-$ and An d48-*orf18*$^-$ mutants were grown in 50 mL M7 medium (0.5% probion,0.1% $CaCl_2 \cdot 2H_2O$, 0.1% $MgSO_4 \cdot 7H_2O$, 0.1% yeast extract, 0.5% soluble starch, 1% HEPES (pH 7.4)) at 30°C in presence of 1% adsorber resin (Amberlite XAD-16, Sigma-Aldrich) for 7 days. Mutants An d48-*orf17*$^-$ and An d48-*orf18*$^-$ were grown in the presence of kanamycin (50 µg/mL). *Cystobacter* SBCb004, SBCb004-*orf17*$^-$ and SBCb004-*orf18*$^-$ mutants were cultivated in 50 mL M-medium (1% papaic digest of soybean meal, 1% maltose, 0.1% $CaCl_2 \cdot 2H_2O$, 0.1 % $MgSO_4 \cdot 7H_2O$, 8 mg/L NaFe-EDTA, 1.19% HEPES, adjusted to pH 7.2) at 30°C in presence of 1% adsorber resin for 7 days. Mutants SBCb004-*orf17*$^-$ and SBCb004-*orf18*$^-$ were grown in the presence of kanamycin (100 µg/mL).

Chromatography and mass spectrometry

For analysis of secondary metabolites, the adsorber resin was separated from the medium and extracted with methanol. The extracts were evaporated to dryness and resuspended in 1 mL methanol before analysis by mass spectrometry. Standard analysis of crude extracts was performed on an HPLC-DAD system from the Agilent 1100 series, coupled to a HCTultra ESI-MS ion trap mass spectrometer (Bruker Daltonics) 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 (Phenomenex)), 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 (Thermo-Fisher) 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 (Waters)), 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 (Advion), a chip-based nano-ESI interface.

Isolation and structure elucidation of tubulysin glycerol esters

A fermentation batch culture (8.5 L) of An d48-*orf18*$^{-}$ mutant was grown in M7 medium in the presence of kanamycin (50 µg/mL) and 2% XAD adsorber resin. During cultivation, 10 mL 20% glycerol per day were added to increase the yield of tubulysin D glycerol ester. The production was monitored and the fermenter was harvested after 8 days when it plateaued. XAD and cells were extracted with 3 L of methanol and the solvent was removed by *in vacuo,* yielding 15.3 g of crude extract. Initial purification was performed by flash chromatography on silica gel using a stepwise gradient of nonpolar to polar solvents: hexane, chloroform, chloroform/methanol (3:1, 2:1 and 1:1), and methanol. Fractions containing significant amounts of tubulysin D glycerol ester were combined and evaporated to dryness. Pure tubulysin D glycerol ester was obtained by purification in a preparative HPLC instrument (Waters), equipped with a 2545 binary gradient module, 515 HPLC pump, SFO system fluidics organizer, 2998 PDA, 3100 mass detector, and 2767 sample manager. An X-Bridge RP-C_{18} column (19 × 100 mm, 5 µm, Waters) was used for separation with a solvent system of H_2O (A) and methanol (B), each containing 0.1% formic acid. The following gradient was applied: 0–1 min, 35% B; 1–6 min 35–55% B; 6–10 min 55–70% B. Fractions were collected using mass-guided fraction collection, with a trigger mass of 902.5 Da. This purification procedure yielded 0.6 mg of a slightly beige amorphous solid. The compound was dissolved in DMSO *d*6 and analyzed by one-dimensional and two-dimensional NMR experiments. NMR spectra were recorded on a Bruker Avance 500 spectrometer.

The SBCb004-*tubZ*$^{-}$ mutant was cultivated in M medium in the presence of kanamycin (50 µg/mL) and 2% XAD adsorber resin at 30°C for 6 days. Cultivation was implemented in a 17 L fermenter and shaking flasks (9 L; 150 rpm), for a total volume of 26 L. Fermentation was performed with stirring at 300 rpm, oxygen saturation of 35%, and at pH of 7.2–7.4. XAD was separated from the supernatant and extracted with 6 L of methanol. This procedure yielded 18.4 g crude extract after solvent evaporation. Purification of pretubulysin A glycerol ester was performed analogously to the protocol described above and yielded 7.5 mg pure compound.

Cloning, expression and purification of ORF17 and ORF18

*Orf*17 and *orf*18 were amplified from genomic DNA of An d48 and SBCb004 using Phusion DNA polymerase (Finnzymes) with primers *orf17*_An d48_NdeI_for (5´-CTA GCA TAT GGC CGC AGC TGC GGC ATC T-3´) and *orf17*_An d48_HindIII_rev (5´-CTA GAA GCT TGC TCC GCG CCT CTC CCG ACT-3´), *orf18*_An d48_NdeI_for (5´-CTA GCA TAT GAC TTC GGG CAC TCG GAA A-3´) and *orf18*_An d48_HindIII_rev (5´-CTA GAA GCT TGC TCC GCG CCT CTC CCG ACT-3´), *orf17*_SBCb004_NdeI_for (5´-CTA GCA TAT GGC CGC GGG TCG CCC T-3´) and *orf17*_SBCb004_HindIII_rev (5´-CTA GAA GCT TAC GGG GAG ACG TCT CGT GGG-3´), *orf18*_SBCb004_NdeI_for (5´-CTA GCA TAT GGC CAC CGG TTC GAC GCC CAA-3´) and *orf18*_SBCb004_HindIII_rev (5´-CTA GAA GCT TAC CTG GAG AGG CCT CGG CTG TCT-3´). The resulting PCR products were purified, digested with *Nde*I and *Hind*III, and ligated into *Nde*I/*Hind*III-digested pET-29 b (+), generating pET-29 b (+)-*orf17*-An d48, pET-29 b (+)-*orf18*-An d48, pET-29 b (+)-*orf17*-SBCb004, and pET-29 b (+)-*orf18*-SBCb004. The ligations were electroporated into *E. coli* DH10B and positive clones were selected on LB agar plates containing kanamycin (50 µg/mL). The identity of the plasmids was confirmed by restriction analysis and sequencing. For production of C-terminal His-tagged fusion proteins, *E. coli* BL21(DE3) harbouring the respective expression plasmid was grown at 37°C in 250 mL LB medium containing kanamycin (50 µg/mL) to an OD_{600} of approximately 0.8 units. For protein expression, the culture was induced with 0.2 mM isopropyl β-D-thiogalactopyranoside (IPTG) and incubated at 16°C for 16 h. For co-expression of ORF17 from both strains with GroES-GroEL chaperones from *E. coli*, the plasmid pGro7 was co-transformed into the expression strain and maintained by addition of chloramphenicol (20 µg/mL). The chaperones were induced by addition of L-arabinose to a final concentration of 2 mM 30 min prior to induction of the proteins of interest. The cells were harvested by centrifugation at 7300 g for 12 min at 4°C in a Beckman Coulter Avanti J-E centrifuge. All subsequent steps were carried out at 4°C. The cell pellet was resuspended in 25 mL binding buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 20 mM imidazole, glycerol 10 % (*v/v*)) and lysed with a French pressure cell press. Cell debris was removed by centrifugation at 25000 g for 45 min. The supernatant was filtered through 1.2 µm Acrodisc syringe filters (PALL Life Sciences) before loading onto a 1 mL Hi-Trap His column using the Äkta purifier system (both from GE Healthcare) at a flow rate of 1 mL/min. The unbound sample was washed away with 5 column volumes binding buffer. To elute the His-tagged heterologous proteins, a stepwise gradient with 7%, 16%, 40% and 100% elution buffer

(50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 500 mM imidazole, glycerol 10% (*v/v*)) with 6 column volumes for each step was applied. Buffer in fractions containing heterologous protein were exchanged for 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, glycerol 10% (v/v) using PD10 columns (Amersham Biosciences) and further concentrated using Amicon ultracentrifugal filter devices. Concentrated protein solutions were frozen in liquid nitrogen and stored at -80°C. Protein concentrations were determined with the Bio-Rad protein assay kit according to the manufacturer's instructions using bovine serum albumin (BSA) as a standard (0.5 – 8 µg/mL). Protein fractions were separated on a 15% sodium dodecyl sulfate (SDS)-polyacrylamide gel by electrophoresis. Proteins were visualized in the gel by staining with Coomassie Brilliant Blue R-250. For matrix assisted laser desorption ionisation time-of-flight mass spectrometry (MALDI-ToF-MS), protein bands were excised from the SDS-polyacrylamide gel and in-gel digestion with trypsin was performed. Samples were prepared using Zip-Tips (Milipore).

Characterization of activity of patatin-like proteins

Assays were performed in 50 mM phosphate buffer, pH 7.0 containing one of the recombinant patatin-like proteins (100 nM) and either pretubulysin D glycerol ester or tubulysin D glycerol ester (25 µM) in a total volume of 100 µL. Control incubations were performed with heat-inactivated patatin-like proteins (10 min at 99°C), cell lysate of *E. coli* BL21(DE3) pET-29 b (+), and without addition of enzyme. Reactions were incubated at 30°C for 1 h and quenched by the addition of 100 µL methanol prior to analysis by HPLC-MS according to the conditions stated above.

Hydrolysis of An d48 cell mass

A portion of cell mass from An d48 was extracted once with acetone and twice with methanol. The remaining cell pellet was divided into two equal portions. The first portion was boiled under reflux in KOH (pH 13.5) for two hours, neutralized, freeze-dried, extracted with methanol, filtered, and evaporated to dryness before HPLC-MS analysis. The second portion was extracted a third time with methanol, filtered, evaporated to dryness, and analysed by HPLC-MS. Another portion of the extract obtained from the first extraction was also hydrolysed and processed as described.

References

[1] H. Steinmetz, N. Glaser, E. Herdtweck, F. Sasse, H. Reichenbach, G. Höfle, *Angew.Chem.Int.Ed.* **2004,** *43,* 4888-4892.

[2] M. W. Khalil, F. Sasse, H. Lünsdorf, Y. A. Elnakady, H. Reichenbach, *ChemBioChem* **2006,** *7,* 678-683.

[3] A. Sandmann, F. Sasse, R. Müller, *Chem.Biol.* **2004,** *11,* 1071-1079.

[4] Y. Chai, D. Pistorius, A. Ullrich, K. J. Weissman, U. Kazmaier, R. Müller, *Chem.Biol.* **2010,** *17,* 296-309.

[5] A. Moraleda-Munoz, L. J. Shimkets, *J.Bacteriol.* **2007,** *189,* 3072-3080.

[6] P. Meiser, K. J. Weissman, H. B. Bode, D. Krug, J. S. Dickschat, A. Sandmann, R. Müller, *Chem.Biol.* **2008,** *15,* 771-781.

[7] M. Kopp, H. Irschik, K. Gemperlein, K. Buntin, P. Meiser, K. J. Weissman, H. B. Bode, R. Müller, *Molecular Biosystems* **2011,** *7,* 1549-1563.

[8] J. D. Swaffield, S. A. Johnston, in *Current protocols in molecular biology*Ed.: F. M. Ausubel)**2011**, p. p. 2.1-2.10.

[9] P. Riggs, in *Current protocols in molecular biology*Ed.: F. M. Ausubel)**2001**, p. p. 16.1-16.14.

[10] J. G. Marblestone, S. C. Edavettal, Y. Lim, P. Lim, X. Zuo, T. R. Butt, *Protein Sci.* **2006,** *15,* 182-189.

[11] K. Nishihara, M. Kanemori, M. Kitagawa, H. Yanagi, T. Yura, *Appl.Environ.Microbiol.* **1998,** *64,* 1694-1699.

[12] K. Nishihara, M. Kanemori, H. Yanagi, T. Yura, *Appl.Environ.Microbiol.* **2000,** *66,* 884-889.

[13] A. de Marco, *Nature Protocols* **2007,** *2,* 2632-2639.

[14] D. L. Andrews, B. Beames, M. D. Summers, W. D. Park, *Biochem.J.* **1988,** *252,* 199-206.

[15] T. J. Rydel, J. M. Williams, E. Krieger, F. Moshiri, W. C. Stallings, S. M. Brown, J. C. Pershing, J. P. Purcell, M. F. Alibhai, *Biochemistry-US* **2003,** *42,* 6696-6708.

[16] S. Banerji, A. Flieger, *Microbiology* **2004,** *150,* 522-525.

[17] R. M. Kohli, C. T. Walsh, *Chem.Commun.* **2003,** 297-307.

[18] M. L. Heathcote, J. Staunton, P. F. Leadlay, *Chem.Biol.* **2001,** *8,* 207-220.

[19] T. A. Keating, D. E. Ehmann, R. M. Kohli, C. G. Marshall, J. W. Trauger, C. T. Walsh, *ChemBioChem* **2001,** *2,* 99-107.

[20] K. Buntin, K. J. Weissman, R. Müller, *ChemBioChem* **2010,** *11,* 1137-1146.

[21] S. C. Tsai, L. J. Miercke, J. Krucinski, R. Gokhale, J. C. Chen, P. G. Foster, D. E. Cane, C. Khosla, R. M. Stroud, *P.Natl.Acad.Sci.USA* **2001,** *98,* 14808-14813.

Chapter 4

Unprecedented anthranilate priming involving two enzymes of the acyl adenylating superfamily in aurachin biosynthesis

Pistorius, D., Li, Y., Mann, S., Müller, R. (2011)
J. Am. Chem. Soc., **133,** 12362-12365

This article is available online at:
http://pubs.acs.org/doi/pdf/10.1021/ja203653w

The supporting information is available online at:
http://pubs.acs.org/doi/suppl/10.1021/ja203653w/suppl_file/ja203653w_si_001.pdf

Chapter 5

Completing the puzzle of aurachin biosynthesis in *Stigmatella aurantiaca* Sg a15

Pistorius, D., Li, Y., Sandmann, A., Müller, R. (2011)
Mol. BioSyst., submitted.

This article is available online at:
http://pubs.rsc.org/en/content/articlepdf/2011/mb/c1mb05328k

The supporting information is available online at:
http://www.rsc.org/suppdata/mb/c1/c1mb05328k/c1mb05328k.pdf

Chapter 6

Biosynthesis of 2-alkyl-4(1*H*)-quinolones in *Pseudomonas aeruginosa*: Potential for therapeutic interference with pathogenicity

Pistorius, D., Ullrich, A., Lucas, S.,
Hartmann, R.W., Kazmaier, U., Müller, R. (2011)
ChemBioChem, **12,** 850-853.

This article is available online at:
http://onlinelibrary.wiley.com/doi/10.1002/cbic.201100014/pdf

The supporting information is available online at:
http://onlinelibrary.wiley.com/store/10.1002/cbic.201100014/asset/supinfo/cbic_201100014_sm_miscellaneous_information.pdf?v=1&s=94b74c3e55c45daf71ad4af886d4c2e590a0f368

Chapter 7

Discovery of the rhizopodin biosynthetic gene cluster in *Stigmatella aurantiaca* Sg a15 by genome mining

Pistorius, D., Müller, R. (2011)
ChemBioChem, to be submitted.

Abstract

The field of bacterial natural product research is currently undergoing a paradigm change concerning the discovery of natural products. Previously most isolation efforts were based on isolation of the most abundant compound in an extract or on the tracking of bioactivity. However, traditional activity guided approaches are limited by the available test panels and frequently lead to the rediscovery of already known compounds. The constantly increasing availability of bacterial genome sequences provides the potential for the discovery of a huge number of new natural compounds by *in silico* identification of biosynthetic gene clusters. Examination of the information on the biosynthetic machinery can further prevent rediscovery of known compounds and identify so far unknown biosynthetic pathways of known compounds. By *in silico* screening of the genome of the myxobacterium *Stigmatella aurantiaca* Sg a15 a *trans*-AT polyketide synthase/non-ribosomal peptide synthetase (PKS/NRPS) gene cluster was identified that could not be correlated to any secondary metabolite known to be produced by this strain. Targeted gene inactivation and analysis of extracts from the resulting mutants by high performance liquid chromatography coupled to high resolution mass spectrometry (HPLC-HRMS) in combination with the use of statistical tools resulted in the identification of a compound which was absent in the mutants extracts. By matching with our in house database of myxobacterial secondary metabolites this compound was identified as rhizopodin. A detailed analysis of the rhizopodin biosynthetic machinery is presented in this manuscript.

Introduction

Myxobacteria are soil-dwelling, Gram-negative bacteria which are a rich source of secondary metabolites (SMs) with multifaceted biological activities, such as antibacterial or antifungal mode of action and cytotoxic or cytostatic effects against eukaryotic cell lines.[1] In general one strain produces not only one but several SMs. This is also the case for the subject of this study - *Stigmatella aurantiaca* Sg a15 which is known to produce aurachins,[2] myxochelins,[3] stigmatellins,[4] myxalamids,[5] myxochromides,[6] and DK-xanthenes.[7] Most of these SMs are built up by non-ribosomal peptide synthethases (NRPSs), polyketide synthases (PKSs) or hybrids of both.[8;9] These systems are giant multifunctional enzymes that are composed of modules. Each module is responsible for one cycle of chain extension of the nascent product,

consisting of a set of domains which dictate the chemical functionalities of the incorporated building block. The minimal NRPS module is composed of an adenylation (A), a condensation (C) and a peptidyl carrier protein (PCP) domain. The A domain activates a specific amino acid that is subsequently linked via its carboxyl function to the 4´-phosphopantetheine (Ppant) cofactor of the PCP, whereas the C domain catalyzes peptide bond formation with the growing chain tethered to the upstream PCP. In addition, optional domains that are responsible for methyl transfer (MT), epimerization (E), heterocyclization (HC) and oxidation (Ox) can raise the structural diversity.[10] In analogy to NRPS, the minimal set of domains within a PKS module consists of an acyltransferase (AT) which loads a coenzyme A activated dicarboxylic acid extender unit onto an acyl carrier protein (ACP), where it undergoes condensation with the nascent polyketide intermediate catalyzed by the upstream ketosynthase (KS) domain. The resulting β-keto function can be further processed by a number of optional functionalities, including ketoreductase (KR), dehydratase (DH) and enoylreductase (ER) domains, to yield hydroxyl, enoyl or fully reduced methylene functions in β-position.[11] Besides these "classical" *cis*-AT PKSs, there exist *trans*-AT PKSs which stand out by a lack of integral AT domains which is complemented freestanding AT enzymes.[12]

NRPS and textbook *cis*-PKS show strong correlation in both the number and the order of modules in the megasynthases and the building blocks incorporated into the final product, also known as colinearity rule.[13] The fact that all of the above mentioned domains show significant sequence conservation with retention of active site motifs and that the genes responsible for the biosynthesis of a SM are usually clustered in the genome, allow for straightforward identification of these biosynthetic gene clusters from complete microbial genome sequences by homology searches. The recent increase of available genome sequences from bacterial SM producers, including several myxobacteria, revealed in all cases a higher biosynthetic potential encoded at the DNA-level, compared to the number of known SMs.[14-18] In some cases the gene clusters may be "silent" under the chosen cultivation conditions, or the amounts of the correlated SM are so low that it escapes detection. Besides screening of new bacterial isolates as novel SM producers, the future of bacterial natural product research lies in the exploitation of this untapped biosynthetic potential by making use of the increasing sequence data, in order to find new chemical entities with potentially novel modes of action, and to identify biosynthetic pathways of known secondary metabolites. To attain this goal a multitude of different approaches has been developed in recent years.[19;20]

Here, we present how the application of targeted gene inactivation in combination with sophisticated analytical and statistical tools led to the identification of rhizopodin as another SM produced by Sg a15 and provide a detailed characterization of the corresponding biosynthetic gene cluster.

Results

Identification of rhizopodin from *S. aurantiaca* Sg a15

The draft genome of *S.aurantiaca* Sg a15 has been obtained by 454 sequencing and was subjected to analysis by a genome scanning pipeline for PKS and NRPS biosynthetic gene clusters.[21] The limited average length of contiguous DNA stretches in the draft genome (approximately 7.8 kb) complicated straightforward genome mining of SM biosynthetic gene clusters, which often exhibit sizes between 30 to 80 kb. In addition to the gene clusters related to the production of known SMs from Sg a15, several fragments of gene clusters were found that could not be correlated to known biosynthetic pathways. Among these two fragments of *trans*-AT PKS-NRPS hybrids. Both of them were obviously exceeding the borders of their respective contigs, with one truncated at the 3'-end the other at the 5'-end. However, in both cases, the essential *trans*-AT functionalities were present. In order to correlate these clusters to their SMs, two single-crossover gene disruption mutants,[22] targeting KS domains in both fragments, were generated. The secondary metabolite profiles of resulting mutants were analyzed by high performance liquid chromatography coupled to mass spectrometry (HPLC-MS) and compared to that of the wild type. In depth manual inspection of chromatograms and MS spectra did not reveal any obvious differences. Consequently, we set out to use statistical tools for identifying minor variations in the mutants' metabolic profiles.[23] Mutants and the wild type were grown and worked up under identical conditions, in quintuplicates. The resulting extracts were measured on a HPLC system coupled to an Orbitrap mass spectrometer. In brief, the obtained HPLC-high resolution (HR)MS data were used to generate a molecular feature inventory. Based on the combination of m/z values of molecular features and the respective retention time, so called buckets are created. These buckets present the input values for the statistical evaluation of the data sets. Using the statistical tool of principle component analysis (PCA), implemented in the ProfileAnalysis software (Bruker Daltonics), we were able to identify two hits with masses of 735.4428 and

1469.8783 *m/z* missing in both the inactivation mutants but present in the wild type (**Figure 1B**). Both masses were recorded at identical retention time. Closer investigation of the isotopic patterns revealed that 735.4428 *m/z* is doubly charged and thus equivalent to the singly charged m/z 1469.8783 *m/z*. Computation of the molecular formula based on the neutral mass of 1468.8705 Da resulted in a number of hits due to the high molecular weight. The most likely molecular formulas, based on biosynthetic considerations, were used in combination with the obtained chromatographic and MS data to scan our in-house myxobacterial SM database. Thereby the compound was identified as rhizopodin. This finding was confirmed by comparative HPLC-MS analysis of extracts of Sg a15 wild type with authentic rhizopodin reference.

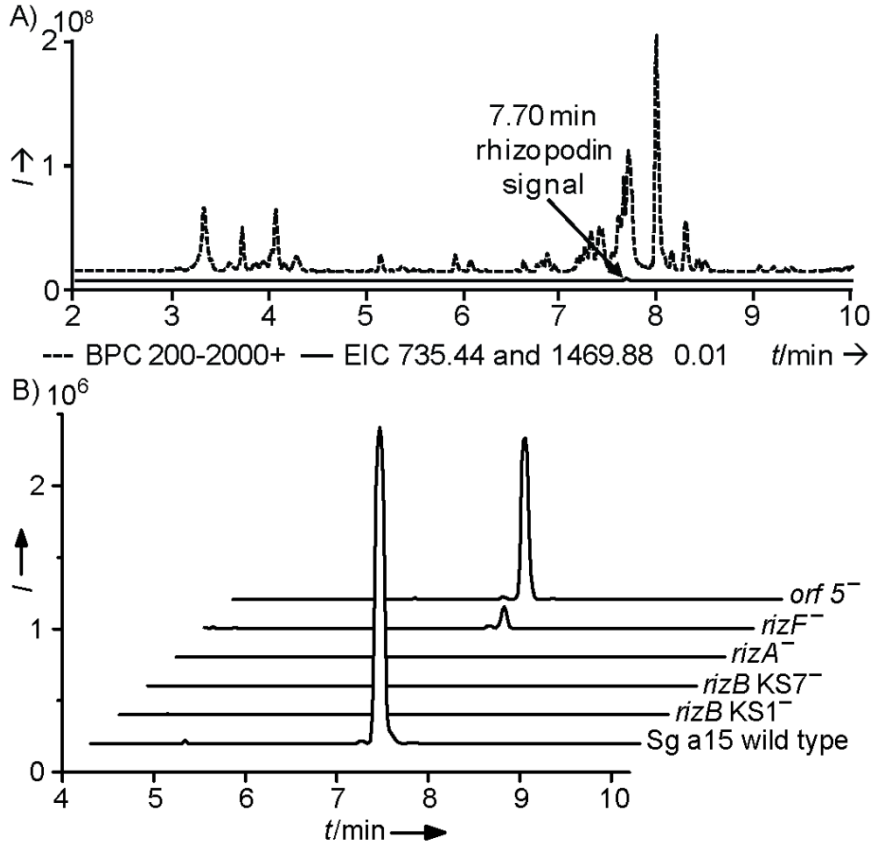


Figure 1. *HPLC-MS chromatograms of Sg a15 extracts.* ***A****) Base peak chromatogram (BPC) from 200-2000 Da in positive ionisation mode (dashed line) and combined extracted ion chromatograms (EIC) of 735.44 and 1469.88 ± 0.01 Da (mass of [rhizopodin + 2H]$^{2+}$ and [rhizopodin + H]$^{+}$, solid line) of extracts from Sg a15 wild type. The rhizopodin signal is highlighted.* ***B****) EICs of 735.44 and 1469.88 ± 0.01 m/z of extracts from Sg a15 wild type, rizB KS1 mutant, rizB KS7 mutant, rizA mutant, rizF mutant and orf5 mutant.*

The very low amounts of rhizopodin produced by Sg a15 (**Figure 1A**) and the fact that it is "hidden" by co-elution with myxalamide B,[5] which is present in the spectra at a 100-fold higher level, have so far prevented the identification of Sg a15 as a producer of rhizopodin

(Figure S1). Rhizopodin is a myxobacterial SM with potent cytostatic activity. It was first isolated and structurally elucidated in the early nineties from *Myxococcus stipitatus*.[24] Fifteen years later it became apparent, during the course of extensive NMR studies,[25;26] that rhizopodin, instead of the originally proposed monolactone, is a C_2-symmetric dilactone. At the same time, the absolute configuration of rhizopodin and its mode of action via inhibition of actin polymerization by dimerization were elucidated by X-ray crystallography.[27] To date, molecular details regarding the biosynthesis of rhizopodin remain unknown.

Sequence analysis of the rhizopodin biosynthetic gene cluster

The results obtained from the inactivation experiments indicate that the two partial DNA sequences coding for *trans*-AT NRPS-PKS hybrids present components of the rhizopodin biosynthetic gene cluster. In order to obtain the whole gene cluster (designated as *riz* cluster), gaps between the two identified contigs were closed by PCR followed by sequencing. Thus 92.2 kb of continuous DNA sequence were obtained that contain 13 genes, among which two encode for hybrid PKS-NRPSs (*rizB* and *rizD*) and two encode PKSs (*rizC* and *rizE*). None of the PKS gene products harbours integral AT domains. Instead, three genes encoding free-standing ATs (*rizA*, *rizF* and *orf5*) were found in the same gene locus. Targeted disruption by insertion mutagenesis of *rizF* located immediately downstream of the last PKS gene (*rizE*) led to an approximately 50-fold decrease in rhizopodin production, indicating that RizF plays a role in the biosynthesis (**Figure 1B**). On the contrary, inactivation of *orf5* located downstream of *rizF* did not significantly affect the production of rhizopodin, and thus allowed to define one of the boundaries of the *riz* cluster. Inactivation of *rizA* completely abolished rhizopodin production. However, a polar effect in this mutant is likely, as *rizA* seems to form an operon with the downstream PKS-NRPS and PKS genes. Regarding the genes (*orf1-4*) upstream of *rizA* no putative functions in the biosynthesis of rhizopodin could be assigned.

The megasynthases RizBCDE encoded in the *riz* cluster are organized into 21 modules (19 PKSs and 2 NRPSs). In order to gain insight into the biosynthetic pathway, all domains were analyzed for the conservation of characteristic motifs and key active site residues. Among the 19 KS domains, seven (KS3, KS7, KS10, KS12, KS14, KS18 and KS20) lack part of or the complete HGTGT motif that is essential for the decarboxylative condensation activity[28] (**Figure 2**).

***Figure 2**. Alignment of the KS domain core regions of KS1-KS20 of RizBCDE. Highly conserved amino acids are highlighted by black background, deviations from the core motifs are marked in bold.*

Therefore the corresponding modules are likely not elongating and may be involved in channeling the polyketide chain to downstream modules as postulated earlier.[29] The presence of 12 active KS domains is in agreement with the retrobiosynthetic analysis that predicts incorporation of 12 acetate-derived C2 building blocks to form one part of the rhizopodin dimer. The *N*-terminus of RizB contains a typical NRPS loading module, composed of an A and a PCP domain together with two additional tailoring domains. The last domain of RizE is a thioesterase (TE) commonly involved in the release of polyketide or polypeptide chains from the megasynthase. These features strongly indicate that RizBCDE are arranged in a colinear manner to the biosynthetic assembly. A domains within the NRPS modules in RizB and RizD show good conservation of the amino acids lining the binding pockets that are predicted to recognize glycine and serine, respectively (**Table 1**).[30]

Table 1. *Comparison of the specificity-conferring code of the A domains in rhizopodin biosynthesis with consensus sequences.*[a]

	Residues									
	235	**236**	**239**	**278**	**299**	**301**	**322**	**330**	**331**	**517**
$Gly_{consensus}$	D	I	L	Q	V	G	L	I	W	K
RizB A_{Load}	D	I	L	Q	C	G	V	I	W	K
$Ser_{consensus}$	D	V	W	H	L	S	L	I	D	K
RizD A11	D	V	W	H	F	S	L	V	D	K

[a] Residue numbering according to the GrsA L-phenylalanine activating A domain.[31]

Analysis of the ten KR domains confirmed in all cases the presence of the Rosmann fold motif GXGXXG and the Lys-Ser-Tyr catalytic triad.[32] The four DH domains show good agreement with the conserved HXXXGXXXXP and DXXX(Q/H) motifs required for activity.[33] The sole ER domain encoded in the rhizopodin gene cluster deviates from the consensus sequence LXHXXXGGVGXXAXXXA in two positions (LIQTAAGGVGLFGMQMA), but the functionality is likely to be retained. The methyl transferase (MT) domains present could be classified into three categories according to their conserved motifs:[34] *N*-MT (in the loading module), *C*-MTs (in modules 2, 5, 8) and *O*-MTs (in modules 3, 7, 14, 18). The NRPS module 11 on RizD contains a tandem HC-domain as distinctive feature. When compared to the eight core motifs of standard HCs (Cy1-Cy7 and C3),[35] both HC domains from the rhizopodin gene cluster showed good agreement for the conserved motifs Cy1, Cy2, Cy6, Cy7 and C3 but only poor to moderate agreement for Cy3, Cy4 and Cy5. The Ox domain downstream of the tandem HCs is in good agreement with the conserved motifs Ox1-3.[36] Based on these results, a detailed model of rhizopodin assembly was proposed (**Table 2** and **Scheme 1**).

Model of rhizopodin biosynthesis

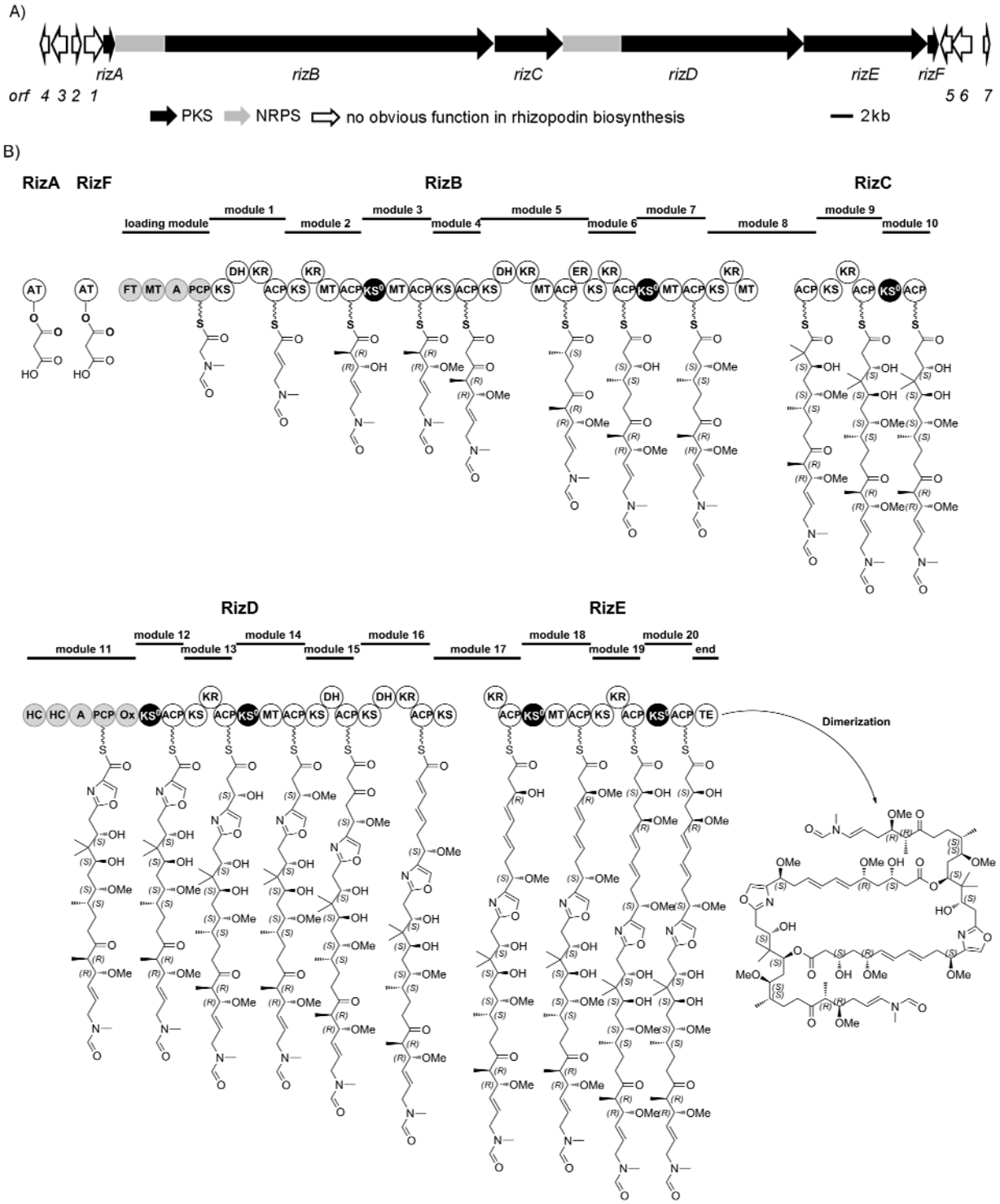
Biosynthesis starts with activation and tethering of glycine to the PCP of the loading module, followed by formylation and *N*-methylation to yield the unusual methylformamide starter unit. The formyltransferase (FT) in the loading module is a rather exceptional feature in NRPS. Examples can be found in the biosynthetic pathways of linear gramicidin,[37] the anabaenopeptilides,[38] and oxazolomycin.[39] The first two rounds of extension are carried out by the successive two PKS modules, whose modular architectures correspond exactly to the degree of β-keto processing, i. e. module 1 generates an acrylyl intermediate while module 2 results in an α-methylated β-hydroxy functionality. Module 3 (KS^0-(*O*-MT)-ACP) does not account for chain extension , due to its inactive KS domain, but mediates *O*-methylation of the β-hydroxy group. Module 4 (KS-ACP) presents the minimal module of *trans*-AT systems and incorporates a malonyl-CoA unit without further reductive processing. Module 5 (KS-DH-KR-(*C*-MT)-ACP-ER) presents an unusual domain organization, with the ER located downstream of the ACP. Nonetheless, the present activities fit with the expected aliphatic α-methylated intermediate. Module 6 (KS-KR-ACP) accounts for a two carbon extension, giving rise to a β-hydroxy intermediate. Module 7 (KS^0-(*O*-MT)-ACP) features a non-elongating KS and thus, like module 3, is solely responsible for *O*-methylation of the β-hydroxyl group. The next module (module 8) is split with KS-KR-(*C*-MT) on RizB and its ACP located on RizC. The *C*-MT is predicted to act twice on the α-carbon thereby installing a dimethyl functionality at thisposition. To the best of our knowledge this unusual course of reaction has only been reported from other *trans*-AT systems (e. g. pederin,[40] onamide,[41] bryostatin,[42] disorazol,[43] oxazolomycin,[39] and elansoloid[44]). For SMs built up by *cis*-AT systems (e. g. epothilone),[45] geminal dimethyl groups are introduced by incorporation of a methylmalonyl-CoA building block followed by a single *C*-methylation. Therefore, the *C*-MTs acting twice seem to be a special feature of *trans*-AT PKS. However, sequence alignment of such *C*-MTs with "classical" *C*-MTs in the *cis*-AT systems did not reveal any obvious differences (data not shown). The assembly continues with one functional PKS module followed by the second NRPS module in the assembly line. The NRPS module (HC-HC-A-PCP-Ox) in RizD attaches a serine to the acyl chain which is subsequently cyclized and oxidized by the concerted action of HC and Ox domains to result in an oxazole moiety. The growing intermediate is then passed through module 12 (KS^0-ACP), extended by module 13 (KS-KR-ACP) and modified by the *O*-MT in module 14 (KS^0-(*O*-MT)-ACP) in analogy to modules 3 and 7. The successive module (module 15) (KS-DH-ACP) lacks the KR

domain needed to act prior to the DH to produce the expected α, β-double bond. Thus, the nascent β-keto group is likely not processed by this module; instead, the acyl chain is transferred to the ACP of module 16 followed by reduction and dehydration to result in the required α,β double bond. Subsequent transfer of this intermediate to KS16 for further extension would be expected. Such reaction sequence has been proposed for the biosynthesis of several myxobacterial SMs including thuggacins,[46] adjudazols,[47] spirangienes,[48] and stigmatellins.[4] However, the detailed mechanisms governing such domains acting out of sequence remain elusive. An alternative scenario including an unknown *trans*-KR that acts in concert with the DH domain in module 15 can currently not be excluded. Module 17 (KS-KR-ACP) is another split module with the KS located on RizD and KR and ACP located on RizE. It catalyzes extension and reduction of the β-keto function to give the expected β-hydroxy intermediate. Module 18 (KS^0-(*O*-MT)-ACP) is the fourth module of this type which yields a non-extended β-methoxy intermediate. The last elongating module (module 19) (KS-KR-ACP) yields the full length monomeric chain of rhizopodin which is passed on by the non-extending module 20 (KS^0-ACP) to the C-terminal TE that is hypothesized to catalyze dimerization by formation of two lactone bonds going along with release of the final product from the assembly line. The presented model for rhizopodin biosynthesis shows good overall agreement with the colinearity rule.

Table 2. *Putative functions and domain organisation of proteins within the rhizopodin biosynthetic gene cluster and the contiguous surrounding region.*

Protein /gene	**Length (bp/aa)**	**Description**	**Modules**	**Domains (coordinates in protein sequence (aa))**
RizA/ *rizA*	855/285	AT	AT	AT (1-285)
RizB/ *rizB*	36942/ 12314	NRPS/PKS	Loading	FT_L (6-342), MT_L(509-778), A_L (905-1415), PCP_L(1420-1504)
			Module 1	KS1 (1561-1996), DH1 (2120-2419), KR1 (2428-2904), ACP1 (2906-2985)
			Module 2	KS2 (3019-3449), KR2 (3624-4110), MT2 (4174-4485), ACP2 (4511-4589)
			Module 3	KS3 (4630-5055), MT3 (5215-5551), ACP3 (5576-5655)
			Module 4	KS4 (5694-6121), ACP4 (6317-6395)
			Module 5	KS5 (6436-6851), DH5 (6969-7269), KR5 (7273-7749), MT5 (7810-8121), ACP5 (8152-8231), ER5 (8276-8594)
			Module 6	KS6 (8626-9052), KR6 (9222-9705), ACP6 (9723-9801)
			Module 7	KS7 (9843-10255), MT7 (10386-10729), ACP7 (10748-10830)
			Module 8	KS8 (10857-11281), KR8 (11445-11923), MT8 (11989-12309)
RizC/ *rizC*	6237/ 2079	PKS	Module 8	ACP8 (24-103)
			Module 9	KS9 (151-577), KR9 (747-1222), ACP9 (1236-1314)
			Module 10	KS10 (1354-1777), ACP10 (1957-2036)
RizD/ *rizD*	23415/ 7805	NRPS/PKS	Module 11	HC11a (37-460), HC11b (488-911), A11 (928-1442), PCP11 (1443-1529), Ox11 (1539-1765)
			Module 12	KS12 (1784-2198), ACP12 (2371-2454)
			Module 13	KS13 (2486-2913), KR13 (3083-3567), ACP13 (3568-3646)
			Module 14	KS14 (3676-4091), MT14 (4226-4559), ACP14 (4579-4658)
			Module 15	KS15 (4684-5109), DH15 (5216-5521), ACP 15 (5546-5624)
			Module 16	KS16 (5669-6101), DH16 (6219-6551), KR16 (6555-7045), ACP16 (7058-7136)
			Module 17	KS17 (7178-7606)
RizE/ *rizE*	11883/ 3961	PKS	Module 17	KR17 (55-540), ACP17 (544-622)
			Module 18	KS18 (665-1097), MT18 (1228-1573), ACP18 (1590-1670)
			Module 19	KS19 (1711-2136), KR19 (2323-2801), ACP19 (2802-2881)
			Module 20	KS20 (2930-3357), ACP20 (3544-3624) TE (3665-3952)
RizF/ *rizF*	861/287	AT	AT	AT (1-287)

orf	**Length (bp/aa)**	**Accession number of closest homologue**	**Similarity/ Identity [%]**	**Putative function/homologue**
1	1764/588	ZP_06890144	44/25	ABC-1 domain protein
2	606/202	ZP_0445617	56/33	hypothetical protein
3	1443/481	CBK82307	64/43	ATPase component of ABC transporter
4	720/240	ZP_07267895	54/32	cobalt transport protein
5	966/322	CAM75127	58/39	Acyltransferase
6	1689/563	ZP_06890144	58/41	ABC-1 domain protein
7	537/179	YP_003953239	96/93	transcription elongation factor GreB



Scheme 1. *Rhizopodin biosynthetic pathway.* ***A****) Organisation of the rhizopodin biosynthetic gene cluster.* ***B****) Model for the biosynthesis of rhizopodin in* S. aurantiaca *Sg a15. NRPS domains are marked in gray, PKS domains in white and non-elongating* KS^0 *domains in black.*

Support of the biosynthetic model by feeding experiments

For experimental support of our biosynthetic model we performed feeding experiments with L-serine-$^{13}C_3$,^{15}N and L-methionine-(*S*-methyl-^{13}C) using liquid cultures. Due to its high molecular weight (elemental composition: $C_{78}H_{124}N_4O_{22}$; monoisotopic mass: 1468.8707), rhizopodin exhibits in MS analysis at least four apparent isotopic signals, with the monoisotopic and the +1 Da signal of equal intensity. Since rhizopodin is composed of two monomers, feeding with L-serine-$^{13}C_3$,^{15}N would result in increase of either +4 Da (single incorporation) and/or +8 Da (double incorporation) in mass. Upon feeding, we observed isotopic signals from +1 to +12 Da with the peak of +4 Da and +8 Da of highest intensity (**Figure 3A**), as expected for incorporation of one or two labeled serine residues. These results confirm that serine is the origin of the oxazole heterocycle in the rhizopodin structure. In the dimeric structure of rhizopodin a total of 18 methyl branches (2x*N*-, 8x*C*-, and 8x*O*-methyl) are to be found. By feeding L-methionine-(*S*-methyl-^{13}C), a maximum increase of +18 Da could be envisioned. However, total labeling is hard to achieve in a complex medium as used in this case to cultivate Sg a15, due to competition with unlabeled methionine. In order to maximize the rate of labeled methionine incorporation into rhizopodin, it was added at high concentration after the culture reached the stationary phase during which secondary metabolites are preferentially produced. Investigation of the mass spectra from extracts of the fed cultures showed labeled rhizopodin with a Gaussian distribution from +10 Da to +19 Da with a maximum at +15 Da (**Figure 3B**), in addition to small amounts of unlabeled rhizopodin. This result proves that all methyl branches in the rhizopodin structure derive from *S*-adenosyl-L-methionine (SAM) by the action of SAM-dependent MTs. Of particular importance, C-methyl branches are not derived from methylmalonyl-CoA, thus pointing out that ATs encoded in the *riz* cluster are specific for malonyl-CoA.

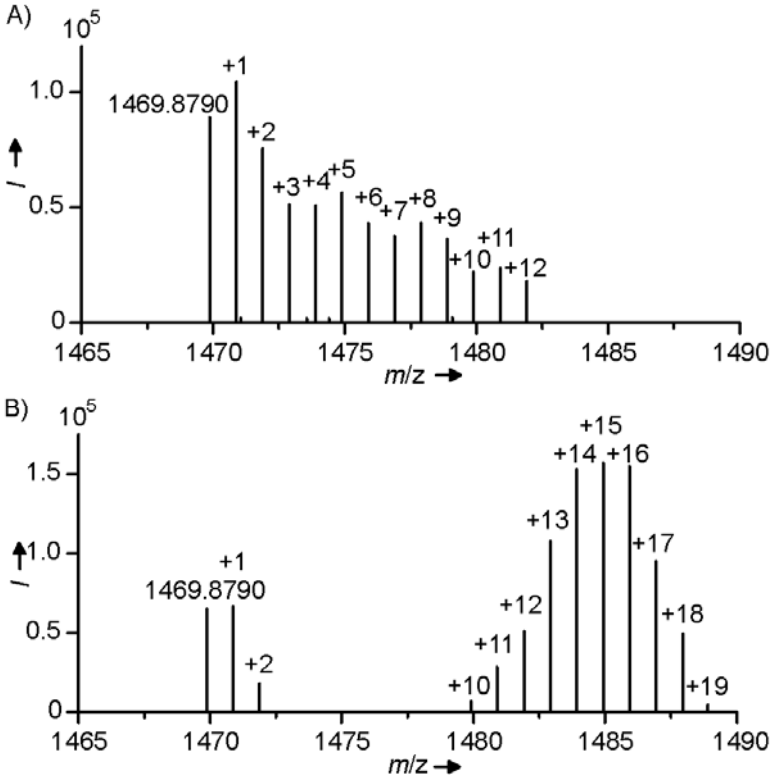


Figure 3. *HR-MS spectra of stable isotope labeled rhizopodin.* ***A****) Feeding of L-serine $^{13}C_3$,^{15}N.* ***B****) Feeding of L-methionine-(methyl-^{13}C).*

Investigations on ACP priming

Both standalone *trans*-ATs, RizA and RizF, show significant similarity to ATs in other *trans*-AT PKS systems (RizA, 51% identity/72% similarity to BryP involved in bryostatin biosynthesis; RizF, 54% identity/70% similarity to BaeE from bacillaene biosynthetic pathway). Their substrate specificity for malonyl-CoA was confirmed by sequence alignments (Table S1).[49] Two or more discrete ATs are less frequent in *trans*-AT PKSs, with six examples identified to date (bacillaene,[50] pederin,[40] etnangien,[51] sorangicin,[49] kirromycin,[52] and elansolid[44]). In some of these cases one of the *trans*-ATs has substrate specificity different from malonyl-CoA (etnangien, sorangicin, and kirromycin), or the AT domain is fused to an ER domain which is able to catalyze enoyl reduction *in trans* (bacillaene,[53] elansolid). Inactivation of *rizF,* which resulted in a 50-fold decrease in rhizopodin production, suggested preferential loading of individual ACPs by one of the ATs, which can be complemented by the other (at least by *rizA*), albeit with lower efficiency. At this stage it is impossible to verify the role of *rizA in vivo*, as insertional mutagenesis is likely to generate a polar effect on the downstream megasynthases and in-frame deletions in Sg a15 have to date not been achieved, despite significant efforts. Therefore, in order to investigate

possible loading preferences of *trans*-ATs for particular ACPs, we expressed both RizA and RizF and several ACPs from the *riz* cluster in *Escherichia coli* as recombinant proteins for *in vitro* loading experiments. Among the selected ACPs, we included two ACPs from non-elongating modules (12 and 18) aiming to investigate their properties concerning malonyl-CoA priming. *Holo*-ACPs were obtained by coexpression with the Ppant-transferase MtaA from the myxothiazol gene cluster[54] and their identities were confirmed by ESI-MS analysis of intact proteins (**Table 3**). Incubation of *holo*-ACP proteins with malonyl-CoA and RizA or RizF resulted in a mass shift of 86 Da, corresponding to loading of a malonyl group. On the contrary, incubation of *holo*-ACPs with malonyl-CoA alone or in combination with cell lysate of *E. coli* BL21 (DE3) carrying the empty expression vector pET28b(+) did not result in any detectable amount of malonyl-ACPs. Both ATs were found to load malonyl-CoA to all tested ACPs with roughly comparable efficiency under the experimental conditions chosen.

Table 3. *Results of HPLC-MS analyses of heterologously expressed* holo-*ACPs from rhizopodin biosynthesis upon incubation with heterologously expressed acetyltransferases RizA and RizF in presence of malonyl-CoA.*[a]

Protein	Down-stream of KS^0	Calc. mass	Obs. mass	RizA malonyl-CoA loading	RizF malonyl-CoA loading
holo-RizACP12	yes	12237	12241	12327(+86)/[87%]	12327(+86)/[54%]
holo-RizACP13	no	11383	11384	11470(+86)/[62%]	11470(+86)/[67%]
holo-RizACP17	no	11648	11650	11736(+86)/[66%]	11736(+86)/[81%]
holo-RizACP18	yes	11676	11692[b]	11778(+86)/[73%]	11778(+86)/[57%]

[a] All given masses in Da. Given in brackets are the percental conversions from *holo*-ACP to malonyl-ACP. [b] Unknown modification; identity of the protein confirmed by MALDI-ToF MS.

The simplified *in vitro* conditions using free-standing ATs and excised ACP domains may not reflect the complex environment within the megasynthase, where multi-domain interactions are involved. Therefore, subtle loading preferences of ATs *in vivo* might be hard to detect *in vitro*. Nevertheless, the observation that ACPs downstream of non-elongating KS^0s can be primed with malonyl-CoA *in vitro* is interesting and suggests that these ACPs are catalytically

active. This finding raises questions regarding mechanisms of chain transfer by non-elongating modules in *trans*-AT PKS systems. If these ACPs are indeed primed with malonyl-CoA *in vivo*, two scenarios would be possible. First, KS^0 would accept the polyketide chain from upstream ACPs and transfer those directly to the next module, a process of module skipping similar to that described in the biosynthesis of myxochromide.[55] The second possibility is that a type II thioesterase is responsible for unloading the misprimed malonyl group from the ACP;[56] yet no such enzyme was identified in the *riz* cluster. It is also possible, that these ACPs are not primed *in vivo* due to constraints imposed by the modular context which prevents their accessibility to the *trans*-ATs. Thus, the KS^0 would channel the acyl intermediate to the next module via its downstream ACP. Future investigations are required to verify these hypotheses.

Discussion

Assignment of orphan biosynthetic pathways

At present, the identification and correlation of SMs from so far unassigned biosynthetic gene clusters represent a significant bottleneck that constrains the exploitation of the genetically encrypted bacterial secondary metabolome. Different approaches based on detailed sequence information have been developed in order to resolve this problem. The identification of a SM by feeding of ^{15}N labelled amino acid building blocks to a culture coupled to ^{1}H-^{15}N HMBC NMR guided fractionation requires only the prediction of the pathway of interest´s A domain specificities, which can be easily extracted from the sequence information.[57] However, the restriction to NRPS containing pathways and the drawbacks of ^{1}H-^{15}N HMBC NMR guided fractionation (compound must be produced in relatively high amounts) along with the fact that different pathways present in the strain might incorporate the same aa building block, are serious drawbacks of this approach. The proteomics based identification of active biosynthetic machineries is helpful to predict the actual presence of the unknown secondary metabolite under the given cultivation conditions,[58] but the tedious workflow and a lack of sensitivity are major hindrances of this approach. The generation of a strain´s in-depth metabolic profile based on HPLC-HRMS data combined with a compound identification algorithm allows for statistics aided comparison with metabolic profiles of closely related strains. This comprehensive secondary metabolomic approach proved to be a

very powerful tool for the identification of nonubiquitous SM candidate compounds.[23;59] Further, this approach can also be used to compare the metabolic profile of a wild type strain to that of a mutant carrying an inactivation of an unknown biosynthetic gene cluster to identify the correlated SM. In this study it was instrumental for the identification of rhizopodin (at very low production levels) and the correlation to its biosynthetic gene cluster in *S. aurantiaca* Sg a15. The biggest drawback of this approach is that it requires the feasibility of genetic manipulation of the strain of interest.

Phylogenetic analysis of the rhizopodin cluster

Rhizopodin joins the family of SMs that are built up by *trans*-AT PKS systems, which currently contains 23 members since the first discovery of such a system from *Bacillus subtilis* in 1993.[60] The *trans*-AT biosyntheses have recently been reviewed by Piel .[61] *Trans*-AT PKSs show a high degree of divergence from textbook colinearity rules established for *cis*-AT systems. Modular architectures with missing or superfluous domains or unusual domain orders are common in *trans*-AT PKSs, thus making correlation of PKS function to product structure a challenge.[51;62;63] Recently, an innovative approach based on phylogenetic clustering of KS sequences in relationship to their substrate specificity was developed and allowed prediction of core structures from *trans*-AT systems with a significantly improved reliability.[49;64] For further validation of the proposed biosynthetic model for rhizopodin, we performed a phylogenetic analysis of the19 KS-domains in the *riz* gene cluster with a large subset of KS sequences used in the initial publication.[64] The grouping of the rhizopodin KS domains in the respective clades of the resulting tree (Figure S2) has been evaluated and deduced substrate predictions have been checked for conformity with the above presented biosynthetic model. Indeed, except for KS1, KS2 and KS5, predicted substrate specificities of the remaining KSs matched in general the ones postulated in the model. For example, the non-extending KS12 subclusters within clade X with KS10 from the chivosazol, KS1 from the leinamycin and KS9 from the disorazol biosynthetic gene clusters, all of which are non-elongating and accept β-oxazole or β-thiazole intermediates. Similarly, non-extending KS3, KS7, KS14 and KS18 that channel β-hydroxylated acyl chains to downstream domains, form a subclade within clade X. Four *riz* KSs (KS4, KS8, KS15 and KS19) form a new subclade within clade I (**Figure 4**), reflecting their specificity for β-methoxylated intermediates which are not found in any other *trans*-AT biosynthetic machineries so far.

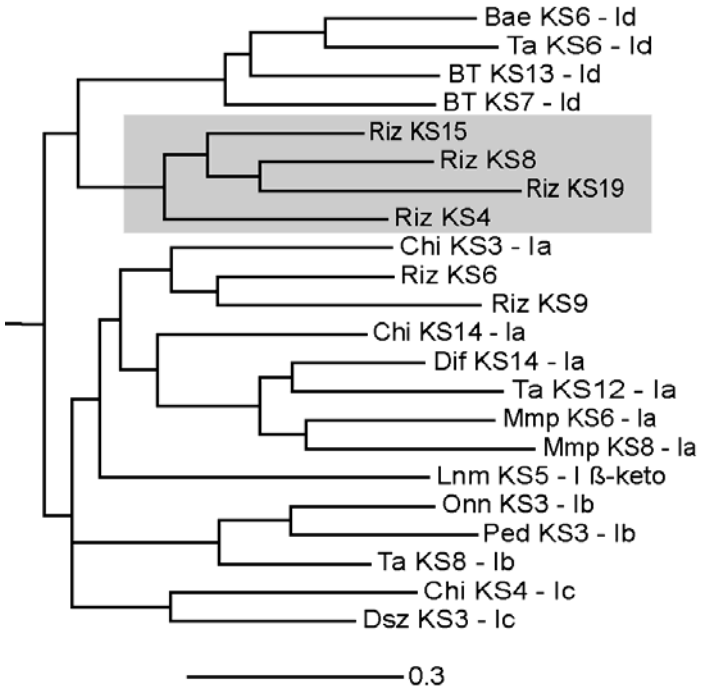


Figure 4. *Resolved tree of Bayesian cladogram of 116 full length KS domains from* trans-*AT PKSs (for the complete cladogram see Figure S2) showing clade I which accepts α-branched or β-methylene and β-branched substrates.*[64] *Riz KS4, Riz KS8, Riz KS15 and Riz KS19 (highlighted in gray) which all accept β-methoxylated substrates form a subclade within clade I. Abbreviations for members of this clade: Bae, bacillaene; BT, thailandamide; Chi, chivosazol; Dif, difficidin; Dsz, disorazol; Lnm, leinamycin; Mmp, mupirocin; Onn, onnamide; Ped, pederin; Riz, rhizopodin; TA, myxovirescin.*

Besides phylogenetic analysis of KS domains, prediction of product stereochemistry by each KR domain according to the conserved motifs[32;65;66] is in full agreement with the actual rhizopodin structure (Figure S3), lending further weight to our proposed biosynthetic model.

Introduction of methoxy groups

The biosynthetic logic for the synthesis of methoxy groups in rhizopodin is special among *trans*-AT systems. 11 core structures (myxovirescin, pederin, onnamide, bryostatin, disorazol, rhizoxin, chivosazol, oxazolomycin, etnangien, kirromycin and thailandamide) out of the 23 reported metabolites produced by *trans*-AT PKS systems, incorporate between one and three methoxy groups. For all of these SMs, except for etnangien and thailandamide, the methoxy groups are introduced into the molecules by free-standing *O*-MTs in post-assembly line tailoring reactions. According to our analysis rhizopodin biosynthesis deviates from this logic by the employment of non-extending modules of KS^0-(*O*-MT)-ACP architecture, which are solely dedicated to methylation of the β-hydroxyl group. A similar organization is found in module 12 of the etnangien assembly line (KS-(*O*-MT)-ACP). Although the KS domain in this module retains the conserved residues required for activity, no elongation step would be

expected.[51] In thailandamide biosynthesis TaiN ((*C*-MT)-ACP-KS^0-(*O*-MT)-ACP-TE) exhibits a module of identical architecture. However, the *O*-MT is speculated to attack, instead of a β-hydroxyl group, a β-keto-function to form an enol ether which will be subsequently reduced.[61]. Interestingly, the KS^0 in this module (BT KS16) clusters with Riz KS3, KS7, KS14 and KS18 (Figure S2). The putative benefit of using embedded *O*-MTs in rhizopodin biosynthesis is that the final product is uniformly methylated at all eight relevant positions. This would avoid production of partially desmethylated derivatives which is observed for many SM with methoxy groups that are introduced by freestanding MTs. Indeed, no partially desmethylated rhizopodin derivatives have been observed in the extract of Sg a15.

Phylogenetic and sequence analysis of tandem HC-domains

Interestingly, in every reported case where SMs are build up by *trans*-AT PKS/NRPS hybrid systems and contain oxazoline/oxazole or thiazoline/thiazole rings the heterocycles are introduced by tandem HC domains (with oxazolomycin as outstanding exception employing no HC domain at all). Examples include leinamycin,[67] disorazol,[43] rhizoxin,[68] chivosazol,[69] and virginamycin.[70] In NRPS or *cis*-AT PKS-NRPS systems a single HC domain is involved in heterocyclization of a serine or cysteine residue tethered on a PCP such as in tubulysin,[71] epothilone,[72] and ajudazol biosynthesis.[73] However, two rare incidences of tandem HC-domains exist for biosynthesis of the siderophores vibriobactin[74] and anguibactin.[75] Biochemical characterization of VibF *in vitro* revealed that HC2 is responsible for the condensation reaction, whereas HC1 catalyzes the heterocycle formation.[76] In order to predict functions of both HC domains in the rhizopodin assembly line, phylogenetic analysis of several HCs domains (tandem and single) were conducted (**Figure 5**).

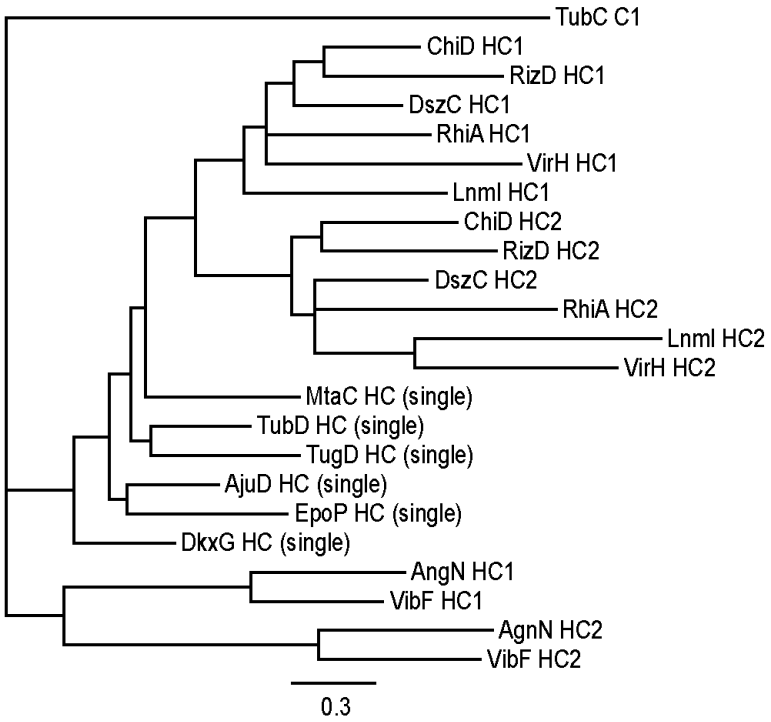


Figure 5. *Bayesian cladogram of tandem HC domains from* trans-*AT PKS-NPRSs, single HC domains from* cis-*PKS-NRPSs, and tandem HC domains found in nonlinear siderophore NRPSs. For tandem HC domains the N-terminal HC domain is labeled HC1 and the following domain is abbreviated as HC2. Abbreviations for members of this tree: Aju, ajudazol; Ang, anguibactin; Chi, chivosazol; Dkx, DKxanthene; Dsz, disorazol; Epo, epothilone; Lnm, leinamycin; Mta, myxothiazol; Rhi, rhizoxin; Riz, rhizopodin; Tub, tubulysin; Tug, thuggacin; Vib, vibriobactin; Vir, virginiamycin.*

This analysis shows clearly that HC1s and HC2s are grouped separately, indicating that tandem HCs in *trans*-AT systems function in a similar manner as in VibF. Intensive sequence analysis of both groups of HCs was performed aiming to identify possible conserved residues that could explain their disctinct functions (condensation or heterocyclization). This revealed certain sequence differences between the members of the HC1s and HC2s (for additional discussion and alignments see Figure S4). Interestingly, the tandem-HCs from the *trans*-AT systems cluster tighter with the "single"-HCs than with the tandem-HCs from vibriobactin and anguibactin. Due to the observed phylogenetic distance between the *trans*-AT tandem HCs and the tandem HCs from vibriobactin and anguibactin biosynthesis it seems likely that they resulted from convergent evolution.

TE-mediated dimerization

Dimeric (or even trimeric) products are very rare among bacterial secondary metabolites, including gramicidin S, enterobactin, thiocoraline (NRPS),[77] disorazol (NRPS-PKS-hybrid), the marinomycins and vermiculine (PKS; biosyntheses not yet elucidated).[78;79] TE-mediated

formation of the dimeric final product has been demonstrated *in vitro* for gramicidin,[80] enterobactin,[81] and thiocoraline.[82] In analogy to these biosyntheses, the following model can be proposed for rhizopodin; The first monomer is transferred to the TE active site serine followed by nucleophilic attack of this acyl-*O*-TE intermediate by the hydroxyl group at C-18 of the second monomer which is tethered to the upstream ACP, yielding a linear dimeric intermediate. The linear dimer would then be transferred to the TE active site, where intramolecular cyclization and product release by formation of a dilactone are catalyzed. Studies on TEs involved in the biosyntheses of gramicidin S and enterobactin showed that there are no obvious sequence differences accounting for their divergent catalytic capabilities. However, it was observed that the linker region between the ACP and the TE was longer than usual in the case of GrsB (41 aa compared to ~15-20 aa).[80] A related long linker is conserved in the rhizopodin and disorazol assembly line (36 aa in RizE/~60 aa in DszC). Furthermore, in both cases a non-elongating KS^0 domain is found upstream of the terminal ACP which may also play a role in the process of dilactone formation. A panel of disorazol derivatives differing in the size of dilactone macrocycles have been identified, indicating that the DszTE exhibits relaxed substrate specificity for dimerization (Buntin, K. and R.M., unpublished results).[83] In the case of rhizopodin biosynthesis, no derivatives of varying macrocycle size have been identified to date. However, in-depth HPLC-MS analysis of extracts from the more prolific rhizopodin producer *Myxococcus stipitatus* Mx s11 (approximately 100-fold higher production compared to Sg a15) revealed the presence of a second peak at a different retention time (7.57 min) of identical mass as rhizopodin (7.70 min), which is upon thorough inspection also seen in the chromatograms from Sg a15 (**Figure 1**; for chromatogram of Mx s11 see Figure S5A). Both compounds show the characteristic +1 and +2 charge states, which led us to the assumption that the earlier eluting compound might correspond to a rhizopodin isomer with lactone bonds between the carboxylic groups at C-1 and C-1' and the hydroxyl groups at C-16 and C-16’ instead of those at C-18 and C-18', resulting in a 34- instead of a 38-membered dilactone (Figure S5B). The minute amount of this compound prevented further isolation and structural characterization. Neither was it possible to confirm the size of the macrocycle by MS/MS fragmentation experiments. The high degree of methylation on hydroxyl groups in rhizopodin may have emerged in order to constrain the formation of more isomers differing in size of the macrocycle, a feature obviously crucial for biological activity.

Conclusion

The identification of *S. aurantiaca* Sg a15 as producer of rhizopodin by means of genome mining in combination with sophisticated analytical and statistical methods and the subsequent elucidation of its biosynthetic gene cluster demonstrate the power of this approach to assign orphan biosynthetic pathways to their respective SMs. We have recently used an analogous approach to identify three novel secondary metabolite classes from *M. xanthus*.[84] This also holds true for very low abundant SMs that might escape visual inspection of the data, as shown for this example.

The rhizopodin assembly line includes several special features, such as an unusual biosynthetic logic for the introduction of methoxy groups and the presence of members of a new KS phylogenetic subclade, both not reported to date from other *trans*-AT PKSs. The provided data will help to expand the knowledge about *trans*-AT PKSs which is still far less comprehensive than that compiled for *cis*-AT systems over many years.

Experimental Section

Bacterial strains and culture conditions

Escherischia coli strains for plasmid preparations were grown in LB at 37 °C. *Stigmatella aurantiaca* Sg a15 and its descendants were cultured in Tryptone-Starch (TS) medium (1% tryptone, 1.19% HEPES, 0.4% soluble starch, 0.2% $MgSO_4 \cdot 7H_2O$; pH adjusted to 7.2) at 30 °C on an orbital shaker. Kanamycin sulfate 50 $\mu g \cdot mL^{-1}$ was added for selection whenever necessary.

Sequence analysis

The DNA sequence of Sg a15 was screened for PKS and NRPS coding sequence using a complete genome scanning pipeline provided by Jacques Ravel (available at: http://nrps.igs.umaryland.edu/nrps/2metdb/Genome_scanning_download.html).

DNA preparations and PCR

S. aurantiaca Sg a15 genomic DNA was prepared with the Puregene® Genomic DNA Purification Kit (Gentra) according to the manufacturer´s protocol. Plasmid DNA was either purified by standard alkaline lysis or by using the NucleoSpin Plasmid Kit (Macherey-Nagel).

For inactivation constructs ~750 bp internal fragments of the target genes were amplified using Taq polymerase (Fermentas) and a set of specific forward and reverse primers. The respective forward primer contained a base pair insertion that caused a frame-shift. The PCR reactions contained 5% dimethylsulfoxid and were carried out in a peqSTAR 96 Universal Gradient (Peqlab) thermocycler. Conditions for PCR reactions: initial denaturation step for 2 min 95 °C, 40 cycles consisting of denaturation at 95 °C for 18 s, annealing at 68 °C for 14 s, elongation at 72 °C for 14 s with terminal a final elongation step at 72 °C for 5 min. The PCR products were cloned into pCR2.1-TOPO (Invitrogen) and sequenced. Primers were *rizB*_KS1ko_for (5′-GCGGATCTCTCCTTGAAGTGGT-3′), *rizB*_KS1ko_rev (5′-TCATCTGCAACAGCACCCTC-3′), *rizB*_KS7ko_for (5′-GCTTCATCGACTGACGTGGAT-3′), *rizB*_KS7ko_rev 5′-GGAATGACACCC TGGCGCAT-3′), *rizA*_ko_for (5′-GTTCTCAGCATGTGGAGGATG-3′), *rizA*_ko_rev (5′-GCTGCTG CAGCAAGTATTGGA-3′), *rizF*_ko_for (5′-CGAGCTGTTTCGACGAGTTTC-3′), *rizF*_ko_rev (5′-G ATGAGGGGCGTCAGGAATTT-3′), *riz_orf5*_ko_for (5′-CCTTGTGACTCACAACTCAAGGA-3′) and *riz_orf5*_ko_rev (5′-CGAGGATCTCTATCGCCAGAAC-3′). The resulting plasmids were introduced into *S. aurantiaca* Sg a15 by electroporation as described previously.[22] The integration of the plasmids in the respective mutants was verified by PCR control (for details see Figure S6). Primers *riz*_connect_for (5′-CTCACGTGCCGCCTGTTATCT-3′) and *riz*_connect_rev (5′-CGCGGATCTGGACCTCTTCT-3′) were used to provide a PCR fragment which was cloned into pCR2.1-TOPO and sequenced in order to close the gap in continuous sequence.

Analysis of SMs

For secondary metabolite production strain Sg a15 and its descendants were cultivated in shake flasks in TS-medium in presence of amberlite XAD-16 (1%; Sigma-Aldrich) for 7 days. Cells and XAD were separated from the culture broth by centrifugation, extracted with methanol and the extracts were concentrated *in vacuo*. To compare the SM profile of the mutants with that of the wild-type strain, the extracts were analyzed on a Dionex Ultimate 3000 RS LC connected to a Thermo Fisher Orbitrap XL operated in positive ionization mode with R = 30000. The HPLC-system was coupled to the Orbitrap by a Triversa Nanomate from Advion, a chip-based nano-ESI interface. Chromatographic conditions: A Waters BEH C_{18} column (50 x 2 mm, 1.7 µm dp, flow 600 µL · min^{-1}) was used for separation with a mobile phase made up of water/acetonitrile (both solvents containing 0.1 % formic acid) with a

gradient from 5 - 95% acetonitrile in 9 min, total runtime 14 min. Since no obvious differences between the SM profiles of mutant versus wild-type were observed, cultures of the KS inactivation mutants and the wild-type were prepared in quintuplicate and extracted and analyzed as described above.

Statistics facilitated data analysis

Briefly, the obtained data were used to calculate a MS-PCA model using the ProfileAnalysis software (version 2.0, Bruker Daltonic GmbH) as described by Cortina *et. al*.[84] Bucket generation was performed for the retention time range from 2 to 11 min and for mass ranges of 200 - 1000 and 1000 – 2000 *m/z* employing an advanced compound-based bucketing approach allowing for deviations of $\Delta m/z = 30$ ppm and ΔRT = 30 s. Parameters for compound detection were a signal-to-noise (S/N) ratio of 5, a minimum compound length of 10 spectra and a smoothing width of 2. In the loadings plot of the resulting MS-PCA model data points (representing buckets) which were located far away from the central cloud were picked and the respective masses were checked.

Feeding experiments

Feeding experiments with L-serine-$^{13}C_3$,^{15}N (Sigma-Aldrich) and L-methionine-(*S*-methyl-^{13}C)(Cambridge Isotope Laboratories) were performed in TS medium with reduced culture volumes (20 mL) in the presence of XAD adsober resin (1%). Starting on the third day of cultivation, 10 mg of L-serine-$^{13}C_3$,^{15}N were fed in three fractions (every 24 hours) to the culture, which was harvested and extracted after five days. For L-methionine-(*S*-methyl-^{13}C), 20 mg were fed following the same schedule.

Construction of expression constructs and protein expression

RizA was amplified from genomic DNA by using Phusion DNA polymerase (Finnzymes) with primer rizA_for_*Nde*I (5´-TTGACATATGCGGATCCAGATTTTTCC-3`) and rizA_rev_*EcoR*I (5´-TTGAGAATTCTCATGACGCCCCCTTCCGCC-3´). Restriction sites are underlined. The resulting PCR product was purified, digested with *Nde*I and *EcoR*I, and ligated into likewise digested pET28b(+) resulting in pET28b(+)-*rizA*, designed for expression of the heterologous protein with a N-terminal hexahistidine tag. The ligation was electroporated into *E. coli* DH10B and positive clones were selected on LB agar plates containing 50 µg mL^{-1} kanamycin. The identity of the plasmids was confirmed by restriction analysis and sequencing. Constructs pET28b(+)-*rizF*, pET28b(+)-*ACP12*, pET28b(+)-*ACP13*,

pET28b(+)-*ACP17*, and pET28b(+)-*ACP18* were prepared analogous with primers *rizF*_for_*Nde*I (5′-TTGACATATGACGAAGATCTTCGTATTT-3′), *rizF*_rev_*EcoR*I (5′-TTGAGAATTCCTACACGCCCGCAGTCCGTTT-3′), *ACP12*_for_*Nde*I (5′-TTGACATATGCGCCCTGCGGCGACCGACCT-3′), *ACP12*_rev_*EcoR*I (5′-TTGAGAATTCTCAGTGCCTTTCGCTCAGGTACT-3′), *ACP13*_for_*Nde*I (5′-TTGACATATGGCGCAGGATGCGCTCCAGGA-3′), *ACP13*_rev_*EcoR*I (5′-TTGAGAATTCTCACTGGCGCGAGGCGAGATGGC-3′), *ACP17*_for_*Nde*I (5′-TTGACATATGCGAGAGGCGCTGAACTCCCG-3′), *ACP17*_rev_*EcoR*I (5′-TTGAGAATTCTCAGTGGTGCTTGAAGAGGTAGT-3′), *ACP18*_for_*Nde*I (5′-TTGACATATGACTGCTTCGTCGGCGGATAT-3′), *ACP18*_rev_*EcoR*I (5-TTGAGAATTCTCATTCCGCCGCACCGAGGTGG-3′). For production of N-terminal His-tagged fusion proteins, *E. coli* BL21(DE3) harbouring the respective expression plasmid was grown at 37 °C in 250 mL LB medium containing 50 µg · mL^{-1} kanamycin to an OD_{600} of approximately 0.8 units. In order to obtain the heterologous ACP proteins as *holo*-ACPs they were expressed in the presence of pSUMtaA.[85] For protein expression the culture was induced with 0.2 mM isopropyl β-D-thiogalactopyranoside (IPTG) and incubated at 16 °C for 16 h. The cells were harvested by centrifugation at 7300 · g for 12 min at 4 °C in a Beckman Coulter Avanti J-E centrifuge. All following steps were carried out at 4 °C. The cell pellet was resuspended in 25 mL binding buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 20 mM imidazole, glycerol 10 % (*v/v*)) and lysed with a French press. Cell debris was removed by centrifugation at 25000 · g for 45 min. The supernatant was filtered through 1.2 µm Acrodisc syringe filters (PALL Life Sciences) before being loaded onto a 1 mL Hi-Trap His column using the Äkta purifier system (both from GE Healthcare) at a flowrate of 1 mL · min^{-1}. The unbound sample was washed away with 5 column volumes binding buffer. For elution of the His-tagged heterologous proteins a stepwise gradient with 7%, 16%, 40% and 100% elution buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 500 mM imidazole, glycerol 10% (*v/v*)) with 6 column volumes for each step was applied. The fractions containing heterologous protein were buffer exchanged to 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, glycerol 10% (v/v) using PD10 columns (Amersham Biosciences) and further concentrated using Amicon ultracentrifugal filter devices. Concentrated protein solutions were frozen in liquid nitrogen and stored at -80 °C. Protein concentrations were determined with Bio-Rad protein assay kit according to the manufacturer's instructions using bovine serum albumin (BSA) as a standard (0.5 – 8 µg · mL^{-1}). Protein fractions were separated on a 15 % sodium dodecyl sulfate (SDS)-polyacrylamide gel by electrophoresis. Proteins were

visualized in the gel by staining with Coomassie Brilliant Blue R-250 (for protein preparations used see Figure S6). For matrix assisted laser desorption ionisation time of flight mass spectrometry (MALDI-ToF-MS) protein bands were cut from the SDS-polyacrylamide gel and in gel digestion with trypsin was performed. Samples were prepared using Zip-Tips (Milipore).

ACP loading assays

Holo-ACPs from different modules of the rhizopodin assembly line were incubated with recombinant RizA or RizF protein in order to investigate malonyl priming activities. Incubations contained 50 mM Tris-HCl pH 8.0, 1 mM malonyl-CoA, 20 µM *holo*-ACP and 0.2 µM trans-AT (either RizA or RizF). Reactions were incubated for 30 min at 30 °C and then analyzed by HPLC-MS on a Jupiter 5 micron C4 300A column (50 x 1.0 mm; Phenomenex) with a water/acetonitrile (both solvents containing 0.1% formic acid) gradient (5 to 50% organic fraction over 20 min), coupled to a HCT-Ultra (Bruker Daltonics) mass spectrometer operating in positive ionization mode. Average MS-spectra were computated and deconvoluted using DataAnalysis software (Bruker Daltonics).

Sequence analysis

Prediction of open reading frames was performed with FramePlot 4.0 (http://nocardia.nih.go.jp/fp4/) and the resulting hits were analyzed by tblastx search.[86] Annotation of the rhizopodin biosynthetic gene cluster and sequence alignments were carried out using Geneious Pro 5.3.4.[87] Computation of phylogenetic trees was performed by Bayesian estimation with the program MrBayes.[88] Settings used: rate matrix: wag; rate variation: gamma; gamma categories: 4; Markov chain Monte Carlo (MCMC) chain length: 1.000.000 generations; heated chains: 4; subsampling frequency: 500 generations.

Accession number

The nucleotide sequence reported here is accessible at the EMBL Nucleotide Sequence Database under accession number FR854394.

References

[1] K. J. Weissman, R. Müller, *Nat.Prod.Rep.* **2010,** *27*, 1276-1295.

[2] B. Kunze, G. Höfle, H. Reichenbach, *J.Antibiot.* **1987,** *40*, 258-265.

[3] B. Silakowski, B. Kunze, G. Nordsiek, H. Blöcker, G. Höfle, R. Müller, *Eur.J.Biochem.* **2000,** *267,* 6476-6485.

[4] B. Kunze, T. Kemmer, G. Höfle, H. Reichenbach, *J.Antibiot.* **1984,** *37*, 454-461.

[5] K. Gerth, R. Jansen, G. Reifenstahl, G. Höfle, H. Irschik, B. Kunze, H. Reichenbach, G. Thierbach, *J.Antibiot.* **1983,** *36*, 1150-1156.

[6] S. C. Wenzel, B. Kunze, G. Höfle, B. Silakowski, M. Scharfe, H. Blöcker, R. Müller, *ChemBioChem* **2005,** *6*, 375-385.

[7] P. Meiser, K. J. Weissman, H. B. Bode, D. Krug, J. S. Dickschat, A. Sandmann, R. Müller, *Chem.Biol.* **2008,** *15*, 771-781.

[8] M. A. Marahiel, *FEBS Lett.* **1992,** *307*, 40-43.

[9] S. Donadio, M. J. Staver, J. B. McAlpine, S. J. Swanson, L. Katz, *Science* **1991,** *252*, 675-679.

[10] C. T. Walsh, H. W. Chen, T. A. Keating, B. K. Hubbard, H. C. Losey, L. S. Luo, C. G. Marshall, D. A. Miller, H. M. Patel, *Curr.Opin.Chem.Biol.* **2001,** *5*, 525-534.

[11] D. E. Cane, C. T. Walsh, *Chem.Biol.* **1999,** *6*, R319-R325.

[12] J. Piel, *P.Natl.Acad.Sci.USA* **2002,** *99*, 14002-14007.

[13] D. E. Cane, C. T. Walsh, C. Khosla, *Science* **1998,** *282*, 63-68.

[14] B. S. Goldman, W. C. Nierman, D. Kaiser, S. C. Slater, A. S. Durkin, J. Eisen, C. M. Ronning, W. B. Barbazuk, M. Blanchard, C. Field, C. Halling, G. Hinkle, O. Iartchuk, H. S. Kim, C. Mackenzie, R. Madupu, N. Miller, A. Shvartsbeyn, S. A. Sullivan, M. Vaudin, R. Wiegand, H. B. Kaplan, *P.Natl.Acad.Sci.USA* **2006,** *103*, 15200-15205.

[15] S. Schneiker, O. Perlova, O. Kaiser, K. Gerth, A. Alici, M. O. Altmeyer, D. Bartels, T. Bekel,S. Beyer, E. Bode, H. B. Bode, C. J. Bolten, J. V. Choudhuri, S. Doss, Y. A. Elnakady, B. Frank, L. Gaigalat, A. Goesmann, C. Groeger, F. Gross, L. Jelsbak, L. Jelsbak, J. Kalinowski, C. Kegler, T. Knauber, S. Konietzny, M. Kopp, L. Krause, D. Krug, B. Linke, T. Mahmud, R. Martinez-Arias, A. C. McHardy, M. Merai, F. Meyer, S. Mormann, J. Munoz-Dorado, J. Perez, S. Pradella, S. Rachid, G. Raddatz, F. Rosenau, C. Rückert, F. Sasse, M. Scharfe, S. C. Schuster, G. Suen, A. Treuner-Lange, G. J. Velicer, F. J. Vorhölter, K. J. Weissman, R. D. Welch, S. C. Wenzel, D. E. Whitworth, S. Wilhelm, C. Wittmann, H. Blöcker, A. Pühler, R. Müller, *Nat.Biotechnol.* **2007,** *25*, 1281-1289.

[16] H. Ikeda, J. Ishikawa, A. Hanamoto, M. Shinose, H. Kikuchi, T. Shiba, Y. Sakaki, M. Hattori, S. Omura, *Nat.Biotechnol.* **2003,** *21*, 526-531.

[17] S. D. Bentley, K. F. Chater, A. M. Cerdeno-Tarraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C. H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill, D. A. Hopwood, *Nature* **2002,** *417*, 141-147.

[18] M. Oliynyk, M. Samborskyy, J. B. Lester, T. Mironenko, N. Scott, S. Dickens, S. F. Haydock, P. F. Leadlay, *Nat.Biotechnol.* **2007,** *25*, 447-453.

[19] S. C. Wenzel, R. Müller, *Nat.Prod.Rep.* **2009,** *26*, 1385-1407.

[20] M. Zerikly, G. L. Challis, *ChemBioChem* **2009,** *10*, 625-633**.**

[21] B. O. Bachmann, J. Ravel, *Methods in Enymology* **2009,** *458*, 181-217.

[22] S. Beyer, B. Kunze, B. Silakowski, R. Müller, *Biochim.Biophys.Acta* **1999,** *1445*, 185-195.

[23] D. Krug, G. Zurek, B. Schneider, R. Garcia, R. Müller, *Anal.Chim.Acta* **2008,** *624* 97-106.

[24] F. Sasse, H. Steinmetz, G. Höfle, H. Reichenbach, *J.Antibiot.* **1993,** *46*, 741-748.

[25] R. Jansen, H. Steinmetz, F. Sasse, W. D. Schubert, G. Hagelueken, S. C. Albrecht, R. Müller, *Tetrahedron Lett.* **2008,** *49*, 5796-5799.

[26] N. Horstmann, D. Menche, *Chem.Commun.(Camb.)* **2008,** 5173-5175.

[27] G. Hagelueken, S. C. Albrecht, H. Steinmetz, R. Jansen, D. W. Heinz, M. Kalesse, W. D. Schubert, *Angew Chem Int Ed Engl* **2009,** *48*, 595-598.
[28] S. W. White, J. Zheng, Y. M. Zhang, Rock, *Annu Rev Biochem* **2005,** *74*, 791-831.
[29] A. K. El-Sayed, J. Hothersall, S. M. Cooper, E. Stephens, T. J. Simpson, C. M. Thomas, *Chem.Biol.* **2003,** *10*, 419-430.
[30] T. Stachelhaus, H. D. Mootz, M. A. Marahiel, *Chem.Biol.* **1999,** *6*, 493-505.
[31] E. Conti, T. Stachelhaus, M. A. Marahiel, P. Brick, *EMBO J.* **1997,** *16*, 4174-4183.
[32] R. Reid, M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V. Santi, C. R. Hutchinson, R. McDaniel, *Biochemistry-US* **2003,** *42*, 72-79.
[33] A. Keatinge-Clay, *J.Mol.Biol* **2008,** *384*, 941-953.
[34] M. Z. Ansari, J. Sharma, R. S. Gokhale, D. Mohanty, *Bmc Bioinformatics* **2008,** *9* .
[35] D. Konz, M. A. Marahiel, *Chem.Biol.* **1999,** *6*, R39-R48.
[36] D. Schwarzer, R. Finking, M. A. Marahiel, *Nat.Prod.Rep.* **2003,** *20*, 275-287.
[37] N. Kessler, H. Schuhmann, S. Morneweg, U. Linne, M. A. Marahiel, *J.Biol.Chem.* **2004,** *279*, 7413-7419.
[38] L. Rouhiainen, L. Paulin, S. Suomalainen, H. Hyytiainen, W. Buikema, R. Haselkorn, K. Sivonen, *Mol.Microbiol.* **2000,** *37*, 156-167.
[39] C. Zhao, J. M. Coughlin, J. Ju, D. Zhu, E. Wendt-Pienkowski, X. Zhou, Z. Wang, B. Shen, Z. Deng, *J.Biol.Chem.* **2010,** *285*, 20097-20108.
[40] J. Piel, *Proceedings of National Academy of Sciences USA* **2002,** *99*, 14002-14007.
[41] J. Piel, D. Hui, G. Wen, D. Butzke, M. Platzer, N. Fusetani, S. Matsunaga, *P.Natl.Acad.Sci.USA* **2004,** *101*, 16222-16227.
[42] S. Sudek, N. B. Lopanik, L. E. Waggoner, M. Hildebrand, C. Anderson, H. B. Liu, A. Patel, D. H. Sherman, M. G. Haygood, *J.Nat.Prod.* **2007,** *70*, 67-74.
[43] M. Kopp, H. Irschik, S. Pradella, R. Müller, *ChemBioChem* **2005,** *6*, 1277-1286.
[44] R. Teta, M. Gurgui, E. J. N. Helfrich, S. Künne, A. Schneider, G. van Echten-Deckert, A. Mangoni, J. Piel, *ChemBioChem* **2011,** *11*, 2506-2512.
[45] B. Julien, S. Shah, R. Ziermann, R. Goldman, L. Katz, C. Khosla, *Gene* **2000,** *249*, 153-160.
[46] K. Buntin, H. Irschik, K. J. Weissman, E. Luxenburger, H. Blöcker, R. Müller, *Chem.Biol.* **2010,** *17*, 342-356.
[47] K. Buntin, S. Rachid, M. Scharfe, H. Blöcker, K. J. Weissman, R. Müller, *Angew.Chem.Int.Ed.* **2008,** *47*, 4595-4599.
[48] B. Frank, J. Knauber, H. Steinmetz, M. Scharfe, H. Blöcker, S. Beyer, R. Müller, *Chem.Biol.* **2007,** *14*, 221-233.
[49] H. Irschik, M. Kopp, K. J. Weissman, K. Buntin, J. Piel, R. Müller, *ChemBioChem* **2010,** *11*, 1840-1849.
[50] X. H. Chen, J. Vater, J. Piel, P. Franke, R. Scholz, K. Schneider, A. Koumoutsi, G. Hitzeroth, N. Grammel, A. W. Strittmatter, G. Gottschalk, R. D. Süssmuth, R. Borriss, *J.Bacteriol.* **2006,** *188*, 4024-4036.
[51] D. Menche, F. Arikan, O. Perlova, N. Horstmann, W. Ahlbrecht, S. C. Wenzel, R. Jansen, H. Irschik, R. Müller, *J.Am.Chem.Soc.* **2008,** *130*, 14234-14243.
[52] E. M. Musiol, T. Härtner, A. Kulik, J. Moldenhauer, W. Wohlleben, T. Weber, *Chemistry & Biology* **2011,** *18*, 438-444.
[53] S. B. Bumpus, N. A. Magarvey, N. L. Kelleher, C. T. Walsh, C. T. Calderone, *J.Am.Chem.Soc.* **2008,** *130*, 11614-11616.
[54] B. Silakowski, H. U. Schairer, H. Ehret, B. Kunze, S. Weinig, G. Nordsiek, P. Brandt, H. Blöcker, G. Höfle, S. Beyer, R. Müller, *J.Biol.Chem.* **1999,** *274*, 37391-37399.
[55] S. C. Wenzel, P. Meiser, T. Binz, T. Mahmud, R. Müller, *Angew.Chem.Int.Ed.* **2006,** *45*, 2296-2301.
[56] M. L. Heathcote, J. Staunton, P. F. Leadlay, *Chem.Biol.* **2001,** *8*, 207-220.
[57] H. Gross, V. O. Stockwell, M. D. Henkels, B. Nowak-Thompson, J. E. Loper, W. H. Gerwick, *Chem.Biol.* **2007,** *14*, 53-63.
[58] S. B. Bumpus, B. S. Evans, P. M. Thomas, I. Ntai, N. L. Kelleher, *Nat.Biotechnol.* **2009,** *27*, 951-956.

[59] D. Krug, G. Zurek, O. Revermann, M. Vos, G. J. Velicer, R. Müller, *Appl.Environ.Microbiol.* **2008,** *74*, 3058-3068.
[60] C. Scotti, M. Piatti, A. Cuzzoni, P. Perani, A. Tognoni, G. Grandi, A. Galizzi, A. M. Albertini, *Gene* **1993,** *130*, 65-71.
[61] J. Piel, *Nat.Prod.Rep.* **2010,** *27*, 996-1047.
[62] V. Simunovic, J. Zapp, S. Rachid, D. Krug, P. Meiser, R. Müller, *ChemBioChem* **2006,** *7*, 1206-1220.
[63] G. L. Tang, Y. Q. Cheng, B. Shen, *Chem.Biol.* **2004,** *11*, 33-45.
[64] T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, J. Piel, *Nat.Biotechnol.* **2008,** *26*, 225-233.
[65] P. Caffrey, *ChemBioChem* **2003,** *4*, 654-657.
[66] A. T. Keatinge-Clay, *Chem.Biol.* **2007,** *14*, 898-908.
[67] Y. Q. Cheng, G. L. Tang, B. Shen, *P.Natl.Acad.Sci.USA* **2003,** *100*, 3149-3154.
[68] L. P. Partida-Martinez, C. Hertweck, *ChemBioChem* **2007,** *8*, 41-45.
[69] O. Perlova, K. Gerth, A. Hans, O. Kaiser, R. Müller, *J.Biotechnol.* **2006,** *121*, 174-191.
[70] N. Pulsawat, S. Kitani, T. Nihira, *Gene* **2007,** *393*, 31-42.
[71] A. Sandmann, F. Sasse, R. Müller, *Chem.Biol.* **2004,** *11*, 1071-1079.
[72] K. Gerth, N. Bedorf, G. Höfle, H. Irschik, H. Reichenbach, *J.Antibiot.* **1996,** *49*, 560-563.
[73] R. Jansen, B. Kunze, H. Reichenbach, G. Höfle, *Eur.J.Org.Chem.* **2002,** 917-921.
[74] C. G. Marshall, M. D. Burkart, T. A. Keating, C. T. Walsh, *Biochemistry-US* **2001,** *40*, 10655-10663.
[75] M. Di Lorenzo, M. Stork, H. Naka, M. E. Tolmasky, J. H. Crosa, *Biometals* **2008,** *21*, 635-648.
[76] C. G. Marshall, N. J. Hillson, C. T. Walsh, *Biochemistry-US* **2002,** *41*, 244-250.
[77] F. Lombo, A. Velasco, A. Castro, F. de la Calle, A. F. Brana, J. M. Sanchez-Puelles, C. Mendez, J. A. Salas, *ChemBioChem* **2006,** *7*, 366-376.
[78] H. C. Kwon, C. A. Kauffman, P. R. Jensen, W. Fenical, *J.Am.Chem.Soc.* **2006,** *128*, 1622-1632.
[79] R. K. Boeckman, J. Fayos, J. Clardy, *J.Am.Chem.Soc.* **1974,** *96*, 5954-5956.
[80] K. M. Hoyer, C. Mahlert, M. A. Marahiel, *Chem.Biol.* **2007,** *14*, 13-22.
[81] C. A. Shaw-Reid, N. L. Kelleher, H. C. Losey, A. M. Gehring, C. Berg, C. T. Walsh, *Chem.Biol.* **1999,** *6*, 385-400.
[82] L. Robbel, K. M. Hoyer, M. A. Marahiel, *FEBS J.* **2009,** *276*, 1641-1653.
[83] R. Jansen, H. Irschik, H. Reichenbach, V. Wray, G. Höfle, *Liebigs Ann Chem* **1994,** 759-773.
[84] N. S. Cortina, Krug, D., Plaza, A., Revermann, O., and Müller, R. **2011,** *Angew. Chem. Int. Ed.*, submitted.
[85] N. Gaitatzis, A. Hans, R. Müller, S. Beyer, *J.Biochem.(Tokyo)* **2001,** *129*, 119-124.
[86] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. H. Zhang, Z. Zhang, W. Miller, D. J. Lipman, *Nucleic Acids Res.* **1997,** *25*, 3389-3402.
[87] Drummond, A. J., Ashton, B., Buxton, S., Cheung, M., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Thierer, T., and Wilson, A. Geneious Pro 4.8.3 . **2010**. Ref Type: Computer Program
[88] J. P. Huelsenbeck, F. Ronquist, *Bioinformatics* **2001,** *17*, 754-755.

Supporting information

Table S1. *Alignment of conserved residues (active-site residues underlined) which correlate with domain specificity of RizA and RizF with those of the malonate specific* trans-*ATs ChiA, DszD and LnmG* [a]

Protein	11	63	90	91	92	93	94	117	200	201	231	250	255	15	58	59	60	61	62	70	72	197	198	199
ChiA	Q	Q	G	H	S	L	G	R	F	H	N	Q	V	R	Q	T	Q	Y	T	A	S	S	G	A
DszD	Q	Q	G	H	S	L	G	R	F	H	N	Q	V	A	Q	T	Q	F	T	A	S	S	A	A
LnmG	Q	Q	G	H	S	L	G	R	F	H	S	Q	V	R	R	T	E	Y	A	A	S	S	A	A
RizA	Q	Q	G	H	S	L	G	R	F	H	N	Q	V	H	Q	T	A	W	T	A	A	S	A	A
RizF	Q	Q	G	H	S	L	G	R	F	H	N	Q	V	K	R	T	E	F	T	A	A	S	A	A

[a] Residue numbering according to Yadav *et al.* [1] from where the consensus active-site residues for malonate specific AT domains were distilled.

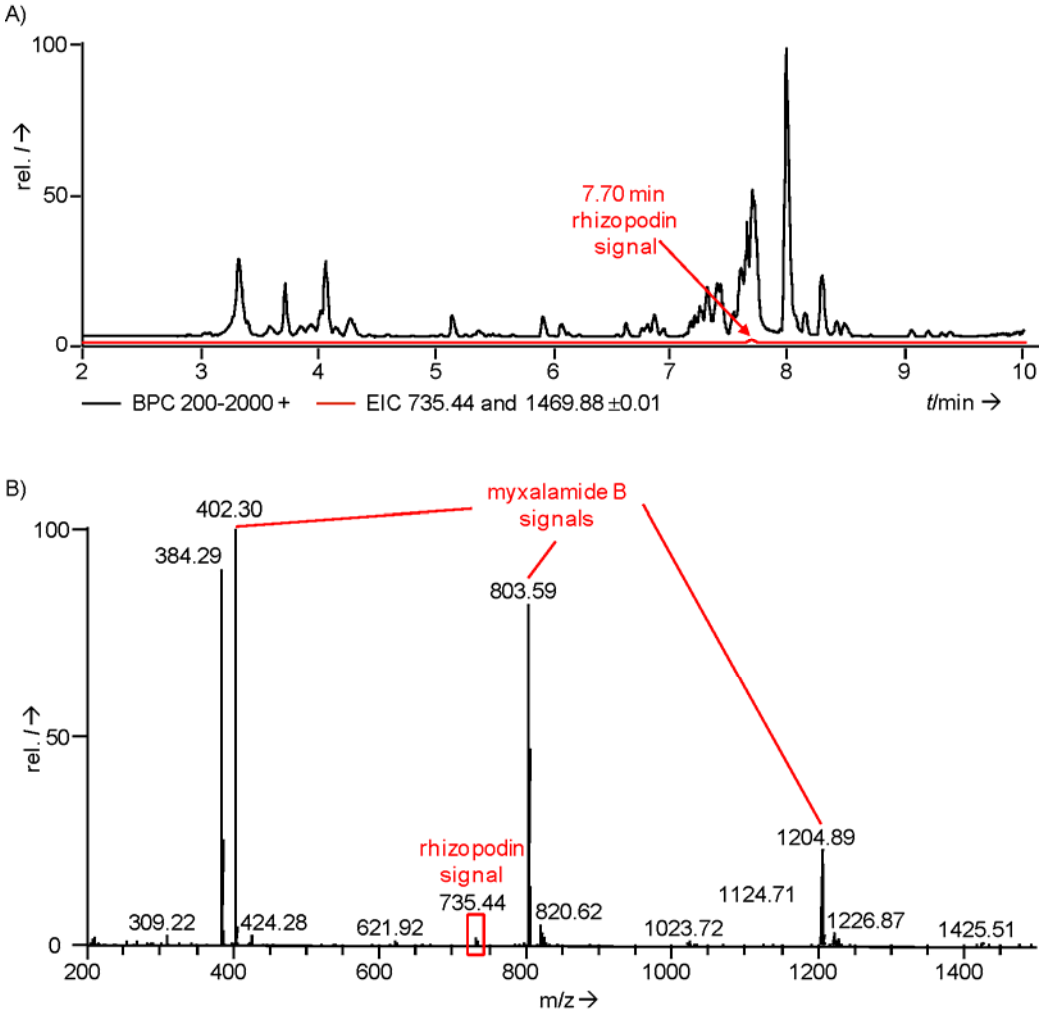


Figure S1. *HPLC-MS data of rhizopodin from an extract of Sg a 15.* A*) HPLC-MS chromatogram of an extract of Sg a15 wild type. Shown in black is the base peak chromatogram (BPC) in positive mode from 200-2000 m/z. Shown in red is the combined extracted ion chromatogram (EIC) of 735.44 and 1469.88 m/z (masses of doubly and singly charged rhizopodin). The low abundant rhizopodin signal is marked by an arrow.* B*) MS-spectrum of retention time 7.70 min, showing the intensities of myxalamide B signals in comparison to that of rhizopodin (marked by a red box).*

Figure S2. *Bayesian cladogram of 116 KS sequences composing from selected members of previously described substrate specific clades and the sequences of the 19 KS domains from the rhizopodin biosynthetic gene cluster (marked in red). Ery KS4 from the canonical erythromycin PKS was used as outgroup. The numbers of the KS domains indicates their position within the respective gene cluster. The subsequent roman numbers indicate their membership to a certain substrate specific clade. Substrate specificities of the respective clades: I : α-branched + either β-OH or β-methylene and β-branched; II : β-OH; III : KS^{0}s, mixed substrates; IV : β-OH (D-configuration); V : ethylene; VI : acetyl starter units; VII : double bond + α-methyl; VIII : β-OH (D-configuration); IX : double bond or heterocycle; X : KS^{0}s, mixed substrates; XI : double bond; XII : double bond; XIII : vinyl or propy; XIV : β-OH, nonelongating KS in dehydrating module; XV : double bond generated by module lacking a DH; XVI : glycine. Abbreviations used for KS domains from biosynthetic gene clusters: Bae, bacillaene; BT, thailandamide; Chi, chivosazol; Dif, difficidin; Dsz, disorazol; GU, uncharacterized PKS from Geobacter uraniumreducens; Kkc, lankacidin; Lnm, leinamycin; Mln, macrolactin; Mmp, mupirocin; Onn, onnamide; Ped, pederin; Riz, rhizopodin; Ta, myxovirescin.*

KR domain	Conserved LDD	Corresponding position in rhizopodin	Predicted stereochemistry	Observed stereochemistry
KR1 (+DH1)	IR**D**	28 and 28´	B-type → reduction to *E*-double bond	*E*-double bond
KR2	LR**D**	26 and 26´	B-type C2 > C4 → R	R
KR5 (+DH5+ER5)	LR**D**	22 and 22´	B-type → complete reduction	-
KR6	VR**D**	20 and 20´	B-type C4 > C2 → S	S
KR8	IR**I**	18 and 18´	A-type C2 > C4 → S	S
KR9	LR**D**	16 and 16´	B-type C4 > C2 → S	S
KR13	LR**D**	11 and 11´	B-type C4 > C2 → S	S
KR16	LR**D**	9 and 9´	B-type → reduction to *E*-double bond	*E*-double bond
		7 and 7´	B-type → reduction to *E*-double bond	*E*-double bond
KR17	AA**Q**	5 and 5´	A-type C4 > C2 → R	R
KR19	ID**S**	3 and 3´	A-type C2 > C4 → S	S

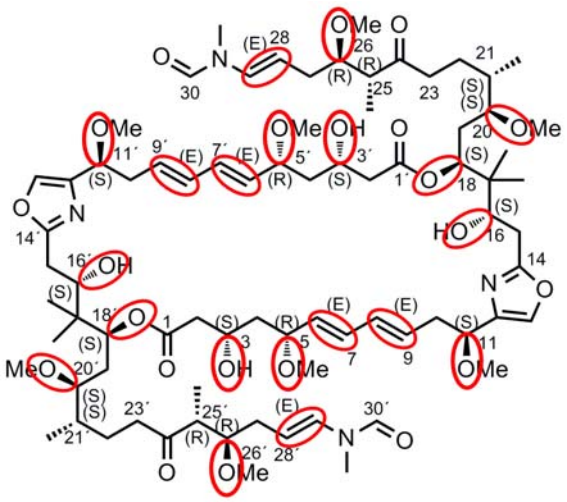


Figure S3. *Prediction of stereospecificities of reduction in rhizopodin biosynthesis.*
Previous work on correlation of observed stereospecificity of ketoreduction with conserved sequence motifs revealed a highly predictive residue.[2;3] *The presence of a conserved Asp residue within a LD**D** motif correlates with B-type ketoreduction (3R, when C2 > C4; 3S when C4 > C2) whereas its absence correlates with A-type ketoreduction (3S, when C2 > C4; 3R when C4 > C2). Further predictive residues, W141 for A-type ketoreduction and P144 and N148 for B-type ketoreduction,*[3] *are not conserved in the ketoreductase sequences of the rhizopodin biosynthtetic gene cluster.*

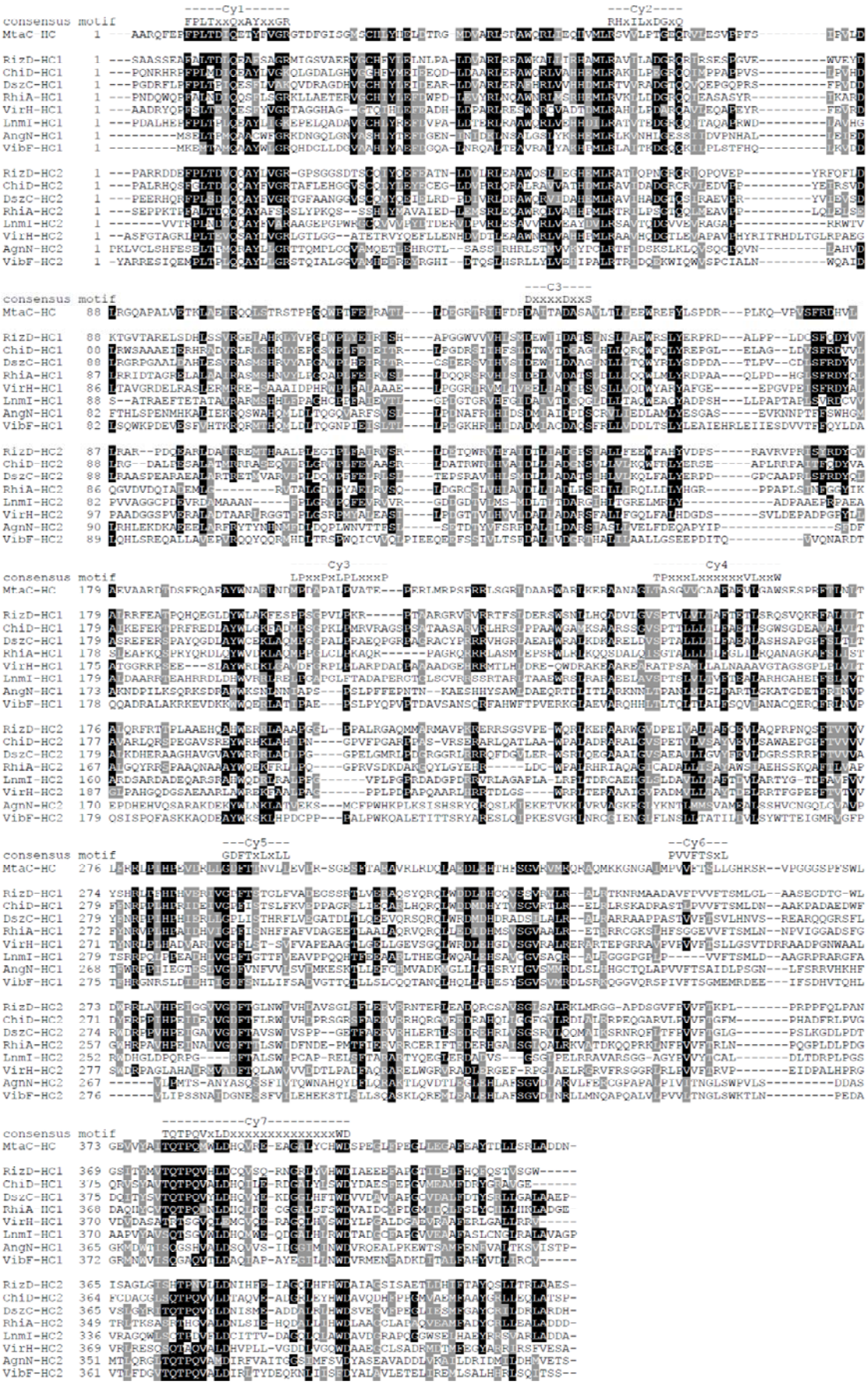


Figure S4.*Alignment of HC1 and HC2 domains from tandem HC domains from* trans*-AT PKS-NRPSs, anguibactin and vibriobactin biosynthesis and the "single" HC domain from myxothiazol. Abbreviations: Ang, anguibactin; Chi, chivosazol; Dsz, disorazol; Lnm, leinamycin; Mta, myxothiazol; Rhi, rhizoxin; Riz, rhizopodin; Vib, vibriobactin; Vir, virginamycin.*

Sequence Analysis of HC domains in order to assign condensation and cyclization half-reactions

The Cy5-motif is much better conserved for the HC2s than for the HC1s, whereas the opposite holds true for the Cy6-motif. Since there are no crystallographic data on HC-domains present we were in our investigations limited to structural data of C-domains[4-6] and data on point mutation experiments performed on HC-domains. Directed amino acid exchanges of the second aspartate in the DXXXXDXXS C3-motif of VibF HC1 and HC2 have shown that this residue is crucial for domain functionality for either condensation or cyclization activity.[7] Both the HC1- and the HC2-domains show a strict conservation of this aspartate residue. By site-directed mutagenesis experiments of specific residues in the HC-domain of the bimodular model system BacA1-2-TE (A-PCP-HC-A-PCP-TE) it has been possible to identify single residues that are absolutely necessary for the condensation or the cyclization half-reactions.[8] The consensus threonine or serine residue at the start of the Cy4-motif and a conserved arginine residue between the Cy4- and the Cy5-motif seem to be essential for the condensation since the respective alanine mutants produced neither linear nor cyclic products. The HC1s show a conserved serine at this postion whereas in the HC2s there is no good conservation. Both HC1s and HC2s show a strict conservation of the just mentioned arginine residue. In the same study it has been shown that a conserved asparagine residue between Cy4 and Cy5 as well as a conserved serine residue within the Cy6-motif are essential for cyclization since the respective alanine mutants produce only linear but no cyclic products. The respective asparagine residue is conserved neither in the HC1s nor in the HC2s but the HC1s show a strictly conserved FTS motif whereas the HC2s show a FTX motif.

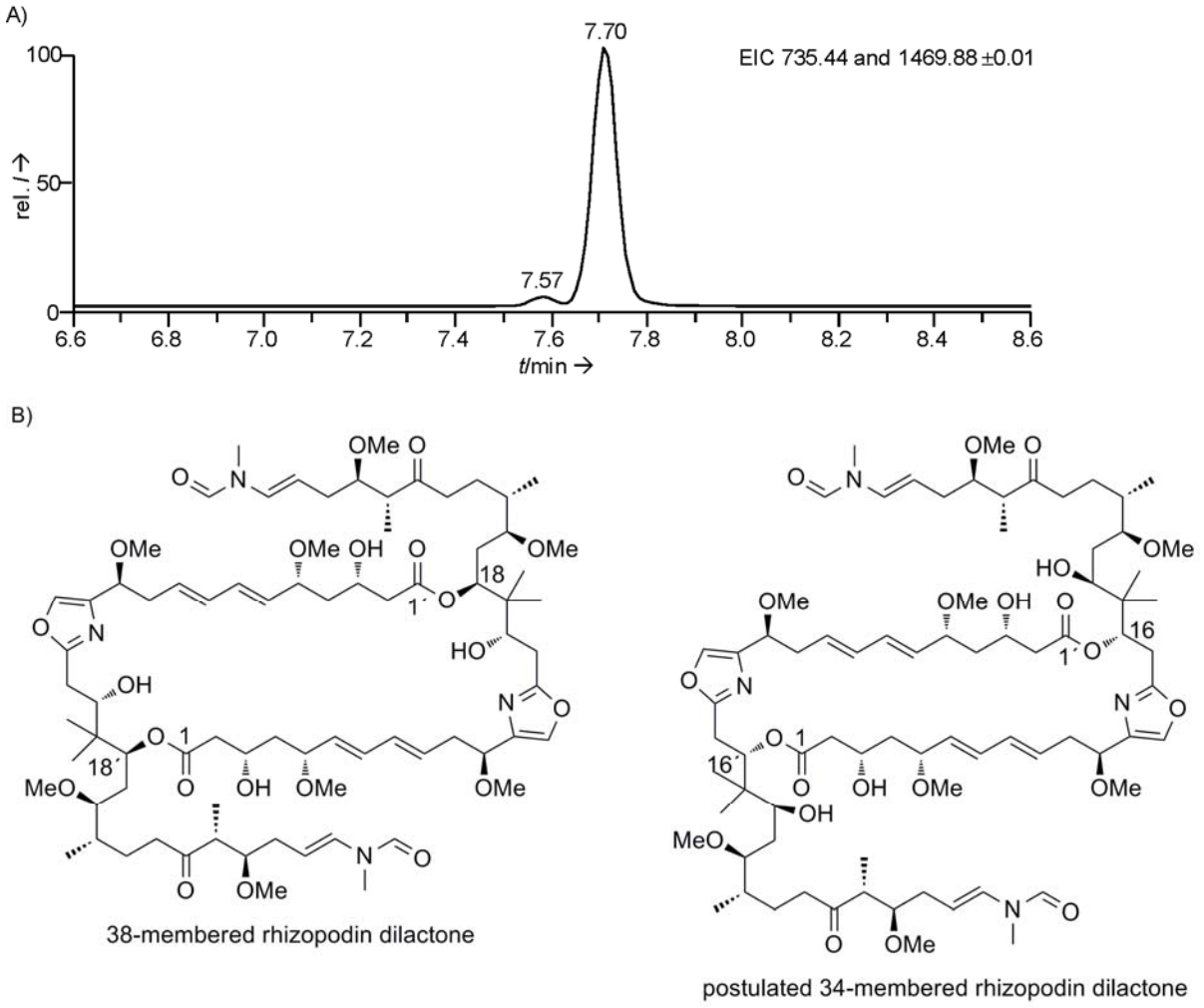


Figure S5. *Postulated rhizopodin isomer.* ***A****) Combined EIC of 735.44 and 1469.88 m/z of an extract of* Myxococcus stipitatus *Mx s11. The chromatogram reveals two peaks with slight differences in retention time.* ***B****) Chemical structures of rhizopodin (38-membered dilactone) and the postulated isomer with a 34-membered dilactone macrocycle.*

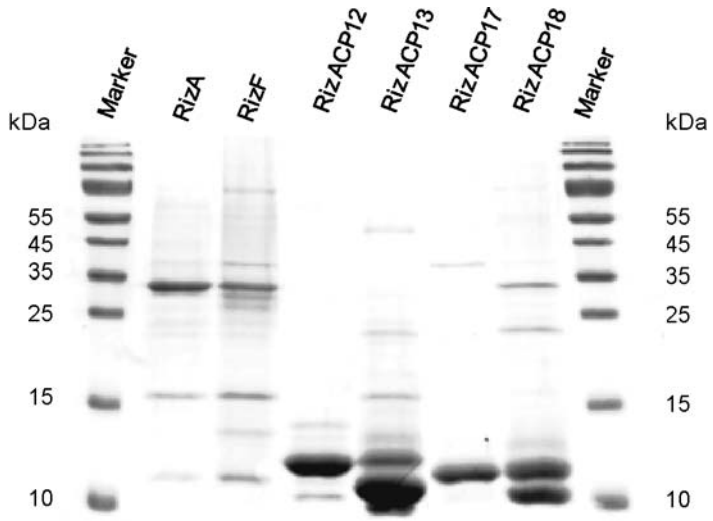


Figure S6. *SDS-PAGE of heterologously expressed and purified proteins used for in vitro ACP loading assays. RizA, calc. size: 33.2 kDa; RizF, calc. size: 34.5 kDa;* holo-*RizACP12, calc. size: 11.8 kDa;* holo-*RizACP13, calc. size: 11.4 kDa;* holo-*RizACP17, calc. size: 11.6 kDa,* holo-*RizACP18, calc. size: 11.7 kDa.*

References

[1] G. Yadav, R. S. Gokhale, D. Mohanty, *Journal of Molecular Biology* **2003,** *328*, 335-363.

[2] R. Reid, M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V. Santi, C. R. Hutchinson, R. McDaniel, *Biochemistry-US* **2003,** *42*, 72-79.

[3] P. Caffrey, *ChemBioChem* **2003,** *4*, 654-657.

[4] T. A. Keating, C. G. Marshall, C. T. Walsh, A. E. Keating, *Nat.Struct.Biol.* **2002,** *9*, 522-526.

[5] S. A. Samel, G. Schoenafinger, T. A. Knappe, M. A. Marahiel, L. O. Essen, *Structure.* **2007,** *15*, 781-792.

[6] A. Tanovic, S. A. Samel, L. O. Essen, M. A. Marahiel, *Science* **2008,** *321,* 659-663.

[7] C. G. Marshall, N. J. Hillson, C. T. Walsh, *Biochemistry-US* **2002,** *41*, 244-250.

[8] T. Duerfahrt, K. Eppelmann, R. Müller, M. Marahiel, *Chem.Biol.* **2004,** *11*, 261-271.

C. Final discussion

1. Tubulysin: new structures and new biosynthetic insights

The tubulysins are a family of secondary metabolites produced by a number of myxobacterial strains including *Angiococcus disciformis* An d48, *Cystobacter* sp. SBCb004, and *Archangium gepyhra* Ar315.[111;112] These natural products are characterized by the presence of a linear tetrapeptide core, consisting of *N*-methyl pipecolic acid (Mep), isoleucine (Ile), a novel amino acid named tubuvaline (Tuv), which is biosynthetically derived from an acetate bridged valine-cysteine building block, and a chain-extended analogue of either phenylalanine or tyrosine, termed tubuphenylalanine (Tup) or tubutyrosine (Tut), respectively. Further modifications by different *O*-acylations expand this tetrapeptide sequence to the naturally occuring tubulysins A-I.

The tubulysins belong to a select group of natural products that are able to interact with the eukaryotic microtubule network.[27;33] Unlike the epothilones that induce tubulin polymerization, the tubulysins inhibit polymerization at concentration lower than 50 pg mL^{-1}.[32] Interestingly, tubulysins show 20 to 100-fold more potency in inhibiting the growth of cancer cells when compared to those of other tubulin modifiers such as epothilones, vinblastine, and taxol. Moreover tubulysins seem to be less affected by multiple drug resistance (MDR) effects, since they present poor substrates for the cellular transporters mainly conferring MDR. Altogether these facts have stimulated significant research into the chemistry and biology of the tubulysins to exploit their potential for therapy of multi drug-resistant tumors. Structure-activity relationship (SAR) studies with tubulysin synthetic analogues have identified the essential structural features related to the cytotoxic activity, as well as suggesting strategies for optimizing the metabolite's pharmacological properties.[113-118] These studies explored the acetoxy moiety at C5 of Tuv, the bis-acyl *N*,*O*-acetal substituent, the Mep heterocycle and the C-terminal Tup/Tut residue for their respective influence on bioactivity (results are sumarized in **Figure C-1**).

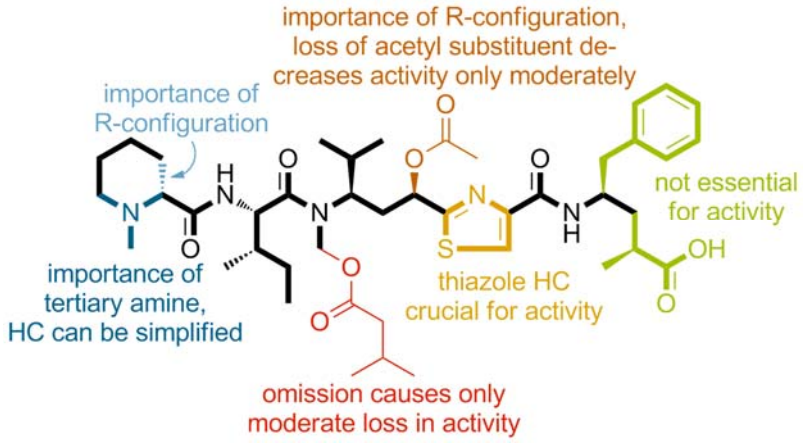


Figure C-1. *Combined insights gained from several SAR studies on tubulysins presented for tubulysin D. The pretubulysin core structure is indicated by bold bonds.*

These results show that the *N*-terminal tertiary amine within the piperidine heterocycle is very important for activity. However, the piperidine heterocycle is not essential for activity and can be simplified. *R*-configuration of the alkyl substituent at the Mep α-carbon it is also important for activity. Only a moderate loss of activity was observed upon omission of the chemically labile *O*-acyl *N*,*O*-acetal group or its replacement with stable alternatives. The same holds true for omission of the Tuv C5-acetoxy group. On the other hand a change in the stereochemistry at the Tuv C5 position decreases activity. Both the benzyl and the 2-methylpropionyl moiety can be deleted without detriment on activity, as long as they are not deleted simultaneously. The recent elucidation of the tubulin-bound structure of tubulysin A allowed for rationalization of these SAR data.[119] In the tubulin-bound state, the thiazole nitrogen atom and the phenyl ring form a basal platform at the bottom of the molecule upon which the *O*-acyl *N*,*O*-acetal side chain of Tuv packs to form a hydrophobic core. In this conformation the alkyl backbone of the Mep piperidine ring, the benzyl and 2-methylpropionyl moieties of Tut and the *O*-acyl *N*,*O*-acetal group, which all have been shown to be not essential for activity, are all situated on the same side of the molecule. The hydrophobic core composed of the thiazole ring, the valine side chain of Tuv and the Ile side chain, seems to be essential, as indicated by the relevance of the stereochemistry at the Tuv C5 atom. The hydrophilic tail of the Mep residue which is situated on the opposite side of this T-shaped core is likely to form a hydrogen bond with side chains of tubulin. Therein it was also shown that simplification of the *N*-terminal Mep in combination with loss of both the acetoxy and the *O*-acyl *N*,*O*-acetal substituent and the loss of Tup C2-methyl resulted in a drastic decrease of activity, suggesting the importance of these functionalities in the absence of the acetoxy and the *N*,*O*-acetal substituents. In addition the importance of the thiazole

heterocycle for activity has been demonstrated as substitutions against phenyl and phenoxy moieties dramatically decreased activity.[120]

In light of these SAR results the previously postulated first enzyme free tubulysin intermediate,[121] termed pretubulysin, should be investigated for its cytotoxic activity. Pretubulysin presents the polyketide-nonribosomal peptide core minus the four post-assembly line oxidative and acylation modifications (structure highlighted by bold bonds in **Figure C-1**). These modifications are most likely introduced by the action of freestanding enzymes, since no corresponding functionalities have been identified in the tubulysin megasynthetases. Pretubulysin was prepared synthetically and the availability of this reference material allowed for the reliable identification of low levels of pretubulysin in culture extracts of *A. disciformis* An d48 by comparison of retention time and MS/MS fragmentation patterns upon HPLC-HRMS analyses, supporting its intermediacy in the pathway. The compound was tested on HL-60 cells to compare its cytotoxicity to that of a Tup C2-desmethyl pretubulysin variant, as well as to tubulysins A and D (**Table C-1**). Tubulysin D is known to be the more toxic than tubulysin A, with an approximately 6-fold higher potency.[32] As predicted from the earlier SAR studies, pretubulysin retained good cytotoxic activity in comparison to those of tubulysins A and D (3- and 5-fold lower cytotoxicity, respectively), while removal of the Tup C2-methyl group led to a 13-fold relative reduction in cytotoxicity. Based on these results, the cytotoxic activity of tubulysins A and D, and pretubulysin and Tup C2-desmethyl pretubulysin was evaluated against two more cell lines, L929 (mouse connective tissue fibroblast) and U937 (human histiocytic lymphoma).

Table C-1. *Cytotoxicity of tubulysins and analogues evaluated by MTT assay.*

IC_{50} (ng mL^{-1})				
Cell line	**Tubulysin A**	**Tubulysin D**	**Pretubulysin**	**Tup C2-desmethyl Pretubulysin**
HL-60	0.01	0.006	0.03	0.39
L929	0.19	0.015	6.5	74
U937	0.003	0.0004	0.08	0.41

Both pretubulysin and Tup C2-desmethyl pretubulysin exhibited activity against both cell lines, albeit reduced relative to tubulysin A, with pretubulysin being reproducibly the more potent. These results reveal several important SAR results, in addition to previously reported data. Notably, the good potency of pretubulysin confirms that neither the *N*,*O*-acetal nor the

acetoxy functionality of Tuv are necessary for cytotoxicity, although there is a modest reduction in activity (ca. 10-fold) relative to a previously characterized analogue which beared the acetoxy group. Comparison of the data obtained on pretubulysin and its Tup C2-desmethyl analogue reveals that the methyl group in Tup also provides an approximately 10-fold enhancement in biological activity. In conclusion, it was demonstrated that the complex structure of the tubulysins can be significantly simplified without a dramatic drop in the biological activity. Pretubulysin, which was identified as a direct biosynthetic precursor of the tubulysins, is less reactive compared to tubulysin A and D, but retains sub-nanomolar activity. Due to the absence of both the acetoxy and the *N*,*O*-acetal substituents, pretubulysin is much easier accessible by chemical synthesis than the elaborate tubulysins. These findings should aid in future efforts to design simplified, yet highly potent analogues of the tubulysins for evaluation as anticancer agents.

As the synthesis of novel tubulysin analogues still remains a laborious procedure, the known tubulysin producing myxobacterial strains *A. disciformis* An d48 and *Cystobacter* sp. SBCb004 were screened for new tubulysin derivatives. By analyzing HPLC-HRMS/MS fragmentation patterns, 23 novel tubulysins derivatives were identified revealing new aspects of the biosynthesis and supporting earlier proposals. The meanwhile achieved isolation of pretubulysin A and its structure elucidation by NMR delivered final proof for the previous structural postulations. The predicted structures of the novel tubulysin derivatives were established based on HRMS/MS data and show that several steps in the tubulysin biosynthetic pathway do not go to completion. The presence of pretubulysins, and compounds **12**, **13**, **19**, **20** and **23**, **24**, **30**, **31** (for numbering see **Chapter 2**) missing one or more of the oxidation and acylation reactions, demonstrates that the post-assembly line modifications are not perfectly efficient. This observation is consistent with the absence of genes coding for obvious tailoring functionalities (oxidases and perhaps an acyltransferase) within the biosynthetic gene 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, giving rise to the *N*-desmethyl compounds (**14**–**18**) and (**25**–**29**). Apparently, the lack of *N*-methylation of the Tuv building block somehow induces skipping of some, if not all, of the reductive steps carried out by the adjacent PKS module. However, these enzymatic activities seem to function occasionally, as the corresponding hydroxyl (**15**, **26**) and methylene (**16**, **27**) analogues are produced at low levels. Processing by post-assembly line enzymes at Tuv C5 occurs to some extent, and seems to be more efficient when full reduction has taken place at

the adjacent C12 position. This mechanism is inconsistent, however, with the discovery *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 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.[122] Unfortunately the yields of these novel tubulysin variants were too low to allow for their isolation and the determination of their respective cytotoxic activities.

It is intriguing to consider if 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'.[123] Indeed, the production of sometimes large families of metabolites which differ in modifications of their core structures appears to be a general feature of myxobacterial secondary metabolism, in which post-assembly line processing is not very abundant.[124] However, relevant information on the biological roles of the various tubulysins for the producing organisms is lacking so far. The fact that the PKS-NRPS can process a range of structurally divergent intermediates encourages the prospect to modify the assembly line by genetic engineering to produce further non-natural tubulysin variants. Similar observations have been made for the epothilone assembly line, from which multiple epothilone derivatives have been generated by directed modification.[125] In the meantime, the novel analogues can be targeted for total synthesis, in order to obtain sufficient material to extend the ongoing structure-activity studies of the tubulysin metabolites.

The comparison of the tubulysin biosynthetic gene clusters in An d48 and SBCb004 revealed the presence of two conserved open reading frames (ORFs), *orf17* and *orf18*, downstream of the last PKS gene *tubF*. The ORFs further downstream did not show any correlation between the strains. *Orf17* and *orf18* both encode patatin-like proteins, which show high mutual sequence homology (in the range of 60% identity and 70% similarity) and suggest evolution by gene duplication. Patatin-like proteins are known to exhibit lipid acyl hydrolase activity,[126] a function not apparently required in the tubulysin pathway. However,

it is worth mentioning that 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,[127] as well as in the leupyrrin gene cluster of *Sorangium cellulosum* So ce690,[128] suggesting that these enzymes may play an unexpected role in natural product biosynthesis. The inactivation of *orf18* in An d48 resulted in a mutant that almost abolished tubulysin D production but accumulated an unknown compound instead, that was not identified in the wild type. This compound was isolated and structure elucidation by 1D- and 2D-NMR experiments identified it as a glycerol ester derivative of tubulysin D. The thereby deduced assumption that the patatin-like proteins encoded downstream of *tubF* are responsible for cleavage of the ester bond between the *C*-terminal carboxy group of tubulysin D and glycerol was verified by heterologous expression and purification of the patatin-like proteins and *in vitro* incubation with tubulysin D glycerol ester. This procedure resulted in the release of tubulysin D. So far it remains unknown how the glycerol moiety is attached to the *C*-terminal carboxy group of tubulysin D in the first place. Due to the site of attachment it is conceivable that this reaction would be catalyzed by the thioesterase (TE) domain, which normally releases the product under hydrolysis or macrolacton/macrolactam formation from the assembly line. However, sequence analysis of the thioesterase domain located at the *C*-terminus of TubF did not reveal obvious differences in amino acid sequence when compared to TE domains from other assembly lines. The only noticeable feature was the fact that the tubulysin TE domain was slightly longer. Besides the so far incompletely resolved biosynthesis of the tubulysin D glycerol ester, also the biological function of this compound remains unknown up to date. Assumptions that the glycerol moiety might be further acylated with fatty acids and thereby serve for membrane associated storage of tubulysin D could not be confirmed experimentally.

The fact that the producer strains dedicate special enzymes to the cleavage of a previously attached glycerol ester to yield tubulysins A and D, is astonishing since the same product could be obtained by hydrolytic release from the TE. The maintenance of this genetic organization suggests a so far not comprised biological function of the tubulysin glycerol esters.

Despite the potent *in vitro* activity of tubulysins against various cancer cell lines, their therapeutic use is hindered due to severe toxicity. Therefore different approaches for targeted delivery of tubulysin derivatives are pursued in order to improve their therapeutic benefits. At the moment tubulysin folate conjugates,[129;130] polymeric tubulysin-peptide nanoparticles,[131]

and tubulysin analogue-dendrimer conjugates[132] are under investigation. For all these approaches the tubulysin core is covalently linked to its carrier via the *C*-terminus of the molecule. For the connections various tubulysin derivatives and linker chemistries are used. In this context the glycerol moiety might present a promising alternative for linkage to a macromolecular carrier, since the ester bond can most likely be cleaved by human esterases to release the cytotoxic tubulysin agent *in situ*.

2. Biosynthesis of the quinoline alkaloids aurachins in *Stigmatella aurantiaca* Sg a15

The aurachins are quinoline alkaloids produced by the myxobacterium *Stigmatella aurantiaca* Sg a15. They exhibit a broad range of biological activities including antimicrobial[133] and anti-plasmodial[134] properties. The targets underlying these properties are the mitochondrial respiration complexes I (aurachin A and aurachin B) and III (aurachin C and aurachin D)[46] as well as the cytochrome b_6/f-complex,[45] both of which are inhibited by the aurachins due to their structural similarity to vitamin K. The structure of the aurachins is composed of an oxygen substituted 2-methylquinoline core and a farnesyl side chain. Aurachins can be classified into two structural types based on the position of the farnesyl substitution. Aurachin A and aurachin B are A-type aurachins with the farnesyl residue at position C4 whereas the C-type aurachins, aurachin C and aurachin D, feature the farnesyl side chain at position C3. A previous study demonstrated the involvement of a type II PKS in formation of the 2-methyl-4-hydroxyquinoline core.[86] The various aurachins are connected via a biosynthetic sequence from aurachin D to aurachin A involving diverse oxidation reactions and an exceptional 3,4-migration of the farnesyl side chain. The biochemical steps involved in the biosynthesis of the quinoline core, as well as the downstream oxidations and the alkyl migration display atypical reactions remote from textbook biochemistry. Both *in vivo* inactivation experiments and *in vitro* characterization of particular enzymes revealed an unprecedented starter unit priming mechanism (**Chapter 4**). The identification of the complete set of genes required for biosynthesis up to the most highly elaborated aurachin A was accomplished by *in vivo* inactivation experiments in Sg a15 (**Chapter 5**). Based on these findings it is now possible for the first time to postulate a comprehensive model for the biosynthesis of all aurachins (**Figure C-2**).

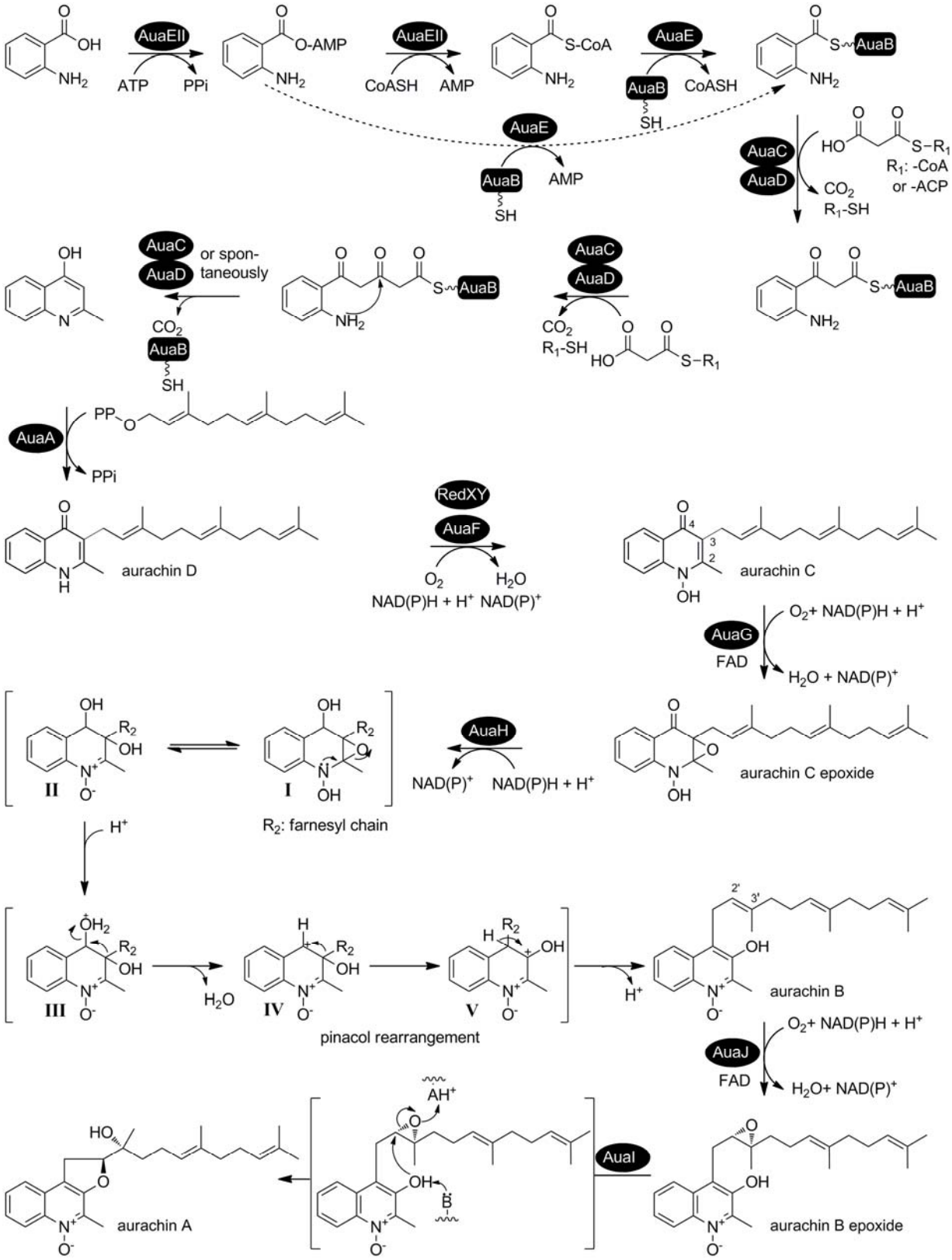


***Figure C-2**. Proposed biosynthetic pathway of aurachins.*

Aurachin biosynthesis is initiated by the priming of an anthranilate starter unit to the sole ACP (AuaB) encoded within the *aua* genes. This proceeds via a novel non-acetate priming

strategy never before observed in thiotemplated biosynthetic machineries. Besides *holo*-AuaB two homologues of aryl:CoA ligases, AuaEII and AuaE, catalyze substrate activation and ACP loading, respectively. AuaEII activates anthranilate in presence of Mg^{2+}, ATP and CoASH via an adenyl (AMP) intermediate to anthraniloyl-CoA, which in turn functions as the substrate for AuaE mediated transfer of anthranilate to *holo*-AuaB by transthioesterification to complete starter unit priming. There is experimental proof that AuaE can also transfer the anthraniloyl-AMP intermediate, formed by AuaEII, to *holo*-AuaB, although with approximately 1000 fold lower efficiency. Based on the homologies of AuaC and AuaD to condensing enzymes it is expected that anthraniloyl-AuaB is extended by two malonyl building blocks, supplied either as malonyl-CoA or malonyl-ACP (-AuaB). The quinoline core is assumed to be formed by nucleophilic attack of the aromatic amine at the β-carbonyl carbon of the twice extended polyketide intermediate, most likely while it is still attached to AuaB and presumably assisted by the AuaC/AuaD complex. Subsequent hydrolytic release and decarboxylation of the terminal carboxyl group yields 2-methyl-4-hydroxyquinoline. Alternative scenarios based on hydrolytic release of the polyketide intermediate from the carrier protein and spontaneous heterocycle formation either prior to, during, or after decarboxylation, cannot be excluded as long as the reaction has not been thoroughly investigated *in vitro*. Complete *in vitro* validation of these steps has been hampered so far by difficulties in obtaining soluble AuaC protein despite numerous efforts. Aurachin biosynthesis proceeds with the membrane-bound prenyltransferase AuaA catalyzing the transfer of a farnesyl moiety from farnesylpyrophosphate (FPP) to position C3 of 2-methyl-4-hydroxyquinoline to yield aurachin D.[135] The Rieske oxygenase AuaF accounts, in conjunction with a still unidentified reductase partner, for the transformation of aurachin D to aurachin C by *N*-hydroxylation. The migration of the farnesyl side chain from C3 to C4 involves monooxygenation, reduction, and subsequent elimination of water to trigger the alkyl rearrangement. The involvement of an oxidation reaction has been established previously by exhaustive feeding experiments. Feeding of [4-$^{13}C^{18}O$] aurachin C, that was isolated beforehand from cultures of Sg a15 fed with [1'-$^{13}C^{18}O_2$] anthranilic acid, to a fresh culture of Sg a15 and subsequent isolation of the final product aurachin A revealed that the oxygen label was lost during the rearrangement.[136] Assuming that the oxygen at position 4 is eliminated in the rearrangement reaction, the introduction of another oxygen atom in the course of the farnesyl migration is required to account for the two oxygen atoms found in aurachin B and aurachin C. Further feeding experiments with $^{18}O_2$ oxygen revealed that the

oxygen substituent at position C3 of aurachin A is derived from molecular oxygen and not from the carbonyl oxygen originating from the anthranilate starter unit.[136] In more detail: AuaG is suspected to catalyze epoxidation of the 2,3 double bond to form an aurachin C 2,3-epoxy intermediate. Subsequent reduction of the C4-keto group by AuaH yields an aurachin C 2,3-epoxy,4-hydroxy intermediate (**I**), for which a tautomeric aurachin C *N*-oxide 3,4-diol intermediate (**II**) can be postulated. Upon protonation of the 4-hydroxy group this intermediate (**III**) is expected to eliminate water, yielding a C4-carbocationic intermediate (**IV**). **IV** is expected to be the direct precursor undergoing the actual alkyl migration from C3 to C4 (**V**) to give the A-type product aurachin B after deprotonation. From a chemical point of view this series of reactions is akin to a pinacol rearrangement. The sequence of reactions is expected to be assisted by the AuaH active site as it is very unlikely to drain of spontaneously without a favorable environment. This reaction mechanism strongly differs from the one postulated by Höfle and Kunze who predicted a migration of the farnesyl side chain to the C4 carbonyl prior to elimination of water, which should be brought about by a push-pull effect between the nitrogen and the carbonyl group.[136] However, with our additional knowledge about the involved functionalities of AuaG and AuaH, the newly postulated mechanism seems to be more likely. For the last biosynthetic conversion AuaJ is expected to catalyze epoxidation of the 2',3' double bond of the aurachin B farnesyl side chain. This 2',3'-epoxy aurachin B intermediate is supposed to undergo intramolecular nucleophilic attack of the 3-hydroxy group at C2', leading to formation of the dihydrofuran heterocycle. The simultaneous opening of the epoxide yields a 3'-hydroxy group and thereby the final product aurachin A. This hypothesis is supported by the feeding studies with ^{18}O-labelled oxygen,[136] although the proposed intermediate III has not been discovered from the cell extracts, either due to its instability in complex solutions, or due to rapid enzymatic turnover.

The completion of the puzzle of aurachin biosynthesis affirmed once more the outstanding features of this biosynthetic pathway. The anthranilate priming is reminiscent of EncN-mediated loading of benzoate in enterocin biosynthesis,[137] however involvement of two distantly located CoA ligase homologues has so far been unprecedented. CoA ligases and A domains of NRPSs, that both belong to the ANL superfamily of adenylating enzymes, carry out the synthesis of acyl- or aryl-thioesters in a two-step reaction. The first half-reaction consists of the adenylation of the carboxylic group of the substrate to form an acyl- or aryl-AMP intermediate. In the second half-reaction, the thiol group of CoA or of the phosphopantetheinyl arm of the PCP attacks the acyl carbon with subsequent elimination of

AMP. Biochemical characterization of recombinant AuaE and AuaEII revealed that only AuaEII was capable to form anthraniloyl-CoA. The inactivity of AuaE could be justified by a C-terminal truncation of the enzyme sequence compared to homologous enzymes, leading to a lack of a conserved lysine residue, which is known to be essential in the adenylate forming half-reaction, and therefore accounts for the enzymes inactivity. The absolute requirement of AuaE for aurachin biosynthesis established by *in vivo* inactivation of the gene in Sg a15 could be rationalized by *in vitro* experiments demonstrating that AuaE mediates anthranilate priming of *holo*-AuaB from either anthraniloyl-AMP or anthraniloy-CoA. The biochemically relevant pathway was determined by kinetic studies, which revealed a 1000 fold higher K_m value for anthraniloyl-AMP compared to anthraniloyl-CoA. AuaE is the first enzyme described in the ANL superfamily that catalyzes a transthioesterification reaction for the loading of a carrier protein involved in polyketide biosynthesis. AuaE, which has obviously evolved from an aryl-CoA ligase ancestor, is functionally equivalent to PKS ATs which belong to the unrelated α/β hydrolase superfamily.[138] Such convergent evolution may be advantageous as existing substrate specificity can be maintained. This view is supported by analysis of the 10-amino acid (aa) code that predicts specificity of NRPS A domains.[61;62] The specificity conferring aa-residues of AuaE are in very good agreement with those extracted from known bacterial anthranilate-activating enzymes.

Polyketide biosynthesis has many parallels to fatty acid biosynthesis. Phylogenetic studies have revealed complex evolutionary relationships among PKSs themselves and prokaryotic and eukaryotic fatty acid synthases.[139;140] The type II fatty acid synthetic pathway found in bacteria and plants consists of a set of discrete proteins, that catalyze discrete steps in chain elongation.[141] In this pathway three different types of condensing enzymes, called β-keto-ACP synthases I, II and III, also known as 3-oxoacyl-(acyl-carrier protein) synthases I, II, and III, or FabB, FabF, and FabH respectively, are involved which perform different condensation reactions on the route to fatty acids.[142] Polyketide synthases show highest homology to KASII (FabF) enzymes. AuaC and AuaD were originally identified as a putative KSα/KSβ pair of type II PKS, presenting the first instance of such a system to be reported from a Gram-negative bacterium.[86] This is still exceptional, as to date only one further example for a type II PKS from a Gram-negative bacterium has been described (*ant*-gene cluster from *Photorhabdus luminescens*).[87] However, the similarities of AuaC and AuaD to KSα and KSβ from actinomycetes are not convincing and based on recent sequence updates in the public sequence databases, the closest homologues identified for AuaC and AuaD were found to be a

pair of proteins encoded by neighbouring genes in *Candidatus Kuenenia stuttgartiensis* CAJ74368 (43% identity/63% similarity) and CAJ74369 (32% identity/49% similariy) respectively, both annotated as β-keto-ACP synthases (KASII). These homologues are part of a cluster, which based on its annotations (CAJ74365, similar to isocitrate dehydrogenase; CAJ74366, ACP; CAJ74367, similar to hydroxymyristoyl ACP dehydratase; CAJ74370, similar to phytoene dehydrogenase), seem to be involved in fatty acid metabolism. In order to assign the phylogenetic relationships of AuaC and AuaD sequence alignments with members of the KASII family and KSα/KSβ pairs have been performed and a phylogenetic tree was generated (**Figure C-3**). KSαs conserve the Cys-His-His catalytical triad identified for KASII,[142] whereas in the KSβs the respective residues are replaced by Gln, Asp, and Arg. The sequence of AuaC shows perfect conservation of the Cys-His-His triad, whereas the AuaD sequence features Thr, Ser, and Glu at the respective positions, thereby differing from the residues usually found in the sequences of KSβs. Remarkably CAJ74369 (Val, Asn, and Glu) and RemB[143] (KSβ from resistomycin biosynthesis; Arg, His, and His) neither show conservation of these residues. The conserved Gln residue, which replaces the active site Cys in KSβs, has previously been postulated to be involved in acetate starter unit priming by promoting the decarboxylation of a malonyl building block. [78;144] This role was recently challenged by the inspection of the crystal structure of the actinorhodin KSα/KSβ heterodimer, attributing this activity rather to the KSα.[92] Thereby the duty of the KSβ is assumed solely in complex formation with KSα to form an amphipatic tunnel at the heterodimer interface that can accommodate the growing polyketide intermediate.

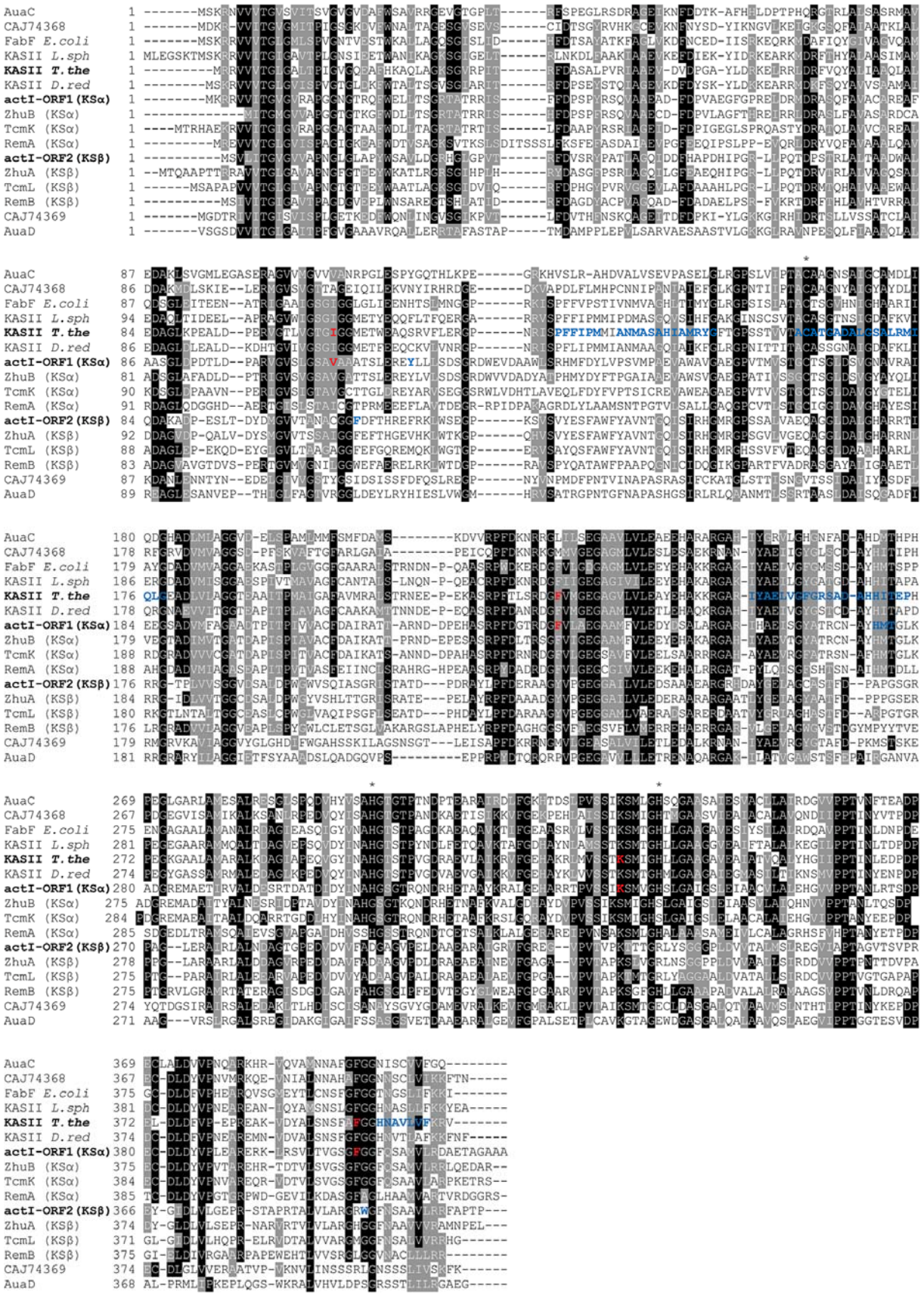


Figure C-3. *Sequence alignment of AuaC, AuaD and their closest homologues with KASII from different species and KSα/KSβ pairs from actinorhodin (act), R1128 substances (zhu), tetracenomycin (tcm) and resistomycin (rem) biosynthetic pathways. Identifiers of sequences with reported crystal*

structures that have been included in the analysis are shown in bold. Active site residues are indicated by asterisks. Further catalytically important residues are marked in red. Residues that have been shown to be involved in dimerization are marked in blue.

In contrast KASII enzymes form homodimers with each monomer capable of growing an acyl chain into a tunnel created by the dimer interface.[145] Structural features accounting for dimerization have been deduced from the crystal structures of both KASII and KSα/KSβ. For KASII it has been described that mainly hydrophobic residues of the loops between α2 and β3 and between β8 and α13, the helices α6, α7, and α8 and the strands β8 and β12 are involved in dimerization (residues marked in blue in the sequence of *Thermus thermophilus* HB8 (*T. the*) KASII).[145] A principle contact is thereby made by a Met or Ile (part of a highly conserved His-(Met/Ile)-Thr motif) that is inserted into a hydrophobic pocket of the other monomer. This His-(Met/Ile)-Thr motif is also conserved for AuaC and the members of the KSα family. The members of the KSβ family on the other hand show no conservation of this motif. The crystal structure of the actinorhodin KSα/KSβ heterodimer revealed that the KSα employs this motif to make contact to the KSβ, whereas the KSβ uses a Trp from a different loop to connect to the KSα.

Current hypothesis is that the KSβs arose from an ancient KSα gene by duplication and subsequent evolution of the KSβ into a structural partner for the KSα resulting in a heterodimer that is able to shield the binding pocket for a highly reactive polyketide intermediate up to a defined maximum size.[146] The fact that KSα and KSβ sequences form distinct clades in phylogenetic analyses suggests that all KSα/KSβ heterodimers are descendants of a common ancestor.[146]

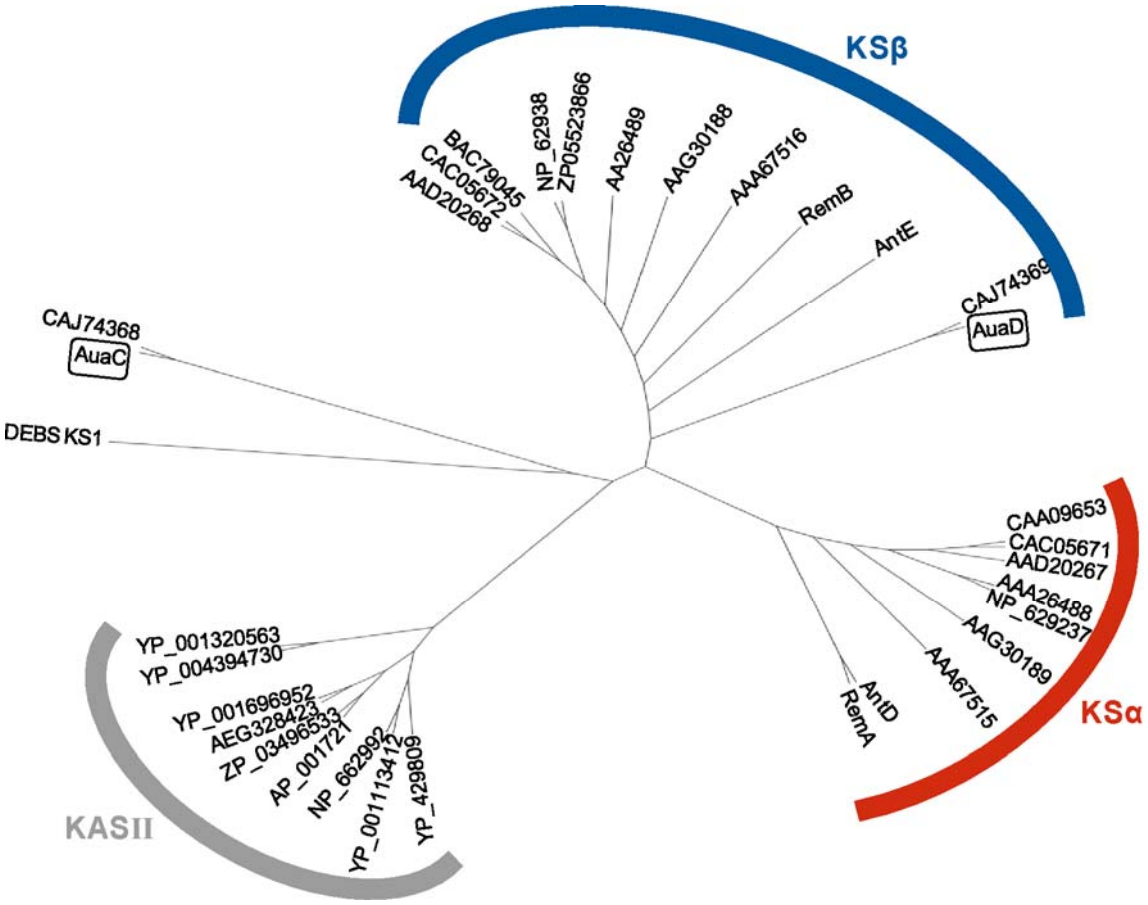


Figure C-4. *Phylogenetic relationships of AuaC and AuaD (framed) with selected members of the KASII (marked in grey), KSα (marked in red), and KSβ (marked in blue) families, inferred by Bayesian estimation. Sequence of DEBS KS1 has been used as outgroup. The tree is unrooted.*

Phylogenetic analysis of sequences of KSα/KSβ pairs and KASIIs along with sequences of AuaC, AuaD, CAJ74368, and CAJ74369 (**Figure C-4**) revealed as expected a clustering of KSα, KSβ, and KASII sequences, respectively. Both AuaD and CAJ74369 fall into the KSβ clade although they form the most distant branch. It has been noted before that KSβ sequences show a higher degree of divergence than KSα sequences, probably in response to formation of products of different chain length.[147] This is supported by the fact that a wider branching of KSβ sequences compared to the KSα sequences is observed. AuaC and CAJ74368 cluster neither with the KSαs from actinobacteria nor with the KASII sequences. Based on these results it can be envisioned that AuaC (and CAJ74368) are not derived from actinobacterial KSαs but stem from an independent convergent evolution originating from KASII. In order to clarify the exact roles of AuaC and AuaD in the formation of the 2-methyl-4-hydroxyquinoline *in vitro* experiments reconstituting its biosynthesis are absolutely required. This approach has been hampered so far by difficulties in obtaining recombinant AuaC protein due to solubility problems. These problems could not be resolved

neither by the use of numerous fusion tags aiming at the improvement of solubility, nor by the screening of several expression conditions. Even attempts to co-express AuaC with its designated partner AuaD were not met with success.

In the transformation of aurachin D to aurachin C, AuaF marks another precedent for the rare cases where Rieske oxygenases are involved in secondary metabolite biosynthetic pathways. Known examples from the literature comprise the biosyntheses of ambruticin,[148] jerangolid,[148] leupyrrin,[128] pyrrolnitrin,[149] streptorubin,[150] and metacycloprodigiosin.[151] The fact that 4 out of the 7 known cases are derived from myxobacteria, further highlights their exploitation of exceptional biochemical reactions for secondary metabolite biosyntheses. Besides the Rieske oxygenase AuaF two FAD-dependent monooxygenases are involved in the late stage oxidations transforming aurachin C via aurachin B to aurachin A. Particularly the conversion of aurachin C to aurachin B, that includes epoxidation of the quinoline core (mediated by AuaG) and subsequent reduction of the C4-carbonyl group (mediated by AuaH) which after elimination of water culminates in the migration of the farnesyl moiety, is unprecedented in secondary metabolite biosynthesis. The investigation of the biochemical details and the putative cooperation of AuaG and AuaH present a very interesting topic for further investigations. The transformation of aurachin B to aurachin A, which is expected to proceed via double bond epoxidation of the farnesyl side chain followed by an intramolecular epoxide opening with simultaneous cyclization is expected to involve AuaJ and AuaI and is both chemically and enzymatically reminiscent of polyether ring formation during monensin biosynthesis.[152] The successive character of tailoring reactions seems to be strict, since the *auaF* mutant accumulated only aurachin D but no aurachin B and aurachin A derivatives missing the *N*-oxide function were found. This indicates that the *N*-hydroxyl/*N*-oxide function might be crucial for acceptance by the downstream tailoring functionalities.

All in all biosynthesis of the aurachins is striking in that it requires the concerted expression of many functionalities encoded at various genetic loci. Besides loci I-III the expression of the previously identified DAHP synthase[153] and presumably the anthranilate synthase complex (TrpE/TrpG) from primary metabolism[154] have to be coordinated. The DAHP synthase gene is part of the operon coding for the biosynthetic machinery of the siderophore myxochelin.[100] This may implicate iron homeostasis in the regulation of aurachins biosynthesis. Additional enzymes such as the reductase partner of AuaF and a malonyl-CoA:ACP transacylase putatively required for 2-methyl-4-hydroxyquinoline

formation remain unidentified. Such split cluster organization raises the question of how these genes are regulated to function together which has yet to be elucidated.

The aurachin gene cluster organization in different loci is most likely the result of successive acquisition by horizontal gene transfer and adaptation of the functionalities that were originally dedicated to other pathways. In support of this hypothesis, biosynthetic genes of aurachins on loci I-III do not have orthologues in *S. aurantiaca* DW4/3-1 (non-producer of aurachins) which is closely related to Sg a15, albeit both genomes share high synteny. In contrast, regions flanking each locus are well conserved in both genomes. In summary, the split organization may not only represent an advantage to facilitate differentiated regulation, but also facilitate acquisition or loss of tailoring genes to generate new aurachin structures. In the future, it will be instructive to study the regulatory mechanism and evolution aspect of aurachins biosynthesis in order to contribute to the general understanding of secondary metabolism.

3. Biosynthesis of the quinoline HHQ in *P. aeruginosa* as potential target for therapeutic interference to attenuate virulence

Pseudomonas aeruginosa is one of the most clinically important pathogens, causing a multitude of nosocomial infections, especially in cystic fibrosis (CF) patients, where it is the main perpetrator of the shortened lifespans of patients.[106] The evoked lung infections can be both acutely life threatening and cause long-term lung damage due to lingering colonization. Therapy of *P. aeruginosa* infections is complicated by widespread antibiotic resistance [155] and the formation of small colony variants and biofilms that contribute to persistence. Biofilms display an unspecific antibiotic resistance mechanism based on greatly reduced penetration of the antibiotic agent into the biofilm matrix. The formation and maturation of these biofilms,[156] as well as the production of the bulk of the many different virulence factors excreted by *P. aeruginosa*, like alkaline proteases, elastase, lipase, lectins A and B, hydrogen cyanide and pyocyanin,[157;158] to name only a few, is controlled by small molecule intercellular signalling, referred to as quorum sensing (QS). QS is based on the production and release of small signalling molecules, called autoinducers, that increase in concentration as a function of cell density and activate corresponding transcriptional regulators after a threshold concentration has been reached.[102] QS is widespread among bacteria, though the nature of specific autoinducers is quite diverse. Gram-positive bacteria mainly use processed oligopeptides, whereas Gram-negative bacteria use acylated homoserine lactones.[103] In *P. aeruginosa*, three different QS systems are known. The *las* and *rhl* systems use acyl-homoserine lactone (AHL) autoinducers and belong to the LuxI/LuxR type systems. The third system is rather unique and restricted to particular *Pseudomonas* and *Burkholderia* strains. Here, 2-alkyl-4(1*H*)-quinolones (AQ) autoinducers such as 2-heptyl-3-hydroxy-4(1*H*)-quinolone (*Pseudomonas* quinolone signal, or PQS) and its direct precursor 2-heptyl-4(1*H*)-quinolone (HHQ) are used.[104] These systems are interconnected with each other and arranged in a certain hierarchy. To sum it up very briefly and simply, the *las* system positively regulates both the *rhl* and *pqs* systems, whereas the *pqs* system positively regulates the *rhl* system, which in turn negatively regulates the *pqs* system (see **Figure** C-**5**). [159;160]

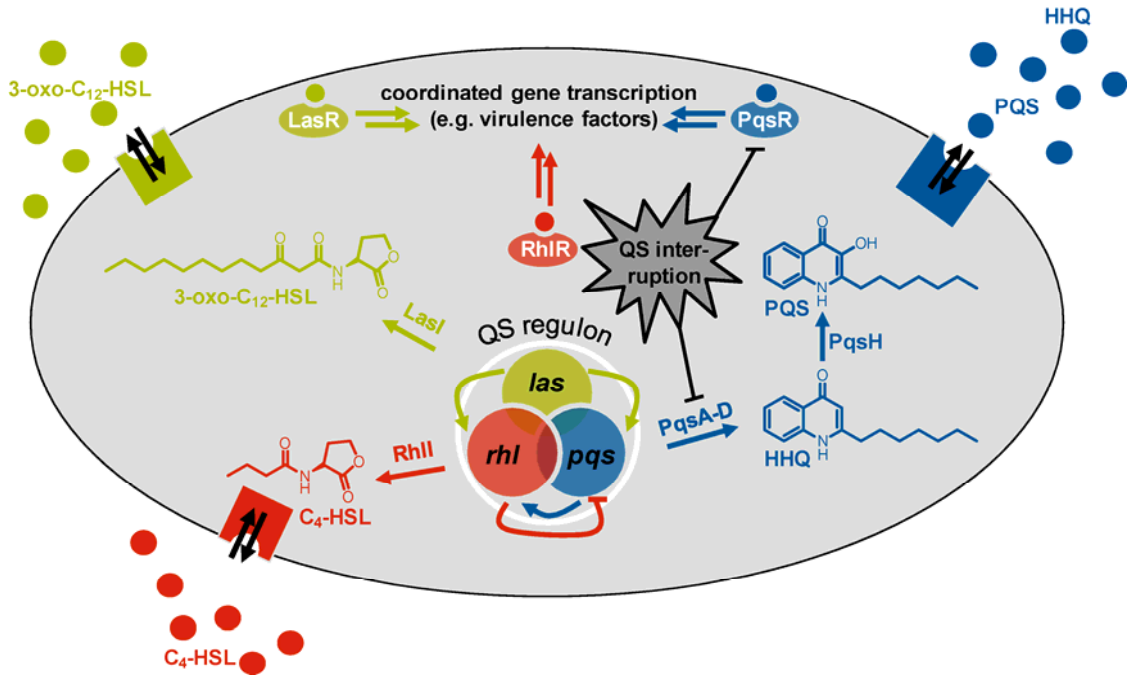


Figure C-5. *Simplified hierarchy of the* las*,* rhl*, and* pqs *quorum sensing systems in* P. aeruginosa *including chemical structures of the autoinducers used by the respective systems. Possible targets for interference with the* pqs *system to either inhibi biosynthesis of the autoinducer or its binding to the receptor are also indicated.*

Transcriptional and proteomics studies have shown that around 400 genes, corresponding to 6-7% of the total genome, are regulated by these three QS systems.[161] The *las*, *rhl*, and *pqs* systems regulate the timing and production of multiple virulence factors and biofilm formation. In several models, *las* and *pqs* signalling deficient mutants were unable to form structurally normal biofilms.[156;162] These effects make the QS system a highly attractive target for drug development through interference with *P. aeruginosa* pathogenicity and biofilm formation. Compounds capable of inhibiting QS signaling could represent the next generation of anti-infective therapeutic agents.[163] It is hoped that an attenuation of bacterial virulence (i.e., inhibition of virulence factor production) without necessarily threatening the single cell would ease pressure on the evolution of resistance mechanisms.[164] The attenuation of virulence may also enable the host`s immune system to eliminate the pathogen on its own. Further, an adjunction with classical antibiotics will most likely exhibit synergistic effects. However, the development of bacterial resistance to QS signalling inhibitors can only be evaluated once such compounds are in broad clinical use. Such drugs might also be of major therapeutic value when used to coat of medical devices that tend to be colonized by QS-signalling pathogens to prevent biofilm formation.

The general validity of QS as target for anti-infective therapy in *P. aeruginsosa* is supported by several animal infection models in which QS-deficient mutants of every

respective system exhibit attenuated virulence levels compared to *P. aeruginosa* wild type.[158;165;166] The assumption that these three QS systems also play a role in *P. aeruginosa* colonization and chronical infection of CF patients is supported by the fact that all three autoinducer molecules 3-oxo-C_{12}-HSL, C_4-HSL and PQS have been detected in sputum of CF patients in significant concentrations.[167;168] These results seem to contradict the finding that QS-deficient *P. aeruginosa* mutants are often isolated from infected CF patients.[169] Thereby, the importance of QS for *P. aeruginosa* during CF infections may be brought into question. The enrichment of QS-deficient mutants during chronic infections in CF patients can be explained by the phenomenon of "social cheating".[170] Although the bacterial population benefits from the concerted gene transcription mediated by QS, the single cell incurs an enormous metabolic burden. Individuals that cease the production of public goods due to a lack of QS signal transduction, called cheaters, can still benefit from the cooperative actions of the remaining population and display a gain in individual fitness. However, "social cheaters" are dependent on a sufficient fraction of QS participants to maintain the overall population benefits, a situation also observed in infected CF patients, where the QS deficient mutants never completely replace the wild type. Therefore, complete interruption of *P. aeruginosa* QS in chronic CF infections by drug therapy should be beneficial for eradication.

In general, there are several strategies for QS signalling interference. Firstly, inhibition of enzymes that catalyze autoinducer synthesis and inhibition of the receptors that bind the autoinducers and mediate the downstream processes are together known as "signal interference". "Quorum quenching" represents the degradation of autoinducer molecules by dedicated enzymes such as AHL acylases and AHL lactonases.[103] While enzyme-dependent "quorum quenching" is, for obvious reasons, likely not a viable therapeutic strategy, both "signal interference" strategies have great potential for the development of small molecule-derived inhibitors.

Biosynthesis of AHLs has been elucidated in detail. They are synthesized by LuxI homologues from acyl-ACP derivatives and *S*-adenosyl-L-methionine (SAM).[103] The most promising strategy for LuxI-mediated AHL biosynthesis inhibition might therefore be to utilize SAM analogs. However, it is questionable if this approach can provide valuable drug candidates for human therapy, since SAM is vital for several processes in both bacterial and human cells and may lead to both rapid emergence of bacterial resistance and severe side effects in humans. The development of receptor antagonists based on AHL derivatives might

be hampered as long as they are recognized as substrates for the above mentioned AHL acylases and AHL lactonases, which are expressed by a broad range of Gram-negative bacteria. Resistance to small molecule inhibitors are more likely to emerge if a broad spectrum of bacteria is affected. Aiming at selective QS interference against *P. aeruginosa*, it should therefore be most promising to target the *pqs* system, which employs HAQ autoinducer molecules and thereby eliminates the aforementioned problems. A first proof of principle that the *pqs* system can be targeted by small molecules was obtained from a mouse infection model, where animals infected with *P. aeruginosa* were treated with halogenated anthranilic acid derivatives inhibiting PQS biosynthesis. This treatment led to a significant increase in survival in comparison to the control group.[171] Knowledge of HAQ biosynthesis compared to that of the AHLs has so far been only rudimentary. Further exploration of this target requires profound knowledge of PQS formation in *P. aeruginosa*. It is known that *pqsA-D* are absolutely required for HHQ biosynthesis,[105] encoding an anthranilate:coenzyme A (CoA) ligase (*pqsA*) and three β-ketoacyl-acyl carrier protein synthase III (KAS III) homologues.[172] An additional gene (*pqsH*) located apart from *pqsA-D* is responsible for the hydroxylation of HHQ to form PQS. Feeding studies *in vivo* have demonstrated that HHQ most likely arises from "head-to-head" condensation of an anthraniloyl precursor and a β-keto fatty acid derivative.[173] However, the details of the enzymatic mechanism of this reaction and the nature of the β-keto fatty acid remained elusive. The formation of 2,4-dihydroxyquinoline (DHQ), another secondary metabolite of *P. aeruginosa* has been reconstituted *in vitro* with recombinant PqsA and PqsD proteins.[174] Based on the structural similarity between DHQ and HHQ, it was rationalized that PqsD might be involved in a similar condensation reaction in HHQ biosynthesis. Indeed, it was shown that recombinant PqsD protein was capable of forming HHQ from anthraniloyl-CoA and β-ketodecanoic acid. Considering the molecular mechanism of HHQ formation by PqsD, two possible reaction mechanisms are conceivable. Either the reaction would proceed via an imine intermediate of anthranilate and β-ketodecanoate followed by decarboxylative Claisen condensation to form the heterocycle and release the final product from the enzyme, or the sequence of reactions would start with decarboxylative Claisen condensation of the β-keto fatty acid and anthraniloyl-CoA, followed by heterocycle formation via intramolecular condensation and elimination of water, presumably assisted by PqsD in anology to the reactions catalyzed by chalcone synthases.[175] By an elaborate experimental approach, it was demonstrated that the latter holds true.

The *in vitro* reconstitution of HHQ biosynthesis by means of recombinant PqsD protein formed the basis for the further development of an *in vitro* assay to screen for inhibitors of PqsD function. For this purpose, PqsD, anthraniloyl-CoA, β-ketodecanoic acid and increasing concentrations of the potential inhibitor were mixed, incubated for a defined time period, and HHQ formation quantified by HPLC-MS/MS. Comparison of HHQ levels formed in presence of various concentrations of putative inhibitors to that formed in control reactions without inhibitor allowed the determination of the half maximal inhibitory concentration (IC_{50}) of the tested compound. The feasibility of this *in vitro* approach was experimentally validated with well-known synthetic FabH inhibitors that were chosen on the basis of structural and functional homologies between FabH and PqsD.

The advantages of this *in vitro* assay are its uncomplicated read-out, since it bypasses the complex and interconnected QS networks in *P. aeruginosa* that would complicate initial hit identification *in vivo*. Further, it enables testing without the opportunistic pathogen, which permits a reduction of safety measures. The assay delivers results quickly and has the potential for adaptation to high-throughput screening (currently under development at our collaborators from Sanofi-Aventis). In addition, it is possible to test compounds with inherent antibiotic activity, such as FabH inhibitors, for inhibition of HHQ formation which is neither possible in *P. aeruginosa* nor in the established reporter gene coupled *in vivo* assays.[176] Though antibiotic activity of a QS inhibitor might not be desirable with regards to the development of bacterial resistance, a PqsD inhibition assay that allows the screening of compounds with inherent antibiotic activity in the initial compound hit identification phase is highly beneficial. This assay might therefore prove to be a valuable tool in the development of small molecule inhibitors of PqsD function.

4. Rhizopodin biosynthesis: Identification and characterization of a trans-AT PKS molecular assembly line

The steady rise of accessible bacterial whole genome sequences in combination with the facile *in silico* identification of NRPS and PKS biosynthetic gene clusters revealed for all so far investigated secondary metabolite (SM)-producing bacteria a much higher genetic potential for SM-biosynthesis then realized by production of the SMs known from the respective strains (**Figure C-6**).[19;20;22;23] At present, the identification and correlation of SMs from so far unassigned biosynthetic gene clusters represent a significant bottleneck that constrains the exploitation of the extensive potential bacterial secondary metabolome.

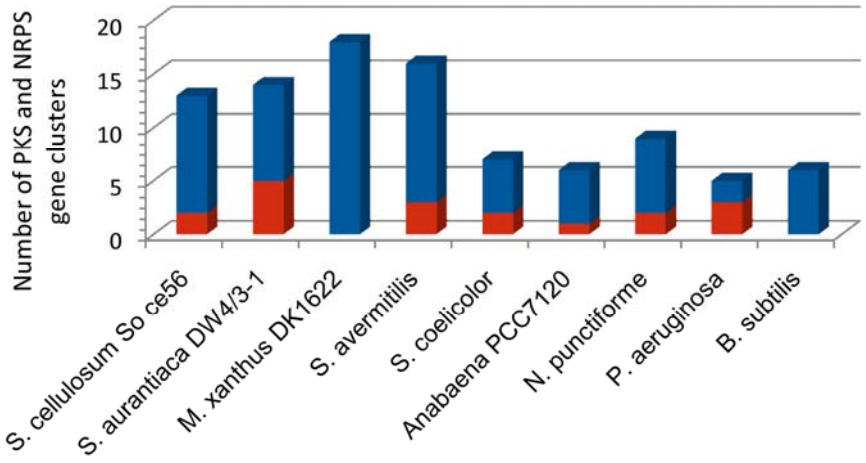


Figure C-6. *Genetic potential for secondary metabolite biosynthesis of selected fully sequenced prokaryotes. The bars indicate* in silico *identified PKS and/or NRPS biosynthetic gene clusters. The red fractions indicate secondary metabolites known to be produced by the strain at the time of genome sequencing (figure modified from Wenzel and Müller,* Nat. Prod. Rep.*, 2009,* ***26****, 1385-1407).*

In this context the myxobacterial strain *Myxococcus xanthus* DK1622 presents a prominent example. Prior to the availability of its genome sequence DK1622 was not known to produce any secondary metabolites at all. Only after in silico analysis of its genome it turned out that DK1622 holds a huge genetic potential for SM biosynthesis, encoding for a total 18 biosynthetic gene clusters of NRPS, PKS, and NRPS-PKS hybrid systems,[177] which corresponds to 8.5% of the total genome size. Until recently only 5 out of the potential 18 SMs from DK1622 were identified (the myxalamids, myxochelins, myxochromides A, myxovirescins, and DKxanthenes; all initially isolated from different myxobacterial strains), leaving 13 unassigned biosynthetic pathways. Transcriptomic and proteomic studies attested the presence of transcripts and/or proteins for all of those 13 assembly lines, thereby

implementing their activity, at least on a very low level.[178;179] The elaboration of an approach, wherein targeted gene inactivations were coupled with a global metabolite mining approach consisting of HPLC-HRMS analysis and data evaluation supported by statistical tools, culminated in the recent identification, isolation, and structure elucidation of the 6th SM from DK1622, called myxoprincomide (Cortina *et al.* in preparation). In addition, masses corresponding to SMs could be assigned to another two biosynthetic pathways. One can assume that most of the 10 remaining unassigned clusters also produce SMs, at least at very low levels, which have so far escaped detection for unknown reasons (e.g. low abundance, instability, extreme (a)polarity, bad ionization). The solely "manual", or more precisely visual, comparison of HPLC-HRMS data from extracts of a mutant versus a wild type (especially in the case of myxobacteria) is a very tedious process that may remain fruitless even for a skilled operator, due to the stunning complexity of the data and the huge differences in signal intensities that easily mask low abundant compounds. The implementation of statistical tools proved instrumental in the elimination of this constraint towards comprehensive comparative HPLC-HRMS data evaluation. Statistical supported HPLC-HRMS data evaluation requires several biological replicates to fade out random variations. The raw data are processed with an algorithm which, based on the presence or absence of isotopic patterns, is able to distinguish between "true" MS signals (molecular features) and background noise. Based on the combination of m/z values of molecular features and the respective retention time, so called buckets are created. These buckets present the input values for the statistical evaluation of the data sets. The ProfileAnalysis software interface (Bruker Daltonics) provides two independent methods for classification of these preprocessed HPLC-HRMS data sets, the principal component analysis (PCA) and the Student`s *t*-test. PCA is a mathematical projection technique designed to extract, display and rank the variance in a data matrix. The goal of this operation is to reduce the dimensionality of multivariate data sets without losing relevant information. Thereby PCA allows identifying the variables that are most influential to the variance of two data sets instead of being obliged to look at a larger number of variables in the original data matrix. In the so called loadings plot, every bucket calculated for the data set (represented by a small sphere) is spotted in a two-dimensional space spanned by PC1 and PC2 (the principal components that explain for the highest variance) based on its relation to the principal components. The further a bucket is away from the central cloud, the more responsible it is for the variance between two data sets. In contrast to the largely unsupervised data evaluation by PCA, the Student`s *t*-test is a

hypothesis driven statistical tool. The *t*-test also requires preprocessing of the data sets into a bucket format, from which it computes a sample inventory listing *p*-values, maximum intensities, average ratios and value counts for each bucket within a sample set. This inventory table is extremely useful since a biosynthetic pathway knockout mutant will exhibit a value count of zero for the bucket corresponding to the respective SM whereas the wild type will exhibit a value count equal to the number of replicate samples analyzed. So far both methods still produce quite a number of false positive hits and therefore require another round of manual verification of the results. However, they greatly reduce the number of molecular features to check, from up to 5000 in the original data set to usually around 30-50 after statistical analysis.

This comprehensive secondary metabolomic approach proved to be a very powerful tool for the identification of non-ubiquitous SM candidate compounds.[180;181] Other approaches for genomics driven secondary metabolite mining described in the literature have severe drawbacks and cannot compete with the output of the approach described above.[182;183] Chapter 7 of this thesis describes how this approach was adapted for the SM correlation of a so far unassigned biosynthetic gene cluster from *Stigmatella aurantiaca* Sg a15, identified from *in silico* analysis of the draft genome sequence of Sg a15. The SM was identified as rhizopodin, which was previously isolated from *Myxococcus stipitatus* Mx s11[184] and whose biosynthetic gene cluster was unknown so far. Neither was *S. aurantiaca* Sg a15 previously known as producer of rhizopodin.

Rhizopodin is produced by a *trans*-AT PKS/NRPS assembly line. Currently 23 SMs are reported to be produced by such special variants of PKS that use freestanding instead of integral AT domains.[185] The *trans*-AT PKSs show a high degree of divergence from textbook colinearity rules established for *cis*-AT systems. Modular architectures with missing or superfluous domains or unusual domain orders are common in *trans*-AT PKSs, thus complicating correlation of PKS function to product structure.[186-188] Recently, an innovative approach based on phylogenetic clustering of KS sequences in relationship to their substrate specificity was developed and allowed prediction of core structures from *trans*-AT systems with a significantly improved reliability.[84;189] Phylogenetic analysis of the 19 KS-domains in the rhizopodin (*riz*) gene cluster with a large subset of KS sequences used in the initial publication revealed good agreement between substrate prediction based on phylogenetic clading and the postulated biosynthetic intermediates. Besides phylogenetic analysis of KS domains, prediction of product stereochemistry by each KR domain according to the

conserved motifs [76;190;191] was in full agreement with the actual rhizopodin structure. The biosynthetic logic for the introduction of methoxy groups in rhizopodin is outstanding among *trans*-AT systems. Eleven (myxovirescin, pederin, onnamide, bryostatin, disorazol, rhizoxin, chivosazol, oxazolomycin, etnangien, kirromycin and thailandamide) out of the 23 reported core structures produced by *trans*-AT PKS systems, incorporate one or up to three methoxy groups. For all of these SMs, except for etnangien and thailandamide, the methoxy groups are introduced into the molecules by free-standing O-methyltransferases (MTs) in post-assembly line tailoring reactions. Rhizopodin biosynthesis deviates from this logic by the employment of non-extending modules of KS^0-(O-MT)-ACP architecture, which are solely dedicated to methylation of the β-hydroxyl group (Riz KS3, KS7, KS14, and KS18). Interestingly, the respective downstream KS domains (Riz KS4, KS8, KS15, and KS19), that will accept β-methoxylated intermediates, form a new subclade in the *trans*-AT KS phylogenetic tree. The benefit of using embedded O-MTs in rhizopodin biosynthesis consists in occurrence of only one final product, which is uniformly methylated at all eight relevant positions. Thereby the production of partially desmethylated derivatives, a case observed for many SMs featuring methoxy groups that are introduced by freestanding MTs, is avoided. In accordance with this theory, no partially desmethylated rhizopodin derivatives could be detected in extracts of Sg a15.

In every reported case where SMs are built up by *trans*-AT PKS/NRPS hybrid systems and contain oxazoline/oxazole or thiazoline/thiazole rings the heterocycles are introduced by tandem HC domains, a so far in the literature overlooked circumstance. Examples include leinamycin,[192] disorazol,[193] rhizoxin,[194] chivosazol,[195] virginamycin,[196] and rhizopodin. In NRPS or *cis*-AT PKS-NRPS systems generally a single HC domain is involved in heterocyclization of a serine or cysteine residue tethered on a PCP such as in tubulysin,[121] epothilone,[197] and ajudazol biosynthesis.[198] However, two rare incidences of tandem HC-domains exist from the biosyntheses of the siderophores vibriobactin and anguibactin. [199;200] Biochemical characterization *in vitro* of VibF revealed that HC2 is responsible for the condensation reaction, whereas HC1 catalyzes the heterocycle formation.[201] Phylogenetic analysis revealed formation of distinct subclades for HC1 and HC2 sequences from *trans*-AT PKS/NRPS hybrid systems, both separated from single HCs and the HC1 and HC2 sequences related to siderophore biosynthesis. Due to the observed phylogenetic distance between the

trans-AT PKS/NRPS tandem HCs and the tandem HCs from vibriobactin and anguibactin biosynthesis it seems likely that they resulted from convergent evolution.

The dimeric nature of rhizopodin is a very rare feature among bacterial secondary metabolites, including gramicidin S, enterobactin (trimer), thiocoraline (NRPS),[202] disorazol (NRPS-PKS-hybrid), the marinomycins and vermiculine (PKS; biosyntheses not yet elucidated).[203;204] TE-mediated formation of the dimeric (or trimeric) final product has been demonstrated *in vitro* for disorazol (Buntin K., unpublished results), gramicidin,[205] enterobactin,[206] and thiocoraline.[207] In analogy to these biosyntheses, the following model can be proposed for rhizopodin: The first monomer is transferred to the TE active site serine followed by nucleophilic attack of this acyl-O-TE intermediate by the hydroxyl group at C-18 of the second monomer which is tethered to the upstream ACP, yielding a linear dimeric intermediate. The linear dimer would then be transferred to the TE active site, where intramolecular cyclization and product release by formation of a dilactone are catalyzed. Studies on TEs involved in the biosyntheses of gramicidin S and enterobactin showed that there are no obvious sequence differences accounting for their divergent catalytic capabilities, but it was observed that the linker region between the ACP and the TE was longer than usual in the case of GrsB (41 aa compared to ~15-20 aa).[205] Similarly, extended linkers are conserved in the rhizopodin and disorazol assembly lines (36 aa in RizE/~60 aa in DszC). Furthermore, in both cases a non-elongating KS^0 domain is found upstream of the terminal ACP which may also play a role in the dimerization process.

The rhizopodin assembly line includes several special features, such as an unusual biosynthetic logic for the introduction of methoxy groups and the presence of members of a new KS phylogenetic subclade, both not reported to date from other *trans*-AT PKSs. The provided data will help to expand the knowledge about *trans*-AT PKSs which is still far less comprehensive than that compiled for *cis*-AT systems over many years.

D. References

[1] A. P. Klockgether-Radke, *Anasthesiol.Intensivmed.Notfallmed.Schmerzther.* **2002,** *37*, 244-249.

[2] D. J. Newman, G. M. Cragg, *J.Nat.Prod.* **2007,** *70*, 461-477.

[3] H. W. Florey, *Br.Med.J.* **1944,** *2,* 169-171.

[4] C. M. Dobson, *Nature* **2004,** *432*, 824-828.

[5] R. I. Sadreyev, N. V. Grishin, *BMC Structural Biology* **2006,** *6* .

[6] F. E. Koehn, G. T. Carter, *Nat.Rev.Drug Discovery* **2005,** *4*, 206-220.

[7] D. Kuhn, N. Weskamp, S. Schmitt, E. Hullermeier, G. Klebe, *Journal of Molecular Biology* **2006,** *359*, 1023-1044.

[8] C. A. Lipinski, *J.Pharmacol.Toxicol.Methods* **2000,** *44*, 235-249.

[9] K. J. Weissman, P. F. Leadlay, *Nat.Rev.Microbiol.* **2005,** *3,* 925-936.

[10] Pharmaceutical Biotechnology. *Drug discovery and clinical applications*, (Eds.: O. Kayser, R. H. Müller) WILEY-VCH, Weinheim **2008**.

[11] J. Kennedy, J. R. Marchesi, A. D. W. Dobson, *Microbial Cell Factories* **2008,** *7* .

[12] R. O. Garcia, D. Krug, R. Müller, in *Methods in Enzymology* **2009**, p. pp. 59-91.

[13] T. Iizuka, Y. Jojima, R. Fudou, M. Tokura, A. Hiraishi, S. Yamanaka, *Syst.Appl.Microbiol*. **2003,** *26*, 189-196.

[14] T. Iizuka, Y. Jojima, R. Fudou, S. Yamanaka, *FEMS Microbiol.Lett*. **1998,** *169*, 317-322.

[15] H. Reichenbach, M. Dworkin, in *The Procaryotes* Eds.: A. Balows, H. G. Trüper, M. Dworkin, W. Harder, K. H. Schleifer), Springer-Verlag, New York **1992**, p. pp. 3416-3487.

[16] H. Reichenbach, in *Bergey's manual of systematic bacteriology* Eds.: D. J. Brenner, N. R. Krieg, J. T. Staley), Springer, **2005**, p. pp. 1059-1144.

[17] W. Dawid, *FEMS Microbiol.Rev.* **2000,** *24*, 403-427.

[18] K. Gerth, S. Pradella, O. Perlova, S. Beyer, R. Müller, *J.Biotechnol.* **2003,** *106*, 233-253.

[19] S. Schneiker, O. Perlova, O. Kaiser, K. Gerth, A. Alici, M. O. Altmeyer, D. Bartels, T. Bekel, S. Beyer, E. Bode, H. B. Bode, C. J. Bolten, J. V. Choudhuri, S. Doss, Y. A. Elnakady, B. Frank, L. Gaigalat, A. Goesmann, C. Groeger, F. Gross, L. Jelsbak, L. Jelsbak, J. Kalinowski, C. Kegler, T. Knauber, S. Konietzny, M. Kopp, L. Krause, D. Krug, B. Linke, T. Mahmud, R. Martinez-Arias, A. C. McHardy, M. Merai, F. Meyer, S. Mormann, J. Munoz-Dorado, J. Perez, S. Pradella, S. Rachid, G. Raddatz, F. Rosenau, C. Rückert, F. Sasse, M. Scharfe, S. C. Schuster, G. Suen, A. Treuner-Lange, G. J. Velicer, F. J. Vorhölter, K. J. Weissman, R. D. Welch, S. C. Wenzel, D. E. Whitworth, S. Wilhelm, C. Wittmann, H. Blöcker, A. Pühler, R. Müller, *Nat.Biotechnol.* **2007,** *25*, 1281-1289.

[20] B. S. Goldman, W. C. Nierman, D. Kaiser, S. C. Slater, A. S. Durkin, J. Eisen, C. M. Ronning, W. B. Barbazuk, M. Blanchard, C. Field, C. Halling, G. Hinkle, O. Iartchuk, H. S. Kim, C. Mackenzie, R. Madupu, N. Miller, A. Shvartsbeyn, S. A. Sullivan, M. Vaudin, R. Wiegand, H. B. Kaplan, *P.Natl.Acad.Sci.USA* **2006,** *103*, 15200-15205.

[21] S. D. Bentley, K. F. Chater, A. M. Cerdeno-Tarraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C. H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill, D. A. Hopwood, *Nature* **2002,** *417*, 141-147.

[22] H. Ikeda, J. Ishikawa, A. Hanamoto, M. Shinose, H. Kikuchi, T. Shiba, Y. Sakaki, M. Hattori, S. Omura, *Nat.Biotechnol.* **2003,** *21*, 526-531.

[23] M. Oliynyk, M. Samborskyy, J. B. Lester, T. Mironenko, N. Scott, S. Dickens, S. F. Haydock, P. F. Leadlay, *Nat.Biotechnol.* **2007,** *25*, 447-453.

[24] H. Reichenbach, *J.Ind.Microbiol.Biotechnol.* **2001,** *27*, 149-156.

[25] H. B. Bode, R. Müller, *J.Ind.Microbiol.Biotechnol.* **2006,** *33*, 577-588.

[26] K. J. Weissman, R. Müller, *Nat.Prod.Rep.* **2010,** *27*, 1276-1295.

[27] G. Höfle, N. Bedorf, H. Steinmetz, D. Schomburg, K. Gerth, H. Reichenbach, *Angew.Chem.Int.Ed*. **1996,** *35*, 1567-1569.
[28] D. M. Bollag, P. A. McQueney, J. Zhu, O. Hensens, L. Koupal, J. Liesch, M. Goetz, E. Lazarides, M. Woods, *Cancer Res.* **1995,** *55*, 2325-2333.
[29] P. Fumoleau, B. Coudert, N. Isambert, E. Ferrant, *Ann.Oncol.* **2007,** *18*, v9-v15.
[30] M. N. Fornier, Clinical Breast *Cancer* **2007,** *7*, 757-763.
[31] H. Reichenbach, G. Höfle, *Drugs in R&D* **2008,** *9*, 1-10.
[32] M. W. Khalil, F. Sasse, H. Lünsdorf, Y. A. Elnakady, H. Reichenbach, *ChemBioChem* **2006,** *7*, 678-683
[33] Y. A. Elnakady, F. Sasse, H. Lünsdorf, H. Reichenbach, *Biochem.Pharmacol.* **2004,** *67*, 927-935.
[34] U. Eggert, R. Diestel, F. Sasse, R. Jansen, B. Kunze, M. Kalesse, *Angew. Chem. Int. Ed.* **2008,** *47*, 6478-6482.
[35] G. Hagelueken, S. C. Albrecht, H. Steinmetz, R. Jansen, D. W. Heinz, M. Kalesse, W. D. Schubert, *Angew. Chem. Int. Ed.* **2009,** *48*, 595-598.
[36] R. Diestel, H. Irschik, R. Jansen, M. W. Khalil, H. Reichenbach, F. Sasse, *Chembiochem*. **2009,** *10,* 2900-2903.
[37] I. Nickeleit, S. Zender, F. Sasse, R. Geffers, G. Brandes, I. Sörensen, H. Steinmetz, S. Kubicka, T. Carlomagno, D. Menche, I. Gütgemann, J. Buer, A. Gossler, M. P. Manns, M. Kalesse, R. Frank, N. P. Malek, *Cancer Res.* **2008,** *14*, 23-35.
[38] L. Pridzun, F. Sasse, H. Reichenbach, in *Antifungal agents* Eds.: G. K. Dixon, C. L.G., D. W. Hollomon), BIOS Scientific Publishers Ltd, Oxford, UK **1995**, p. pp. 99-109.
[39] H. F. Vahlensieck, L. Pridzun, H. Reichenbach, A. Hinnen, *Curr.Genet.* **1994,** *25*, 95-100.
[40] S. C. Weatherly, S. L. Volrath, T. D. Elich, *Biochem.J.* **2004,** *380*, 105-110.
[41] H. Irschik, R. Jansen, K. Gerth, G. Höfle, H. Reichenbach, *J.Antibiot.* **1987,** *40*, 7-13.
[42] H. Irschik, R. Jansen, G. Höfle, K. Gerth, H. Reichenbach, *J.Antibiot.* **1985,** *38*, 145-152.
[43] H. Irschik, H. Augustiniak, K. Gerth, G. Höfle, H. Reichenbach, *J.Antibiot.* **1995,** *48*, 787-792.
[44] A. O'Neill, B. Oliva, C. Storey, A. Hoyle, C. Fishwick, I. Chopra, *Antimicrobial Agents Chemotherapy* **2000,** *44*, 3163-3166.
[45] B. Meunier, S. A. Madgwick, E. Reil, W. Oettmeier, P. R. Rich, *Biochemistry-US* **1995,** *34*, 1076-1083.
[46] T. Friedrich, P. van Heek, H. Leif, T. Ohnishi, E. Forche, B. Kunze, R. Jansen, W. Trowitzsch-Kienast, G. Hofle, H. Reichenbach, et al, *Eur.J.Biochem.* **1994,** *219*, 691-698.
[47] G. Thierbach, H. Reichenbach, *Biochim.Biophys.Acta* **1981,** *638*, 282-289.
[48] M. Ritter, H. Palsdottir, M. Abe, W. Mäntele, C. Hunte, H. Miyoshi, P. Hellwig, *Biochemistry-US* **2004,** *43,* 8439-8446.
[49] M. A. Marahiel, *FEBS Lett.* **1992,** *307*, 40-43.
[50] S. Donadio, M. J. Staver, J. B. McAlpine, S. J. Swanson, L. Katz, *Science* **1991,** *252*, 675-679.
[51] S. C. Wenzel, B. Kunze, G. Höfle, B. Silakowski, M. Scharfe, H. Blöcker, R. Müller, *ChemBioChem* **2005,** *6*, 375-385.
[52] S. Gaisser, A. Trefzer, S. Stockert, A. Kirschning, A. Bechthold, *J.Bacteriol.* **1997,** *179*, 6271-6278.
[53] C. T. Walsh, *Science* **2004,** *303*, 1805-1810.
[54] D. E. Cane, C. T. Walsh, *Chem.Biol.* **1999,** *6*, R319-R325.
[55] R. H. Lambalot, A. M. Gehring, R. S. Flugel, P. Zuber, M. LaCelle, M. A. Marahiel, R. Reid, C. Khosla, C. T. Walsh, *Chem.Biol.* **1996,** *3*, 923-936.
[56] H. D. Mootz, R. Finking, M. A. Marahiel, *J.Biol.Chem.* **2001,** *276*, 37289-37298.
[57] R. M. Kohli, C. T. Walsh, *Chem.Commun.* **2003,** 297-307.
[58] M. A. Marahiel, T. Stachelhaus, H. D. Mootz, *Chem Rev* **1997,** *97*, 2651-2674.
[59] R. Finking, M. A. Marahiel, Annual *Review of Microbiology* **2004,** *58*, 453-488.
[60] E. Conti, T. Stachelhaus, M. A. Marahiel, P. Brick, *EMBO J.* **1997,** *16*, 4174-4183.
[61] T. Stachelhaus, H. D. Mootz, M. A. Marahiel, *Chem.Biol.* **1999,** *6*, 493-505.
[62] G. L. Challis, J. Ravel, C. A. Townsend, *Chem.Biol.* **2000,** *7*, 211-224.
[63] A. M. Gulick, *ACS Chem Biol* **2009,** *4*, 811-827.
[64] T. Stachelhaus, H. D. Mootz, V. Bergendahl, M. A. Marahiel, *J.Biol.Chem.* **1998,** *273*, 22773-22781.

[65] C. T. Walsh, H. W. Chen, T. A. Keating, B. K. Hubbard, H. C. Losey, L. S. Luo, C. G. Marshall, D. A. Miller, H. M. Patel, *Curr.Opin.Chem.Biol.* **2001,** *5*, 525-534.
[66] M. A. Fischbach, C. T. Walsh, *Chem.Rev.* **2006,** *106*, 3468-3496.
[67] L. Luo, R. M. Kohli, M. Onishi, U. Linne, M. A. Marahiel, C. T. Walsh, *Biochemistry-US* **2002,** *41*, 9184-9196.
[68] W. L. Kelly, N. J. Hillson, C. T. Walsh, *Biochemistry-US* **2005,** *44*, 13385-13393.
[69] L. Du, M. Chen, C. Sanchez, B. Shen, *FEMS Microbiol.Lett.* **2000,** *189*, 171-175.
[70] S. F. Haydock, J. F. Aparicio, I. Molnar, T. Schwecke, A. König, A. F. A. Marsden, I. S. Galloway, J. Staunton, P. F. Leadley, *FEBS Lett.* **1995,** *374*, 246-248.
[71] G. Yadav, R. S. Gokhale, D. Mohanty, *Journal of Molecular Biology* **2003,** *328*, 335-363.
[72] Y. A. Chan, M. T. Boyne, A. M. Podevels, A. K. Klimowicz, J. Handelsman, N. L. Kelleher, M. G. Thomas, *P.Natl.Acad.Sci.USA* **2006,** *103*, 14349-14354.
[73] Y. A. Chan, M. G. Thomas, *Biochemistry-US* **2010,** *49,* 3667-3677.
[74] S. C. Wenzel, R. M. Williamson, C. Grünanger, J. Xu, K. Gerth, R. A. Martinez, S. J. Moss, B. J. Carroll, S. Grond, C. J. Unkefer, R. Müller, H. G. Floss, *J.Am.Chem.Soc.* **2006,** *128*, 14325-14336.
[75] K. J. Weissman, M. Timoney, M. Bycroft, P. Grice, U. Hanefeld, J. Staunton, P. F. Leadlay, *Biochemistry-US* **1997,** *36*, 13849-13855.
[76] R. Reid, M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V. Santi, C. R. Hutchinson, R. McDaniel, *Biochemistry-US* **2003,** *42*, 72-79.
[77] J. Staunton, K. J. Weissman, *Nat.Prod.Rep.* **2001,** *18*, 380-416.
[78] C. Bisang, P. F. Long, J. Cortés, J. Westcott, J. Crosby, A. L. Matharu, R. J. Cox, T. J. Simpson, J. Staunton, P. F. Leadlay, *Nature* **1999,** *401*, 502-505.
[79] J. Ligon, S. Hill, J. Beck, R. Zirkle, I. Molnar, J. Zawodny, S. Money, T. Schupp, *Gene* **2002,** *285*, 257-267.
[80] H. Ikeda, T. Nonomiya, S. Omura, *J.Ind.Microbiol.Biotechnol.* **2001,** *27*, 170-176.
[81] B. Silakowski, G. Nordsiek, B. Kunze, H. Blöcker, R. Müller, *Chemistry & Biology* **2001,** *8*, 59-69.
[82] J. Piel, *P.Natl.Acad.Sci.USA* **2002,** *99*, 14002-14007.
[83] D. E. Cane, C. T. Walsh, C. Khosla, *Science* **1998,** *282*, 63-68.
[84] T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, J. Piel, *Nat.Biotechnol.* **2008,** *26*, 225-233.
[85] B. Silakowski, B. Kunze, R. Müller, *Gene* **2001,** *275*, 233-240.
[86] A. Sandmann, J. Dickschat, H. Jenke-Kodama, B. Kunze, E. Dittmann, R. Müller, *Angew.Chem.Int.Ed.* **2007,** *46*, 2712-2716.
[87] A. O. Brachmann, S. A. Joyce, H. Jenke-Kodama, G. Schwär, D. J. Clarke, H. B. Bode, *ChemBioChem* **2007,** *8*, 1721-1728.
[88] F. Malpartida, D. A. Hopwood, *Nature* **1984,** *309* 462.
[89] B. Shen, *Top.Curr.Chem.* **2000,** *209*, 1-51.
[90] C. Hertweck, A. Luzhetskyy, Y. Rebets, A. Bechthold, *Nat.Prod.Rep.* **2007,** *24*, 162-190.
[91] R. McDaniel, S. Ebert-Khosla, D. A. Hopwood, C. Khosla, *Science* **1993,** *262*, 1546-1550.
[92] A. T. Keatinge-Clay, D. A. Maltby, K. F. Medzihradszky, C. Khosla, R. M. Stroud, *Nat Struct Mol Biol* **2004,** *11*, 888-893.
[93] M. B. Austin, J. P. Noel, *Nat.Prod.Rep.* **2003,** *20*, 79-110.
[94] B. S. Moore, J. N. Hopke, *ChemBioChem* **2001,** *2*, 35-38.
[95] Y. Li, R. Müller, Phytochemistry *Phytochemistry* **2009,** *70*, 1850-1857.
[96] P. M. Dewick, *Medicinal Natural Products*, 3th ed. Wiley, **2009,** p. 507.
[97] M. Mentel, E. G. Ahuja, D. V. Mavrodi, R. Breinbauer, L. S. Thomashow, W. Blankenfeldt, *ChemBioChem* **2009,** *10,* 2295-2304.
[98] P. M. Dewick, *Nat.Prod.Rep.* **2002,** *19*, 181-222.
[99] S. Haeder, R. Wirth, H. Herz, D. Spiteller, *Proc.Natl.Acad.Sci.U.S A* **2009,** *106*, 4742-4746.
[100] B. Silakowski, B. Kunze, G. Nordsiek, H. Blöcker, G. Höfle, R. Müller, *Eur.J.Biochem.* **2000,** *267,* 6476-6485.
[101] L. P. Partida-Martinez, C. Hertweck, *Nature* **2005,** *437*, 884-888.

[102] M. B. Miller, B. L. Bassler, *Annual Review of Microbiology* **2001,** *55*, 165-199.
[103] J. S. Dickschat, *Nat.Prod.Rep.* **2010,** *27*, 343-369.
[104] E. C. Pesci, J. B. Milbank, J. P. Pearson, S. McKnight, A. S. Kende, E. P. Greenberg, B. H. Iglewski, *P.Natl.Acad.Sci.USA* **1999,** *96*, 11229-11234.
[105] L. A. Gallagher, S. L. McKnight, M. S. Kuznetsova, E. C. Pesci, C. Manoil, *J.Bacteriol.* **2002,** *184*, 6472-6480.
[106] C. Winstanley, J. L. Fothergill, *FEMS Microbiol.Lett.* **2009,** *290*, 1-9.
[107] M. Juhas, L. Eberl, B. Tummler, *Environ.Microbiol.* **2005,** *7*, 459-471.
[108] J. Sambrook, D. W. Russell, Molecular cloning: A laboratory manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY **2001**.
[109] S. Beyer, B. Kunze, B. Silakowski, R. Müller, *Biochim.Biophys.Acta* **1999,** *1445*, 185-195.
[110] N. Gaitatzis, A. Hans, R. Müller, S. Beyer, *J.Biochem.(Tokyo)* **2001,** *129*, 119-124.
[111] F. Sasse, H. Steinmetz, J. Heil, G. Höfle, H. Reichenbach, *J.Antibiot.* **2000,** *53*, 879-885.
[112] H. Steinmetz, N. Glaser, E. Herdtweck, F. Sasse, H. Reichenbach, G. Höfle, *Angew.Chem.Int.Ed.* **2004,** *43*, 4888-4892.
[113] R. Balasubramanian, B. Raghavan, A. Begaye, D. L. Sackett, R. A. Fecik, *J.Med.Chem* **2009,** *52*, 238-240.
[114] R. Balasubramanian, B. Raghavan, J. C. Steele, D. L. Sackett, R. A. Fecik, *Bioorg Med Chem Lett* **2008,** *18*, 2996-2999.
[115] B. Raghavan, R. Balasubramanian, J. C. Steele, D. L. Sackett, R. A. Fecik, *J.Med.Chem.* **2008,** *51*, 1530-1533.
[116] A. W. Patterson, H. M. Peltier, F. Sasse, J. A. Ellman, *Chemistry* **2007,** *13*, 9534-9541.
[117] Z. Y. Wang, P. A. McPherson, B. S. Raccor, R. Balachandran, G. Y. Zhu, B. W. Day, A. Vogt, P. Wipf, *Chemical Biology & Drug Design* **2007,** *70*, 75-86.
[118] A. Dömling, B. Beck, U. Eichelberger, S. Sakamuri, S. Menon, Q. Z. Chen, Y. Lu, L. A. Wessjohann, *Angew.Chem.Int.Ed.* **2006,** *45*, 7235-7239.
[119] K. Kubicek, S. K. Grimm, J. Orts, F. Sasse, T. Carlomagno, *Angew.Chem.Int.Ed.* **2010,** *49*, 4809-4812.
[120] J. L. Burkhart, R. Müller, U. Kazmaier, *Eur.J.Org.Chem.* **2011,** 2011 3050-3059.
[121] A. Sandmann, F. Sasse, R. Müller, *Chem.Biol.* **2004,** *11*, 1071-1079.
[122] J. D. Kittendorf, B. J. Beck, T. J. Buchholz, W. Seufert, D. H. Sherman, *Chemistry & Biology* **2007,** *14*, 944-954.
[123] R. D. Firn, C. G. Jones, *Nat.Prod.Rep.* **2003,** *20*, 382-391.
[124] K. J. Weissman, R. Müller, *ChemBioChem* **2008,** *9*, 826-848.
[125] J. Mulzer, K. H. Altmann, G. Hofle, R. Mueller, K. Prantz, *Comptes Rendus Chimie* **2008,** *11*, 1336-1368.
[126] A. Moraleda-Munoz, L. J. Shimkets, *J.Bacteriol.* **2007,** *189*, 3072-3080.
[127] P. Meiser, K. J. Weissman, H. B. Bode, D. Krug, J. S. Dickschat, A. Sandmann, R. Müller, *Chem.Biol.* **2008,** *15*, 771-781.
[128] M. Kopp, H. Irschik, K. Gemperlein, K. Buntin, P. Meiser, K. J. Weissman, H. B. Bode, R. Müller, *MolBioSyst* **2011,** *7,* 1549-1563.
[129] C. P. Leamon, J. A. Reddy, M. Vetzel, R. Dorton, E. Westrick, N. Parker, Y. Wang, I. Vlahov, *Cancer Res.* **2008,** *68*, 9839-9844.
[130] J. A. Reddy, R. Dorton, A. Dawson, M. Vetzel, N. Parker, J. S. Nicoson, E. Westrick, P. J. Klein, Y. Wang, I. R. Vlahov, C. P. Leamon, *Molecular Pharmaceutics* **2009,** *6*, 1518-1525.
[131] T. Schluep, P. Gunawan, L. Ma, G. S. Jensen, J. Duringer, S. Hinton, W. Richter, J. Hwang, *Clin.Cancer Res.* **2009,** *15*, 181-189.
[132] W. C. Floyd, G. K. Datta, S. Imamura, H. M. Kieler-Ferguson, K. Jerger, A. W. Patterson, M. E. Fox, F. C. Szoka, J. M. J. Frechet, J. A. Ellman, *ChemMedChem* **2011,** *6*, 49-53.
[133] B. Kunze, G. Höfle, H. Reichenbach, *J.Antibiot.* **1987,** *40*, 258-265.
[134] G. Höfle, B. Böhlendorf, T. Fecker, F. Sasse, B. Kunze, *J.Nat.Prod.* **2008,** *71,* 1967-1969.
[135] E. Stec, D. Pistorius, R. Müller, S. M. Li, *ChemBioChem* **2011,** *12,* 1724-1730.
[136] G. Höfle, B. Kunze, *J.Nat.Prod.* **2008,** *71*, 1843-1849.

[137] M. Izumikawa, Q. Cheng, B. S. Moore, *J.Am.Chem.Soc.* **2006,** *128*, 1428-1429.
[138] A. T. Keatinge-Clay, A. A. Shelat, D. F. Savage, S. C. Tsai, L. J. W. Miercke, J. D. O'Connell, C. Khosla, R. M. Stroud, *Structure* **2003,** *11*, 147-154.
[139] S. Kroken, N. L. Glass, J. W. Taylor, O. C. Yoder, B. G. Turgeon, *P.Natl.Acad.Sci.USA* **2003,** *100*, 15670-15675.
[140] H. Jenke-Kodama, A. Sandmann, R. Müller, E. Dittmann, *Mol.Biol.Evol.* **2005,** *22*, 2027-2039.
[141] C. O. Rock, S. Jackowski, J. E. Cronan, Jr., in *Biochemistry of lipids* Eds.: D. E. Vance, J. E. Vance), Elsevier Sciences, **1996**, p. pp. 35-74.
[142] S. W. White, J. Zheng, Y. M. Zhang, Rock, *Annu Rev Biochem* **2005,** *74*, 791-831.
[143] K. Jakobi, C. Hertweck, *J.Am.Chem.Soc.* **2004,** *126*, 2298-2299.
[144] J. Dreier, C. Khosla, *Biochemistry-US* **2000,** *39*, 2088-2095.
[145] B. Bagautdinov, Y. Ukita, M. Miyano, N. Kunishima, *Acta Crystallogr Sect F Struct Biol Cryst Commun* **2008,** 64 358-366.
[146] C. P. Ridley, H. Y. Lee, C. Khosla, *P.Natl.Acad.Sci.USA* **2008,** *105*, 4595-4600.
[147] R. McDaniel, K. Ebert, D. A. Hopwood, C. Khosla, *Nature* **1995,** *375*, 549.
[148] B. Julien, Z. Q. Tian, R. Reid, C. D. Reeves, *Chem.Biol.* **2006,** *13*, 1277-1286.
[149] J. Lee, M. Simurdiak, H. Zhao, *J.Biol.Chem.* **2005,** *280*, 36719-36727.
[150] S. Mo, P. K. Sydor, C. Corre, M. M. Alhamadsheh, A. E. Stanley, S. W. Haynes, L. Song, K. A. Reynolds, G. L. Challis, *Chemistry & Biology* **2008,** *15*, 137-148.
[151] P. K. Sydor, S. M. Barry, O. M. Odulate, F. Barona-Gomez, S. W. Haynes, C. Corre, L. J. Song, G. L. Challis, *Nature Chemistry* **2011,** *3*, 388-392.
[152] A. R. Gallimore, C. B. Stark, A. Bhatt, B. M. Harvey, Y. Demydchuk, V. Bolanos-Garcia, D. J. Fowler, J. Staunton, P. F. Leadlay, J. B. Spencer, *Chem.Biol.* **2006,** *13*, 453-460.
[153] B. Silakowski, B. Kunze, R. Müller, *Arch.Microbiol.* **2000,** *173*, 403-411.
[154] T. Knöchel, A. Ivens, G. Hester, A. Gonzalez, R. Bauerle, M. Wilmanns, K. Kirschner, J. N. Jansonius, *P.Natl.Acad.Sci.USA* **1999,** *96*, 9479-9484.
[155] H. W. Boucher, G. H. Talbot, J. S. Bradley, J. E. Edwards, D. Gilbert, L. B. Rice, M. Scheld, B. Spellberg, J. Bartlett, *Clin.Infect.Dis.* **2009,** *48*, 1-12.
[156] M. Allesen-Holm, K. B. Barken, L. Yang, M. Klausen, J. S. Webb, S. Kjelleberg, S. Molin, M. Givskov, T. Tolker-Nielsen, *Mol.Microbiol.* **2006,** *59*, 1114-1128.
[157] J. F. Dubern, S. P. Diggle, *Mol.Biosyst.* **2008,** *4*, 882-888.
[158] E. Deziel, S. Gopalan, A. P. Tampakaki, F. Lepine, K. E. Padfield, M. Saucier, G. P. Xiao, L. G. Rahme, *Mol.Microbiol.* **2005,** *55*, 998-1014.
[159] D. S. Wade, M. W. Calfee, E. R. Rocha, E. A. Ling, E. Engstrom, J. P. Coleman, E. C. Pesci, *J.Bacteriol.* **2005,** *187*, 4372-4380.
[160] E. C. Pesci, J. P. Pearson, P. C. Seed, B. H. Iglewski, *J.Bacteriol.* **1997,** *179*, 3127-3132.
[161] G. Rampioni, C. Pustelny, M. P. Fletcher, V. J. Wright, M. Bruce, K. P. Rumbaugh, S. Heeb, M. Camara, P. Williams, *Environ.Microbiol.* **2010,** *12*, 1659-1673.
[162] D. G. Davies, M. R. Parsek, J. P. Pearson, B. H. Iglewski, J. W. Costerton, E. P. Greenberg, *Science* **1998,** *280*, 295-8.
[163] S. Raina, D. De Vizio, M. Odell, M. Clements, S. Vanhulle, T. Keshavarz, *Biotechnol.Appl.Biochem.* **2009,** *54*, 65-84.
[164] H. O. Sintim, J. Al Smith, J. X. Wang, S. Nakayama, L. Yan, Future *Medicinal Chemistry* **2010,** *2*, 1005-1035.
[165] H. Cao, G. Krishnan, B. Goumnerov, J. Tsongalis, R. Tompkins, L. G. Rahme, *P.Natl.Acad.Sci.USA* **2001,** *98*, 14613-14618.
[166] R. S. Smith, B. H. Iglewski, *Curr.Opin.Microbiol.* **2003,** *6*, 56-60.
[167] P. K. Singh, A. L. Schaefer, M. R. Parsek, T. O. Moninger, M. J. Welsh, E. P. Greenberg, *Nature* **2000,** *407*, 762-764.
[168] D. L. Erickson, R. Endersby, A. Kirkham, K. Stuber, D. D. Vollman, H. R. Rabin, I. Mitchell, D. G. Storey, *Infect.Immun.* **2002,** *70,* 1783-1790.

[169] R. Le Berre, S. Nguyen, E. Nowak, E. Kipnis, M. Pierre, F. Ader, R. Courcol, B. P. Guery, K. Faure, *Clin Microbiol Infect* **2008,** *14*, 337-343.
[170] C. N. Wilder, S. P. Diggle, M. Schuster, *ISME J.* **2011,** *5*, 1332-1343.
[171] B. Lesic, F. Lepine, E. Deziel, J. Zhang, Q. Zhang, K. Padfield, M. H. Castonguay, S. Milot, S. Stachel, A. A. Tzika, R. G. Tompkins, L. G. Rahme, *PLoS Pathog.* **2007,** *3*, 1229-1239.
[172] J. P. Coleman, L. L. Hudson, S. L. McKnight, J. M. Farrow, M. W. Calfee, C. A. Lindsey, E. C. Pesci, *J.Bacteriol.* **2008,** *190*, 1247-1255.
[173] F. Bredenbruch, M. Nimtz, V. Wray, M. Morr, R. Müller, S. Haussler, *J.Bacteriol.* **2005,** *187*, 3630-3635.
[174] Y. M. Zhang, M. W. Frank, K. Zhu, A. Mayasundari, C. O. Rock, *J.Biol.Chem.* **2008,** *283*, 28788-28794.
[175] I. Abe, H. Morita, *Nat.Prod.Rep.* **2010,** *27*, 809-838.
[176] M. P. Fletcher, S. P. Diggle, M. Camara, P. Williams, *Nat.Protoc.* **2007,** *2*, 1254-1262.
[177] H. B. Bode, R. Müller, *Angew.Chem.Int.Ed.* **2005,** *44*, 6828-6846.
[178] H. B. Bode, M. W. Ring, G. Schwär, M. O. Altmeyer, C. Kegler, I. R. Jose, M. Singer, R. Müller, *ChemBioChem* **2009,** *10*, 128-140.
[179] C. Schley, M. O. Altmeyer, R. Swart, R. Müller, C. G. Huber, *J Proteome Res* **2006,** *5*, 2760-2768.
[180] D. Krug, G. Zurek, O. Revermann, M. Vos, G. J. Velicer, R. Müller, *Appl.Environ.Microbiol.* **2008,** *74*, 3058-3068.
[181] D. Krug, G. Zurek, B. Schneider, R. Garcia, R. Müller, *Anal.Chim.Acta* **2008,** *624*, 97-106.
[182] H. Gross, V. O. Stockwell, M. D. Henkels, B. Nowak-Thompson, J. E. Loper, W. H. Gerwick, *Chem.Biol.* **2007,** *14*, 53-63.
[183] S. B. Bumpus, B. S. Evans, P. M. Thomas, I. Ntai, N. L. Kelleher, *Nat.Biotechnol.* **2009,** *27,* 951-956.
[184] F. Sasse, H. Steinmetz, G. Höfle, H. Reichenbach, *J.Antibiot.* **1993,** *46*, 741-748.
[185] J. Piel, *Nat.Prod.Rep.* **2010,** *27*, 996-1047.
[186] V. Simunovic, J. Zapp, S. Rachid, D. Krug, P. Meiser, R. Müller, *ChemBioChem* **2006,** *7*, 1206-1220.
[187] D. Menche, F. Arikan, O. Perlova, N. Horstmann, W. Ahlbrecht, S. C. Wenzel, R. Jansen, H. Irschik, R. Müller, *J.Am.Chem.Soc.* **2008,** *130*, 14234-14243.
[188] G. L. Tang, Y. Q. Cheng, B. Shen, *Chem.Biol.* **2004,** *11*, 33-45.
[189] H. Irschik, M. Kopp, K. J. Weissman, K. Buntin, J. Piel, R. Müller, *ChemBioChem* **2010,** *11*, 1840-1849.
[190] P. Caffrey, *ChemBioChem* **2003,** *4*, 654-657.
[191] A. T. Keatinge-Clay, *Chem.Biol.* **2007,** *14*, 898-908.
[192] Y. Q. Cheng, G. L. Tang, B. Shen, *P.Natl.Acad.Sci.USA* **2003,** *100*, 3149-3154.
[193] M. Kopp, H. Irschik, S. Pradella, R. Müller, *ChemBioChem* **2005,** *6*, 1277-1286.
[194] L. P. Partida-Martinez, C. Hertweck, *ChemBioChem* **2007,** *8*, 41-45.
[195] O. Perlova, K. Gerth, A. Hans, O. Kaiser, R. Müller, *J.Biotechnol.* **2006,** *121*, 174-191.
[196] N. Pulsawat, S. Kitani, T. Nihira, *Gene* **2007,** *393*, 31-42.
[197] K. Gerth, N. Bedorf, G. Höfle, H. Irschik, H. Reichenbach, *J.Antibiot.* **1996,** *49*, 560-563.
[198] R. Jansen, B. Kunze, H. Reichenbach, G. Höfle, *Eur.J.Org.Chem.* **2002,** *2002,* 917-921.
[199] C. G. Marshall, M. D. Burkart, T. A. Keating, C. T. Walsh, *Biochemistry-US* **2001,** *40* 10655-10663.
[200] M. Di Lorenzo, M. Stork, H. Naka, M. E. Tolmasky, J. H. Crosa, *Biometals* **2008,** *21*, 635-648.
[201] C. G. Marshall, N. J. Hillson, C. T. Walsh, *Biochemistry-US* **2002,** *41*, 244-250.
[202] F. Lombo, A. Velasco, A. Castro, F. de la Calle, A. F. Brana, J. M. Sanchez-Puelles, C. Mendez, J. A. Salas, *ChemBioChem* **2006,** *7*, 366-376.
[203] H. C. Kwon, C. A. Kauffman, P. R. Jensen, W. Fenical, *J.Am.Chem.Soc.* **2006,** *128*, 1622-1632.
[204] R. K. Boeckman, J. Fayos, J. Clardy, *J.Am.Chem.Soc.* **1974,** *96*, 5954-5956.
[205] K. M. Hoyer, C. Mahlert, M. A. Marahiel, *Chem.Biol.* **2007,** *14*, 13-22.
[206] C. A. Shaw-Reid, N. L. Kelleher, H. C. Losey, A. M. Gehring, C. Berg, C. T. Walsh, *Chem.Biol.* **1999,** *6,* 385-400.
[207] L. Robbel, K. M. Hoyer, M. A. Marahiel, *FEBS J.* **2009,** *276*, 1641-1653.

E. Appendix

Author`s effort in publications presented in this work

Chapter 1: The author was involved in the feeding experiments and in the execution and analysis of the MS/MS fragmentation experiments.

Chapter 2: The author performed the purifications of pretubulysin and *N*-desmethyl 12-keto pretubulysin and their structural elucidation by NMR. In addition the author performed part of the feeding studies and was involved in the analysis of the MS/MS fragmentation experiments.

Chapter 4: The author performed the cloning, heterologous expression, and purification of the characterized proteins and generated the inactivation mutants in Sg a15. Further the author planned, executed, measured, and analyzed all of the *in vitro* experiments.

Chapter 5: The author generated and analyzed the inactivation mutants and was involved in the annotation of the genes coding for the biosynthetic functionalities and the elaboration of the biosynthetic model.

Chapter 6: The author performed the cloning, heterologous expression, and purification of PqsD and planned, executed, measured, and analyzed the *in vitro* experiments for the reconstitution of HHQ biosynthesis as well as the further development towards the *in vitro* inhibitor assay.

Chapter 7: The author screened the draft genome sequence of Sg a15 for unassigned PKS and NRPS clusters and generated and analyzed the inactivation mutants. In addition the author performed the PCA analysis, interpreted the results and postulated the biosynthetic model for rhizopodin.

Further publications

In addition to the publications printed in this thesis I contributed to the following publications:

Zöllner, A., Dragan, C.A., **Pistorius, D.**, Müller, R., Bode, H.B., Peters, F.T., Maurer, H.H., Bureik, M. (2009) Human CYP4Z1 catalyzes the in-chain hydroxylation of lauric acid and myristic acid. *Biol Chem.,* **390,** 313-317.

Khatri, Y., Hannemann, F., Ewen, K.M., **Pistorius, D.**, Perlova, O., Kagawa, N., Brachmann, A.O., Müller, R., Bernhardt, R. (2010) The CYPome of Sorangium cellulosum So ce56 and identification of CYP109D1 as a new fatty acid hydroxylase. *Chem Biol.,* **17**, 1295-1305.

Garcia, R., **Pistorius, D.**, Stadler, M., Müller, R. (2011) Fatty acid related phylogeny of myxobacteria as an approach to discover polyunsaturated omega-3/6 fatty acids. *J Bacteriol.,* **193,** 1930-1942.

Stec, E., **Pistorius, D.**, Müller, R., Li, S.M. (2011) AuaA, a membrane-bound farnesyltransferase from Stigmatella aurantiaca, catalyzes the prenylation of 2-methyl-4-hydroxyquinoline in the biosynthesis of aurachins. *ChemBioChem,* **12,** 1724-1730.

Conference contributions

Pistorius, D., September 2010. Oral presentation - Exploring quorum sensing in *Pseudomonas aeruginosa* as putative target for anti-infective compounds. International workshop: *Biology of bacteria producing natural products*, Tübingen, Germany

Pistorius, D., Li, Y., Sandmann, A., Müller, R. September 2008. Poster presentation - Investigation on biosynthesis of the quinolone core of the aurachins, a family of secondary metabolites from *Stigmatella aurantiaca* Sg a15. International workshop: *Biology and chemistry of antibiotic-producing bacteria*, Berlin, Germany

Pistorius, D., Müller, R. September 2010. Poster presentation - Discovery and *in silico* characterization of the rhizopodin biosynthetic gene cluster. International conference: *Biology of myxobacteria - Myxomeeting 2010*, Otzenhausen, Germany

Pistorius, D., Chai. Y., Hermann, J.E., Ullrich, A., Burkhart, J.L., Kazmaier, U., Müller, R. February 2011. Poster presentation - Tubulysins: special biosynthetic features and chemically synthesized derivatives. Symposium: *23. Irseer Naturstofftage – Current trends in natural product research*, Irsee, Germany